



République Algérienne Démocratique et
Populaire Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université ABOU BEKR BELKAID –TLEMCEM

Faculté des Sciences de la Nature et de la Vie, et Sciences de la Terre et de l'Univers

Département de Biologie



MÉMOIRE

Présenté par

ZIGH Sana

BENABDALLAH Ikram

En vue de l'obtention du

Diplôme de MASTER

En Immunologie

Thème

**Analyse *in silico* des SNPs à conséquences fonctionnelles sur TGF- β R2 et SMAD7 dans le cancer colorectal :
Prédiction des modifications post traductionnelles et modélisation moléculaire**

Soutenu le 26 juin, devant le jury composé de:

Président	Aribi Mourad	Professeur	Université de Tlemcen
Encadrante	Hadjidj Zeyneb	MCB	Université de Tlemcen
Examinatrice	El Mezouar Chahrazed	MCA	Université de Tlemcen

Année universitaire 2022/2023



République Algérienne Démocratique et
Populaire Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université ABOU BEKR BELKAID –TLEMCEM

Faculté des Sciences de la Nature et de la Vie, et Sciences de la Terre et de l'Univers

Département de Biologie



MÉMOIRE

Présenté par

ZIGH Sana

BENABDALLAH Ikram

En vue de l'obtention du

Diplôme de MASTER

En Immunologie

Thème

**Analyse *in silico* des SNPs à conséquences fonctionnelles sur TGF- β R2 et SMAD7 dans le cancer colorectal :
Prédiction des modifications post traductionnelles et modélisation moléculaire**

Soutenu le 26 juin, devant le jury composé de:

Président	Aribi Mourad	Professeur	Université de Tlemcen
Encadrante	Hadjidj Zeyneb	MCB	Université de Tlemcen
Examinatrice	El Mezouar Chahrazed	MCA	Université de Tlemcen

Année universitaire 2022/2023

Dédicace

J'ai le grand plaisir de dédier ce modeste travail :

A la lumière de ma vie, la source de tendresse, mon première supporteur et mon amour éternelle, ma mère que j'adore **Zoulikha**.

A mon très cher père **Tayeb**, pour ses encouragements, son soutien, son amour et son sacrifice, Qu'ALLAH le tout puissant te préserve, t'accorde Santé, bonheur et te protège de tout mal.

A mon frère **Adel**, et mes sœurs **Hanaa** et **Sara**, Tous les mots ne sauraient exprimer la gratitude, l'amour, la reconnaissance et le Respect que j'ai pour vous.

A mon soutien moral et source de joie et de bonheur, mon fiancé **Bilal**, qui a une place spéciale dans mon cœur, merci pour l'encouragement et l'aide que vous m'avez toujours accordé.

A toute ma famille et mes amis.

A mon binôme et mon bras droite **Ikram** pour tous les souvenirs pendant les années d'études ensemble.

Tout simplement, à tous ceux que j'aime et qui m'aiment.

ZIGH Sana

Dédicace

Je dédie ce mémoire

À mes chers parents qui ont été toujours près de moi et m'ont toujours soutenu
tout au long de ces longues années d'études.

EN signe de contestation, l'énonciation de ma profonde gratitude pour tout ce
qu'ils ont admettre d'effort et de moyens pour me voir réussir dans mes études.

A toute ma famille

A mes chères sœurs Raja et Manar

A mon cher frère Hicham

Et A toutes mes amies

A tous les gens qui me connaissent et que je connais

Et à tous ceux qui aiment le bon travail et ne suspendre pas devant les
complications de la vie.

Enfin, je remercie mon binôme Sanae qui a collaboré à la réalisation de ce
modeste travail

BENABDALLAH Ikram

Résumé

Résumé

Dans cette étude, nous avons réalisé des analyses *in silico* des Single Nucleotide Polymorphisms (SNPs) ayant des conséquences fonctionnelles sur les protéines TGFBR2 et SMAD7 dans le contexte du cancer colorectal. Nous avons réalisé des prédictions des modifications post-traductionnelles associées aux SNPs identifiés comme potentiellement délétères, ainsi que la modélisation moléculaire des structures tridimensionnelles des protéines pour une visualisation plus précise des altérations.

Nous avons exploité plusieurs outils bioinformatiques tels que GPS6.0, MusiteDeep, NetPhos 3-1, GPS-MSP, GPS-UBER, PRmePRed, GPS.PBS, CarSite-II, GPS-YNO2 et GPS-SNO 1.0 pour prédire les modifications post-traductionnelles. Ensuite, nous avons utilisé les logiciels Phyre2 et SWISS-MODEL pour réaliser la modélisation moléculaire de ces protéines mutées et ERRAT/PROCHECK pour évaluer la qualité de nos résultats.

Nos résultats indiquent que certaines variations dans la protéine TGFBR2 (R448G, R520L, R522Q et K277N) peuvent potentiellement induire diverses modifications post-traductionnelles telles que la méthylation et l'ubiquitination.

En conclusion notre étude *in silico* fournit une base préliminaire intéressante concernant l'impact possible de certains SNPs sur TGFBR2 dans le contexte du cancer colorectal. Ces résultats pourront être utilisés comme base de départ pour des études expérimentales visant à valider les effets de ces SNPs sur les protéines et leur rôle dans le développement de cette maladie.

Mots clés : cancer colorectal, modélisation des protéines, modifications post traductionnelles, nsSNP, SMAD7, TGFBR2.

Abstract

Abstract

In this study, we performed *in silico* analyses of Single Nucleotide Polymorphisms (SNPs) with functional consequences on TGFBR2 and SMAD7 proteins in the context of colorectal cancer. We predicted post-translational modifications associated with SNPs identified as potentially deleterious, as well as performing molecular modeling of the three-dimensional structures of proteins for more precise visualization of alterations.

We used several bioinformatics tools such as GPS6.0, MusiteDeep, NetPhos 3-1, GPS-MSP, GPS-UBER, PRmePRed, GPS.PBS, CarSite-II, GPS-YNO2 and GPS-SNO 1.0 to predict post-translational modifications. Then we used Phyre2 and SWISS-MODEL software to perform molecular modeling of these mutated proteins and ERRAT/PROCHECK to evaluate the quality of our results.

Our results indicate that certain variations in TGFBR2 protein (R448G, R520L, R522Q, and K277N) can potentially induce various post-translational modifications such as methylation and ubiquitination.

In conclusion our *in silico* study provides an interesting preliminary basis regarding the possible impact of some SNPs on TGFBR2 within the context of colorectal cancer. These results could be used as a starting point for experimental studies aiming to validate the effects of these SNPs on proteins and their role in disease development.

Keywords: colorectal cancer, protein modeling, post-translational modifications, nsSNP, SMAD7, TGFBR2.

ملخص

في هذه الدراسة ، أجرينا تحليلات سيليكو لتعدد أشكال النوكليوتيدات المفردة (SNPs) مع عواقب وظيفية على بروتينات TGFBR2 و SMAD7 في سياق سرطان القولون والمستقيم. لقد أجرينا تنبؤات بتعديلات ما بعد الترجمة المرتبطة بـ SNPs التي تم تحديدها على أنها ضارة ، بالإضافة إلى النمذجة الجزيئية للهياكل ثلاثية الأبعاد للبروتينات للحصول على تصور أكثر دقة للتغيرات.

استفدنا من العديد من أدوات المعلوماتية الحيوية مثلًا GPS6.0 و MusiteDeep و NETPhos 3-1 و GPS-MSP و GPS-UBER و PRmePRed و GPS.PBS و CarSite-II و GPS-YNO2 و GPS-SNO 1.0 للتنبؤ بالتغيرات اللاحقة. -ترجمة. بعد ذلك ، استخدمنا برنامج Phyre2 و SWISS-MODEL لإجراء النمذجة الجزيئية لهذه البروتينات المتحولة و ERRAT / PROCHECK لتقييم جودة نتائجننا.

تشير نتائجننا إلى أن بعض الاختلافات في بروتين TGFBR2 (R448G و R520L و R522Q و K277N) يمكن أن تؤدي إلى تعديلات مختلفة بعد الترجمة مثل المثيلة والانتشار.

في الختام ، توفر دراستنا في السيليكو أساسًا أوليًا مثيرًا للاهتمام فيما يتعلق بالتأثير المحتمل لبعض SNPs على TGFBR2 في سياق سرطان القولون والمستقيم. يمكن استخدام هذه النتائج كنقطة انطلاق للدراسات التجريبية التي تهدف إلى التحقق من صحة آثار هذه النيوكليوتايد على البروتينات ودورها في تطور هذا المرض.

الكلمات المفتاحية: سرطان القولون والمستقيم ، نمذجة البروتين ، تعديلات ما بعد الترجمة ، nsSNP ، SMAD7 ، TGFBR2.

Remerciement

Avant tout on remercie dieu tout puissant de nous avoir donné le privilège, la chance d'étudier et de nous avoir donné force, courage, et patience pour accomplir ce travail. Sans oublier nos parents qui ont veillé sur nous durant toute notre vie.

Nous remercions chaleureusement notre encadreur **Mme HADJIDJ Zeyneb**, docteur en immunologie à l'Université ABOU BEKR BELKAID - Tlemcen d'avoir suivi et dirigé ce travail. Et pour Son aide, ses conseils, ses orientations et sa gentilles.

Nous tenons à remercier avec plus grande gratitude **Pr ARIBI Mourad**, Professeur à l'université de Tlemcen de l'honneur qu'il nous a fait d'avoir accepté de présider le jury de ce mémoire.

Nous aimerions également remercier **Mme El Mezouar Chahrazed**, docteur à l'université de Tlemcen, d'avoir accepté de se joindre à ce jury comme examinateur.

Nos vifs remerciements vont aux l'ensemble de l'équipe du laboratoire de recherches de **BIOMOLIM**.

Nos remerciements vont également à Nos collègues, **BENCHABAN Amaria** et **BELKAID Adnan** et à tous ceux et celles qui ont contribué de près ou de loin, par leurs conseils, leurs suggestions et par leurs encouragements, à la réalisation de ce travail.

Liste des figures

Liste des figures

- Figure I.1:** Carte montrant la distribution mondiale des taux d'incidence estimés, normalisés selon l'âge, du CCR en 2020 pour les deux sexes et tous les âges
- Figure I.2:** Carte montrant la distribution mondiale des taux de mortalité estimés, normalisés selon l'âge, du CCR en 2020 pour les deux sexes et tous les âges
- Figure I.3:** Les principaux facteurs de risque associés au cancer colorectal
- Figure I.4:** Mécanismes d'échappement à la réponse immunitaire
- Figure I.5:** la structure du TGF β R2
- Figure I.6:** le récepteur du facteur de croissance transformant R2 (TGF- β R2)
- Figure I.7:** la structure 3D de SMAD 7
- Figure I.8:** les domaines MH1 et MH2 de SMAD 7
- Figure I.9:** La voie de signalisation TGF β
- Figure II.1:** study design
- Figure II.2:** Recherche des séquences sur la base de données UniProt.
- Figure II.3:** Prédiction des sites de phosphorylation par GPS 6.0
- Figure II.4:** Prédiction des sites de phosphorylation par NetPhos 3.1
- Figure II.5:** Prédiction des sites de phosphorylation et d'ubiquitination par MusiteDeep
- Figure II.6:** Prédiction des sites de méthylation par GPS-MSP 1.0
- Figure II.7:** Prédiction des sites de méthylation par PRmePRed
- Figure II.8:** Prédiction des sites d'ubiquitination par GPS-Uber
- Figure II.9:** Prédiction des domaines de liaison aux phosphoprotéines par GPS-PBS
- Figure II.10:** Prédiction des sites de carbonylation par iCarPS
- Figure II.11:** Prédiction des sites de carbonylation par CarSite-II
- Figure II.12:** Prédiction des sites de nitration des tyrosines par GPS-YNO2
- Figure II.13:** Prédiction des sites de S-nitrosylation et de nitration des tyrosines par DeepNitro
- Figure II.14:** Prédiction des sites de S-nitrosylation des protéines par GPS-SNO
- Figure II.15:** Modélisation des structures tridimensionnelles des protéines par Phyre2
- Figure II.16:** Modélisation des structures tridimensionnelles des protéines par SWISS-MODEL
- Figure II.17:** alignement des structures des protéines par TM-align
- Figure II.18:** Évaluations de la qualité des structures des protéines par ERRAT
- Figure II.19:** Évaluations de la qualité des structures des protéines par PROCHEK
- Figure II.20:** Modélisation des structures tridimensionnelles des protéines par Chimera

Liste des tableaux

Tableau III.1 : Résultats de la prédiction des sites de modification post-traductionnelle

Tableau III.2 : examen des effets des nsSNPs sur la structure de TGFBR2 (test par Phyre2)

Tableau III.3 : examen des effets des nsSNPs sur la structure de SMAD7 (test par Phyre2)

Tableau III.4 : examen des effets des nsSNPs sur la structure de TGFBR2 (test par SWISS-MODEL)

Tableau III.5 : examen des effets des nsSNPs sur la structure de SMAD7 (test par SWISS-MODEL)

Tableau III.6 : Estimations de qualités des protéines modélisées par SWISS-MODEL

Tableau III.7 : Comparaison structurelle entre la séquence sauvage de TGFBR2 et leurs séquences mutées par Chimera.

Tableau III.8 : Comparaison structurelle entre la séquence sauvage de SMAD7 et leurs séquences mutées par Chimera.

Liste des abréviations

CCR : Cancer colorectal

TGF- β : Facteur de croissance transformant beta

TGF- β R2 : Récepteur de Facteur de croissance transformant beta

AcM: Anticorps monoclonaux

NK : Natural killer cell

IDA : Anémie ferriprive

SMAD: Small mothers against decapentaplegic

EGF: Epidermal Growth Factor

SNP: Polymorphisme nucléotidique simple

nsSNP: non-Synonyme polymorphisme nucléotidique simple

ARNm: acide ribonucléique messenger

HNPCC: cancer colorectal héréditaire sans polypose

PAF : polypose adénomateuse colique familiale

IRM: imagerie par résonance magnétique

TDM: tomodensitométrie

ATP: adenosine triphosphate

ADP: adenosine diphosphate

AA : acide aminé

TCD4 : Lymphocyte T cluster de différenciation 4.

T-reg : T régulateur.

TCD8 : Lymphocyte T cluster de différenciation 8

Table des matières

Dédicace.....	I
Résumé.....	II
Abstract.....	III
ملخص.....	IV
Remerciement.....	V
Liste des figures.....	VI
Liste des tableaux.....	VII
Liste des abréviations.....	VIII
Introduction.....	1
Chapitre I : Revue de la littérature.....	4
1. Cancer colorectal.....	5
1.1. Généralités.....	5
1.2. Facteurs de risque.....	6
1.3. Développement du cancer colorectal.....	7
1.4. Symptômes et diagnostic.....	8
1.5. Traitement.....	9
1.6. Immuno-édition.....	10
2. TGFβR2.....	12
2.1. Généralités.....	12
2.2. Structure.....	12
2.3. Fonctions.....	13
2.4. TGF βR2 et cancer colorectal.....	14
3. SMAD 7.....	14
3.1. Généralité.....	14
3.2. Structure.....	15
3.3. La relation entre SMAD 7 et TGF β dans le cancer colorectal.....	15
3.4. Voies de signalisation.....	16
4. Polymorphisme nucléotidique simple.....	17
4.1. Généralité.....	17
4.2. Types de polymorphisme nucléotidique.....	17
4.3. Avantage des polymorphismes nucléotidiques.....	18
4.4. Le polymorphisme nucléotidique dans le cancer colorectal.....	18
4.5. Analyse <i>in silico</i> des polymorphismes nucléotidiques.....	18
5. Modifications post traductionnelles.....	19
5.1. Généralités.....	19

Table de matières

5.2. Prédiction <i>in silico</i> des modifications post traductionnelles	21
6. Problématique et objectif.....	21
6.1. Problématique.....	21
6.2. Objectif	21
6.3. But	21
Chapitre II : Matériels et méthodes.....	22
1. Plan de l'étude.....	23
2. Collecte de données et sélection de séquences de protéines	24
3. Prédiction des sites de modification post-traductionnelle	24
3.1. Prédiction des sites de phosphorylation.....	25
3.2. Prédiction des sites de méthylation.....	27
3.3. Prédiction des sites d'ubiquitination.....	28
3.4. Prédiction des domaines de liaison aux phosphoprotéines	28
3.5. Prédiction des sites de carbonylation.....	29
3.6. Prédiction des sites de nitration des tyrosines	31
3.7. Prédiction des sites de S-nitrosylation	31
4. Modélisation moléculaire	33
5. Comparaison entre les structures tridimensionnelles des protéines	34
6. Évaluations de la qualité des structures des protéines modélisées	35
7. Comparaison et superposition des structures 3D	37
Chapitre III : Résultats et interprétation.....	38
1. Résultats de la Prédiction des sites de modification post-traductionnelle	39
2. Résultats examen des effets des nsSNPs par modélisation 3D des protéines	40
3. Comparaison et superposition des structures 3D	52
Chapitre IV : Discussion	56
Chapitre V : Conclusion et perspectives	59
Chapitre VI : Bibliographies	61
Chapitre VII : Annexes	66

Introduction

Introduction

Introduction :

Le cancer colorectal (CCR) est un problème de santé publique majeur dans le monde entier. Il représente la troisième forme de cancer la plus fréquente, après celui du poumon et du sein. En termes de mortalité, il se classe deuxième après les cancers pulmonaires (**Lotfollahzadeh et al., 2023**). Selon les estimations récentes, près de 1,9 million de nouveaux cas ont été diagnostiqués en 2020 avec environ 930 000 décès dans le monde (**Morgan et al., 2023**).

Les symptômes de cette maladie sont peu fréquents, voire inexistants. Même lorsqu'ils se manifestent, ils sont souvent non spécifiques et peuvent être associés à d'autres pathologies. Cette situation complique considérablement la détection de la tumeur et retarde le diagnostic (**Vega et al., 2015**), ce qui réduit les chances de traitement efficace pour les patients atteints.

Le traitement conventionnel standard du CCR comprend généralement la chirurgie, la chimiothérapie et la radiothérapie (**Johdi and Sukor, 2020**). Cependant, il existe également des traitements plus récents tels que l'immunothérapie anticancéreuse qui renforce le système immunitaire en restaurant les capacités des lymphocytes à éliminer les cellules cancéreuses. Ces thérapies utilisent des inhibiteurs de points de contrôle immunitaires tels que les Anti-CTLA4, les Anti PD-L1 et les Anti PD-1, ainsi que des anticorps monoclonaux et des cytokines. La combinaison de ces approches avec les traitements conventionnels existants, le pronostic des patients atteints de CCR peut être amélioré (**Mukherji et al., 2022**).

Le système immunitaire joue un double rôle dans le cancer : il peut non seulement supprimer la croissance tumorale en détruisant les cellules cancéreuses, mais aussi promouvoir la progression de la tumeur en sélectionnant les cellules tumorales ou en créant un microenvironnement tumoral immunosuppresseur (**Markman and Shiao, 2015**).

Le facteur de croissance transformant bêta (TGF- β) est une cytokine pléiotropique qui exerce des effets biologiques complexes et cruciaux sur plusieurs processus physiologiques des cellules, tels que la prolifération, l'apoptose, la différenciation, la survie et la migration cellulaire (**Syed, 2016**). Cependant, dans certaines conditions pathologiques telles que les cancers, le TGF- β peut également contribuer au développement de ces maladies en favorisant l'invasion et la migration des cellules cancéreuses. Ces propriétés oncogéniques sont souvent associées à une régulation altérée du signal transducteur SMAD7 qui perturbe les voies de signalisation en aval impliquées dans le contrôle de la croissance tumorale (**Stolfi et al., 2013**). De plus, il a été démontré que le TGF- β pouvait altérer les réponses cellulaires aux traitements anticancéreux classiques tels que la chimiothérapie en induisant une résistance

Introduction

accrue aux médicaments utilisés pour traiter les tumeurs (**Gu and Feng, 2018**).

Les polymorphismes mononucléotidiques non synonymes (nsSNP) sont des mutations génétiques d'une seule base qui entraînent une modification de la séquence d'acides aminés de la protéine codée (**Yates and Sternberg, 2013**). Ces variations peuvent avoir un impact sur la structure et la fonction des protéines, ce qui peut conduire à l'apparition de maladies complexes et fréquentes telles que le cancer colorectal (CCR) (**Butler, 2012**). Les nsSNPs ont donc émergé comme des biomarqueurs diagnostiques et thérapeutiques potentiels dans le CCR (**Deng et al., 2017**).

L'identification précise des nsSNPs impliqués la maladie est un défi majeur en raison du grand nombre possible de variantes génétiques différentes. Pour y remédier, les scientifiques ont désormais recours aux analyses *in silico* en utilisant différentes techniques computationnelles et différents logiciels pour prédire les effets fonctionnels des mutations sur les protéines codées par les gènes porteurs de ces mutations. Cette approche permet aux chercheurs d'évaluer rapidement les conséquences potentielles des variantes génétiques identifiées sans avoir besoin de réaliser des expériences coûteuses au laboratoire (**Lim et al., 2021**).

En 2022, Zouaimia Nahla et Senouci Meriem Radia ont réalisé une analyse *in silico* du polymorphisme nucléotidique ayant des conséquences fonctionnelles sur TGFBR2 et SMAD7 afin d'identifier les variations délétères à risque élevé qui peuvent être impliquées dans le cancer colorectal ou d'autres pathologies et explorer des nouvelles perspectives dans le diagnostic et la prévention de la maladie. Dans ce travail, nous allons étudier *in silico* l'effet de ces variations sur la structure tridimensionnelle de TGFBR2 et SMAD7 et leurs modifications post traductionnelles.

Chapitre I : Revue de la littérature

1. Cancer colorectal

1.1. Généralités

Le CCR est une maladie caractérisée par une prolifération cellulaire maligne qui se développe aux dépens des constituants histologiques du colon et du rectum (**Kuipers et al., 2015**).

En raison de sa fréquence élevée et de sa gravité, le CCR représente un problème majeur de santé publique dans le monde entier. Selon les données de GLOBOCAN, plus de 1,9 million nouveaux cas ont été diagnostiqués en 2020, faisant ainsi du CCR le troisième cancer le plus courant et la deuxième cause de décès par cancer avec près de 935000 décès estimés (**Hossain et al., 2022**). L'incidence et la mortalité liées au CCR varient considérablement entre les différents pays à travers le monde. Les pays développés tels que l'Australie, l'Amérique du Nord et l'Europe sont les plus exposés au risque tandis que l'incidence tend à être basse en Asie, en Amérique latine et en Afrique. Ces variations peuvent être associées à des niveaux socio-économiques différents dans chaque pays (**GLOBOCAN 2020**).

En Algérie également, Le CCR a connu une augmentation alarmante ces dernières années. Selon les statistiques fournies par l'**Institut National de Santé Publique** pour l'année 2021 près de 6 500 nouveaux cas ont été recensés dont environ 3.500 cas chez les hommes et 3.000 cas chez les femmes.

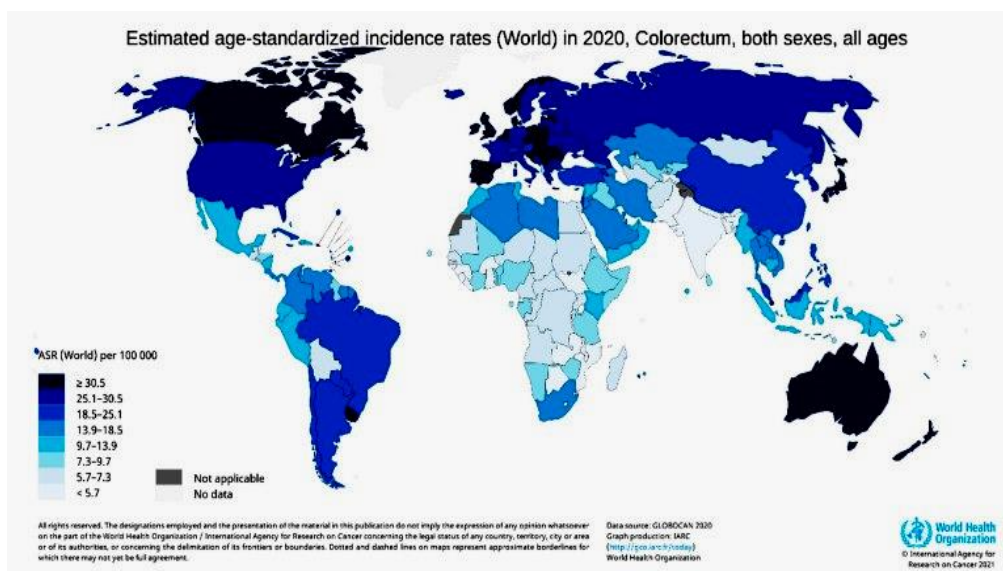


Figure 1: Carte montrant la distribution mondiale des taux d'incidence estimés, normalisés selon l'âge, du CCR en 2020 pour les deux sexes et tous les âges (**GLOBOCAN**).

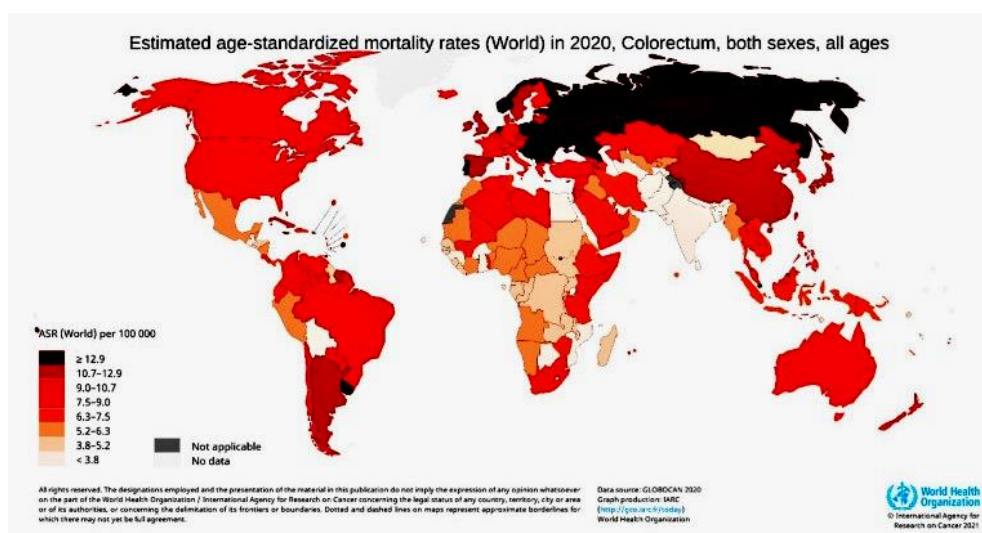


Figure 2: Carte montrant la distribution mondiale des taux de mortalité estimés, normalisés selon l'âge, du CCR en 2020 pour les deux sexes et tous les âges (GLOBOCAN).

1.2. Facteurs de risque

Le CCR est une maladie multifactorielle (**Duan et al., 2022**), et associée à des facteurs de risque modifiables qu'ils sont liés aux habitudes ou aux modes de vie individuels (l'obésité, tabagisme, consommation d'alcool, Le régime pauvre en fibres végétales ...), et des facteurs non modifiables qui ne peuvent pas être contrôlés par les individus, notamment les Antécédents médicaux familiaux et personnels, l'âge, sexe, race....(**Mármol et al., 2017**).

L'âge est considéré comme l'un des facteurs les plus importants influençant le risque de développer un CCR. Avant 40 ans, les cancers colorectaux sont rares. Le risque augmente à partir de 50 ans et s'accroît jusqu'à 80 ans (**Rawla et al., 2019**).

Dans une étude comparative, les hommes ont un risque environ 30 % plus élevé de développer un CCR que les femmes. Cette disparité peut être due à une différence d'exposition aux facteurs de risque, aux habitudes alimentaires et aux hormones sexuelles (**Keum and Giovannucci, 2019**).

L'incidence du cancer colorectal varie considérablement en fonction de la race. Les Noirs non hispaniques ont l'un des taux les plus élevés de tous les groupes raciaux, ce qui constitue un facteur important dans la prévalence globale de cette maladie(**Sharma et al., 2020**).

Le risque est également augmenté chez les personnes souffrant de maladies inflammatoires de l'intestin (MII) comme la maladie de Crohn et la colite ulcéreuse (**Shah and Itzkowitz, 2022**). Et aussi chez les personnes atteintes de diabète de type 2 en raison de divers mécanismes physiopathologiques potentiels comprennent l'hyperinsulinémie, l'axe du facteur de croissance analogue à l'insuline (IGF), l'hyperglycémie, l'inflammation induite par un

dysfonctionnement du tissu adipeux, un trouble de la motilité gastro-intestinale et une surveillance immunologique altérée (Yu et al., 2022).

D'autres résultats d'études épidémiologiques suggèrent des risques accrus modestes et faibles de CCR dû à l'obésité, la viande rouge / transformée, le tabac, l'alcool, la thérapie de privation d'androgènes et la cholécystectomie....(Lotfollahzadeh et al., 2023).

Le CCR peut avoir des causes génétiques dans environ 2 à 8 % des cas. Les syndromes héréditaires représentent donc une proportion significative de ces cancers. Parmi les différents types de syndromes, le syndrome HNPCC (cancer colorectal héréditaire sans polypose) et le syndrome PAF (polypose adénomateuse colique familiale) sont les plus courants (Valle et al., 2019).

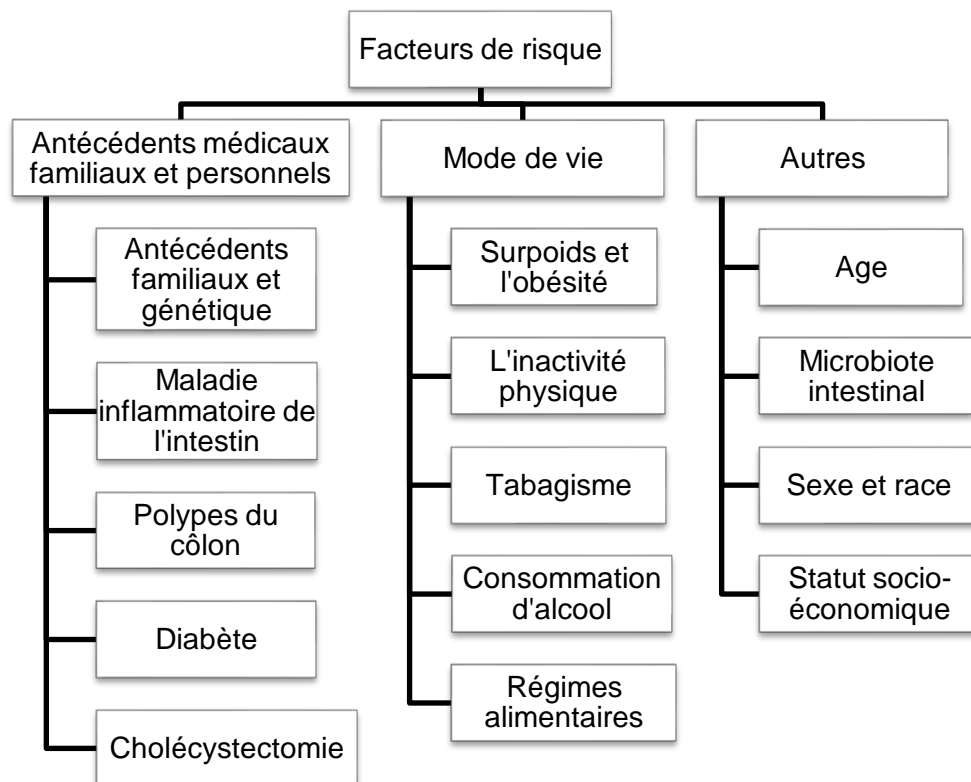


Figure 3: Les principaux facteurs de risque associés au cancer colorectal (Sawicki et al., 2021).

1.3. Développement du cancer colorectal

Dans la grande majorité des cas (96 %), le cancer colorectal se développe dans les muqueuses de l'intestin sous forme d'adénocarcinomes (Hossain et al., 2022).

Au niveau moléculaire, le développement du cancer colorectal implique une accumulation progressive de mutations génétiques au sein des cellules épithéliales coliques

ainsi que l'activation d'oncogènes et l'inactivation de gènes suppresseurs de tumeurs. Plusieurs voies distinctes ont été identifiées, notamment l'instabilité microsatellitaire (MSI), l'instabilité chromosomique (CIN) et la voie dentelée associée à un phénotype méthylateur CpG (CIMP). L'instabilité microsatellitaire est souvent observée chez les patients atteints du syndrome HNPCC tandis que l'instabilité chromosomique est plus courante chez ceux atteints du syndrome de PAF. La voie dentelée CIMP peut être causée par des facteurs environnementaux ou génétiques spécifiques (**Malki et al., 2020**).

Le CCR se développe lorsque les cellules épithéliales acquièrent une série de modifications génétiques ou épigénétiques qui leur permettent de proliférer de manière excessive. Cette prolifération incontrôlée peut donner naissance à un polype bénin ou adénome (stade 0) qui peut évoluer en cancer invasif. Environ 10 % des polypes adénomateux peuvent ainsi progresser vers un stade malin, formant un adénocarcinome qui envahit la musculature propria (stade I). Au fur et à mesure que la tumeur grossit, elle envahit davantage les tissus environnants jusqu'à atteindre la séreuse (stade II) puis le péritoine viscéral (stade III). À ce stade, il est également possible que des métastases lymphatiques ou vasculaires apparaissent dans d'autres parties du corps (stade IV) (**Pernot et al., 2014**).

Le stade détermine la gravité de la maladie et influence les options thérapeutiques disponibles pour chaque patient. Bien que la chirurgie soit généralement le traitement standard pour les stades 0 à II du cancer colorectal, le stade III nécessite souvent une intervention chirurgicale suivie d'une chimiothérapie adjuvante tandis que le traitement combiné comprenant l'intervention chirurgicale, une chimiothérapie et une thérapie ciblée est recommandé pour les patients atteints d'un cancer colorectal récurrent ou au stade IV (**Slattery et al., 1998**).

1.4. Symptômes et diagnostic

Le cancer colorectal reste longtemps asymptomatique. Ainsi, lorsque les symptômes apparaissent, ils sont souvent le signe d'une maladie déjà évoluée. Plus le diagnostic tarde à être posé et plus ces symptômes sont nombreux et fréquents (**Balchen and Simon, 2016**). Plusieurs symptômes courants associés au cancer colorectal ont été identifiés (**Holtedahl et al., 2021**) :

- Saignement de rectum.
- Présence de sang dans les selles (il peut être rouge clair ou très foncé).
- Masse dans l'abdomen ou le rectum.

- Douleur ou malaise au niveau de l'abdomen ou du rectum.
- Diarrhées ou constipations.
- Perte d'appétit inexplicée.
- Faiblesse.
- Anémie ferriprive.
- Troubles respiratoires.

Le cancer colorectal est souvent difficile à détecter car il n'a pas de signes spécifiques. Ces symptômes sont généralement associés à d'autres affections médicales, ce qui peut retarder le diagnostic jusqu'à un stade avancé de la maladie (**Vega et al., 2015**).

Plusieurs techniques peuvent être utilisées pour détecter le cancer colorectal. L'une des plus importantes est la coloscopie, qui permet au médecin d'examiner visuellement la muqueuse intestinale et de détecter les polypes précancéreux afin de faciliter leur suppression si nécessaire (**Kaminski et al., 2010**).

Une biopsie est souvent requise pour confirmer avec certitude un diagnostic de cancer colorectal. Cette intervention consiste à prélever des cellules ou une petite quantité de tissu du colon ou du rectum afin qu'ils soient examinés au microscope pour identifier toute présence cancéreuse dans les échantillons prélevés. La biopsie peut être réalisée lors d'une coloscopie ou d'une sigmoïdoscopie lorsque des polypes ont été identifiés (**Mahmoud, 2022**).

D'autres examens complémentaires sont également utilisés pour évaluer l'étendue du cancer colorectal. La radiographie, l'échographie, la tomodensitométrie (TDM) ou encore l'imagerie par résonance magnétique (IRM) permettent d'obtenir des images détaillées des tissus, des organes et des os afin de détecter toute anomalie éventuelle. Ces techniques permettent également de mesurer la taille de la tumeur et son extension aux autres parties du corps si nécessaire pour une prise en charge plus adaptée à chaque cas individuel (**Duan et al., 2022**).

1.5. Traitement

Le traitement du cancer colorectal dépend de plusieurs facteurs, tels que le stade et la localisation de la maladie, ainsi que l'état général du patient. Les options thérapeutiques incluent la chirurgie, la chimiothérapie, la radiothérapie, les thérapies ciblées et l'immunothérapie (**LAOUAR and DAOUDI, 2016**).

La chirurgie est considérée comme le pilier du traitement curatif pour le cancer colorectal (**Röhrl et al. 2019**), Elle a connu des améliorations majeures grâce à l'utilisation d'instruments plus sophistiqués et à une évaluation préopératoire plus minutieuse. La laparoscopie est une

technique mini-invasive couramment utilisée qui offre plusieurs avantages par rapport à la chirurgie ouverte traditionnelle (**Lee et al. 2017**).

La chimiothérapie est un traitement qui consiste à administrer des médicaments cytotoxiques pour détruire les cellules cancéreuses que les examens ne peuvent pas détecter. Elle peut être utilisée en complément d'une intervention chirurgicale pour éliminer les cellules cancéreuses résiduelles et éviter une récurrence ou la formation de métastases (chimiothérapie adjuvante), pour réduire la taille du cancer colorectal avant une intervention chirurgicale (chimiothérapie néo-adjuvante) ou encore pour améliorer la survie et le confort du patient atteint d'un cancer avancé (chimiothérapie palliative) (**Bretin, 2019**) (**Correard et Estève, 2020**).

La radiothérapie utilise des rayons X ou des particules de haute énergie pour détruire les cellules cancéreuses dans le rectum ou dans tout autre endroit concerné par ce type de cancer. Cette méthode permet notamment de réduire le risque qu'un nouveau foyer cancéreux apparaisse après un premier traitement (**Bastien, 2009**).

Les thérapies ciblées ont été une avancée importante dans le traitement des patients atteints de cancer colorectal métastatique, avec en particulier le ciblage de la voie du récepteur de l'EGF (Epidermal Growth Factor) par des anticorps monoclonaux anti-récepteurs de l'EGF (EGFr). La liaison de l'EGF à son récepteur active des voies de signalisation intracellulaires conduisant à la prolifération des cellules tumorales. Le blocage de la voie de signalisation du récepteur de l'EGF est ciblé par le cétuximab (Erbix®), panitumumab (Vectibix®) et le bévacizumab (Avastin®) (**Di Fiore, 2009**).

L'immunothérapie est de plus en plus reconnue comme une modalité thérapeutique clé pour traiter le cancer (**Das et al. 2017**), elle s'est rapidement imposée comme une modalité de traitement majeure pour plusieurs types de cancers solides, y compris un sous-ensemble de cancers colorectaux (**Ganesh et al. 2019**). L'objectif principal moderne est d'une part la stimulation de la réponse immunitaire anti tumorale et d'autre part, la désactivation des mécanismes suppresseurs qui soutiennent la croissance des tumeurs (**Abakushina et al. 2019**).

1.6. Immuno-édition

L'Immunoédition est un processus qui permet à l'organisme de reconnaître et d'éliminer les cellules tumorales. Dans le cancer colorectal, ce processus peut être altéré, ce qui entraîne une progression de la maladie (**Lasek, 2022**).

Chapitre I : Revue de littérature

Elle est connue sous le nom de théorie des 3E, car elle se déroule en trois phases : l'élimination, l'équilibre et l'échappement.

La phase d'élimination est caractérisée par une réponse immunitaire forte et efficace contre les cellules tumorales. Les cellules tumorales sont éliminées par les cellules immunitaires telles que les cellules NK (Natural Killer) et les cellules T cytotoxiques. Cette phase est cruciale pour la prévention de la formation de tumeurs malignes (**Markman and Shiao, 2015**).

La phase d'équilibre est caractérisée par une lutte constante entre les cellules tumorales et le système immunitaire. Les cellules tumorales sont maintenues en échec par le système immunitaire, mais elles ne sont pas éliminées. Cette phase peut durer plusieurs années. Si les cellules tumorales ne sont pas éliminées, elles peuvent évoluer en tumeurs malignes (**Markman and Shiao, 2015**).

La phase d'échappement est caractérisée par l'échec du système immunitaire à contrôler la croissance des cellules tumorales. Les cellules tumorales échappent à la surveillance immunitaire en développant des mécanismes qui inhibent la réponse immunitaire. Ces mécanismes peuvent inclure l'expression de molécules qui inhibent l'activation des cellules immunitaires ou la production de cytokines qui suppriment la réponse immunitaire. Cette phase est associée à la progression de la maladie et à la formation de CCR (**Markman and Shiao, 2015**).

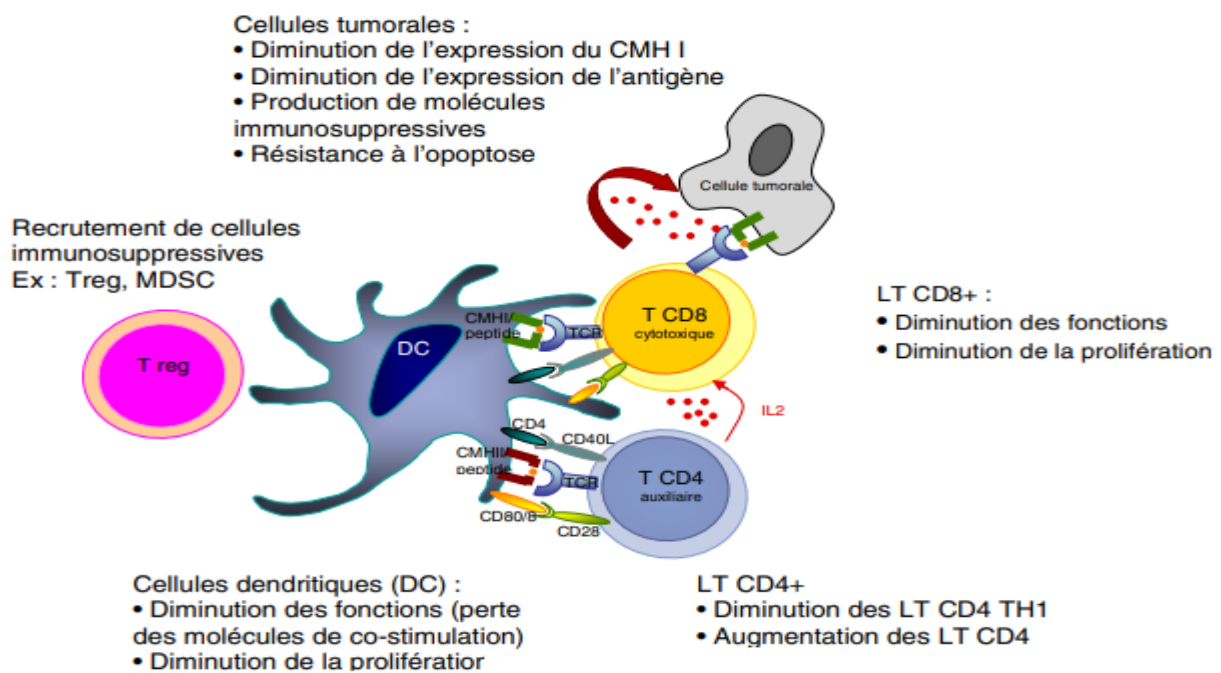


Figure 4: Mécanismes d'échappement à la réponse immunitaire (**Lescaille et al, 2014**).

Les mécanismes d'échappement à la réponse immunitaire sont multiples. Ils sont soit directement liés à la cellule tumorale, soit liés à une modulation de la réponse immunitaire par la tumeur (**Zhang et al., 2020**).

Parmi ces mécanismes on trouve :

- ✓ La perte d'expression du CMH de classe I
- ✓ La dérégulation de la présentation antigénique
- ✓ L'augmentation de l'expression de molécules immuno-inhibitrices
- ✓ Les cellules tumorales colorectales sécrètent des facteurs immunosuppresseurs tels que le TGF- β et favorisent le recrutement de Tregs et MDSCs
- ✓ Les cellules tumorales ont la capacité de déréguler les signaux de mort cellulaire et d'échapper à la lyse.

2. TGF β R2

2.1. Généralités

Le TGF- β est considéré comme un acteur clé dans les processus physiologiques essentiels. Cependant, il est également impliqué dans certaines pathologies humaines telles que le cancer colorectal, où son rôle peut varier en fonction du stade de développement de la tumeur. La réponse cellulaire au TGF- β est initiée par sa fixation sur les récepteurs TGF- β R2 qui activent ensuite les facteurs de transcription SMAD (**Jung et al., 2017**).

Le TGF β R2 joue un rôle important dans le système immunitaire et est également impliqué dans la progression du cancer. La liaison entre le TGF- β et le TGF β R2 permet la formation d'un complexe hétéromérique qui conduit à la phosphorylation de TGF β R1. La phosphorylation de TGF β R1 est une étape clé pour activer les cascades de signalisation cellulaires qui régulent divers processus biologiques tels que la prolifération cellulaire, l'apoptose ou encore l'inflammation. Cependant, des altérations génétiques ou épigénétiques peuvent conduire à une dysrégulation de ces voies de signalisation et favoriser ainsi le développement tumoral (**Meulmeester and ten Dijke, 2011**).

2.2. Structure

Le TGF β R2 possède une structure rigide appelée nœud de cystéine, qui explique la résistance du TGF β à la chaleur, aux agents dénaturants et au pH externe. La quatrième cystéine contribue à la liaison intermoléculaire entre deux peptides pour former un dimère (**Javelaud and Mauviel, 2004**).

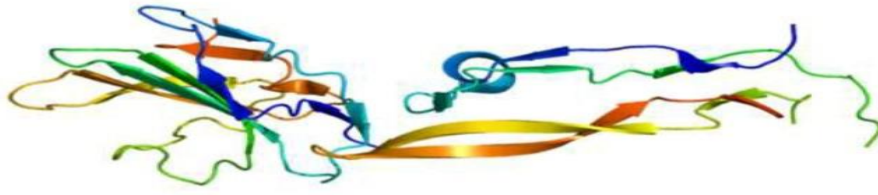
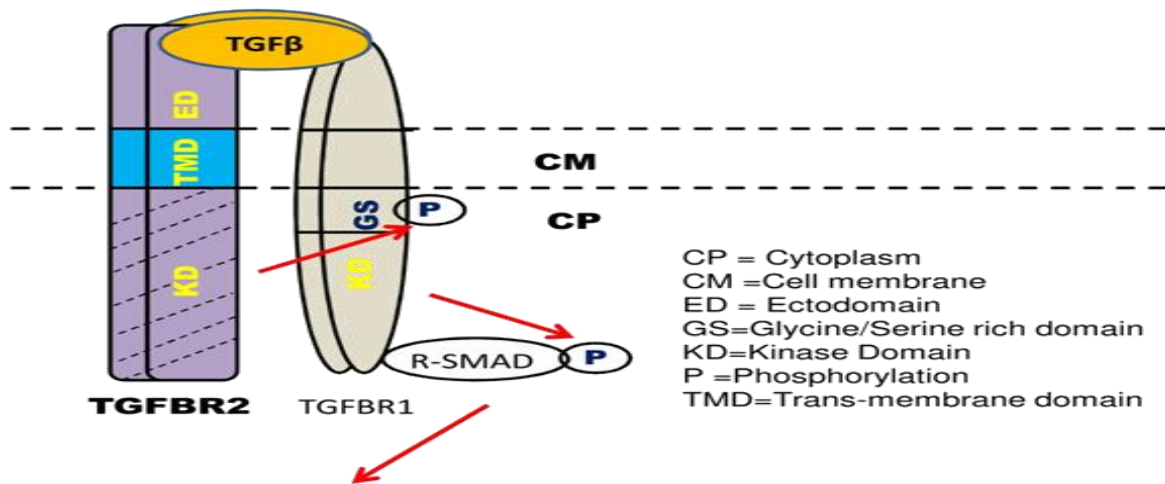


Figure 5: la structure du TGF β R2 (Javelaud and Mauviel, 2004)

Le récepteur TGF- β R2 est composé d'un domaine extracellulaire riche en cystéine et d'un domaine kinase intracellulaire. Il possède une activité kinase particulière qui permet l'activation et la phosphorylation du domaine GS (glycine/serine) du récepteur de type 1. La formation du complexe entre le récepteur de type 1 et le récepteur TGF- β R2 est essentielle pour engendrer une réponse au TGF- β (Serra, 2002)..



Activation/repression of TGF β target genes

Figure 6: le récepteur du facteur de croissance transformant R2 (TGF- β R2) (Serra, 2002).

2.3. Fonctions

2.3.1. Régulation du cycle cellulaire

Le cycle cellulaire est inhibé par le TGF- β au niveau de la phase G1 et la phase S. Son mécanisme d'action implique notamment l'inhibition du facteur de transcription c-MYC (hélice-boucle-hélice) ainsi que la phosphorylation de la protéine du rétinoblastome (pRb), qui joue un rôle important dans la régulation des cellules à travers la phase G1. Le récepteur TGF- β R2 active le domaine kinase intracellulaire du récepteur TGF- β R1. Cette activation conduit ensuite à une cascade complexe d'événements moléculaires qui aboutissent finalement à l'inhibition de l'activité cycline-dépendante kinases (cDKs) dans les cellules (Serra, 2002).

2.3.2. Suppression de tumeur

Le TGF- β est un facteur de croissance inhibiteur des cellules épithéliales. Son mécanisme d'action complexe implique notamment la régulation de l'expression des gènes impliqués dans la prolifération et la différenciation cellulaires. Cependant, bien que le TGF- β puisse avoir un effet suppressif sur la progression tumorale *in vivo* en réduisant la prolifération et en induisant l'apoptose des cellules cancéreuses, il peut également favoriser l'apparition de tumeurs plus tardives en stimulant l'invasion et la métastase. En effet, les altérations moléculaires associées aux stades avancés du cancer peuvent perturber les voies de signalisation du TGF- β normalement inhibitrices, conduisant ainsi à une réponse pro-tumorale plutôt qu'à une réponse anti-tumorale. D'autre part, le récepteur TGF- β R2 joue un rôle crucial dans cette voie et ses altérations génétiques ou épigénétiques peuvent affecter significativement l'efficacité du signal transmis par le TGF- β (Serra, 2002).

2.4. TGF β R2 et cancer colorectal

Le TGF- β R2 joue un rôle important dans l'activation de l'immunité antitumorale et la suppression des tumeurs. Les patients atteints de cancer colorectal (CCR) avec une perturbation du TGF- β R2 présentent généralement un taux plus élevé de métastases par rapport aux autres types de CCR. D'autre part, l'absence d'expression du facteur de transcription SMAD, qui joue un rôle important dans cette voie de signalisation TGF- β , peut entraîner une évolution défavorable après une chirurgie curative (Jung et al., 2017).

3. SMAD 7

3.1. Généralité

Le gène SMAD7 code pour une protéine humaine appartenant à la famille des protéines SMAD. La fonction principale de SMAD 7 est d'inhiber et de réguler négativement la voie de signalisation TGF- β afin de préserver les effets anti-inflammatoires associés à cette voie (Nassiri et al., 2013).

Les études ont montré que la suppression du gène SMAD7 était présente dans environ 43% des échantillons biopsiques provenant de patients atteints de CCR, ce qui suggère que ce marqueur pourrait être un indicateur pronostique négatif chez ces patients (Troncone and Monteleone, 2019).

3.2. Structure

Le SMAD 7 possède deux domaines, le premier c'est MH1 qui représente un domaine amino-terminal à une structure en épingle à cheveux β engagé dans les liaisons spécifiques, de plus le deuxième domaine MH2 est un carboxy-terminaux conservé leur liaisons assurées par les régions divergentes, il contient des constituants structurels importants pour transmettre le signal, donc la boucle L3 est un capital des interactions spécifiques.

Les domaines MH1 et MH2 interfèrent avec plusieurs collaborateurs de liaison qui permettent la modification des propriétés de signalisation des SMADs. Ces indications fournissent une base structurelle pour la création du complexe SMAD quelque soit homo ou hétérotrimérique (Li et al., 2017).

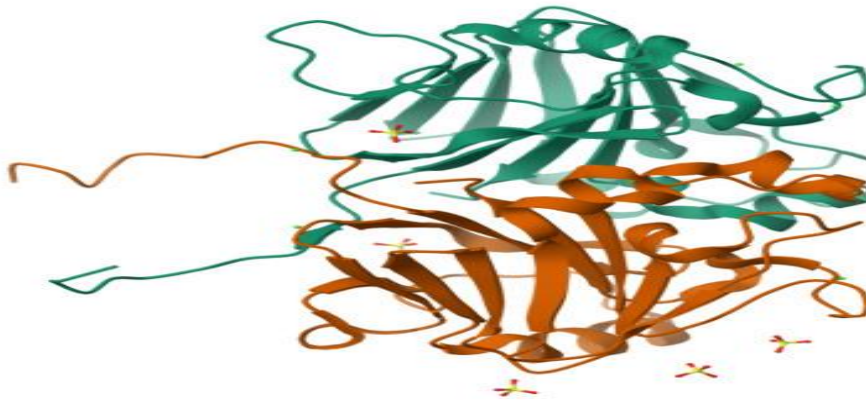


Figure 7: la structure 3D de SMAD 7 (Li et al., 2017, p. 7).

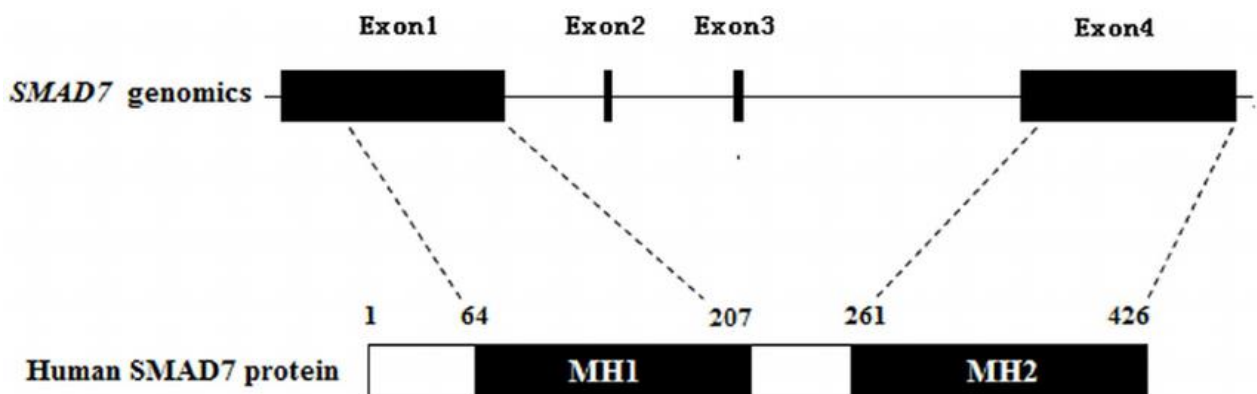


Figure 8: les domaines MH1 et MH2 de SMAD 7 (Gomes et al, 2021).

3.3. La relation entre SMAD 7 et TGF β dans le cancer colorectal

Le SMAD 7 joue un rôle clé dans la signalisation de TGF- β dont le cadre de la régulation des

processus biologiques. l'inhibiteur SMAD 7 actionner comme des antagonistes des TGF β qui résoudre positivement dans le cancer colorectal .en plus le SMAD 7 inhibe les réponses transcriptionnelles provoquer par le TGF- β qui bloque la formations des complexes entre les SMAD résulte l'inhibition de la régulation négative médiée par le TGF- β du phosphorylation de Rb donc le SMAD 7 augmente la tumorigénicité des cellules et bloque la croissance par le TGF- β donc cela entrainer une tumorigénicité accrue (Li et al. 2017).

3.4. Voies de signalisation

Le ligand TGF- β se lie spécifiquement au récepteur TGF- β R2, qui contient des domaines de sérine et de thréonine kinase. Ce récepteur forme ensuite un complexe hétéromérique avec le récepteur TGF- β R1. La phosphorylation du domaine GS (glucine-sérine) du récepteur TGF- β R1 par le récepteur TGF- β R2 conduit à l'activation de l'activité kinase sérine/thréonine, ce qui entraîne la phosphorylation des molécules de signalisation impliquées dans la voie de signalisation TGF β (Javelaud and Mauviel, 2004).

Les SMADs sont phosphorylés par le récepteur TGF- β R1, en particulier SMAD2 et SMAD3. Ces derniers forment ensuite un complexe hétéromérique avec SMAD4, qui est transloqué vers le noyau de la cellule. Une fois dans le noyau, les SMADs agissent comme des facteurs de transcription capables de se lier directement à l'ADN ou en collaboration avec d'autres facteurs tels que FAST-1, AP-1, SP-1 ou CBP/P300. La régulation génique induite par TGF- β implique plusieurs gènes clés tels que l'inhibiteur de l'activateur plasminogène 1P (AI -1), les inhibiteurs de kinase cycline dépendante P15 et P21 ainsi que les facteurs de transcription homéotiques mix2. Les différents types de SMADs peuvent rivaliser pour la liaison au récepteur TGF- β R1 afin d'éviter une surexpression inappropriée des gènes cibles. Cette compétition permet donc une régulation fine du signal grâce à la prévention simultanée d'une activation excessive ou insuffisante (Javelaud et Mauviel 2004).

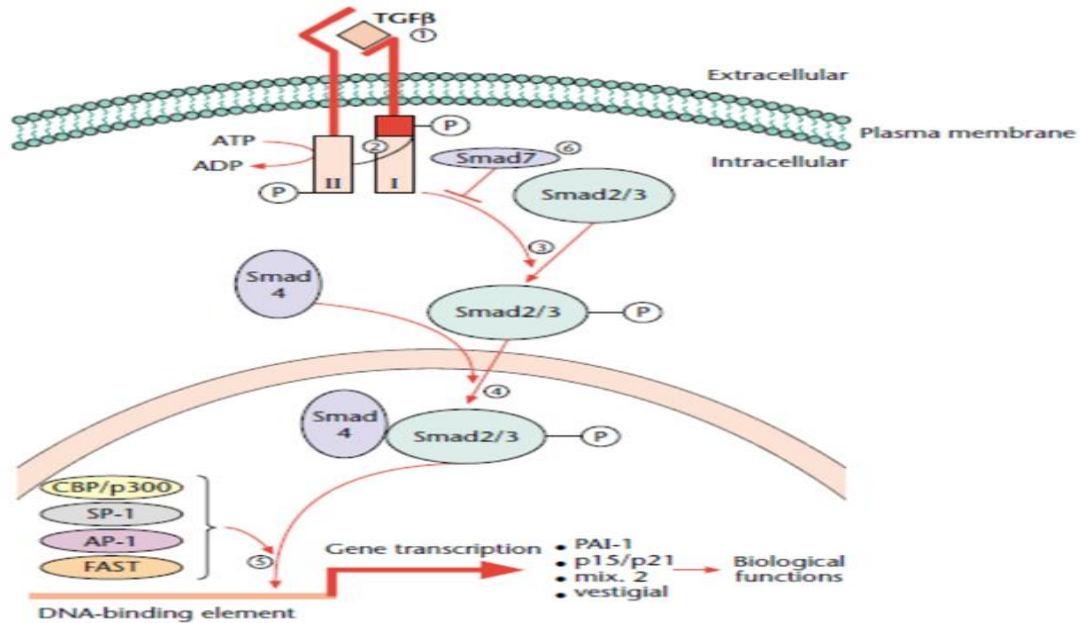


Figure 9: La voie de signalisation TGF β (Javelaud et Mauviel 2004)

4. Polymorphisme nucléotidique simple

4.1. Généralité

Les SNP, ou polymorphismes nucléotidiques simples, sont des variations génétiques courantes qui consistent en un changement d'une seule base dans une séquence d'ADN. Ces variations peuvent se produire dans des régions codantes, non-codantes ou intergénomiques. Ces SNP fonctionnels ont un impact sur la diversité génétique. Lorsqu'ils se produisent dans des régions codantes, ils peuvent entraîner des modifications biologiques au niveau de la protéine codée. En revanche, lorsque les SNP se produisent dans des régions non-codantes, ils peuvent affecter la régulation spécifique de l'expression génique selon l'allèle porté par l'individu (Nassiri et al. 2013).

4.2. Types de polymorphisme nucléotidique

Les SNPs peuvent être classés en deux types différents: les SNPs synonymes et les SNPs non synonymes.

4.2.1. Les SNPs synonymes

les polymorphisme nucléotidique synonymes c'est le résultat d'une mutation par la

substitution d'un codon par un autre qui peut donner le même acide aminé, des études montrent que le polymorphisme synonyme capable de changer l'épissage d'ARNm ainsi que sa stabilité et sa structure de plus que ce polymorphisme collaboré à l'émergence des pathologies restent encore inconnus (**Korzeniewski et al., 2013**).

4.2.2. Les SNPs non synonymes

Les SNP non-synonymes sont des mutations qui résultent d'une substitution entraînant l'ajout d'un acide aminé différent ou un arrêt prématuré de la traduction, ce qui induit une modification dans la structure du polypeptide. Ces SNPs peuvent ainsi générer des perturbations moléculaires, cellulaires et physiologiques significatives (**Korzeniewski, Hofman, et Brest 2013**).

4.3. Avantage des polymorphismes nucléotidiques

Les SNPs sont la forme la plus fréquente de variation génétique. Il existe plus de 13000 SNPs associés à des pathologies différentes, cette variation doit être située à un endroit spécifique du génome et se manifeste sur une proportion supérieure à 1 % de la population pour être caractérisée comme un SNP, donc en concluons que la position des SNPs c'est l'un des avantages très utiles. Les variabilités des SNPs sont découvertes sur le génome qui donne une grande importance pour la découverte de gènes dans une pathologie (**Kamiza et al., 2018**).

4.4. Le polymorphisme nucléotidique dans le cancer colorectal

Le polymorphisme nucléotidique peut être associé à une augmentation du risque de cancer colorectal (CCR). L'identification des SNPs présente donc une importance majeure, car les variations génétiques associées aux SNPs peuvent être utilisées comme marqueurs pour le dépistage et la planification du traitement du CCR (**Nassiri et al. 2013**).

Une nouvelle étude a montré que les SNPs liés à une prédisposition au cancer colorectal, plus récemment une méta-analyse de 15 études publiées du total 4475 cas et 6373 témoins a excité une participation révélatrice entre les polymorphismes et le risque de CCR (**Wang et al., 2020**).

4.5. Analyse *in silico* des polymorphismes nucléotidiques

L'analyse *in silico* des polymorphismes nucléotidiques est une méthode de recherche qui permet d'étudier l'impact potentiel des variations génétiques à un niveau moléculaire et

cellulaire. Cette approche utilise des techniques informatiques pour examiner les données génomiques et identifier les SNP ou autres mutations associées à certaines maladies. Les analyses *in silico* offrent plusieurs avantages par rapport aux méthodes traditionnelles telles que les tests *in vivo* ou *in vitro*. Elles sont plus rapides, moins coûteuses et peuvent être utilisées pour prédire avec précision l'effet potentiel d'un SNP sur une protéine donnée, ce qui peut aider à mieux comprendre les liens entre variations génétiques et maladies humaines (Pogodina 1975).

5. Modifications post traductionnelles

5.1. Généralités

Une modification post traductionnelle (PTM) est une modification de nature chimique covalente d'une protéine affectent la structure des acides aminés qui composent les protéines et régulent leur fonction. les PTM modifient tous les paramètres qui nécessaire a la fonction d'une protéine, son activité, sa localisation, sa stabilité. une modification post traductionnelle c'est-à-dire une protéines après le libération par les ribosomes (Poulard et al., 2010).

5.1.1. La méthylation

la méthylation c'est une modification poste traductionnelle biochimique qui permet l'ajout d'un groupement méthyles sur un azote du l'arginine ou la lysine. la lysine peut méthylée une fois, deux fois, jusqu'à trois fois et l'arginine peut être méthylée au max deux fois. la méthylation des protéines joue un divers rôle dans la régulations des fonctions cellulaire en conséquent une dérégulations qui est impliqué dans divers pathologies (Righini et al., 2008).

5.1.2. La phosphorylation

La phosphorylation est un processus biologique assuré par des enzymes de type kinases qui ajoutent un groupement phosphate sur une molécule, en utilisant l'énergie fournie par l'hydrolyse d'une molécule d'ATP pour libérer une molécule d'ADP. Ce mécanisme permet de réguler l'action biologique des protéines cibles. C'est une modification post-traductionnelle commune effectuée sur les protéines, et elle implique la fixation d'un groupement phosphate sur certains acides aminés spécifiques. Cette modification peut avoir plusieurs effets différents sur la structure et/ou la fonctionnalité de la protéine cible. Des études ont montré que des variations dans les niveaux ou le motif de phosphorylation pouvaient être associées à

différentes maladies comme les cancers (**Righini et al. 2008**).

5.1.3. L'ubiquitination

L'ubiquitination est une modification post-traductionnelle biochimique qui implique plusieurs étapes pour permettre la fixation covalente d'une protéine ubiquitine sur une lysine accepteuse de la protéine substrat. Cette modification peut avoir plusieurs fonctions, dont la dégradation de la protéine substrat par le protéasome. La voie d'ubiquitination commence par l'activation d'un conjugué ubiquitine-enzyme E1, qui transfère ensuite l'ubiquitine à un conjugué enzyme E2 spécifique. Enfin, l'énergie fournie par l'ATP est utilisée pour catalyser le transfert de l'ubiquitine du conjugué E2 vers le résidu lysine accepteur de la protéine cible. Cette modification joue un rôle crucial dans divers processus biologiques tels que les niveaux de signalisation cellulaire, mais elle est surtout connue pour son implication dans le contrôle qualité des protéines en marquant celles qui sont endommagées ou non fonctionnelles pour leur destruction via le système protéasomal. Les recherches ont montré que des anomalies dans ce processus pouvaient être associées à différentes maladies humaines telles que les cancers et les maladies neurodégénératives (**Sun 2003**).

5.1.4. La carbonylation

La carbonylation est un processus chimique ou une contre-révolution dans laquelle une molécule de monoxyde de carbone, capable d'être seules ou avec d'autres réactifs a pour insérer une molécule d'un substrat organique. certaines carbonylations présentes un ample intérêt c'est l'hydroformylations et la réaction de Reppe (**Peng et al., 2019**).

5.1.5. tyrosines nitration

la tyrosine est un acide aminé dont l'énantiomère L est l'un des 22 acide aminé qui est sur les ARNm, cet acide aminé capable d'être produire une quantité suffisante par l'organisme qui nécessitent un rapport alimentaire. la nitrations des tyrosines collabore à la synthèse des catécholamines tels que l'adrélines, la noradréline, la dopamine et la dopa. d'autre part est un précurseur de la mélanine sous le rôle d'être un régulateur physiologique et amplificateur des processus toxiques qui liés au stress oxydant (**Radi, 2013**).

5.1.6. S-nitrosylation

La nitrosylation est la liaison covalente de monoxyde d'azotes sur une molécule pour

former un nitroso. Une classe singulière de nitrosylation qui joue un rôle très important en biologie moléculaire qui est engagé dans les groupes thiol-SH pour créer des S –nitrosothiols –SNO. il -s’accomplit de la S-nitrosylation, cette dernière est une modification post traductionnelle des plusieurs protéines sur un des résidus de cystéine, sont intéressés et impliqués dans les processus différents telles que la signalisation cellulaire (**Plenchette et al., 2016**).

5.2. Prédiction *in silico* des modifications post traductionnelles

Les modifications post-traductionnelles (PTM) des protéines c’est l’un des nouvelles étapes de la biosynthèse des protéines qui est référent aux modifications chimiques des protéines. Les PTM sont des changements aux niveaux de la chaîne polypeptidique donc c’est détérioration vont du clivage enzymatique dans les liaisons peptidiques. D’autre part, ces modifications chimiques de la chaîne polypeptidique après son anabolisme être tenues élargir la gamme des structures et les propriétés des acides aminés (**Delcourt 2007**).

6. Problématique et objectif

6.1. Problématique

En 2022, Zouaïmia Nahla et Senouci Meriem Radia ont réalisé une analyse *in silico* du polymorphisme nucléotidique ayant des conséquences fonctionnelles sur TGFBR2 et SMAD7 afin d’identifier les variations délétères à risque élevé qui peuvent être impliquées dans le cancer colorectal ou d’autres pathologies et explorer des nouvelles perspectives dans le diagnostic et la prévention de la maladie. Dans ce travail, nous allons étudier *in silico* l’effet de ces variations sur la structure tridimensionnelle de TGFBR2 et SMAD7 et leurs modifications post traductionnelles.

6.2. Objectif

L’objectif principal de notre étude c’est de réaliser une prédiction *in silico* de l’effet des SNPs probablement délétères sur les modifications post traductionnelles et la structure tridimensionnelle de TGF-βR2 et SMAD7.

6.3. But

Le but de notre étude est d’identifier des SNPs à conséquences fonctionnelles sur TGFBR2 et SMAD7 et potentiellement impliqués dans le cancer colorectal.

Chapitre II : Matériels et méthodes

1. Plan de l'étude

Les travaux de mémoire de fin d'études réalisés en 2022 par **Snouci Meriem Radia** et **Zouaimia Nahla Ibtissam**, sous la direction du **Dr. Hadjidj Zeyneb**, ont permis d'identifier 13 variations potentiellement délétères dans TGFBR2 (K277N,R279H,T334M,R338W,R338Q,R338P,R338L,G356S,G356D,G394W,R448G, R520L,R522Q) ainsi que 12 autres dans SMAD7 (R426W ,N424K, P416L,R409H, R409L,Y407S, G402V,S336C,R314Q,R200Q,K175T, P166R).

Ce travail se poursuit actuellement avec deux axes principaux : la recherche des modifications post-traductionnelles engendrées ou supprimées par ces variations et la modélisation tridimensionnelle des variantes protéiques afin d'identifier les différences structurelles susceptibles d'être causées par ces variations.

Ces recherches ont été menées au sein du Laboratoire Biologie Moléculaire Appliquée et Immunologie dirigé par le Professeur ARIBI Mourad.

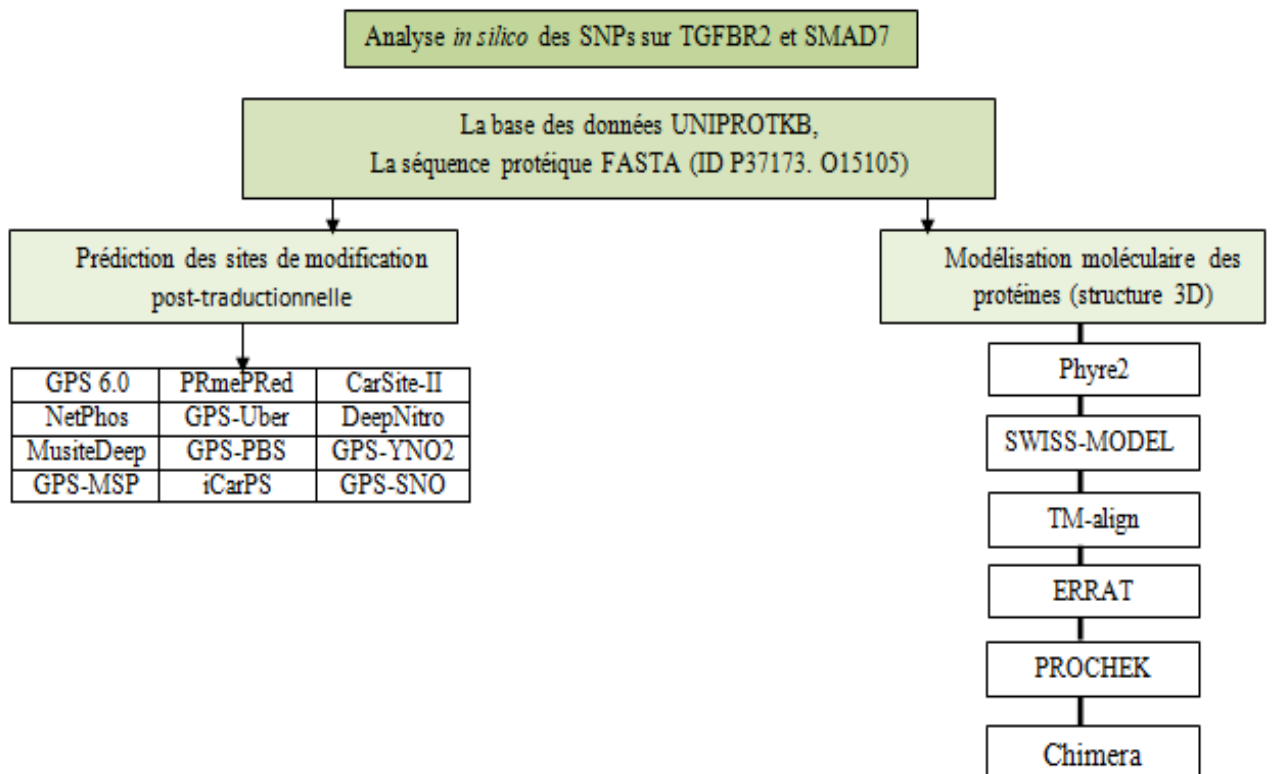


Figure 1: study design

Chapitre II : Matériels et méthodes

2. Collecte de données et sélection de séquences de protéines

Nous avons utilisé la base de données UniProt (<https://www.uniprot.org/>) pour récupérer les séquences primaires des protéines TGFBR2 et Smad7 en forma FASTA (numéro d'accèsion : P37173, O15105)

Afin d'obtenir les séquences mutées des protéines étudiées, nous avons tout d'abord obtenu les séquences primaires à partir de la base de données UniProt (<https://www.uniprot.org/>). Ensuite, ces séquences ont été modifiées pour inclure les mutations spécifiques que nous souhaitons étudier. Les acides aminés variant ont ainsi été substitués par leur équivalent muté dans chaque protéine.

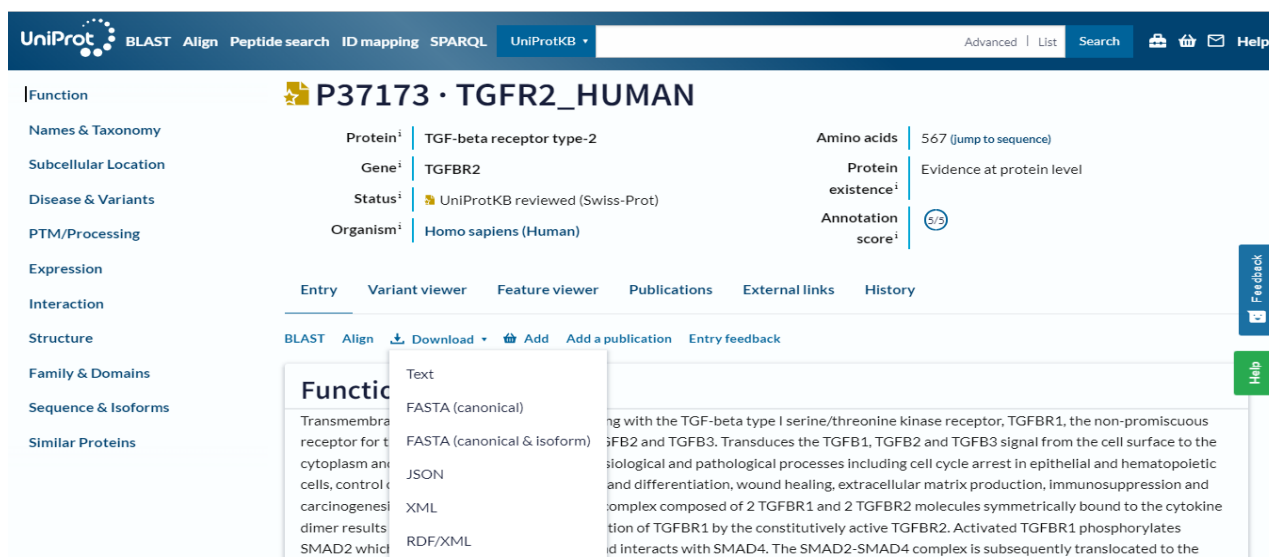


Figure 2: Recherche des séquences sur la base de données UniProt.

3. Prédiction des sites de modification post-traductionnelle

Différents types de modifications post-traductionnelles ont été étudiées, notamment la **phosphorylation** par GPS 6.0 (<http://gps.biocuckoo.cn/>), NetPhos 3.1 (<https://services.healthtech.dtu.dk/services/NetPhos-3.1/>) et MusiteDeep (<https://www.musite.net/>). La **méthylation** par GPS-MSP 1.0 (<http://msp.biocuckoo.org/>) et PRmePRed (<http://bioinfo.icgeb.res.in/PRmePRed>). Pour l'**ubiquitination** par GPS-Uber (<http://gpsuber.biocuckoo.cn/>), et également le MusiteDeep. Pour prédire les **domaines de liaison aux phosphoprotéines** nous avons utilisé GPS-PBS (<http://pbs.biocuckoo.cn/>). Les sites de **carbonylation** ont été prédits par iCarPS (<http://lin-group.cn/server/iCarPS/webServer.html>) et CarSite-II (<http://47.100.136.41:8081/>). Les logiciels GPS-YNO2 1.0 (<http://yno2.biocuckoo.org/>) et DeepNitro

Chapitre II : Matériels et méthodes

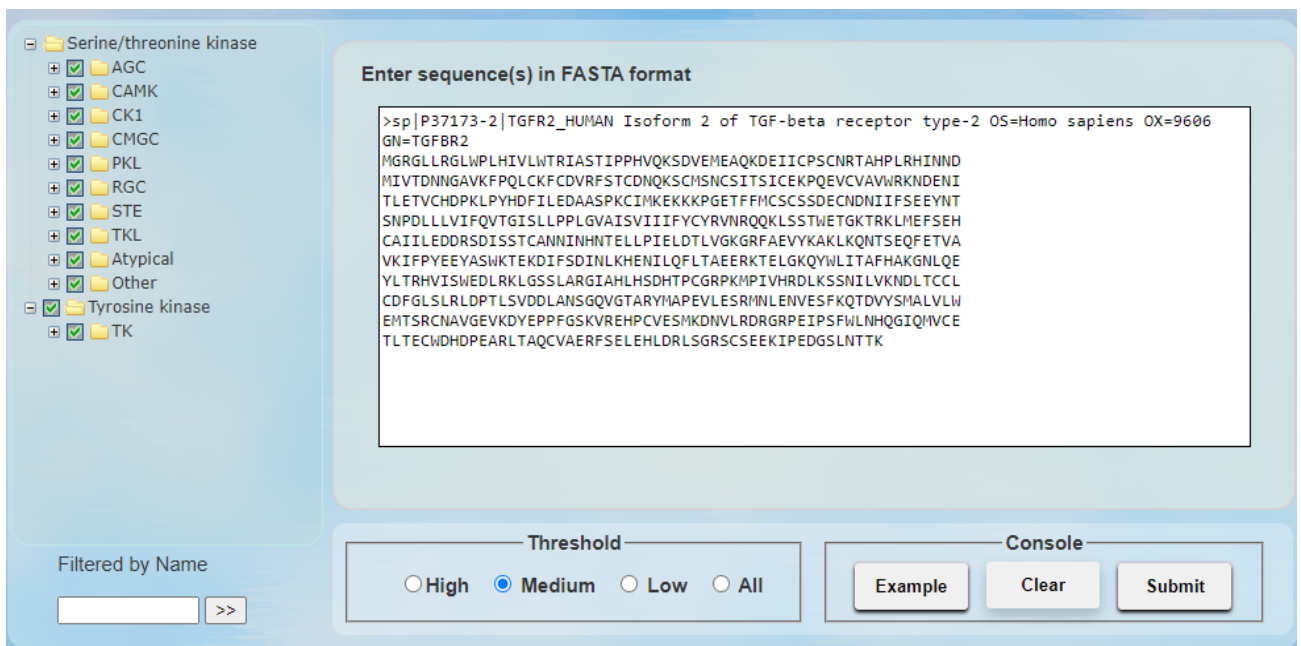
(<http://deepnitro.renlab.org/webserver.html?fbclid=IwAR3IXaPoaCnCqNy0iaZjEaJ5uEiyeo3hxzGiGixz5hvEx7P6GwLsxO36Xbc#result>) ont été utilisé pour la prédiction des **sites de nitration de la tyrosine**. Et le GPS-SNO 1.0 (<http://sno.biocuckoo.org/>) pour les **sites de S-nitrosylation**.

3.1. Prédiction des sites de phosphorylation

✓ GPS 6.0

C'est un outil de prédiction des modifications post-traductionnelles (PTM) des protéines, avec pour objectif principal la découverte de sites de phosphorylation sur les acides aminés tels que la sérine, la thréonine et la tyrosine. Le fonctionnement de cet outil implique l'utilisation d'une séquence de protéine au format FASTA. Les éléments de déclaration doivent être organisés en fonction du score pour examiner chaque site potentiellement modifié et déterminer sa véracité (Chen et al., 2023).

Site web: <http://gps.biocuckoo.cn/>



The screenshot displays the GPS 6.0 web interface. On the left, there is a sidebar with a tree view of kinase types: Serine/threonine kinase (expanded) with sub-categories AGC, CAMK, CK1, CMGC, PKL, RGC, STE, TKL, Atypical, and Other; and Tyrosine kinase (expanded) with sub-category TK. The main area is titled 'Enter sequence(s) in FASTA format' and contains a text box with a FASTA sequence for TGF-beta receptor type-2. Below the text box, there is a 'Threshold' section with radio buttons for High, Medium (selected), Low, and All. To the right of the threshold section is a 'Console' section with buttons for Example, Clear, and Submit. At the bottom left, there is a 'Filtered by Name' input field with a search icon.

Figure 3: Prédiction des sites de phosphorylation par GPS 6.0

✓ NetPhos 3.1

C'est un outil de prédiction des sites de phosphorylation pour les protéines. Son algorithme recherche les sites de phosphorylation des acides aminés tels que la sérine, la thréonine et la tyrosine.

Pour obtenir une liste des sites de phosphorylations, il suffit tout d'abord d'entrer une séquence de protéine au format FASTA et d'adapter les éléments de déclaration en

Chapitre II : Matériels et méthodes

conséquence (Biswas et al., 2010).

Site web: <https://services.healthtech.dtu.dk/services/NetPhos-3.1/>

The screenshot shows the 'Submission' tab of the NetPhos 3.1 web interface. It includes a navigation bar with 'Submission', 'Instructions', 'Output format', 'PhosphoBase', and 'Downloads'. The main content area is titled 'Submission' and contains the following elements:

- Sequence submission: paste the sequence(s) and/or upload a local file**
- Paste a single sequence or several sequences in *FASTA* format into the field below:
- A text input field containing a FASTA sequence:

```
CDFGLSRLDPTLSVDDLANSQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLW  
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWLNLHQGIQMVCE  
TLTECWDHDPPEARLTAQCAERFSELEHLDRLSGRSCSEEKIPEDGSLNTTK
```
- Submit a file in *FASTA* format directly from your local disk:
- A file selection button labeled 'Choisir un fichier' and the text 'Aucun fichier choisi'.
- Residues to predict** with radio buttons for Serine, Threonine, Tyrosine, and 'all three' (selected).
- For each residue display only the best prediction** with an unchecked checkbox.
- Display only the scores higher than** with a text input field containing '0.5'.
- Output format** with radio buttons for 'classical' (selected) and 'GFF'.
- Generate graphics** with a checked checkbox.
- 'Submit' and 'Clear fields' buttons.
- Restrictions:**
At most 2000 sequences and 200,000 amino acids per submission; each sequence not less than 15 and not more than 4,000 amino acids.

Figure 4: Prédictions des sites de phosphorylation par NetPhos 3.1

✓ MusiteDeep

Le logiciel MusiteDeep est un outil informatique qui permet de prédire les sites de modification post-traductionnelle des protéines, en mettant l'accent particulièrement sur la phosphorylation et l'acétylation. Il utilise des réseaux neuronaux profonds et des algorithmes d'apprentissage automatique avancés pour analyser les séquences protéiques et prédire les emplacements potentiels de modification (Wang et al., 2017).

Site web: <https://www.musite.net/>

The screenshot shows the MusiteDeep web interface. It features a navigation bar with 'HOME', 'Job History', 'Download', 'API', 'Contact', 'Help', and 'Musite(SVM version)'. The main content area is titled 'MusiteDeep: a deep-learning framework for protein post-translational modification site prediction'. It includes the following elements:

- Statistics: 5337 unique visitors, 1756725 processed proteins, 319350658 processed amino acids.
- A 'Submit your sequence(s)' button.
- A dropdown menu to 'Please select a prediction model:' with 'Phosphorylation (S,T)' and 'Phosphorylation (Y)' selected.
- A text input field for 'Paste input FASTA sequence(s) (up to 10 sequences or 5000 residues in total) in the area below:' with a 'Load example FASTA' link.
- A text area containing a FASTA sequence:

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606 GN=TGFR2  
MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRHTAHLRHHND  
MIVTDNNGAVKFPQLCKFCDFRSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI  
TLETVCHDPKLPYHDFILEDAASPKCIMKEKKKPGETFFMCSCSSDECDNIIIFSEYNT  
SNPDLVVIFQVTGISLLPPLGVAISVIIIIFCYRVNRQQKLSSTWETGTRKLMFSEH  
CAIILEDDRSDISSTCANNINHTELLPIELDTLVGGRFAEVYKAKLKQNTSEQFETVA  
VKIFPYEYASWKTEKDFSDINLKHENILQFLTAERKTELQKQYWLITAFHAKGNLQE  
YLTRHVISWEDLRKLGSSLARGIAHLHSDHTPCGRPKMPIVHRDLKSSNIIKNDLTCCL  
CDFGLSRLDPTLSVDDLANSQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLW  
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWLNLHQGIQMVCE  
TLTECWDHDPPEARLTAQCAERFSELEHLDRLSGRSCSEEKIPEDGSLNTTK
```
- A link 'For larger job, upload a FASTA file'.
- A 'Start prediction' button.

Figure 5: Prédictions des sites de phosphorylation et d'ubiquitination par MusiteDeep

3.2. Prédiction des sites de méthylation

✓ GPS-MSP 1.0

Cet outil est utilisé pour la prédiction de sites de méthylation de la lysine et l'arginine dans la protéine, il se base sur les motifs des séquences conservées dans les protéines similaires.

Le GPS-MSP se base sur les prédictions informatiques et les résultats doivent d'abord être validés empiriquement pour confirmer la présence des sites de méthylation dans la protéine (Wankun Deng et al).

Site web: <http://msp.biocuckoo.org/>

Methylation types: All K.mono K.di K.tri K.all R.mono R.s.di R.a.di R.all R.di

Enter sequence(s) in FASTA format

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606 GN=TGFB2
MGRGLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHLRHNND
MIVTDNNGAVKFPQLCKFCDFRSTCDNQKSCMNSCSITICEKPEVCVAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKIMKEKKPGETFFHCSCSSDECNDNIIIFSEYNT
SNPDLLLVIFQVTGISLPPPLGVAISVIIIFYCYRVNRQKLSSTWETGKTRKLMFSEH
CAIILEDSDISSTCANNINHTELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEYASWKTEKIDIFSDINLKHENILQFLTAERKTELKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSSLAGIAHLHSDHTPCGRPKMPIVHRDLKSSNII LVKNDLTCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLESRMNLENVESFKQTDVYSHALVLW
EMTSRNCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSFWLNHQGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEKIPEDGSLNTTK
```

Or, upload file(<2M) Aucun fichier choisi

Threshold: High Medium Low All

Organisms: All *H. sapiens* *M. musculus* *S. cerevisiae* *R. norvegicus* *L. interrogans* *S. solfataricus*
 D. vulgaris

Figure 6: Prédiction des sites de méthylation par GPS-MSP 1.0

✓ PRmePRed

PRmePRed est un logiciel spécialement conçu pour prédire les sites de méthylation des protéines, en se concentrant principalement sur les résidus de lysine. Il utilise des algorithmes d'apprentissage automatique avancés et des modèles prédictifs basés sur des caractéristiques spécifiques des séquences protéiques pour réaliser ses prédictions. Grâce à cette approche, PRmePRed offre une méthode précise et fiable pour prédire les emplacements potentiels de méthylation des protéines, ce qui permet d'améliorer notre compréhension de leur régulation et de leurs fonctions biologiques.

Site web: <http://bioinfo.icgeb.res.in/PRmePRed>

PRmePred
Outil de prédiction de la méthylation de l'arginine protéique

Maison À propos Base de données Aider Contactez-nous

Collez ou téléchargez vos séquences de protéines (jusqu'à 25) au format **FASTA**.
La limite maximale pour le téléchargement de fichiers est de 2 Mo. Conseils

```
YLTRHVISWEDLRKLGSSLRGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCCL
CDFGLSLRLDPTLSVDDLANSGQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVL
W
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSFWLNHQIQMV
CE
TLTECWDHDPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNTTK
```

Charger les séquences de démonstration
Effacer les séquences

(OU) Téléchargez le fichier de séquence(s) au format **FASTA** :
(jusqu'à 2 Mo)

Choisir un fichier Aucun fichier choisi

Sélectionnez Seuil : Valeur de seuil

Exécuter la prédiction Reset

Figure 7: Prédiction des sites de méthylation par PRmePred

3.3. Prédiction des sites d'ubiquitination

✓ GPS-Uber

Cadre d'apprentissage hybride pour la prédiction des sites d'ubiquitination de la lysine généraux et spécifiques à E3.

Site web: <http://gpsuber.biocuckoo.cn/>

Selection option

Ubiquitination

Filtered by Name >>

Enter sequence(s) in FASTA format

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TFGBR2
MGRGLRGLNPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHLRHNND
HIVTDNNGAVKFPQLCKFCVRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKCIHKEKKKPGETFFMCSGSSDECNDNIIIFSEEYNT
SNPDLLLVIFQVGTGSLPPLGVAISVIIIFYCYRVNRQQKLSSTMETGTRKLNHFSEH
CAIILEDSDISSTCANNINHNTELPIELDTLVGKGRFAEVYKAKLKONTSEQFETVA
VKIFPPEEYASUKTEKDFISDINKHENILQFLTAERKTELKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSSLRGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCCL
CDFGLSLRLDPTLSVDDLANSGQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSFWLNHQIQMVCE
TLTECWDHDPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNTTK
```

View experimental information

GPS-Uber

Threshold High Medium Low

Console

Figure 8: Prédiction des sites d'ubiquitination par GPS-Uber

3.4. Prédiction des domaines de liaison aux phosphoprotéines

✓ GPS-PBS

L'outil GPS-PBS a été utilisé pour prédire les domaines de liaison aux phosphoprotéines.

Chapitre II : Matériels et méthodes

C'est un cadre d'apprentissage profond pour prédire les sites de phosphorylation qui interagissent spécifiquement avec des domaines de liaison aux phosphoprotéines (Guo et al. 2020).

Site web: <http://pbs.biocuckoo.cn/>

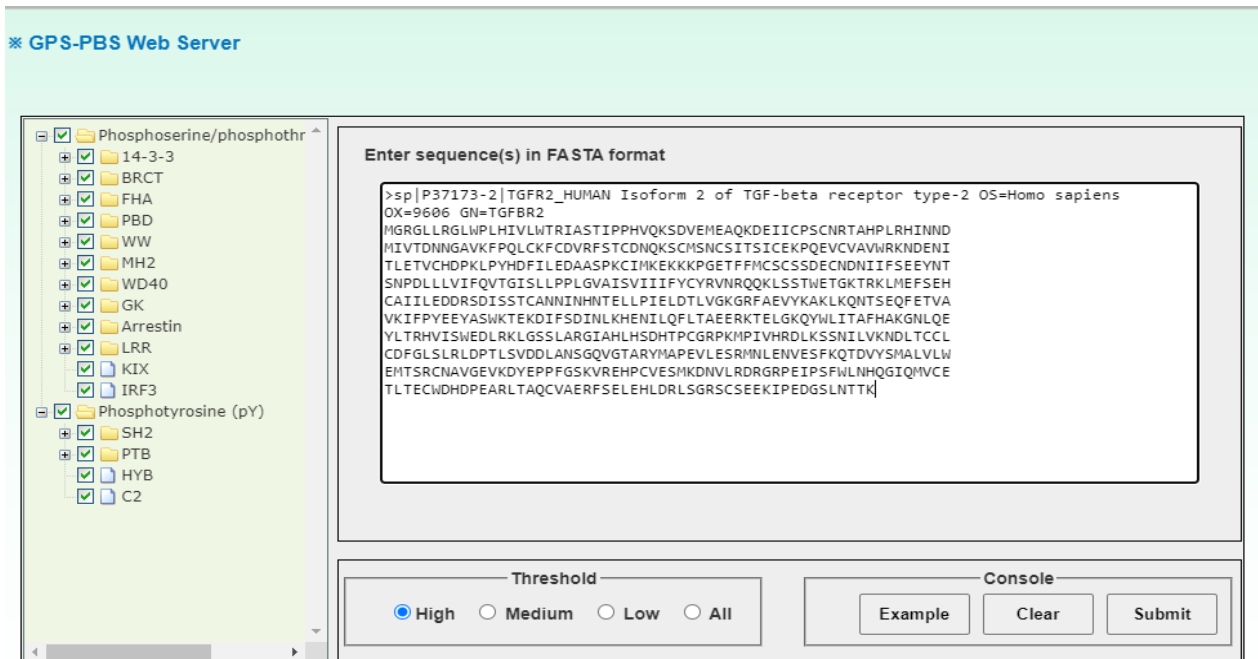


Figure 9: Prédiction des domaines de liaison aux phosphoprotéines par GPS-PBS

3.5. Prédiction des sites de carbonylation

✓ ICarPS

Un outil informatique pour identifier les sites de carbonylation des protéines par de nouvelles caractéristiques codées.

Site web: <http://lin-group.cn/server/iCarPS/webServer.html>

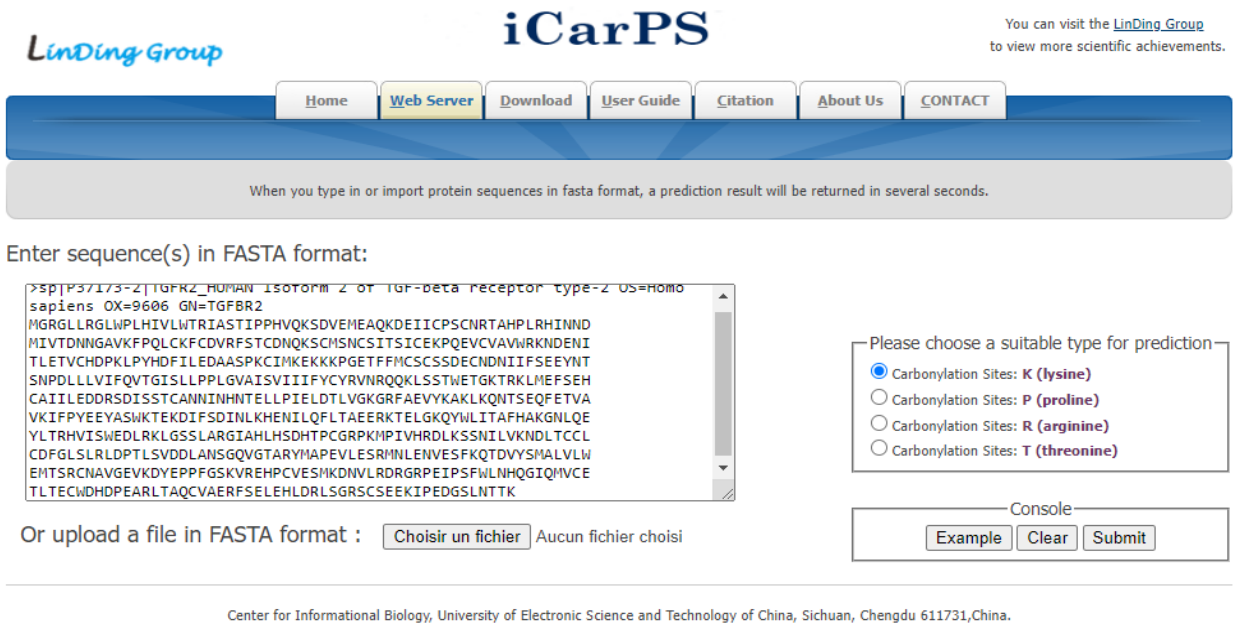


Figure 10:Prédiction des sites de carbonylation par iCarPS

✓ CarSite-II

C'est un algorithme de classification intégré pour identifier les sites carbonylés basé sur des techniques de sous-échantillonnage basées sur la similarité K-means et de sur l'échantillonnage de minorités synthétiques.

Site web : <http://47.100.136.41:8081/>

CarSite-II: an integrated classification algorithm for identifying carbonylated sites based on K-means Similarity undersampling and synthetic minority oversampling technique

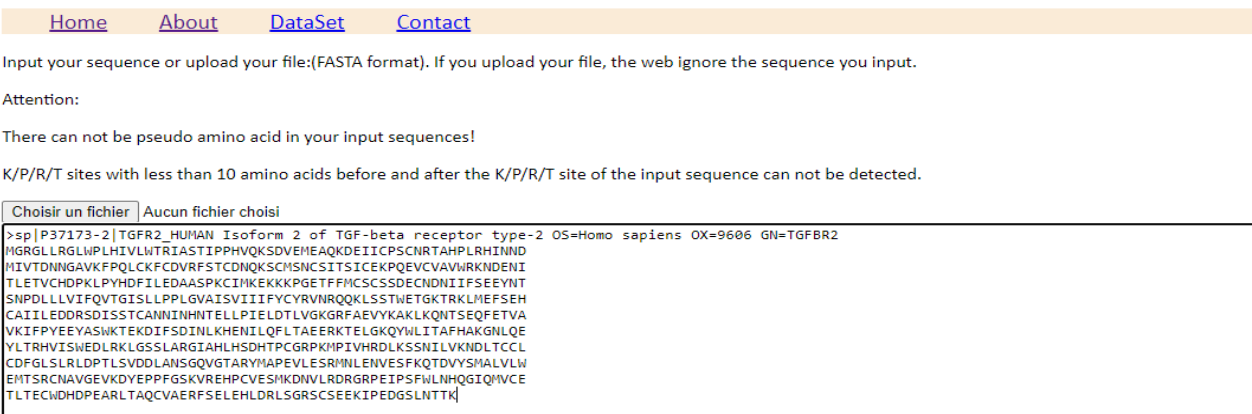


Figure 11: Prédiction des sites de carbonylation par CarSite-II

Chapitre II : Matériels et méthodes

3.6. Prédiction des sites de nitration des tyrosines

✓ GPS-YNO2

Ce logiciel génère un ensemble de prédictions pour les résidus de nitrate dans les protéines humaines (Liu et al., 2011).

Site web: <http://yno2.biocuckoo.org/>

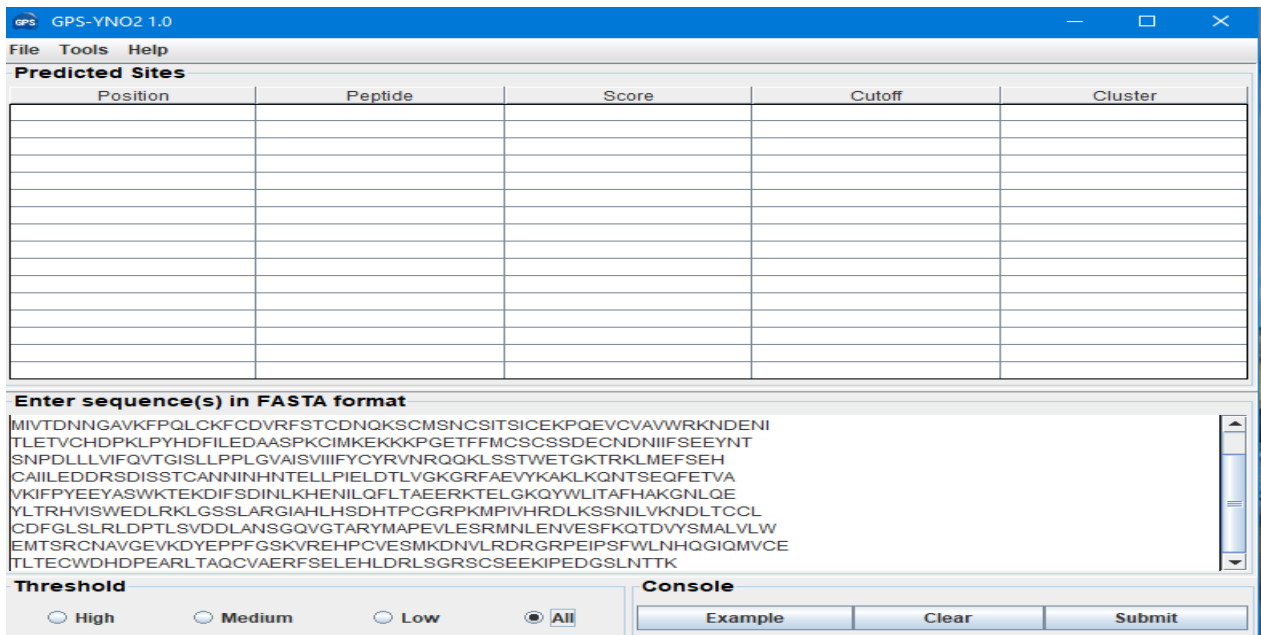


Figure 12: Prédiction des sites de nitration des tyrosines par GPS-YNO2

3.7. Prédiction des sites de S-nitrosylation

✓ DeepNitro

La nitrosylation est une modification post-traductionnelle dans laquelle un groupe nitroso (-NO) est ajouté à un résidu d'acide aminé, généralement une cystéine, et elle joue un rôle crucial dans la régulation des fonctions biologiques. DeepNitro est un logiciel informatique utilisé pour prédire les sites de nitrosylation dans les protéines. Il offre une approche précise et fiable pour prédire ces sites de nitrosylation (Xie et al., 2018).

Site web:

<http://deepnitro.renlab.org/webserver.html?fbclid=IwAR3IXaPoaCnCqNy0iaZjEaJ5uEiyeo3hxyzGiGixz5hvEx7P6GwLsxO36Xbc#result>

Figure 13: prédiction des sites de S-nitrosylation et de nitration des tyrosines par DeepNitro

✓ GPS-SNO

C'est un programme utilisé pour prédire les sites de S-nitrosylation des protéines. Il s'agit d'un outil informatique qui utilise un algorithme modifié basé sur le système GPS (Xue et al., 2010).

Site web: <http://sno.biocuckoo.org/>

Figure 14: Prédiction des sites de S-nitrosylation des protéines par GPS-SNO

4. Modélisation moléculaire

✓ Modélisation par Phyre 2

Le logiciel Phyre 2 est largement utilisé comme une solution informatique essentielle, il est capable de prédire la structure tridimensionnelle d'une protéine inconnue en se basant sur la similarité de séquence avec des protéines dont la structure est déjà connue. Pour améliorer la précision de ses prédictions, Phyre 2 intègre également des informations provenant de diverses bases de données et d'outils de prédiction de structures secondaires (Kelley et al., 2015).

Site web: http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index_advanced

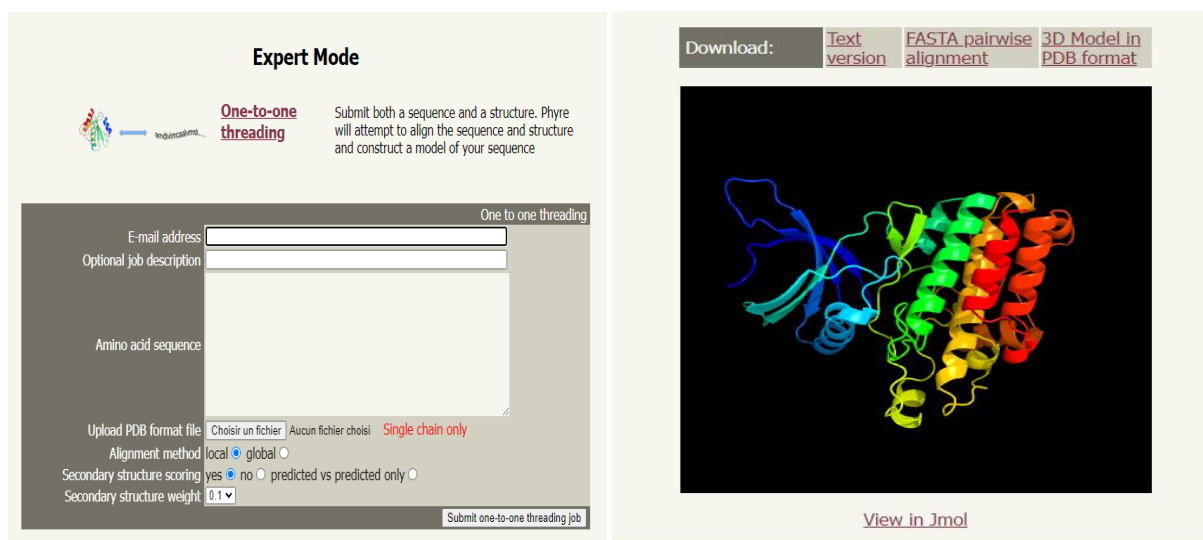


Figure 15: Modélisation des structures tridimensionnelles des protéines par Phyre2

✓ Modélisation par SWISS-MODEL

C'est un serveur de modélisation de structure de protéine en ligne qui requiert les modèles d'homologie pour découvrir la structure 3D des protéines. Swiss-Model contient plusieurs outils de modélisation parmi lequel le SWISS MODEL Work Space qui permet d'acquérir les séquences de protéine et de génère des nouvelles modèles (Schwede et al., 2003).

Site web: <https://swissmodel.expasy.org/>

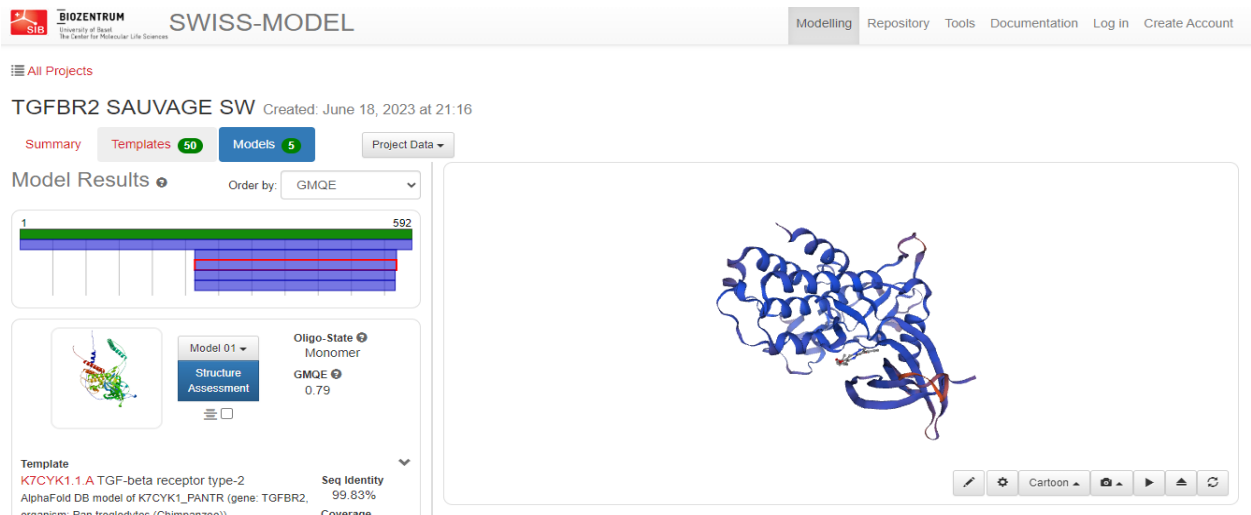


Figure 16: Modélisation des structures tridimensionnelles des protéines par SWISS-MODEL

5. Comparaison entre les structures tridimensionnelles des protéines

✓ Tm-align

Le logiciel TM-align est un outil informatique utilisé dans le domaine de la biologie structurale pour évaluer et comparer la similarité structurale entre deux protéines. Il se base sur l'algorithme TM-align, qui mesure l'alignement optimal entre les atomes des deux structures protéiques et calcule un score des similarités appelé TM-score.

TM-align est capable d'aligner et de superposer les structures tridimensionnelles de protéines, en prenant en compte les déformations flexibles et les différences de taille entre les molécules. Cet outil est particulièrement utile pour détecter des similitudes structurales entre des protéines qui ont des séquences d'acides aminés différentes, mais qui ont une fonction ou une structure globale similaires.

Le logiciel TM-align fournit des informations sur la superposition des structures, les déviations entre les résidus, les régions structurellement conservées et les régions flexibles. Il est utilisé pour l'annotation structurale, l'analyse comparative de structures protéiques, la prédiction de la fonction des protéines et la recherche de relations évolutive entre les familles de protéines (Zhang et Skolnick 2005).

Site web: <https://zhanggroup.org/TM-align/>

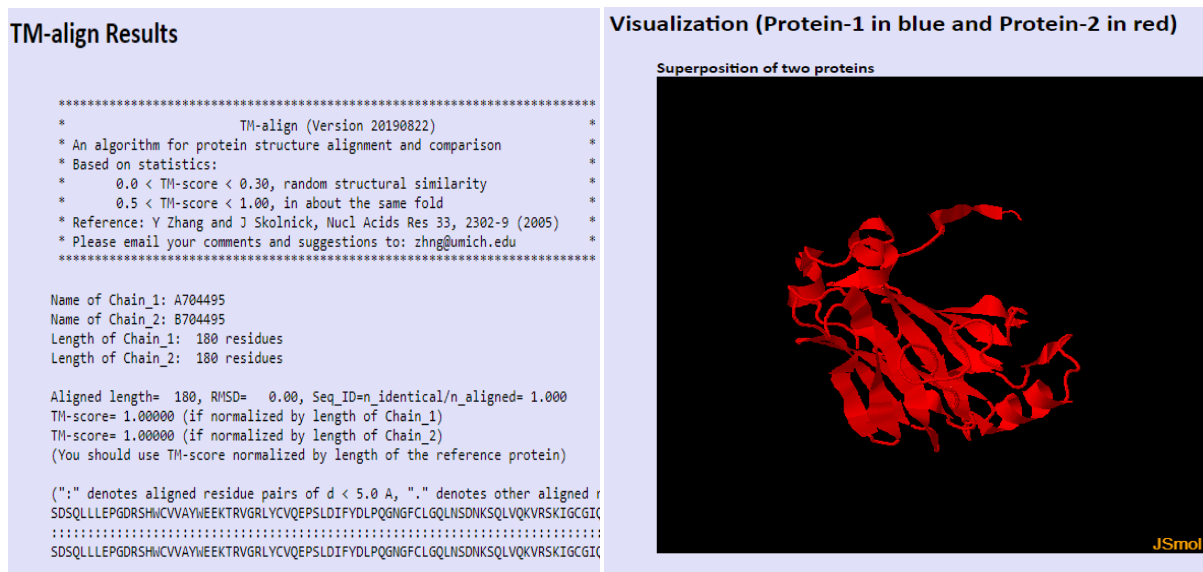


Figure 17: alignement des structures des protéines par TM-align

6. Évaluations de la qualité des structures des protéines modélisées

✓ ERRAT

Le logiciel ERRAT est un outil informatique largement utilisé dans le domaine de la biologie structurale pour évaluer la qualité des structures tridimensionnelles de protéines. En se basant sur une analyse statistique des facteurs d'erreur, ERRAT compare la distribution des résidus dans la structure modélisée avec celle attendue pour une structure de haute qualité.

En utilisant les références de structures protéiques de haute qualité, ERRAT calcule un score d'erreur global qui mesure la correspondance entre la structure modélisée et les références.

Un score ERRAT élevé indique une meilleure correspondance et une meilleure qualité structurale.

L'objectif principal d'ERRAT est de détecter les régions erronées ou mal modélisées dans une structure protéique. Il est capable d'identifier les résidus mal placés, les régions déformées ou les erreurs dans les angles de liaison, fournissant ainsi des informations précieuses pour améliorer la qualité des modèles de structures protéiques (ERRAT – UCLA-DOE Institute s. d.).

Site web: <https://saves.mbi.ucla.edu/>

UCLA-DOE LAB — SAVES v6.0

[← Control panel](#)
[↑ New Job](#)

Job #1378798: TGFBR2_SV_PHYRE2 (2).pdb

[ERRAT](#)
[PROCHECK](#)

ERRAT

Overall Quality Factor

79.3919

[Log](#)
[PostScript](#)
[PDF](#)

Figure 18: Évaluations de la qualité des structures des protéines par ERRAT

✓ PROCHEK

PROCHEK est un outil de vérification de la qualité des structures protéiques. Il permet d'identifier les erreurs, d'améliorer la fiabilité des modèles et de fournir une assurance de la qualité des données structurales utilisées dans les études de biologie structurale et de conception de médicaments (Laskowski, MacArthur, et Thornton 2006).

Site web: <https://saves.mbi.ucla.edu/>

[ERRAT](#)
[PROCHECK](#)

PROCHECK

Out of 8 evaluations

- Errors: 1
- Warning: 2
- Pass: 5

The evaluations are the '+' (Warning) and '' (Error) in the summary. The categories on the left do not always correspond in number due to PROCHECK output documents.*

Summary
Ramachandran plot Warning
All Ramachandrans Pass
Chi1-chi2 plots Pass
Main-chain params
Side-chain params Error
Residue properties Pass
Bond len/angle Pass
Mis bond lengths

```

+-----<<< P R O C H E C K   S U M M A R Y >>>-----+
| /var/www/SAVES/Jobs/1378798/saves.pdb   1.5                304 residues |
| Ramachandran plot:  95.3% core   4.7% allow   0.0% gener   0.0% disall |
+ All Ramachandrans:  1 labelled residues (out of 302)
| Chi1-chi2 plots:    0 labelled residues (out of 194)
| Side-chain params:  5 better     0 inside     0 worse
* Residue properties: Max.deviation:   2.3                Bad contacts:   11
+ Bond len/angle:    4.6      Morris et al class:  1  1  2
| G-factors          Dihedrals:   0.30  Covalent:   0.27  Overall:   0.30
| Planar groups:    100.0% within limits  0.0% highlighted
+-----+
+ May be worth investigating further.  * Worth investigating further.
    
```

Figure 19: Évaluations de la qualité des structures des protéines par PROCHEK

Chapitre II : Matériels et méthodes

7. Comparaison et superposition des structures 3D

✓ Chimera

Chimera est un système d'imagerie et d'analyse moléculaire hautement évolutif et interactif. Chimera peut lire les structures moléculaires et les données associées plusieurs formats, afficher les structures dans différentes représentations et générer des images et des animations de haute qualité adaptées à la publication et à la présentation. De plus, Chimera fournit des outils pour : afficher des cartes de densité et analyser les données microscopiques ; utiliser des informations de symétrie pour montrer les structures d'ordre supérieur ; afficher plusieurs alignements de séquences, avec diaphonie entre les séquences et les structures ; et permettre l'analyse des trajectoires de dynamique moléculaire et des résultats d'amarrage (Vincent Zoete, 2007).

Site web: <https://www.cgl.ucsf.edu/chimera/download.html>



Figure 20:Modélisation des structures tridimensionnelles des protéines par Chimera

Chapitre III : Résultats et interprétation

Chapitre III : Résultats et interprétation

1. Résultats de la Prédiction des sites de modification post-traductionnelle

Les modifications post-traductionnelles (PTMs) affectent la structure des acides aminés composant les protéines et régulent ainsi leurs fonctions et leurs activités.

A l'aide des logiciels GPS-MSP et PRmePRed, Nous avons prédit 3 sites de méthylation dans la protéine TGFBR2 sauvage (R448. R520. R522) et qui ont été supprimés lorsque nous les avons remplacés par des nsSNPs (R448G. R520L. R522Q). Et nous n'avons pas obtenu des résultats positifs pour la protéine SMAD7.

Les logiciels GPS-Uber et MusiteDeep ont trouvé un site d'ubiquitination au niveau dans un résidu de lysine dans TGFBR2 type sauvage à la position 277. Les résultats sont nuls chez le SMAD7.

En utilisant GPS6.0 et NetPhos, nous avons prédit les sites de phosphorylation au niveau des résidus sérine (S), la thréonine (T) et la tyrosine (Y) de certaines protéines, mais rien n'a été trouvé.

De même pour d'autres prédictions tels que la carbonylation, sites de nitration des tyrosines, les sites de S-nitrosylation et les domaines de liaison aux phosphoprotéines les résultats sont toujours négatifs.

Tableau 1 : Résultats de la prédiction des sites de modification post-traductionnelle

<i>Phosphorylation</i>										
Protéine	Position d'AA	Sauvage			SNP ID	Position d'AA	Mutée			+/-
		GPS 6.0	NetPhos	Musite Deep			GPS 6.0	NetPhos	Musite Deep	
TGFBR2	T334	0.0117	0.535	0.09	rs202168735	T334M	Nulle	Nulle	Nulle	-
	G356	Nulle	Nulle	Nulle	rs759215875	G356S	0.0002	Nulle	Nulle	-
SMAD7	Y407	Nulle	Nulle	Nulle	rs755148320	Y407S	0.0290	0.574	0.05	-
	S336	0.1015	Nulle	Nulle	rs944090062	S336C	Nulle	Nulle	Nulle	-
	K175	Nulle	Nulle	Nulle	rs762811026	K175T	0.0002	Nulle	Nulle	-
<i>Méthylation</i>										
Protéine	Position d'AA	Sauvage		SNP ID	Position d'AA	Mutée		+/-		
		GPS-MSP 1.0	PRmePRed			GPS-MSP 1.0	PRmePRed			
TGFBR2	K277	0.0605	Nulle	rs1160016010	K277N	Nulle	Nulle	-		
	R279	0.0870	0.751386	rs751948498	R279H	Nulle	Nulle	-		
	R338	0.0929	Nulle	rs55751315	R338W	Nulle	Nulle	-		
	R338	0.0929	Nulle	rs200361387	R338Q	Nulle	Nulle	-		
	R338	0.0929	Nulle	rs200361387	R338P	Nulle	Nulle	-		

Chapitre III : Résultats et interprétation

	R338	0.0929	Nulle	rs200361387	R338L	Nulle	Nulle	-
	R448	0.3667	0.804149	rs771094273	R448G	Nulle	Nulle	+
	R520	0.4186	0.799707	rs1305853447	R520L	Nulle	Nulle	+
	R522	0.4119	0.77732	rs200958264	R522Q	Nulle	Nulle	+
SMAD7	R426	0.0209	Nulle	rs770772156	R426W	Nulle	Nulle	-
	N424	Nulle	Nulle	rs776390568	N424K	0.0165	Nulle	-
	R409	0.0160	Nulle	rs756158031	R409H	Nulle	Nulle	-
	R409	0.0160	Nulle	rs756158031	R409L	Nulle	Nulle	-
	R314	0.2399	Nulle	rs766483132	R314Q	Nulle	Nulle	-
	R200	0.1184	0.580663	rs779292276	R200Q	Nulle	Nulle	-
	K175	0.0749	Nulle	rs762811026	K175T	Nulle	Nulle	-
	P166	Nulle	Nulle	rs759769003	P166R	0.1229	Nulle	-
Ubiquitination								
Protéine	Position d'AA	Sauvage		SNP ID	Position d'AA	Mutée		+/-
		GPS-Uber	Musite Deep			GPS-Uber	Musite Deep	
TGFβR2	K277	0.4531	0.43	rs1160016010	K277N	Nulle	Nulle	+
SMAD7	N424	Nulle	Nulle	rs776390568	N424K	Nulle	Nulle	-
	K175	Nulle	Nulle	rs762811026	K175T	Nulle	Nulle	-
domaines de liaison aux phosphoprotéines								
Protéine	Position d'AA	Sauvage		SNP ID	Position d'AA	Mutée		+/-
		GPS-PBS				GPS-PBS		
Nulle								
Carbonylation								
Protéine	Position d'AA	Sauvage		SNP ID	Position d'AA	Mutée		+/-
		iCarPS	CarSite-II			iCarPS	CarSite-II	
Nulle								
tyrosine nitration sites								
Protéine	Position d'AA	Sauvage		SNP ID	Position d'AA	Mutée		+/-
		GPS-YNO2 1.0	DeepNitro			GPS-YNO2 1.0	DeepNitro	
SMAD7	Y407	0.058	0.025	rs755148320	Y407S	Nulle	Nulle	-
S-nitrosylation sites								
Protéine	Position d'AA	Sauvage		SNP ID	Position d'AA	Mutée		+/-
		GPS-SNO 1.0	DeepNitro			GPS-SNO 1.0	DeepNitro	
SMAD7	S336	Nulle	Nulle	rs944090062	S336C	0.623	0.17	+

2. Résultats examen des effets des nsSNPs par modélisation 3D des protéines

Une modélisation 3D prédictive a été réalisée avec les comparaisons structurales entre les modèles de type sauvage et mutant pour déterminer si les nsSNPs à haut risque modifient considérablement les protéines résultantes.

Chapitre III : Résultats et interprétation

Le Template 7dv6.1.A a été utilisée pour prédire la protéine TGFBR2 de sauvage et leurs modèles mutants. Et le Template 7cd1.1.A pour la SMAD7 sauvage et leurs modèles mutants.

Pour la modélisation des structures protéiques on utilisé Phyre2 et SWISS-MODEL, ce dernier donner les scores suivants :

- GMQE Score : nous avons obtenu un GMQE score compris entre 0.43 pour TGFBR2 et 0.34 pour SMAD7 ce qui suggère que nos modèles sont basés sur de bons alignements.
- QMEAN Z-Score : Dans TGFBR2 nous avons obtenu un Z-score compris entre -1.09 et -0.71 indique que l'énergie de nos modèles des séquences mutées est plus élevée que la moyenne des structures similaires utilisées comme référence. Et dans SMAD7 on a obtenu un z-score positif dans P416L (0.06) et P166R (0.02) qui suggère une bonne concordance énergétique entre nos modèles mutées et la structure de référence.
- Ramachandran value : On a obtenu des scores entre 96-98% pour les séquences mutées cela indique que la grande majorité des angles de torsion des résidus de nos protéines se situe dans les régions autorisées du diagramme de Ramachandran.

On a comparé entre les structures des protéines mutées et des protéines sauvages par TM-align, Nous avons obtenu un TM Score pour TGFBR2 et SMAD7 de 1 et 0,99 ce qui indique une correspondance parfaites et presque parfaites entre deux structure, Et RMSD compris entre 0 et 0,06. Il est considéré comme très bon, Cela indique que les positions des atomes dans notre modèle ne diffèrent pas par rapport à la structure de référence.

Un score ERRAT élevé (score>79) indique une meilleure correspondance et une meilleure qualité structurale.

Par rapport au diagramme de Ramachandran en logiciel PROCHEK, nos résultats suggèrent une bonne qualité de la conformation des résidus de nos modèles de séquences mutées

Chapitre III : Résultats et interprétation

Tableau 2 : examen des effets des nsSNPs sur la structure de TGFBR2 (test par Phyre2)

TGFBR2								
SNP ID	Position	TM-align Phyre2		ERRAT Phyre2	PROCHEK Phyre2			
		SCORE	RMSD	Value	Residues in most favoured regions	Residues in additional allowed regions	Residues in generously allowed regions	Residues in disallowed regions
TGFBR2 sauvage		1	0	79.3919	262 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs1160016010	K277N	1	0	79.7297	262 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs751948498	R279H	0.99671	0	80.6122	261 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs202168735	T334M	1	0	79.3919	262 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs55751315	R338W	1	0	80.4054	262 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs200361387	R338Q	1	0	83.1081	262 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs200361387	R338P	1	0	81.7568	261 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs200361387	R338L	1	0	82.9932	262 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs759215875	G356S	0.99671	0	79.322	261 (94.9%)	13 (4.7%)	1 (0.4%)	0 (0%)
rs1332138742	G356D	1	0	79.3919	262 (94.9%)	13 (4.7%)	1 (0.4%)	0 (0%)
rs765447250	G394W	1	0	79.3919	262 (94.9%)	13 (4.7%)	0 (0%)	1 (0.4%)
rs771094273	R448G	1	0	77.3649	261 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs1305853447	R520L	1	0	79.3919	262 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)
rs200958264	R522Q	1	0	79.7297	262 (95.3%)	13 (4.7%)	0 (0%)	0 (0%)

Chapitre III : Résultats et interprétation

Tableau 3 : examen des effets des nsSNPs sur la structure de SMAD7 (test par Phyre2)

SMAD7								
SNP ID	Position	TM-align Phyre2		ERRAT Phyre2	PROCHEK Phyre2			
		SCORE	RMSD	Value	Residues in most favoured regions	Residues in additional allowed regions	Residues in generously allowed regions	Residues in disallowed regions
SMAD7 sauvage		1	0	85.5346	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs770772156	R426W	1	0	84.9057	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs776390568	N424K	1	0	85.5346	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs1193638981	P416L	1	0	80.5031	146 (92.4%)	10 (6.3%)	0 (0%)	2 (1.3%)
rs756158031	R409H	1	0	86.7925	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs756158031	R409L	1	0	86.7925	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs755148320	Y407S	1	0	85.5346	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs185241768	G402V	1	0	76.7296	146 (92.4%)	10 (6.3%)	0 (0%)	2 (1.3%)
rs944090062	S336C	1	0	86.1635	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs766483132	R314Q	1	0	85	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs779292276	R200Q	1	0	85.5346	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs762811026	K175T	1	0	85.5346	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)
rs759769003	P166R	1	0	85.5346	145 (92.4%)	10 (6.4%)	0 (0%)	2 (1.3%)

Chapitre III : Résultats et interprétation

Tableau 4 : examen des effets des nsSNPs sur la structure de TGFBR2 (test par SWISS-MODEL)

TGFBR2											
SNP ID	Position	SWISS-MODEL			TM-ALIGN SWISS-MODEL		ERRAT	PROCHEK			
		GMQE	QMEAN	Ramachandran	SCORE	RMSD	Value	Residues in most favored regions	Residues in additional allowed regions	Residues in generously allowed regions	Residues in disallowed regions
TGFBR2 sauvage		0.43	-0.72	97.02%	1	0	99.3007	257 (93.5%)	18 (6.5%)	0 (0%)	0 (0%)
rs11600 16010	K277N	0.43	-0.80	97.02%	1	0	99.3007	257 (93.5%)	18 (6.5%)	0 (0%)	0 (0%)
rs75194 8498	R279H	0.43	-0.71	97.02%	1	0	99.3007	257 (93.5%)	18 (6.5%)	0 (0%)	0 (0%)
rs20216 8735	T334M	0.43	-0.72	97.02%	1	0	99.3007	257 (93.5%)	18 (6.5%)	0 (0%)	0 (0%)
rs55751 315	R338W	0.43	-0.73	97.02%	1	0	99.2982	257 (93.5%)	18 (6.5%)	0 (0%)	0 (0%)
rs20036 1387	R338Q	0.43	-0.80	97.02%	1	0.01	99.2982	257 (93.5%)	18 (6.5%)	0 (0%)	0 (0%)
rs20036 1387	R338P	0.43	-0.88	97.02%	0.99999	0.02	99.3007	255 (93.1%)	19 (6.9%)	0 (0%)	0 (0%)
rs20036 1387	R338L	0.43	-0.74	97.02%	1	0	98.5915	257 (93.5%)	18 (6.5%)	0 (0%)	0 (0%)
rs75921 5875	G356S	0.43	-1.09	96.69%	1	0.01	99.3007	257 (93.1%)	18 (6.5%)	1 (0.4%)	0 (0%)
rs13321 38742	G356D	0.43	-1.05	96.69%	1	0.01	99.3007	257 (93.1%)	18 (6.5%)	1 (0.4%)	0 (0%)
rs76544 7250	G394W	0.43	-1.02	96.69%	1	0.01	99.3007	257 (93.1%)	18 (6.5%)	0 (0%)	1 (0.4%)
rs77109 4273	R448G	0.43	-0.89	97.02%	1	0.01	99.3007	257 (93.8%)	17 (6.2%)	0 (0%)	0 (0%)
rs13058 53447	R520L	0.43	-0.79	97.35%	1	0	99.3007	257 (93.5%)	18 (6.5%)	0 (0%)	0 (0%)
rs20095 8264	R522Q	0.43	-0.78	97.35%	1	0.01	99.3007	256 (93.1%)	19 (6.9%)	0 (0%)	0 (0%)

Chapitre III : Résultats et interprétation

Tableau 5 : examen des effets des nsSNPs sur la structure de SMAD7 (test par SWISS-MODEL)

SMAD7											
SNP ID	Position	SWISS-MODEL			TM-ALIGN SWISS-MODEL		ERRAT	PROCHEK			
		GMQE	QMEAN	Ramachandran	SCORE	RMSD	Value	Residues in most favored regions	Residues in additional allowed regions	Residues in generously allowed regions	Residues in disallowed regions
SMAD7 sauvage		0.33	-0.00	97.75%	1	0	99.0625	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)
rs77077 2156	R426W	0.33	-0.13	97.74%	0.99444	0	99.0566	285 (91.3%)	23 (7.4%)	0 (0%)	4 (1.3%)
rs77639 0568	N424K	0.34	-0.29	97.75%	0.99995	0.04	99.6875	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)
rs11936 38981	P416L	0.33	0.06	97.75%	0.99998	0.02	98.75	289 (91.5%)	23 (7.3%)	0 (0%)	4 (1.3%)
rs75615 8031	R409H	0.33	-0.15	98.03%	1	0	98.125	285 (90.8%)	25 (8%)	0 (0%)	4 (1.3%)
rs75615 8031	R409L	0.33	-0.24	98.03%	1	0.01	99.0596	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)
rs75514 8320	Y407S	0.33	-0.08	97.75%	1	0	99.0566	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)
rs18524 1768	G402V	0.33	-0.27	97.19%	0.99984	0.06	92.638	286 (90.5%)	26 (8.2%)	0 (0%)	4 (1.3%)
rs94409 0062	S336C	0.34	-0.15	97.75%	1	0.01	99.0625	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)
rs76648 3132	R314Q	0.33	-0.01	97.75%	1	0	98.75	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)
rs77929 2276	R200Q	0.33	-0.2	97.75%	1	0	99.0625	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)
rs76281 1026	K175T	0.34	-0.06	97.75%	1	0	99.0625	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)
rs75976 9003	P166R	0.34	0.02	97.75%	1	0	99.0625	286 (91.1%)	24 (7.6%)	0 (0%)	4 (1.3%)

Par logiciel SWISS MODEL nous avons obtenu aussi deux graphes :

Le graphe « local quality estimate » montre que nos séquences sont très élevées. Une indication que ces structures sont bien conçues et que leur position et leur environnement chimique dans la structure sont conformes aux contraintes expérimentales ou aux règles géométriques attendues pour un résidu de ce type.

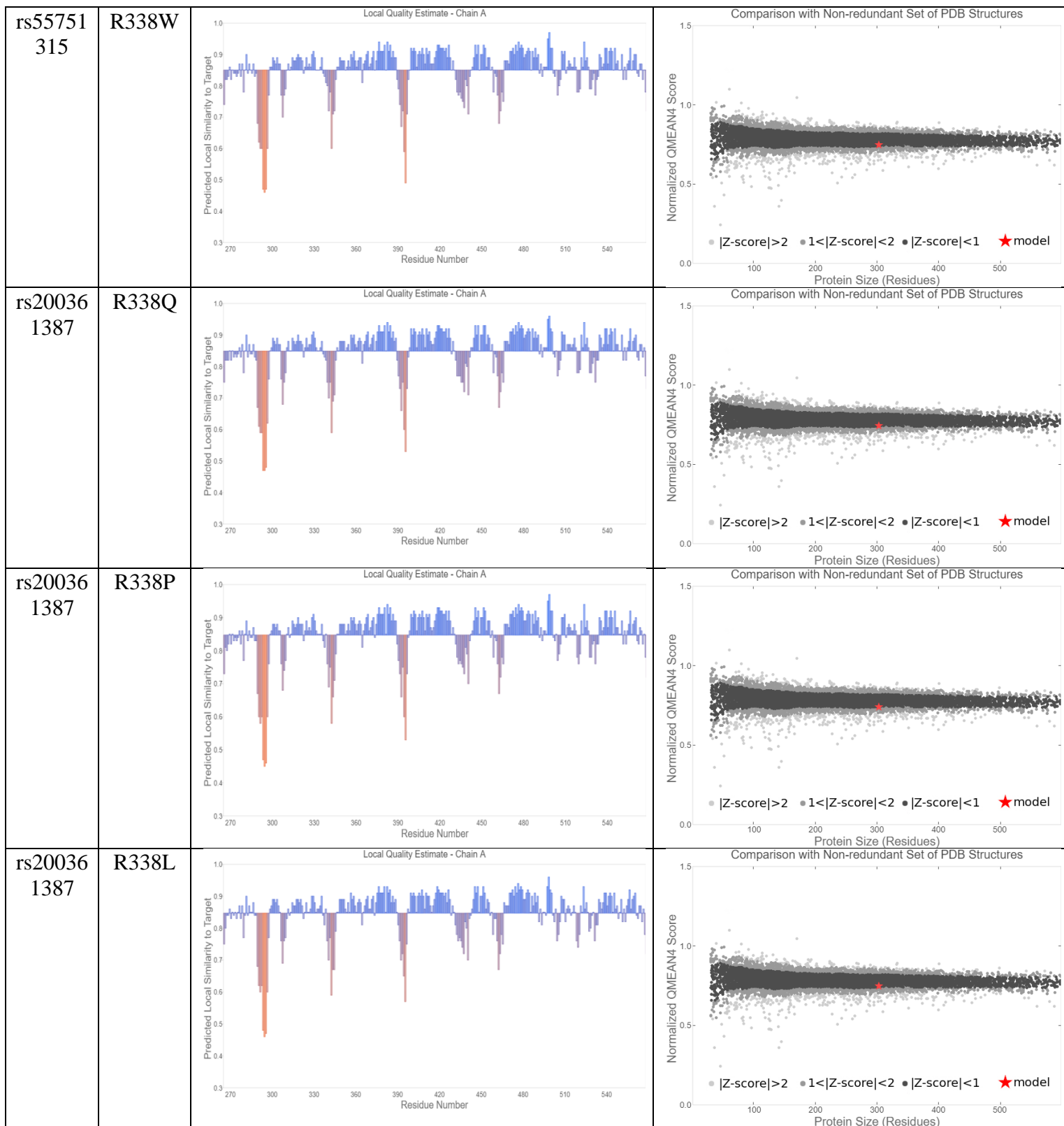
Le graphe « comparison with redundant Set of PDB structures » représente nos modèles avec une étoile dans un champ noir, indiquant une bonne prédiction de la modélisation qui a été effectuée dans toutes nos séquences mutantes.

Chapitre III : Résultats et interprétation

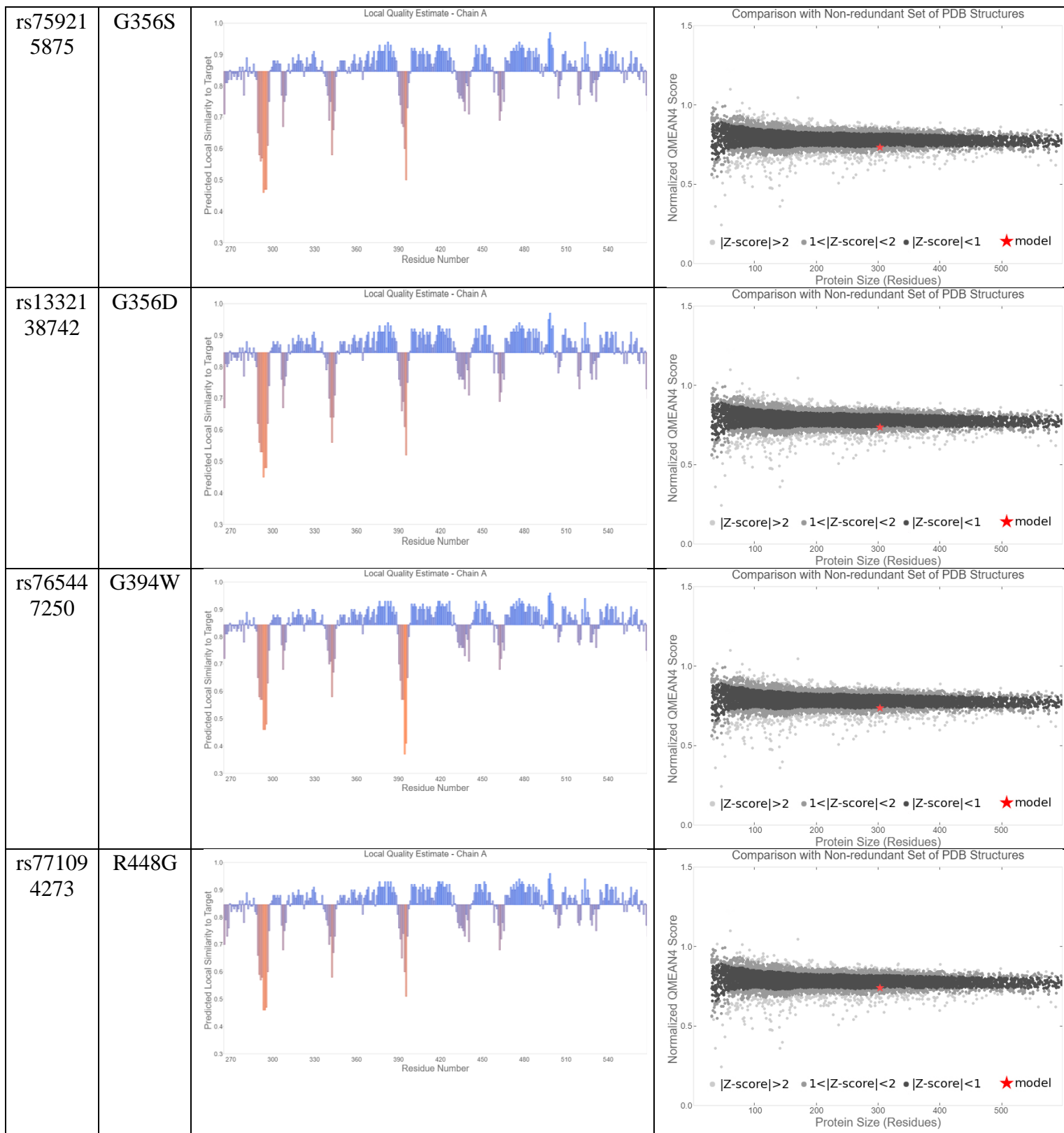
Tableau 6 : Estimations de qualités des protéines modélisées par SWISS-MODEL

SNP	AA	Local qualité estimate	Comparison redondant Set of PDB structures
TGFBR2			
TGFBR2 sauvage			
rs11600 16010	K277N		
rs75194 8498	R279H		
rs20216 8735	T334M		

Chapitre III : Résultats et interprétation



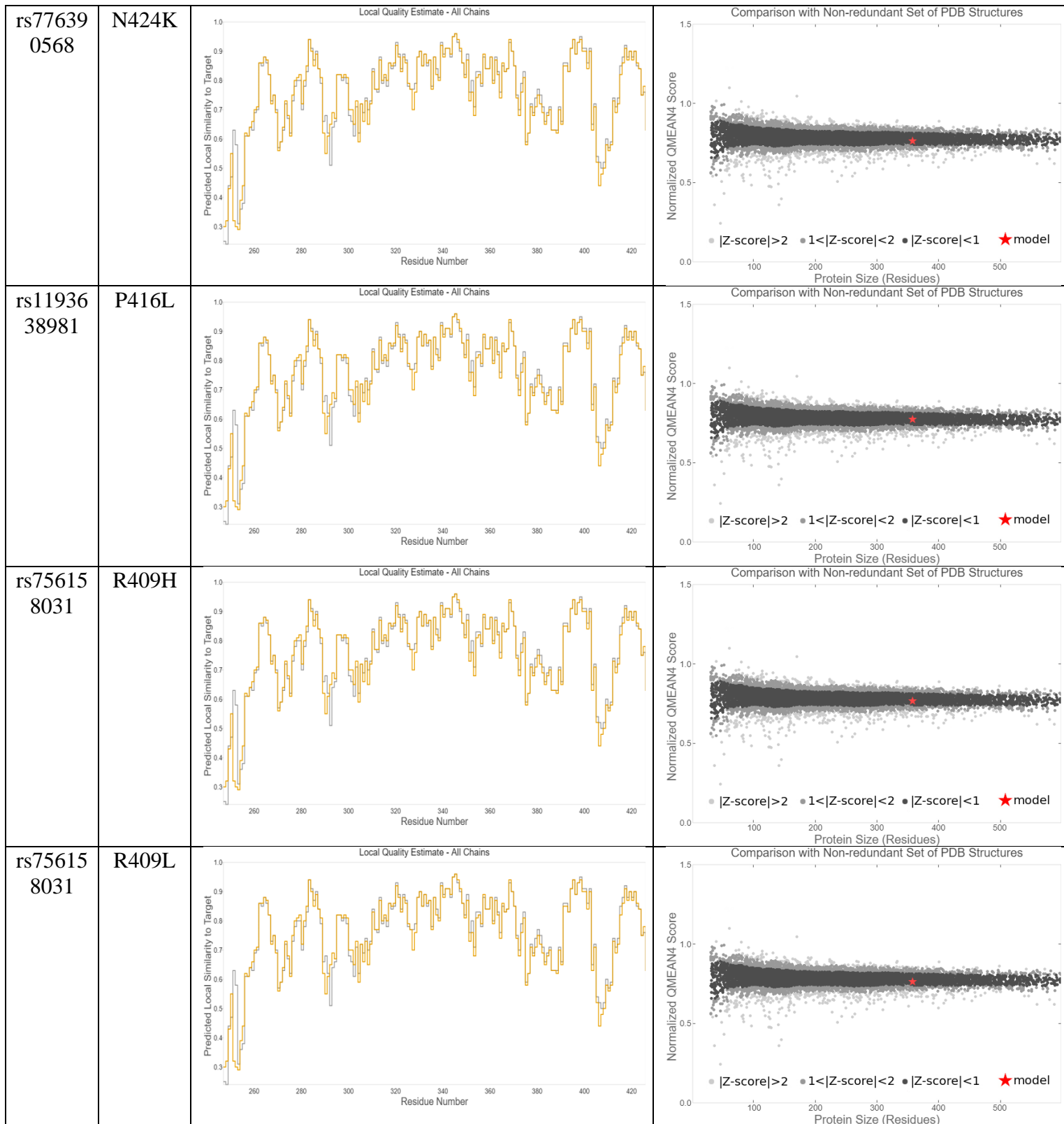
Chapitre III : Résultats et interprétation



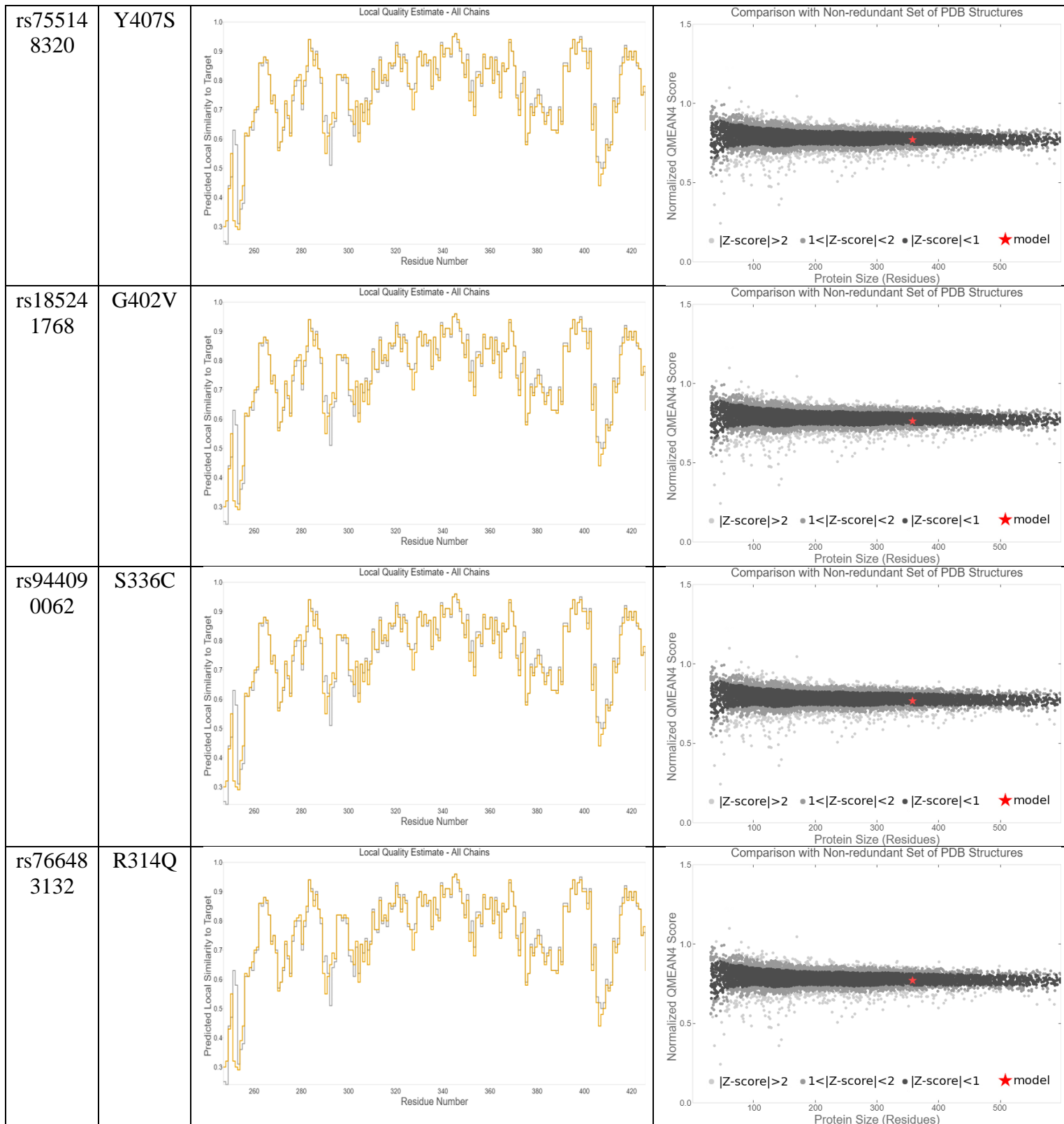
Chapitre III : Résultats et interprétation

rs13058 53447	R520L		
rs20095 8264	R522Q		
SMAD7			
SMAD7 sauvage			
rs77077 2156	R426W		

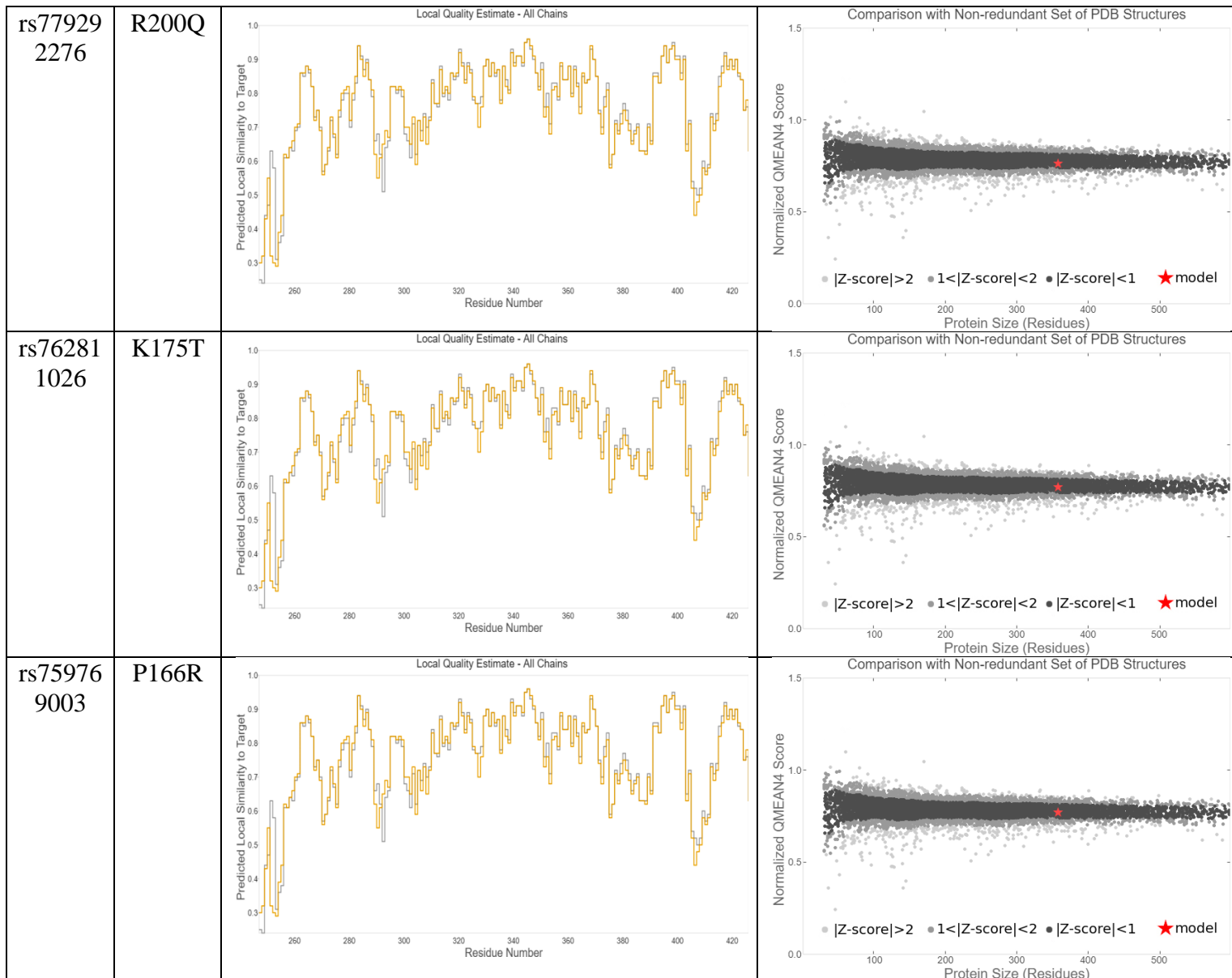
Chapitre III : Résultats et interprétation



Chapitre III : Résultats et interprétation



Chapitre III : Résultats et interprétation

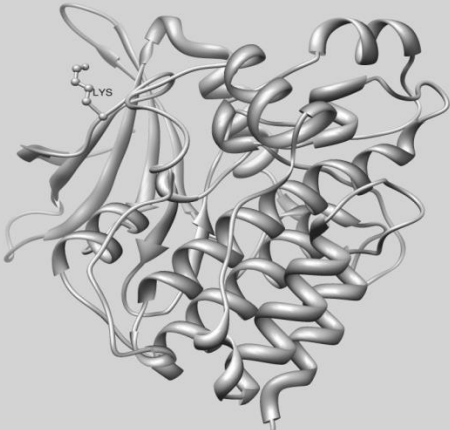
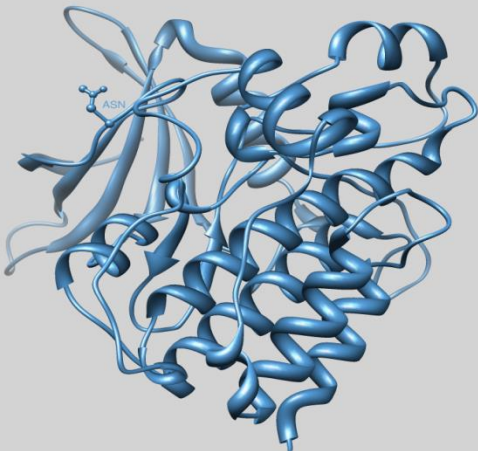
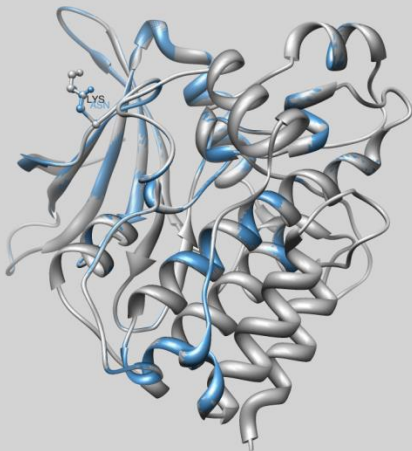
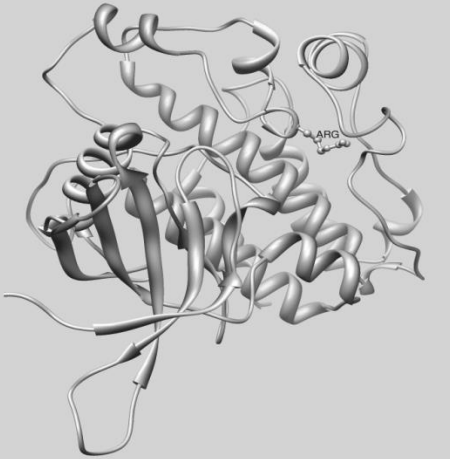
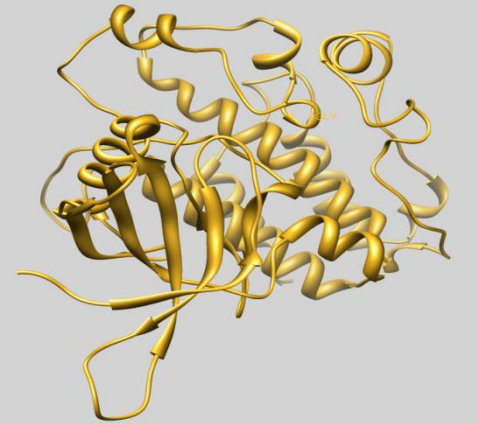
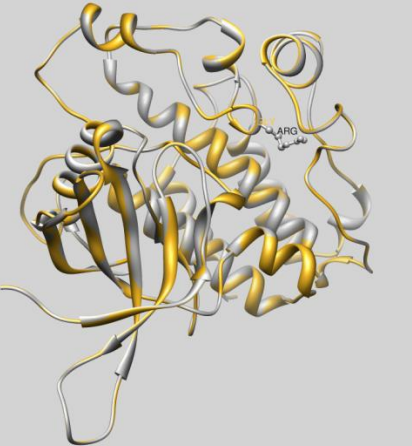


3. Comparaison et superposition des structures 3D

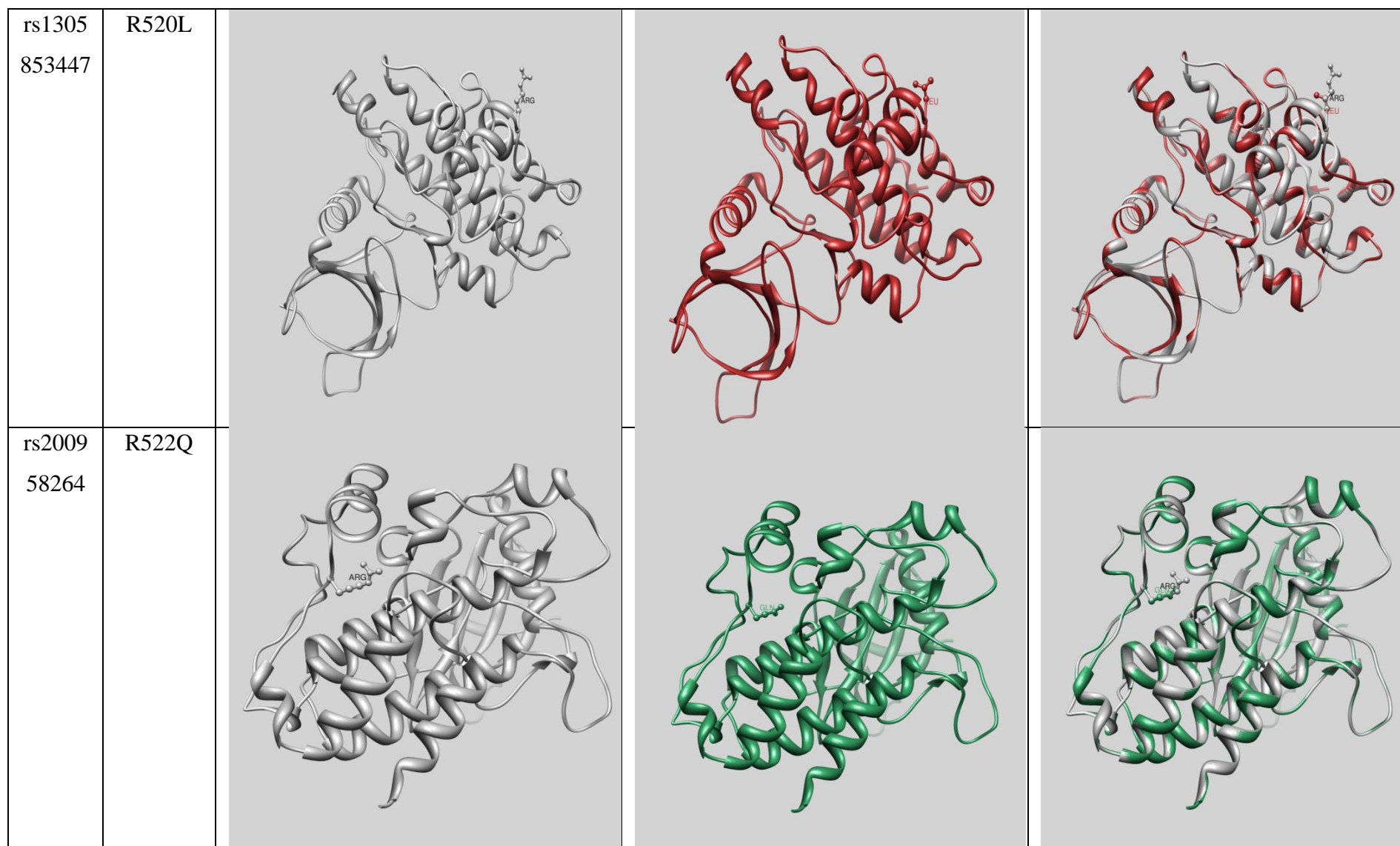
Pour terminer le travail, nous avons réalisé des structures 3D pour les séquences validées avec un résultat significatif de protéine TGFBR2 (K277N, R448G, R520L, R522Q), et quelques substitutions pour le SMAD7 (Y407S, S336C). Ensuite, nous avons confirmé les résultats et visualisé par la méthode de superposition, configuré par l'outil bioinformatique Chimera.

Chapitre III : Résultats et interprétation

Tableau 7 : Comparaison structurale entre la séquence sauvage de TGFBR2 et leurs séquences mutées par Chimera.


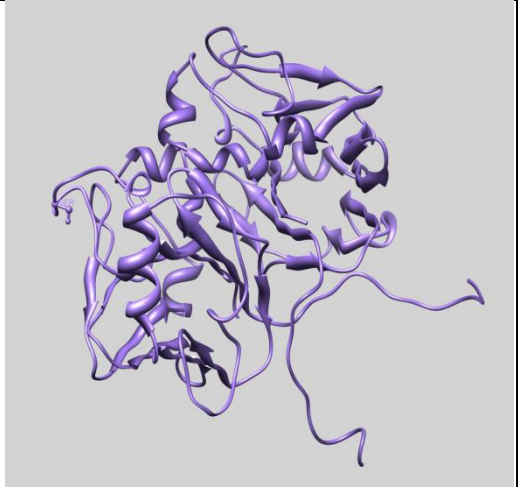
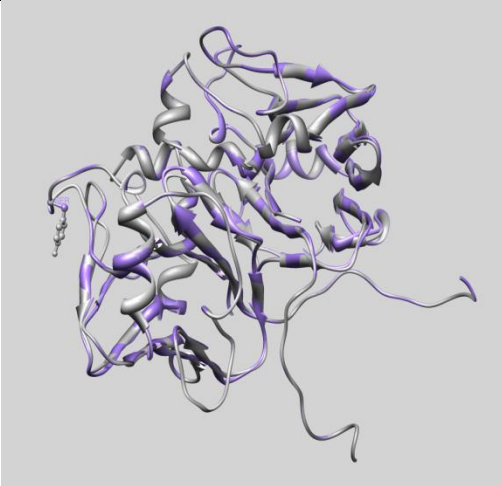


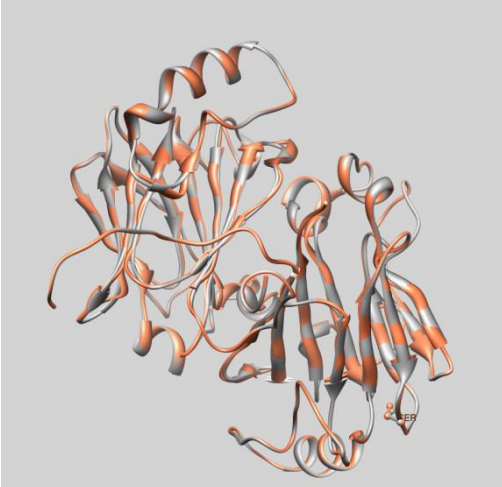
TGFBR2				
SNP ID	Position d'AA	Site sauvage	Site muté	Site merged
rs1160 016010	K277N			
rs7710 94273	R448G			

Chapitre III : Résultats et interprétation



Chapitre III : Résultats et interprétation

Tableau 8 : Comparaison structurale entre la séquence sauvage de SMAD7 et leurs séquences mutées par Chimera.

SMAD7				
SNP ID	Position d'AA	Site sauvage	Site muté	Site merged
rs7551 48320	Y407S			
rs9440 90062	S336C			

Chapitre IV : Discussion

Chapitre IV : Discussion

Discussion :

La voie du TGF β -SMAD7 est l'une des voies clés et les plus importantes dans le développement du CCR, Elle joue un rôle essentiel dans l'invasion et la migration des cellules cancéreuses, et ça grâce des variantes génétiques qui peuvent affecter l'expression du récepteur TGFBR2.

L'année précédente, Snouci Meriem Radia et Zouaimia Nahla Ibtissam ont identifié les nsSNPs à conséquences fonctionnelles à haut risque dans les protéines de TGFBR2 et SMAD7 afin de comprendre leur association avec le CCR. Elles ont pu déterminer 13 nsSNPs dans le gène TGFBR2 (K277N, R279H, T334M, R338W, R338Q, R338P, R338L, G356S, G356D, G394W, R448G, R520L, R522Q) ainsi que 12 autres dans SMAD7 (R426W, N424K, P416L, R409H, R409L, Y407S, G402V, S336C, R314Q, R200Q, K175T, P166R)

Nous avons poursuivi ce travail en réalisant une analyse *in silico* de l'effet de ces variations sur la structure tridimensionnelle de TGF β R2 et SMAD7 et leurs modifications post traductionnelles et Nous avons identifié 4 nsSNPs à conséquences fonctionnelles à haut risque dans TGF β R2.

Les analyses de prédiction des modifications post-traductionnelles dans les protéines, en utilisant 12 outils bioinformatiques, ont révélé la disparition de trois sites de méthylation. Cette disparition a été causée par les variations suivantes : R448G, R520L, R522Q. On a trouvé aussi un site d'ubiquitination on position K277 de TGFBR2 sauvage qui est également supprimé par la variation K277N.

Pour la modélisation moléculaire, deux outils de modélisation d'homologie distincts ont été utilisés (Phyre2 et SWISS-MODEL) afin de créer la structure 3D de nos protéines sauvages et de chaque protéine mutée selon des modèles sélectionnés « Template ». Les similitudes entre les structures protéiques modélisées de type sauvage et mutantes ont été ensuite analysées par TM-align. En fin toutes les structures ont été visualisées et superposé à l'aide de Chimera.

Les résultats de cette étude montrent que les variantes génétiques dans les protéines TGFBR2 et SMAD7 peuvent potentiellement affecter le récepteur TGF β R2, qui est une voie clé et importante dans le développement du CCR.

Notre étude a révélé que certaines mutations pourraient supprimer des modifications post-traductionnelles dans TGF β R2 telles que la méthylation ou l'ubiquitination. Ces PTM peuvent influencer la régulation des fonctions de TGF- β et ces activités.

La modélisation moléculaire nous a permis de constater que ces mutations sont responsables

Chapitre IV : Discussion

d'une légère modification structurelle de TGF β R2.

En somme, cette étude fournit un nouvel éclairage sur le rôle possible de certains nsSNPs identifiés précédemment sur TGF β R2 et SMAD7 dans le contexte du cancer colorectal.

Chapitre V : Conclusion et perspectives

Conclusion :

Notre étude a permis d'analyser les conséquences fonctionnelles des Single Nucleotide Polymorphisms (SNPs) sur les protéines TGFBR2 et SMAD7 dans le contexte du cancer colorectal. Nous avons utilisé différentes méthodes bioinformatiques pour prédire les modifications post-traductionnelles associées aux SNPs identifiés comme potentiellement délétères, ainsi que la modélisation moléculaire pour visualiser plus précisément ces altérations.

Nos résultats ont montré que certaines variantes de la protéine TGFBR2 (R448G, R520L, R522Q et K277N) peuvent potentiellement induire diverses modifications post traductionnelles telles que la méthylation et l'ubiquitination.

En somme, notre étude fournit une base préliminaire intéressante concernant l'impact possible de certains SNPs sur TGFBR2 dans le contexte du cancer colorectal. Ces résultats pourront être utilisés comme point de départ pour des études expérimentales visant à valider les effets de ces SNPs sur les protéines et leur rôle dans le développement de cette maladie.

Chapitre VI : Bibliographies

Bibliographies

B

- Balchen, V., Simon, K., 2016. Colorectal cancer development and advances in screening. *Clin. Interv. Aging* Volume 11, 967–976. <https://doi.org/10.2147/CIA.S109285>
- Biswas, A.K., Noman, N., Sikder, A.R., 2010. Machine learning approach to predict protein phosphorylation sites by incorporating evolutionary information. *BMC Bioinformatics* 11, 273. <https://doi.org/10.1186/1471-2105-11-273>
- Butler, J.M., 2012. Single Nucleotide Polymorphisms and Applications, in: *Advanced Topics in Forensic DNA Typing*. Elsevier, pp. 347–369. <https://doi.org/10.1016/B978-0-12-374513-2.00012-9>

C

- Chen, M., Zhang, W., Gou, Y., Xu, D., Wei, Y., Liu, D., Han, C., Huang, X., Li, C., Ning, W., Peng, D., Xue, Y., 2023. GPS 6.0: an updated server for prediction of kinase-specific phosphorylation sites in proteins. *Nucleic Acids Res.* gkad383. <https://doi.org/10.1093/nar/gkad383>

D

- Deng, N., Zhou, H., Fan, H., Yuan, Y., 2017. Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget* 8, 110635–110649. <https://doi.org/10.18632/oncotarget.22372>
- Duan, B., Zhao, Y., Bai, J., Wang, J., Duan, X., Luo, X., Zhang, R., Pu, Y., Kou, M., Lei, J., Yang, S., 2022. Colorectal Cancer: An Overview, in: *Morgado-Diaz, J.A. (Ed.), Gastrointestinal Cancers*. Exon Publications, Brisbane (AU).

G

- Gu, S., Feng, X.-H., 2018. TGF- β signaling in cancer. *Acta Biochim. Biophys. Sin.* 50, 941–949. <https://doi.org/10.1093/abbs/gmy092>

H

- Holtedahl, K., Borgquist, L., Donker, G.A., Buntinx, F., Weller, D., Campbell, C., Månsson, J., Hammersley, V., Braaten, T., Parajuli, R., 2021. Symptoms and signs of colorectal cancer, with differences between proximal and distal colon cancer: a prospective cohort study of diagnostic accuracy in primary care. *BMC Fam. Pract.* 22, 148. <https://doi.org/10.1186/s12875-021-01452-6>
- Hossain, M.S., Karuniawati, H., Jairoun, A.A., Urbi, Z., Ooi, D.J., John, A., Lim, Y.C., Kibria, K.M.K., Mohiuddin, A.K.M., Ming, L.C., Goh, K.W., Hadi, M.A., 2022. Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. *Cancers* 14, 1732. <https://doi.org/10.3390/cancers14071732>

J

- Javelaud, D., Mauviel, A., 2004. Transforming growth factor- β s : signalisation et rôles physiopathologiques. *Pathol. Biol.* 52, 50–54. <https://doi.org/10.1016/j.patbio.2003.10.002>
- Johdi, N.A., Sukor, N.F., 2020. Colorectal Cancer Immunotherapy: Options and Strategies. *Front. Immunol.* 11, 1624. <https://doi.org/10.3389/fimmu.2020.01624>
- Jung, B., Staudacher, J.J., Beauchamp, D., 2017. Transforming Growth Factor β Superfamily Signaling in Development of Colorectal Cancer. *Gastroenterology* 152, 36–52. <https://doi.org/10.1053/j.gastro.2016.10.015>

K

- Kaminski, M.F., Regula, J., Kraszewska, E., Polkowski, M., Wojciechowska, U., Didkowska, J., Zwierko, M., Rupinski, M., Nowacki, M.P., Butruk, E., 2010. Quality indicators for colonoscopy and the risk of interval cancer. *N. Engl. J. Med.* 362, 1795–1803. <https://doi.org/10.1056/NEJMoa0907667>
- Kamiza, A.B., Wang, W.-C., You, J.-F., Tang, R., Wang, Y.-T., Chien, H.-T., Lai, C.-H., Chiu, L.-L., Lo, T.-P., Hung, K.-Y., Hsiung, C.A., Yeh, C.-C., 2018. *EGFR*, *SMAD7*, and

TGFBR2 Polymorphisms Are Associated with Colorectal Cancer in Patients with Lynch Syndrome. *Anticancer Res.* 38, 5983–5990. <https://doi.org/10.21873/anticancer.12946>

- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. <https://doi.org/10.1038/nprot.2015.053>
- Keum, N., Giovannucci, E., 2019. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* 16, 713–732. <https://doi.org/10.1038/s41575-019-0189-8>
- Korzeniewski, S., Hofman, P., Brest, P., 2013. Des polymorphismes silencieux plutôt bruyants. *médecine/sciences* 29, 124–126. <https://doi.org/10.1051/medsci/2013292003>
- Kuipers, E.J., Grady, W.M., Lieberman, D., Seufferlein, T., Sung, J.J., Boelens, P.G., van de Velde, C.J.H., Watanabe, T., 2015. Colorectal cancer. *Nat. Rev. Dis. Primer* 1, 15065. <https://doi.org/10.1038/nrdp.2015.65>

L

- LAOUAR, H., DAOUDI, S., 2016. Le cancer colorectal Profil épidémiologique-anatomopathologique-immunohistochimique.
- Lasek, W., 2022. Cancer immunoediting hypothesis: history, clinical implications and controversies. *Cent.-Eur. J. Immunol.* 47, 168–174. <https://doi.org/10.5114/ceji.2022.117376>
- Li, J., Zou, L., Zhou, Y., Li, L., Zhu, Y., Yang, Y., Gong, Y., Lou, J., Ke, J., Zhang, Y., Tian, J., Zou, D., Peng, X., Chang, J., Gong, J., Zhong, R., Zhou, X., Miao, X., 2017. A low-frequency variant in SMAD7 modulates TGF- β signaling and confers risk for colorectal cancer in Chinese population. *Mol. Carcinog.* 56, 1798–1807. <https://doi.org/10.1002/mc.22637>
- Lim, S.W., Tan, K.J., Azuraiddi, O.M., Sathiy, M., Lim, E.C., Lai, K.S., Yap, W.-S., Afizan, N.A.R.N.M., 2021. Functional and structural analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the MYB oncoproteins associated with human cancer. *Sci. Rep.* 11, 24206. <https://doi.org/10.1038/s41598-021-03624-x>
- Liu, Z., Cao, J., Ma, Q., Gao, X., Ren, J., Xue, Y., 2011. GPS-YNO2: computational prediction of tyrosine nitration sites in proteins. *Mol. Biosyst.* 7, 1197. <https://doi.org/10.1039/c0mb00279h>
- Lotfollahzadeh, S., Recio-Boiles, A., Cagir, B., 2023. Colon Cancer, in: *StatPearls*. StatPearls Publishing, Treasure Island (FL).

M

- Mahmoud, N.N., 2022. Colorectal Cancer: Preoperative Evaluation and Staging. *Surg. Oncol. Clin. N. Am.* 31, 127–141. <https://doi.org/10.1016/j.soc.2021.12.001>
- Malki, A., ElRuz, R.A., Gupta, I., Allouch, A., Vranic, S., Al Moustafa, A.-E., 2020. Molecular Mechanisms of Colon Cancer Progression and Metastasis: Recent Insights and Advancements. *Int. J. Mol. Sci.* 22, 130. <https://doi.org/10.3390/ijms22010130>
- Markman, J.L., Shiao, S.L., 2015. Impact of the immune system and immunotherapy in colorectal cancer. *J. Gastrointest. Oncol.* 6, 208–223. <https://doi.org/10.3978/j.issn.2078-6891.2014.077>
- Mármol, I., Sánchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E., Rodriguez Yoldi, M.J., 2017. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int. J. Mol. Sci.* 18, 197. <https://doi.org/10.3390/ijms18010197>
- Meulmeester, E., ten Dijke, P., 2011. The dynamic roles of TGF- β in cancer: TGF- β in cancer. *J. Pathol.* 223, 206–219. <https://doi.org/10.1002/path.2785>
- Morgan, E., Arnold, M., Gini, A., Lorenzoni, V., Cabasag, C.J., Laversanne, M., Vignat, J., Ferlay, J., Murphy, N., Bray, F., 2023. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* 72, 338–344. <https://doi.org/10.1136/gutjnl-2022-327736>
- Mukherji, R., Weinberg, B.A., Pedersen, K.S., 2022. Immunotherapy for Colorectal Cancer. *Hematol. Oncol. Clin. North Am.* 36, 603–626. <https://doi.org/10.1016/j.hoc.2022.02.010>

N

Chapitre VI : Bibliographies

- Nassiri, M., Kooshyar, M.M., Roudbar, Z., Mahdavi, M., Doosti, M., 2013. Genes and SNPs Associated with Non-hereditary and Hereditary Colorectal Cancer. *Asian Pac. J. Cancer Prev.* 14, 5609–5614. <https://doi.org/10.7314/APJCP.2013.14.10.5609>

P

- Peng, J.-B., Geng, H.-Q., Wu, X.-F., 2019. The Chemistry of CO: Carbonylation. *Chem* 5, 526–552. <https://doi.org/10.1016/j.chempr.2018.11.006>
- Pernot, S., Terme, M., Voron, T., Colussi, O., Marcheteau, E., Tartour, E., Taieb, J., 2014. Colorectal cancer and immunity: what we know and perspectives. *World J. Gastroenterol.* 20, 3738–3750. <https://doi.org/10.3748/wjg.v20.i14.3738>
- Plenchette, S., Romagny, S., Laurens, V., Bettaieb, A., 2016. Itinéraire d'un agent double: NO, S-nitrosylation et cancer. *médecine/sciences* 32, 625–633. <https://doi.org/10.1051/medsci/20163206027>
- Poulard, C., Bouchekioua-Bouzaghrou, K., Sentis, S., Corbo, L., Le Romancer, M., 2010. Les modifications post-traductionnelles orchestrent l'action du récepteur des œstrogènes ϵ R α dans les tumeurs mammaires. *médecine/sciences* 26, 636–640. <https://doi.org/10.1051/medsci/2010266-7636>

R

- Radi, R., 2013. Protein Tyrosine Nitration: Biochemical Mechanisms and Structural Basis of Functional Effects. *Acc. Chem. Res.* 46, 550–559. <https://doi.org/10.1021/ar300234c>
- Rawla, P., Sunkara, T., Barsouk, A., 2019. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Przegląd Gastroenterol.* 14, 89–103. <https://doi.org/10.5114/pg.2018.81072>
- Righini, C.A., de Fraipont, F., Timsit, J.-F., Dassonville, O., Milano, G., Moro-Sibilot, D., 2008. Étude de la méthylation des GST dans la salive en cas de cancer des VADS. *Rev. Stomatol. Chir. Maxillofac.* 109, 226–232. <https://doi.org/10.1016/j.stomax.2008.06.007>

S

- Schwede, T., Kopp, J., Guex, N., Peitsch, M.C., 2003. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381–3385. <https://doi.org/10.1093/nar/gkg520>
- Serra, R., 2002. Transforming Growth Factor beta (TGF β), in: John Wiley & Sons, Ltd (Ed.), *ELS*. Wiley. <https://doi.org/10.1038/npg.els.0000934>
- Shah, S.C., Itzkowitz, S.H., 2022. Colorectal Cancer in Inflammatory Bowel Disease: Mechanisms and Management. *Gastroenterology* 162, 715-730.e3. <https://doi.org/10.1053/j.gastro.2021.10.035>
- Sharma, I., Kim, S., Sridhar, S., Basha, R., 2020. Colorectal Cancer: An Emphasis on Factors Influencing Racial/Ethnic Disparities. *Crit. Rev. Oncog.* 25, 151–160. <https://doi.org/10.1615/CritRevOncog.2020035174>
- Slattery, M.L., Edwards, S.L., Samowitz, W., 1998. Stage of colon cancer at diagnosis: implications for risk factor associations? *Int. J. Epidemiol.* 27, 382–387. <https://doi.org/10.1093/ije/27.3.382>
- Stolfi, C., Marafini, I., De Simone, V., Pallone, F., Monteleone, G., 2013. The Dual Role of Smad7 in the Control of Cancer Growth and Metastasis. *Int. J. Mol. Sci.* 14, 23774–23790. <https://doi.org/10.3390/ijms141223774>
- Syed, V., 2016. TGF- β Signaling in Cancer. *J. Cell. Biochem.* 117, 1279–1287. <https://doi.org/10.1002/jcb.25496>

T

- Tronccone, E., Monteleone, G., 2019. Smad7 and Colorectal Carcinogenesis: A Double-Edged Sword. *Cancers* 11, 612. <https://doi.org/10.3390/cancers11050612>

V

- Valle, L., Vilar, E., Tavtigian, S.V., Stoffel, E.M., 2019. Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. *J. Pathol.* 247, 574–588. <https://doi.org/10.1002/path.5229>

Chapitre VI : Bibliographies

- Vega, P., Valentín, F., Cubiella, J., 2015. Colorectal cancer diagnosis: Pitfalls and opportunities. *World J. Gastrointest. Oncol.* 7, 422. <https://doi.org/10.4251/wjgo.v7.i12.422>

W

- Wang, D., Zeng, S., Xu, C., Qiu, W., Liang, Y., Joshi, T., Xu, D., 2017. MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction. *Bioinforma. Oxf. Engl.* 33, 3909–3916. <https://doi.org/10.1093/bioinformatics/btx496>
- Wang, H.-B., Ma, L.-H., Zhang, T., Huang, K.-C., Zhao, Y.-D., Liu, T.-C., 2020. Simple and accurate visual detection of single nucleotide polymorphism based on colloidal gold nucleic acid strip biosensor and primer-specific PCR. *Anal. Chim. Acta* 1093, 106–114. <https://doi.org/10.1016/j.aca.2019.09.048>

X

- Xie, Y., Luo, X., Li, Y., Chen, L., Ma, W., Huang, J., Cui, J., Zhao, Y., Xue, Y., Zuo, Z., Ren, J., 2018. DeepNitro: Prediction of Protein Nitration and Nitrosylation Sites by Deep Learning. *Genomics Proteomics Bioinformatics* 16, 294–306. <https://doi.org/10.1016/j.gpb.2018.04.007>
- Xue, Y., Liu, Z., Gao, X., Jin, C., Wen, L., Yao, X., Ren, J., 2010. GPS-SNO: Computational Prediction of Protein S-Nitrosylation Sites with a Modified GPS Algorithm. *PLoS ONE* 5, e11290. <https://doi.org/10.1371/journal.pone.0011290>
- Yates, C.M., Sternberg, M.J.E., 2013. The Effects of Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) on Protein–Protein Interactions. *J. Mol. Biol.* 425, 3949–3963. <https://doi.org/10.1016/j.jmb.2013.07.012>

Y

- Yu, G.-H., Li, S.-F., Wei, R., Jiang, Z., 2022. Diabetes and Colorectal Cancer Risk: Clinical and Therapeutic Implications. *J. Diabetes Res.* 2022, 1747326. <https://doi.org/10.1155/2022/1747326>

Z

- Zhang, Y., Rajput, A., Jin, N., Wang, J., 2020. Mechanisms of Immunosuppression in Colorectal Cancer. *Cancers* 12, 3850. <https://doi.org/10.3390/cancers12123850>

Chapitre VII : Annexes

Annexes

Les séquences FASTA des protéines sauvage et mutées

- **TGFBR2 sauvage:**

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2
MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKCMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLLLVIQVVTGISLLPPLGVAVISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAIILED RSDISSTCANNINHNTPELLDITLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAERKTELKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSSLRGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWLNLHQGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTTK
```

- **TGFBR2 SNP: K277N**

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2
MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKCMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLLLVIQVVTGISLLPPLGVAVISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAIILED RSDISSTCANNINHNTPELLDITLVGNRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAERKTELKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSSLRGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWLNLHQGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTTK
```

- **TGFBR2 SNP: R279H**

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2
MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKCMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLLLVIQVVTGISLLPPLGVAVISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAIILED RSDISSTCANNINHNTPELLDITLVGKGHFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAERKTELKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSSLRGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWLNLHQGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTTK
```

- **TGFBR2 SNP: T334M**

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2
MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKCMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLLLVIQVVTGISLLPPLGVAVISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
```


Chapitre VII : Annexes

CAILED RSDISSTCANNINHNTPELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLMAEERKTELGKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSS LARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLES RMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWL NHQGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTTK

- **TGFBR2 SNP: R338W**

>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2

MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI
TLETVCHDPKLPYHDFILED AASP KCIMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLLL VIFQVTGISLLPPLGVAISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAILED RSDISSTCANNINHNTPELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEWKTELGKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSS LARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLES RMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWL NHQGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTTK

- **TGFBR2 SNP: R338Q**

>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2

MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI
TLETVCHDPKLPYHDFILED AASP KCIMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLLL VIFQVTGISLLPPLGVAISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAILED RSDISSTCANNINHNTPELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEQKTELGKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSS LARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLES RMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWL NHQGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTTK

- **TGFBR2 SNP: R338P**

>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2

MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI
TLETVCHDPKLPYHDFILED AASP KCIMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLLL VIFQVTGISLLPPLGVAISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAILED RSDISSTCANNINHNTPELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEKPTELGKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSS LARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLES RMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWL NHQGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTTK

- **TGFBR2 SNP: R338L**

>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2

MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRFSTCDNQKSCMSNCSITSICEKPQEVCAVWRKNDENI

Chapitre VII : Annexes

TLETVCHDPKLPYHDFILEDAAAPKCMKEKKKPKGETFFMCS CSSDECNDNIIFSEEYNT
SNPDL LLVIFQVTGISLLPPLGV AISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAIILED RSDISSTCANNINHNT ELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEELKTEL GKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSS LARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANS GQVGTARYMAPEVLES RMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWL NHQGIQMVCE
TLTECWDHHDPEARLTAQCVAERFSELEHLDR LSGRSCSEEKIPEDGSLNNTK

- **TGFBR2 SNP: G356S**

>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2

MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFC DVRFSTCDNQKSCMSNCSITSICEKPQEV CVAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKCMKEKKKPKGETFFMCS CSSDECNDNIIFSEEYNT
SNPDL LLVIFQVTGISLLPPLGV AISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAIILED RSDISSTCANNINHNT ELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEERKTEL GKQYWLITAFHAKSNLQE
YLTRHVISWEDLRKLGSS LARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANS GQVGTARYMAPEVLES RMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWL NHQGIQMVCE
TLTECWDHHDPEARLTAQCVAERFSELEHLDR LSGRSCSEEKIPEDGSLNNTK

- **TGFBR2 SNP: G356D**

>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2

MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFC DVRFSTCDNQKSCMSNCSITSICEKPQEV CVAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKCMKEKKKPKGETFFMCS CSSDECNDNIIFSEEYNT
SNPDL LLVIFQVTGISLLPPLGV AISVIIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAIILED RSDISSTCANNINHNT ELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEERKTEL GKQYWLITAFHAKDNLQE
YLTRHVISWEDLRKLGSS LARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANS GQVGTARYMAPEVLES RMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWL NHQGIQMVCE
TLTECWDHHDPEARLTAQCVAERFSELEHLDR LSGRSCSEEKIPEDGSLNNTK

- **TGFBR2 SNP: G394W**

>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2

MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRTAHPLRHINND
MIVTDNNGAVKFPQLCKFC DVRFSTCDNQKSCMSNCSITSICEKPQEV CVAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAAPKCMKEKKKPKGETFFMCS CSSDECNDNIIFSEEYNT
SNPDL LLVIFQVTGISLLPPLGV AISVIIIFYCYRVNRQQKLSSTWETGKTRKLMFSEH
CAIILED RSDISSTCANNINHNT ELLPIELDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEERKTEL GKQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSS LARGIAHLHSDHTPCWRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANS GQVGTARYMAPEVLES RMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWL NHQGIQMVCE
TLTECWDHHDPEARLTAQCVAERFSELEHLDR LSGRSCSEEKIPEDGSLNNTK

- **TGFBR2 SNP: R448G**

>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo sapiens OX=9606
GN=TGFBR2

Chapitre VII : Annexes

```
MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRSTCDNQKSCMSNCSITSICEKPQEVAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAASPKCIMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLVVIFQVGTGISLLPPLGVAISVIIIFYCYRVNRQKLSSTWETGKTRKLMFSEH
CAIILEDSDISSTCANNINHNTPELLIEDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEYASWKTEKDFSDINLKHENILQFLTAERKTELKGQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSSSLARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANSQVGTAGYMAPEVLESRMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGRPEIPSWLNLHOGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTK
```

- **TGFBR2 SNP: R520L**

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo
sapiens OX=9606 GN=TGFR2
MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRSTCDNQKSCMSNCSITSICEKPQEVAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAASPKCIMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLVVIFQVGTGISLLPPLGVAISVIIIFYCYRVNRQKLSSTWETGKTRKLMFSEH
CAIILEDSDISSTCANNINHNTPELLIEDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEYASWKTEKDFSDINLKHENILQFLTAERKTELKGQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSSSLARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDLGRPEIPSWLNLHOGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTK
```

- **TGFBR2 SNP: R522Q**

```
>sp|P37173-2|TGFR2_HUMAN Isoform 2 of TGF-beta receptor type-2 OS=Homo
sapiens OX=9606 GN=TGFR2
MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSDVEMEAQKDEIICPSCNRRTAHPLRHINND
MIVTDNNGAVKFPQLCKFCDFRSTCDNQKSCMSNCSITSICEKPQEVAVWRKNDENI
TLETVCHDPKLPYHDFILEDAAASPKCIMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNT
SNPDLVVIFQVGTGISLLPPLGVAISVIIIFYCYRVNRQKLSSTWETGKTRKLMFSEH
CAIILEDSDISSTCANNINHNTPELLIEDTLVGKGRFAEVYKAKLKQNTSEQFETVA
VKIFPYEYASWKTEKDFSDINLKHENILQFLTAERKTELKGQYWLITAFHAKGNLQE
YLTRHVISWEDLRKLGSSSLARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCL
CDFGLSLRLDPTLSVDDLANSQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLW
EMTSRCNAVGEVKDYEPFPGSKVREHPCVESMKDNVLRDRGQPEIPSWLNLHOGIQMVCE
TLTECWDHDPPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNNTK
```

- **SMAD7 sauvage :**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo
sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLRSRAPGGEDEEEGAGGGGGGGELRGEGATDSRAHGAGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDLRHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSOLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQLVQKVRISKIGCGIQLTREVGDVWVYNRSSYPYIFIKSATLDNPDRTLLVHKV
FPGFSIKAFDYEKAYSLSQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSPCWLE
VIFNSR
```

- **SMAD7 SNP:R426W**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo
sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLRSRAPGGEDEEEGAGGGGGGGELRGEGATDSRAHGAGGGGPGRAGC
```

Chapitre VII : Annexes

```
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDRLHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQLVQKVRISKIGCGIQLTREVDGVVWVYNRSSYPIFIKSATLDNPDSTRLLVHKV
FPGFSIKAFDYEKAYSLSQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSPCWLE
VIFNSW
```

- **SMAD7 SNP:N424k**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo
sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGEEDEEEGAGGGGGGGELRGEGATDSRAHGAGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDRLHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQLVQKVRISKIGCGIQLTREVDGVVWVYNRSSYPIFIKSATLDNPDSTRLLVHKV
FPGFSIKAFDYEKAYSLSQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSPCWLE
VIFKSR
```

- **SMAD7 SNP:P416L**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo
sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGEEDEEEGAGGGGGGGELRGEGATDSRAHGAGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDRLHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQLVQKVRISKIGCGIQLTREVDGVVWVYNRSSYPIFIKSATLDNPDSTRLLVHKV
FPGFSIKAFDYEKAYSLSQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSCLWLE
VIFNSR
```

- **SMAD7 SNP:R409H**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo
sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGEEDEEEGAGGGGGGGELRGEGATDSRAHGAGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDRLHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQLVQKVRISKIGCGIQLTREVDGVVWVYNRSSYPIFIKSATLDNPDSTRLLVHKV
FPGFSIKAFDYEKAYSLSQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTHQFISSPCWLE
VIFNSR
```

- **SMAD7 SNP:R409L**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo
sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGEEDEEEGAGGGGGGGELRGEGATDSRAHGAGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDRLHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQLVQKVRISKIGCGIQLTREVDGVVWVYNRSSYPIFIKSATLDNPDSTRLLVHKV
FPGFSIKAFDYEKAYSLSQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTLQFISSPCWLE
VIFNSR
```

Chapitre VII : Annexes

- **SMAD7 SNP:Y407S**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGGEDEEEEGAGGGGGGGGELRGEGATDSRAHGAGGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDLRHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQVLVQKVRISKIGCGIQLTREV DGVVWVYNRSSYP IFIKSATLDNPD SRTLLVHKV
FPGFSIKAFDYEKAYSLQRPNDHEFMQQPWTGFTVQISFVKGWGQCSTRQFISSPCWLE
VIFNSR
```

- **SMAD7 SNP: G402V**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGGEDEEEEGAGGGGGGGGELRGEGATDSRAHGAGGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDLRHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQVLVQKVRISKIGCGIQLTREV DGVVWVYNRSSYP IFIKSATLDNPD SRTLLVHKV
FPGFSIKAFDYEKAYSLQRPNDHEFMQQPWTGFTVQISFVKVWGQCYTRQFISSPCWLE
VIFNSR
```

- **SMAD7 SNP:S336C**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGGEDEEEEGAGGGGGGGGELRGEGATDSRAHGAGGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDLRHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQVLVQKVRISKIGCGIQLTREV DGVVWVYNRCSYPIFIKSATLDNPD SRTLLVHKV
FPGFSIKAFDYEKAYSLQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSPCWLE
VIFNSR
```

- **SMAD7 SNP: R314Q**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGGEDEEEEGAGGGGGGGGELRGEGATDSRAHGAGGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDLRHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPOGNGFCLGQ
LNSDNKSQVLVQVQSKIGCGIQLTREV DGVVWVYNRSSYP IFIKSATLDNPD SRTLLVHKV
FPGFSIKAFDYEKAYSLQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSPCWLE
VIFNSR
```

- **SMAD7 SNP:R200Q**

```
>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo sapiens OX=9606 GN=SMAD7 PE=1 SV=1
MFRTKRSALVRRLLWRSRAPGGGEDEEEEGAGGGGGGGGELRGEGATDSRAHGAGGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
```

Chapitre VII : Annexes

TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDRLRHSSEVKRLCCC
ESYGKINPELVCCNPHHLSQLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPQNGFCLGQ
LNSDNKSQLVQKVRSKIIGCGIQLTREV DGVVWVYNRSSYPIFIKSATLDNPD SRTLLVHKV
FPGFSIKAFDYEKAYSLQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSPCWLE
VIFNSR

- **SMAD7 SNP:K175T**

>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo sapiens OX=9606 GN=SMAD7 PE=1 SV=1

MFRTKRSALVRRLLWRSRAPGGEDEEEGAGGGGGGGELRGEGATDSRAHGAGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWPDRLRHSSEVTRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPQNGFCLGQ
LNSDNKSQLVQKVRSKIIGCGIQLTREV DGVVWVYNRSSYPIFIKSATLDNPD SRTLLVHKV
FPGFSIKAFDYEKAYSLQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSPCWLE
VIFNSR

- **SMAD7 SNP:P166R**

>sp|O15105|SMAD7_HUMAN Mothers against decapentaplegic homolog 7 OS=Homo sapiens OX=9606 GN=SMAD7 PE=1 SV=1

MFRTKRSALVRRLLWRSRAPGGEDEEEGAGGGGGGGELRGEGATDSRAHGAGGGGPGRAGC
CLGKAVRGAKGHHHPHPPAAGAGAAGGAEADLKALTHSVLKKLKERQLELLLQAVESRGG
TRTACLLLPGRLDCRLGPGAPAGAQAQPPSSYSPLLLLCKVFRWRDLRHSSEVKRLCCC
ESYGKINPELVCCNPHHLSRLCELESPPPPYSRYPMDFLKPTADCPDAVPSSAETGGTNY
LAPGGLSDSQLLLEPGDRSHWCVVAYWEEKTRVGRLYCVQEPSLDIFYDLPQNGFCLGQ
LNSDNKSQLVQKVRSKIIGCGIQLTREV DGVVWVYNRSSYPIFIKSATLDNPD SRTLLVHKV
FPGFSIKAFDYEKAYSLQRPNDHEFMQQPWTGFTVQISFVKGWGQCYTRQFISSPCWLE
VIFNSR

Résumé

Dans cette étude, nous avons réalisé des analyses *in silico* des Single Nucleotide Polymorphisms (SNPs) ayant des conséquences fonctionnelles sur les protéines TGFBR2 et SMAD7 dans le contexte du cancer colorectal. Nous avons réalisé des prédictions des modifications post-traductionnelles associées aux SNPs identifiés comme potentiellement délétères, ainsi que la modélisation moléculaire des structures tridimensionnelles des protéines pour une visualisation plus précise des altérations.

Nous avons exploité plusieurs outils bioinformatiques tels que GPS6.0, MusiteDeep, NETPhos 3-1, GPS-MSP, GPS-UBER, PRmePRed, GPS.PBS, CarSite-II, GPS-YNO2 et GPS-SNO 1.0 pour prédire les modifications post-traductionnelles. Ensuite, nous avons utilisé les logiciels Phyre2 et SWISS-MODEL pour réaliser la modélisation moléculaire de ces protéines mutées et ERRAT/PROCHECK pour évaluer la qualité de nos résultats.

Nos résultats indiquent que certaines variations dans la protéine TGFBR2 (R448G, R520L, R522Q et K277N) peuvent potentiellement induire diverses modifications post-traductionnelles telles que la méthylation et l'ubiquitination.

En conclusion notre étude *in silico* fournit une base préliminaire intéressante concernant l'impact possible de certains SNPs sur TGFBR2 dans le contexte du cancer colorectal. Ces résultats pourront être utilisés comme base de départ pour des études expérimentales visant à valider les effets de ces SNPs sur les protéines et leur rôle dans le développement de cette maladie.

Mots clés : cancer colorectal, modélisation des protéines, modifications post traductionnelles, nsSNP, SMAD7, TGFBR2.

Abstract

In this study, we performed *in silico* analyses of Single Nucleotide Polymorphisms (SNPs) with functional consequences on TGFBR2 and SMAD7 proteins in the context of colorectal cancer. We predicted post-translational modifications associated with SNPs identified as potentially deleterious, as well as performing molecular modeling of the three-dimensional structures of proteins for more precise visualization of alterations.

We used several bioinformatics tools such as GPS6.0, MusiteDeep, NetPhos 3-1, GPS-MSP, GPS-UBER, PRmePRed, GPS.PBS, CarSite-II, GPS-YNO2 and GPS-SNO 1.0 to predict post-translational modifications. Then we used Phyre2 and SWISS-MODEL software to perform molecular modeling of these mutated proteins and ERRAT/PROCHECK to evaluate the quality of our results.

Our results indicate that certain variations in TGFBR2 protein (R448G, R520L, R522Q, and K277N) can potentially induce various post-translational modifications such as methylation and ubiquitination.

In conclusion our *in silico* study provides an interesting preliminary basis regarding the possible impact of some SNPs on TGFBR2 within the context of colorectal cancer. These results could be used as a starting point for experimental studies aiming to validate the effects of these SNPs on proteins and their role in disease development.

Keywords: colorectal cancer, protein modeling, post-translational modifications, nsSNP, SMAD7, TGFBR2.

ملخص

في هذه الدراسة ، أجرينا تحليلات سيليكو لتعدد أشكال النوكليوتيدات المفردة (SNPs) مع عواقب وظيفية على بروتينات TGFBR2 و SMAD7 في سياق سرطان القولون والمستقيم. لقد أجرينا تنبؤات بتعدلات ما بعد الترجمة المرتبطة بـ SNPs التي تم تحديدها على أنها ضارة ، بالإضافة إلى النمذجة الجزيئية للهياكل ثلاثية الأبعاد للبروتينات للحصول على تصور أكثر دقة للتغيرات.

استفدنا من العديد من أدوات المعلوماتية الحيوية مثلًا GPS6.0 و MusiteDeep و NETPhos 3-1 و GPS-MSP و GPS-UBER و PRmePRed و GPS.PBS و CarSite-II و GPS-YNO2 و GPS-SNO 1.0 للتنبؤ بالتغيرات اللاحقة. ترجمة. بعد ذلك ، استخدمنا برنامج Phyre2 و SWISS-MODEL لإجراء النمذجة الجزيئية لهذه البروتينات المتحولة و ERRAT / PROCHECK لتقييم جودة نتائجنا. تشير نتائجنا إلى أن بعض الاختلافات في بروتين TGFBR2 (R448G و R520L و R522Q و K277N) يمكن أن تؤدي إلى تعدلات مختلفة بعد الترجمة مثل المثيلة والانتشار.

في الختام ، توفر دراستنا في السيليكو أساسًا أوليًا مثيرًا للاهتمام فيما يتعلق بالتأثير المحتمل لبعض SNPs على TGFBR2 في سياق سرطان القولون والمستقيم. يمكن استخدام هذه النتائج كنقطة انطلاق للدراسات التجريبية التي تهدف إلى التحقق من صحة آثار هذه النيوكليوتيد على البروتينات ودورها في تطور هذا المرض.

الكلمات المفتاحية: سرطان القولون والمستقيم ، نمذجة البروتين ، تعدلات ما بعد الترجمة ، nsSNP ، SMAD7 ، TGFBR2