

Université de Saad Dahleb 1

Faculté des sciences de la nature et de la vie
Département de Biologie et Physiologie Cellulaire



MEMOIRE DE MASTER

En Sciences biologiques
Spécialité : Bioinformatique

Etude par outils bioinformatiques de différents variant XPC et de leur impact dans Xeroderma Pigmentosum

Par

Ichrak MEDDAH

Karima AISSANI

Devant le jury composé de :

Mme Benmensour	MCA	Présidente	USDB
Mme Eddaikra A.	MCA	Examinatrice	USDB
Mr Bessaad M.	MCB	Promoteur	USDB
Mme Khemili Talbi S.	MCA	Co-promotrice	UMBB
Mme Gussaibia N	MCB	Invité	USDB

Blida, Septembre 2017

ملخص

هذا البحث جزء من الدراسة في السيليكو (*in silico*) باستعمال برامج معلوماتية و نمذجة جزئية نهدف لتحليل مختلف الطفرات وتأثيرها في (*xeroderma pigmentosum*). نهدف إلى تدريج قائمة هذه الطفرات التي يمكن ان تغير خصائص البروتين (*XPC*) و تصميم نموذج هيكله ثلاثي الأبعاد الذي ليس محددًا بعد. لهذا الغرض، استخدمنا البرامج المعلوماتية (*UCSC*) و (*NCBI*) لتدريج قائمة الطفرات ودراسة تأثيرها على هذا البروتين، و سيتم استغلال هذه القائمة في جزء النمذجة الجزيئية حيث سنتابع ثلاثة نهج مختلفة لتصميم النموذج ثلاثي الأبعاد (النمذجة بالمقارنة باستعمال البرنامج *SwisModel* ، النمذجة بالإعتراف على الطيات باستعمال البرنامج *Phyre2* ، و النمذجة بالإعتراف على الطيات مشتركة مع آب إينيسيو (*Ab initio*) باستعمال البرنامج *I-TASSER*). بالإضافة إلى استعمال البرنامج *Pymol* الذي سمح لنا بقياس التشابه الهيكلي بين نماذج *XPC* المصممة وشظايا البروتين (*XPC*) التي تم تحديدها تجريبيا من خلال حساب جذر متوسط التربيع (*rms*) ، ما دل على أن النموذج المصمم من طرف البرنامج *I-TASSER* قد أعطى أفضل تشابه هيكلي مع جذر متوسط التربيع (*rms*) منخفض جدا مقارنة بالهياكل المصممة الأخرى ، ما سمح لنا الاحتفاظ به واعتباره كنموذج ثلاثي الأبعاد محتمل للبروتين *XPC*. قد تم إجراء الالتحام الجزيئي لحساب التوجه المفضل من جزيئين من البروتين من أجل تشكيل مجمع مستقر بين (*XPC*) والبروتينات الأخرى). البروتين (*XPC*) المتحول وفقا للطفرات المرتبطة بهذا المرض يسمح بتحليل تأثير الطفرة على خصائص البروتين (وظيفتها، وتشكيل المجمعات)

Résumé

Cette recherche fait partie de l'étude *in silico* (utilisant des approches bioinformatiques et une modélisation moléculaire) pour analyser les différents variants XPC et leur impact sur Xeroderma pigmentosum. Nous visons à énumérer les SNP qui peuvent modifier les propriétés de la protéine XPC et prédire sa structure 3D qui est inconnue. À cette fin, nous avons utilisé le serveur UCSC et NCBI pour énumérer et étudier les SNP affectant notre protéine, ces derniers seront exploités dans la partie de modélisation moléculaire où nous suivrons trois approches de prédiction de structure 3D différentes par trois programmes différents (la modélisation comparative par le programme SwissModel, modélisation par la reconnaissance de repliements par le programme Phyre2, et reconnaissance des repliements combinés avec *ab initio* par le programme I-TASSER); en outre l'utilisation du logiciel Pymol nous a permis de mesurer la similarité structurelle entre les modèles XPC prédits et les fragments XPC expérimentaux en calculant les rms ce qui montre que le modèle prédit de la XPC par I-TASSER a donné la meilleure similarité structurelle avec des rms très faible par rapport aux autres modèles cela nous a permis de le retenir comme une structure probable de la XPC, ainsi faire une mutagenèse de la XPC prédite par I-TASSER selon les SNPs étudiés, un amarrage moléculaire a été effectué pour calculer l'orientation préférée de deux molécules afin de former un complexe stable (protéines XPC) par le serveur ClusPro. XPC muté selon les SNPs associés à la maladie a permis l'analyse de l'impact de la mutation sur les propriétés de la protéine (sa fonction, la formation des complexes).

Mots clés : Xeroderma pigmentosum, XPC, SNP, modélisation moléculaire, prédiction structurale.

Abstract

This research is part of the *in silico* study (using bioinformatic approaches and molecular modeling) to analyze the different XPC variants and their impact on Xeroderma pigmentosum. We aim to enumerate SNPs that can modify the properties of the XPC protein and predict its 3D structure which is unknown. To this end, we used the UCSC and NCBI server to enumerate and study SNPs affecting our protein, the latter will be exploited in the molecular modeling part where we will follow three different 3D structure prediction approaches by three different programs (modeling comparative study by the SwissModel program, modeling by recognition of Phyre2 plots, and recognition of combinations with *ab initio* by the I-TASSER program); the use of the Pymol software allowed us to measure the structural similarity between the predicted XPC models and the experimental XPC fragments by computing the rms, which shows that the predicted model of the XPC by I-TASSER gave the merileure structural similarity with very small rms compared to other lmodels this allowed us to retain it as a probable structure of the XPC, thus making a mutagenesis of the XPC predicted by I-TASSER according to the SNPs studied, a molecular ammarage was performed to calculate the preferred orientation of two molecules to form a stable complex (XPC proteins) by the ClusPro server. XPC mutated according to the SNP associated with the disease allowed the analysis of the impact of the mutation on the properties of the protein (its function, complex formation)

Key words : Xeroderma pigmentosum, XER, SNP, molecular modéling, structural prediction.

Remerciements.

Ce mémoire n'aurait jamais pu voir le jour sans le courage et la volonté que nous a donné le bon Dieu.

A notre promoteur Mr BESSAAD Mohamed El amine qui nous a honoré de leur confiance et nous a encadré et nous a bien orienté, on vous remercie énormément.

Nos remerciements à Mme KHEMILI-TALBI Souad de nous avoir accueilli avec gentillesse et de nous offrir son aide avec simplicité et ouverture d'esprit.

A Monsieur SENINET Mohamed El Hadi qui nous a mis en disposition des données de séquençage et qui nous a accompagné dans le départ de notre travail.

Nous remercions nos familles qui nous ont accompagnés tout au long de notre carrière académique et leur soutien à nous.

A tous mes enseignants ainsi que les membres de jury qui nous ont fait l'immense honneur de présider et d'examiner ce travail.

Nos remerciements les plus sincères à toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce travail.

Liste des abréviations.

ADN : Acide Désoxyribonucléique.

XP : Xeroderma Pigmentosum.

EMG : Électromyogramme.

IRM : Imagerie par Résonance Magnétique.

UDS : Unscheduled DNA Synthesis.

PCR-RFLP : Restriction Fragment Length Polymorphism

PCR-SSCP : Single Strand Conformation Polymorphism.

UV : Ultraviolet.

DSB : Double Strand Break.

NHEJ : Non-Homologous End-Joining.

DNA-Pkcs : DNA-dependent Protein Kinase.

XRCC4 : X-ray repair cross-complementing protein 4.

HR : Homologue Recombination.

ATM : Ataxia Telangiectasia Mutated.

RPA : Replication Protein A.

MMR : DNA Mismatch Repair.

PCNA : Proliferating Cell Nuclear Antigen.

RFC : Replication Factor C.

MSH : Melanocyte-Stimulating Hormone.

BER : Base Excision Repair.

NER : Nucléotide Excision Repair.

6-4PP : Pyrimidine 6-4 Pyrimidine.

GG-NER : Global Genome NER.

TC-NER : Transcription Coupled NER.

CSA/B : Cockayne Syndrome A/B.

TFIIH : Transcription Factor II H.

TTD : Trichothiodystrophie.

ERCC : Excision Repair Complementing Cross.

XPA : Xeroderma Pigmentosum groupe de complémentation A.

XPB : Xeroderma Pigmentosum groupe de complémentation B.

XPC : Xeroderma Pigmentosum groupe de complémentation C.

XPD : Xeroderma Pigmentosum groupe de complémentation D.

XPE : Xeroderma Pigmentosum groupe de complémentation E.

DDB2 : DNA Damage-Binding protein 2.

XPF : Xeroderma Pigmentosum groupe de complémentation F.

XPG : Xeroderma Pigmentosum groupe de complémentation G.

XPV : Xeroderma Pigmentosum groupe de complémentation Variet.

POLH : Polymérase éta.

UV-DDB

SNP : Single Nucleotide Polymorphisms.

CNV : Copy Number Variation.

UCSC : Université Californie Santa Cruz.

NCBI : National Center for Biotechnology Information.

3D : Tridimensionnelle.

BLAST : Basic Local Aligement Search Tool.

PDB : Protein Data Bank.

Phyre2 : Protein Homology/analogY Recognition Engine V 2.0.

I-TASSER : Iterative Threading ASSEmby Refinement.

Liste des figures

Figure 1: Réparation des cassures double brin.	7
Figure 2: Mécanisme de réparation par mésappariement MMR.	9
Figure 3: Modèle du mécanisme de réparation par excision de nucléotide NER.	11
Figure 4: Localisation chromosomique du gène XPA.	12
Figure 5: Localisation chromosomique du gène ERCC3	13
Figure 6: Localisation chromosomique du gène XPC.	13
Figure 7: Localisation chromosomique du gène ERCC2	14
Figure 8: Localisation chromosomique du gène DDB2.	14
Figure 9: Localisation chromosomique du gène ERCC4	15
Figure 10: Localisation chromosomique du gène ERCC5	16
Figure 11: Localisation chromosomique du gène de POLH	16
Figure 12: Protéines qui interagissent avec XPC dans la reconnaissance des dégâts de l'ADN.	18
Figure 13: Structure du complexe entre XPC (en rose) et TFIIH (en vert) (code pdb : 2rvb).	19
Figure 14 : Méthodologie de notre travail de fin de cycle.	21
Figure 15 : Page d'accueil UCSC.	22
Figure 16 : Interface de Génome Browser permettant la sélection de paramètres de filtre des SNPs ..	23
Figure 17 : Page d'accueil de <i>Table Browser</i> et de paramètres de filtrage des SNPs annotés.	24
Figure 18 : Page d'accueil de NCBI.	24
Figure 19 : Page d'accueil de Uniprot.	25
Figure 20 : Page d'accueil de Blastp.	26
Figure 21 : Page d'accueil de PDB.	27
Figure 22 : Interface d'interaction de la séquence protéique de XPC dans le serveur Swiss-Model	28
Figure 23 : Page d'accueil du Phyre2	29
Figure 24 : Page d'accueil de I-TASSER	30
Figure 25 : Page d'accueil de STRING.	31
Figure 26 : Page d'accueil de ClusPro	32
Figure 27 : Page d'application Pymol (mutagenèse).	33
Figure 28 : Page d'accueil Interpro.	34
Figure 29 : Localisation chromosomique du gène XPC (3p25)	35
Figure 30 : La liste de tous les SNPs positionnés sur le gène XPC via Génome Browser.	35
Figure 31 : les listes des SNPs dans Génome Browser.	36
Figure 32 : Listes des SNPs annotés dans Table Browser.	37
Figure 33 : La séquence protéique du XPC sous format fasta tiré de la banque de données Uniprot. ..	38
Figure 34 : (A) Structure du complexe Centrine2-XPC déterminée par RMN (code PDB 2a4j) ; (a) fragment XPC et (b) Centrine-2.	40
Figure 35 : (A) Structure du complexe Centrine2-XC déterminée par diffraction des rayons X (code PDB 2GGM; (a) fragment XPC et (b) Centrine2.	41
Figure 36 : (A) Structure du complexe TFIIH-XPC déterminée par RMN (code PDB 2rvb) ; (a) fragment XPC et (b) TFIIH.	41
Figure 37 : Modèles de structures tridimensionnelles du XPC prédits par SwissModel. ;	44
Figure 38 : Modèles de structures tridimensionnelles du XPC prédits par Phyre2 (XPC-Phyre2.pdb).	45
Figure 39 : Modèles de structures tridimensionnelles du XPC prédits par I-TASSER ; (a) XPC-model1-TASSER.pdb, (b) XPC-model2-TASSER.pdb et (c) XPC-model3-TASSER.pdb.	46

Figure 40 : Superposition des structures des fragments déterminées expérimentalement avec la structure du modèle 3 prédit par I-TASSER .	48
Figure 41 : Interactions XPC-Protéines d'après la base des données String.	49
Figure 42 : Superposition de la structure du complexe XPC-CETN prédite par ClusPro (vert-rouge) avec celle déterminée expérimentalement (code PDB 2ggm) en jaune fragment XPC et en Cyan la chaîne de Centrine2.	52
Figure 43 : La structure du complexe XPC-TFIIH prédite par ClusPro	53
Figure 44 : Les domaines protéiques du XPC d'après la base de données InterPro.	54
Figure 45 : L'ensemble des rs pathogènes (en sticks roses) positionnés sur la structure du modèle 3 de la XPC prédit par I-TASSER (en cartoon vert).	56
Figure 46 : rs (en rouge) positionnés dans le domaine d'interaction entre la structure de la protéine XPC prédite par I-TASSER (en vert) et le domaine TFIIH (en bleu).	57

Liste des tableaux

Tableau I: Les groupes de complémentation de XP.....	17
Tableau II : Signification des couleurs Genome Browser.....	37
Tableau III : Résultats d'alignement de séquence de XPC contre la PDB généré par Blastp.....	39
Tableau IV : Tableau récapitulatif des résultats de la prédiction de la structure 3D de la XPC	43
Tableau V : Résultats de la mesure de la similarité structurale des modèles prédits avec les fragments de la XPC déterminés expérimentalement par le calcul des rms.....	47
Tableau VI : Valeurs de la fonction de score de ClusPro des complexes prédits et valeurs du rms obtenues après superpositions de ces modèles avec les complexes respectifs déterminés expérimentalement	51
Tableau VII : Tableau des rs associés à la maladie XPC et leurs localisations dans ses domaines protéiques fonctionnels.....	55

Sommaire

Introduction	1
I. Synthèse bibliographique	
I.1 Xeroderma Pigmentosum.....	3
I. 2 Epidémiologie.....	3
I.3 Symptomatologie.....	4
I.4 Diagnostic.....	5
I.4.1 Tests biologiques.....	5
I.4.2 Diagnostic de la biologie moléculaire.....	5
I.5 Etiologie.....	5
I.6 Lésions sur l'ADN et leurs réparations.....	5
I.6.1 Lésions sur l'ADN.....	5
I.6.2 Systèmes de réparation.....	6
I.7 Physiopathologie.....	12
I.7.1 Xeroderma Pigmentosum groupe de complémentation XPA.....	12
I.7.2 Xeroderma Pigmentosum groupe de complémentation XPB (ERCC3).....	12
I.7.3 Xeroderma Pigmentosum groupe de complémentation XPC.....	13
I.7.4 Xeroderma Pigmentosum groupe de complémentation XPD(ERCC2).....	13
I.7.5 Xeroderma Pigmentosum groupe de complémentation XPE(DDB2).....	14
I.7.6 Xeroderma Pigmentosum groupe de complémentation XPF(ERCC4).....	15
I.7.7 Xeroderma Pigmentosum groupe de complémentation XPG(ERCC5).....	15
I.7.8 Xeroderma Pigmentosum groupe de complémentation XPV(POLH).....	16
I.8 Protéomique de XPC.....	17
I.8.1 Reconnaissance des dégâts d'ADN dans GG-NER.....	18
I.8.2 Complexe de formation d'incision à TR-NER.....	19
II. Matériels et méthodes	
II.1 Matériels.....	20
II.2 Méthodes.....	20
II.2.1 Partie 1 : Listes des SNPs.....	22
a) UCSC.....	22
b) Le serveur NCBI.....	24
II.2.2 Partie 2 : Modélisation moléculaire de XPC.....	25
1. Prédiction de structure 3D de XPC.....	25
2. Amarrage protéine-protéine.....	30

3. Visualisation moléculaire de XPC.....	32
4. Recherche de domaines protéiques et localisation des SNPs	33
III. Résultats et discussions	
III.1 Liste des SNPs par Génome Browser	35
III.2 Modélisation moléculaire de la protéine XPC.....	37
III.2.1 Recherche de similarité de la séquence cible/patrons par Blastp/PDB.....	38
III.2.2 Collection des structures expérimentales de XPC.....	40
III.2.3 Prédiction de structure 3D de la protéine XPC	42
a) Modélisation comparative de la structure de la XPC par Swiss-Model.....	43
b) Modélisation par reconnaissance de repliements de la structure de la XPC par Phyre2.....	44
c) Modélisation de la structure de la XPC par ab initio combinée à la reconnaissance de repliements par I-TASSER	45
d) Analyse des modèles des structures prédits pour la XPC.....	46
III.2.4 Amarrage protéine-protéine (XPC-protéine) par ClusPro	48
III.2.5 Localisation des rs dans la structure de XPC prédite	54
Conclusion	58
Références bibliographiques	60
Annexes	65

Introduction

Xeroderma pigmentosum (XP) est une maladie à transmission autosomique récessive dont la prévalence est très importante en **Afrique de Nord** et en **Moyen Orient** où la consanguinité est plus fréquente.

Elle est due à une sensibilité accrue de l'ultraviolet et de la lumière du soleil, une inefficacité du système de réparation (NER) qui implique un *Xeroderma pigmentosum*.

Cette maladie constitue un groupe de maladies dont huit gènes différents (situés sur des chromosomes différents) qui lorsqu'ils sont mutés entraînent un *Xeroderma pigmentosum*.

Il a été démontré que l'hétérogénéité clinique de cette maladie est liée à un défaut dans les gènes du système de réparation par excision-resynthèse des nucléotides (*Nucleotide Excision Repair*, NER) pour les sept premiers groupes génétiques (A-G), et à une anomalie des gènes de la transcription pour le huitième groupe (*Xeroderma pigmentosum* variant, XPV).

A l'heure actuelle, Il n'existe pas encore de traitement curatif permettant de soigner les malades souffrant de XP : les dommages d'ADN sont cumulés et irréversibles, ainsi le diagnostic précoce de la maladie est capital pour mettre en route sans délais les mesures de photo protection, seul moyen disponible, pour prévenir ou tout au moins retarder l'apparition de néoplasies cutanées et oculaires.

Cette étude s'inscrit dans le cadre de l'étude de différents variant de XPC et de leurs impacts dans *Xeroderma pigmentosum* par outils bioinformatiques dont le but est de lister les SNPs et de modéliser la structure de la protéine XPC par différentes approches de modélisation moléculaire, ainsi étude de son interaction avec d'autres protéines et analyse de l'impact de quelques SNPs sur la fonction de la protéine XPC.

Pour pouvoir couvrir tous les aspects de ce thème, le mémoire a été structuré en trois chapitres :

- ✓ Après une introduction générale, le premier chapitre présente des généralités sur la maladie de *Xeroderma pigmentosum* et une synthèse bibliographique qui présente les différents types de XP et la protéomique de la protéine XPC.
- ✓ Le deuxième chapitre présente la méthodologie du travail. En premier lieu, le listing des SNPs à partir d'un fichier de séquençage d'ADN et d quelques outils bioinformatiques, ensuite la modélisation moléculaire de la protéine XPC.

- ✓ Le troisième chapitre porte sur l'analyse des résultats dont le but est de déterminer la structure 3D de la protéine XPC et de la valider par quelques SNPs associés à la maladie *Xeroderma pigmentosum*.
- ✓ Et enfin les résultats sont récapitulés sous forme d'une conclusion générale avec quelques recommandations pour les futurs travaux de recherches.

Synthèse

Bibliographique

I.1 Xeroderma Pigmentosum.

Les premiers cas décrits de *Xeroderma Pigmentosum* remontent à 1863 par le **Docteur Mortiz Kaposi** dermatologue hongrois qui lui donna son nom. « **Xeroderma** » signifie « peau sèche » et le terme « **Pigmentosum** » fait référence à l'aspect très pigmenté de la peau exposé au soleil (*Encyclopédie Orphanet Grand Public.,2008*), elle est appelée aussi la maladie des enfants de la lune puisque que ce sont les enfants qui subissent.

Xeroderma Pigmentosum est une maladie génétique autosomique récessive caractérisée par une sensibilité accrue à l'ultraviolet et à la lumière du soleil (*Ozoux., 2013*) due à une inefficacité de système de réparation de l'ADN endommagé.

Pour la plupart des enfants elle est caractérisée par des cancers cutanés et de dégénérescences neurologiques, des changements pigmentaires oculaires tels que des taches de rousseur, photophobie, conjonctivite. (*Bhutto et Kirk., 2008*).

Xeroderma Pigmentosum constitue un groupe de maladies, on dénombre en tous huit gènes lorsqu'ils sont mutés entraînent un *Xeroderma Pigmentosum* (*Ozoux, 2013*), elle est subdivisée en huit groupes de complémentation, selon le gène affecté : XP classique (XPA à XPG) et XP variant (XPV).

La gravité des symptômes, l'âge de leur apparition dépend du gène muté. Dans les formes les plus graves de la maladie, outre l'hypersensibilité à la lumière, les malades peuvent présenter des neuropathies et des anomalies du développement.

I. 2 Epidémiologie.

La prévalence de *Xeroderma Pigmentosum* est estimée à 1/1 000 000 aux **Etats-Unis** et en **Europe** (*Kleijer et al., 2008*), alors que certaines populations présentent une prévalence beaucoup plus élevée comme au **Japon** où elle est estimée à 1/40 000 à 1/100 000(*Hirai et al., 2006*).

En Afrique du Nord (Tunisie, Algérie, Maroc, Libye et Egypte) et **au Moyen-Orient** (Turquie et Syrie) la prévalence est également augmentée, notamment dans les communautés où la consanguinité est fréquente (*Kraemer et DiGiovanna.,2016*).

Dr Kamel Boucherma a indiqué " ...A ce jour, il n'existe pas des statistiques exactes en Algérie qui renseignent sur le nombre de personnes atteintes du *Xeroderma Pigmentosum* mais il dépasse toutefois les 500 cas...".

ALGER, un spécialiste en ophtalmologie à l'hôpital Beni Messous (Alger) a annoncé " ...Plus de 500 enfants souffrant de *Xeroderma Pigmentosum* ou « enfants de la lune » ont été recensés au plan national...".

Dr Boucherma avant de souligner que les enfants atteints de cette maladie dans les régions éloignées n'ont pu être recensés a précisé" ...Ce chiffre a été établi à partir des enfants soumis à des traitements dans les différents hôpitaux du pays...".

D'après une étude qui a été faite en Algérie les groupes XPA et XPC sont les plus réponsus en **Afrique du Nord** (*Bensenouci et al.,2016*).

I.3 Symptomatologie.

Xeroderma Pigmentosum (XP) doit être suspecté chez les personnes présentant les symptômes suivants : peau, œil, système nerveux et antécédents familiaux (*Kraemer et DiGiovanna.,2016*).

➤ **Peau.**

-Sensibilité aiguë au soleil.

-Pigmentation marquée des taches de rousseur sur le visage avant l'âge de deux ans.

-Cancer de la peau au cours de la première décennie de la vie.

➤ **Yeux.**

-Photophobie avec injection conjonctivale importante.

-Une kératite sévère, entraînant parfois une opacification cornéenne et une vascularisation.

-Augmentation de la pigmentation des paupières avec perte de cils.

-Atrophie de la peau des paupières.

➤ **Système neuronale.**

-Diminution ou absence des réflexes d'étirement tendineux profond. Les vitesses d'EMG et de conduction nerveuse peuvent montrer une neuropathie axonale (ou mixte).

-Perte auditive progressive.

-Microcéphalie acquise. La tomодensitométrie et l'IRM du cerveau peuvent montrer des ventricules agrandis avec l'amincissement du cortex et l'épaississement des os du crâne.

-Trouble cognitif progressif.

I.4 Diagnostic.

Le diagnostic du XP repose sur des observations cliniques de la peau, des yeux et de possibles manifestations neurologiques. Lorsque de jeunes enfants consultent pour une extrême réactivité suite à une exposition au soleil avec l'apparition de nombreuses taches de rousseur sur le visage, le médecin peut suspecter un cas de XP, des tests de réparation de l'ADN sont réalisés après exposition aux UV (*Bensenouci.,2016*).

I.4.1 Tests biologiques.

- UDS, synthèse d'ADN non programmée.
- Mesure de l'hypersensibilité cellulaire aux rayons ultraviolets.
- Le test HCR.

I.4.2 Diagnostic de la biologie moléculaire.

Il est possible d'identifier aisément les mutations des gènes XP soit par PCR-RFLP en utilisant plusieurs enzymes de restriction ou bien par PCR-SSCP pour détecter de nouvelles mutations ponctuelles. La biologie moléculaire nécessite l'utilisation de la radioactivité (*Orita et al., 1989*).

D'autres techniques plus performantes et qui sont l'actualité d'aujourd'hui telles que le séquençage et les outils bioinformatiques.

I.5 Etiologie.

Le XP est dû à des mutations de huit gènes impliqués dans la réparation de l'ADN. Sept d'entre eux, XPA à XPG sont impliqués dans la réparation par excision de nucléotides (NER).

Le gène XPV (POLH) code une ADN polymérase ϵ qui permet la réplication de l'ADN endommagé par les UV (*Fassihi.,2011*).

I.6 Lésions sur l'ADN et leurs réparations.

I.6.1 Lésions sur l'ADN.

Divers types de lésions apparaissent sur l'ADN elles peuvent être d'origine endogène ou exogène tels que les agents physiques ou chimiques, les cancérogènes chimiques, les radiations ultraviolets ou ionisantes qui entraînent l'apparition des deux lésions majeures entre deux pyrimidines adjacentes (les dimères pyrimidines (6-4) pyrimidines) et de nombreuses lésions mineures (*Ziani., 2014*).

Afin de maintenir l'intégrité du génome, il existe différents systèmes de réparation prenant en charge de façon spécifique une ou plusieurs lésions.

I.6.2 Systèmes de réparation.

Les cassures doubles brins DSB sont induites par les radiations ionisantes et les rayons X et par la réplication de certaines cassures simple brin.

La détection d'un DSB déclenche une cascade de réaction ayant pour but l'arrêt de cycle cellulaire et le recrutement de facteur de réparation (*Ziani., 2014*).

La réparation des DSB se fait par deux principales voies.

a. Recombinaison non homologue par jonction.

La liaison de l'hétérodimère KU70-KU80 aux extrémités de l'ADN, suivie par le recrutement et l'activation de la kinase DNA.PKcs dont le rôle est de maintenir à proximité les deux extrémités de cassures, et de recruter et phosphoryler Ardenis, ce qui stimule son activité exonucléasique nécessaire à la préparation des extrémités de l'ADN, avant ligation par le complexe XRCC4-ligase IV (*Panier et Boulton, 2014*) (**Figure 1a**).

b. Recombinaison homologue.

-Cette voie implique le recrutement du complexe MRN constitué des trois molécules Mre11, Nbs1 et Rad50 au niveau des extrémités de la cassure. La phosphorylation de ce complexe par ATM stimule son activité 5'-3' exonucléasique et permet la formation de deux extrémités simple brin 3' qui sont immédiatement protégées par recouvrement de RPA (*Ozoux., 2013*).

-RAD51, en collaboration avec ses paralogues (RAD51B, C, D, XRCC2 et XRCC3), RAD52 et RAD54, qui forment des nucléofilaments autour de ces mêmes extrémités, permet le déplacement de RPA. Ces étapes sont dépendantes des protéines BRCA1 et BRCA2 (*Ozoux., 2013*).

-Rad51 par son activité recombinase permet ensuite les étapes de reconnaissance de la séquence homologue sur la chromatide sœur et l'invasion de brin. La nouvelle synthèse de brin est induite à l'extrémité 3' de chaque brin et s'étend au-delà de l'endroit de la lésion. Elle donne naissance à une structure intermédiaire appelée **jonction de Holliday** dont la particularité nécessite l'action d'enzymes spécifiques, les resolvases, pour permettre la restauration des deux fragments d'ADN originaux (*Ozoux., 2013*) (**Figure 1b**).

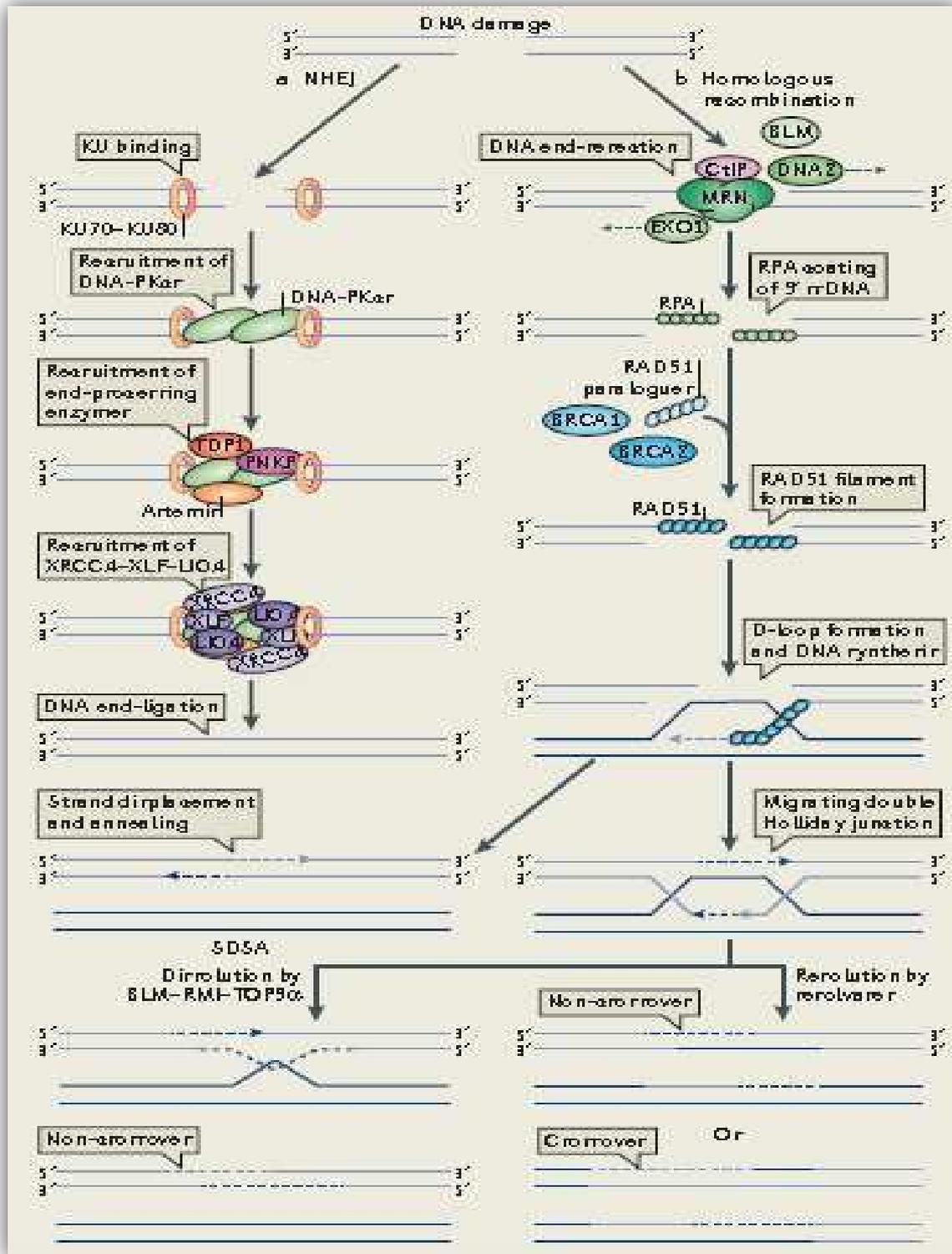


Figure 1: Réparation des cassures double brin.

1a- recombinaison non homologue par jonction NHEJ, 1b-Réparation par recombinaison homologue HR et (Panier and Boulton., 2014).

Une excision de l'altération et le remplacement.

a. Réparation d'un mésappariement.

Cette voie de réparation est activée par la présence des bases mal appariées provenant des erreurs d'incorporation par les ADN polymérase, et des insertions/délétions (d'une à plusieurs bases) résultant de glissements réplicatifs au niveau des séquences répétitives, ou lors de recombinaisons. Elle est organisée en quatre étapes principales (Figure 2) :

- **Reconnaissance du mésappariement** : L'hétérodimère hMutS α (hMSH2/6) reconnaît les mésappariements et les insertions d'une base, et le dimère hMutS β (hMSH2/3) les insertions/délétions de plusieurs bases (*Larrea et al, 2010*)
- **Recrutement d'autres facteurs MMR** : L'hétérodimère MutL α (Hmlh1/Hpms2) permet le recrutement des autres protéines impliquées dans les phases d'excision et de resynthèse (PCNA, exo nucléase 1, RPA, RFC, pol δ , ligase I) Il interagit aussi avec les complexes MSH et la machinerie de réplication, ce qui est important pour la discrimination du brin néo synthétisé : pour le brin direct au cours de la réplication (*Ziani., 2014*).
- **Identification du brin (néo synthétisé) portant l'erreur et dégradation** au niveau du mésappariement (*Ziani., 2014*).
- **Resynthèse du brin excisé et ligation** : PCNA oriente l'incision induite par hMutL α et stimule son activité endonucléase. Pour le brin indirect, il a été proposé que la discrimination se fasse par la reconnaissance des terminaisons (4 des fragments d'Okazaki existant) (*Ziani., 2014*).

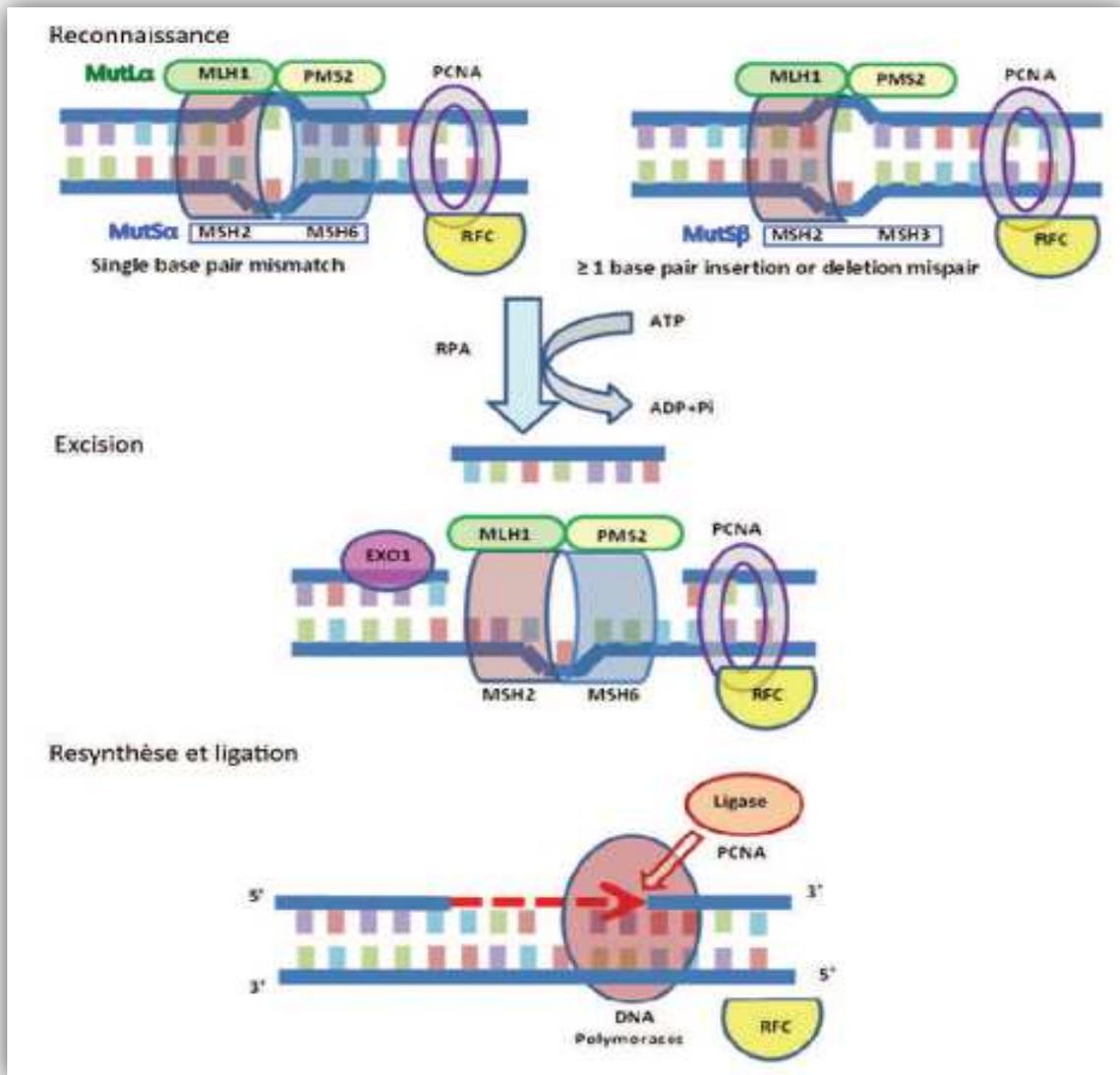


Figure 2: Mécanisme de réparation par mésappariement MMR.

(Ziani, 2014).

b. Réparation par excision de base.

Ce mécanisme de réparation de l'ADN intervient principalement dans la réparation des lésions oxydatives liées au métabolisme cellulaire, mais prend aussi en charge les cassures simple brin induites par les rayonnements ionisants.

Dans le système BER (Réparation par excision de base), des glycosylases spécifiques enlèvent de l'ADN les bases azotées endommagées. Les sites apuriques/aprimidiques sont excisés, et l'intervalle simple brin (de 1 à 8 nucléotides) est comblé par une ADN polymérase. Ce système répare principalement les adduits de petites tailles (alkylation, oxydation et désamination) et

quelques mésappariements pouvant provenir des dommages chimiques ou de l'incorporation erronée d'uracile dans l'ADN (U/G, T/G et peut-être C/C) (*Ozoux., 2013*).

c. Réparation par excision de nucléotide.

-Le système NER (Réparation par excision de nucléotide) cible plusieurs types de lésions, incluant les 6-4PP générées par les ultraviolets. La caractéristique commune à ces lésions est la génération d'une distorsion de la double hélice d'ADN, ce qui permet leur reconnaissance par le système NER (*Gillet et Sharer, 2006*)

-Il existe deux voies de réparation NER, la GG-NER qui agit sur la totalité du génome et la TC-NER qui répare les lésions apparaissant sur les régions activement transcrites, ces deux voies se différencient par l'étape de reconnaissance de la lésion de l'ADN et se rejoignent ensuite en un mécanisme en commun, ouverture de l'ADN autour de la lésion, double incision du brin endommagé, resynthèse et ligation. (*Ziani.,2013*) (Figure 3).

-La reconnaissance des dégâts pendant TC-NER ne nécessite pas de XPC, mais se produit lorsque la machine de transcription est bloquée sur le site de la blessure. Le complexe ARN polymérase bloqué est déplacé afin de permettre aux protéines NER d'accéder à l'ADN endommagé. Ce déplacement est aidé par l'action des protéines CSA et CSB, ainsi que d'autres facteurs spécifiques au TC-NER. Les étapes suivantes de GG- et TC-NER se déroulent de manière essentiellement identique. XPA et le RPA hétérotrimérique se lient alors sur le site de blessure et d'autres aides à la reconnaissance des dégâts.

-Ensuite, les hélicases XPB et XPD, composants du facteur de transcription multi-sous-unités TFIIH dévoilent le duplex ADN au voisinage immédiat de la lésion pour l'enzyme RNA Polymérase-II pour commencer la transcription. Les Endonucléases XPG et ERCC1 (Excision Repair Cross-Complementing Group-1) / XPF clivent ensuite un brin de l'ADN aux positions 3 et 5 du dommage respectivement, générant un oligonucléotide d'approximativement 30 bases contenant la lésion. Cet oligonucléotide est déplacé, ouvrant la voie à la synthèse de la réparation des écarts (réalisée par DNA Pol Delta / Epsilon, ainsi que plusieurs facteurs d'accessoires de réplication). Enfin, le nick dans le brin réparé est scellé par une ADN Ligase, complétant ainsi le processus NER(*Hanawalt.,2000*).

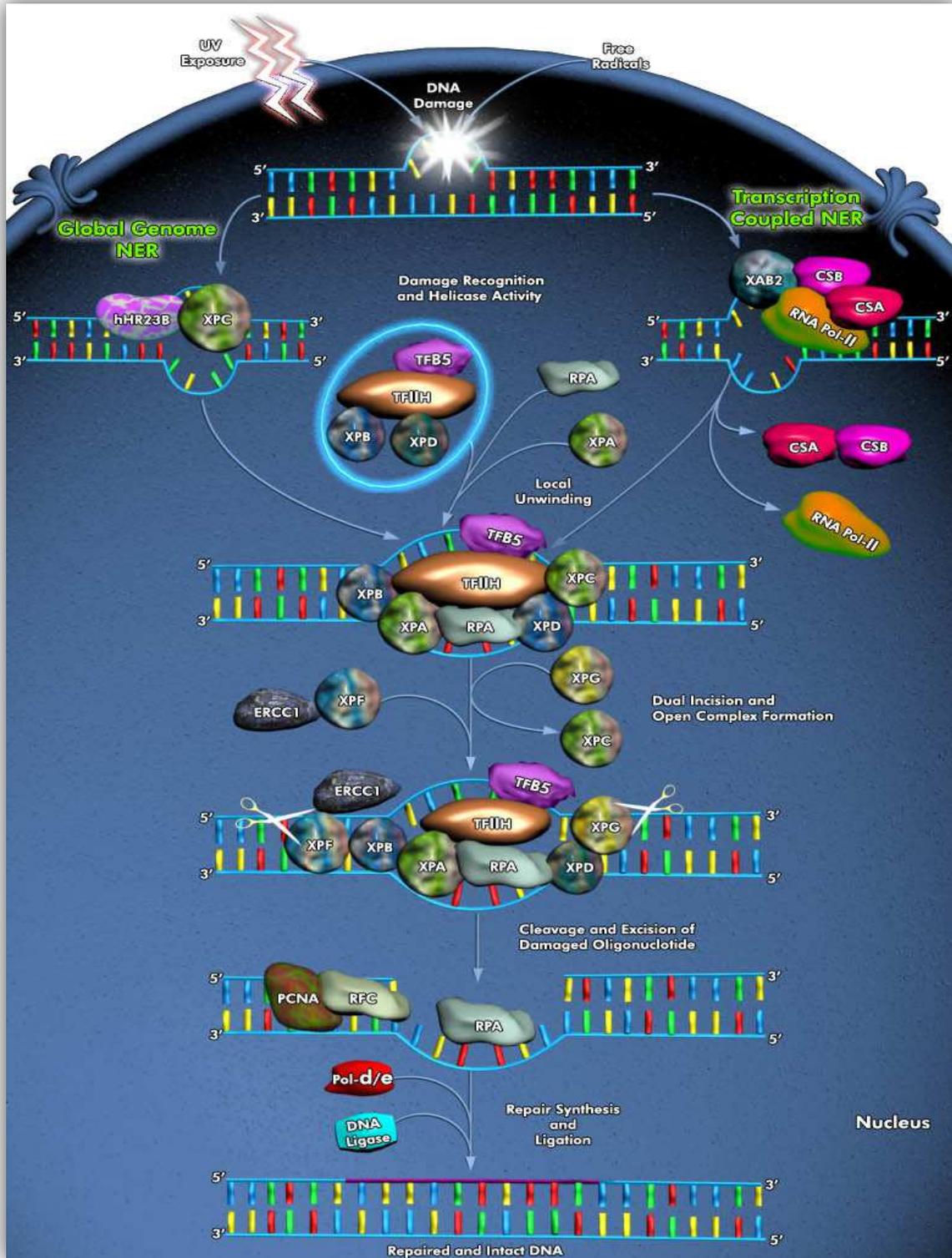


Figure 3: Modèle du mécanisme de réparation par excision de nucléotide NER.

(<https://corporate.qiagen.com/>).

I.7 Physiopathologie.

Les huit groupes de *Xeroderma Pigmentosum* sont dus à des mutations au niveau de système de réparation par excision de nucléotide NER.

I.7.1 Xeroderma Pigmentosum groupe de complémentation XPA.

Le groupe de complémentation correspond à une forme sévère de XP. Sa prévalence globale est inconnue mais il représente 25% de tous les cas de XP et il s'agit de la forme la plus courante de XP au **Japon**. Les patients présentent les lésions typiques du XP (photosensibilité de la peau avec sensation de brûlure, taches de rousseur, et cancers de la peau) associées à des anomalies des troubles du système nerveux central et périphérique, parfois très sévères (détérioration cognitive, dysarthrie, troubles de l'équilibre, aréflexie). Des retards de croissance et du développement sexuel peuvent également être observés chez ces patients (*Kanada et al.,1990*).

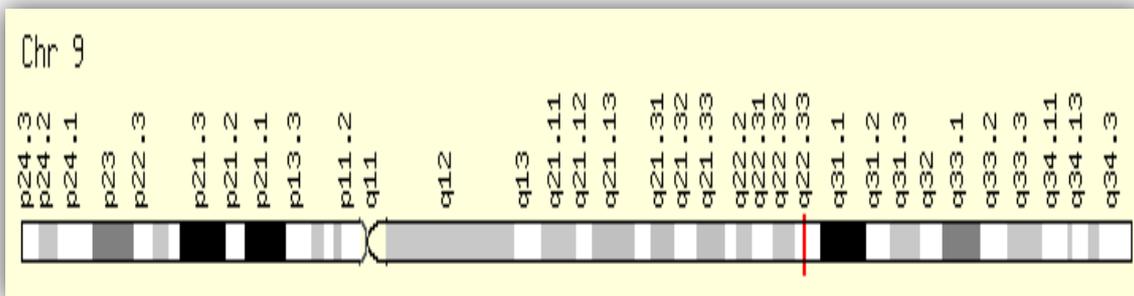


Figure 4: Localisation chromosomique du gène XPA

(*Gene Cards* : <http://www.genecards.org/>).

I.7.2 Xeroderma Pigmentosum groupe de complémentation XPB (ERCC3).

Le groupe de complémentation B est un sous-type extrêmement rare de XP. Cette maladie autosomique récessive est décrite chez moins de 10 familles.

Certains patients XPB présentent les caractéristiques classiques du XP (photosensibilité de la peau avec brûlures et taches de rousseur, tumeurs cutanées et oculaires) avec une gravité variable et de légères anomalies neurologiques. D'autres combinent les caractéristiques classiques du XP avec des manifestations systémiques et neurologiques du syndrome de Cockayne (CS) telles que petite taille, perte d'audition bilatérale et hyperreflexie (complexe XP/CS). XPB est due à des mutations dans le gène XPB (ERCC3), impliqué dans le mécanisme de réparation par excision de nucléotides (NER) (*Fassih,2011*).

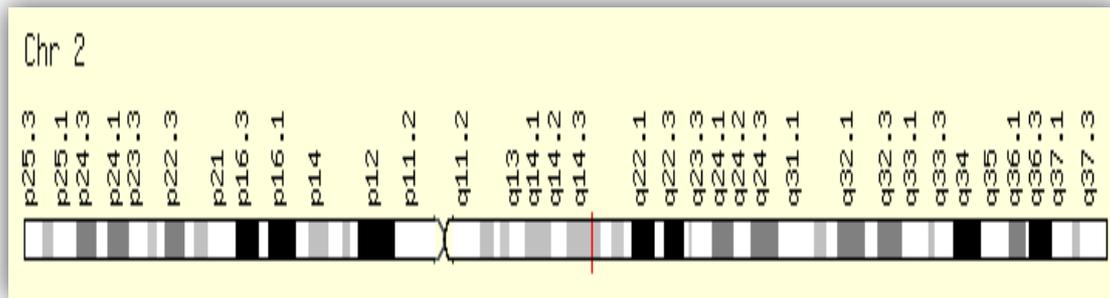


Figure 5: Localisation chromosomique du gène ERCC3

(Gene Cards : : <http://www.genecards.org/>).

I.7.3 Xeroderma Pigmentosum groupe de complémentation XPC.

Le groupe de complémentation C est le sous-type le plus fréquent de XP dans la population caucasienne. Les patients représentent environ 25% de tous les cas de XP et plus de 50 cas de XPC ont été rapportés dans la littérature.

Les patients présentent des lésions de la peau typique du XP (taches de rousseur progressives, sècheresse de la peau, cancers de la peau, tumeurs malignes oculaires), mais en général ils ne présentent aucune réactions aiguës au soleil (types coup de soleil). Par ailleurs, les patients XPC ne souffrent pas de troubles neurologiques. La maladie est due à des mutations dans le gène XPC (locus 3p25), impliqué dans la reconnaissance des dommages à l'ADN lors de la NER. La transmission est autosomique récessive(*Fassihi.,2011*).

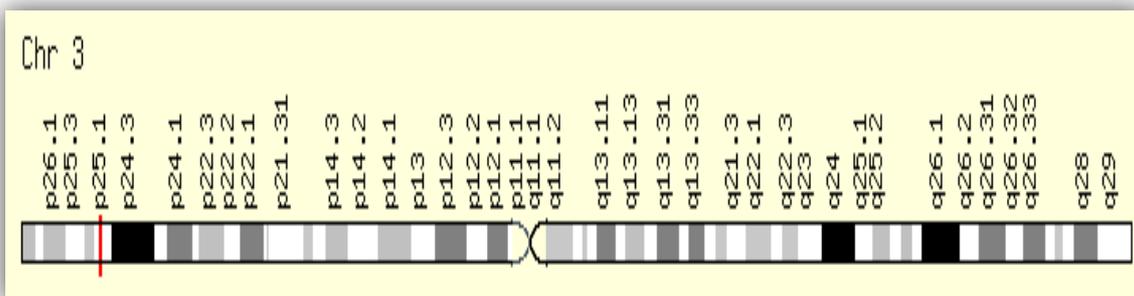


Figure 6: Localisation chromosomique du gène XPC

(Gene Cards : : <http://www.genecards.org/>).

I.7.4 Xeroderma Pigmentosum groupe de complémentation XPD(ERCC2).

Le groupe de complémentation D est un sous-type de XP qui représente environ 15% des cas de XP. Plus de 30 cas ont été rapportés dans la littérature pour cette maladie dont la transmission est autosomique récessive. Il s'agit d'une forme cliniquement hétérogène qui présente les

manifestations typiques du XP (photosensibilité de la peau avec sensation de brûlure, taches de rousseur, sécheresse de la peau, cancers de la peau) associées ou non à des anomalies neurologiques de gravité variable. Certains patients XPD combinent les caractéristiques classiques du XP avec des manifestations systémiques et neurologiques du syndrome de Cockayne (complexe XP/CS), d'autres présentent encore des manifestations de la trichothiodystrophie (syndrome XP/TTD). L'affection XPD est provoquée par des mutations dans le gène XPD (ERCC2) (locus 19q13.2-q13.3), impliqué dans la phase d'ouverture de la molécule d'ADN lors le NER(*Fassih.,2011*).

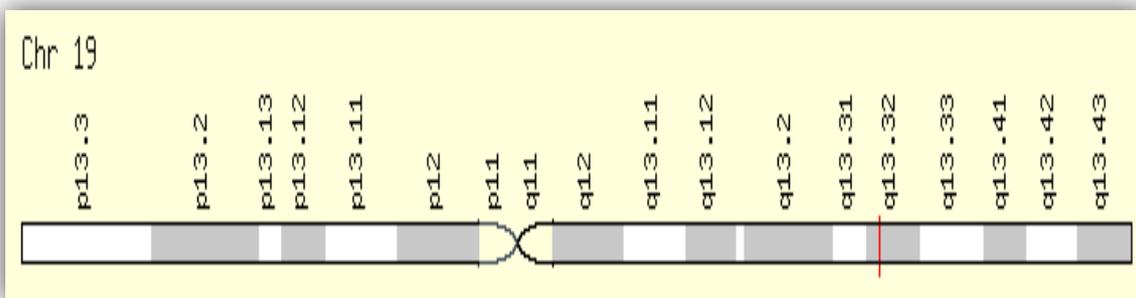


Figure 7:Localisation chromosomique du gène ERCC2

(*Gene Cards* : : <http://www.genecards.org/>).

I.7.5 Xeroderma Pigmentosum groupe de complémentation XPE(DDB2).

Le groupe de complémentation E est un sous-type extrêmement rare de XP, moins de 10 cas ont été rapportés dans la littérature. La transmission est autosomique récessive. Les patients présentent des symptômes légers du XP et aucune anomalie neurologique. Comme leurs symptômes cutanés sont légers, les patients XPE n'ont pas souvent une photo protection rigoureuse et développent des cancers cutanés graves. La maladie est due à des mutations dans le gène XPF (DDB2) (locus 11p12-p11), impliqué dans l'étape de reconnaissance des dommages lors de la réparation de l'ADN par le NER(*Fassih.,2011*).

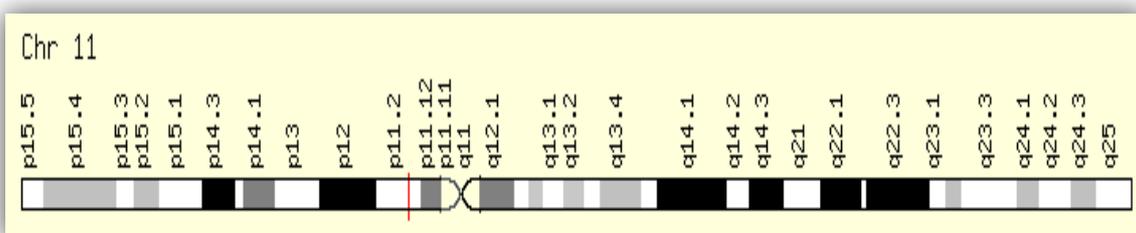


Figure 8: Localisation chromosomique du gène DDB2 (*GeneCards* : :

<http://www.genecards.org/>).

I.7.6 Xeroderma Pigmentosum groupe de complémentation XPF(ERCC4).

Le groupe de complémentation F est un sous-type de XP de sévérité généralement moyenne et dont la transmission est autosomique récessive. Il est représenté presque exclusivement dans la population japonaise. Les patients XPF représentent 6% de tous les cas de XP et moins de 15 cas ont été rapportés dans la littérature. La plupart des patients montrent des signes cutanés très légers et aucune maladie oculaire ou neurologique. La maladie est due à des mutations dans le gène XPF (ERCC4) (locus 16p13.3-p13.13), impliqué dans l'étape de coupure du brin d'ADN endommagé lors du mécanisme de réparation de l'ADN par NER (*Fassih,2011*).

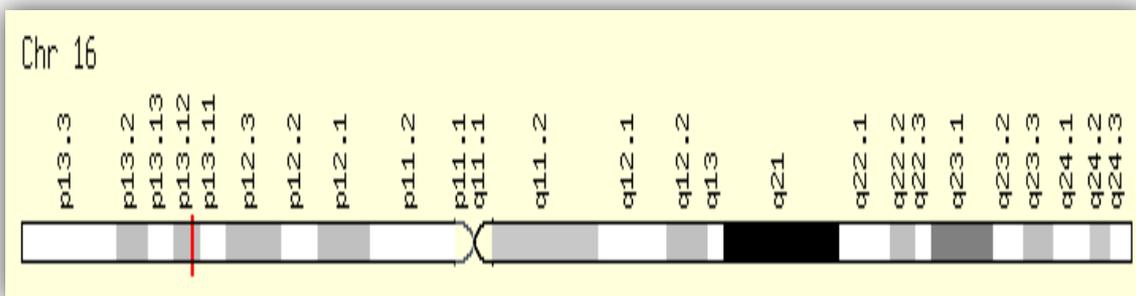


Figure 9:Localisation chromosomique du gène ERCC4

(GeneCards : : <http://www.genecards.org/>).

I.7.7 Xeroderma Pigmentosum groupe de complémentation XPG(ERCC5).

Le groupe de complémentation G est un sous-type extrêmement rare de XP. Les patients souffrant de XPG représentent 6% de tous les cas de XP et environ 10 cas ont été rapportés dans la littérature, la plupart en Europe. Les manifestations cliniques peuvent varier. Certains patients présentent un phénotype léger de XP (sensibilité aux UV, lésions cutanées, hyper ou hypo-pigmentation et une incidence accrue de cancer de la peau), d'autres combinent les symptômes de XP avec des manifestations systémiques et neurologiques du syndrome de Cockayne (complexe XP/CS). La maladie est autosomique récessive, elle est due à des mutations dans le gène XPG (ERCC5) (locus 13q33), impliqué dans l'étape de coupure du brin d'ADN endommagé lors de la réparation de l'ADN par NER (*Fassih,2011*).

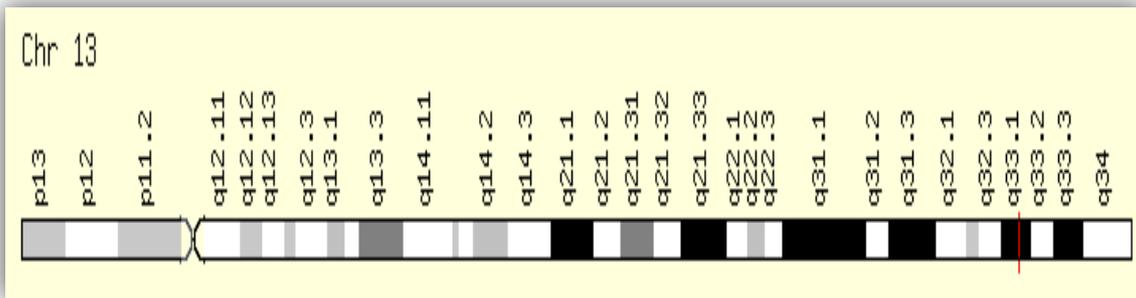


Figure 10:Localisation chromosomique du gène ERCC5

(GeneCards : : <http://www.genecards.org/>).

I.7.8 Xeroderma Pigmentosum groupe de complémentation XPV(POLH).

Xeroderma pigmentosum Variant (XPV) est un sous-type de XP à symptomatologie légère. Pour cette raison, il est souvent diagnostiqué plus tardivement, à la fin de l'adolescence ou vers l'Age de 20 ans. Il est observé chez environ 20% des patients XP et environ 50 cas ont été rapportés dans la littérature, la transmission est autosomique récessive(*Fassih.,2011*).

Les patients XPV présentent des lésions de la peau, généralement au niveau des parties du corps exposé au soleil, qui évoluent en cancer de la peau aux environ de 20-30 ans avec une fréquence 1.000 fois plus élevée que celle de la population générale. Aucune manifestation neurologique n'est observée. La maladie est due à des mutations dans le gène POLH (XPV) (locus 6p21.1-p12) codant pour l'ADN-polymérase η ($\text{pol-}\eta$) qui est chargé de synthétiser la zone excisée et restaurée la continuité de l'ADN lors de la réparation par NER. (*Kraemer et DiGiovanna,2016*). Les cellules XP-V sont en effet capables de mener à bien le processus de réparation par excision de nucléotides mais présentent un défaut dans la réplication des ADN endommagés par les ultraviolets.

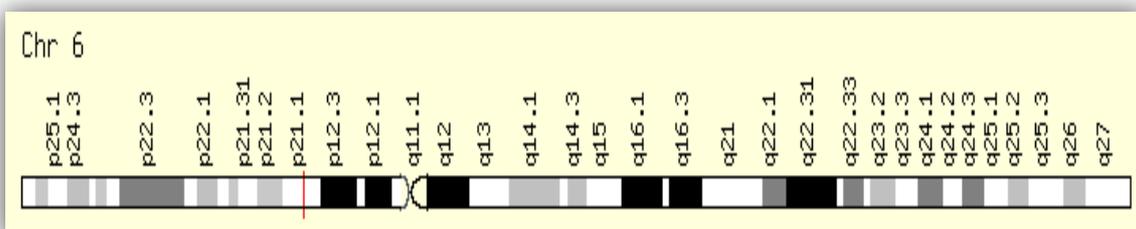


Figure 11:Localisation chromosomique du gène de POLH

(GeneCards : : <http://www.genecards.org/>).

Tableau I: Les groupes de complémentation de XP (*Fassih.,2012*).

Groupe	Gène muté	Chromosome/ Locus	Fréquence	Sévérité
XPA	XPA	9/9q22.32	25%	Très sévère avec anomalies neurologique importantes.
XPB	ERCC3	2/2q21.2	Très rare (-de 10 cas au monde)	Recouvrement avec le syndrome de Cockayne.
XPC	XPC	3/3p25	25%	Absence de problèmes neurologiques
XPD	ERCC2	19/19q13.3	15%	Toujours accompagné d'anomalies neurologiques plus ou moins importantes.
XPE	DDB2	11/11p12	Rare	Symptômes relativement légers sans troubles neurologique.
XPF	ERCC4	16/16p13.3	6%	Réparation de l'ADN est totale mais extrêmement lente.
XPG	ERCC5	13/13q33.1	6% Très rare	Recouvrement avec le syndrome de Cockayne.
XPV	POLH	6/6p12	21%	Symptômes en fonction du type de peau. Pas d'affection neurologique.

*Dans cette étude on s'intéresse sur le gène XPC qui est le plus réponsus en **Afrique du Nord** et plus précisément sur la protéine XPC qui a un rolé très important dans le système de réparation par excision de nucléotifde NER.

I.8 Protéomique de XPC.

XPC est une protéine de 940 acides aminés, et une masse de 105,953 Da de la famille des Protéine de réparation d'ADN Rad4 avec des domaines qui peuvent se lier à un ADN

endommagé et à des facteurs de réparations. Elle peut être En complexe avec RAD23B et CETN2 dans la voie GG-NER et avec TFIIH dans la voie TR-NER (*Araki et al.,2001*)

I.8.1 Reconnaissance des dégâts d'ADN dans GG-NER.

Les dommages à l'ADN sont reconnus par deux complexes de protéines. Le premier complexe se compose de XPC/RAD23A ou RAD23B ou CETN2. Ce complexe teste l'hélice d'ADN et reconnaît les dommages qui perturbent l'appariement normal de la base, ce qui entraîne la liaison du complexe XPC/RAD23 ou CETN2 au brin d'ADN non endommagé (*Oh et al.,2011*) (Figure12).

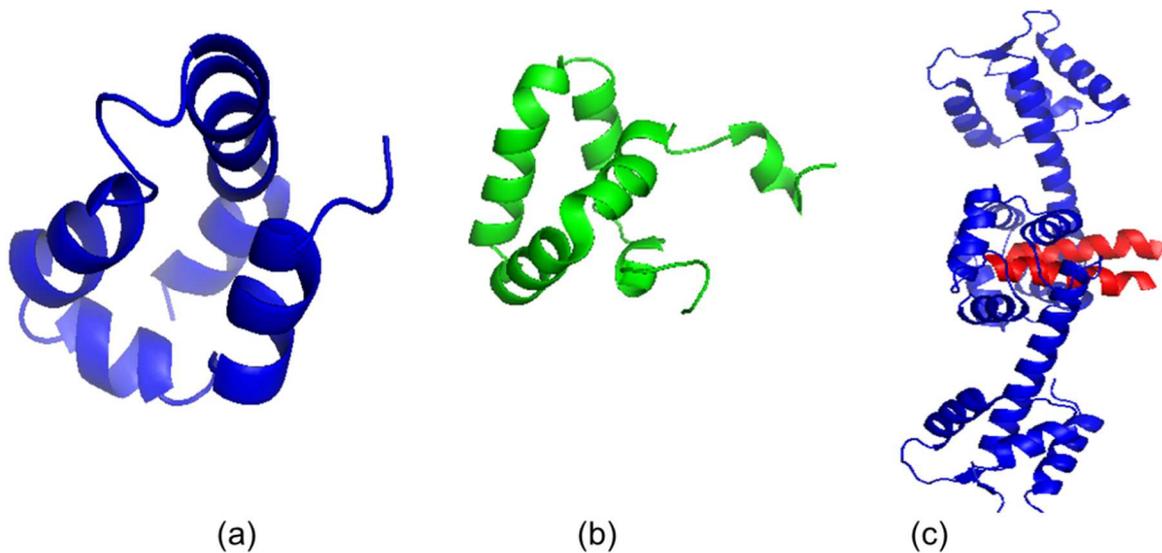


Figure 12: Protéines qui interagissent avec XPC dans la reconnaissance des dégâts de l'ADN

(<https://www.rcsb.org/pdb/home/home.do>).

(a): Structure de la protéine RAD23A en bleu avec le code pdb 1tp4

(b): Structure de la protéine RAD23B en vert avec le code pdb 1pve

(c): Structure du complexe Centrine-XPC (centrine en bleu et le fragment de XPC en rouge) avec le code pdb 2ggm.

Le deuxième complexe est une UV-DDB d'ubiquitine ligase qui se compose de DDB2, DDB1, CUL4A ou CUL4B et RBX1. Le complexe UV-DDB est nécessaire pour la reconnaissance des dommages causés par l'ADN induit par UV et peut contribuer à la rétention du complexe XPC : RAD23 : CETN2 sur le site de dommages à l'ADN. Le complexe UV-DDB lie le brin d'ADN endommagé (*Oh et al.,2011*).

I.8.2 Complexe de formation d'incision à TR-NER.

Après que le complexe XPC et le complexe UV-DDB lient un ADN endommagé, un facteur de transcription TFIID est recruté sur le site de réparation d'excision nucléotidique (NER) (*figure 13*) (*Riedl et al., 2003*). DNA hélicases ERCC2 (XPD) et ERCC3 (XPB) sont des sous-unités du complexe TFIID. ERCC2 déroule l'ADN autour des dégâts en concert avec l'activité ATPase d'ERCC3, créant une bulle ouverte (*Coin et al., 2007*). Simultanément, la présence du dommage est vérifiée par XPA (*Camenisch et al., 2006*). Deux endonucléases d'ADN, ERCC5 (XPG) et le complexe de ERCC1 et ERCC4 (XPF) sont recrutés dans la structure de bulle ouverte pour former le complexe d'incision qui accise l'oligonucléotide endommagé du brin d'ADN affecté (*Orelli et al., 2010*).

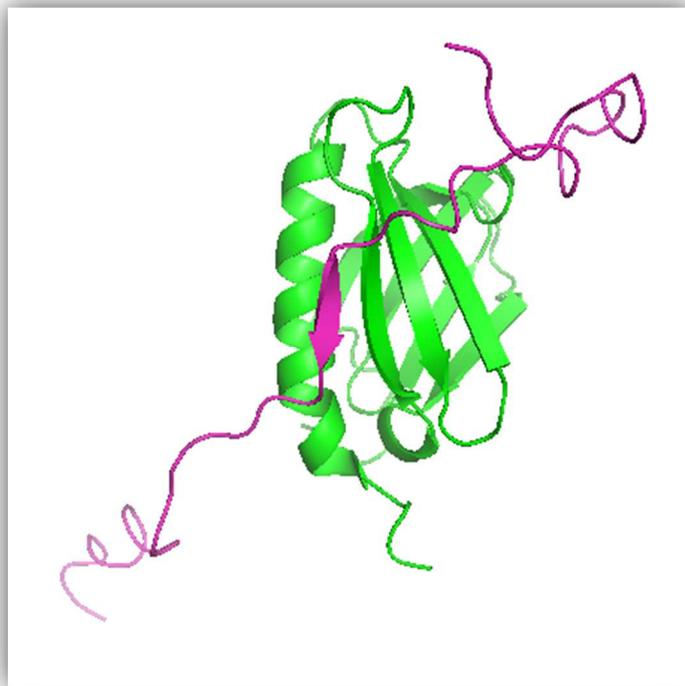


Figure 13: Structure du complexe entre XPC (en rose) et TFIID (en vert) (code pdb : 2rvb) (<https://www.rcsb.org/pdb/home/home.do>).

Matériels

et

Méthodes

II.1 Matériels

Sur les huit gènes connus du *Xeroderma pigmentosum*, nous avons choisi de travailler sur le gène XPC qui est le plus réparti dans les pays du Maghreb. La liste des SNPs que nous exploitons sont tirés de résultats de séquençage d'ADN de quelques patients tunisiens présentés sous formes des SNPs de des CNVs en position génomiques.

II.2 Méthodes

La méthodologie de notre travail est résumée comme suit selon l'organigramme présenté dans la figure 14 :

En premier lieu, nous avons utilisé des plates formes et des bases de données bioinformatiques pour avoir les listes des SNPs. UCSC a permis d'identifier tous les SNPs du gène XPC connues en utilisant *Génome Browser* et qui par la suite de déterminer ceux qui sont annotés dans *Table Browser*. Le filtrage du fichier du séquençage d'ADN (chromosome 3 : gène XPC) a donné quelques rs en tirant toutes les informations nécessaires de ces derniers dans la base de données SNP de NCBI. Notre protéine d'intérêt est la protéine XPC dont le code d'accèsion de sa séquence est Q01831 dans la base de données protéiques Uniprot.

Dans une deuxième partie, nous nous sommes intéressés à la prédiction de la structure tridimensionnelle du XPC par différentes approches de modélisation moléculaire. Un blastp de la séquence de la protéine de références XPC (code d'accèsion Q011831) contre la banque de données PDB a été effectué pour la recherche de similarité de séquences cible (XPC)/ patrons (séquences dont la structure est connue).

La prédiction de la structure 3D de XPC s'est faite par trois programmes qui se basent sur trois approches différentes :

-Swiss-Model : Méthode de prédiction 3D par modélisation comparative.

-Phyre2 : Méthode de prédiction 3D par reconnaissance de repliements.

-I-TASSER : Méthode de prédiction 3D par *ab initio* combinée à la méthode de reconnaissance de repliements.

Par ailleurs, *d'après la base de données STRING, la protéine XPC interagit avec plusieurs protéines. Nous nous sommes donc intéressés à l'étude de

*l'amarrage protéines-protéine par le programme ClusPro pour : recherches d'éventuels complexes entre la structure de XPC prédites et la structure des protéines connues en interaction avec cette dernière.

*Les structures issues de PDB, et des modèles d’XPC prédits par SwissModel, Phyre2 et I-TASSER ont été visualisées et leur similarité structurale après superposition a été mesurée après alignement par Pymol.

Dans une troisième partie * Interpro a été utilisé pour déterminer la localisation des rs associés à la maladie XP issue du *Table Browser*. Ces rs ont été étudiés sur le modèle prédit de XPC afin d’analyser l’impact de ces derniers sur les propriétés de la protéine (sa fonction, formation de complexes).

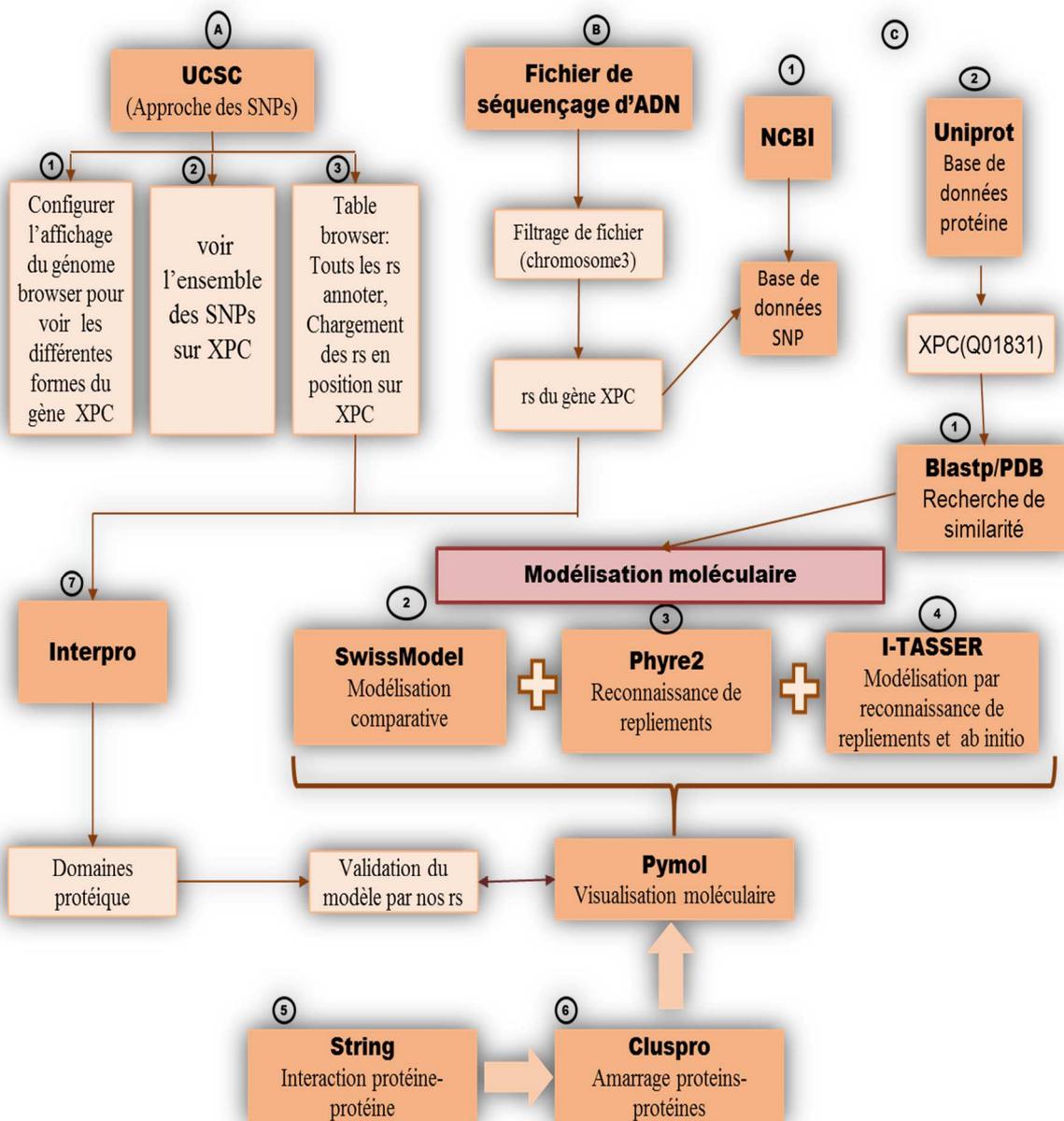


Figure 14 : Méthodologie de notre travail de fin de cycle.

Organisation et traitement des échantillons

Le fichier de séquençage sur lesquels on vas travailler contient 299141 SNP et CNV. Après avoir exclue les CNVs, nous avons filtré par EXCEL les SNPs du chromosome 3 positionnés dans l'intervalle [14145147-14178672] là où se positionne le gène XPC confirmé par NCBI.

Notre travail se subdivise en deux parties une pour étudier les SNPs du gène XPC et l'autre pour la prédiction de la structure 3D de la protéine XPC.

II.2.1 Partie 1 : Listes des SNPs

Connaissant la base de données UCSC qui comporte plusieurs gènes et une masse importante d'information, nous avons opté pour travailler avec UCSC pour ressortir les SNPs connus du gène XPC. L'objectif est d'avoir la lista des SNPs du gène XPC dans *Génome Browser* ensuite ceux qui sont annotés dans *Table Browser*.

a) UCSC

Le site universitaire de UCSC (<https://genome.ucsc.edu/>) est organisé par l'Université de Californie Santa Cruz (UCSC). Il s'agit d'une offre d'un site web interactif avec accès aux données de la séquence de génomes d'une variété d'espèces de vertébrés et invertébrés et principaux organismes modèles, intégrés à une vaste collection d'annotations alignés.



Figure 15 : Page d'accueil UCSC (<https://genome.ucsc.edu/>).

-Génome Browser

Génome Browser est utilisé pour déterminer la liste des SNPs. Notre recherche a été limitée en utilisant les critères suivants :

All SNPs (147) ; Common SNPs (147) ; Flagged SNPs (147) : full (Figure 16).

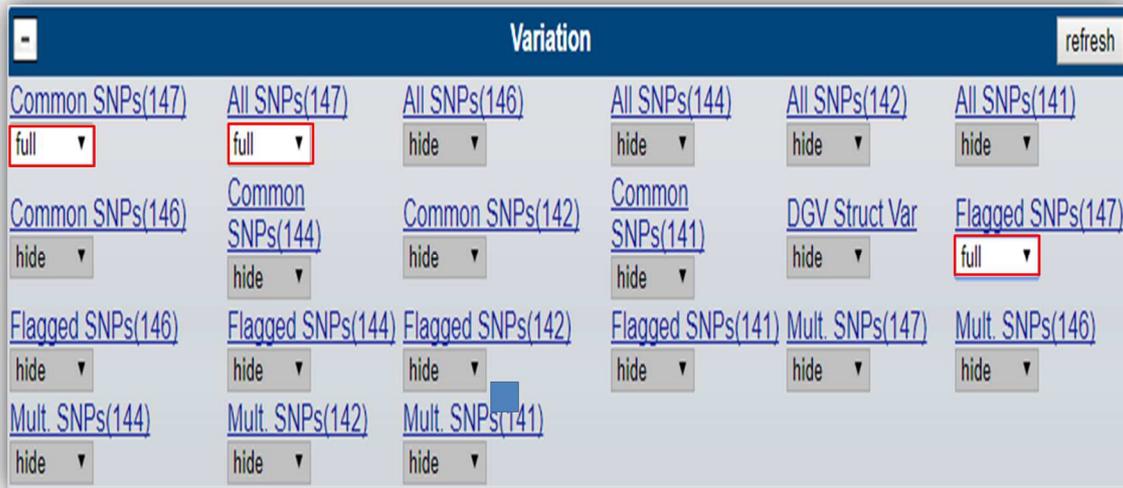


Figure 16 : Interface de *Génome Browser* permettant la sélection de paramètres de filtre des SNPs (https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr3%3A14145147-14178783&hgsid=607250681_7fs3hq8iwWu2HXHIfAptFnQeazs)

-Table Browser.

Table Browser (<https://genome.ucsc.edu/cgi-bin/hgTables>) fournit une interface graphique puissante et flexible pour interroger et manipuler les tableaux d'annotations de *Génome Browser*. Etant donné que *Table Browser* utilise la même base de données que *Génome Browser*, les deux vues sont toujours cohérentes.

Les rs recherchés ont été chargés dans *Table Browser* (Figure 17).

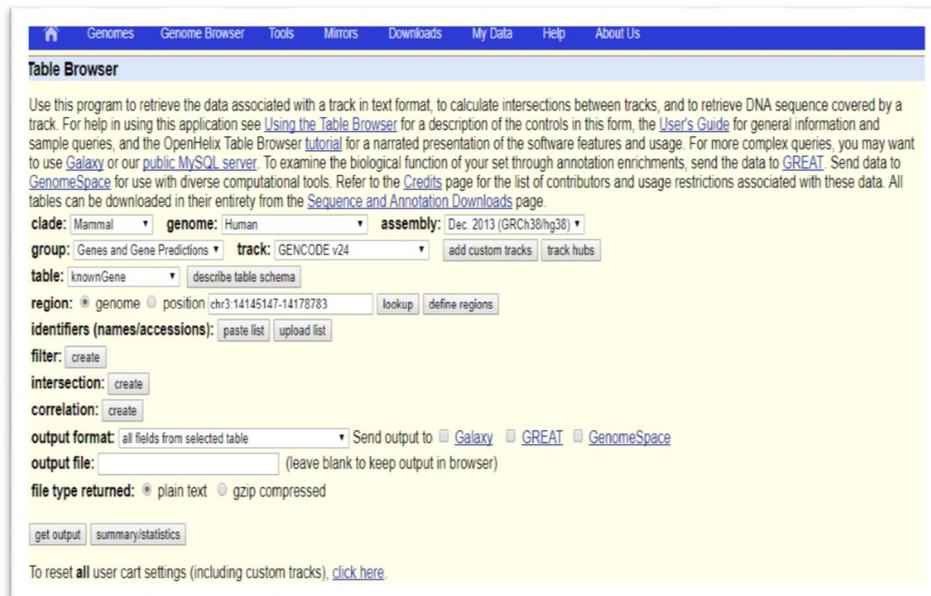


Figure 17 : Page d'accueil de *Table Browser* et de paramètres de filtrage des SNPs annotés

(<https://genome.ucsc.edu/cgi-bin/hgTables>)

b) Le serveur NCBI.

C'est une base de données (<https://www.ncbi.nlm.nih.gov/>), qui permet de mener des recherches en bioinformatique et d'analyser des données *via* des logiciels. Elle permet la diffusion des travaux de recherche dans le domaine biomédical. Ce serveur constitue une toile communicante vers d'autres serveurs spécialisés dans la génomique fonctionnelle ou comparative.

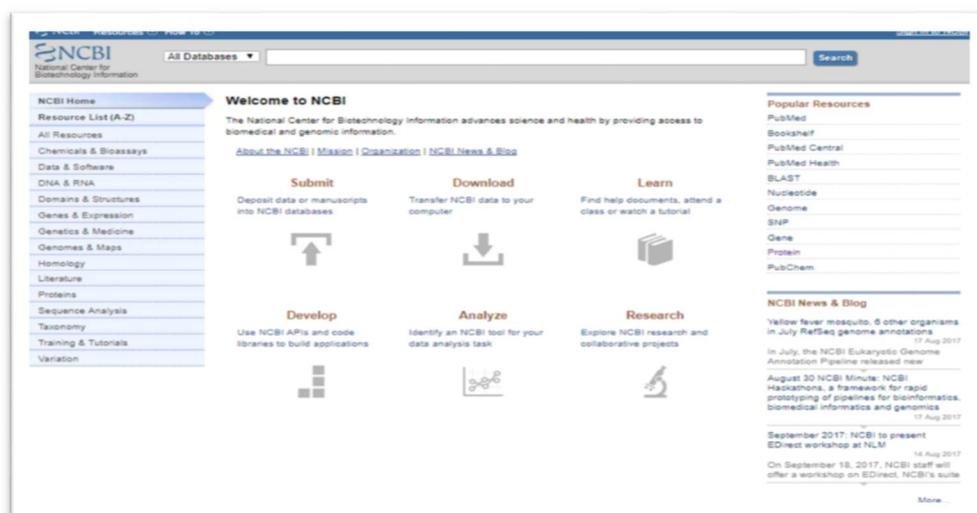


Figure 18 : Page d'accueil de NCBI (<https://www.ncbi.nlm.nih.gov/>).

Ce serveur a été utilisé dans le but de trouver le détail de nos SNPs identifiés, pour vérifier les résultats issus de traitement des SNPs du fichier des résultats de séquençage, les SNPs correspondant au gène XPC et les détails des rs annotés dans *Table Browser*.

II.2.2 Partie 2 : Modélisation moléculaire de XPC

Dans une deuxième partie de notre travail, nous nous sommes intéressés à prédire et à étudier la structure 3D de XPC. En effet, une voie permettant la compréhension de la fonction biologique et des propriétés d'une protéine passe par l'étude de sa structure tridimensionnelle (3D) (*Khemili., 2013*).

1. Prédiction de structure 3D de XPC

a) La séquence protéique XPC

La séquence protéique XPC a été tirée de la base de données Uniprot (<http://www.uniprot.org>) avec le code d'accèsion « Q01831 », elle contient 940 acide aminés avec une masse de 105,953 Da.

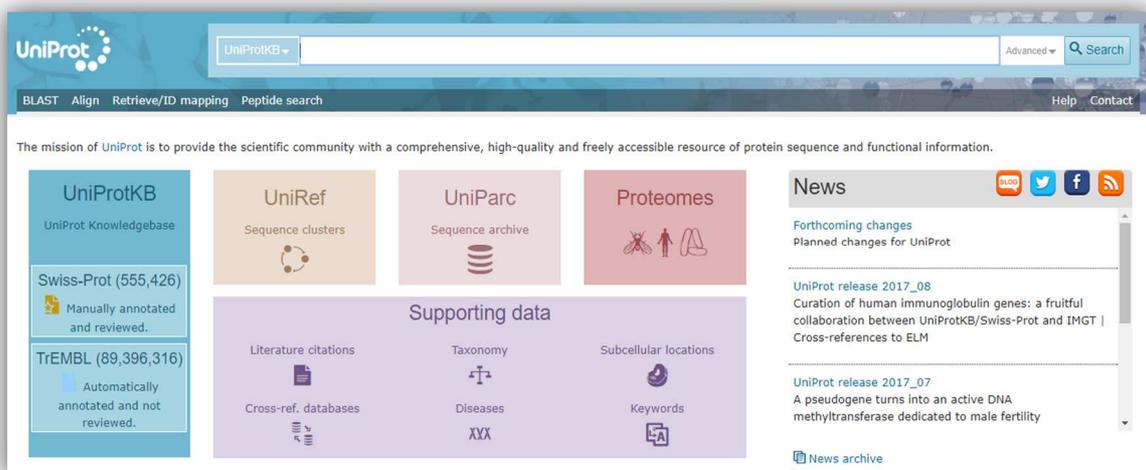


Figure 19 : Page d'accueil de Uniprot (<http://www.uniprot.org/>).

b) Recherche de similarité cible/patrons

La recherche de similarité entre la cible (séquence de XPC Q0131 dont la structure est inconnue) avec les éventuels patrons trouvés dans la *Protein Data Bank* PDB (séquences dont la structures 3D a été déterminée expérimentalement) a été effectuée grâce au programme Blastp. Cette étape est cruciale pour le choix de la méthode de modélisation moléculaire qui pourrait être utilisée pour la prédiction de la structure 3D de notre protéine d'intérêt XPC (Q01831).

1-Blastp

Un outil de recherche d'alignement locale de base (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Il trouve des régions de similarité locale entre les séquences, compare les séquences de nucléotides ou de protéines aux bases de données séquentielles et calcule la signification statistique des correspondances.

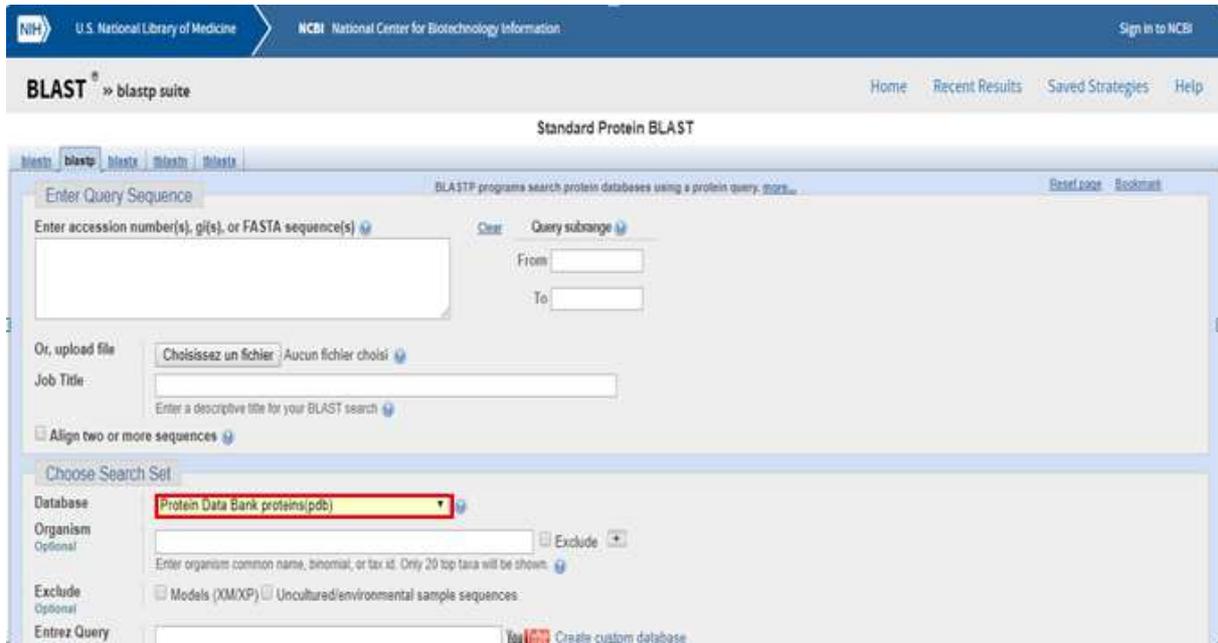


Figure 20 : Page d'accueil de Blastp.

(<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastendPAGETYPE=BlastSearchendLINK LOC=blasthome>).

2-Collection de structures expérimentales des patrons de la *Protein Data Bank*

Cette ressource est alimentée par l'archive des données de Protein Data Bank (<https://www.rcsb.org>) sur les formes 3D des protéines, des acides nucléiques et des assemblages complexes qui aident à comprendre tous les aspects de la biomédecine et de l'agriculture, de la synthèse des protéines à la santé et à la maladie

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 132905 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PDB-101 Worldwide Protein Data Bank EMDatabank Worldwide Protein Data Bank Foundation

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Video: How Enzymes Work

VIDEO: HOW ENZYMES WORK

August Molecule of the Month

Glutathione Transferases

Figure 21 : Page d'accueil de PDB. (www.rcsb.org)

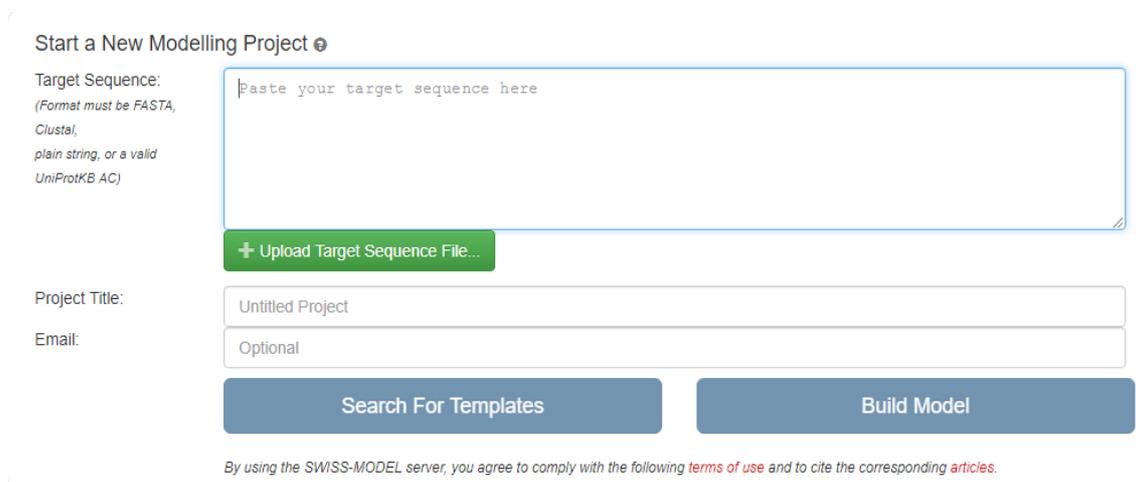
c) Prédiction de la structure 3D de XPC

La prédiction de la structure de XPC a été effectuée en utilisant trois programmes SwissModel, Phyre2 et I-TASSER qui se basent sur trois approches de modélisation moléculaires différentes

1-Swiss-Model

SWISS-MODEL (<https://swissmodel.expasy.org/>) est accessible via le serveur Web ExPASy, il se base sur ma méthode de prédiction par modélisation comparative. La prédiction de la structure 3D par la modélisation comparative est fondée sur le fait que deux protéines qui partagent une identité de séquence élevée (plus de 30%) adoptent un même repliement (*Rst, 1999 ; Baker and Sali, 2001*). Son principe consiste à aligner la séquence d'une protéine donnée dont la structure est inconnue (cible, dans notre cas la XPC) avec la séquence d'une

ou plusieurs protéines dont la structure a été déterminée expérimentalement par RMN ou Rayons-X (patrons ou références collectées de la PDB) (*Sali and Blundell,1993*). Le serveur Web Swiss-Model prend plus ou moins 30 minutes pour créer des modèles de structure de la cible (en fonction de sa composition en acides aminés)



The screenshot shows the 'Start a New Modelling Project' interface. It includes a text input field for the 'Target Sequence' with a placeholder 'Paste your target sequence here' and a note that the format must be FASTA, Clustal, plain string, or a valid UniProtKB AC. Below this is a green button labeled '+ Upload Target Sequence File...'. There are also input fields for 'Project Title' (containing 'Untitled Project') and 'Email' (containing 'Optional'). At the bottom, there are two blue buttons: 'Search For Templates' and 'Build Model'. A small red text at the bottom states: 'By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the corresponding articles.'

Figure 22 : Interface d'interaction de la séquence protéique de XPC dans le serveur Swiss-Model (<http://swissmodel.expasy.org/interactive>)

2-Phyre2

Le programme Phyre2 (**Protein Homology/analogY Recognition Engine V 2.0**

(<http://www.sbg.bio.ic.ac.uk/phyre2>)) se base sur la méthode de prédication de structure 3D par la modélisation de reconnaissance de repliements est utilisée pour modéliser par analogie, les protéines qui ont les mêmes repliements que des protéines de structures connues, mais qui n'ont pas de protéines homologues de structure connue. Elle consiste à enfile la séquence de la cible sur une bibliothèque de repliement afin de déterminer les structures qui correspondent le mieux à la séquence sur la base d'un critère énergétique ou de score. Phyre2 est une suite d'outils disponibles sur le Web pour prédire et analyser la structure, la fonction et les mutations des protéines. L'accent de Phyre2 est de fournir aux biologistes une interface simple et intuitive aux outils de bioinformatique de protéines à la fine pointe de La technologie (*Kelly et al,2015*). Le serveur Web Phyre2 prend plus ou moins 1 heure pour créer des modèles de structures de la cible (en fonction de sa composition en acides aminés)

Figure 23 : Page d'accueil du Phyre2 . ([Http://www.sbg.bio.ic.ac.uk/phyre2](http://www.sbg.bio.ic.ac.uk/phyre2))

3-I-Tasser

Iterative Threading ASSEMBLY Refinement (<http://zhanglab.cmb.med.umich.edu/I-TASSER/>), est une approche hiérarchique de la structure des protéines et de la prédiction de la fonction. Le programme I-TASSER se base sur la modélisation *ab initio* ainsi que la reconnaissance de repliements. Si peu de similarité de séquence cible/patrons est détectée les méthodes *ab initio* sont utilisées pour prédire la structure de la cible à partir de sa séquence en acides aminés en se basant sur leurs interaction physicochimiques (électrostatiques, van der Waals, liaisons hydrogènes, etc....) entre les atomes des résidus (*Cornell et al., 1995*). Ces méthodes se basent sur l'hypothèse que la structure native des protéines correspond à l'énergie libre minimale en fonction de la longueur de la séquence protéique, ce serveur peut prendre 1 journée à 1 semaine pour créer des modèles de structures de la cible.

I-TASSER
Protein Structure & Function Predictions

(The server completed predictions for **348468** proteins submitted by **85060** users from **132** countries)
(The template library was updated on 2017/07/16)

I-TASSER (Iterative Threading ASSEMBly Refinement) is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template fragment assembly simulations. Function insights of the target are then derived by threading the 3D models through protein function database BioLiP. I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, and CASP12 experiments. It was also ranked as the best for function prediction in CASP9. The server is in active development with the goal to provide the most accurate structural and function predictions using state-of-the-art algorithms. Please report problems and questions at [I-TASSER message board](#) and our members will study and answer the questions asap. (> [More about the server...](#))

[Queue] [Forum] [Download] [Search] [Registration] [Statistics] [Remove] [Potential] [Decoys] [News] [Annotation] [About] [FAQ]

I-TASSER On-line Server ([View an example of I-TASSER output](#)):

Copy and paste your sequence below ([10, 1500] residues in FASTA format). [Click here for a sample input.](#)

Or upload the sequence from your local computer:
 Aucun fichier choisi

Email: (mandatory, where results will be sent to)

Password: (mandatory, please click [here](#) if you do not have a password)

ID: (optional, your given name of the protein)

- ▶ [Option I: Assign additional restraints & templates to guide I-TASSER modeling.](#)
- ▶ [Option II: Exclude some templates from I-TASSER template library.](#)
- ▶ [Option III: Specify secondary structure for specific residues.](#)

Keep my results public (uncheck this box if you want to keep your job private. A key will be assigned for you to access the results)

(Please submit a new job only after your old job is completed)

Figure 24 : Page d'accueil de I-TASSER (<http://zhanglab.cmb.med.umich.edu/I-TASSER/>),

2. Amarrage protéine-protéine

Afin d'étudier les éventuelles interactions de la protéine XPC (code d'accèsion Uniprot **Q01831**) avec d'autres protéines nous avons interrogé à la base de données String

a-String

La base de données STRING (<https://string-db.org/>) vise à favoriser et fournir une évaluation critique et l'intégration des interactions protéine-protéine y compris direct (physique) et indirect (fonctionnel).

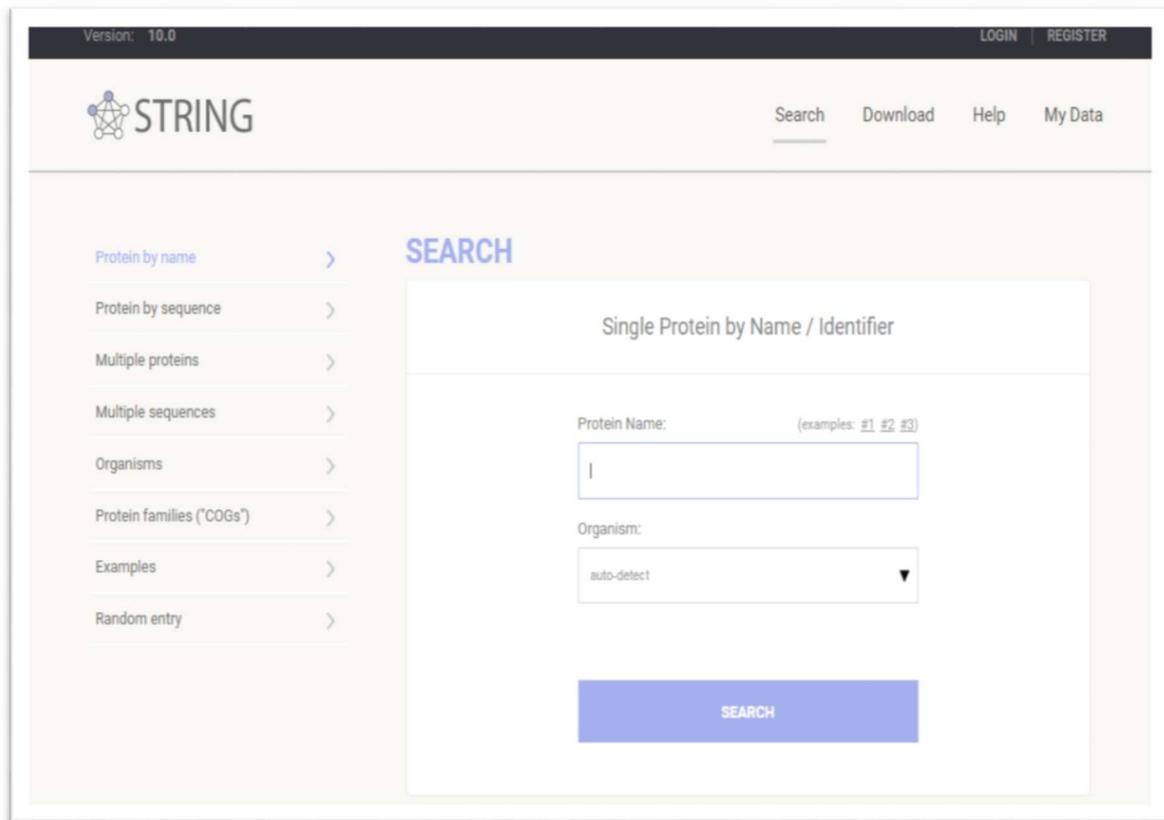


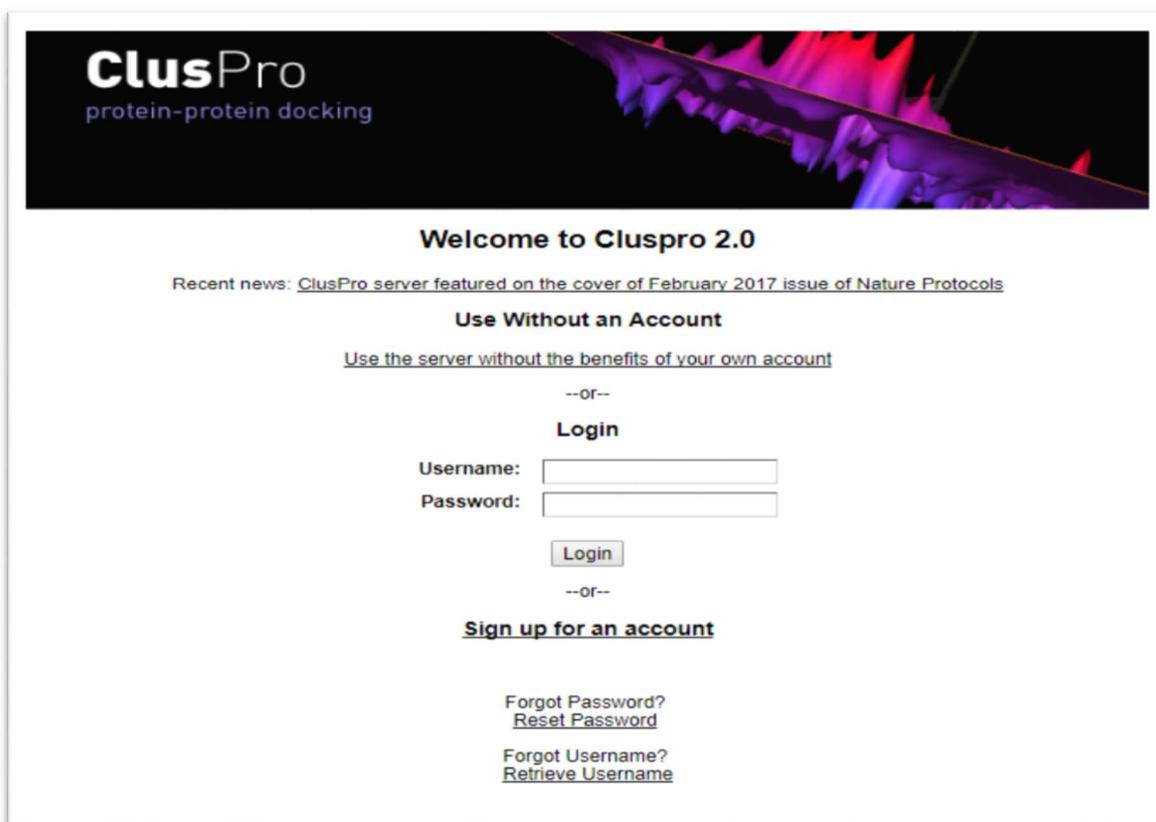
Figure 25 : Page d'accueil de STRING. (<https://string-db.org/>)

Cette recherche nous permet de déterminer des interactions de XPC avec d'autres protéines qui ont un rôle crucial dans le système de réparation de l'ADN.

Après avoir recensé les protéines qui pourraient interagir avec notre XPC, le programme Cluspro a été utilisé pour prédire les éventuels complexes entre XPC et ces dernières

b-ClusPro

Cluspro (<http://cluspro.bu.edu/login.php>) représente le premier programme entièrement automatisé basé sur le Web pour l'amarrage informatique des structures protéiques. Les utilisateurs peuvent télécharger les fichiers de coordonnées de deux structures protéiques *via* l'interface Web de ClusPro ou entrer les codes PDB des structures respectives que ClusPro va ensuite télécharger à partir du serveur PDB.



ClusPro
protein-protein docking

Welcome to Cluspro 2.0

Recent news: [ClusPro server featured on the cover of February 2017 issue of Nature Protocols](#)

Use Without an Account

[Use the server without the benefits of your own account](#)

--or--

Login

Username:

Password:

Login

--or--

[Sign up for an account](#)

[Forgot Password?](#)
[Reset Password](#)

[Forgot Username?](#)
[Retrieve Username](#)

Figure 26 : Page d'accueil de ClusPro .(<https://cluspro.bu.edu/login.php>)

Après avoir déterminé les protéines en interactions avec XPC et leurs structures PDB, une modélisation de ses interactions par ClusPro est ensuite réalisée avec le modèle de structure prédit retenu de XPC.

3. Visualisation moléculaire de XPC

PyMOL (<https://pymol.org/dsc/>) est un logiciel de visualisation moléculaire, utilisé pour visualiser les structures des protéines étudiées. Grâce à sa fonction « *align* », PyMol a été utilisé pour mesurer la similarité structurale en calculant le « rms » entre les modèles de XPC prédites (XPC-Swiss-Model1, XPC-Phyre2 et XPC-model3-TASSER) les fragments de structure de XPC déterminés expérimentalement tirés de *Protein Data Bank*

La commande « *mutagenesis* » de Pymol, nous a, en outre, permis d'obtenir des structures d'XPC mutés selon les rs étudiés. Ces structures mutées vont permettre l'analyse de l'impact de la mutation sur les propriétés de la protéine XPC (notamment ses sites de formation de complexe protéine –protéine)

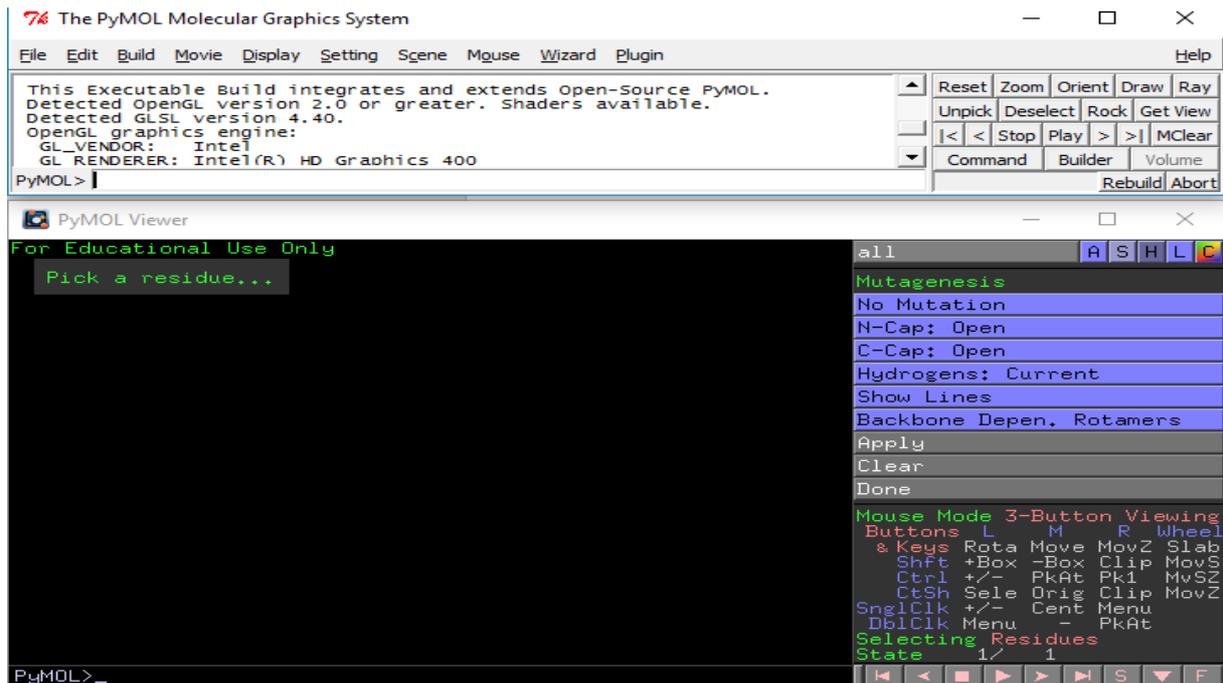


Figure 27 : Page d'application Pymol (mutagenèse)

4. Recherche de domaines protéiques et localisation des SNP(rs)

Pour localiser les rs dans les domaines protéiques de XPC nous avons interrogé à la base de donnée Interpro.

Interpro (<https://www.ebi.ac.uk/interpro/>) est une ressource qui fournit une analyse fonctionnelle des séquences de protéines en les classant dans les familles et en prédisant la présence des domaines et des sites importants. Pour classer les protéines de cette façon, InterPro utilise des modèles prédictifs, connus sous le nom des signatures, fournis par plusieurs bases de données différentes (appelées bases de données membres) qui composent le consortium InterPro.

The screenshot shows the InterPro website homepage. At the top, there is a navigation bar with links for EMBL-EBI, Services, Research, Training, and About us. The main header features the InterPro logo and the tagline "Protein sequence analysis & classification". A search bar is located in the top right corner, with a search icon and a magnifying glass. Below the search bar, there are examples of search terms: IPR020406, kinase, P51587, PF02032, GO:0007165. A secondary navigation bar includes links for Home, Search, Release notes, Download, About InterPro, Help, and Contact. The main content area is divided into two tabs: "By sequence" (selected) and "By domain architecture". The "InterProScan sequence search" section is the primary focus, containing a form for entering a protein sequence in FASTA format. The form includes a text input field and buttons for "Submit", "Clear", and "Example protein sequence". To the right of the form, there is a section titled "InterProScan" with a diagram illustrating the process: a sequence is analyzed and then applied to nucleotide and protein sequences. Below this, there is a "Need more help?" section with a question mark icon and a link to "More about InterProScan".

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

InterPro
Protein sequence analysis & classification

Search InterPro...
Examples: IPR020406, kinase, P51587, PF02032, GO:0007165

Home Search Release notes Download About InterPro Help Contact

By sequence By domain architecture

InterProScan sequence search

This form allows you to scan your sequence for matches against the InterPro protein signature databases, using InterProScan tool.

Enter or paste a protein sequence in FASTA format (complete or not - e.g. PMPIGSKERPTFFEIFKTRCNKADLGPISLN), with a maximum length of 40,000 amino acid long.

Please note that you can only scan one sequence at a time.

Analyse your protein sequence

Advanced options

Submit Clear Example protein sequence

InterProScan

InterProScan is a sequence analysis application (nucleotide and protein sequences) that combines different protein signature recognition methods into one resource.

[More about InterProScan.](#)

Need more help?

If you need more info on InterProScan, you can either

Figure 28 : Page d'accueil Interpro (<https://www.ebi.ac.uk/interpro/>)

Résultats

et

Discussions

Avant de filtrer les SNPs du gène XPC du fichier de séquençage d'ADN, on avait 17 737 SNPs positionnés sur le chromosome 3. En triant ces SNPs, on a trouvé 4 SNPs touchant le gène XPC ce sont : rs4685074 rs1043943, rs2228000 et rs2733533 afin de les étudier nous avons commencé par voir les listes des SNPs connues sur UCSC.

III.1 Liste des SNPs par Génome Browser

Pour avoir la liste des SNPs on a commencé d'abord par répertorier le gène XPC ; les résultats représentent toutes les formes connues de ce gène. Nous avons choisi celles concernant XPC de l'*Homo Sapiens*, auxquelles on accède par le biais du lien : [XPC \(uc011ave.3\) au chr3 :14145147-14178783](#) : *Homo sapiens Xeroderma pigmentosum*, groupe de complémentation C(XPC), transcrit variant 1, mRNA. (De RefSeq NM_004628) (figure 27). On obtient la localisation chromosomique et tous les variations existantes dans ce gène, des listes des SNPs sont apparues, liste des SNPs communs ; liste Flagged SNPs et liste de tous les SNPs par Génome Browser (Figure 29).



Figure 29 : Localisation chromosomique du gène XPC (3p25)

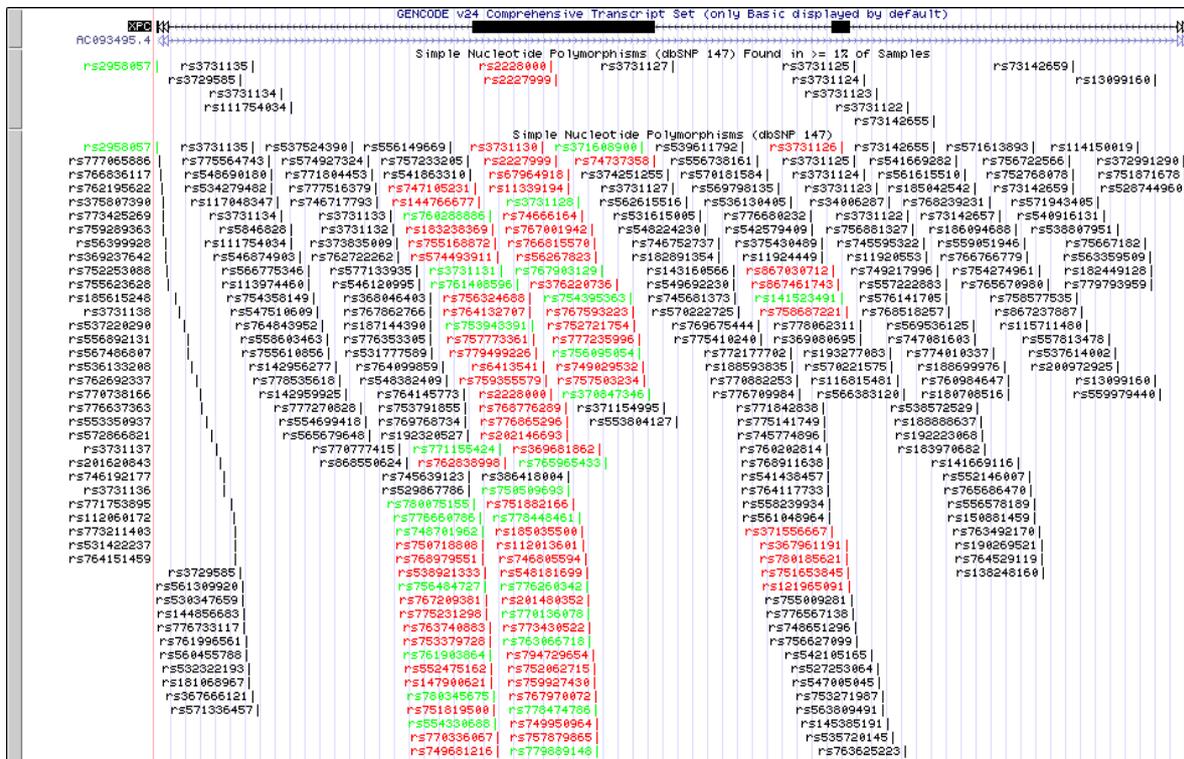


Figure 30 : La liste de tous les SNPs positionnés sur le gène XPC via Génome Browser.

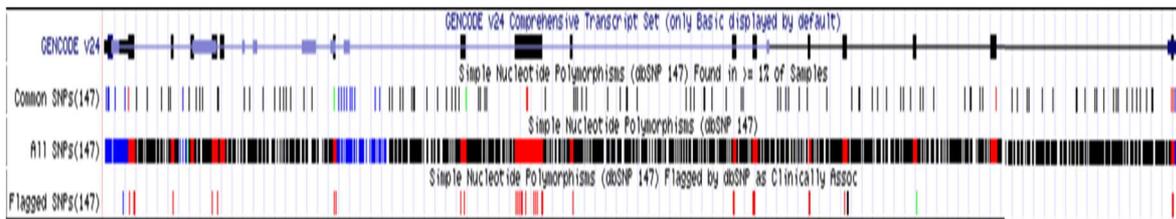


Figure 31 : les listes des SNPs dans *Génome Browser*.

La première ligne de la Figure 31 représente le gène XPC avec 16 exons et 15 introns.

La deuxième ligne représente les SNPs communs $\geq 1\%$ MAF qui sont mappés en un seul emplacement.

La troisième ligne représente tous les SNPs de dbSNP mappés en un seul emplacement.

La dernière ligne représente les SNPs marqué $< 1\%$ MAF qui sont mappés en un seul emplacement, Signalé dans dbSNP comme "cliniquement associé".

Dans la section *Options* de coloration de la page de contrôle des pistes, les termes de la fonction sont regroupés en plusieurs catégories, illustrées ici avec des couleurs par défaut :

Tableau II : Signification des couleurs *Genome Browser*.

Noir	Locus	En amont/en aval du gène variant.
	Intron	Intron variant
Vert	Codage synonyme	Variant synonyme
Rouge	Codage non Synonyme	Gain stop, faux sens variant, perte stop, changement du cadre dans la cadre indel. Site d'épissage: splice accepteur variant, splice donneur variant.
Bleu	Non traduit	5 prime UTR variant, 3 prime UTR variant. Non-coding (ncRNA) : (nctranscript variant).

Table Browser

Après avoir lister la totalité des SNPs dans *Genome Browser*, on passe à identifier les SNPs annotés du génome browser dans table browser (figures 32).

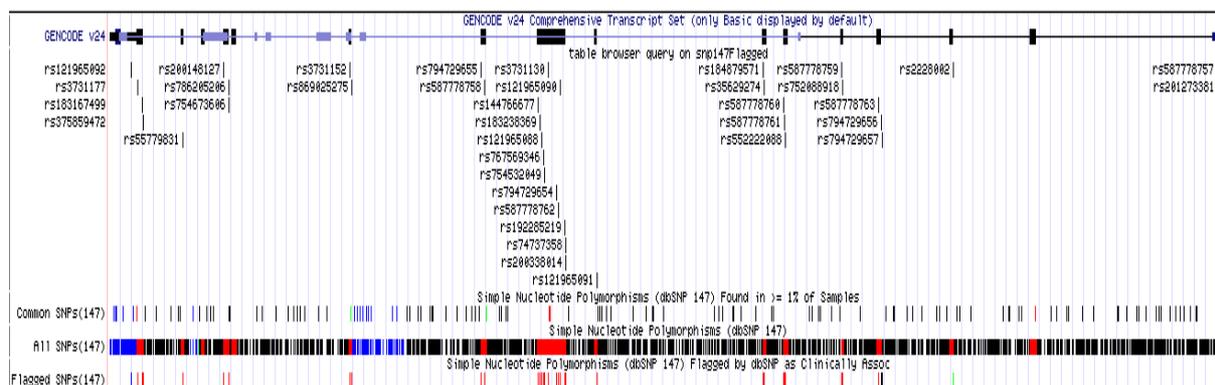


Figure 32 : Listes des SNPs annotés dans Table Browser.

La figure 32 représente l'ensemble des résultats de *Genome Browser* et *Table Browser*. La première ligne c'est ceux de Table Browser qui représente celles qui sont annotés parmi Flagged SNP et les trois dernières lignes ceux de Genome Browser.

En allant dans NCBI pour révéler la nature de chaque SNP de nos rs obtenues du filtrage nous avons réalisé que deux de nos rs sont introniques (rs4685074 et rs2733533) l'autre dans la partie UTR 3 prime (rs1043943) et un qui est faux sens variant (rs2228000). Deux de ces SNP (rs2733533 et rs2228000) sont annotés dans *Table Browser*. Nous allons nous intéresser sur le rs2228000 faux sens dont l'allèle ancestral C est remplacé par T (GCG [Ala]=GTG [val]).

Les trois autres (rs4685074, rs2733533 et rs1043943) pourront faire l'objectif d'une autre étude.

III.2 Modélisation moléculaire de la protéine XPC

En premier, la séquence protéique XPC a été récoltée de la banque de données Uniprot (code d'accèsion Q01831, avec 940aa) (figure 33).

```
>sp|Q01831|XPC_HUMAN DNA repair protein complementing XP-C cells OS=Homo sapiens GN=XPC PE=1 SV=4
MARKRAAGGEPGRGLRSQKSAKSKARREEEEDAFEDKPPKSLLSKVSQGRKRKGC
SHPGGSADGPAKKVAVKVTVKSLENLKVIKDEALSDGDDLDFPSDLKKAHHLKRGATHNE
DSNEEEEESENDWEEVEELSEPVLDVRETAFAFSRLLPVKPVIEIEIETPEQAKTRERSE
KIKLEFETYLRRAMKRFNKGVHEDTHKVLHLLCLLANGFYRNINCSQPDLHAIGLSIIPAR
FTRVLPRDVTYYLSNLVKWFIGTFTVNAELSAEQDNLQTTLERRFAIYSARDDEELVH
IFLLILRALQLLTRLVLSLQPIPLKSATAKGGKPSKERLTADPGGSSETSSQVLENHTKP
KTSKGTQKEETFAGKTCRPSAKGKRKNGGRKKRSKPSSSEDEGPGDKQEKATQRRPHGR
ERRVASRVSYKEESGSDAEGSGSDFELSSGEASDPDESEPGPPKQRKAPAPQRTKAGS
KSASRTHRGSHRKDPSLPAASSSSSSKRGKMCSDGEKAEKRSIAGIDQWLEVFCEQEE
KHWCVDCVHGVVQPLTCYKYATKPMYVVGIDSDGIVRDVTRQYDPVWHTVTRKCRVDA
ENWAETLRPYQSPFMDREKKEDLEFQAKHMDQPLPTAIGLYKNHPLYALKRHLLKYEAIY
PETAAIILGYCRGEAVYSRDCVHTLHSRDTWLKARVVRLEGEVYKMWKGFNSNRARKARLA
EPQLREENDLGLFGYWTTEEYQPPVAVDGKVPNRNEFGNVYFLPSMMPIGCVLNLNLH
RVARKLDIDCVQAITGDFHGGYSHPVTDGYIVCEEFKDVLLTAWENEQAVIERKEKEKK
EKRALGNKLLAKGLLIRERLKRRYGPKSEAAAPHTDAGGGLSSDEEEGTSSQAEAAARIL
AASWPQNRDEEEKQKLGKGGPKTKREKKAASHLPFEQL
```

Figure 33 : La séquence protéique du XPC sous format fasta tiré de la banque de données Uniprot.

La modélisation moléculaire proprement dite s'est basée sur différentes approches de prédiction de la structure tridimensionnelle de XPC (Q01831): la modélisation comparative, la reconnaissance de repliements et l'approche *ab initio*.

III.2.1 Recherche de similarité de la séquence cible/patrons par Blastp/PDB

Un alignement local du XPC (Q01831) s'est fait par le programme blastp contre les séquences stockées dans la *Protein Data Bank*. Cette étape est primordiale pour le choix de la méthode de prédiction de la structure 3D de XPC.

Les résultats sont apparus en trois parties, la première partie est un diagramme avec des barres colorées représentant les séquences possédant les scores les plus élevés avec notre séquence cible.

La deuxième partie représente une liste des séquences avec leurs numéros d'accension, le score de l'alignement, et l'E-value (Tableau III); la troisième partie, des alignements proprement dits de notre séquence cible avec chacune des séquences listées dans la partie 2.

Le tableau représente sept résultats (Tableau III):

- Les trois premières structures représentent la structure de Rad4-Rad23 avec le code pdb (2qsh,2yir,2qsf) avec une similarité de 26% sur 45% de la séquence du XPC.

- Une structure du complexe XPC-TFIIH avec le code pdb (2rvb) avec une similarité de 100% sur 5% de la séquence du XPC.

Résultats et discussions

-Deux structures du complexe XPC-CENTRINE avec le code pdb (2obh,2ggm) avec une similarité de 100% sur 1% de la séquence du XPC.

-Une structure du Passenger Domain Of Plasmid Ecoded Toxin avec le code pdb (2om9) avec une similarité de 40% sur 3% de la séquence du XPC.

Tableau III : Résultats d'alignement de séquence de XPC contre la PDB généré par Blastp.

Séquence similaires à la cible	Score total	Pourcentage des résidus alignés	Evalue	Pourcentage d'identité
Chaîne A de la structure Rad4-Rad23 liée à un ADN désapparié de <i>Saccharomyces cerevisiae</i> (2QSH.pdb)	129	45% de 207 à 531 aa	1 ^e -20	26%
Chaîne A de la Rad4-rad23 réticulé à un ADN normal de <i>Saccharomyces cerevisiae</i> (4YIR.pdb)	129	45% de 207 à 531 aa	1e-20	26%
Chaîne A de la structure du complexe Rad4-Rad23 de <i>Saccharomyces cerevisiae</i> (2QSF.pdb)	128	45% de 202 à 526 aa	1 ^e -20	27%
structure du complexe entre le domaine acide XPC et P62 domaine de TFIIH de <i>Homo sapiens</i> (2RVB.pdb)	59.7	5% de 5 à 52 aa	4 ^e -11	100%
Peptide Centrine2-XPC <i>synthetic construct</i> (2OBH.pdb)	37.0	1% de 2 à 18 aa	0,003	100%
complexes Centrine2-XPC (2GGM) <i>synthetic construct</i>	37.0	1% de 1 à 17 aa	0,003	100%
Passenger Domain Of Plasmid Ecoded Toxin <i>Escherichia coli</i> (2OM9.pdb)	33.5	3% de 214 à 243 aa	1,4	40%

D'après ce résultat, nous pouvons déduire que la protéine XPC n'a pas de structure 3D déterminée expérimentalement. Cependant, la structure de quelques fragments qui représentent 5% (2RVB.pdb), 1% (2OBH.pdb) et 1% (2GGM.pdb) de la séquence de XPC ont été déterminées par Cristallographie aux rayons X et par RMN, respectivement. Elles représentent un alignement significatif avec des E-value très petites.

Les structures (2QSH.pdb) 26 % (4YIR.pdb) 26% (2QSF.pdb) 27 % représentent une faible identité de séquences avec la XPC avec des valeurs de E-values très petites de 1e-20/4^e-11.

Par contre, la structure de Passenger Domain Of Plasmid Ecoded Toxin d'Escherichia coli représente un alignement de séquence non significatif (aléatoire) avec un E-value supérieurs à 1.

Ce résultat peut nous aider quant au choix de la méthode de modélisation moléculaire à utiliser pour prédire la structure 3D de la XPC. Cette faible identité entre cible et patrons nous a poussé à utiliser les trois approches de prédiction de la structure des protéines afin de pouvoir couvrir toute la séquence de XPC.

III.2.2 Collection des structures expérimentales de XPC

D'après la PDB, la structure de quatre fragments de la XPC en interaction avec d'autres protéines ont été téléchargées. Ces structures pourront servir pour la validation des modèles de structure de XPC prédits qui pourraient présenter une bonne similarité structurale avec les fragments déterminés expérimentalement :

*Structure RMN du domaine C-terminal (T94-Y172) de la centrine humaine 2 en complexe avec un peptide de 17 résidus (N847-R863, que nous avons nommé P1-2a4j) de la protéine du groupe C de *Xeroderma pigmentosum* (code PDB_2A4J)_(Figure 34).

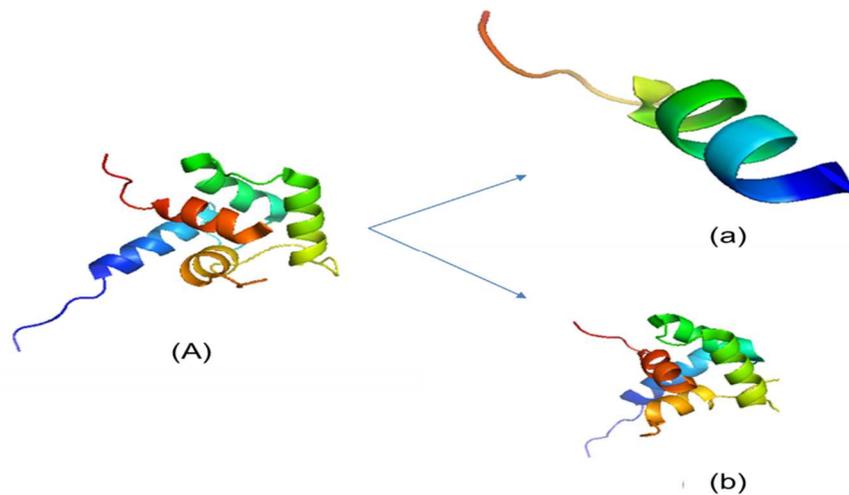


Figure 34 : (A) Structure du complexe Centrine2-XPC déterminée par RMN (code PDB 2a4j) ; (a) fragment XPC et (b) Centrine-2.(Yang et al., 2015).

*Structure cristallographique de la Centrine2 en complexe avec deux peptide de 17 résidus (N847-R863, que nous avons nommé P2c-P2d-2ggm) de la protéine du groupe C de *Xeroderma pigmentosum* (code PDB 2OBH et 2GGM) _(Figure35).

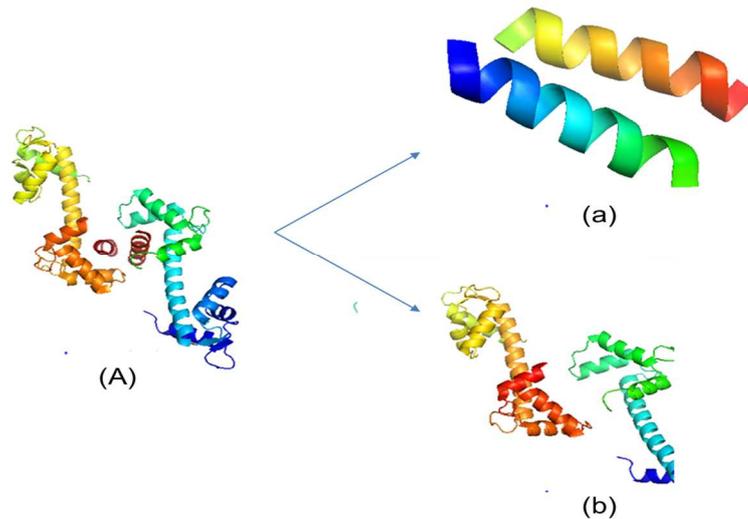


Figure 35 : (A) Structure du complexe Centrine2-XC déterminée par diffraction des rayons X (code PDB 2GGM; (a) fragment XPC et (b) Centrine2 (Thompson et Charbonier, 2006).

*Structure de la solution RMN du complexe entre le domaine acide de la XPC et le domaine TFIIH p62 PH (code PDB 2rvb). Cette dernière est une protéine de liaison d'ADN/transcription chez l'*Homo sapiens* (Figure 36).

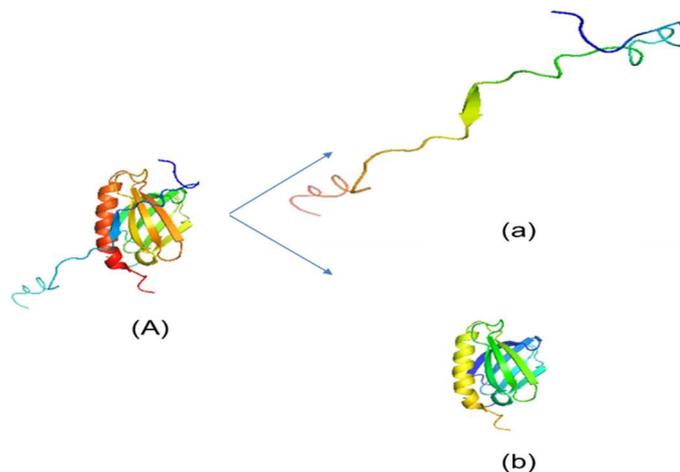


Figure 36 : (A) Structure du complexe TFIIH-XPC déterminée par RMN (code PDB 2rvb); (a) fragment XPC et (b) TFIIH. (Okuda et Nishimura, 2007).

A partir de ces trois fichiers pdb avec les codes pdb (2a4j, 2ggm/2obh et 2rvb), nous a pu extraire les coordonnées pdb des fragments de la XPC à partir des complexes déterminés expérimentalement. Ces fragments ont été nommés :

- P1-2a4j.pdb: fragment du XPC avec 17 aa (de 1 à 16) de la chaîne B issus du code pdb 2a4j (Figure 34a) ;
- P2c-P2d-2ggm/ P2c-P2d-2obh.pdb : fragment du XPC avec deux chaînes C et D du 847 à 826 aa issus du code pdb 2ggm ou 2obh (Figure 35a) ;
- P3-2rvb.pdb: fragment du XPC de la chaîne A issus du code pdb 2rvb du 109 à 156 aa (Figure 36a).

De cette manière, nous avons pu également obtenir les coordonnées pdb des structures des protéines qui interagissent avec la XPC. Ces dernières ont été nommées :

- ✓ 2a4jSANSxpc-CETN.pdb qui représente la structure de la centrine issue du code pdb 2a4j du 136 à 166 aa (Figure 34b) ; La centrine est une protéine qui appartient à la superfamille de protéines à mains-EF qui lie le calcium (Ca^{2+}). La centrine 2 régule la division cellulaire lors du cycle du centrosome.
- ✓ 2ggmSANSxpc-CETN.pdb qui représente la structure de la centrine issue du code pdb 2ggm (Figure 35b) ;
- ✓ 2rvbSANSxpc-TFIIH.pdb qui représente la structure de TFIIH issues du code pdb 2rvb (Figure 36b). TFIIX est un facteur général de transcription pour l'ARN polymérase II possédant une activité hélicase et jouant un rôle pendant l'initiation de la transcription chez les eucaryotes

III.2.3 Prédiction de structure 3D de la protéine XPC

Compte tenu de la faible identité de séquence entre la cible XPC et ses patrons, nous avons opté pour la prédiction de sa structure 3D l'utilisation des trois approches de modélisation moléculaire, à savoir la modélisation comparative, modélisation par reconnaissance de repliements et l'approche *ab initio*. Cette prédiction a été effectuée en utilisant les trois programmes Swiss-Model, Phyre2 et I-TASSER, respectivement. Les résultats de la modélisation de la structure 3D de la XPC avec ces trois programmes est résumé dans le tableau IV.

Tableau IV : Tableau récapitulatif des résultats de la prédiction de la structure 3D de la XPC

Programme de prédiction de la structure 3D de la cible	Méthode de prédiction de la structure 3D	Nombre des modèles prédits	Nombre d'acides aminés (aa)	Nomenclature du modèle
Swiss-Model	Modélisation comparative	Deux modèles	176 à 831 aa 109 à 156 aa	- XPC-SwissModel1.pdb (patron 2qsg.pdb, structure de la RAD4, une protéine impliquée dans la réparation d'ADN) - XPC-SwissModel2.pdb (patron 2rvb.pdb)
Phyre2	Modélisation par reconnaissance de repliements	Un seul modèle	185 à 826 aa	- XPC-Phyre2.pdb (patron 2qsh.pdb (Min et Pavletich,2007), structure de la RAD4, une protéine impliquée dans la réparation d'ADN,)
I-TASSER	Modélisation par <i>ab initio</i> combinée à la reconnaissance de repliements	Trois modèles	940 aa	- XPC-model1-TASSER.pdb - XPC-model2-TASSER.pdb - XPC-model3-TASSER.pdb

a) Modélisation comparative de la structure de la XPC par Swiss-Model

La modélisation comparative de notre protéine par Swiss-Model s'est basée sur la construction d'un modèle de la protéine cible XPC à partir de sa séquence d'acides aminés sur la base d'une structure expérimentale tridimensionnelle d'une protéine présentant une bonne conservation de séquences. Ce programme a généré deux modèles de structures probables de la XPC :

a) -XPC-SwissModel1 : le patron est la protéine de réparation d'ADN RAD4 (code pdb 2qsg) avec une similarité de séquence de 23,63% du 176 à 831 aa de la protéine XPC.

b) - XPC-SwissModel2 : le patron est un fragment de la protéine XPC déterminé expérimentalement (code pdb 2rvb) avec une similarité de 100 % du 109 à 156 aa de la protéine XPC

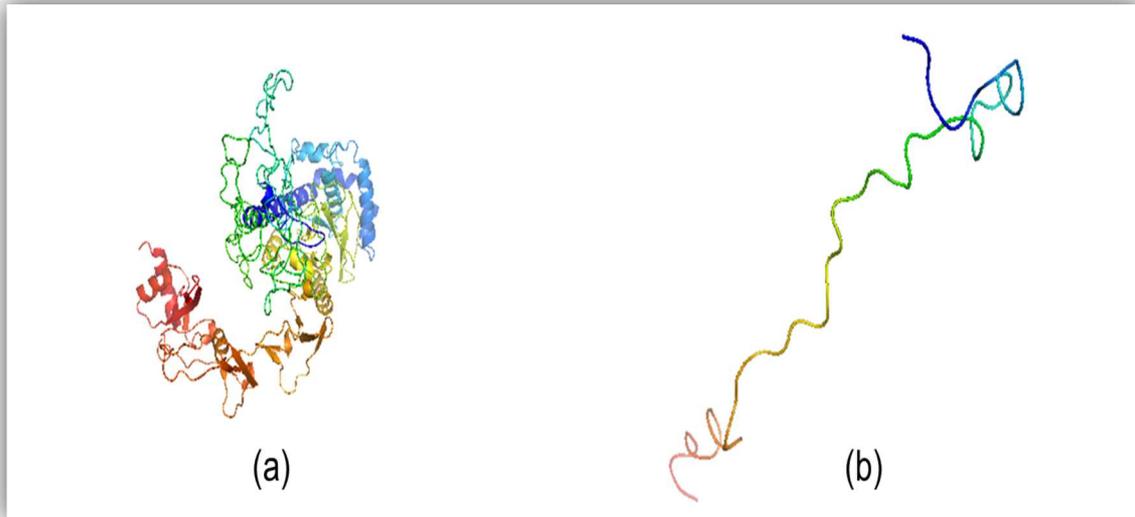


Figure 37 : Modèles de structures tridimensionnelles du XPC prédits par SwissModel. ;

(a) XPC-SwissModel1.pdb, (b) XPC-SwissModel2.pdb.

b) Modélisation par reconnaissance de repliements de la structure de la XPC par Phyre2

La reconnaissance de repliements est un procédé de modélisation utilisé pour modéliser par analogie, les protéines qui ont les mêmes repliements que des protéines de structures connues, mais qui n'ont pas de protéines homologues de structure connue. Elle consiste à enfile la séquence de la cible sur une bibliothèque de repliements afin de déterminer les structures qui correspondent le mieux à la séquence sur la base d'un critère énergétique ou de score.

Une prédiction de structure de la protéine XPC par reconnaissance de repliements s'est faite par le programme Phyre2 ; un modèle de structure a été prédit (XPC-Phyre2.pdb) sur la base d'un seul patron la structure de la RAD4, une protéine impliquée dans la réparation d'ADN ayant le code pdb 2qsh (Min et Pavletich,2007). Le modèle généré couvre la séquence de la XPC de 185 à 826 aa (figure 38)

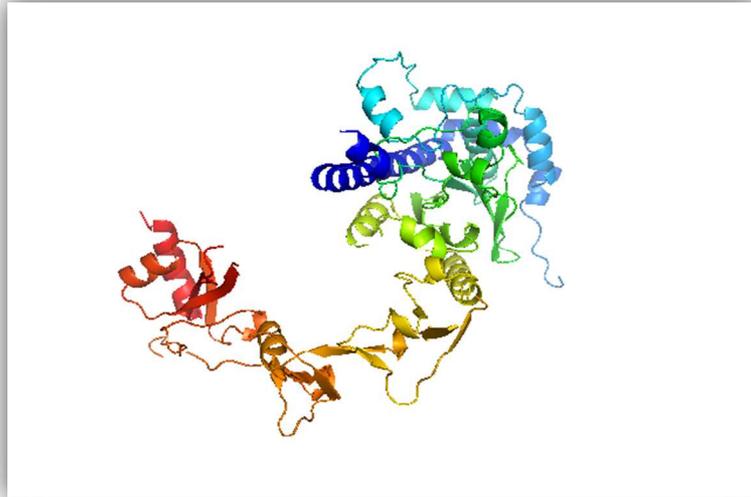


Figure 38 : Modèles de structures tridimensionnelles du XPC prédits par Phyre2 (XPC-Phyre2.pdb).

c) Modélisation de la structure de la XPC par ab initio combinée à la reconnaissance de repliements par I-TASSER

XPC-model1-TASSER.pdb, (b) XPC-model2-TASSER.pdb (figure b) et (c) XPC-model3-TASSER.pdb. La modélisation de la structure de la XPC par I-TASSER a généré trois modèles (figure 38): XPC-model1-TASSER.pdb (figure a); XPC-model2-TASSER.pdb (figure b); XPC-model3-TASSER.pdb (figure c). Ces trois modèles de structures ont couvert toute la longueur de la séquence de XPC (940aa).

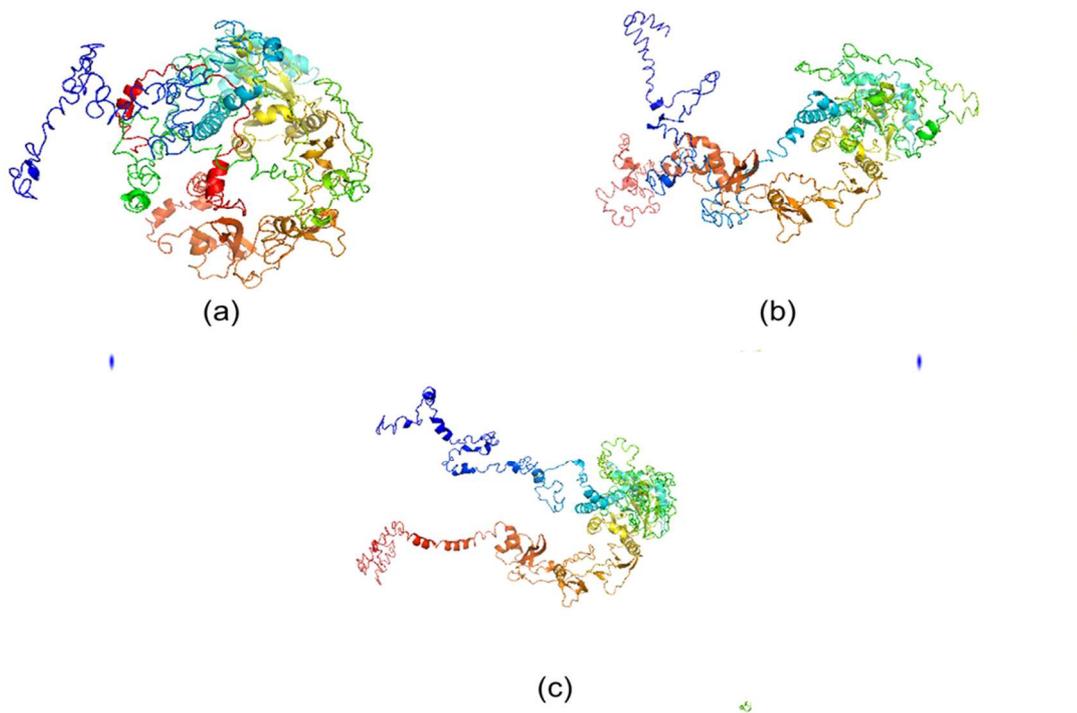


Figure 39 : Modèles de structures tridimensionnelles du XPC prédits par I-TASSER ; (a) XPC-model1-TASSER.pdb,(b) XPC-model2-TASSER.pdb et (c) XPC-model3-TASSER.pdb.

d) Analyse des modèles des structures prédits pour la XPC

Toutes les structures issues de la modélisation moléculaire ont été visualisées par PyMol. Pour mesurer leur similarité structurale par le calcul des rms, un alignement de structure de ces modèles a été effectué par PyMol. Le RMS mesure la distance entre les $C\alpha$ après superposition optimale des deux fragments ; plus le rms est petit plus les modèles de structures sont proches Tableau V.

Tableau V : Résultats de la mesure de la similarité structurale des modèles prédits avec les fragments de la XPC déterminés expérimentalement par le calcul des rms

	XPC-model1-TASSER.pbb	XPC-model2-TASSER.pdb	XPC-model3-TASSER.pdb	XPC-Phyre2.pdb	XPC-SwissModel1.pdb	
P1-2a4j.pdb	17 vs 940	17 vs 940	17 vs 940	17 vs 465	17 vs 655	Nombre d'aa
	17 atomes	16 atomes	14 atomes	7 atomes	13 atomes	Atomes alignés
	5.033	3.418	1.897	3.168	5.008	RMS
P2c-P2d-2ggm.pdb	33 vs 940	33 vs 940	33 vs 940	33 vs 465	33 vs 655	Nombre d'aa
	17 atomes	17 atomes	17 atomes	12 atomes	15 atomes	Atomes alignés
	4.487	2.650	1.672	7.254	7.242	RMS
P3-2rvb.pdb	48 vs 940	48 vs 940	48 vs 940	48 vs 465	48 vs 655	Nombre d'aa
	48 atomes	48 atomes	45 atomes	17 atomes	22 atomes	Atomes alignés
	15.081	15.957	15.554	7.111	7.136	RMS
XPC-Phyre2.pdb	465 vs 940	465 vs 940	465 vs 940	/	465 vs 465	Nombre d'aa
	359 atomes	343 atomes	344 atomes	/	325 atomes	Atomes alignés
	2.868	0.325	0.310	/	0.430	RMS
XPC-SwissModel1.pdb	655 vs 940	655 vs 940	655 vs 940	655 vs 465	/	Nombre d'aa
	449 atomes	423 atomes	411 atomes	325 atomes	/	Atomes alignés
	3.485	1.676	1.027	0.430	/	RMS

D'après les résultats des rms calculés entre les structures des modèles prédits par les différents programmes de prédiction de structures 3D avec celles des fragments déterminés expérimentalement (P1-2a4j, P2c-P2d-2ggm et P3-2rvb), nous pouvons suggérer que le modèle de structure **3** de la protéine XPC généré par le programme I-TASSER est celui qui présente une meilleure similarité structurale avec les structures expérimentales avec des valeurs de rms très faibles comparés aux autres modèles prédits. La comparaison des structures des modèles prédits avec celles des fragments déterminés expérimentalement de la XPC permet de discriminer certains modèles de structures. C'est pour cette raison que pour la suite de notre

travail, nous avons retenu la structure du modèle 3 prédit par I-TASSER comme structure 3D probable de la XPC. L'alignement de structures est illustré dans la Figure 39 ; la structure de XPC-modèl3-ITASSER en bleu est alignés avec les fragments de XPC P1-2a4j en jaune, P2c-P2d-2ggm en magenta et p3-2rvb en vert.

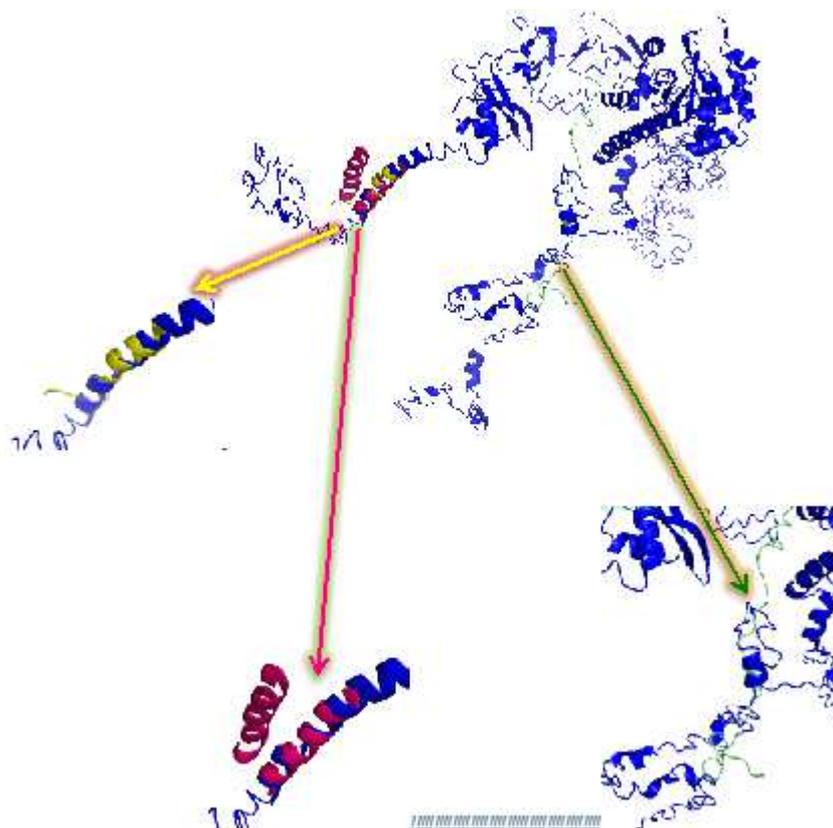


Figure 40 : Superposition des structures des fragments déterminés expérimentalement avec la structure du modèle 3 prédit par I-TASSER .

III.2.4 Amarrage protéine-protéine (XPC-protéine) par ClusPro

L'amarrage de protéines (appelé aussi le *docking* ou l'ancrage) consiste à prédire la conformation optimale stable d'un complexe de protéines ou protéine-ligand à partir des structures individuelles des partenaires dans leur état non lié. La stabilité de ces complexes est essentiellement évaluée selon la complémentarité géométrique (ou structurelle) et la complémentarité chimique. La combinaison de ces paramètres conduit à la formation du complexe moléculaire (Khemili, 2013).

Pour connaître les éventuelles protéines qui pourrait interagir avec notre protéine XPC, nous avons interrogé la base de données String. Les résultats de cette recherche ont montré que, la XPC interagit avec une dizaine de protéines (Figure 41) ;

- ✓ RAD23A, RAD2B : Le récepteur de la chaîne multi ubiquitine impliqué dans la modulation de la dégradation protéasomique, lié aux chaînes de poly ubiquitine, proposé d'être capable de se lier simultanément au protéasome 26S et aux substrats poly ubiquités et à délivrer des protéines ubiquitines au protéasome.
- ✓ GTFIIH : Facteur général de transcription IIH, polypeptide 1, 62 kDa; Composant du facteur de transcription basal du noyau-TFIIH impliqué dans la réparation de l'excision nucléotidique (NER) de l'ADN et lorsqu'il est en complexe avec CAK, dans la transcription de l'ARN par ARN polymérase II.
- ✓ CETN2 : Centrine, EF-main protéine, 2; Joue un rôle fondamental dans la structure et la fonction du centre d'organisation des microtubules, nécessaire pour la duplication de centriole et la formation correcte de la broche, avoir un rôle dans la régulation de la cytokine et de la stabilité du génome grâce à la coopération avec CALM1 et CCP110.

Toutes ces interactions s'agissent des interactions directes avec la XPC mais il existe d'autres interactions par des protéines intermédiaires tel que : XPA, DDB2, ERCC2, ERCC3, ERCC4 et ERCC5.

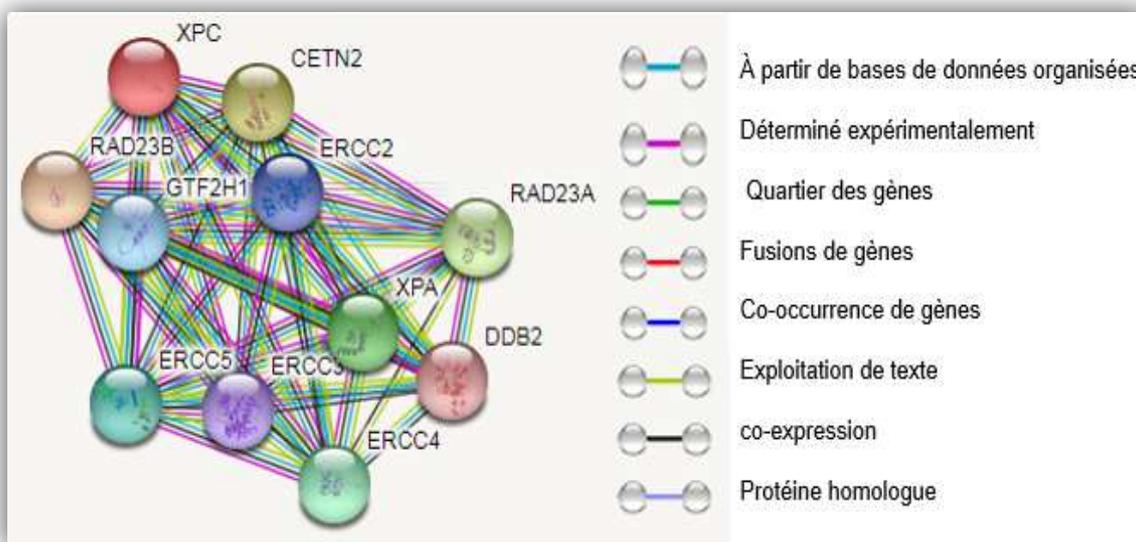


Figure 41 : Interactions XPC-Protéines d'après la base des données String.

Compte tenu de la disponibilité du complexe expérimental du XPC-Centrine2, nous avons utilisé le serveur ClusPro2.0 pour prédire les éventuelles conformations de ce complexe en prenant le modèle 3 généré par I-TASSER comme récepteur 1 et la structure de la centrine 2 (issus du code pdb 2ggm) comme récepteur 2. Cette étude permet :

- ✓ De tester la fiabilité du programme ClusPro à déterminer le bon complexe protéine-protéine (dans notre cas XPC-centrine2 du code pdb 2ggm, et XPC- TFIIH1 du code pdb 2rvb),
- ✓ De valider la structure du modèle 3 générée par I-TASSAR si un des complexes prédits par ClusPro correspond à celui déterminé expérimentalement.

Nous avons finalement retenu 30 modèles pour chacun des complexes XPC-centrine2 et XPC-TFIIH1. La comparaison entre les complexes prédits avec ceux déterminés expérimentalement (codes pdb 2ggm et 2rvb) a été effectuée par la mesure du rms après superposition des structures par PyMol (tableau VI).

Ces résultats suggèrent que ClusPro est en mesure de prédire de façon fiable la structure des complexes et que le modèle de la structure 3 généré par I-TASSER pourrait correspondre à la structure native de la XPC. Le Tableau représente RMS dont les modèles XPC-CETN22 et XPC-TFIIH23 sont les modèles les plus probable d'où ils représentent une similarité avec leurs modèles expérimentales 2GGM et 2RVB respectivement et ils représentent un RMS petit (RMS XPC-CETN22 = 0.701 ; XPC-TFIIH23 = 0.644. Le score énergétique de ClusPro de chaque complexe est très petit -808.3 et -748.9 de XPC-CETN22 et XPC-TFIIH23 respectivement.

Tableau VI : Valeurs de la fonction de score de ClusPro des complexes prédits et valeurs du rms obtenues après superpositions de ces modèles avec les complexes respectifs déterminés expérimentalement

Les modèles	2GGM	Le score énergétique	Les modèles	2RVB	Le score énergétique
XPC-CETN1	0.780	-1002.0	XPC-TFIIH1	0.891	-874.4
XPC-CETN2	0.798	-914.3	XPC-TFIIH2	1.081	-853.2
XPC-CETN3	0.814	-982.4	XPC-TFIIH3	0.932	-846.3
XPC-CETN4	17.44	-942.4	XPC-TFIIH4	1.066	-815.8
XPC-CETN5	0.800	-947.9	XPC-TFIIH5	1.131	-893.8
XPC-CETN6	0.864	-992.3	XPC-TFIIH6	1.021	-759.6
XPC-CETN7	0.836	-860.3	XPC-TFIIH7	0.869	-766.6
XPC-CETN8	0.866	-944.8	XPC-TFIIH8	0.779	-783.6
XPC-CETN9	0.784	-942.1	XPC-TFIIH9	1.047	-747.6
XPC-CETN10	0.913	-861.8	XPC-TFIIH10	0.937	-828.3
XPC-CETN11	0.918	-854.0	XPC-TFIIH11	0.795	-738.9
XPC-CETN12	0.742	-866.8	XPC-TFIIH12	0.930	-729.6
XPC-CETN13	0.870	-884.0	XPC-TFIIH13	0.888	-810.9
XPC-CETN14	0.874	-824.8	XPC-TFIIH14	0.981	-802.0
XPC-CETN15	0.893	-861.4	XPC-TFIIH15	0.834	-805.4
XPC-CETN16	0.762	-876.7	XPC-TFIIH16	0.986	-755.0
XPC-CETN17	0.918	-865.9	XPC-TFIIH17	1.271	-750.8
XPC-CETN18	0.849	-822.6	XPC-TFIIH18	0.777	-747.4
XPC-CETN19	0.835	-965.0	XPC-TFIIH19	0.816	-854.6
XPC-CETN20	0.866	-824.6	XPC-TFIIH20	0.906	-728.0
XPC-CETN21	0.735	-909.1	XPC-TFIIH21	1.041	-805.3
XPC-CETN22	0.701	-808.3	XPC-TFIIH22	0.790	-809.7
XPC-CETN23	0.779	-862.4	XPC-TFIIH23	0.644	-748.9
XPC-CETN24	0.814	-840.9	XPC-TFIIH24	0.942	-791.5
XPC-CETN25	0.816	-885.1	XPC-TFIIH25	1.029	-749.8
XPC-CETN26	0.829	-970.0	XPC-TFIIH26	0.972	-792.4
XPC-CETN27	0.948	-837.3	XPC-TFIIH27	1.013	-762.2
XPC-CETN28	0.904	-839.0	XPC-TFIIH28	0.902	-718.4
XPC-CETN29	1.072	-918.0	XPC-TFIIH29	1.128	-718.0
XPC-CETN30	0.922	-886.8	XPC-TFIIH30	0.776	-773.7

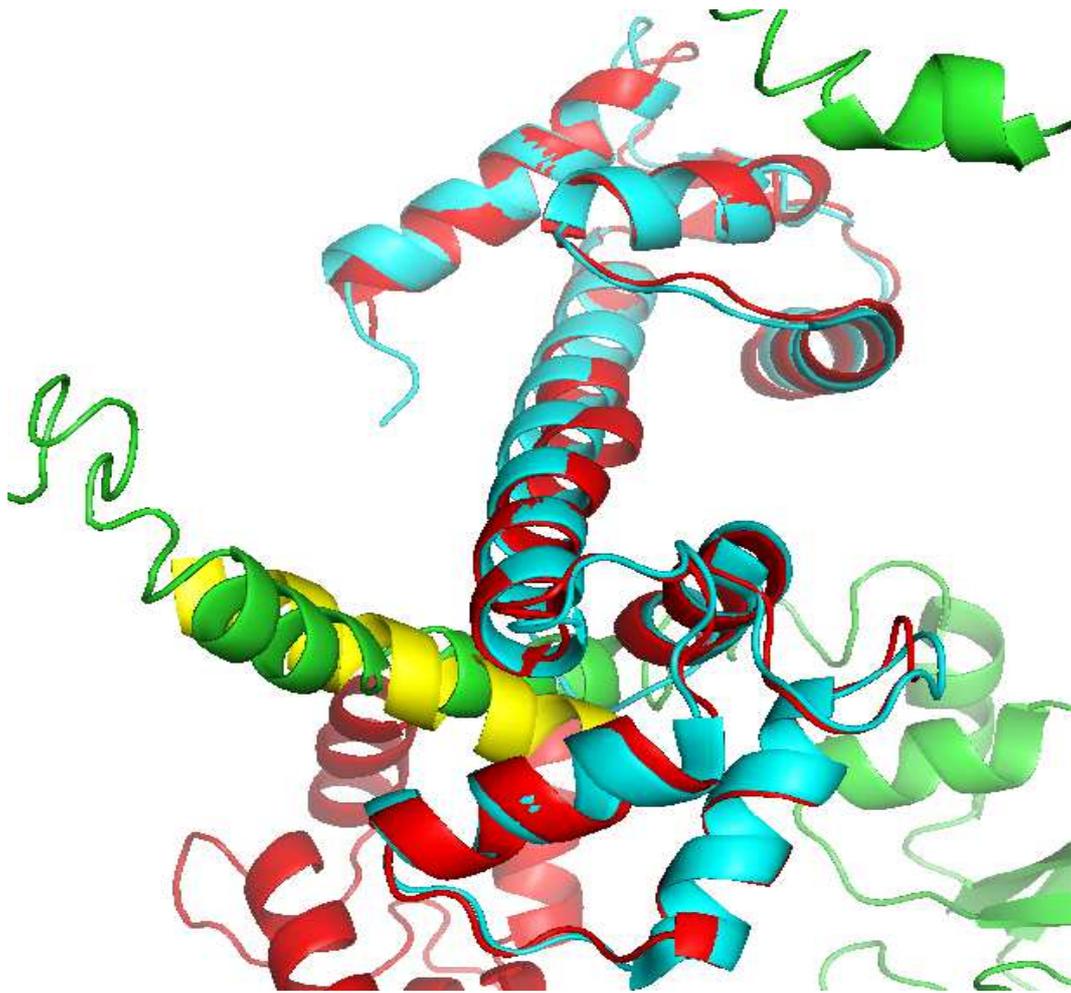


Figure 42 : Superposition de la structure du complexe XPC-CETN prédite par ClusPro (vert-rouge) avec celle déterminée expérimentalement (code PDB 2ggm) en jaune fragment XPC et en Cyan la chaîne de Centrine2.

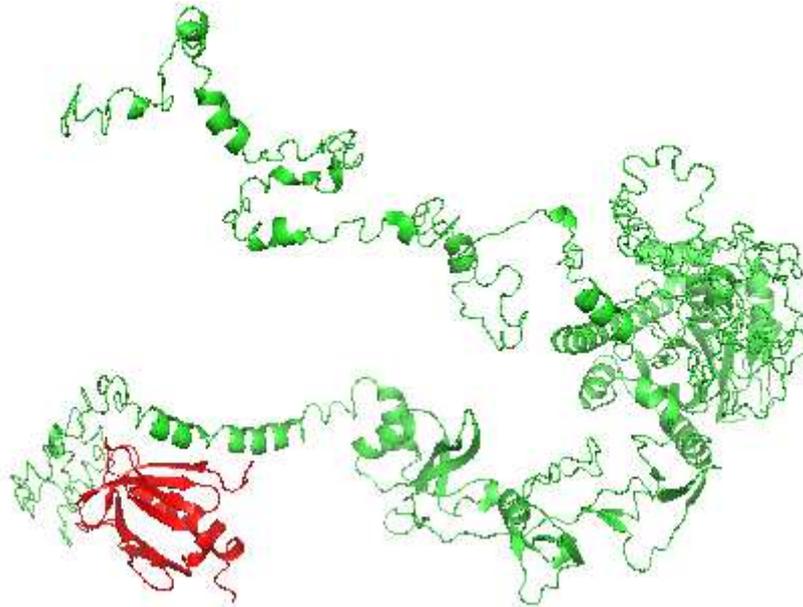


Figure 43 : La structure du complexe XPC-TFIIH prédite par ClusPro .

Pour mieux comprendre les propriétés structurales et fonctionnelles de la XPC, nous nous sommes intéressés à la localisation des rs SNPs annotés qui sont associé à la maladie XP (Flagged SNPs) (Voir annexe V).

III.2.5 Localisation des rs dans la structure de XPC prédite

Pour pouvoir localiser nos rs dans les domaines protéiques fonctionnels dans notre protéine XPC nous avons interrogé la base de données InterPro (figure 45, Tableau VII).

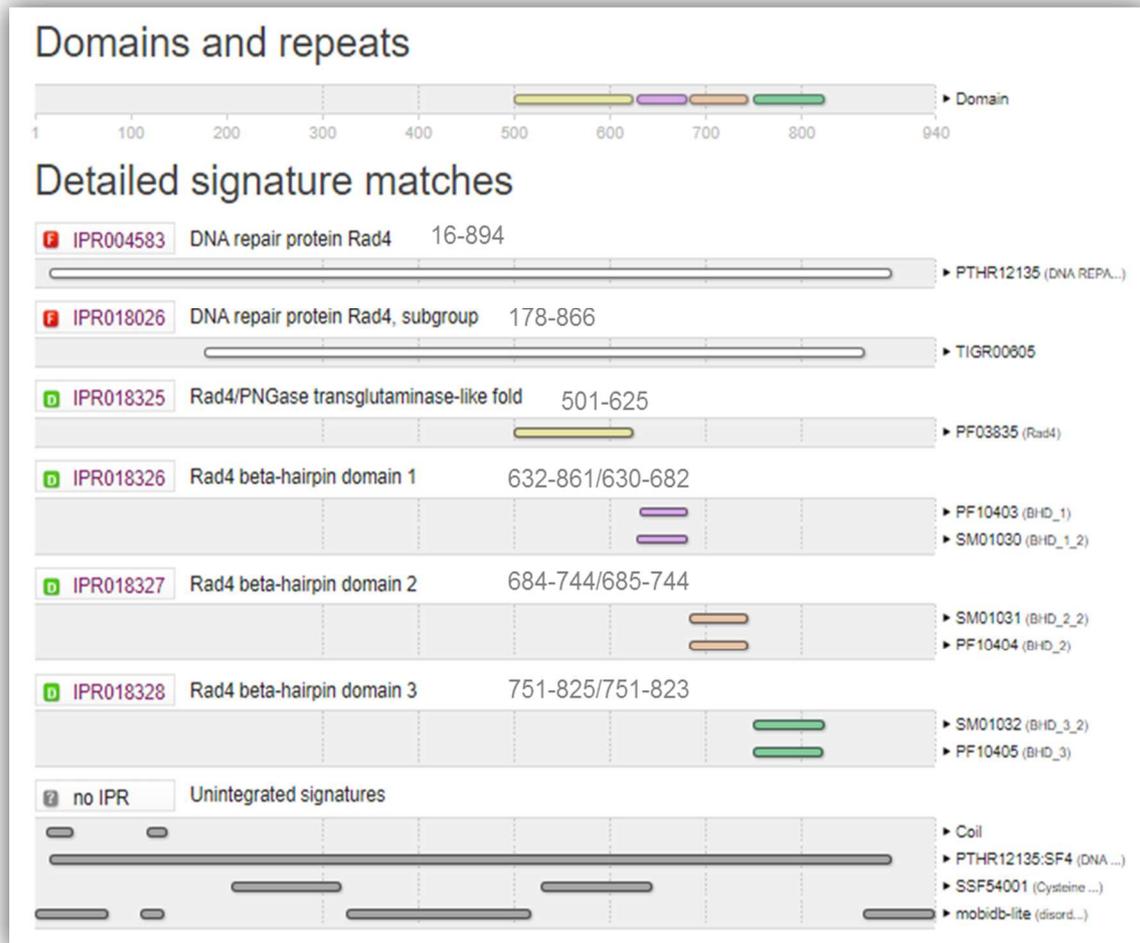


Figure 44 : Les domaines protéiques du XPC d’après la base de données InterPro.

Tableau VII : Tableau des rs associés à la maladie XPC et leurs localisations dans ses Domaines protéiques fonctionnels.

rs	Résidu change	Positions dans la séquence protéique	Domaines
rs587778757	S[Ser] ⇒ RArg]	18	Protéine de réparation d'ADN Rad4
rs2228002	R [Arg] ⇒ R [Arg]	100	
rs587778763	V[Val] ⇒ L[Leu]	160	
rs587778759	K[Lys] ⇒ N[Asn]	199	
rs552222088	R[Arg] ⇒ C[Cys]	240	
rs587778761	P [Pro] ⇒ S [Ser]	246	
rs587778760	Y [Tyr] ⇒ C [Cys]	252	
rs35629274	F[Phe] ⇒ C[Cys]	287	
rs184879571	S[Ser] ⇒ C[Cys]	291	
rs74737358	P[Pro]⇒H[His]/R[Arg] /L[Leu]	334	
rs200338014	P[Pro] ⇒ A[Ala] /S[Ser]	334	
rs192285219	A [Ala] ⇒ V [Val]	341	
Rs121965090	R [Arg] ⇒ W [Trp]	393	
rs587778762	E [Glu] ⇒ G [Gly]	410	
rs794729654	K [Lys] ⇒ R [Arg]	431	
rs2228000	R[Arg]⇒ V [Val]	499	Le pli de type transglutaminase Rad4 / PNGase
rs3731130	M [Met] ⇒ I [Ile]	513	
rs754532049	V [Val] ⇒ A[Ala]	548	
rs121965088	R [Arg] ⇒ Ter	579	
rs183238369	R [Arg] ⇒ C [Cys]	594	
rs144766677	F [Phe] ⇒ S [Ser]	614	
rs587778758	Y [Tyr] ⇒ H [His]	641	
rs3731152	T [Thr] ⇒ M [Met]	689	Domaine Rad4 beta-hairpin 2
rs786205206	N [Asn] ⇒ K [Lys]	754	Domaine Rad4 beta-hairpin 3
rs200148127	G [Gly] ⇒ S [Ser]	802	Domaine Rad4 beta-hairpin 3 et domaine d'interaction avec TFIIH
rs375859472	P [Pro] ⇒ L [Leu]	874	
rs55779831	G [Gly] ⇒ E [Glu]	846	
rs183167499	A [Ala] ⇒ G [Gly] /V [Val]	878	

La localisation de ces rs pathogènes dans la structure du modèle 3 de la XPC prédit par I-TASSER est illustrée dans la figure 45.

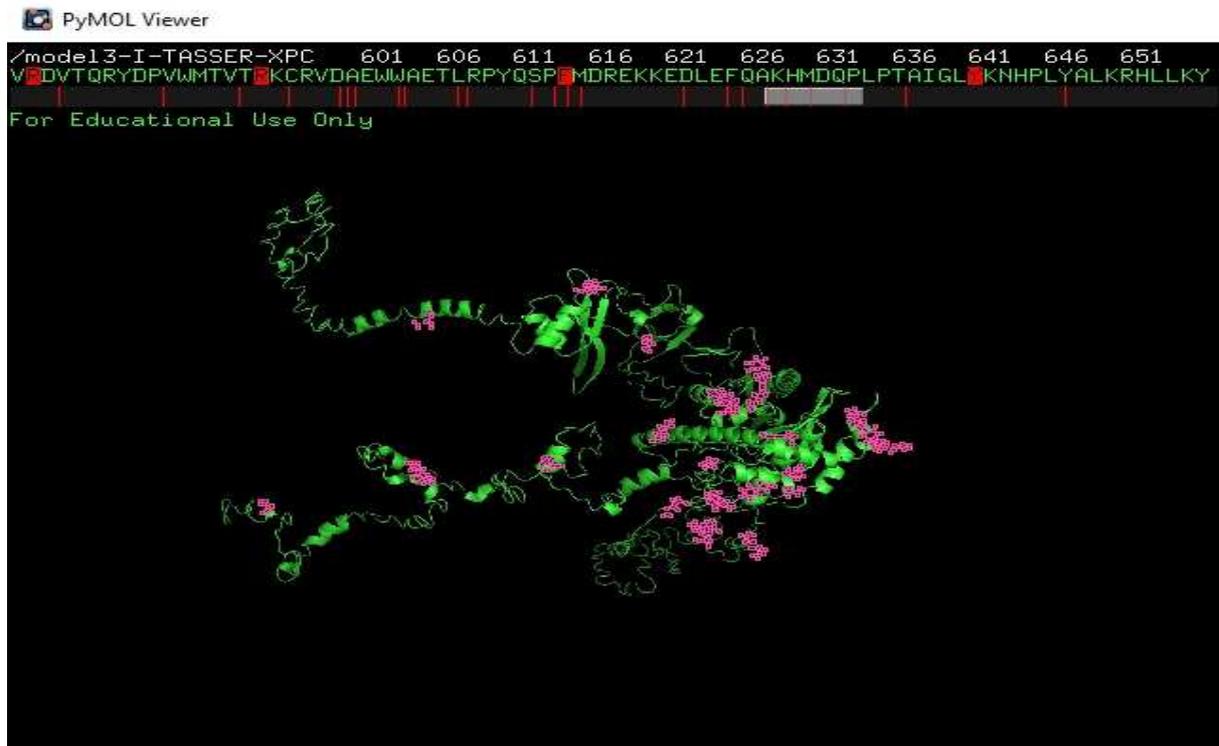


Figure 45 : L'ensemble des rs pathogènes (en sticks roses) positionnés sur la structure du modèle 3 de la XPC prédit par I-TASSER (en cartoon vert).

D'après le tableau VI nous remarquons que la position des rs pathogènes touchent principalement les domaines fonctionnels de la Protéine de réparation d'ADN Rad4. Le domaine de pli de type transglutaminase Rad4 / PNGase, domaine Rad4 beta-hairpin 1, le domaine Rad4 beta-hairpin 2 , le domaine Rad4 beta-hairpin 3

Par ailleurs, aucun rs pathogène n'est trouvé dans le domaine d'interaction de la XPC avec la centrine 2 (du 847 à 866aa). Cependant, 3 rs ont été observés dans le domaine d'interaction de la XPC avec le TFIIF (de 816 à 940 aa) ; le rs375859472 où le résidu P [Pro] en position 874 est substitué par une L [Leu], le rs55779831 où le résidu G846 [Gly] est substitué en E [Glu] et le rs183167499 où le résidu A878 [Ala] est substitué par une G [Gly]. Ces mutations jouent un rôle primordial dans le fonctionnement et la pathogénicité de la XPC par la modification de ses propriétés structurales. (Figure 46).

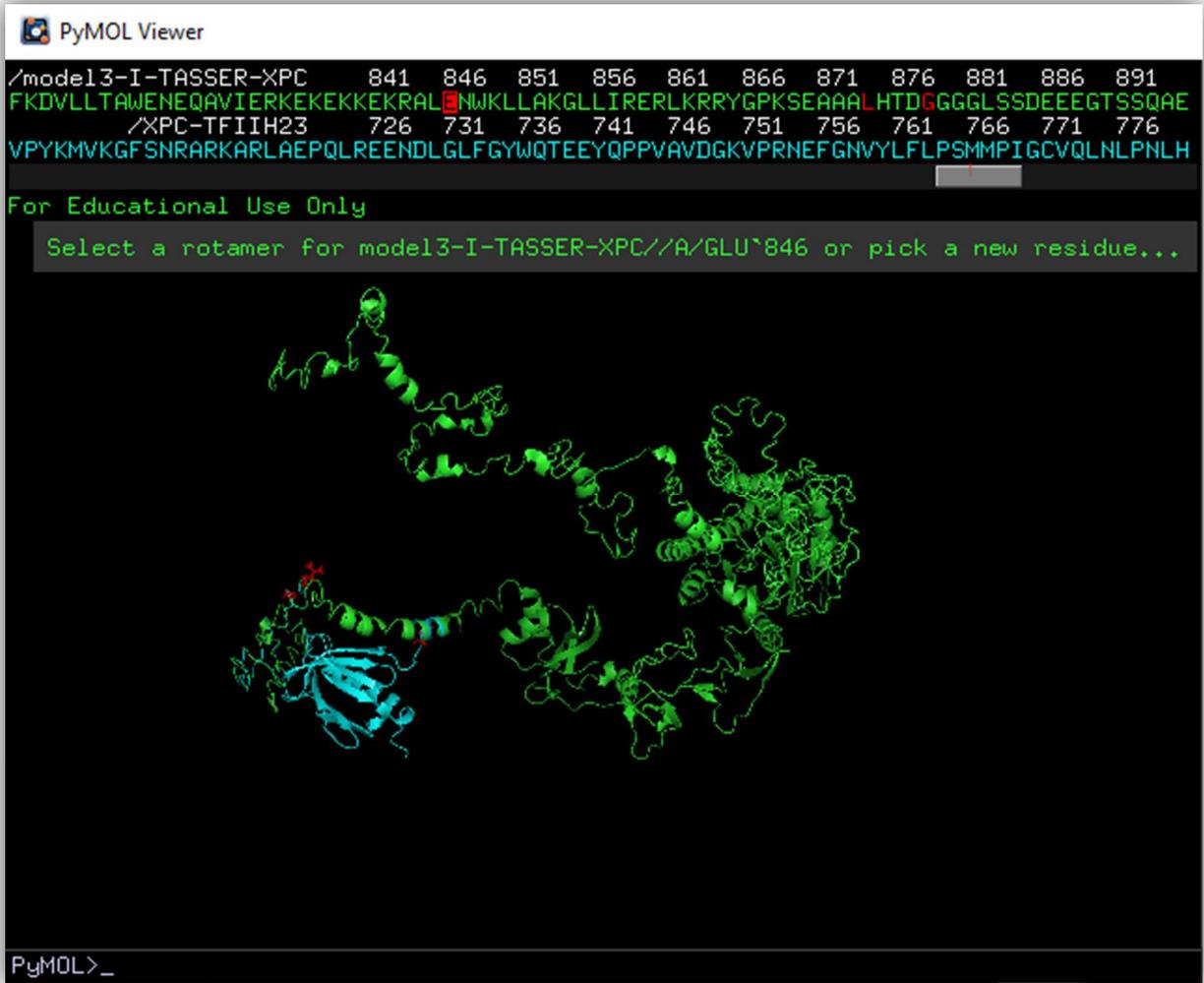


Figure 46 : rs (en rouge) positionnés dans le domaine d'interaction entre la structure de la protéine XPC prédite par I-TASSER (en vert) et le domaine TFIIH (en bleu).

Conclusion

et

Perspectives

Xeroderma pigmentosum est une maladie de transmission autosomique récessive, elle est due à une sensibilité accrue de l'ultraviolet et de la lumière du soleil. A ce jour on dénombre huit gène de XP : XP classique (de A à G) et XPV dont le plus répons dans les pays du Maghreb est le XPC. Cette protéine est impliquée dans le système de réparation de l'ADN (*Nucleotide Excision Repair* NER). La structure 3D des protéines est très importante puisqu'elle détermine ses propriétés biochimiques et fonctionnelles. Cependant, il n'existe pas encore de structure 3D de la protéine XPC jusqu'à ce jour.

L'objectif de ce mémoire est de lister les SNPs touchant la XPC et de voir leur localisation sur un modèle de structure 3D de cette protéine prédit *in silico*. L'étude s'est basée sur deux étapes ; une pour lister les SNPs et l'autre pour la modélisation de la structure 3D de la protéine XPC. Cette dernière étape a comporté deux phases; la prédiction de la structure de la XPC et l'amarrage protéine-protéine (XPC-protéine).

D'après nos résultats, compte tenu de la faible similarité de séquences entre la cible (XPC) et les patrons (issus de la protein Data Bank) effectuée par Blastp, trois méthodes de prédiction de la structure 3D ont été utilisées ; la modélisation comparative par le serveur web SwissModel, la reconnaissance de repliements par le serveur web Phyre2 et la méthode *ab initio* combinée à la reconnaissance de repliements par le serveur I-TASSER.

Plusieurs modèles de structure 3D ont été générés par ces trois méthodes et nous n'avons retenu qu'un seul modèle d'après sa forte similarité structurale mesurée par le calcul des rms après superposition par PyMol avec les fragments de XPC dont la structure a été déterminée expérimentalement. Ce modèle a été généré par le programme I-TASSER. Il est important de souligner que I-TASSER a été démontré étant le meilleur serveur de prédiction de structure 3D d'après les concours d'évaluation CASP 2006-2016 (CASP7, CASP8, CASP9, CASP10, CASP11 et CASP12) (*Critical Assessment of techniques for protein Structure Prediction*, <http://www.predictioncenter.org/casp10/index.cgi>).

D'autre part, la localisation des sites de liaisons ou d'interactions entre les protéines ou entre les protéines et des ligands est d'une importance fondamentale pour la compréhension de la fonction biochimique des protéines et la conception des médicaments. Dans cette optique, nous avons effectué des amarrages entre le modèle de la structure de la XPC prédite avec I-TASSER et les protéines connues en interaction avec la XPC, notamment la centrine 2 et la TFIID par le serveur ClusPro2.0. Les résultats montrent la fiabilité de ce serveur de prédire les bons

complexes XPC-centrine2 et XPC-TFIIH, et donc, la validation du modèle de structure de la XPC prédit.

En outre, une mutagenèse effectuée sur la structure la protéine XPC prédite par PyMol nous a permis de localiser les différents rs pathologique de la XPC. Les observations effectuées permettent la compréhension de l'impact de la mutation sur les propriétés de la protéine XPC.

Tous ces résultats suscitent de nouvelles interrogations, et ouvrent donc la voie à des futures études dans le volet bioinformatique et/ou expérimental.

La structure 3D de la protéine XPC permet la connaissance de leurs fonctions et de mieux comprendre les mécanismes physiopathologiques du XPC et peut être à long terme d'envisager des perspectives thérapeutiques. Dans l'avenir le diagnostic moléculaire de cette pathologie et la prise en charge des patients peut être faciliter par la compréhension du fonctionnement de la protéine XPC et sa formation de complexes.

Références bibliographiques

A

Araki M, Masutani C, Takemura M, Sugawara K, Kondoh J, Ohkuma Y, Hanaoka F, 2001, 'Centrosome Protein Centrin 2/Caltractin 1 Is Part of the Xeroderma Pigmentosum Group C Complex That Initiates Global Genome Nucleotide Excision Repair', *The journal of Biological Chemistry*, Vol 276, n°22, pp.18665-18672.

B

Baker, D., & Sali, a. (2001). Protein structure prediction and structural genomics. *Science (New York, N.Y.)*, 294(5540), 93–6. doi:10.1126/science.1065659

Bensenouci Salima., Louhibi Lotfi., Hubert De Verneuil., Mahmoudi Khadidja., Saidi-Mekhtar Nadhira., 2016, 'Diagnosis of xeroderma pigmentosum groups A and C by detection of two prevalent mutations in west Algerian population : A rapid genotyping tool for the frequent XPC mutation c.1649_1644delTG', *BioMed Research International*, vol2016, pp.1-7.

Bensenouci Salima, 2016, 'Etude moléculaire de la maladie xeroderma pigmentosum dans la région de l'Ouest Algerian', Thèse de Doctorat LMD, Université des Sciences et de la technologie d'Oran Mohammed boudiaf, Oran. pp.170.

Bhutto Abdul Manan and Kirk Sandra H, 2008, 'population distribution of xeroderma pigmentosum', *edit moléculaire mécanismes of xeroderma pigmentosum*, Pakistan, pp.138-143.

C

Camenisch Ulrike, Dip Ramiro, Sylvie Briand Schumacher, Schuter Benjamin, Naegeli Hanspeter, 2006, 'Recognition of helical kinks by xeroderma pigmentosum group A protein Triggers DNA excision repair', *nature structural and molecular biology*, vol 13, n°3, pp.278-284.

Coin Frederic, Oksenysh Valentyn, Egly Jean-Marc, 2007, 'Distinct Roles for the XPB/p52 and XPD/p44 Subcomplexes of TFIIH in Damaged DNA Opening during Nucleotide Excision Repair', *Molecular Cell*, Vol 26, pp.246-256.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., & Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19), 5179–5197.
doi:10.1021/ja00124a002

E

Encyclopédie Orphanet Grand Public, [www.orpha.net/data/patho/Pub/fr/Xeroderma Pigmentosum -FRfrPub3253v01.pdf](http://www.orpha.net/data/patho/Pub/fr/Xeroderma_Pigmentosum-FRfrPub3253v01.pdf) | Juin 2008.

F

FASSIHI, Dr Hiva. Xeroderma pigmentosum. Orpha.net. [En ligne] Mai 2011. [Citation : 18 Novembre 2012.] <http://www.orpha.net> .

G

GeneCards : The Human Gene Database. <http://www.genecards.org/> .

Gillet Ludovic C J and Scharer Oriando D,2006,'Molécular Mechanisms of Mammalian Global Genome Nucleotide Excision Repair',chemical reviews, Vol 106, n°2,pp.253-276.

H

Hanawalt Philipe C, 2000,'The bases for Cockayne syndrome', Nature, vol 405, 415-416.

Hirai Yuku, Kodama Yoshiaki, Moriwaki Shin-Ichi, Noda Asao,Cullings Harry M,MacPhee Donald G,Kodama Kazunori, Mabuchi Kiyohiko,Kreamer Kenneth H,Land Charles E,Nakamura Nori,2006,'Heterozygous individuals bearing a founder mutation in the XPA DNA repair gene comprise nearly 1% of the Japanese population',Fundamental and molecular mechanisms of mutagenesis, vol601,pp.171-178.

K

Kanda T, Oda M, Yonezawa M, Tamagawa K, Isa F, Hanakago R, Tsukagoshi H,1990, 'Peripheral neuropathy in xeroderma pigmentosum', journal of Neurology.

Kelley Lawrence A,Mezulis Stefans, Yates Christopher M, Wass Mark N, Sternberg Michael J E,2015,'The Phyre web portal for protein modeling prediction and analysis',Nature protocols, vol 10, n°6,pp 845-858.

Khemili Souad, 2013, 'Prédiction des épitopes B de l'allergène Der p 5 et étude in silico de la structure et de l'état oligomérique des allergènes d'acariens des familles 5 et 21', Thèse de Doctorat universite m'hamed bougara de boumerdes, pp.189.

Kleijer Win J,Laugel Vincent,Berneburg Mark,Nadro Tiziana, Fawcett Heather,Gratchev Alexei,Jaspers Nicolas G J,Sarasin Alain,Stefanini Miria, Lehmann Alan R,2008,'Incidence of DNA repair defeciency disorders in western Europe : xeroderma pigmentosum Cockayne syndrome and trichothiodystrophy,DNA repair 7,pp.744-750.

Kraemer Kenneth, DiGiovanna John J,2016,'Xeroderma Pigmentosum'GeneReviews,[En ligne], <https://www.ncbi.nlm.nih.gov/books/NBK1397/> .

L

Larrea Andres A, Lujan Scott A, Kurkel Thomas A,2010,'DNA Mismatch Repair',cell 141,pp.730-731.

O

Oh Kyu-Seon, Imoto Kyoko, Emmert Steffen, Tamura Deborah, DiGivanna Jhon J, Kreamer Kenneth H,2001,'Nucleaotide Excision Repair Proteins Rapidly Accumulate but fail to Persist in Human XP-E(DDB2 Mutant) Cells,Photochemistry and photobiology,vol 87,729-733.

Références bibliographiques

Orelli Barbara, McCledon Brooke, Tsodikov Oleg V, Ellenberger Tom, Niedrnhofer Laura J, Scha Orlando, 2010, 'The XPA-binding domain of ERCC1 is Required for Nucleotide Excision Repair but Not Other DNA Repair Pathways', *Biological Chemistry*, Vol 285, n°6, pp.3705-3712.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T, 1989, 'Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms', *Proceedings of the National Academy*, Vol 86, pp.2766-2770.

Ozoux Barbara, 2013, 'xeroderma pigmentosum, Actualisation des connaissances sur la maladie et les traitements associés', Thèse de Doctorat en pharmacie, Université Toulouse III Paul Sabatier, Toulouse, pp.104.

P

Panier Stephanie and Boulton Simon J, 2014, 'Double strand break repair : 53BP1 comes into focus', *nature reviews molecular cell biology*, Vol 15, pp.7-18.

Q

QIAGEN un fournisseur mondial d'échantillons moléculaires et de technologies de dosage.
<https://www.qiagen.com/dz/>

R

Riedl Thilo, Hanaoka Fumio and Egly Jean-Marc, 2003, 'The comings and goings of nucleotide excision repair factors on damaged DNA', *The EMBO journal*, Vol 12, n°19, pp.5293-5303.

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein engineering*, 12(2), 85–94. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10195279>

S

Sali, A., & Blundell, T. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Protein Structure by Distance Analysis*, 234, 779–815. Retrieved from <http://books.google.com/books?hl=en&lr=&id=LKxFL1ksNH8C&oi=fnd&pg=PA64&dq=Comparative+Protein+Modelling+by+Satisfaction+of+Spatial+Restraints&ots=gd7>

RSC5Vwf&sig=Z42X6_bvfobz_hnfLleJFBJoUd0

Z

Ziani Salim,2014,'etude du complexe de réparation par excision de nucléotide,Thèse de Doctorat,Université de Strasbourg,pp.87.

Annexes

Annexe I : démarche Génome Browser.

La page d'accueil Génome Browser.

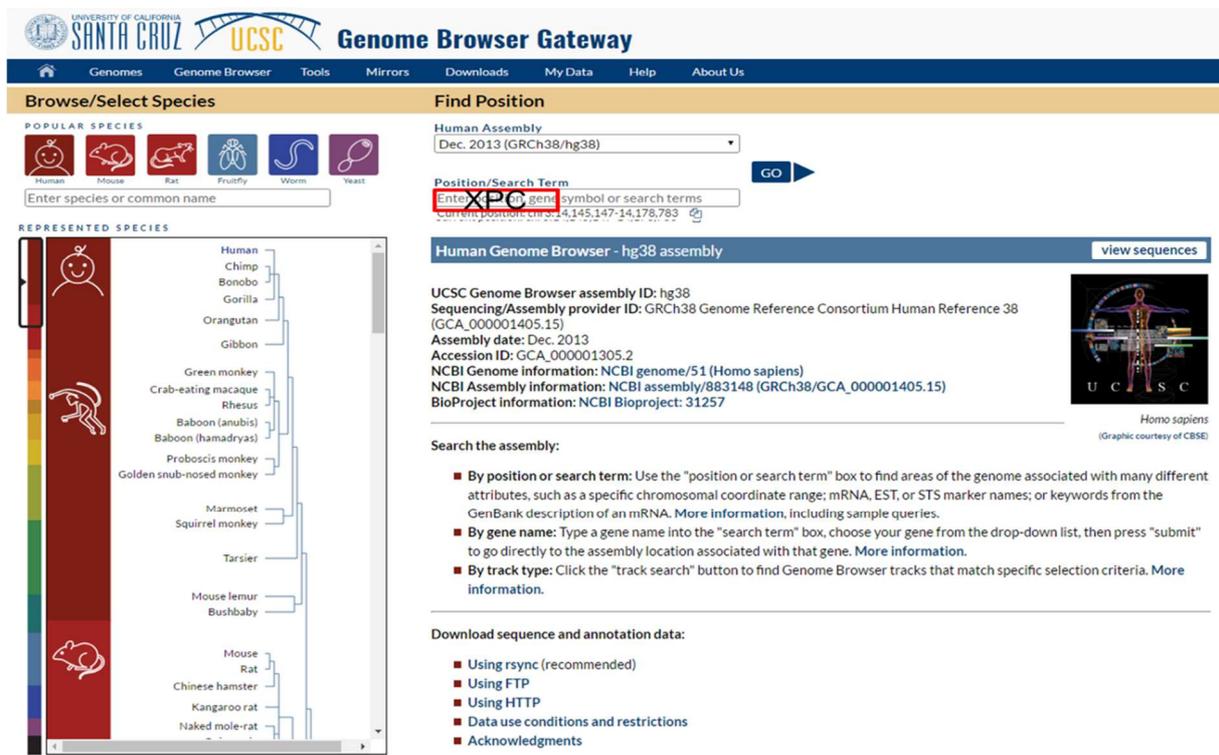


Our tools

- **Genome Browser**
interactively visualize genomic data
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Gene Sorter**
find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBiB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **VisiGene**
interactively view in situ images of mouse and frog

[More tools...](#)

Taper le nom du gène : XPC.



Find Position

Human Assembly
Dec. 2013 (GRCh38/hg38)

Position/Search Term
Enter **XPC** gene symbol or search terms
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=4,145,147-14,178,783

Human Genome Browser - hg38 assembly [view sequences](#)

UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: GRCh38 Genome Reference Consortium Human Reference 38 (GCA_000001405.15)
Assembly date: Dec. 2013
Accession ID: GCA_000001305.2
NCBI Genome information: NCBI genome/51 (Homo sapiens)
NCBI Assembly information: NCBI assembly/883148 (GRCh38/GCA_000001405.15)
BioProject information: NCBI Bioproject: 31257

Search the assembly:

- **By position or search term:** Use the "position or search term" box to find areas of the genome associated with many different attributes, such as a specific chromosomal coordinate range; mRNA, EST, or STS marker names; or keywords from the GenBank description of an mRNA. [More information](#), including sample queries.
- **By gene name:** Type a gene name into the "search term" box, choose your gene from the drop-down list, then press "submit" to go directly to the assembly location associated with that gene. [More information](#).
- **By track type:** Click the "track search" button to find Genome Browser tracks that match specific selection criteria. [More information](#).

Download sequence and annotation data:

- Using rsync (recommended)
- Using FTP
- Using HTTP
- Data use conditions and restrictions
- Acknowledgments

Après avoir les résultats des gènes connus.

Known Genes

[XPC \(uc011ave.3\) at chr3:14145147-14178781](#) - Homo sapiens xeroderma pigmentosum, complementation group C (XPC), transcript variant 1, mRNA. (from RefSeq NM_004628)

[XPC \(uc061gen.1\) at chr3:14168271-14178641](#) - The sequence shown here is derived from an Ensembl automatic analysis pipeline and should be considered as preliminary data. (from UniProt)

[XPC \(uc061gzn.1\) at chr3:14164907-14170614](#) - xeroderma pigmentosum, complementation group C (from HGNC XPC)

[XPC \(uc061gzn.1\) at chr3:14164366-14170469](#) - xeroderma pigmentosum, complementation group C (from HGNC XPC)

[XPC \(uc061gzk.1\) at chr3:14158807-14167267](#) - xeroderma pigmentosum, complementation group C (from HGNC XPC)

[XPC \(uc061gze.1\) at chr3:14147467-14148928](#) - xeroderma pigmentosum, complementation group C (from HGNC XPC)

[XPC \(uc061gzd.1\) at chr3:14145729-14178801](#) - Involved in global genome nucleotide excision repair (GG-NER) by acting as damage sensing and DNA-binding factor component of the XPC comp

[RAD23A \(uc002mw.3\) at chr19:12945888-12953642](#) - Homo sapiens RAD23 homolog A, nucleotide excision repair protein (RAD23A), transcript variant 1, mRNA. (from RefSeq NM_005053)

[RAD23A \(uc031tp.2\) at chr19:12945892-12953591](#) - Homo sapiens RAD23 homolog A, nucleotide excision repair protein (RAD23A), transcript variant 3, mRNA. (from RefSeq NM_001270363)

[RAD23A \(uc002mv.3\) at chr19:12945876-12953623](#) - Homo sapiens RAD23 homolog A, nucleotide excision repair protein (RAD23A), transcript variant 2, mRNA. (from RefSeq NM_001270362)

[CUL4A \(uc059ymh.1\) at chr13:113208836-113267108](#) - Core component of multiple cullin-RING-based E3 ubiquitin-protein ligase complexes which mediate the ubiquitination of target protein

[GTF2H1 \(uc001mh.3\) at chr11:18322596-18367044](#) - Homo sapiens general transcription factor IIH subunit 1 (GTF2H1), transcript variant 1, mRNA. (from RefSeq NM_005316)

[GTF2H1 \(uc001mi.3\) at chr11:18322295-18366261](#) - Homo sapiens general transcription factor IIH subunit 1 (GTF2H1), transcript variant 2, mRNA. (from RefSeq NM_001142307)

[GTF2H1 \(uc009hm.3\) at chr11:18322567-18366813](#) - Component of the core-TFIIB basal transcription factor involved in nucleotide excision repair (NER) of DNA and, when complexed to CAK, DDB2 (uc058bb.1) at chr11:47215137-47238849 - Required for DNA repair. Binds to DDB1 to form the UV-damaged DNA-binding protein complex (the UV-DDB complex). The UV-DDB complex may

[DDB2 \(uc058bb.1\) at chr11:47214992-47239189](#) - Required for DNA repair. Binds to DDB1 to form the UV-damaged DNA-binding protein complex (the UV-DDB complex). The UV-DDB complex may

[DDB2 \(uc009li.2\) at chr11:47215137-47238849](#) - Required for DNA repair. Binds to DDB1 to form the UV-damaged DNA-binding protein complex (the UV-DDB complex). The UV-DDB complex may

[CUL4A \(uc021rmv.3\) at chr13:113209544-113263582](#) - Homo sapiens cullin 4A (CUL4A), transcript variant 1, mRNA. (from RefSeq NM_001008895)

[CUL4A \(uc032apo.2\) at chr13:113208193-113265072](#) - Homo sapiens cullin 4A (CUL4A), transcript variant 3, mRNA. (from RefSeq NM_001278513)

[CUL4A \(uc021rmv.2\) at chr13:113208772-113265077](#) - Homo sapiens cullin 4A (CUL4A), transcript variant 2, mRNA. (from RefSeq NM_003589)

[CETN2 \(uc004fgg.4\) at chrX:152826973-152830777](#) - Homo sapiens centrin 2 (CETN2), mRNA. (from RefSeq NM_004344)

[ERCC3 \(uc002toh.1\) at chr2:127257208-127294176](#) - Homo sapiens excision repair cross-complementation group 3 (ERCC3), transcript variant 3, mRNA. (from RefSeq NM_001303418)

[RAD23B \(uc004bde.4\) at chr9:107283279-107332194](#) - Homo sapiens RAD23 homolog B, nucleotide excision repair protein (RAD23B), transcript variant 1, mRNA. (from RefSeq NM_002874)

[RAD23B \(uc022b1f.2\) at chr9:107284053-107332192](#) - Homo sapiens RAD23 homolog B, nucleotide excision repair protein (RAD23B), transcript variant 3, mRNA. (from RefSeq NM_001244724)

[DDB2 \(uc001neb.3\) at chr11:47214942-47239240](#) - Homo sapiens damage specific DNA binding protein 2 (DDB2), transcript variant 1T, mRNA. (from RefSeq NM_000107)

[DDB2 \(uc001nee.3\) at chr11:47215137-47238849](#) - Homo sapiens damage specific DNA binding protein 2 (DDB2), transcript variant D1, mRNA. (from RefSeq NM_001300734)

[DDB1 \(uc001nrc.6\) at chr11:61299451-61333366](#) - Homo sapiens damage specific DNA binding protein 1 (DDB1), mRNA. (from RefSeq NM_001923)

[SIRT1 \(uc001jnd.3\) at chr10:67884669-67918390](#) - Homo sapiens sirtuin 1 (SIRT1), transcript variant 1, mRNA. (from RefSeq NM_012238)

[HWGB1 \(uc058udx.1\) at chr13:30456097-30465242](#) - DNA binding proteins that associates with chromatin and has the ability to bend DNA. Binds preferentially single-stranded DNA. Involved

[HWGB1 \(uc058udy.1\) at chr13:30456097-30465242](#) - DNA binding proteins that associates with chromatin and has the ability to bend DNA. Binds preferentially single-stranded DNA. Involved

[HWGB1 \(uc058udy.1\) at chr13:30456098-30465310](#) - DNA binding proteins that associates with chromatin and has the ability to bend DNA. Binds preferentially single-stranded DNA. Involved

[HWGB1 \(uc001uxv.5\) at chr13:30456204-30465365](#) - Homo sapiens high mobility group box 1 (HWGB1), transcript variant 2, mRNA. (from RefSeq NM_002128)

[SLC16A2 \(uc031tjv.2\) at chrX:74421461-74533917](#) - Homo sapiens solute carrier family 16 member 2 (SLC16A2), mRNA. (from RefSeq NM_006517)

En précisant ce que nous voulons avoir.

Mapping and Sequencing						refresh
Base Position	Alt Map...	Assembly	Centromeres	Chromosome Band	Clone Ends	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
FISH Clones	Gap	GC Percent	GRC Contigs	GRC Incident	GRC Patch Release	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Hg19 Diff	INSDC	LRG Regions	Restr Enzymes	Scaffolds	Short Match	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
STS Markers						
hide ▼						
Genes and Gene Predictions						refresh
GENCODE v24	NCBI RefSeq	All GENCODE...	AUGUSTUS	CCDS	CRISPR...	
full ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Geneid Genes	Genscan Genes	IKMC Genes Mapped	LRG Transcripts	MGC Genes	Non-coding RNA...	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Old UCSC Genes	ORFeome Clones	Other RefSeq	Pfam in UCSC Gene	RetroGenes V9	SGP Genes	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
SIB Genes	TransMap...	UCSC Alt Events	UniProt			
hide ▼	hide ▼	hide ▼	hide ▼			
Phenotype and Literature						refresh
OMIM Alleles	ClinGen CNVs	ClinVar Variants	Coriell CNVs	COSMIC Regions	Development Delay	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Gene Interactions	GeneReviews	GWAS Catalog	OMIM Genes	OMIM Pheno Loci	UniProt Variants	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
mRNA and EST						refresh
Human ESTs	Human mRNAs	Other ESTs	Other mRNAs	SIB Alt-Splicing	Spliced ESTs	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Expression						refresh
GTEX	Affy GNF1H	Affy U133	Affy U95	GNF Atlas 2	GWIPS-viz Riboseq	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Regulation						refresh
ENCODE Regulation...	CpG Islands...	OREgAnno				
hide ▼	hide ▼	hide ▼				
Comparative Genomics						refresh
Conservation	Cons 7 Verts	Cons 20 Mammals	Primate Chain/Net	Placental Chain/Net	Vertebrate Chain/Net	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Variation						refresh
Common SNPs(147)	All SNPs(147)	All SNPs(146)	All SNPs(144)	All SNPs(142)	All SNPs(141)	
pack ▼	pack ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Common SNPs(146)	Common SNPs(144)	Common SNPs(142)	Common SNPs(141)	DGV Struct Var	Flagged SNPs(147)	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	pack ▼	
Flagged SNPs(146)	Flagged SNPs(144)	Flagged SNPs(142)	Flagged SNPs(141)	Mult. SNPs(147)	Mult. SNPs(146)	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Mult. SNPs(144)	Mult. SNPs(142)	Mult. SNPs(141)				
hide ▼	hide ▼	hide ▼				
Repeats						refresh
RepeatMasker	Interrupted Rpts	Microsatellite	RepeatMasker Viz.	Segmental Dups	Self Chain	
hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	hide ▼	
Simple Repeats	WM + SDust					
hide ▼	hide ▼					

Annexe II : démarche Table Browser.

Page d'accueil Table Browser et la sélection des SNPFlagged.

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: genome: assembly:

group: track:

table:

region: genome position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: Send output to [Galaxy](#) [GREAT](#) [GenomeSpace](#)

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Select Fields from hg38.snp147Common

<input type="checkbox"/>	bin	
<input type="checkbox"/>	chrom	Reference sequence chromosome or scaffold
<input type="checkbox"/>	chromStart	Start position in chrom
<input type="checkbox"/>	chromEnd	End position in chrom
<input checked="" type="checkbox"/>	name	dbSNP Reference SNP (rs) identifier
<input type="checkbox"/>	score	Not used
<input type="checkbox"/>	strand	Which DNA strand contains the observed alleles
<input type="checkbox"/>	refNCBI	Reference genomic sequence from dbSNP
<input type="checkbox"/>	refUCSC	Reference genomic sequence from UCSC lookup of chrom,chromStart,chromEnd
<input type="checkbox"/>	observed	The sequences of the observed alleles from rs-fasta files
<input type="checkbox"/>	molType	Sample type from exemplar submitted SNPs (ss)
<input type="checkbox"/>	class	Class of variant (single, in-del, named, mixed, etc.)
<input type="checkbox"/>	valid	Validation status of the SNP
<input type="checkbox"/>	avHet	Average heterozygosity from all observations. Note: may be computed on small number of samples.
<input type="checkbox"/>	avHetSE	Standard Error for the average heterozygosity
<input type="checkbox"/>	func	Functional category of the SNP (coding-synon, coding-nonsynon, intron, etc.)
<input type="checkbox"/>	locType	Type of mapping inferred from size on reference; may not agree with class
<input type="checkbox"/>	weight	The quality of the alignment: 1 = unique mapping, 2 = non-unique, 3 = many matches
<input type="checkbox"/>	exceptions	Unusual conditions noted by UCSC that may indicate a problem with the data
<input type="checkbox"/>	submitterCount	Number of distinct submitter handles for submitted SNPs for this ref SNP
<input type="checkbox"/>	submitters	List of submitter handles
<input type="checkbox"/>	alleleFreqCount	Number of observed alleles with frequency data
<input type="checkbox"/>	alleles	Observed alleles for which frequency data are available
<input type="checkbox"/>	alleleNs	Count of chromosomes (2N) on which each allele was observed. Note: this is extrapolated by dbSNP from submitted frequencies and total sample 2N, and is not always an integer.
<input type="checkbox"/>	alleleFreqs	Allele frequencies
<input type="checkbox"/>	bitfields	SNP attributes extracted from dbSNP's SNP_bitfield table

La listes des rs obtenues.

rs121965092
rs3731177
rs183167499
rs375859472
rs55779831
rs200148127
rs786205206
rs754673606
rs3731152
rs869025275
rs794729655
rs587778758
rs144766677
rs183238369
rs121965088
rs767569346
rs754532049
rs3731130
rs794729654
rs587778762
rs121965090
rs192285219
rs74737358
rs200338014
rs121965091
rs184879571
rs35629274
rs587778760
rs587778761
rs552222088
rs587778759
rs752088918
rs587778763
rs794729656
rs794729657
rs2228002
rs587778757
rs201273381

Paste In Identifiers for Common SNPs(147)

Please paste in the identifiers you want to include. The items must be values of the **name** field of the currently selected table, **snp147Common**. (The "describe table schema" button shows more information about the table fields.) Some example values:

rs10027781
rs10020600
rs10015607
rs10013607
rs10077230

Copier coller les rs

submit clear cancel

Table Browser (Input Identifiers)

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal **genome:** Human **assembly:** Dec. 2013 (GRCh38/hg38)
group: Variation **track:** Flagged SNPs(147)
table: snp147Flagged
region: genome position
identifiers (names/accessions):
filter:
intersection:
correlation:
output format: custom track Send output to [Galaxy](#) [GREAT](#) [GenomeSpace](#)
output file: (leave blank to keep output in browser)
file type returned: plain text gzip compressed

Pour avoir les rs annotés dans Génome Browser.

Output snp147Flagged as Custom Track

[Custom track header:](#)

name=
 description=
 visibility=
 url=

Create one BED record per:

- Whole Gene
- Upstream by bases
- Downstream by bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Annexe III : démarche Blastp.

Page d'accueil Blast.

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

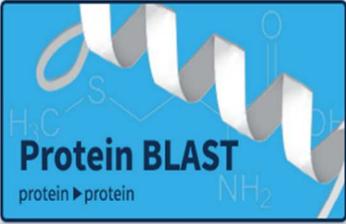
QuickBLASTP
Try **QuickBLASTP** for a fast protein search of nr.
Tue, 23 May 2017 13:00:00 EST [More BLAST news...](#)

Web BLAST



Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein



Protein BLAST
protein ▶ protein

BLAST Genomes

[Human](#) [Mouse](#) [Rat](#) [Microbes](#)

Copier-coller la séquence protéique afin de modifier les paramètres.

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) [Query subrange](#)

Copier coller la séquence protéique

From:
To:

Or, upload file: Aucun fichier choisi

Job Title:
Enter a descriptive title for your BLAST search

[Align two or more sequences](#)

Choose Search Set

Database: Protein Data Bank proteins(pdb)

Organism: [Exclude](#)
Enter organism name or id—completions will be suggested. Only 20 top taxa will be shown.

Exclude: Models (XM/XP) Uncultured/environmental sample sequences

Entrez Query: [You Tube](#) [Create custom database](#)
Enter an Entrez query to limit search

Program Selection

Algorithm:

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

BLAST Search database Protein Data Bank proteins(pdb) using Blastp (protein-protein BLAST)

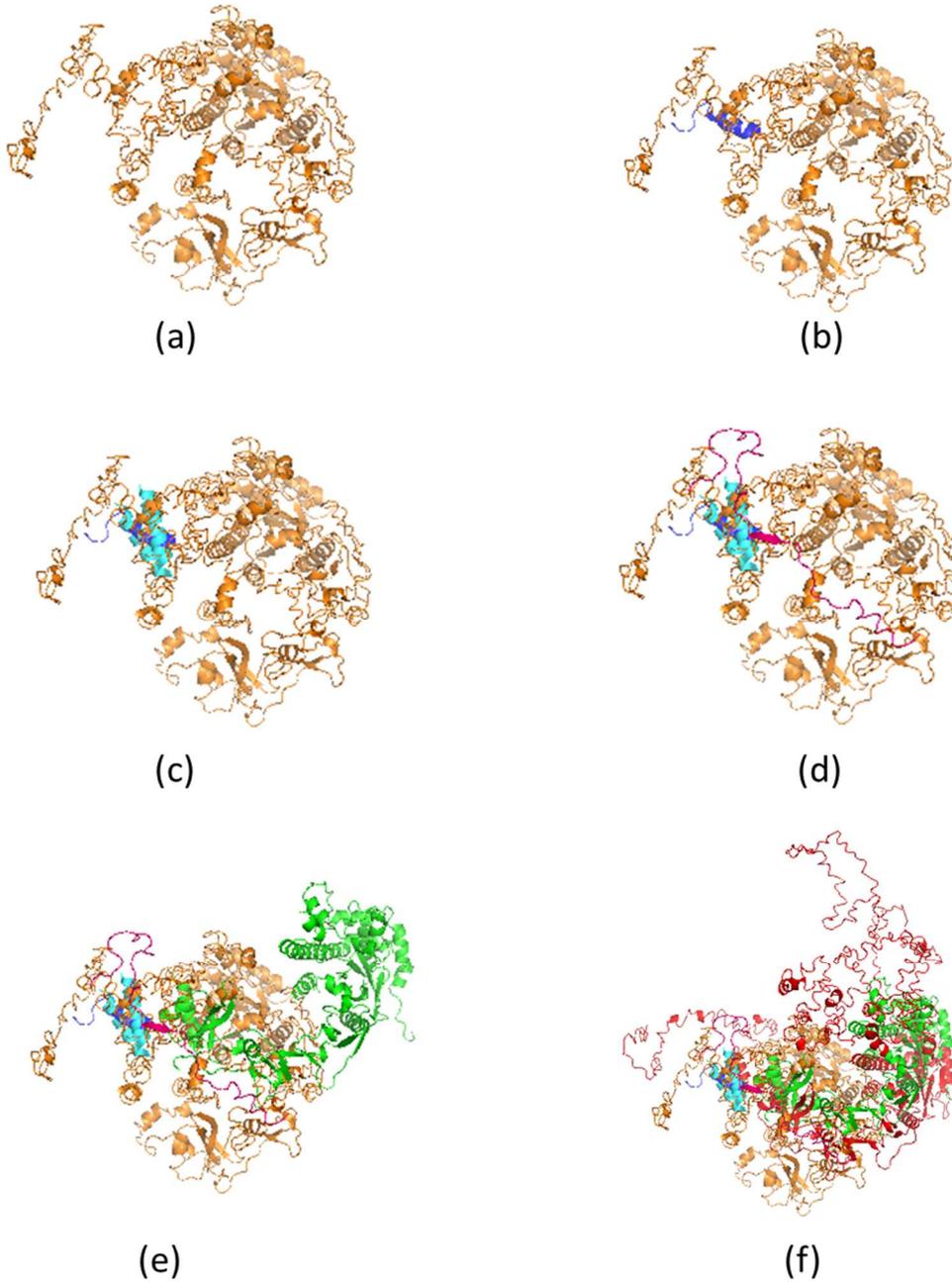
Show results in a new window

Résultats de SwissModel : XPC-SwiwwModell.

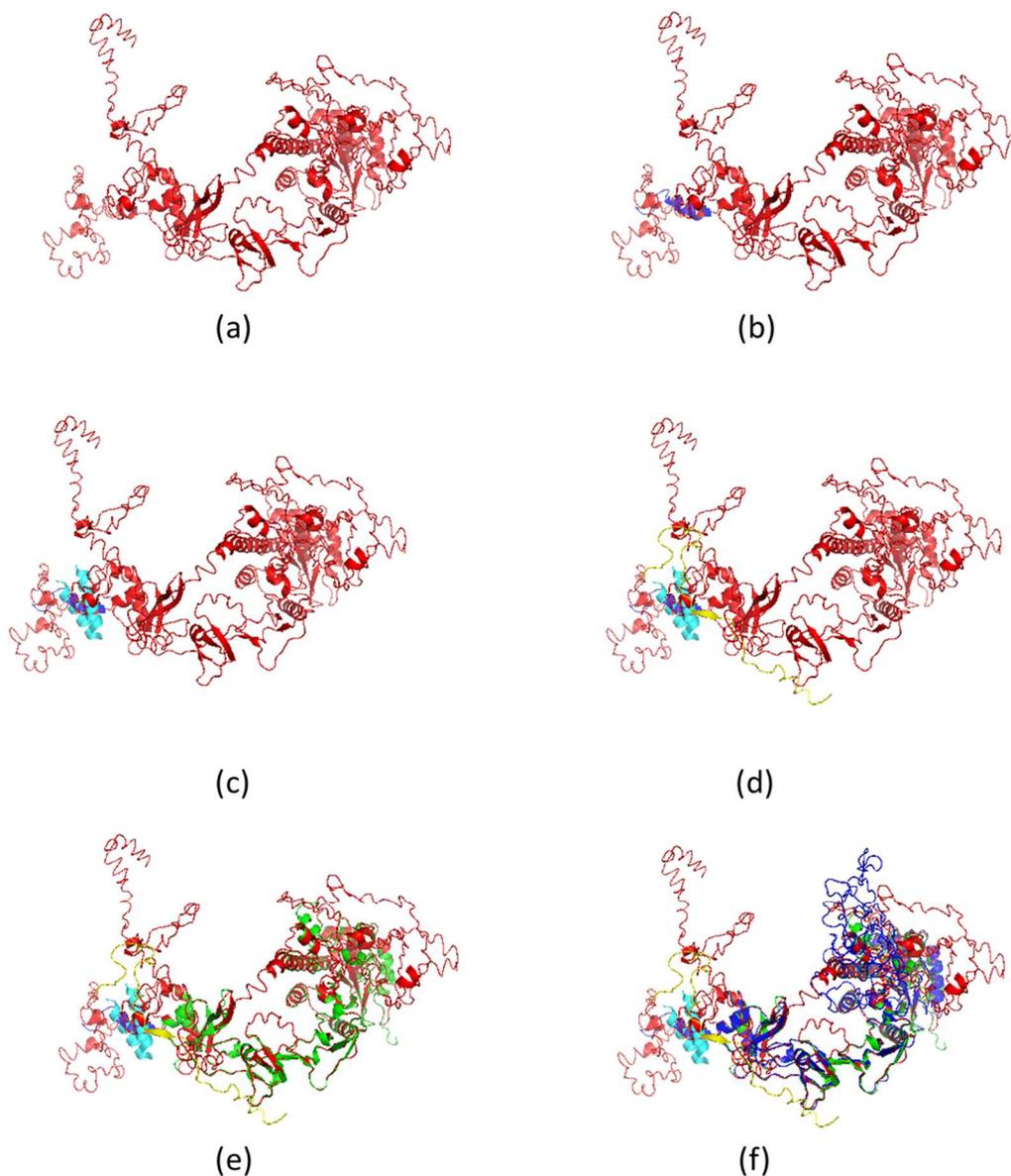


Annexe IV : alignement pymol des structures prédites de I-TASSER avec les autres structures.

L'alignement est illustré dans la Figure la structure de XPC-modèl1-ITASSER (a) est alignés avec des fragments de XPC P1-2a4j (b), P2c-P2d-2obh (c) et p3-2rvb (d) et ensuite avec XPC-Phyre2 et enfin la structure de XPC-SwissModel1



L'alignement est illustré dans la Figure la structure de XPC-modèle2-ITASSER (a) est alignés avec des fragments de XPC P1-2a4j (b), P2c-P2d-2obh (c) et p3-2rvb (d) et ensuite avec XPC-Phyre2 et enfin la structure de XPC-SwissModel1



Annexe V : Tableau des rs associés à la maladie XPC ainsi leurs détails

La première colonne représente les rs, la deuxième colonne représente leur localisation dans l'exon ensuite les trois colonnes qui se suit représentent l'allèle et la position protéique de chaque aa et la dernière colonne représente la s séquences fasta de chaque rs.

rs	Exn	Allèle change	Position protéique	Résiduel change	séquence fasta
rs587778757	01	AGC ⇒ CGC	18	S [Ser] ⇒ R [Arg]	>gnl dbSNP rs587778757 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='A/C' mol=Genomic build=142 TGGCTCGGAA ACGCGCGGCC GCGGGGAGC GCGGGGACG CGAACTGCGC M GCCAGAAATC CAAGGCCAAG AGCAAGGCC GCGGTGAGGA GGAGGAGGAG
rs201273381	01	GGA ⇒ CGA	13	G [Gly] ⇒ R [Arg]	>gnl dbSNP rs201273381 allelePos=201 totalLen=401 taxid=9606 snpclass=1 alleles='C/G' mol=Genomic build=150 CGGGCTGCC CCACGGCGCG GAACGCGCG AGCAACTCC ACCAGGCCTC CGCTGTGGA CTCCCGCCT GCCTCTGGGC CTCTCCGCC CACCGCGGC GTCTCCGGG AAGCCCGCTG GGCTCGCTC TCACCTCTC CTCTCTCTC ACGCCGGGCC TTGCTCTGG CCTTGGATT CTGGCTGCG AGTTCGCGT S CCGCGGTCC CCGCGCGCG CGCGTTCG AGCCAATGTT CTGTCTGG CAAATCCAC TTCGCGAGT ACGACCCCG CCGCATGCG CTAGAACGCC GGCCCCCG CGAGCGGTG CGTTGGCA CGCCCTTCG TTGGAGGCT AGTACGCC CTAAGGAAG CTCACCCCA TACCCGTGGC CACTAAAAA
rs2228002	03	AGG ⇒ AGA	100	R [Arg] ⇒ R [Arg]	>gnl dbSNP rs2228002 allelePos=256 totalLen=511 taxid=9606 snpclass=1 alleles='A/G' mol=Genomic build=150 TCITTCACAC GATCTAGATT GTAGTTTT TAACAGACA TGTGTAGTA ATTAGACAGA TGATCAGAGC ACATCTTAA GAGATTGGA ACATGTTG ATGGAGGAAG TGAGGCTCAG AGCTTACTAA TCATCTTAA ATAGTTGCA CAGATTCTA ATAAATAGGA GGGGTTTTAA AAATTCATT TAGCTGAAT GGAACACTAG GATGCTGAG TCTTCTCAT CTCCTCTC GTGGTCTTA CACAG R GACTTCCAA GTGACCTCAA GAAGGCACAC CATCTGAAGA GAGGGCTAC CATGAATGA GACAGCAATG AAGAAGAGGA AGAAAGTGA AATGATTGG AAGAGGTGA AGGTGAAACA TCTTCTTTG TTTACAGAAC TGTTGGTTC TGTTGTTTT TTGTTTGA GTCAGATCAC TAATTGCAAT CCTTTGTT TCTTTGAG TAATTAGACC ACAGATAAGG TTGACTAGG GACTT
rs587778763	04	GIG ⇒ ITG	160	V [Val] ⇒ L [Leu]	>gnl dbSNP rs587778763 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='G/T' mol=Genomic build=144 TGCTGGGTGA CGTGAGAGAA AGTACAGCCT TCTCTGATC TCTCTGCCT K TGAAGCCAGT GGAGATAGAG ATTGAAACGC CAGAGCAGC GAAGACAAGA
rs587778759	05	AAA ⇒ AAC	199	K [Lys] ⇒ N [Asn]	>gnl dbSNP rs587778759 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='A/C' mol=Genomic build=150 AAACTGGAGT TTGAGACATA TCTTCGGAG GCGATGAAAC GTTCAATAA M GGGGTCCATG AGGACACACA CAAGTAAGG GCAAGGAATG ATGGGGAAGG
rs587778761	06	CCT ⇒ TCT	246	P [Pro] ⇒ S [Ser]	>gnl dbSNP rs587778761 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='C/T' mol=Genomic build=150 TGCAIGCTAT TGGCCTGTC ATCAATCCAG CCGCTTAC CAGAGTGTG Y CTCGAGATG GGACACCTAC TACCTCTCAA ACCTGGTGA GIGGTAAGGC
rs587778760	06	TAC ⇒ TGC	252	Y [Tyr] ⇒ C [Cys]	>gnl dbSNP rs587778760 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='A/G' mol=Genomic build=150 CATCATCCA GCCCGTTA CCAGAGTGCT GCCTCGAGAT GTGGACCT R CTACCTCTCA AACCTGGTA AGTGGTAAG CCCTCCGCTT GTCCTGAGA
rs55222088	06	CGC ⇒ TGC	240	R [Arg] ⇒ C [Cys]	>gnl dbSNP rs55222088 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='A/G' mol=Genomic build=150 GTTTGAGAG TAGTAGGTG CCACATCTCG AGGCAGCACT CTGGTAAAGC R GGCTGGGATG ATGGACAGC CAATAGCATG CAGATCTGG TGGCTGAGA
rs184879571	07	TCT ⇒ TGT TCT ⇒ TTT	291	S [Ser] ⇒ C [Cys]	>gnl dbSNP rs184879571 allelePos=201 totalLen=401 taxid=9606 snpclass=1 alleles='A/C/G' mol=Genomic build=150 AGACCTTGA GGCATCATGT AGCTATTCTG GGCTCAGCA GCTATCACG TCATTATGAG GACTGAATAA GGTAGTGTG GTAACACACC TGGAAATGCA TCTGGCATT GGCTGCCATT ATCATTGTT AACAGTACTG ATAAAAACA GTGATCCGGG AGGATCACTT ACATGGACCA ATTCTCATC ATCTCGAGCA V AGTAAATAGC AAATCTCCTT TCCAATGTAG TCTGAGGTT ATCTGTTC CTGGTGAAA GTTCTGATT AACTGTAAT GTTCAATGA ACCTGGGAG AAGCAGGCA TTCCTGTGT CAGAGTCA GGCAGAGGG AATTTTATT TCCAGCAAG AAAATAAAA GAGGAGTGA ATCTGAGAC CCGCTGCT
rs35629274	07	TTT ⇒ TGT	287	F [Phe] ⇒ C [Cys]	>gnl dbSNP rs35629274 allelePos=301 totalLen=601 taxid=9606 snpclass=1 alleles='A/C' mol=Genomic build=150 CCACAAC TAG AGCGATAGG GAGGAAAGCA TACAGGCCCT CCACTCAGAC TCTGATCCCA ATCTCGAATC TCAGGCTCAC CAAGGAAAG ACCTTAGAGG CATCATGAG CTATTCTGG CTTCAGCAGC TATCAACGTC ATTATGAGGA CTGAATAAGG TAGTGTGGT AACACACTG GAATGGCATC TGGACATAGG CTGCCATTAT CATTAGTAA CAGTACTGAT AAAAAACAGT GATCCGGAG GATCACTTAC ATGGACCAAT TCCTCATCAT CTCGAGCAGA GTAATAGCA M ATCTCCTTC CAATGATGTC TGCAGTTT CTGTTCAT GGCTGAAAGT TCTGCAATA CTGTAATGT TCCAATGAA CTGGGAGAA AGCAGGCATT CCTGTGTCA GAGTCAAGG CAAAGGGGAA TTTTCAATTC CAGCCAAGAA AATAAAAAGA GGGAGTGAAT CGTGAGACCC GCCTGCCTC GTCCTACTT CCCCAGCCG TCTAGCTAAC TCCTATCAT ACTGCAGACC CAGCTCAGAC TTTACTCTC TGGGGAGCCT TCCAGACCT CCGAGTGA AAAAAAGCC

rs	Exon	Allèle change	Position protéique	Résiduel change	séquence fasta
rs192285219	09	GCG ⇒ GTG	341	A [Ala] ⇒ V [Val]	>gnl dbSNP rs192285219 allelePos=201 totalLen=401 taxid=9606 snpclass=1 alleles='A/C/G/T' mol=Genomic build=150 TGCTCTCTGG GCCCTCATCT TCCTCGCTGG AGGAGGGCTT GCTCCGTTTC TTCTGCCTC CCTGTCTCT CTCCCTTTG GCACTTGGCC TGCAGGTGCC CTTAGCAAAG GTTCTCTCT GTTGGTTCC TTGCTGTGC TTGGTTTGG TGTGGTTTC TAGAACTTGG CTGGAAGTTT CTGAGGAGCC TCCTGGATCC V CAGTCAATCT TCCTTGGAA GGTTCCTTC CCTTAAACAG AATAAGAAAT TTGCTTTTT TTCTCCCC CTCTTTTCT AATGATAGA TAGAAATCT GTAATCAAT AGGGTTAAGT CACCAGTAG AGAACCAATA CATCAAGATG TCCCGAGCTC AGACAGTGT CAACCCAGTA TCCTCAGGG TTGATGAT
rs74737358	09	CCT ⇒ CAT CCT ⇒ CGT CCT ⇒ CTT	334	P [Pro] ⇒ H [His] P [Pro] ⇒ R [Arg] P [Pro] ⇒ L [Leu]	>gnl dbSNP rs74737358 allelePos=251 totalLen=501 taxid=9606 snpclass=1 alleles='A/C/G/T' mol=Genomic build=150 CGGACGTCGC TGGGTTCCT TCCTCTGCTT GTCTCTGGG CCTCATCTT CCTCGCTGGA GGAGGGCTTG CTCCGTTCT TCCTGCCTCC CTGTCTCTC TTCCTTTGG CACTTGGCT GCAGGTGCC ITAGCAAAGG TTCTCTCTG TTGGTTCTT TTGGTTCTT TTGGTTTGT GTGGTTTCT AGAACTTGGC TGGGAAGTTC TGAGGAGCCT CTGGATCCG CAGTCAATCT TTCTTGGAA N GTTTCTTCC CTTAAACAGA ATAAGAAAT TTGCTTTTT TTCTCCCCC TCCTTTGCTA ATGATATG AGAAATCTG TAATCAATA GGGTAAAGT ACCAGCTAGA GAACCAATC ATCAAAGAT CCCAGCTCA GACAGTATC AACCAAGCTAT CCTCAGGGT TGTATGATG TATTGGCTAA AATCTGTAGC ATCTAGATA AGGCCACAGT ATAATCTTT TCACCTAAC TGTAAGAGA
rs144766677	09	TTT ⇒ TCT	614	F [Phe] ⇒ S [Ser]	>gnl dbSNP rs144766677 allelePos=201 totalLen=401 taxid=9606 snpclass=1 alleles='A/G' mol=Genomic build=150 ACTAAAATA AGGCAGCAAC TGCCCGAGCT TTATATGGCT GTGACAATA AATGAGACAA CCCAATAAAA ACACCAACATAGTGTGGG CATATAAAG GTGCTCAAAA ACAGGAATAA TTTTAATAAC CTGACTGTGT CTGGAGGCC CTGGCAGCCA AGGCCTTACC TCCAAGTCTT CTTCTTCTC CTGTCCATA R ATGGGCTCTG GTATGTCTC AAGGTCTCG CCCACCCTC AGCATCAACC CGGCATTCG GGTCACTGT CATCCAGACT GGGTCTGACC TCTGTGTGAC ATCTCGGACC CAAGCTCTAC TGCAATGCC CACCACATAG GTCATGGCT TGGTGGCGTA CTGTAAACAG GTACAGGGCT GGCCACCAC ACCGTGACA
rs183238369	09	CGC ⇒ TGC	594	R [Arg] ⇒ C [Cys]	>gnl dbSNP rs183238369 allelePos=201 totalLen=401 taxid=9606 snpclass=1 alleles='A/G' mol=Genomic build=150 CCATTAATAA CACCAACAT AGTGTGGG ATATATAAG TGCTCAAAA CAGGAATAA TTAATAACC TGACTGTGT TTGGAGCCC TGGCAGCCAA GGCCTTACT CAAGCTCTC TTTCTTCTC CTGCTCAAA ATGGGCTCTG GTATGTCTC AAGGTCTCG CCCACCCTC AGCATCAACC CGGCATTGC R GGTCACTGT ATCCAGACTG GGTCTACTC CTGTGTGACA TCTCGGACC AGCCGTCCT GTCAATGCC ACCACATAG TCATGGGCTT GGTGGCTAC TTGTAACAGG TCAGAGGCTG GCCACCACA CCGTGCACAC AGCTACACA TACCACCTT TCCTCTGCT CAGAGAACAC CTTAGCCAC TGGTCTATC
rs121965088	09	CGA ⇒ TGA	579	R [Arg] ⇒ Ter* [OPA]	>gnl dbSNP rs121965088 allelePos=251 totalLen=501 taxid=9606 snpclass=1 alleles='C/T' mol=Genomic build=150 AAGCTTGCCA GCGGCATCT CAAGCTCTC AAGCAGTAAA AGAGGCAAGA AATGTGCAG CGATGGTGG AAGGCAGAAA AAAGAAGCAT AGCTGGTATA GACCAGTGGC TAGAGGTGT CTGTGAGCAG GAGGAAAAGT GGGTATGTG AGACTGTGTG CACGGTGTGG TGCCGCCACC TCTGACCTGT TACAAGTACG CCACCAAGCC CATGACCTAT GTGTGGGGA TTGACAGTGA CGGCTGGGTC Y GAGATGTAC ACAGAGGTAC GACCCAGTCT GGATGACAGT GACCCGAAAG TGCCGGGTTG ATGCTGAGTG GTGGCCGAG ACCTTGAGAC CATAACAGAG CCCATTATG GACAGGAGTA AGAAAGAAGA CTGGAGGTA AGGCCTTGGC TGCCAGGCC TCCAAGACAC AGTCAGGTA TAAAATAT TCCTGTTTT GACACCTTA TATATGCCA GACTATGTT GGGTGTGTTT AATGGTGTG
rs767569346	09	TAC ⇒ TAA TAC ⇒ TAG TAC ⇒ TAT	559	Y [Tyr] ⇒ Ter* [OCH] Y [Tyr] ⇒ Ter* [AMB] Y [Tyr] ⇒ Y [Tyr]	>gnl dbSNP rs767569346 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='A/C/G/T' mol=Genomic build=150 CCGTCACTGT CAATGCCAC CACATAGGTC ATGGGCTTGG TGGCGTACTT N TAACAGGTCA GAGGCTGGCC CACCACACCG TGCACACAGT CTACACATAC
rs754532049	09	GTG ⇒ G	548	V [Val] ⇒ A [Ala]	>gnl dbSNP rs754532049 allelePos=51 totalLen=101 taxid=9606 snpclass=2 alleles='- /CA' mol=Genomic build=144 GGCTTGTGG CACTTGTGA ACAGGTCAGA GGCTGGCCCA CCACACCGTG N CACAGTCTAC ACATACCCAC TTTCTCTCT GCTCACAGAA CACTCTAGC
rs3731130	09	ATG ⇒ ATA ATG ⇒ ATC	513	M [Met] ⇒ I [Ile] M [Met] ⇒ I [Ile]	>gnl dbSNP rs3731130 allelePos=256 totalLen=511 taxid=9606 snpclass=1 alleles='A/C/G' mol=Genomic build=150 GTCTATAAA GAGGAGAGTG GGAGTGATGA GGCTGGCAGC GGCTCTGATT TTGAGCTCTC CAGTGGAGAA GCCTCTGATC CCTCTGATGA GGATTCGAA CTGGCCCTC CAAAGCAGAG GAAAGCCCC GCTCTCAGA GGACAAAGGC TGGTCCAAG AGTGCTCCA GACCCATCG TGGAGCCAT CGTAAGGACC CAAGCTTGGC AGCGGATCC TCAAGCTCTT CAAGCAGTAA AAGAGGCAAG AAAAT V TGCAGCGATG GTGAGAAGGC AGAAAAAGA AGCATAGCTG GTATAGACCA GTGGCTAGAG GTGTTCTGTG AGCAGGAGGA AAGTGGGTA TGTGTAGACT GTGTGACCG TGTGGTGGGC CAGCCTCTGA CCTGTACAA GTACGCCACC AAGCCATGA CTAATGTGT GGGCATGAC AGTGACGGCT GGGTCCGAGA TGTACACAG AGGTACGACC CAGTCTGGAT GACAGTGACC CGCAAGTCCC GGGTT
rs794729654	09	AAA ⇒ A	431	K [Lys] ⇒ R [Arg]	>gnl dbSNP rs794729654 allelePos=51 totalLen=101 taxid=9606 snpclass=2 alleles='- /AA' mol=Genomic build=144 GCGACGTCGG CATGGCCGG AGCGGCGGGT GGCTCCAGG GTGTCTTATA N GAGGAGAGTG GGAGTGATGA GGCTGGCAGC GGCTCTGATT TTGAGCTCTC
rs587778762	09	GAG ⇒ GGG	410	E [Glu] ⇒ G [Gly]	>gnl dbSNP rs587778762 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='A/G' mol=Genomic build=150 GAGCAAGCCC TCCTCCAGG AGGAAGATGA GGGCCAGGA GACAAGCAGG R GAAGGCAACC CAGCGAGCTC CGCATGGCCG GGAGCGGCGG GTGGCTCCA
Rs121965090	09	CGG ⇒ TGG	393	R [Arg] ⇒ W [Trp]	>gnl dbSNP rs121965090 allelePos=251 totalLen=501 taxid=9606 snpclass=1 alleles='C/T' mol=Genomic build=150 TTTATCATA TCATTAGCAA AAGAGGGGGG AGAAAAAAA GCAAAATTTT TTAATCTGTT TAAGGGAAG AAACCTTCCA AGGAAGATT GACTGCGGAT CAGGAGGCT CCTCAGAAAC TTCCAGCCAA GTTCTAGAAA ACCACACCAACCAAGACC AGCAAAGGAA CCAAACAAGA GAAACCTTT GCTAAGGGCA CCGCAGGCC AAGTGCCAAA GGAAGAGGA ACAAGGGAGG CAGAAAGAAA Y GGAGCAAGCC CTCTCCAGC GAGGAAGAT AGGGCCAGG AGACAAGCAG GAGAAGGCA CCCAGCAGC TCCGATGGC CGGAGCGGC GGGTGGCTC CAGGTGTCT TATAAGAGG AGAGTGGAG TGATGAGGCT GGCAGCGCT CTGATTTGA GCTCTCAGT GGAGAAGCCT CTGATCCCTC TGATGAGGAT TCCGAACCTG GCCTCCAAA GCAGAGGAAA GCCCCGCTC CTCAGAGGAC
rs200338014	09	CCT ⇒ GCT CCT ⇒ TCT	334	P [Pro] ⇒ A [Ala] P [Pro] ⇒ S [Ser]	>gnl dbSNP rs200338014 allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='A/C/G' mol=Genomic build=150 GGAAGTTTCT GAGGAGCCTC CTGGATCCG AGTCAATCTT TCCTTGGAAAG V TTTCTTCC TAAAACAGAA TAAGAAATTT TGCTTTTTT TCTCCCTCT

rs	Exon	Allele change	Position protéique	Résiduel change	séquence fasta
rs587778758	10	TAT ⇒ CAT	641	Y [Tyr] ⇒ H [His]	>gnl[dbSNP]rs587778758[allelePos=51 totalLen=101 taxid=9606 snpclass=1 alleles='C/T' mol=Genomic build=150 AGTTTCAGGC TAAACACATG GACCAGCCTT TGCCCACTGC CATTGGCTTA Y ATAAGAACCA CCCTCTGTAT GCCCTGAAGC GGCATCTCCT GAAATATGAG
rs3731152	11	ACG ⇒ ATG	689	T [Thr] ⇒ M [Met]	>gnl[dbSNP]rs3731152[allelePos=256 totalLen=511 taxid=9606 snpclass=1 alleles='C/T' mol=Genomic build=150 TGCAGATTAG GGGTTGTAAG TGGACACATC CCCAAAAGA CAGGCAGTCC ACGTTCAAGG CTGTTTGCTT AGCAGACGTT CTCTGGGCC CCTAGGAGCT CGGCCTGCTC CAGGGTGGCT GAACCTGAGG AGGGAGCCTG CAGGGCTGCG CTCTCCACTG TCCCGCATC CCGCAGTGGG GCTGTGGACT GAGTACCTT TGTGTCCTGT GTTGGTTCCA CAGGGATTGT GTGCACACTC TGCAITCCAG GGACA Y GTGGCTGAAG AAAGCAAGAG TGGTGAGGCT TGGAGAAGTA CCCTACAAGG TAACCTGGAGC TGGGGGGTCC CTGTGGGGAG TGGTGACACA GGCCTGTAGC TGAGCGGGAG AGCTGGGGGC AGAGAAGTGA TGATGAGCCT CCCAGTCTG CCCAGGATCC TGATCAGAAA TGTCCTACTT TCCCGGCTT AGCAGGATTC AAGGAGGAGA GAAACAATTA AATTGTCTTA TTAATAAGAC ATTGTGAAGA AATGG
rs200148127	13	GGC ⇒ AGC	802	G [Gly] ⇒ S [Ser]	>gnl[dbSNP]rs200148127[allelePos=201 totalLen=401 taxid=9606 snpclass=1 alleles='C/T' mol=Genomic build=150 CGTTCCTCTT GGAGGGAAGC CAGAATGGT AAAGCACTGA CTTTGCAAAAT CCAGTGTAAAC ATCTGAAAA TTGGAGCCAC CAGGCCTCAA CTCCAGCAG CCCATGCCA CCGTTCATC CCCATCTCTG GAGCACCCCC TCCCATCCC TGTGTTAGC CTCCATCGAA GGCCCTCAC GCACACGGGATGGGAGTAC Y GCCATGAAAA TCAAAGCCAG TGATGGCCTG GACACAGTCG ATGTCCAGCT TGCGGGCCAC GCGGTGTAGA TTGGCAGGT TCAGCTGGAC ACAGCCAATA GGCATCATGC TGGGCAGGAA GAGGTACACA TTCCAAACT CGTCCGGGG CACTGTGTC GGGTGAGCAA GTCAGCAITT GGCCAGCAGG GGAACAAGGC
rs786205206	13	AAC ⇒ AA	754	N [Asn] ⇒ K [Lys]	>gnl[dbSNP]rs786205206[allelePos=51 totalLen=101 taxid=9606 snpclass=2 alleles='- /C' mol=Genomic build=144 CCCCGTCTGG CCAAATGCTG ACTTGCTCAC CCGACACAGG TGCCCCGGAA N GAGTTGGGA ATGTGTACTT CTCCTGCC AGCATGATGC CTAITGGCTG
rs55779831	15	GGG ⇒ GAG GGG ⇒ GCG	846	G [Gly] ⇒ E [Glu] G [Gly] ⇒ A [Ala]	>gnl[dbSNP]rs55779831[allelePos=121 totalLen=241 taxid=9606 snpclass=1 alleles='C/G/T' mol=cDNA build=144 TCTCAAGAA ACTGACCTG AGCTTGTCTT CTCTGCAAAG AACCTGCACT GACCTTGGGC CCGTAGGCAC GCTTCAGCCT CTCCCTGATG AGCAGACCTT TGGCCAGCAA CTCCAGITC B CTAGAGCCCG CTCTCCTTT TCTGCAGGC AAAAATGAAG TGGGAGAAAA GTGTTAAGCA CTGACATTT CAGGAAAAAT ATTAAGATGG AAACCTGTGAA TGTAAGACA GATATATA
rs3731177	16	AAA ⇒ CAA	928	K [Lys] ⇒ Q [Gln]	>gnl[dbSNP]rs3731177[allelePos=256 totalLen=511 taxid=9606 snpclass=1 alleles='A/C' mol=Genomic build=150 TGCTGCCTCT TCATGGGGCT TCTGGTATC TGATTACTGA CCCTCGCCTG TGTCCTCCA CCACTGCCAC CTGTCCAGAG TGAGGCAGCA GCTCCCCACA CAGATGCAGG AGGTGGACTC TCTTCTGATG AAGAGGAGGG GACCAGCTCT CAAGCAGAAG CGGCCAGGAT ACTGGCTGCC TCCTGGCTC AAAACCGAGA AGATGAAGAA AAGCAGAAGC TGAAGGGTGG GCCAAAGAAG ACCAAAGGG AAAAG M AAGCAGCAGC TTCCACCTG TTCCATTG AGAAGCTGTG AGCTGAGCGC CCACTAGAGG GGACCCACC AGTTGCTGCT GCCCCACTAC AGGCCCCACA CTGCTCCTGG GCATGCCAG CCCCTGGTGG TGGGGGCTTC TCTGCTGAGA AGGCAAAGT AGGCAGCATG CACGGAGGCG GGGTACGGGG AGACGAGGCC AAGCTGAGGA GGTGCTGACG GTCCCTCTG CTCCAGCCC TTGTCAGAIT CACC
rs183167499	16	GCA ⇒ GGA GCA ⇒ GTA	878	A [Ala] ⇒ G [Gly] A [Ala] ⇒ V [Val]	>gnl[dbSNP]rs183167499[allelePos=201 totalLen=401 taxid=9606 snpclass=1 alleles='A/C/G' mol=Genomic build=144 GGCGCTCAGC TCACAGTGC TCAAATGGGA ACAGGTGGGA AGTGTCTGCT TCTTTTCCC TTTTGGTCTT CTGGGGCCA CCCTCAGCT TCTGCTTTC TTCACTTCT CGGTTTGGAG GCCAGGAGC AGCCAGTATC CTGGCCGCTT CTGCTGAGA GCTGGTCCCC TCCTTTCAT CAGAAGAGAG TCCACTCCT V CATCTGTGTG GGGAGCTGCT GCCTACTCT GGACAGGTGG CAGTGGTGGG AGGACACAGG CGAGGGTTAG TAATCAGATA CCAGGAAGCC CCATGAAGAG GCAGCAAGTT CCCAGCTGC CCTAAGCTGT GTAACCTGG GAGGACAAGG GCATGGGACA GGCAGGGAG AGAGCCAGCA CCCTTCTAC TGAGCCCTCC
rs375859472	16	CCC ⇒ CTC	824	P [Pro] ⇒ L [Leu]	>gnl[dbSNP]rs375859472[allelePos=101 totalLen=201 taxid=9606 snpclass=1 alleles='A/G' mol=Genomic build=150 GTTTTGAGGC CAGGAGGCAG CCAATATCTT GGCCGCTTCT GCTTGAGAGC TGGTCCCCTC CTCTCATCA GAAGAGAGTC CACCTCTGC ATCTGTGTGG R GAGCTGCTG CTACTCTGG ACAGTGGCA GTGGTGGAG GACACAGGCGAGGGTAGTA ATCAGATACC AGGAAGCCCC ATGAAGAGGC AGCAAGTTCC

Annexe VI : Tableau d'énergie de Cluspro.

Tableau d'énergie de Cluspro de modèle XPC-modèle3-I-TASSER avec centrine.

Cluster	Members	Representative	Weighted Score
0	95	Center	-890.7
		Lowest Energy	-1002.0
1	72	Center	-891.4
		Lowest Energy	-914.3
2	60	Center	-945.5
		Lowest Energy	-982.4
3	53	Center	-842.0
		Lowest Energy	-942.4
4	42	Center	-775.3
		Lowest Energy	-947.9
5	35	Center	-992.3
		Lowest Energy	-992.3
6	31	Center	-774.1
		Lowest Energy	-860.3
7	30	Center	-768.0
		Lowest Energy	-944.8
8	25	Center	-821.1
		Lowest Energy	-942.1
9	25	Center	-861.8
		Lowest Energy	-861.8
10	21	Center	-792.5
		Lowest Energy	-854.0
11	20	Center	-790.7
		Lowest Energy	-866.8
12	19	Center	-818.8
		Lowest Energy	-884.0
13	19	Center	-766.8
		Lowest Energy	-824.8
14	18	Center	-818.6
		Lowest Energy	-861.4
15	18	Center	-876.7
		Lowest Energy	-876.7
16	17	Center	-858.1
		Lowest Energy	-865.9
17	17	Center	-822.6
		Lowest Energy	-822.6
18	14	Center	-860.3
		Lowest Energy	-965.0
19	14	Center	-824.6
		Lowest Energy	-824.6
20	14	Center	-797.0
		Lowest Energy	-909.1
21	13	Center	-772.7
		Lowest Energy	-808.3
22	13	Center	-781.2
		Lowest Energy	-862.4
23	13	Center	-836.9
		Lowest Energy	-840.9
24	13	Center	-858.0
		Lowest Energy	-885.1
25	12	Center	-794.8
		Lowest Energy	-970.0
26	11	Center	-765.7
		Lowest Energy	-837.3
27	11	Center	-791.0
		Lowest Energy	-839.0
28	11	Center	-823.3
		Lowest Energy	-918.0
29	10	Center	-816.4
		Lowest Energy	-886.8

Tableau d'énergie de Cluspro de modèle XPC-modèle3-I-TASSER avec TFIH.

Cluster	Members	Representative	Weighted Score
0	64	Center	-838.4
		Lowest Energy	-874.4
1	51	Center	-719.5
		Lowest Energy	-853.2
2	51	Center	-811.4
		Lowest Energy	-846.3
3	49	Center	-750.5
		Lowest Energy	-815.8
4	40	Center	-893.8
		Lowest Energy	-893.8
5	40	Center	-726.2
		Lowest Energy	-759.6
6	39	Center	-754.5
		Lowest Energy	-766.6
7	38	Center	-716.7
		Lowest Energy	-783.6
8	30	Center	-747.6
		Lowest Energy	-747.6
9	28	Center	-783.8
		Lowest Energy	-828.3
10	26	Center	-713.1
		Lowest Energy	-738.9
11	25	Center	-710.9
		Lowest Energy	-729.6
12	23	Center	-690.0
		Lowest Energy	-810.9
13	23	Center	-802.0
		Lowest Energy	-802.0
14	23	Center	-767.9
		Lowest Energy	-805.4
15	23	Center	-739.4
		Lowest Energy	-755.0
16	20	Center	-704.2
		Lowest Energy	-750.8
17	18	Center	-706.9
		Lowest Energy	-747.4
18	18	Center	-736.1
		Lowest Energy	-854.6
19	17	Center	-728.0
		Lowest Energy	-728.0
20	16	Center	-745.2
		Lowest Energy	-805.3
21	16	Center	-809.7
		Lowest Energy	-809.7
22	15	Center	-748.9
		Lowest Energy	-748.9
23	14	Center	-706.8
		Lowest Energy	-791.5
24	14	Center	-703.0
		Lowest Energy	-749.8
25	13	Center	-718.2
		Lowest Energy	-792.4
26	12	Center	-728.0
		Lowest Energy	-762.2
27	11	Center	-718.4
		Lowest Energy	-718.4
28	11	Center	-706.7
		Lowest Energy	-718.0
29	10	Center	-691.4
		Lowest Energy	-773.7

Résumé

Cette recherche fait partie de l'étude *in silico* (utilisant des approches bioinformatiques et une modélisation moléculaire) pour analyser les différents variants XPC et leur impact sur Xeroderma pigmentosum. Nous visons à énumérer les SNP qui peuvent modifier les propriétés de la protéine XPC et prédire sa structure 3D qui est inconnue. À cette fin, nous avons utilisé le serveur UCSC et NCBI pour énumérer et étudier les SNP affectant notre protéine, ces derniers seront exploités dans la partie de modélisation moléculaire où nous suivrons trois approches de prédiction de structure 3D différentes par trois programmes différents (la modélisation comparative par le programme SwissModel, modélisation par la reconnaissance de repliements par le programme Phyre2, et reconnaissance des repliements combinés avec *ab initio* par le programme I-TASSER); en outre l'utilisation du logiciel Pymol nous a permis de mesurer la similarité structurale entre les modèles XPC prédits et les fragments XPC expérimentaux en calculant les rms ce qui montre que le modèle prédit de la XPC par I-TASSER a donné la meilleure similarité structurale avec des rms très faible par rapport aux autres modèles cela nous a permis de le retenir comme une structure probable de la XPC, ainsi faire une mutagenèse de la XPC prédite par I-TASSER selon les SNPs étudiés, un amarrage moléculaire a été effectué pour calculer l'orientation préférée de deux molécules afin de former un complexe stable (protéines XPC) par le serveur ClusPro. XPC muté selon les SNPs associés à la maladie a permis l'analyse de l'impact de la mutation sur les propriétés de la protéine (sa fonction, la formation des complexes).

Mots clés : Xeroderma pigmentosum, NER, SNP, modélisation moléculaire, prédiction structurale.