UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Estudo conformacional do complexo proteico DDB2-DDB1 e suas diferentes variantes mutantes na doença Xeroderma Pigmentosum

BRUNO CÉSAR FELTES

Porto Alegre

18 de agosto de 2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Estudo conformacional do complexo proteico DDB2-DDB1 e suas diferentes variantes mutantes na doença Xeroderma Pigmentosum

MSC. BRUNO CÉSAR FELTES

Tese submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular (PPGBCM) da UFRGS como requisito parcial para obtenção do grau de Doutor em Biologia Celular e Molecular.

Orientador: Prof. Dr. Diego Bonatto

Co-orientador: Prof. Dr. Hugo Verli

Porto Alegre

18 de agosto de 2017

INSTITUIÇÕES E FONTES FINANCIADORAS

Agências financiadoras

Conselho Nacional de Desenvolvimento Científico e Tecnológico

CAPES (Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior).

Instituição de origem

Laboratório de Biologia Molecular e Computacional, sala 219

Centro de Biotecnologia da UFRGS

Universidade Federal do Rio Grande do Sul

“Ay, caramba!”

- Bart Simpson

AGRADECIMENTOS

Em minha dissertação superei minhas capacidades filosóficas pela forma como transmiti meus agradecimentos e não gostaria de repetir-me, assim como não gostaria de transmitir os novos agradecimentos de forma ortodoxa. Dessa forma, fiquei olhando para este documento como um pateta, esperando uma inspiração. Acho que perdi minha licença de poeta nos últimos quatro anos e estou tendo dificuldades de encontrá-la. Porém, chego à conclusão de que não preciso de uma inspiração cósmica, ou uma epifania para escrever algo para o que basta sinceridade. Então, caso você encontre minha licença de poeta, por favor, faça bom uso dela.

Viajo toda semana para minha cidade natal por um motivo: ver os dois seres humanos que me criaram. Falando assim sei que parece que sou nada mais do que um constructo de carbono programado para falar, mas esse não é o caso. À minha mãe, Heloísa Pedroso de Moraes Feltes, devo o mais complexo agradecimento: o de estimular ao livre-pensar. Ela me forneceu espaço para pensar, ser e existir, criando minha personalidade livremente, mas puxando-me para a realidade quando necessário. Hoje sou resoluto em minhas decisões e me dou o direito de ser quem sou por causa dessa liberdade. Ao meu pai, Paulo César Feltes, devo o exemplo de superação e força de vontade de seguir em frente, mesmo em cenários desfavoráveis, o que ajudou a me mostrar que meu conhecimento de nada serve sem força de vontade para lutar.

Trabalhando do meu lado pelos últimos nove anos, encontra-se minha ex-namorada. Chamo-a assim porque hoje chamo-a de esposa. A minha esposa, Joice de Faria Poloni, agradeço por algo tão raro ao ponto que muitas vezes custo a acreditar que tenha recebido tal tesouro: o amor mais resistente e colossal que já vi. Sua personalidade, competência, inteligência e carinho me tornaram uma pessoa que eu nem sabia que poderia ser. Eu achava que conhecia a felicidade antes de conhecê-la em 2008. Adivinhem? Eu estava redondamente enganado.

Chegamos ao ponto em que aquele constructo de carbono fala, pensa, tem força vontade e é feliz. Mas nenhuma construção fica de pé sem suporte. E esse suporte veio na forma dos mais sensacionais amigos e colegas de trabalho que já cruzaram a galáxia. Primeiro vou agradecer a meus colegas de trabalho, que nos

últimos seis anos permaneceram ao meu lado e me forneceram ajuda no trabalho e na vida pessoal. O maior conhecimento que adquiri no meu doutorado não foi científico, foi pessoal, e eles foram responsáveis por grande parte disso. Então, à Beatriz, ao Chapola, que alguns dizem se chamar Henrique, mas tenho certeza de que isso é um engano, aos Gabri(h)eis, à Larissa e à Raquel, o meu mais sincero obrigado. Deixei um espaço especial ao Kendi, pois foi ele que entrou nos assuntos mais filosóficos possíveis comigo e se esforçou para entender minha mente mais do que qualquer amigo em minha história, além de ser uma agência de fomento extra.

Aos meus amigos de Caxias do Sul, os quais são responsáveis por manter minha sanidade por todos esses anos, eu devo muitos obrigados. A presença deles não só me deu suporte, mas alegrou minha existência: Armando (Sísifo/O Lich), Edna, Letícia, Manuel (Siggard/Rolen), Miguel (Skull/Flick), Nicolas (Seraphiel/Razael), Tahila e Tito (Eban/Luigi). A todos, um grande obrigado, e àqueles com nomes em parênteses: nós sempre teremos uma ilha!

Um agradecimento especial devo aos meus orientadores Diego Bonatto e Hugo Verli, que me forneceram as condições técnicas e suporte científico para a produção do trabalho que vocês lerão a seguir. Da mesma forma, um enorme obrigado é necessário ao grupo do professor Hugo, por ter compartilhado seu conhecimento comigo nesses últimos quatro anos, não só de forma muito profissional, mais em meio a ótimas conversas descontraídas. Um obrigado especial devo ao Conrado Pedebos, vulgo Science Man, e ao Rodrigo Ligabue-Braun.

Um agradecimento à minha família pelo carinho e apoio. Em especial à minha madrinha, Berenice Pedroso de Moraes, a minha tia Miriam Feltes e ao meu padrinho Ivan Carlos Feltes, pelo impacto direto em minha vida pessoal. Também agradeço a meus sogros, João Poloni e Marlene Faria Poloni, pelo grande apoio e atenção.

À Thara e à Vicky, por serem os seres mais fofos e queridos do universo.

Por fim, sempre gosto de lembrar que todos são formados por centenas de variáveis. Essas variáveis podem ser fatos, mensagens, músicas, livros, jogos e, especialmente, pessoas. Aqueles que acham que podem viver sem o conhecimento de outros seres humanos e não possui a humildade de compartilhá-lo e de recebê-lo, espero que um dia encontrem em si a vontade de mudar.

Espere... aquela ali é minha licença de poeta?

ESTRUTURA DA TESE

Esta tese de doutorado é dividida em uma introdução geral, dois capítulos redigidos em forma de artigo, sendo um publicado e um a ser submetido, uma discussão geral e conclusões. Da mesma forma, é incluído um item de apêndices contendo outras produções desenvolvidas durante o período de doutorado que, embora dentro do tema de reparo de DNA e Xeroderma Pigmentosum, não eram adequadas para o corpo principal da tese.

A introdução geral consiste em uma apresentação histórica do descobrimento da doença Xeroderma pigmentosum e os caminhos desbravados pelos primeiros pesquisadores responsáveis pelos principais avanços no entendimento desta doença. Juntamente ao apanhado histórico, as manifestações clínicas e bases moleculares da doença são abordadas. Este tópico também disserta sobre a via de atuação das proteínas, cuja disfunção acarreta no desenvolvimento da doença. A introdução geral termina com uma explicação sobre dinâmica molecular e redes dinâmicas, e suas vantagens no estudo da biologia estrutural.

O Capítulo 1 trata-se de uma revisão sobre estrutura das proteínas XPA até XPG, suas funções biológicas, e interações proteína-proteína. A revisão discute todas as mutações encontradas nessas proteínas, tanto naturais como artificialmente induzidas e como elas afetam sua estrutura e função biológica. Da mesma forma, todas as modificações pós-traducionais e resíduos cruciais para o papel biológico das proteínas são listas e discutidas. Esse artigo foi publicado na revista *Mutation Research Reviews* e, no presente momento, é a revisão mais atual sobre o tópico.

O Capítulo 2 consiste no objeto de estudo central desta tese, apresentando uma profunda análise de dinâmica molecular do comportamento da proteína DDB2 (produto do gene XPE) e suas diferentes variantes mutantes encontradas em pacientes com XP, tipo E. Da mesma forma, o comportamento do complexo entre a proteína DDB2 e DDB1, sua principal proteína de interação, é estudado tanto em seu estado nativo, quanto composto pelas distintas variantes mutantes de DDB2. O efeito das mutações estudadas é explorado por diferentes ferramentas de análise estrutural e múltiplas simulações. O comportamento comparativo dos complexos é discutido juntamente com o impacto de cada mutação nas proteínas e complexos estudados.

Este artigo é o primeiro trabalho computacional que atenta explicar como as mutações vistas em pacientes com XP, tipo E, afetam a estrutura das proteínas envolvidas na patologia. O artigo será submetido para a revista *Nucleic Acids Research*.

Os dois capítulos são seguidos de uma discussão geral dos resultados obtidos e uma conclusão do tema, focando no impacto do trabalho desenvolvido na tese no entendimento da doença. Perspectivas são levantadas e discutidas.

Por fim, a tese finaliza com um item de apêndices, contendo trabalhos produzidos durante o período de doutorado que, embora relacionados ao tema da doença Xeroderma pigmentosum, não eram adequados para fazer parte do corpo principal da tese. Desta forma, o primeiro trabalho consiste em um capítulo de livro que trata doenças de caráter genético, como Xeroderma pigmentosum, Síndrome de Cockayne, Canceres, Anemia Fanconi, dentre outras. Ademais o trabalho discute, de forma atualizada, suas características, vias moleculares e peculiaridades. O capítulo foi publicado pela editora Elsevier, em 2016. O segundo trabalho é uma análise do impacto da mutação no gene *xpa-1* no nematoide *Caenorhabiditis elegans* e como ele afeta seu desenvolvimento. Essas observações são discutidas focando em possíveis relações com o desenvolvimento de Xeroderma pigmentosum em humanos, uma vez que *xpa-1* é homólogo de XPA em *Homo sapiens*. Esse artigo encontra-se submetido na revista *Functional and Integrative Genomics*.

SUMÁRIO

LISTA DE ABREVIATURAS

6,4PP = (6,4)-Pirimidina-Pirimidona

ATM = *Ataxia Telangiectasia Mutated*

ATR = *Ataxia Telangiectasia And Rad3-Related Protein*

BER = *Base Excision Repair*

CDK7 = *Cyclin Dependent Kinase* 7

CEP-164: Centrosomal Protein-164kDa

CHO = *Chinese Hamster Ovary*

CPD = *Cyclobutane Pyrimidine Dimer*

CSA e B = *Cockayne Syndrome WD Repeat Protein* A e B

CUL4A = *Culin* 4A

DDB1-2 = *DNA-Damage Binding Protein* 1 e 2

DM = Dinâmica Molecular

E2F1 = E2F *Transcription Factor* 1

ERCC1 = *Excision Repair Cross-Complementation Group* 1

FANCA = *Fanconi Anemia Complementation Group* A

GG-NER = *Global Genome-Nucleotide Excision Repair*

HR23B = XP-C *Repair Complementing Protein*

ILCR = *Interstrand Cross-Link Repair*

IPP = Interação Proteína-Proteína

MAT1 = *Menage A Trois* 1 (é isso mesmo)

NER = *Nucleotide Excision Repair*

P8 = também chamada de NUPR1, *Nuclear Protein, Transcriptional Regulator* 1

P44 = também chamada de MAPK3, *Mitogen-Activated Protein Kinase* 3

P52 = também chamada de NFKB2, *Nuclear Factor Kappa B Subunit* 2

P53 = também chamada de TP53, *Tumor Protein* P53

P62 = também chamada de SQSTM1, *Sequestosome 1*

P300 = também chamada de EP300, E1A *Binding Protein* P300

PCNA = *Proliferating Cell Nuclear Antigen*

RAD51-2 = RAD51 *Recombinase* 1 e 2

RCC1 = *Regulator of chromatin condensation* 1

RD = Redes Dinâmicas

RFC = *Replication Factor* C

ROC1 = *Ring Box* 1

RPA = *Replication Protein* A1

STAGA = SPT3-TAF9-GCN5 *acetylase complex*

TC-NER = *Transcription Coupled-Nucleotide Excision Repair*

TDG = *Thymine* DNA *Glycosylase*

TFIIH = *Transcription fator* II H

TRF2 = *Telomeric Repeat Binding Factor* 2

XAB1-2 = *XPA Binding Protein* 1 e 2

XP = Xeroderma Pigmentosum;

XPA até XPG = Xeroderma Pigmentosum *complementation group* A até G;

RESUMO

Uma das dificuldades no entendimento da doença Xeroderma Pigmentosum (XP) é a falta de dados estruturais das proteínas XPA até XPG, cuja inatividade ou disfunção levam ao seu desenvolvimento. Nesse sentido, a proteína DDB2, produto do gene *XPE*, atua no reconhecimento de danos ao DNA, nos primeiros passos da via de Reparo por Excisão de Nucleotídeos, a principal via molecular afetada na doença. Da mesma forma, pacientes com desregulação em DDB2 possuem uma das maiores incidências de câncer de pele dentre os subtipos de XP, tornando seu entendimento importante para as ciências biomédicas. Atualmente, DDB2 é a única proteína XP cuja estrutura tridimensional está 90% elucidada, contudo, pouco de sabe sobre seu comportamento molecular e como as mutações encontradas em pacientes com XP afetam seu funcionamento e sua ligação com a proteína DDB1, sua principal parceira molecular.

Sendo assim, esta tese de doutorado tem como objetivo caracterizar o comportamento molecular de DDB2 e as variantes mutantes R273H, L350P e K244E, assim como DDB2 complexada com DDB1 e como as diferentes mutações afetam o complexo DDB2-DDB1. Para tanto, análises por dinâmica molecular e redes dinâmicas foram empregadas para cada sistema selvagem e mutante. Os dados revelaram que todas as variantes mutantes apresentam um aumento de rigidez em DDB2. Adicionalmente, uma região específica, resíduos 354 até 371, mostrou-se fortemente afetada nos mutantes e parece ter conexão com seu estado conformacional. Quando complexada, todos mutantes de DDB2 geraram uma proteína mais rígida. A mutação L350P alterou a interação entre DDB2 e DDB1, originando uma interação mais fraca. A mutação R273H, assim como a L350P, alterou o domínio de ligação ao DNA de DDB2, e este mesmo domínio pareceu ser instável comparado à forma selvagem. A mutação K244E mostrou uma alteração de ligação entre o domínio de ligação do DNA de DDB2 e DDB1, indicando um complexo com perda de função de reconhecimento. Sendo assim, os dados forneceram informações críticas para o entendimento de como as mutações vistas em XP, tipo E, afetam DDB2 e seu complexo com DDB1.

ABSTRACT

A major challenge in the understanding of the Xeroderma Pigmentosum (XP) disease is the lack of structural data of XPA to XPG proteins, whose inactivity or dysfunction, lead to XP. In this sense, the DDB2 protein acts in the DNA-damage recognition, on the first steps of the Nucleotide Excision Repair pathway, the main molecular pathway affects in XP. Patients with impairments in DDB2 protein have one of the highest incidences of skin cancer among the XP subtypes, making its understanding an important matter in the biomedical field. Nowadays, DDB2 is the only XP protein which its tridimensional structure is 90% elucidated; however, little is known about how the mutations found in the DDB2 protein affect its binding proprieties to with protein DDB1 in XP patients.

This work aims to characterize the molecular dynamics of wild-type DDB2 and the XP variants R273H, L350P and K244E. In addition, the dynamic of WT- and mutant-associated DDB2-DDB1 complex was performed. For this purpose, molecular dynamics and dynamic networks were employed for each WT and mutant DDB2, alone and complexed with DDB1. The data revealed that the all mutated variants show an increase in structural rigidity in DDB2. Additionally, a specific region, residues 354 to 371, was altered in the mutants and appears to be linked to the DDB2 conformational state. When complexed, all DDB2 mutants display a less dynamic structure. In this sense, the L350P mutation diminished the potential of interaction between DDB2 and DDB1. The R273H mutation, as well as the L350P, modified the DNA-binding domain of DDB2, and the same domain seems to be unstable when compared to the WT. The K244E mutants affects the DNA-binding domain of DDB2 as well as the interaction with DDB1, denoting a complex with loss of the DNA-recognition function. Thus, the data offers critical information for the understanding of how mutations found in XPE affect DDB2 and its complex with DDB1.

1. INTRODUÇÃO

1.1. Xeroderma Pigmentosum: seu descobrimento, bases moleculares e impactos nos estudos de reparo de DNA

Descoberta em 1874 pelo dermatologista austríaco Ferdinand Ritter von Hebra e seu genro, de mesma profissão, o húngaro Moritz Kaposi, a doença Xeroderma Pigmentosum (XP) (do latim *parchment-like skin with pigmentation abnormalities*) ganhou seu nome baseada na observação de duas de suas características fenotípicas clássicas, sendo elas uma pele ressecada e hiperpigmentada (Cleaver, 2008; Gillet & Schärer, 2006). A XP é uma doença genética rara, de característica autossômica e recessiva (**Fig. 1**), com uma incidência global de aproximadamente um caso a cada 200-250 mil habitantes, sem restrições étnicas ou de sexo. Contudo, uma incidência maior em comunidades japonesas e do norte da África, de um caso a cada 40 mil habitantes, já foi evidenciada (Cleaver, 2008; Gillet & Schärer, 2006). Ao nascerem, os indivíduos não apresentam sintomas aparentes, porém as primeiras manifestações morfológicas clássicas começam em torno em torno dos três anos, acompanhadas de atrofia, ceratose e telangiectasia (**BOX 1**) (Setlow et al., 1969).

Apesar das características que geraram o nome da doença, os sintomas que mais caracterizam XP são a hipersensibilidade à luz solar e a alta incidência de cânceres de pele (Gillet & Schärer, 2006; Schärer, 2013). A hipersensibilidade à luz solar pode ocasionar lesões nas pálpebras e ulcerações nas córneas que normalmente levam à cegueira (Setlow et al., 1969). Embora esses sintomas já fossem claros na época da descoberta, a investigação dos mecanismos moleculares subjacentes a XP só tiveram início em torno de 90 anos depois, na década de 1960, através dos trabalhos de James E. Cleaver, quando ela foi reconhecida como uma doença resultante de defeitos no reparo de DNA (Cleaver, 1963, 1969 e 2008).

BOX 1

Atrofia: Doença que provoca uma diminuição das células que compõe um tecido, através do processo de apoptose. Desta forma, a Atrofia provoca disfunção, diminuição no volume e tamanho dos órgãos, assim como nos demais tecidos.

Ceratose: Lesão na pele causada, normalmente, pela exposição à luz do sol (UV). Esta doença é considerada um estado pré-cancerígeno de câncer de pele. Manifesta-se como uma camada de pele seca, pigmentada e com textura anormal.

Telangiectasia: Se refere a más-formações causadas por dilatações em vasos sanguíneos que podem aparecer em qualquer parte do corpo.

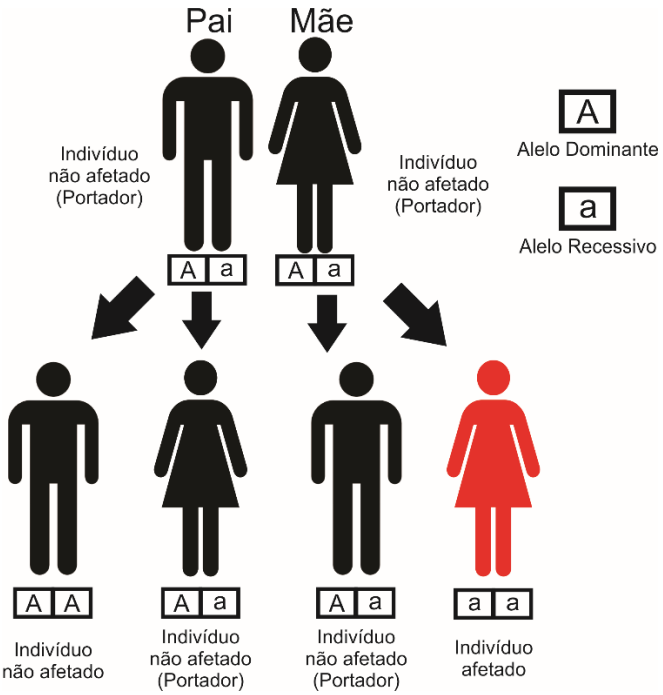


Fig. 1. Herança genética de característica recessiva. Uma doença autossômica indica que não existe relação entre a doença e cromossomos sexuais. A propriedade de recessividade indica que é necessário que a prole possua as duas cópias alélicas recessivas para que a doença se manifeste. Como indicado na figura, qualquer mistura que não resulte em duas cópias do alelo recessivo não ocasionará um indivíduo com XP.

Tendo conhecimento da alta sensibilidade à luz solar compartilhada pelos indivíduos com XP, Cleaver submeteu culturas de fibroblastos derivados da pele de três pacientes com XP e da pele de indivíduos normais à irradiação por luz ultravioleta (UV). A luz UV é conhecida por causar danos à molécula de DNA, e Cleaver desejava averiguar se as células eram capazes de sintetizar DNA após a irradiação por luz UV. Utilizando nucleotídeos marcados por isótopos radioativos, ele analisou os níveis de produção de DNA após a irradiação. O resultado foi que os fibroblastos derivados da pele saudável conseguiram reparar os danos de DNA causados pela luz UV, enquanto os fibroblastos de pacientes com XP reparavam o dano numa velocidade extremamente inferior, ou não possuíam função de reparo (Cleaver, 1963). Entretanto, esse fato não demonstrava quais proteínas estavam envolvidas na doença ou em quais vias de reparo de DNA as mesmas poderiam estar atuando. Era um fato estabelecido que culturas de células, oriundas de diferentes indivíduos com XP, se comportavam de

formas diferentes quando irradiadas com luz UV, mostrando diferentes níveis de atividade de reparo de DNA (Cleaver, 1963). Essa observação estabeleceu a hipótese de que defeitos em proteínas distintas poderiam ser o fator determinante desse comportamento. A descoberta sobre quais proteínas poderiam estar envolvidas com XP veio anos depois da descoberta de Cleaver, com De Weerd-Kastelein e colaboradores, em 1972, utilizando a técnica de fusão de células somáticas e revolucionando as pesquisas com XP. Nesse experimento, os autores fusionavam, por meio da fusão de seus núcleos, células originárias de diferentes amostras de pacientes com distintos fenótipos de XP e observavam a cultura de células resultantes (De Weerd-Kastelein et al., 1972). Dessa forma, quando a cultura A, fusionada com a cultura B, gerava células sem hipersensibilidade à luz UV, isso significava que elas provinham de grupos complementares diferentes, uma vez que a cultura resultante agora possuía ambas as proteínas que antes elas individualmente não possuíam (De Weerd-Kastelein et al., 1972; Gillet & Schärer, 2006) (**Fig. 2**).

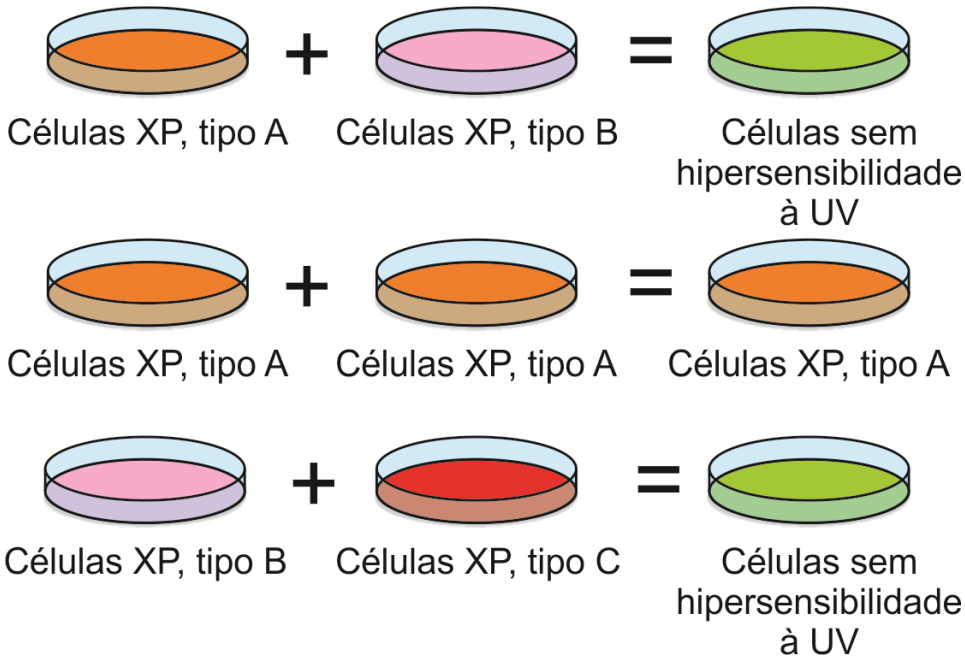


Fig. 2. Experimento feito por De Weerd-Kastelein e colaboradores. Culturas de diferentes indivíduos com XP eram fusionadas e expostas à luz UV. Caso a cultura resultante não fosse hipersensível a irradiação, significava que as culturas individuais pertenciam a grupos complementares. Caso ainda fossem hipersensíveis, indicava que elas pertenciam ao mesmo grupo.

A fusão de diversas culturas de células, conjuntamente com a criação de células mutantes de ovário de hamster chinês (CHO), que correspondiam a células humanas com XP, determinou a existência de sete proteínas, então chamadas de *xeroderma pigmentosum, complementation group*, nomeadas de A a G (XPA-XPG) (Busch et al., 1980; Thompson et al., 1980; Cleaver, 2008). O nome "*complementation group*" foi baseado nos resultados obtidos pela técnica de fusão.

O impacto do descobrimento de uma doença de caráter dermatológico que estaria intrinsicamente relacionada ao reparo de DNA estimulou os estudos nas áreas da radiobiologia e da biologia molecular do reparo, e não tardou para que estudos sobre as principais lesões ao DNA, causadas pela luz UV, fossem realizados e redescobertos, adicionando mais informações sobre sua formação, características e cinética de reações (Fisher & Johns, 1976; Franklin et al., 1982, Wood et al., 1988). Da mesma forma, estudos com *Escherichia coli* foram fundamentais para compreender a dinâmica da excisão de lesões ao DNA dado à alta semelhança do processo em bactéria e humanos (Bohr, 1991).

Foi verificado que pirimidinas adjacentes, como timinas e citosinas, poderiam sofrer uma reação quando a célula era irradiada com luz UV, transformando-se em um novo nucleotídeo danoso ao DNA. Duas lesões principais são comumente discutidas: os dímeros de ciclobutano de pirimidina (do inglês *cyclobutane pyrimidine dimer* – CPD) e os fotoprodutos (6,4)-pirimidina-pirimidona (6,4PP). A **Fig. 3** exemplifica, através de duas timinas, como são geradas essas duas lesões. Duas timinas adjacentes em conformação *cis*, quando irradiadas com luz UV, reagem através dos carbonos nas posições 5 e 6 da instauração de ambas estruturas cíclicas, gerando uma lesão CPD, agora com as estruturas cíclicas em conformação *trans* (Gillet & Schärer, 2006). A posição 6 da estrutura cíclica de uma timina também pode reagir com a posição 4 da outra, criando um 6,4PP. No caso de duas timinas, uma em conformação *cis* e outra em *trans*, o resultado será um 6,4PP, cuja única diferença do 6,4PP gerada pelas timinas em *cis* será a troca de um radical na lesão resultante (Gillet & Schärer, 2006) (**Fig. 3**).

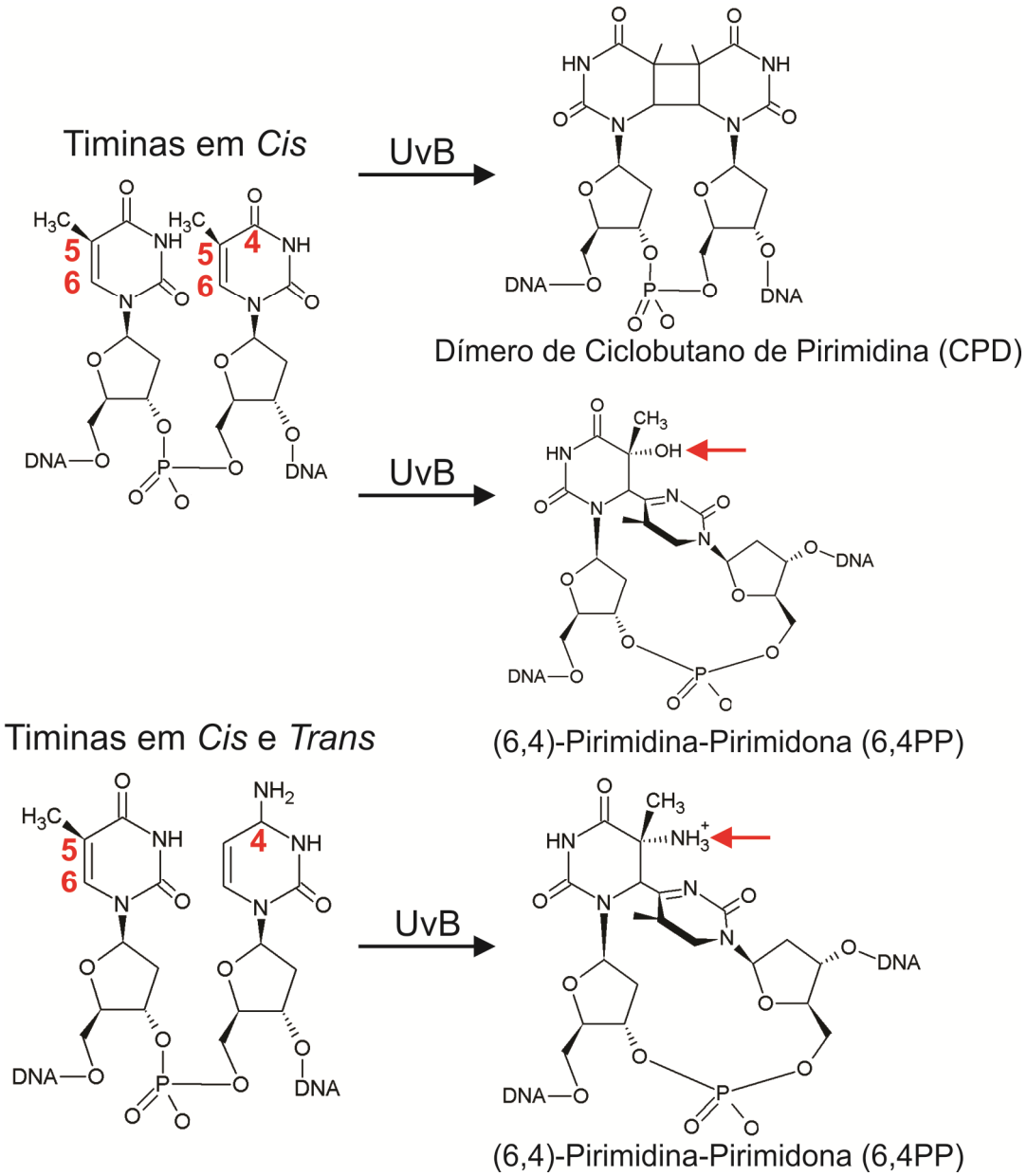


Fig. 3. Principais lesões no DNA geradas por irradiação à luz UV. Duas pirimidinas adjacentes, neste caso timinas, podem gerar uma nova forma de nucleotídeo (lesão) na molécula de DNA. Quando as duas timinas estão adjacentes na conformação *cis*, a exposição a UvB pode gerar tanto um CPD quanto um 6,4PP. Quando as timinas adjacentes se encontram, uma na conformação *cis* e outra na conformação *trans*, o resultado é um 6,4PP, cuja única diferença é a troca de um radical hidroxila (OH) por um radical amônia (NH_3^+ (setas vermelhas). Os números em vermelho correspondem à posição dos carbonos que reagem para gerar a lesão. As moléculas foram desenhadas utilizando o programa ChemSketch (http://www.acdlabs.com/resources/freeware/chemsketch/).

Estava claro que XP estava intimamente ligada ao reparo de DNA. Estudos demonstravam que células de paciente com XP eram deficientes em reparar o dano ao DNA após irradiação por luz UV (Cleaver, 1963), e possuíam uma atividade muito menor da remoção de CPD por endonucleases, que clivam o DNA para a remoção do dano (Setlow et al., 1969; Cleaver, 1969). Os caminhos bioquímicos da doença claramente apontavam para uma disfunção do reparo por excisão de nucleotídeos (do inglês, *Nucleotide Excision Repair* - NER) (Cleaver & Bootsma, 1975), e a ideia de que proteínas XP formavam complexos com outras proteínas já estava sendo considerada (Giannelli & Pawsey, 1974). Após décadas de uso de culturas de células, experimentos com bactérias e células CHO, o modelo murino finalmente foi disponibilizado, criando uma nova plataforma para o estudo de XP (Friedberg & Meira, 2000).

Finalmente estabelecida as bases do conhecimento de XP, a comunidade científica engajada no estudo dessa doença compreendeu que havia sete diferentes manifestações fenotípicas, cada uma envolvendo um gene diferente, com suas próprias distinções.

1.2. Manifestações clínicas da doença Xeroderma Pigmentosum

Uma vez estabelecido o descobrimento e como isso levou ao reconhecimento de sete variantes principais dos genes XP, o detalhamento de suas manifestações clínicas foi facilitado, especialmente para desenvolver seu diagnóstico.

A extrema sensibilidade à luz UV é o primeiro sintoma observado para elaboração do diagnóstico da doença, em que queimaduras solares são observadas nas primeiras semanas de vida, embora o diagnóstico errôneo seja comum nesses casos, pois as queimaduras podem ser consideradas casos de impetigo e celulite (Lehmann et al., 2011) (**BOX 2**). Outros 40% dos casos podem não mostrar esse sintoma (Lehmann et al., 2011). Embora queimaduras solares severas sejam comuns nos grupos XP tipo A-B, -D e F-G, os grupos –C e –E não apresentam esse grau de severidade. A partir dos 2-3 anos, a formação de lentigos é observada, acompanhada de ceratose, telangiectasia e ataxia (Setlow et al., 1969; Lehmann et al., 2011) (**BOX 1 e 2**). Todos esses sintomas se concentram em regiões expostas ao sol, mas podem afetar o corpo inteiro (**Fig. 4**). Dentre todos os casos de XP, em torno de 40% sofrem de problemas oftalmológicos como fotofobia, conjuntivite e ulcerações nas córneas são mais

comuns, porém, blefatite, simbléfaro e ectrópio podem ocorrer (**BOX 3**) (Hengge & Emmert, 2008).

Pacientes com XP podem apresentar problemas neurológicos, especialmente em casos de mutação no gene XPA, XPB, XPD e XPG, compreendendo a 20-30% dos casos de XP (Hengge & Emmert, 2008; Menk & Munford, 2014, 2014; Fassihi, 2013). O impacto dos problemas neurológicos incluem neurodegeneração, microcefalia, defeitos cognitivos e oculares (FASSIHI, 2013; Hengge & Emmert, 2008; Lehmann et al., 2011; Menk & Munford, 2014) (**BOX 2**). Dentre os únicos dois locais com uma maior frequência de XP, o Japão é o que possui a maior incidência de indivíduos com XP tipo A (Hengge & Emmert, 2008). Esse subtipo de XP é o mais prevalente, consistindo 30% dos casos, sendo essa variante a mais comum por ocasionar problemas neurológicos, enquanto os indivíduos com XP tipo E e tipo C não mostram essa característica (Hengge & Emmert, 2008).

Fig.4. Indivíduo com um caso de XP, apresentando os fenótipos clássicos: pele seca, lentigo, ceratose e telangiectasia. O paciente também possui um caso de carcinoma basocelular. Fonte: http://dermatologyoasis.net/multiple-basal-cell-carcinomas-in-xp/.

BOX 2

Impetigo: Doença de origem bacteriana, causada por *Staphylococcus aureus* ou *Streptococcus pyogenes* (ou ambos), que provoca vermelhidão, erupções cutâneas, formação de crosta e secreções purulentas.
Lentigo: Manchas na pele, semelhantes sardas, que não possuem relação com câncer de pele ou qualquer outra condição patogênica. Normalmente causada por exposição ao sol, mas podem ocorrer ao longo do tempo. Contudo, manchas lentiginosas podem ser malignas.
Microcefalia: Doença na qual o cérebro não se desenvolve devidamente durante o desenvolvimento, acarretando em um cérebro menor e com menos massa. Indivíduos com esta condição apresentam problemas cognitivos, motores, de fala, assim como convulsões.

Apesar de possuir uma ampla gama de sintomas, a característica mais agressiva de XP é a alta incidência de diferentes tipos de cânceres de pele, uma vez que indivíduos afetados pela doença possuem 10.000x maior probabilidade de desenvolver cânceres de pele não-melanoma até os 9 anos de idade e 2.000x maior probabilidade de desenvolver melanomas até os 20 anos (Fassihi, 2013). O câncer na cavidade oral é comum, especialmente câncer de células escamosas e basais, assim como melanomas lentiginosos (Fassihi, 2013; Hengge & Emmert, 2008). Vale ressaltar que, dentre todos os tipos de XP, os pacientes com XP tipo C e E são os que possuem a maior incidência de cânceres de pele, com 1000 x maior probabilidade de desenvolver cânceres induzidos por exposição solar, assim como tumores internos, tornando-os os mais envolvidos com a biologia do câncer (Marteijn et al., 2014).

Tendo dissertado sobre seus aspectos gerais, clínicos e sua história, torna-se indispensável elucidar a via de NER. Os estudos com essa via molecular de reparo foram um passo fundamental não só para entender os aspectos moleculares subjacentes a XP, mas também para a classificação dos subtipos de manifestações clínicas da doença.

1.3. A via de Reparo por Excisão de Nucleotídeos e as proteínas XP (NER)

Ainda em 1974, os estudos com reparo de DNA envolvendo XP levaram à descoberta de que a via de NER ocorria em diferentes partes do genoma, e esse conhecimento levou à caracterização de duas ramificações da via, hoje conhecidas como *transcription coupled* NER (TC-NER), e a ramificação mais intrinsicamente relacionada à doença XP, a *global genome* NER (GG-NER) (Mansbridge & Hanawalt, 1983; Cleaver, 2008; Feltes et al., 2016). TC-NER é apenas responsável pela remoção de danos em sítios ativos de transcrição, enquanto GG-NER é encarregado da remoção de lesões em sítios ativos de transcrição, regiões de heterocromatina e genes inativos (Gillet & Schärer, 2006; Schärer, 2013; Feltes et al., 2016).

Os genes XP estão localizados em diferentes cromossomos (Menk & Munford, 2014) (**Fig. 4**), e a via de NER, possui cinco etapas principais, nas quais cada proteína XP atua em um determinado passo: (1) reconhecimento do dano à molécula de DNA, sendo este passo a única diferença entre ambas as ramificações. No TC-NER o reconhecimento tem início quando a RNA polimerase II é atida no sítio da lesão,

provocando uma sinalização para as proteínas CSA, CSB e XAB2 que se ligam no local onde a RNA polimerase foi detida e dando inicio aos próximos passos da via. Já no GG-NER o reconhecimento é feito pelo heterodimero XPC-HR23B(RAD23B) ou o Complexo-DDB, composto por DDB2, produto do gene XPE, que faz complexo com as proteínas DDB1, CUL4A e ROC1. XPC-HR23B possui uma maior afinidade por 6,4PP, enquanto o Complexo-DDB detém uma maior afinidade por CPD. Entretanto, mesmo após detectar o dano, o Complexo-DDB recruta XPC-HR23B para o sítio da lesão; (2) abertura da dupla hélice de DNA através das proteínas helicases XPB e XPD, ambas componentes do complexo de helicase TFIIH; (3) recrutamento do heterodimero XPA-RPA, que se liga na fita simples (não danificada) e compõe uma plataforma de ligação proteína-proteína que recruta as proteínas subsequentes; (4) excisão do sítio danificado, executado pelo heterodimero XPF-ERCC1 e a proteína XPG; e (5) síntese de uma nova fita simples de DNA onde antes havia o sítio danificado (Gillet & Schärer, 2006; Schärer, 2013; Feltes et al., 2016) (**Fig. 5**).

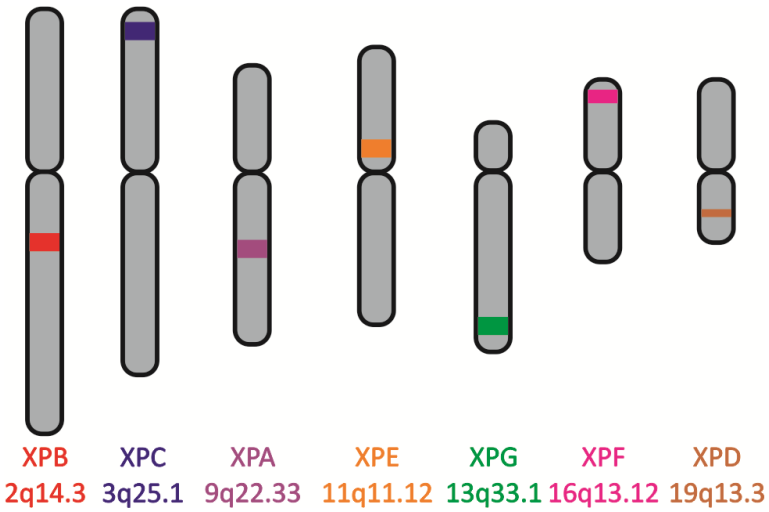


Fig. 5. Localização cromossomal dos genes XPA-XPG. Cada gene é localizado em um cromossomo diferente.

BOX 3
Blefarite: É uma doença de caráter inflamatório nas pálpebras cujos sintomas variam entre prurido, lacrimejamento e irritação. Essa doença pode levar ao crescimento bacteriano e problemas de visão. **Simbléfaro:** Se refere a uma condição onde a superfície conjuntival das pálpebras se adere ao globo ocular, podendo causar inflamações e ulcerações nos olhos. **Ectrópio:** É uma condição onde o interior da pálpebra inferior torce-se para fora. Ectrópio pode causar irritações e infecções.

.

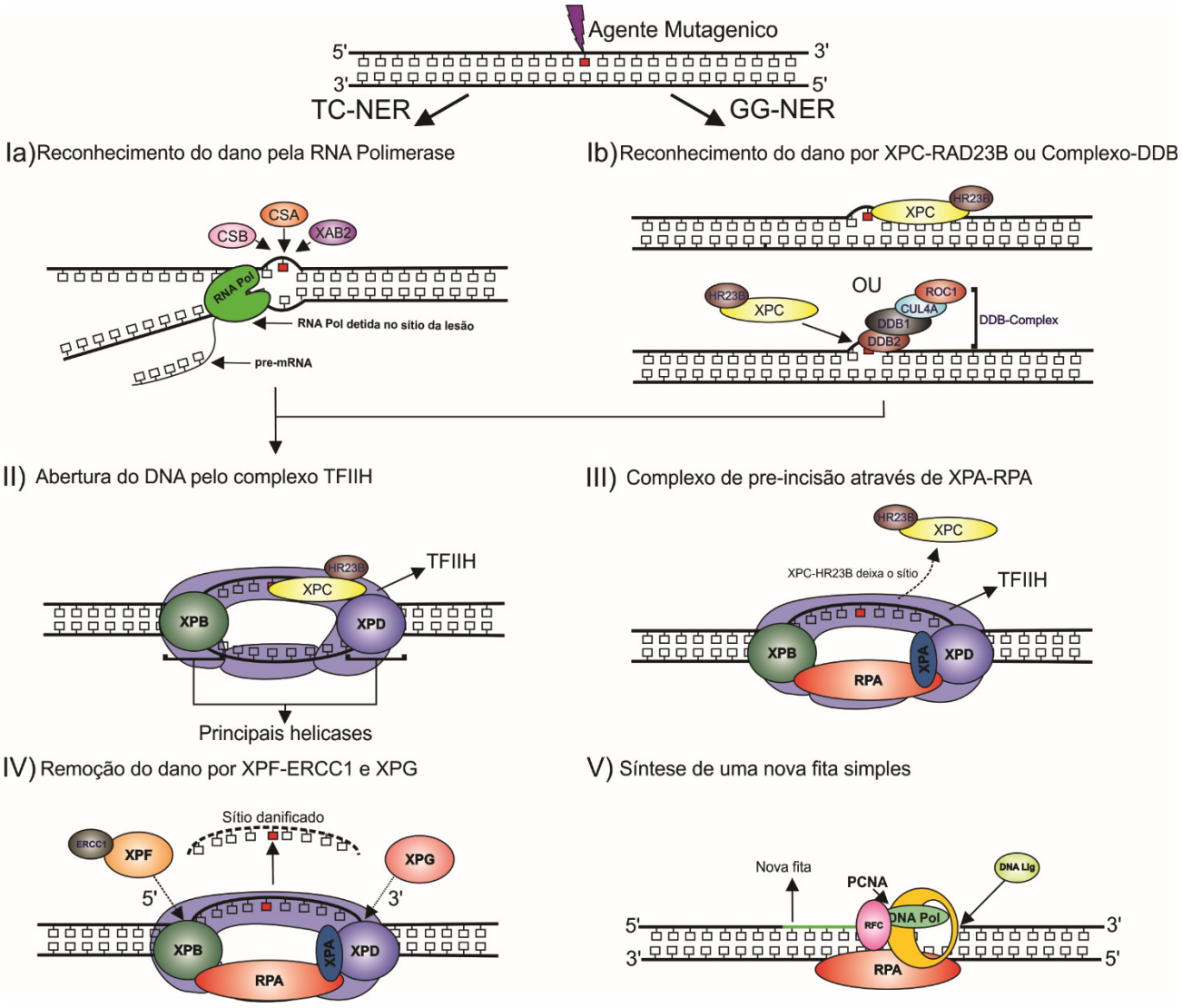


Fig. 6. A via de NER. Ela começa após o reconhecimento de uma lesão ao DNA. (Ia) No caso de TC-NER, o dano é reconhecido pela RNA polimerase II em um sítio de transcrição ativo. Esse fenômeno recruta as proteínas CSA, CSAB e XAPB2 na região. (Ib) Em GG-NER, a lesão é reconhecida por XPC-HR23B ou o Complexo-DDB. Caso o Complexo-DDB reconheça, ele recruta XPC-HR23B para a região do dano. Após o reconhecimento da lesão, ambas as ramificações convergem para os mesmos passos da via de NER. (II) XPC-HR23B recruta o complexo TFIIH, onde suas principais helicases, XPB e XPC provocam a abertura da dupla hélice. (III) XPC-HR23B deixa o sítio e o heterodimero XPA-RPA se liga na fita simples, não danificada, provocando uma plataforma de interação proteína-proteína. (IV) As endonucleases XPF (complexada com a proteína ERCC1) e XPG são recrutadas para o local da excisão e removem a fita danificada. XPF possui função de endonucleases 5', enquanto XPG possui função 3'. (V)

Por fim, com exceção de RPA, todas as proteínas deixam a região e a maquinaria de síntese, composta pela DNA polimerase, PCNA e RFC, sintetizam uma nova fita e a DNA ligase conecta a nova fita no DNA pré-existente. **Essa imagem foi traduzida de um capítulo de livro publicado pelo autor, durante a obtenção do título de doutor, e se encontra no ANEXO 1, no final desta tese (ver Feltes et al., 2016).**

Posteriormente, outras proteínas foram relatadas por serem ligadas às proteínas XP, expandindo ainda mais sua importância na manutenção do genoma (Shell & Zou, 2008). Seguindo a ordem de explicação pela via de NER, XPC foi relatada por se ligar à ATM, uma proteína chave na parada de ciclo celular, assim como à TDG, uma glicosilase que remove danos ao DNA do tipo *mismatch*, na via de reparo por excisão de bases (do inglês, *base excision repair* – BER) (Colton, et al., 2006, Shimizu, et al., 2003). XPE interage com p300 e o complexo STAGA, ambos com funções de histona acetilase, sendo essenciais para o relaxamento da cromatina (Datta, et al., 2001, Kulaksiz, et al., 2005). Outro estudo reporta sua ligação com E2F1, um cofator de transcrição ligado à parada de ciclo celular (Hayes, et al., 1998). Tanto XPB quanto XPD fazem parte do complexo multiproteico TFIIH, interagindo, assim, com várias outras proteínas. XPB liga-se as subunidades p8, p44, p52 e p62 do complexo, enquanto XPD cdk7, ciclina H e MAT1 (Coin, et al., 2006, Hall, et al., 2006, Sandrock & Egly, 2001). Ambas as proteínas interagem com a proteína p53, que inativa suas atividades, e RAD52, conectando NER com a via de reparo por recombinação homóloga (Leveillard, et al., 1996, Liu, et al., 2002). XPA foi observada por interagir com ATR, uma proteína ligada à parada de ciclo que auxilia a translocação de XPA para o núcleo, à XAB1, que, por sua vez, possui a mesma função de ATR no auxílio da translocação de XPA, e XAB2, que liga XPA a RNA pol II na via TC-NER (Nakatsu, et al., 2000, Nitta, et al., 2000, Wu, et al., 2007). No último estágio da via, XPF interage com RAD51 e 52, assim como com FANCA, uma proteína envolvida com a via de pontes intercadeias de DNA (do inglês, *Interstrand crosslink repair* – ICLR), e TRF2, uma proteína essencial para a proteção de telômeros (Motycka, et al., 2004, Sridharan, et al., 2003, Zhang, et al., 2005). Por fim, XPG foi reportada por interagir com PCNA, e NTH1, uma proteína que também faz parte da via de BER (Bessho, 1999, Gary, et al., 1997).

Ter elucidado a via de NER foi fundamental para compreender a função de cada proteína XP no reparo de DNA e dar início aos estudos mais profundos sobre suas estruturas.

1.4. Estrutura das proteínas XP e DDB1

Tendo esclarecido suas manifestações clínicas, os caminhos que levaram ao seu descobrimento e suas funções na via de NER, torna-se essencial explicar como as proteínas XP se comportam em nível estrutural. Ainda que a doença venha sendo estudada há mais de 100 anos, é surpreendente que os estudos em nível de estrutura de proteína sejam escassos. As mutações nos genes XPA-XPG geram proteínas defeituosas ou sem função, e estudar sua estrutura é um passo crítico no entendimento da doença.

Nesse sentido, o Capítulo 1 (pag. 35) consiste em uma revisão sobre estruturas das proteínas XP. No presente momento, o artigo é a mais atual revisão do assunto, abordando todas as mutações que ocorrem naturalmente em indivíduos com XP, assim como as artificialmente induzidas com o propósito de estudo da estrutura. Ademais, a função das modificações pós-traducionais que ocorrem em certas proteínas XP são abordadas, e os resíduos críticos tanto para sua função quanto para a IPP são discutidos. Domínios de ligação com DNA e outras proteínas são mapeados juntamente com todos os resíduos atualmente discutidos, por serem essenciais para a função das proteínas.

Dentre essas proteínas, destaca-se a proteína derivada do gene XPE, DDB2. Atualmente, DDB2 é a única proteína XP com mais de 90% de sua estrutura mapeada, sendo um excelente alvo para análises estruturais, além de ser importante para entender como as mutações listadas para o subtipo E de XP podem afetar seu complexo com a proteína DDB1, sua principal parceira na via de NER, e afetar o reconhecimento do dano de DNA causado pela exposição à luz UV.

1.4.1. DDB1

A proteína DDB1 é composta por 1140 aminoácidos organizados em três estruturas *β-propeller* (BPA até BPC) em torno de um cluster de α-hélices (He, et al., 2006, Iovone, et al., 2011, Wittschieben & Wood, 2003) (**Fig. 7**).

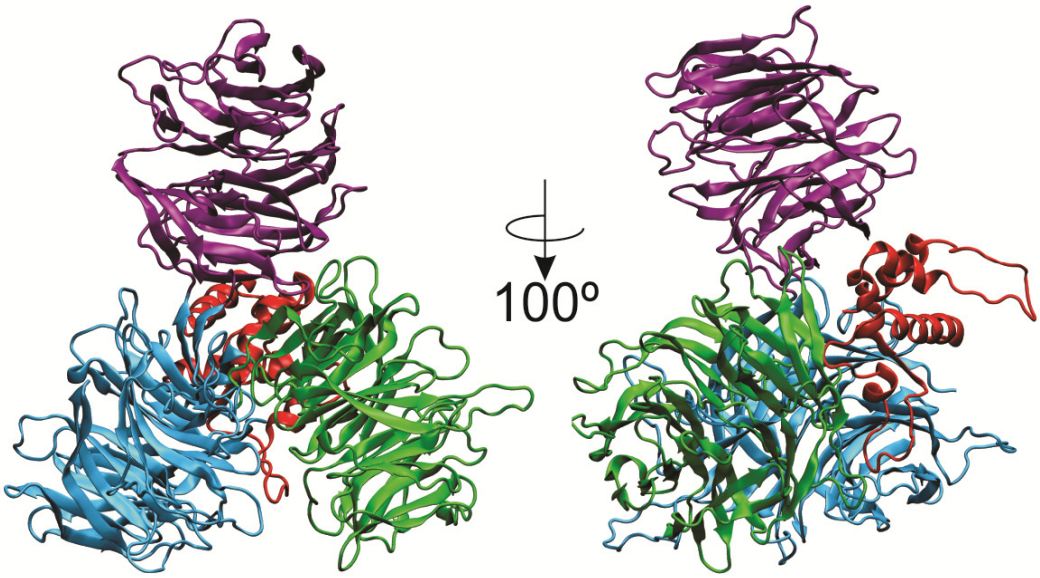


Fig. 7. Estrutura da proteína DDB1. Esta proteína é composta por três estruturas *β-propeller* dispostas em torno de um agrupamento de α-hélices. O BPA, BPB e BPC encontram-se coloridos em verde, azul e roxo, respectivamente. As estruturas de α-hélices são representadas pela cor vermelha. Esta figura foi modificada da **Fig. 1**, encontrada no Capítulo II, desta tese.

Como mencionado anteriormente, na via de NER, o papel da proteína DDB1 é formar um complexo com DDB2, CUL4A e ROC1, para formar um complexo de reconhecimento de dano de DNA, com maior especificidade para CPD (He, et al., 2006, Schärer, 2013). DDB1 é uma proteína citosólica que é translocada para o núcleo após a irradiação por luz UV, ligando-se a DDB2 e formando o complexo de reconhecimento (Wittschieben & Wood, 2003).

DDB1 não só atua na via de GG-NER, mas também de TC-NER, pois após sua translocação para o núcleo, ela pode se ligar não só a DDB2, mas também à proteína CSA, que é essencial para o começo do reparo de DNA na via de TC-NER (Wittschieben & Wood, 2003). Ela também já foi observada se ligando ao complexo multiproteico de

remodelação de cromatina STAGA, indicando que é uma proteína essencial para a manutenção do material genômico (Wittschieben & Wood, 2003).

Dentro do complexo de reconhecimento DDB, a proteína DDB2 é a efetiva reconhecedora do dano ao DNA, contudo, ela necessita de DDB1 para se ligar devidamente à dupla hélice (He, et al., 2006, Iovone, et al., 2011, Schärer, 2013, Wittschieben & Wood, 2003), fazendo com que ela seja igualmente essencial nos primeiros passos da via de NER quanto DDB2.

Levando esses fatos em consideração, para entender como as mutações encontradas em pacientes com XP tipo E afetam o reconhecimento do dano pelo complexo DDB, não basta apenas estudar a proteína DDB2, mas também seu impacto no heterodimero DDB2-DDB1.

1.5. Dinâmica molecular

Dispor do conhecimento da estrutura das proteínas XP é crítico para o estudo da doença. Contudo, estudar as estruturas em si é de um desafio completamente diferente. No Capítulo 1 veremos que dezenas de estudos foram realizadas com todas as proteínas XP, porém, por mais que eles evidenciassem o efeito final das mutações que ocorrem naturalmente nos indivíduos afetados pela patologia, eles não explicam como ocorrem ou como as proteínas estão se comportando. Interações entre aminoácidos, energias de ligação, perda ou ganho de estruturas secundárias, dobramento de proteínas e complexização de biomoléculas, entre outros aspectos da biologia estrutural de proteínas são fenômenos dinâmicos que impactam em sua função.

Nesse sentido, uma das técnicas amplamente utilizadas no estudo do comportamento de estruturas proteicas é a Dinâmica Molecular (DM). A DM analisa a variação do comportamento molecular de uma biomolécula em função do tempo (Verli, 2014). Através dessa técnica, todas as mudanças de orientação e interação entre resíduos, conteúdo de estruturas secundárias e informações sobre proteínas e suas interações com outras biomoléculas são analisados em função de um tempo pré-definido (Verli, 2014). Essas propriedades conformacionais analisadas pela DM são baseadas na segunda Lei de Newton, também chamada de Princípio Fundamental da Dinâmica, conforme a equação (1):

$$(1)\ Fx_i = \frac{d^2x_i}{dt^2}m_i = \frac{\Delta v_i}{\Delta t}m_i = a_i m_i$$

Onde Fx_i é a força aplicada no átomo *i* na posição *x*, *t* corresponde ao tempo, *v* equivale à velocidade e a_i a aceleração do átomo *i*. Dessa forma, a força empregada em cada átomo muda em função de sua posição e da posição dos outros átomos com que ele interage (Leach, 2001a).

Uma análise por DM incorpora uma diversidade de fenômenos estruturais cujo nível de informação não é possível na visão estática da proteína obtida por cristalografia, como a flexibilidade, obtida através da variação temporal de propriedades físico-químicas, e a temperatura, resultante da aceleração dos átomos (Karplus & McCammon, 1983, Verli, 2014). Dois tipos principais de simulações computacionais são empregadas para o estude do comportamento de biomoléculas: a mecânica quântica e a mecânica clássica (Monticelli & Tieleman, 2013).

A mecânica quântica permite a análise de todos os aspectos das interações moleculares que podem ser obtidas. Contudo, esse método possui um custo computacional elevado, limitando suas aplicações. Uma forma alternativa de análise do comportamento de biomoléculas é pela mecânica clássica, através do uso de um conjunto de equações matemáticas, chamado campos de força (Schlink, 2010, Verli, 2014). Campos de força, ao contrário da mecânica quântica, ignoram as movimentações eletrônicas de um sistema e baseia-se apenas na localização atômica, o que diminui consideravelmente o tempo computacional (Leach, 2001b). Em outras palavras, campos de força partem do princípio de que a energia de qualquer sistema pode ser interpretada pela soma de diferentes potenciais físicos, como deformação de ligações, eletrostáticas, ângulo de ligações, entre outras. Adicionalmente, campos de força partem do princípio de que potenciais de energia aplicados a um pequeno grupo de moléculas com grupos químicos similares podem ser aplicados a um vasto conjunto de moléculas com as mesmas características (Gunteren, et al. 2006, Monticelli & Tieleman, 2013). Independentemente de suas peculiaridades, os campos de força partem de um cálculo de energia potencial como descrito nas equações (2) até (6), para interações físicas, e na equação (7), para interações eletrostáticas:

(2) Ligações $= k_b(b - b_0)^2$

(3) Ângulos $= k_\theta(\theta - \theta_0)^2$

(4) Torções $= k_\emptyset[\cos(n\emptyset + \delta) + 1]$

(5) Pares não diretamente ligados $= \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right]$

Onde *b* é a distância interatômica, k_b e b_0 representam a rigidez e o equilíbrio de distância da ligação, θ é o ângulo formado por dois vetores de ligação, e os valores de θ_0 e k_0 avaliam a rigidez e o equilíbrio geométrico do ângulo. O potencial de torção no equilíbrio é calculado por uma função cujo gráfico é uma cossenóide, onde ϕ representa o ângulo torcional, δ a fase, e *n* o potencial diédrico (Vlachakis et al., 2014).

Levando em consideração que todas essas variáveis são necessárias para determinar o grau de movimentação de cada átomo e, consequentemente, da molécula, definimos a energia potencial resultante como a soma dos somatórios de cada um dos parâmetros previamente definidos (equação (6)):

(6) $V(r) = \sum Ligações \ + \sum Ângulos + \sum Torções + \sum Paresnão\ diretamente\ ligados$

(7) $V(r) = \sum Pares\ não\ diretamente\ ligados \left(\varepsilon_{ij}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2 \times \left(\frac{R_{min,ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}}\right)$

A última equação diz respeito a ligações eletrostáticas, em que ε_{ij} é um parâmetro baseado nas interações entre um átomo *i* e *j*, q_i e q_j representam as cargas efetivas nesses dois átomos e $R_{min,ij}$ indicam a distância onde a equação de Lennard-Jones, que descreve a interação entre um par de átomos neutros, está em seu mínimo (Guvench & MacKerell, 2008, Vlachakis et al., 2014).

Por fim, a equação simplificada final da energia total é descrita na equação (8):

(8) $\sum Total = \sum Ligações\ físicas + \sum Sem\ interação + \sum Outros$

Na equação acima, $\Sigma outros$ se refere a interações de repulsão, van der Walls e de Coulomb.

Somando todos esses fatores, a DM permite a observação de processos a nível atomístico e molecular em um nível de detalhamento não fornecido por outras técnicas experimentais. As simulações por DM também permitem um controle do ambiente experimental, como temperatura, pressão, pH, entre outros (Vlachakis et al., 2014).

Uma das áreas cujo interesse por análises de DM tem crescido é a medicina personalizada. Nesse sentido, ela tem acrescentado informações pertinentes na detecção de pequenas mudanças na estrutura e função de ácidos nucleicos, SNPs e proteínas de interesse médico, auxiliando no descobrimento de novas drogas e terapias (Sneha & Doss, 2016). Entender como as proteínas se comportam em nível estrutural em resposta ao ambiente ou em resposta à ligação com outras proteínas também é fundamental para analisar o impacto de mutações em sua estrutura (Goh, et al., 2004, Sneha & Doss, 2016).

1.6. Redes Dinâmicas

Um sistema pode ser definido como um conjunto de elementos interdependentes que, de maneira direta ou indireta, atuam juntos para formar um todo. Todo sistema possui uma função a ser cumprida, e cada elemento que o forma tem um impacto menor ou maior no exercício dessa função. Nos últimos anos, uma das formas mais utilizadas para o estudo de sistemas complexos é a análise de redes. Todo e qualquer sistema pode ser observado pelas partes individuais que o compõe e como estão conectadas entre si, sendo, assim, passíveis de uma análise de redes. Essa análise tem sido empregada na análise de IPP (Barabási & Oltvai, 2004), interações sociais (Palla et al., 2007), relações entre doenças humanas (Goh, et al., 2007), características e distribuição da internet (Yook, et al., 2002), sinalização celular (Wuchty, et al., 2006), alvos de drogas de interesse médico (Yildirim, et al., 2007), diagnóstico clínico (Gustafsson, et al., 2014), e até mesmo diversidade de temperos culinários (Ahn, et al., 2011). Proteínas também são passíveis da análise de redes, pois são sistemas compostos por aminoácidos interligados entre si, cuja interação resulta em uma estrutura dinâmica, tridimensional, que exerce diversas funções celulares (Amitai et al., 2004, Greene & Higman, 2003, del Sol & O'Meara, 2005, del Sol, et al., 2006, Taverna & Goldstein, 2002, Vendruscolo et al., 2002).

As redes ou grafos de interação são constituídos de dois elementos principais: (i) vértices ou nós (*nodes*) (**Fig. 8**), que representam as partes interagentes do sistema (Barabási & Oltavai, 2004; Newman, 2003) e, (ii) conectores (*edges*) (**Fig. 8**), que indicam o tipo de conexão entre os nós (Barabási & Oltavai, 2004; Newman, 2003). Na visão clássica da análise de redes para estruturas de proteínas, o nó é representado por um aminoácido, enquanto o conector corresponde ao tipo de ligação entre eles, podendo representar ligações de hidrogênio, ligação iônica, pontes salinas, atração de van der Walls, dentre outras (Doncheva, et al., 2011 e 2012).

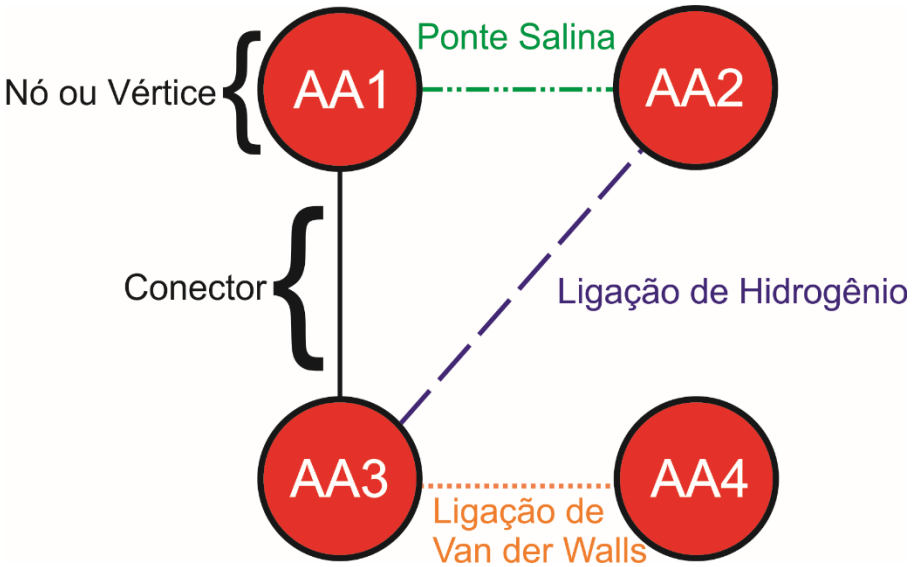


Fig. 8. Exemplo de rede voltada para estrutura de proteína em sua visão clássica. Cada nó representa um aminoácido (AA) hipotético, e cada conector indica um tipo e conexão.

A visão clássica de análises de redes de estrutura de proteínas é útil para analisar pontos críticos de IPP, aminoácidos estruturalmente essenciais para a estrutura da molécula e até mesmo no dobramento da proteína (Doncheva, et al., 2011 e 2012, Vendruscolo et al., 2002). Contudo, proteínas são sistemas dinâmicos que mudam em função do tempo e, graças a esse ponto, a visão clássica possui uma limitação. Ela não leva em consideração o tempo, apenas um momento estrutural específico bidimensional. Como mencionado anteriormente, as proteínas possuem uma característica flexível, movimentando-se em relação ao tempo, e diversas conformações são possíveis dependendo do momento observado. Levando esse fato

em consideração, para analisar um sistema derivado de uma DM, uma abordagem que abranja o tempo como variável é necessária.

Dessa forma, as chamadas Redes Dinâmicas (RD), fornecem uma solução para a limitação da visão clássica da análise de redes para estrutura de proteínas. No entanto, nessa análise, o nó se refere ao carbono alfa de cada aminoácido, e o conector representa a correlação de movimento entre os aminoácidos ao longo do tempo de simulação (**Fig. 9A-B**). Quanto maior a correlação de movimento entre os nós, mais grosso é o conector, enquanto uma correlação baixa é representada por um conector fino. A análise de RD também permite que se observem comunidades (**Fig. 9C**). A análise de comunidades é baseada em outro parâmetro comum na análise de redes, os chamados módulos. O parâmetro de modularidade se baseia em um princípio de agrupamento entre partes individuais de um sistema, ou seja, na formação de regiões altamente conectadas (Wagner *et al*., 2007). Esse princípio de união é observado nas relações entre proteínas na formação de bioprocessos, agrupamento de tópicos de internet, associação entre indivíduos para formar grupos sociais, entre outros (Newman, 2003 e 2006). Na análise de RD, as comunidades são grupos de aminoácidos que possuem uma alta correlação de movimento entre si, ao longo do tempo simulado. Na visão de redes de estrutura de proteínas, essas comunidades podem indicar domínios e regiões de interesse para IPP (Sethi, et al., 2009).

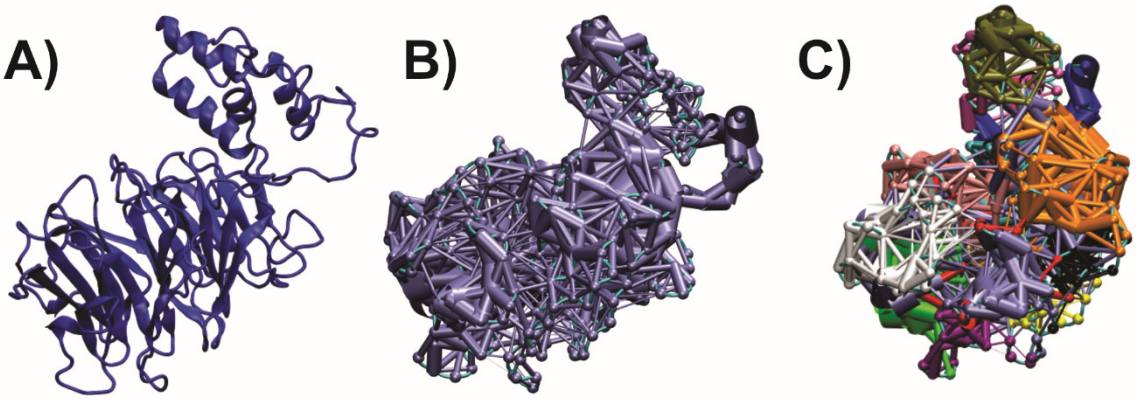


Fig.9. Visão da análise de RD. **A)** Visão canônica tridimensional da proteína DDB2 (alvo deste estudo). **B)** A estrutura tridimensional da proteína é convertida para uma visão de redes, onde cada nó refere-se ao carbono alfa de um aminoácido, e cada conector representa a correlação de movimento entre eles. Conectores grossos

indicam maior correlação de movimento entre os aminoácidos durante um tempo de simulação, enquanto um conector fino representa o oposto. **C)** Comunidades de aminoácidos com alta correlação de movimento entre si. Cada cor representa uma comunidade que apresentou uma alta correlação de movimento durante a simulação.

Dessa forma, RD permitem compreender o comportamento molecular de uma proteína de uma maneira distinta de outras análises computacionais, possibilitando observações pontuais sobre a correlação de movimento entre aminoácidos e a possível perda ou ganho de elementos estruturais.

2. OBJETIVOS

2.1. Objetivo geral

Estudar os impactos de três mutações encontradas em pacientes com Xeroderma Pigmentosum tipo E, no comportamento molecular da proteína DDB2, assim como no complexo DDB2-DDB1.

2.2. Objetivos específicos

- Obter e modelar a estrutura tridimensional das proteínas DDB2 e DDB1.
- Construir três variantes mutantes da proteína DDB2: $DDB2^{R273H}$, $DDB2^{L350P}$ e $DDB2^{K244E}$.
- Analisar e comparar o comportamento molecular da proteína DDB2 selvagem e suas variantes mutantes.
- Analisar o comportamento molecular da proteína DDB1.
- Analisar o complexo DDB2-DDB1 selvagem, assim como o mesmo complexo contendo cada uma das variantes mutantes de DDB2, e comparar os diferentes impactos de cada mutação em sua estrutura e interação proteína-proteína.
- Analisar cada sistema obtido anteriormente através de redes dinâmicas para maior compreensão dos efeitos de cada mutação na estrutura de DDB2 e do complexo DDB2-DDB1.

CAPÍTULO I

Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications

Artigo publicado na revista *Mutation Research Reviews*

Mutation Research 763 (2015) 306–320

Contents lists available at ScienceDirect

Mutation Research/Reviews in Mutation Research

journal homepage: www.elsevier.com/locate/reviewsmr
Community address: www.elsevier.com/locate/mutres

Review

Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications

Bruno César Feltes, Diego Bonatto *

Biotechnology Center of the Federal University of Rio Grande do Sul, Department of Molecular Biology and Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form 8 December 2014
Accepted 9 December 2014
Available online 16 December 2014

Keywords:
Xeroderma pigmentosum
Protein structure
TC-NER
GG-NER
DNA repair
Post-translational modifications

ABSTRACT

The xeroderma pigmentosum complementation group proteins (XPs), which include XPA through XPG, play a critical role in coordinating and promoting global genome and transcription-coupled nucleotide excision repair (GG-NER and TC-NER, respectively) pathways in eukaryotic cells. GG-NER and TC-NER are both required for the repair of bulky DNA lesions, such as those induced by UV radiation. Mutations in genes that encode XPs lead to the clinical condition xeroderma pigmentosum (XP). Although the roles of XPs in the GG-NER/TC-NER subpathways have been extensively studied, complete knowledge of their three-dimensional structure is only beginning to emerge. Hence, this review aims to summarize the current knowledge of mapped mutations and other structural information on XP proteins that influence their function and protein–protein interactions. We also review the possible post-translational modifications for each protein and the impact of these modifications on XP protein functions.

Contents

1. Introduction

Exposure to a range of environmental hazards, such as ultraviolet (UV) light and various toxic pollutants, can induce the formation of bulky DNA lesions. Among these bulky DNA lesions, both 6-4 photoproducts and cyclobutane pyrimidine dimers (CPDs) represent the most common UV-induced nucleotide damage. 6-4 photoproducts are formed from the reaction between

* Corresponding author at: Centro de Biotecnologia da UFRGS – Sala 219, Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul – UFRGS, Avenida Bento Gonçalves 9500 – Prédio 43421, Caixa Postal 15005, Porto Alegre 91509-900, RS, Brazil. Tel.: +55 51 3308 6080; fax: +55 51 3308 7309.
E-mail address: diegobonatto@gmail.com (D. Bonatto).

http://dx.doi.org/10.1016/j.mrrev.2014.12.002
1383-5742/

the C5–C6 bond of a 5′-pyrimidine residue and the C4 carbonyl group of a 3′-thymine, while CPDs are induced by the cycloaddition reaction of adjacent pyrimidine bases [1,2].

To protect the genome against these bulky DNA lesions, different molecular pathways were evolutionarily selected in eukaryotes, including DNA repair mechanisms such as nucleotide excision repair (NER) [1].

The NER pathway is divided into the transcription-coupled repair (TCR) and global genome repair (GGR) subpathways [1–4]. TCR is responsible for the repair of bulky DNA lesions from transcribed strands of active genes, whereas GGR is associated with the removal of DNA damage from transcribed strands of active genes, heterochromatin areas and non-active genes [1–4].

Although several proteins in complex are required to recognize bulky DNA lesions and promote NER, the xeroderma pigmentosum complementation group proteins (XPs) play critical roles. XPs include seven proteins termed XPA through XPG [1,4–6]. These proteins coordinate and promote the following basic steps in the NER pathway: (i) recognition of the distortion in the DNA molecule caused by a bulky nucleotide lesion, (ii) assembly of the repair machinery (the "preincision complex"), (iii) excision of the damaged nucleotide *via* dual incision, and (iv) synthesis of a new DNA fragment followed by ligation of the synthesized fragment with the pre-existing DNA molecule [6]. Additionally, XP proteins interact with a variety of proteins critical for genome organization and maintenance that are not directly involved in NER (reviewed in [4]).

Due to their critical role, mutations in XP proteins are commonly found in various cancer types and are responsible for a rare autosomal recessive disorder known as xeroderma pigmentosum (XP). Affected individuals exhibit a high predisposition for development of skin cancers and neurological degeneration, among other clinical symptoms; although not all XP patients exhibit the same effects [7]. Defects in NER are also related to Cockayne Syndrome (CS) and trichothiodystrophy (TTD), which are two conditions that lead to progeria [8]. The symptoms related to XP can be mild or severe, where severe affected XP patients may display combined phenotypes with CS or TTD diseases, sometimes leading to premature aging phenotype, developmental impairments and severe neurological abnormalities [9].

Considering the importance of XP proteins in NER and that their dysfunction results in XP disease, an understanding of the structure of XP proteins is necessary. However, few studies have examined the available information concerning the impact of mutations on XP proteins function and how post-translational modifications affect the function of these XP proteins.

Therefore, we aim to review the current knowledge of the structure and post-translational modifications of XP proteins, as well mutations that may affect their function and protein–protein interactions (PPIs). A review of the primary function of each XP is beyond the scope of this review, as this topic has previously been broadly reviewed (see [1,6]). In addition, we focus on the structure of XP proteins from XPA to XPG, not considering the XP variant DNA polymerase eta (XP-V).

2. The pathfinders: XPC and the UV–DDB complex

2.1. Mutations that affect XPC structure and critical regions

The first step of the GG-NER pathway is the recognition of bulky distortions in DNA promoted by the tight interaction between XPC and RAD23B. In general, XPC patients show no severe phenotypes, since they do not develop neurological abnormalities or exaggerated sunburn, but they can develop skin cancers [9]. However, a report of a male patient showing abnormal neuronal features and strong cell carcinomas and melanomas was reported by Khan et al. in 1998 [10].

Human XPC is a 940-residue DNA-binding protein that exhibits high affinity for 6-4 photoproducts and *N*-acetoxy-2-acetylaminofluorene adducts [11]. XPC forms a heterodimer with RAD23B, which is a multiubiquitin chain receptor, to recognize bulky DNA lesions in GG-NER [11]. It also forms a heterodimer with centrin 2, which is an essential protein for chromosome segregation and microtubule organization [12]. Centrin 2 is an important stabilizer of the XPC-RAD23B heterodimer and assists in XPC DNA binding [13,14] (Fig. 1; Table 1).

XPC is composed of a C-terminal region (residues 492–940) that contains the major PPI sites and the DNA-binding domain and an N-terminal region (residues 1–491) that contains an XPA-binding region (Fig. 1) [15]. To date, the first study to identify XPC protein structure was accessed using the XPC yeast orthologue Rad4 [16]. Rad4 was crystalized bound to a CPD lesion and the authors observed three 50–90 residues structurally related α/β domains characterized by a long β-hairpin (BHD1-BHD3) [16]. BHD1 is responsible for binding undamaged DNA, whereas BHD2 and BHD3 bind to the CPD lesion [16]. In humans, BHD1 corresponds to the regions 632–689, BHD2 to region 690–741 and BHD3 to region 742–833 (Fig. 1). The authors identified numerous residues that are responsible for Rad4 DNA-binding and are described in Table 1 and Fig. 1.

Site-directed mutagenesis of four conserved residues (Phe799Ala, Phe797Ala, Phe756Ala and Asp754Ala) in human XPC was used to evaluate DNA binding and DNA damage recognition [17] (Table 2). Nearly all of the heterologously expressed XPC mutants exhibited a reduced ability to bind undamaged DNA, with the exception of XPC (Phe799Ala), which retained DNA-binding properties [17]. These four mutants also exhibited decreased levels of accumulation at UV-damaged sites compared with wild-type (WT) XPC. FRAP measurements reveals that these mutations restricted XPC nuclear mobility and prevented the formation of stable recognition complexes [17].

XPC recognition of DNA damage is the first step of GG-NER and requires the subsequent recruitment the of the TFIIH complex to promote DNA denaturation and the assembly of the preincision complex [1,4,6]. Remarkably, the XPC (Phe797Ala) and XPC (Phe799Ala) mutants exhibited a decreased ability to recruit the TFIIH complex [17].

Another study utilized site-directed mutagenesis and DNA-binding assays to assess the role of different residues in XPC function [18]. The XPC (Phe733Ala), XPC (Trp690Ser) and XPC (Trp690Ala) mutants were unable to bind DNA lesions [18]. Other XPC mutations generated using site-directed mutagenesis did not affect the ability of XPC to bind DNA lesions. These authors also identified regions in XPC that play a critical role in the recognition of UV-damaged DNA regions by generating truncation mutants (Table 1; Fig. 1).

Based on the analysis of truncated XPC proteins, these authors also identified a 25-residue region (residues 742–766) that corresponds to a β-turn near BHD3 that appears to be crucial for XPC activity. Generation of truncation mutants ($XPC_{607–741}$ and $XPC_{607–766}$) confirmed that this region is necessary for XPC recruitment to UV-damaged sites [18]. Another study demonstrated that residues 606–742 constitute the primary region required for DNA binding, which is consistent with other findings that indicated that mutations inside and near this region lead to a loss in the ability of XPC to bind DNA [15]. Moreover, several other regions of XPC were proposed to be PPI sites [15] (Table 3 and Fig. 1).

Other mutations in the *XPC* gene were found in lung and colorectal cancer patients and in XP patients (Table 2; Fig. 1) [19–21]. Interestingly, a majority of the mutations found in XPC are located in the PPI and DNA-binding regions.

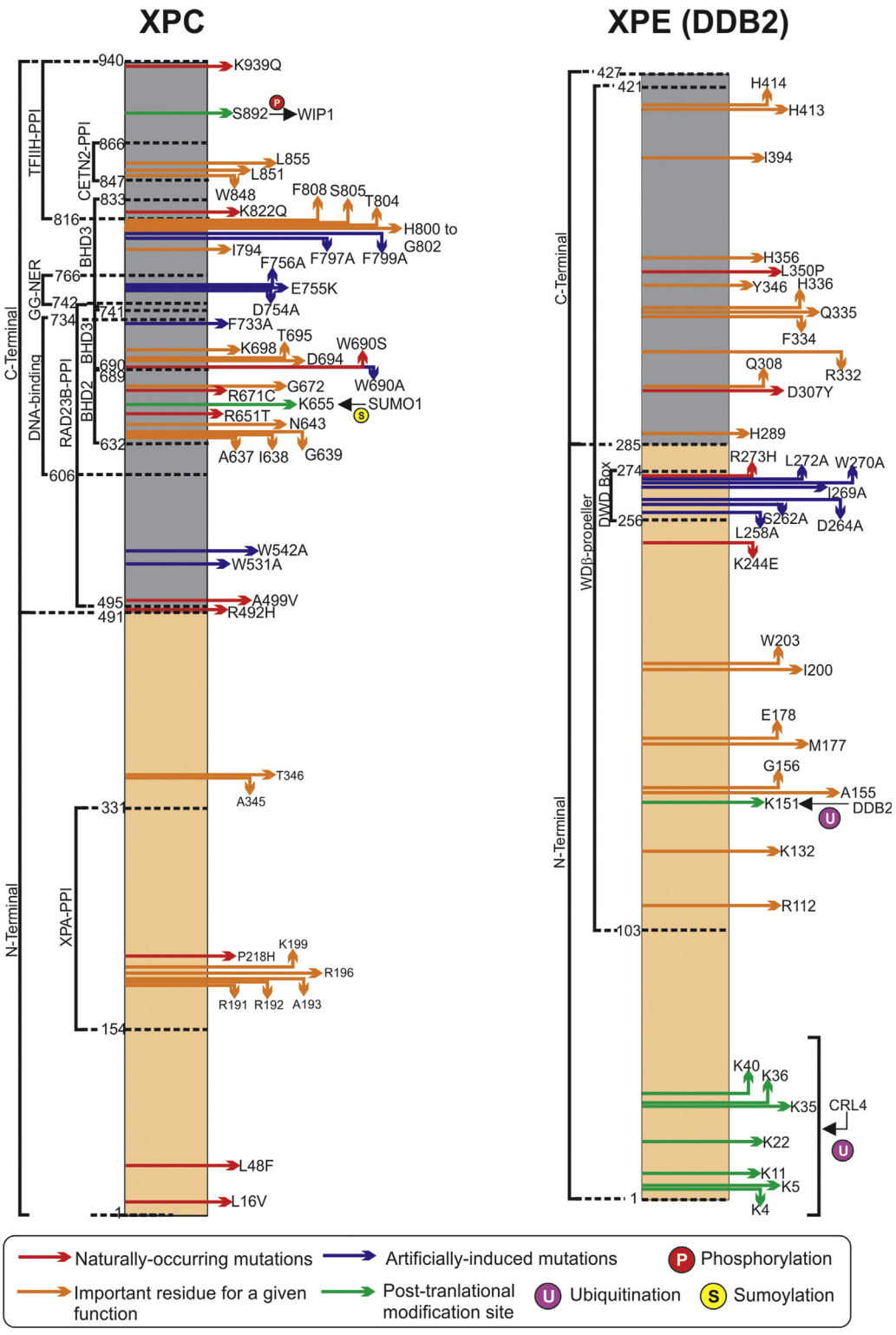


Fig. 1. Schematic model showing regions necessary for protein interaction, important amino acids residues, post-translational sites, known naturally occurring and artificially-induced mutations, and protein domains of XPC (A) and XPE (DDB2) (B). It is necessary to observe that "important amino acid residues" are all those needed for the maintenance of XP proteins-associated function and structure. Substitution of those amino acids by natural-occurring mutations or by artificial mutagenesis protocols is indicated in the figure. All domains were defined by taking into account the data present in scientific literature. Deleted residues indicate specific regions in the protein that were excluded for functional studies. For information about the role of each residue or region, please address Tables 1-3 and the main text of the manuscript.

Additional studies on the XPC structure are necessary to understand the specific residues required for PPIs, particularly between XPC and known binding partners, such as RAD23B, the TFIIH complex and the DDB complex [1,4,6].

2.2. Post-translational modifications in XPC

Of the known XP proteins, XPC is the only member for which post-translational modifications have been well characterized.

Table 1
A summary of the important residues in each XP protein. This includes residues important for function and protein–protein interaction (PPI). The residues in parenthesis are those that correspond to the human XP.

XP	Residue and relevance[a]
XPA	**Cys105, Cys108, Cys126, Cys129**: residues that coordinate Zn^{2+} in XPA and required for DNA binding [61,103]; **Ala229 Trp235 and Lys236**: described to be necessary for XPA DNA-binding [66]; **Glu147, Tyr148, Leu149, Leu150, Lys151, Asp152, Cys153, Asp154, Leu162, Lys163, Phe164, Ile165, Lys168, Asn169, Pro170 and His171**: residues that, when deleted, promote reduced XPA binding to RPA [67]; **Gly72, Gly73 and Gly74**: insert into a V-shaped groove in XPF; necessary for the XPA–XPF interaction [68]; **Asp70, Thr71, Phe75, Ile76 and Leu77**: residues that interact with ERCC1 to stabilize DNA binding [68]; **Glu78-Glu84**: involved in XPA-ERCC1 binding and cell survival [69]; **Residues 163–167 (APIM region)**: evolutionarily conserved region that is necessary for XPA binding to PCNA [71]; **Residues 4–97**: involved in the CEP164–XPA interaction [73].
XPB	**Arg11 (Ser73), Arg30 (Tyr121) and Arg35 (Ala125)**: are involved in Af-XPB binding to DNA [37]; **Lys264, Lys267, Lys285, Arg271, Arg278, Arg279 and Arg307 (Lys529)**: sequence-independent DNA binding of Af-XPB [37]; **Tyr163 (Tyr410), Glu167 (Trp429), Lys168 (Leu430), Glu187 (Lys449), Gln191 (Arg453), Gln194 (Tyr456) and Arg214 (Lys476)**: are located in HD1 and may be responsible for the Af-XPB protein–DNA interaction [37]; **$XPB_{198–305}$**: region described to interact with the C-terminal domain of XPG [85]; **$XPB_{305–387}$**: region described to interact with the N-terminal domain of XPG [85].
XPC	**Trp848, Leu851 and Leu855**: essential for anchoring centrin2 to XPC [13] and [14]; **Arg129 (Arg191), Asn130 (Arg192), Val131 (Ala193), Asn134 (Arg196), Arg137 (Lys199), Met194 (Ala345), Lys195 (Thr346), Ser437 (Ala637), Val438 (Ile638), Gln439 (Gly639), Asn443 (Asn643), Val471 (Gly672), His472, Lys477, Arg494 (Asp694), Gln495 (Thr695), Met498 (lys698), Asn554 (Asn754), Phe556 (Phe756), Val594 (Ile794), Phe597 (Phe797), Phe599 (Phe799), Glu600 (His800), Arg601, Gly602, Ser603, Thr604, Vel605 and Pro607**: residues related to DNA-binding in the XPC yeast orthologue Rad4 [16]; **$XPC_{154–331}$**: are involved in XPC–XPA binding [15]; **$XPC_{606–742}$**: proposed region for DNA binding [15]; **$XPC_{495–734}$**: proposed region for RAD23B binding [15]; **$XPC_{816–940}$**: proposed region for TFIIH binding [15].
XPD	**Cys88 (Cys116), Cys102 (Cys134), Cys105 (Cys157) and Cys137 (Cys192)**: residues that coordinate the 4FeS cluster in Sulfolobus acidocaldarius [51]. The H. sapiens corresponding residues in parenthesis, can be found in [52]; **$XPD_{1–162}$**: necessary for PPI with the XPG C-terminal domain [85]; **$XPD_{324–434}$**: necessary for PPI with the XPG N-terminal domain [85].
XPE (DDB2)	**Phe371 (Phe334), Gln372 (Gln335), Pro191 (Ala155), Tyr393 (Tyr157) and His373 (His356):** responsible for recognition of DNA damage, DNA binding and displacement of CPDs and 6-4 photoproducts from double-stranded DNA [1,27,29,30]; **Trp239 (Trp203), Trp236 (Ile200), Ile213 (Met177), Arg214 (Glu178) and Gly192 (Gly156):** responsible for DDB2 binding to CPDs and 6-4 photoproducts in DNA [30]; **Arg148 (Arg112), Arg369 (Arg332), Arg404 (Arg371), Lys168 (Lys132), Lys280 (Lys244), Tyr393 (Tyr346), Gln345 (Gln308), Ile428 (Ile394), Asn448 (His414), Phe447 (Tyr413) and Lys325 (His289):** residues necessary for zebrafish DDB2 interaction with the DNA phosphate backbone [28,30]; **Residues 256–274:** correspond to the DWD box [34]; **Trp239 (Trp203), Ile213 (Met177), Gly192 (Gly156), Asp237 (Asn201) and Arg214 (Glu178):** stabilize the interaction between CPD and the zebrafish DDB2 WD40 β-propeller [28,30].
XPF	**Asp676, Glu679, Asp704 and Glu714:** mutations of these residues to Ala indicates a role in the XPF metal-binding site [97]; **Tyr833, Pro837, Gln838, Asp389, Phe840, Leu841, Leu842, Lys843, Met844, Met856, His857, Asn861, Ile862, Ala863, Glu864, Leu865, Ala866, Ala867, Phe889, Ile890, His891, Thr892, Phe894 and Ala895**: residues involved in the XPF-ERCC1 PPI [98,99]; **Lys850, Arg853, Ser854, His857 and His858, Ser875, Ile876, Leu877, Gly878 and Asn879:** mediate the interaction between XPF and ssDNA [101];
XPG	**Glu791 and Asp812**: proposed to coordinate the Mg^{2+} ions in the XPG structure [79]; **$XPG_{1–337}$ (N-terminal and spacer region):** region described to participate in PPIs with XPB and XPD [81,85]; **$XPG_{747–928}$ (C-terminal and D2 box):** region described to participate in PPIs with XPB, XPD, p62 and p44 [81,85]; **$XPG_{668–747}$ (C-terminal and spacer region):** region defined to participate in the PPI with RPA [81,84]; **$XPG_{554–730}$ (spacer region):** deletion of this region results in diminished DNA bubble cleavage and generation of DNA excision products [81]. This deletion also results in decreased recruitment of XPG to UV-damaged sites and PPI with XPB [81]; **$XPG_{184–210}$ (spacer region-D1 box):** deletion of this region results in diminished DNA bubble cleavage and generation of DNA excision products [81]. This deletion also results in decreased recruitment of XPG to UV-damaged sites and PPI with XPB [81]; **$XPG_{225–231}$ (spacer region):** deletion of this region results in diminished PPI with TFIIH, impairments in NER and prevents the restoration of UV resistance [81]; **$XPG_{111–550}$ (spacer region):** deletion of this region results in diminished DNA bubble cleavage and completely abolishes the generation of DNA excision products and the PPI with the TFIIH complex [81]; **Residues 981–1009:** responsible for the PPI with PCNA [86].

[a] Structurally important residues that affect a given XP function or protein interaction.

Studies have shown that XPC is ubiquitinated (Table 3; Fig. 1) in a UV–DDB complex-dependent manner (following UV irradiation of a cell) [11,22]. The ubiquitination of XPC by the UV–DDB complex appears to facilitate the recruitment of XPC to 6-4 photoproducts and cyclobutane pyrimidine lesions sites, although this post-translational modification is not necessary for XPC to bind lesion sites on its own [22].

The interplay between XPC and the UV–DDB complex is necessary for the proper recognition of CPDs, as XPC alone exhibits a low affinity for these lesions compared with the UV–DDB complex, which exhibits a high affinity for CPDs [11,22]. The UV–DDB complex binds to the CPD adducts and recruits XPC-RAD23B to the site. The interaction between these two complexes is mediated by XPC polyubiquitination *via* the E3 ubiquitin ligase activity of CUL4A and ROC1 [22].

In addition to ubiquitination, sumoylation of XPC by SUMO1 upon UV irradiation has also been described [23]. Sumoylation protects XPC from degradation following exposure to UV radiation [23], indicating that both sumoylation and ubiquitination are necessary for stabilization of XPC during DNA repair. A study prospected the possible sumoylations sites of SUMO1 for XPC, and determined a mutation Lys655Ala affected the capability of XPC to

Table 2
A summary of the natural occurring and artificially-induced mutations in each XP protein.

XP	Residue and relevance[a]
XPA	**Cys261Ser/Cys264Ser**: double mutant that is incapable of binding the TFIIH-complex [64]; **Lys141Glu/Lys179Glu**: double mutant that exhibits a substantial decrease in DNA binding and DNA repair and fails to recruit the XPF–ERCC1 heterodimer [65]; **Arg207Gly**: exhibits reduced XPA binding to DNA and diminished XPA–DDB2 association [65,70]; **Lys145Gly/Glu, Lys151Gly/Glu and Lys204Gly/Glu**: mutations that result in diminished XPA binding to DNA [65]; **Phe75Ala**: abolishes DNA repair activity [68]; **His244Arg**: naturally occurring mutation in XP patients [105]; **Cys108Phe and Cys108Ser**: inhibit XPA binding to DNA [103,104].
XPB	**Arg210Ala, Glu211Ala and Asp212Ala**: greatly decrease Af-XPB helicase activity [37]. The corresponding residues in H. sapiens are Arg472, Glu473 and Asp474, respectively; **Lys346Arg, Glu473Ala and residues 516–526**: human XPB carrying these mutations are unable to bind to UV-damaged sites and repair damaged DNA, exhibit decreased ATPase activity or disrupt the helicase activity of TFIIH [39]; **Thr469Ala and Gln638Ala**: disrupt the helicase activity of XPB [39,42]; **Phe99Ser and Thr119Pro**: these mutations exhibit diminished NER activity. However, Phe99Ser exhibits a more severe defect than Thr119Pro [40] due to the loss of XPB interaction with the p52 subunit of the TFIIH complex [42]; **Lys359Glu and Ser382Pro**: result in decreased NER activity during CPD repair and decreased RNA synthesis following UV irradiation in Chinese hamster ovary cells [43]; **Thr232Ala, Lys449Arg and Ser704Leu**: missense mutations found in lung cancer patients [20]; **Lys346Arg**: impairs the recruitment of the human TFIIH complex to CPD sites [39].
XPC	**Asp754Ala, Phe756Ala, Phe797Ala and Phe799Ala**: DNA-binding defect, unable to bind single-stranded oligonucleotides in undamaged DNA (except Phe799Ala); these mutations result in low accumulation of XPC at UV-damaged sites; Phe797Ala and Phe799Ala result in the inability of XPC to recruit the TFIIH complex [17]; **Trp690Ser, Trp690Ala, Trp531Ala and Trp542Ala**: DNA-binding defect [18,106]; **Phe733Ala**: nearly complete loss of XPC binding to DNA [18]; **$XPC_{427–940}$ and $XPC_{607–940}$**: loss of XPC binding to DNA [18]; **$XPC_{1–495}$ and $XPC_{1–718}$**: loss of XPC recognition of DNA damage [18]; **Residues 742–766**: essential for XPC GG-NER activity [18]; **Glu755Lys**: exhibits reduced nuclear mobility [18]; **Ala499Val**: exhibits lower XPC levels [19]; **Leu16Val, Leu48Phe, Arg492His, Ala499Val, Arg651Trp, Arg671Cys and Lys939Gln**: missense mutations found in lung cancer and colorectal cancer patients [19,20]; **Lys822Gln and Pro218His**: missense mutations found in XP patients [21].
XPD	**Asp234Asn:** reduces DNA repair [49]; **Arg658Cys:** affects the thermostability of the TFIIH complex [54]; **Arg112His:** Mutation present in patients with XP and TTD that impairs DNA repair, most likely through defects in DNA binding, present in patients with XP and TTD [49,54,107]; **Gly675Arg and Arg722Trp:** are localized in HD2; impairs p44-XPD PPI; Gly675Arg also reduces HD2 stability [51,108]; **Arg683Trp/Gln, Lys48Arg, Arg666Trp and Gly47Arg:** inhibit ATP binding and ATP hydrolysis [51,108]; **Asp312Asn and Lys751Glu:** missense mutations found in lung cancer and melanoma patients [20,109]; **Cys259Tyr^{D}, Leu461Val^{A,D}, Arg487Gly^{B}, Arg511Gln^{B}, Ala594Pro^{B}, Arg616Pro/Trp^{A,C,D}, Arg658His/Cys^{D}, Cys663Arg^{B}, Gly713Tr^{D} and Ala725Pro^{D}:** mutations found in XP (marked with "A"), CS (marked with "B"), COFSS (marked with "C") and TTD patients (marked with "D") [49,54,55,107]; **Residues 716–730:** this region is found to be absent in a TTD patient [110]; **Asp681Asn:** proposed in a SaXPD model to affect DNA binding, helicase activity and possibly ATP binding [51]; Found as a mutation in a patient with COFSS [55]; **Ser541Arg, Tyr542Cys, Arg601Leu/Trp and Gly602Asp:** known XPD mutations [49,54,107]. In the SaXPD model, the original residues interact with ssDNA [51]; **Arg592Pro:** reduces HD2 stability [51]; **Arg592Pro and Asp673Gly:** decrease TFIIH stability and XPD helicase activity [51].
XPE (DDB2)	**Arg273His and Lys244Glu:** impair the binding of DDB2 to damaged DNA [111]. Arg273His also disrupts the interaction between DDB1and DDB2 [111]; **Leu258Ala, Ser262Ala, Asp264Ala, Iso269Ala, Trp270Ala, Leu272Ala and Arg273Ala**: abolish the DDB1-DDB2 PPI [34]; **Leu350Pro and Asp307Tyr:** appear to be involved in the DDB2–DDB1 interaction, generates an unstable DDB-complex [112].
XPF	**Ser3Leu, Ala13Ser, Gln21Glu, Phe379Ser, Arg415Gln, Leu401Phe and Phe6Ser:** missense mutations found in lung cancer patients [20]; **Asp676Ala, Arg678Ala, Asp704Ala, Arg715Ala, Lys716Ala and Asp720Ala:** mutations of these residues indicate that they are responsible for XPF nuclease activity [97]; **Lys716Ala:** inhibits the XPF catalytic site [65,97]; **Arg689Ser and Leu230Pro:** missense mutations found in Fanconi anemia patients [113]. Arg689Ser exhibits abnormal endonuclease activity, whereas Leu230Pro exhibits residual excision activity and reduced mobility from the cytoplasm to the nucleus [113]; **Arg153Pro, Arg454Trp, Arg490Gln, Glu502Lys[b], Gly513Arg, Ile529Thr, Thr567Ala, Leu608Pro, Arg788Trp, Arg589Trp, Glu491Lys, Ile518Thr, Thr556Ala, Arg479Gln, Leu599Pro, Ile214Met, Gly502Arg and Arg443Trp:** mutations found in CS and XP patients [114–116]; **Leu801Pro:** impairs the XPF-ERCC1 PPI [102]; **Pro85Ser:** abolishes the XPF-RPA PPI [102]; **Cys236Arg:** a mutation found in a CS patient that exhibits reduced endonuclease activity [114].

Table 2 (*Continued*)

XP	Residue and relevance[a]
XPG	**Cys794Ala/Ser:** reduces endonuclease activity [79]; **Met254Val, Ala1119Val, Glu399Val, Asp1104His, Cys529Ser, Thr971Ala, Val597Leu, Arg819Trp and Gln622His:** missense mutations found in lung cancer patients [20]; **Pro72His:** a mutation found in an XP patient that impairs XPG endonuclease function [117]; **Leu858Pro:** a mutation found in an XP patient, that results in reduced endonuclease activity [118]; **Ala874Thr:** diminishes XPG activity to repair DNA damage in an XP patient [119]; **Ala792Val:** a mutation found in an XP patient XPG that eliminates XPG endonuclease activity and repair function [79] and results in undetectable 5′ incision, most likely by disrupting the PPI with XPF-ERCC1 [79]; **Asp812Ala:** abolishes XPG endonuclease activity [79,120]; **Leu65Pro:** impairs DNA repair [121]; **Asp77Ala/Glu and Glu791Ala/Asp:** these mutations abolish XPG endonuclease activity [79,81]. The substitutions to Ala also abolish DNA repair and result in decreased XPG activity in the repair bubble, whereas substitutions to Glu and Asp, respectively, do not exhibit activity in the bubble substrate [79]; **Arg992Ala and Arg992Glu:** necessary for XPG binding to PCNA [87]. Cells expressing these mutations were unable to rescue NER activity *in vivo* [87].

[a] The Table includes relevant XP mutants with altered functions, with the exception of mutant sequences generated by nonsense mutations. Site-directed mutations were also included because they examined the role of important residues in XP proteins. The primary amino acid sequences of XP proteins from organisms used as models were aligned using the online version of the program T-Coffee [tcoffee.crg.cat/apps/tcoffee/do:regular], and the corresponding residues for *H. sapiens* are accordingly provided. Figs. 1, 3–5 illustrate the differences between the observed mutations.

[b] A sequence conflict was detected for Gly502Arg and Glu502Lys when considering references [115] and [116], respectively. However, the primary sequence of XPF present in the NCBI entry for XPF contains Glu at position 502; therefore, we retained Glu502Lys in the representation.

resist degradation after UV-irradiation and impaired its ability to restore DNA-damage [24]. Interestingly, this residue is located within the BHD2 domain and RAD23B PPI region (Fig. 1).

XPC can be dephosphorylated at Ser892 by the wild-type p53-induced phosphatase 1 protein (WIP1/PPM1D), resulting in XPC inactivation [25]. WIP1 overexpression was found to inhibit CPD repair, whereas WIP1 null mice exhibited increased CPD repair and less UV-prone apoptosis [25]. However, how additional post-translational modifications alter XPC function remains unclear, although it appears that ubiquitination and sumoylation are required for XPC stabilization and protection against degradation. Phosphorylation appears to activate XPC, whereas dephosphorylation can inactivate it, strengthening the idea that phosphorylation is required for XPC activation during cell cycle checkpoints. In addition, it is interesting to observe that the phosphorylation site in XPC is in the same region described to interact with the TFIIH complex (Fig. 1). However, whether these modifications affect the physical interactions between XPC and HB23B or with other NER components is unknown.

2.3. Structure of XPE (DDB2) and post-translational modifications

The UV–DDB complex comprises the damage-specific DNA binding proteins 1 and 2 (DDB1 and DDB2), cullin 4A (CUL4A) and the Ras-like without CAAX protein (ROC1) [1] and [26]. The proteins CUL4A and ROC1 form a ubiquitin–ligase complex that is inhibited by the COP9 signalosome [1,26].

DDB1 does not directly interacts with damaged DNA; however, it is essential for bridging DDB2, which is a damage-recognition factor in the DDB-complex, to the other three subunits that hold the ubiquitin–ligase complex, CUL4A and ROC1 together [27].

DDB2 is a DNA-binding protein consisting of 427 residues and is the major DNA-damage recognition factor in the DDB-complex [1,6] (Figs. 1 and 2B). DDB2 plays a critical role in GG-NER; without DDB2, XPC-RAD23B is unable to bind CPD lesions, although it can still recognize 6-4 photoproducts [1]. Similar to XPC, XPE individuals show no severe phenotype, no neurological abnormalities, no serious sunburns, but can develop skin cancers [9]. This relation indicates that due to the fact that DDB2 and XPC are both proteins related to DNA-damage recognition, a possible compensation mechanism lies within this step in GG-NER. DNA-lesion could still be located by one of the proteins, thus a more severe phenotype is not developed.

A study using DDB2 isolated from *Danio rerio* indicated that this protein requires multiple residues to bind UV-induced DNA lesions [27]. DDB2 contains a WD40 β-propeller in which four residues are responsible for DNA binding [27–31] (Table 1; Fig. 1). A WD40 β-propeller is a circular arranged motif containing four-stranded

Table 3
A summary of the residues related to post-translational modifications in each XP protein.

XP	Residues
XPA	**Lys63 and Lys67**: XPA is deacetylated at these two residues by SIRT1, and this modification is necessary for optimal DNA repair. Acetylation of these two residues also impairs XPA–RPA binding [74]; **Ser173**: XPA is phosphorylated at this residue *in vitro*, but Ser173 phosphorylation does not appear to be as crucial as Ser196 phosphorylation *in vivo* [75]; **Ser196**: is phosphorylated by ATR and increases cell survival following UV-induced DNA damage [75]. XPA is dephosphorylated at this residue by WIP1, inactivating XPA [25].
XPB	**Ser751**: phosphorylation at this residue results in the loss of NER activity [44].
XPC	**Ser892**: dephosphorylation at this residue by WIP1 inactivates XPC [25]; **Lys655**: sumoylation of this residue protects XPC from degradation upon UV-irradiation [24].
XPD	**Lys701**: possible ubiquitination site [56].
XPE: DDB2	**Lys4Arg, Lys5Arg, Lys11Arg, Lys22Arg, Lys35Arg, Lys36Arg and Lys40Arg**: exhibit increased stability of human DDB2 against degradation, potential residues for ubiquitination by CRL4 [30]; **Lys151**: target for DDB2 autoubiquitination [30].
XPF	Not described
XPG	Not described

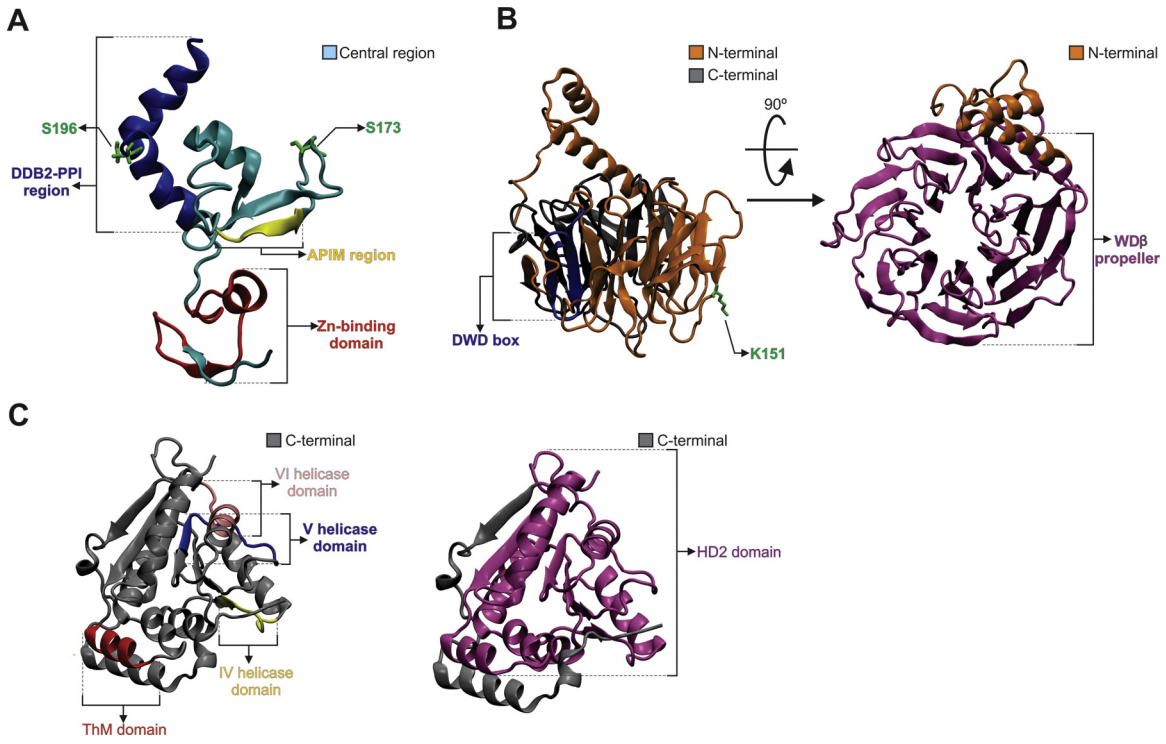


Fig. 2. Three-dimensional representation of the available structures of human XP proteins. The structures were colored according to the 2D representations of Figs. 1, 3–5. (A) Fragment corresponding to the central domain of XPA (PDB code 1XPA) showing the Zn-binding domain in red, the APIM region in yellow, the DDB2-PPI region in dark blue and the phosphorylation sites in green. (B) Structure of DDB2 extracted from the DDB2-DDB1 heterodimer (PDB code 3EI4). In the left side is a front-view representation of DDB2 showing the DWD box in blue and the ubiquitination site in green. In the right side is an upper-view representation of the WDβ-propeller region. (C) Fragment of a C-terminal region of XPB (PDB code 4ERN). In the left side is an upper-view representation showing the ThM domain in red, the IV, V and VI helicase domains in yellow, blue and pink, respectively. And in the right side is a representation of the HD2 domain region.

anti-parallel β-sheets that form a structure usually referred to as blades [32,33]. DDB2 possesses seven blades, but a WD40 β-propeller can have 6–10 blades [32].

The residues Gln372, His373 and Phe371 in the propeller insert into the DNA molecule and displace the CPD lesion for XPC recognition and promote GG-NER [29] (the equivalent residues in human DDB2 are summarized in Table 1). Additionally, five DDB2-associated residues stabilize the exposed CPD lesion inside the hydrophobic region of the WD40 β-propeller while eleven residues coordinate the DDB2 interaction with the DNA backbone [28,30] (Table 1; Fig. 1). Notably, DDB2 does not alter its conformation upon CPD binding.

The DWD (DDB1 binding and WD40 repeat) box, which is a small 16-residue region, is another important region in DDB2 [34]. This region is recognized by DDB1 and is necessary for DDB1-mediated CUL4A binding to DDB2 [34].

Naturally occurring mutations have been identified in DDB2 that concentrate in the WD40 β-propeller region (Table 2; Figs. 1 and 2B). Interestingly, the missense mutations Arg273His, Asp307Tyr and Leu350Pro are responsible for disruption of the DDB1-DDB2 interaction, demonstrating the importance of proper PPIs in XP.

Following recognition of the DNA lesion and recruitment of XPC, DDB2 is degraded by the proteasome *via* protein auto-ubiquitination, allowing the preincision complex to be assembled in the XPC region [30,35]. Deletion of the DDB2 N-terminus (residues 1–40) abolishes DDB2 degradation. Mapping of the residues essential for DDB2 autoubiquitination and degradation indicated that seven Lys residues located at the beginning of the N-terminus (Lys4, 5, 11, 22, 35, 36 and 40) are fundamental for this process. Mutation of these seven Lys residues to Arg (Table 2; Fig. 1) prolonged DDB2 stabilization compared with WT DDB2 upon cycloheximide treatment in normal human fibroblast cells, although DDB2 retained its association with DDB1-CUL4A [30]. Lysine acetylation has also been described for DDB2 by means of global analysis of the acetylome; however, no detailed information regarding its impact on DDB2 function or participation in DNA repair has been explored [36].

Although the DDB2–DDB1 structure and its protein partners within the DDB-complex have been extensively described, studies that provide a deeper understanding of PPIs of DDB1–DDB2 proteins are essential to expand our knowledge of how this complex can modulate genomic stability, particularly in the development of cancer.

3. The unwinding companions: XPB and XPD structures and post-translational modifications

3.1. XPB structural information and mutations

XPB is a 782-residue protein with an ATP-dependent 3′ → 5′ helicase domain that participates in the TFIIH complex. Different from XPC and XPE, mutations in XPB can not only occasioned XP, but also CS, where affected individuals normally present both diseases, but can developed only XP [9]. XPB patients show skin cancers, severe sunburns and can also present neurological abnormalities [9].

The structure of *Archaeoglobus fulgidus* XPB has been determined by Fan et al. [37]. *A. fulgidus* XPB (referred to by the authors as Af-XPB) is composed of a damage recognition domain (DRD), two RecA-like helicase domains (HD1 and HD2), a flexible thumb motif (ThM) and a RED motif (Arg210–Glu211–Asp212, which corresponds to Arg472–Glu473–Asp474 in human XPB) [37–39], (Figs. 2C and 3). The DRD uses three Arg residues to bind DNA. In

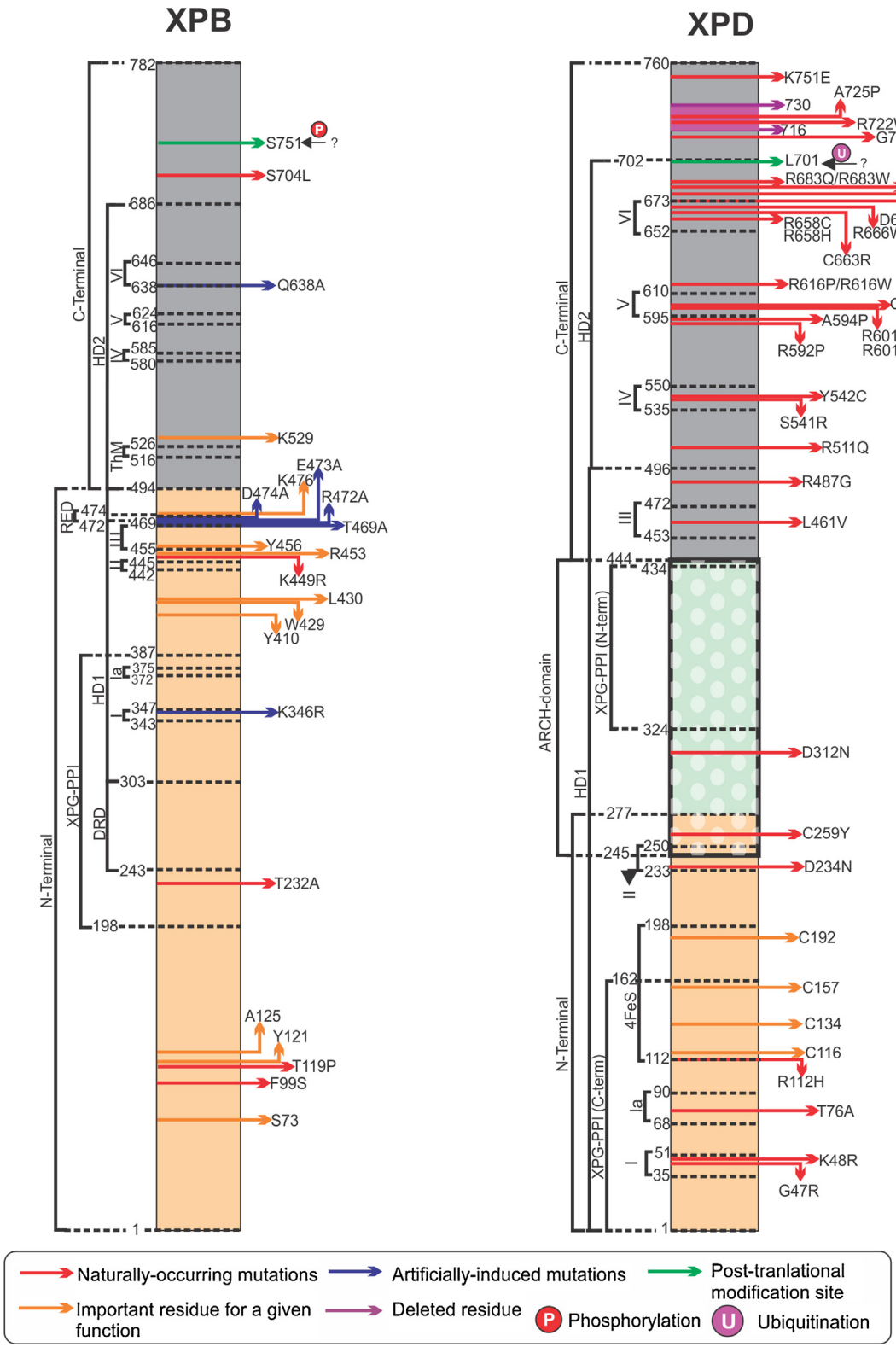


Fig. 3. Schematic model showing regions necessary for protein interaction, important amino acids residues, post-translational sites, known naturally occurring and artificially-induced mutations, and protein domains of XPB (A) and XPD (B). It is necessary to observe that "important amino acid residues" are all those needed for the maintenance of XP proteins-associated function and structure. Substitution of those amino acids by natural-occurring mutations or by artificial mutagenesis protocols is indicated in the figure. Deleted residues indicate specific regions in the protein that were excluded for functional studies. For information about the role of each residue or region, please address Tables 1-3 and the main text of the manuscript.

human XPB, these residues are not located in the DRD but rather at the beginning of the N-terminus (Table 1; Fig. 3). On the other side of the groove, seven residues in HD1 are responsible for protein-DNA interactions (Table 1; Fig. 3) [37]. In human XPB, these residues are also present in HD1. The ThM is also required for DNA binding and possesses seven conserved residues involved in this binding (Table 1; Figs. 2C and 3) [37]. Lys529 is the only residue in the ThM with a corresponding residue in human XPB that is located

in HD2. The RED motif is small region composed of three residues that are conserved in all XPB homologs from archaea to humans and are crucial for XPB helicase activity [37].

Studies have shown that a Glu473Ala mutation in the human RED motif and a deletion of residues 516–526 in the ThM are responsible for numerous impairments in XPB activity, such as a reduced cell survival rate following UV exposure and reduced DNA repair, helicase activity and ATPase activity [39]. An induced XPB mutant (Lys346Arg) exhibited similar behavior and negatively affected the recruitment of TFIIH to the CPD sites [39]. Different polymorphisms in the *XPB* gene have been found in lung cancer patients; however, their impact on XPB structure and function are unknown (Table 2; Fig. 3) [20].

Another noteworthy study assessed several mutations in individuals with XP and CS that result in impaired XPB function using genetic analyses [40]. Of these mutations, two missense mutations in *XPB* were already observed to reduce NER activity by affecting XPB protein interaction with the TFIIH complex [40,41] (Table 3). These results are also supported by another study that showed that a Phe99Ser mutation impaired the interaction between XPB and the p52 subunit of the TFIIH complex through the disruption of XPB ATPase activity [42]. These authors demonstrated that residues 305–358 in XPB interact with p52 [42]. Furthermore, expression of XPB containing two mutations (Lys359Glu and Ser382Pro) in Chinese hamster ovarian (CHO) cells impairs protein interactions between XPG and the p62 and p44 subunits of the TFIIH complex [43]. A Lys359Glu mutation affects single-stranded DNA (ssDNA) binding, whereas a Ser382Pro mutation disrupts 5′ incision in NER [43]. Additional regions of XPB have also been described that interact with XPG (Table 2; Fig. 3).

Structural studies on XPB PPIs are scarce. Although the structure of XPB has been determined and several mutations are known, little is known regarding the role of XPB in disorders such as XP and CS and how XPB interacts with other subunits in the TFIIH complex or other XP proteins.

To date, only one report describes post-translational phosphorylation of Ser751 in XPB [44]. This modification prevents 5′ incision by XPF-ERCC1, whereas dephosphorylation of Ser751 stimulates 5′ incision [44]. Although this post-translational modification has been shown to impair NER, it does not alter the opening of the DNA molecule around the damaged site or the transcriptional activity of the TFIIH complex [44].

3.2. XPD structure and mutations

XPD is an ATP-dependent $5' \rightarrow 3'$ helicase and consists of 760 residues. XPD individuals also show skin cancers and serious sunburns. Neurological abnormalities are rare, but there are studies that report cases with delayed neural development and decreased intelligence [45,46].

XPD contains two RecA-like helicase domains (HD1 and HD2), a 4-FeS cluster coordinated by four cysteines residues near the N-terminus of the protein and an Arch domain that is involved in protein interactions, most likely with different subunits of the TFIIH complex (Table 1; Fig. 3) [47–50].

The first study addressing the structure of human XPD used the UvrB helicase from *Escherichia coli* as a template (62% similarity with XPD) and threading algorithms to generate a molecular model of XPD [49]. Residues on UvrB were mutated to correspond to mutations in human XPD that are associated with XP, CS and TTD to examine the effects of these mutations on the structure and function of human XPD [49].

The functional analysis of the UvrB Glu110Ala/Arg mutant indicated diminished ATPase activity, despite the low sequence conservation between the UvrB β-hairpin region and XPD [49]. However, despite the high sequence similarity, the identity between UvrB and XPD is low (15% identity), which could explain possible experimental differences between the two proteins.

The 4-FeS cluster is a major structural characteristic that is crucial for XPD function [51,52]. The FeS cluster of XPD contains four cysteines residues that coordinate four iron ions [4Fe-4S] controlled by ATP binding and hydrolysis [51–53]. This cluster is sensitive to oxidation and may be responsible for XPD instability [51]. Arg112His is a common mutation that affects the 4-FeS cluster in the β-hairpin region and induces severe UV sensitivity; Arg112 is responsible for DNA binding [49,51,54]. Moreover, different regions of XPD have been described to interact with XPG (Table 1; Fig. 3).

Other studies have also assessed the impact of several mutations in XPD [51]. The effects of known human XPD mutations were examined using the determined structure of *Sulfolobus acidocaldarius* XPD (SaXPD) as a model (summarized in Table 2 and organized according to the effect and the location of the mutation). In summary, this study showed that mutations in XPD for: (i) XP patients primarily occur in residues that lie between HD2 and the other domains, including HD1, and impair ATP hydrolysis and ssDNA binding; (ii) in CS patients, the mutations occur in residues that lie within HD1 and HD2; and (iii) in TTD patients, naturally occurring XPD mutations occur in the four domains, primarily in the C-terminal half of the protein [51]. Interestingly, the mutation Asp681Asn, which has been proposed to impair SaXPD ATP hydrolysis and DNA binding, was found in a patient with cerebro-oculo-facio-skeletal syndrome (COFSS), which is a disorder that results in neurological and visual impairment [55]. The majority of the mutations found in XPD are located in HD2 (Table 2; Fig. 3).

Studies on post-translational modifications of XPD are lacking. A study that proposed 281 ubiquitination sites in 252 different proteins that are specific for the human lineage suggested that XPD featured an ubiquitination site at Lys701, which is located in the highly conserved C-terminus (Table 3; Fig. 3) [56]. However, *in vitro* or *in vivo* ubiquitination of XPD has not been reported.

Studies on *S. acidocaldarius* and *Thermoplasma acidophilum* XPD proteins considerably contribute to the understanding of the structure and function of human XPD [51,57]. Nevertheless, details of XPD interactions with other proteins (*e.g.*, XPG) remain to be elucidated. XPG has been suggested to participate in the TFIIH complex [8], and studies have indicated that XPG forms a stable complex with TFIIH by functioning in concert with XPD [58]. However, this interaction has not been structurally characterized.

4. Preparing the site: XPA structure and post-translational modifications

Following DNA-damage recognition and opening of the dsDNA, XPA is recruited to the damaged site to serve as a platform for the endonucleases XPF and XPG. XPA individuals equally develop skin cancers as do other XP patients. However, these individuals can also show serious sunburns and a syndrome named DeSanctis-Cacchione Syndrome (DSC), described for the first time in 1932 [59]. DSC individuals show neurological degeneration with age [9].

XPA is a 273-residue DNA-binding protein that forms a complex with the replication protein A1 (RPA1) to exert its function in NER (Fig. 4). The entire tertiary structure of XPA has not been determined, and only the central domain of the protein has been defined [60]. Four XPA Cys residues coordinate a Zn^{2+} cation (Table 1; Fig. 2A and Fig. 4), present in the fragment related to DNA-binding [61]. However, the zinc-finger motif appears necessary for proper XPA binding to RPA, but, in contrast, has no role in DNA-binding [62]. Moreover, the XPA N-terminus (residues 1–97) and central region (residues 98–225) are fundamental for RPA binding

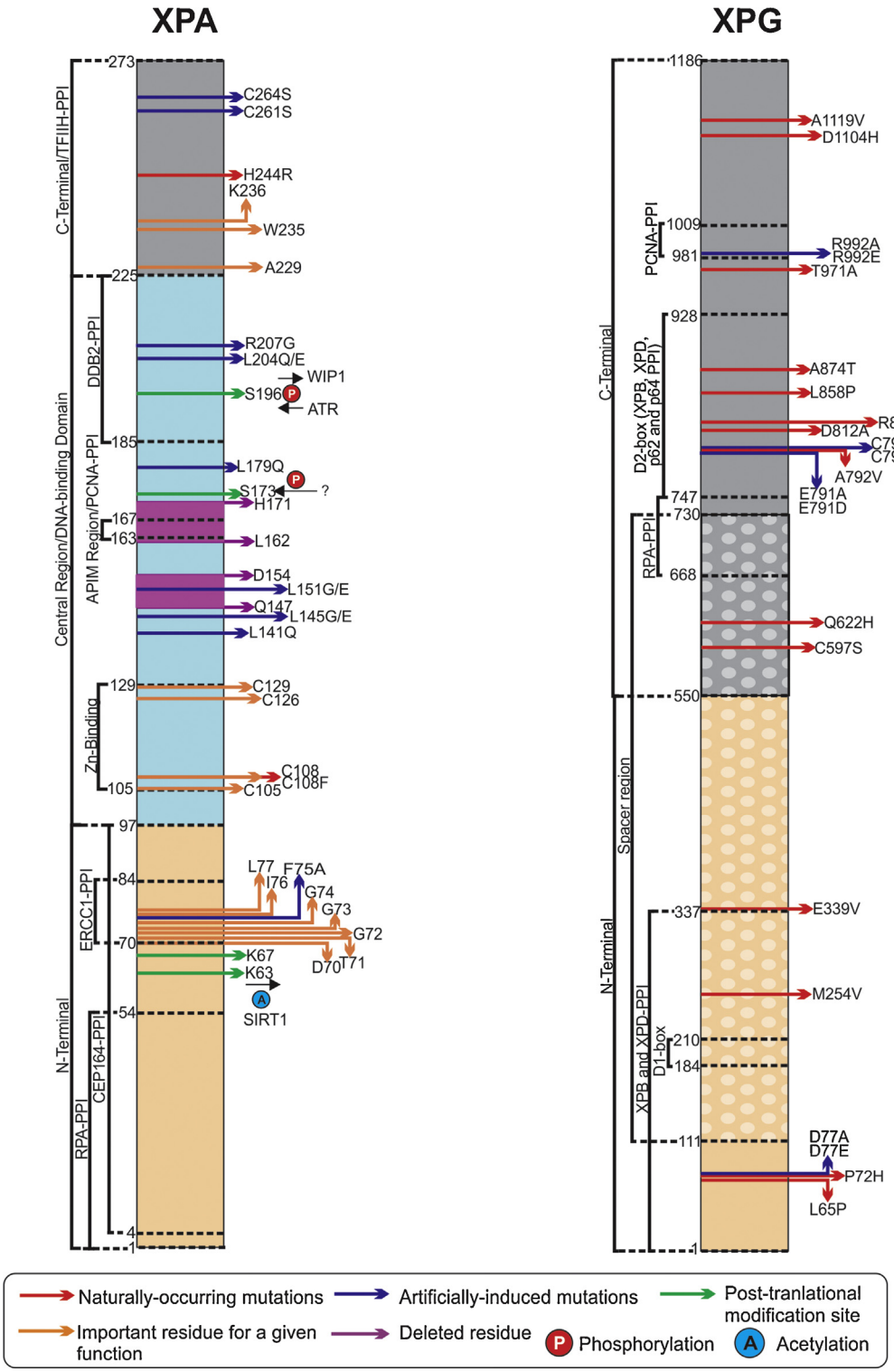


Fig. 4. Schematic model showing regions necessary for protein interaction, important amino acids residues, post-translational sites, known naturally occurring and artificially-induced mutations, and protein domains of XPA (A) and XPG (B). It is necessary to observe that "important amino acid residues" are all those needed for the maintenance of XP proteins-associated function and structure. Substitution of those amino acids by natural-occurring mutations or by artificial mutagenesis protocols is indicated in the figure. For information about the role of each residue or region, please address Tables 1-3 and the main text of the manuscript.

and [63]. Additionally, Cys261 in the C-terminal region (residues 226–273) is necessary for XPA binding to the TFIIH complex [64].

Of the several induced mutations in XPA, a double mutant Lys141Glu/Lys179Glu exhibited several defects in DNA binding) [65]. These authors also propose that these two residues are crucial for recognition of DNA-damaged regions. In addition, a most recent study identified three aa residues (Ala229, Trp235 and Lys236) that appear to be important for XPA DNA-binding capacity, in the C-terminal of XPA protein [66].

Site-directed mutagenesis studies on XPA have also demonstrated that removal of the conserved regions disrupts binding to the 70-kDa subunit, but not the 34-kDa subunit, of RPA [67]. Thus,

the N-terminal and central regions of XPA are necessary for interaction with RPA, whereas XPA DNA binding is regulated by residues found throughout the protein. In contrast, a study that used a small XPA peptide (XPA_{67-80}) demonstrated that three glycine residues are necessary for XPA binding to ERCC1 and XPF [67,68]. These results are consistent with a previous study that examined a recombinant XPA lacking these glycine residues and Phe75, which was unable to bind ERCC1 [69]. Other residues appear to be responsible for stabilization and strengthening of the XPA–ERCC1 interaction [5,68,69] (Table 1; Fig. 4).

XPA has also been shown to interact with DDB2 [70] (Figs. 2A and 4). The XPA Arg207Gly mutation has been shown to impair the formation of the XPA–DDB2 heterodimer and results in a failure to promote CPD excision [70]. Another study has demonstrated that XPA can interact with PCNA, which is a protein responsible for genome maintenance [71]. PCNA binds to the so-called APIM (XPA residues 163–167), which is an evolutionarily conserved motif found in proteins involved in DNA integrity and maintenance [71,72]. A Phe164Ala mutation in XPA impaired the XPA-PCNA interaction [71]. Moreover, XPA interacts with CEP164, which is a centrosomal protein that is involved in the response to DNA damage [73]. Upon UV damage, CEP164 is phosphorylated by ATR and leads to CHK1 accumulation [73].

XPA is also target of post-translational modifications, specifically deacetylation by the NAD-dependent acetyl-transferase sirtuin 1 (SIRT1) [74]. *In vitro* studies have shown that XPA can be acetylated at Lys63 and Lys67 (Table 3; Fig. 4) and that SIRT1 was able to deacetylate these residues. This finding was confirmed using a SIRT1 mutant (His393Tyr), which lacks deacetylase activity [74]. This study also indicated that XPA is acetylated in the early stages of UV-induced DNA damage, a process that decreases with time and is consistent with an increase in SIRT1 expression. Additionally, acetylated XPA can impair the interaction between XPA and RPA, thus negatively regulating NER [74]. Overall, deacetylation of XPA is necessary to prime DNA repair.

The ataxia telangiectasia-related and Rad3-related (ATR) protein has also been shown to phosphorylate XPA [75] (Table 3; Figs. 2A and 4). XPA colocalizes and physically interacts with ATR in the nucleus following UV irradiation, promoting resistance to UV-induced DNA damage and NER activity [75]. The XPA (Ser196Ala) mutation, which impairs phosphorylation, increased UV sensitivity; however, how XPA phosphorylation affects NER remains unclear. Ser196 is located in the XPA central region, which contains residues that interact with RPA and DDB2 (Table 3; Figs. 2A and 4). It is possible that XPA phosphorylation alters its PPIs or, most likely, its activity. This speculation is reasonable because XPA has been reported to be dephosphorylated by WIP1 (Table 3) [25]. WIP1-deficient cells exhibited increased CPD repair, whereas WIP1 overexpression inhibited NER [25].

Finally, XPA has also been reported to be ubiquitinated by HERC2. Downregulation of HERC2 by siRNA completely abolished this XPA ubiquitination [76]. Aside from its proteolytic degradation [76], no other function was assigned for the ubiquitination of XPA.

To date, only a small portion of XPA has been modeled. Due to its importance in NER, it is surprising that a complete three-dimensional structure of XPA remains unavailable.

5. Hack-and-slash: XPF and XPG structures

5.1. XPG structural information and post-translational modifications

XPG is the first NER-associated 3′-endonuclease recruited to the lesion site and consists of 1186 residues. XPG individuals show, in addition to severe sunburns and skin cancers, developmental abnormalities like microcephaly and bilateral pes cavus [77,78].

The first XPG structure was modeled and studied using structurally similar endonucleases, including the bacteriophage proteins T4 RNase H and T5 D15, FEN1 (DNAse IV) and Taq DNA polymerase, all of which possess an evolutionarily conserved N-terminal region and an internal region called the "I region" [79]. However, XPG contains a 600-residue spacer region separating the N-terminal and the I region [80] and [81], and this feature differs from the other related proteins. Dunand-Sauthier et al. [81] have assessed the effect of different spacer region lengths ($XPG_{111-550}$, $XPG_{184-210}$ and $XPG_{554-730}$) on human XPG activity and function (Table 1; Fig. 4).

The XPG D1 box (residues 184–210) is evolutionarily conserved that together with the D2 box (residues 890–984) were first studied in *Drosophila melanogaster* [81] and [82]. The spacer region is generally not involved in catalysis but is required for proper XPG function in DNA cleavage, recruitment of XPG to UV-damaged sites, PPIs and DNA excision [80,81,83,84], most likely due to its ability to properly position XPG at the damaged site [81]. Other residues in the spacer region are required for XPG PPIs with different subunits of the TFIIH complex (Table 1; Fig. 4). Biochemical studies have also indicated a role for Mg^{2+} or Mn^{2+} in XPG endonuclease activity [80,85]; although XPG exhibits a higher affinity for Mn^{2+}, it will use whichever cation is more abundant [79]. A naturally occurring mutation derived from an XP patient that demonstrated reduced endonuclease activity exhibited a greater ability to cleave DNA in the presence of Mn^{2+} than in the presence of Mg^{2+} [79]. Glu791 and Asp812 are predicted to coordinate these divalent metal ions [79]. XPG can also interact with PCNA *via* a conserved motif (residues 981–1009) [86,87].

XPG, in addition to RPA, may be responsible for coordinating and signaling PCNA to the damaged site [88]. The importance of PCNA binding to XPG has been examined in a recent work that used a mutant Rad2p, which is the yeast homolog of human XPG, to demonstrate the relationship of XPG-PCNA during the response to DNA damage [89]. Induced mutations in the Rad2p binding site of PCNA [Rad2 (Phe1001Ala) and Rad2 (Phe1002Ala)] were found to impair cell cycle regulation but were not necessary for NER activity [87,89].

As shown in Table 2 and Fig. 4, several naturally occurring mutations have been observed in lung cancer and XP patients that affect XPG protein interactions. Because XPG is also necessary for proper recruitment of XPF-ERCC1, understanding how different mutations affect the XPG–XPF interaction is of considerable interest.

XPG has been shown to be post-translationally modified by acetylation [73]. A study has demonstrated that XPG association with p300 and CBP increases following UV radiation [90]. XPG has been shown to be acetylated by p300/CBP and that this acetylation can be stimulated by p21 [90]. However, knockdown of p300/CBP resulted in an accumulation of XPG at CPD-damaged sites [90]. Acetylation is apparently not required for XPG recruitment to the damaged sites but is required for XPG release from the lesions.

5.2. XPF structural information and post-translational modifications

XPF is a 5′ endonuclease that consists of 916 residues. XPF forms a heterodimer with ERCC1, and this complex is crucial for proper XPF binding to XPA (Fig. 5). XPF patients do not diverge from the symptoms seen in XPG affected individuals, although XPF patients can develop CS, COFS and XPF-ERCC1 progeroid (XFE) and neurological abnormalities [91].

XPF is a member of the evolutionarily conserved and structurally related endonuclease family, which also includes the yeast homolog Rad1p and the nuclease Mus81p [92–94]. XPF/Rad1/Mus81 contain a conserved signature sequence often

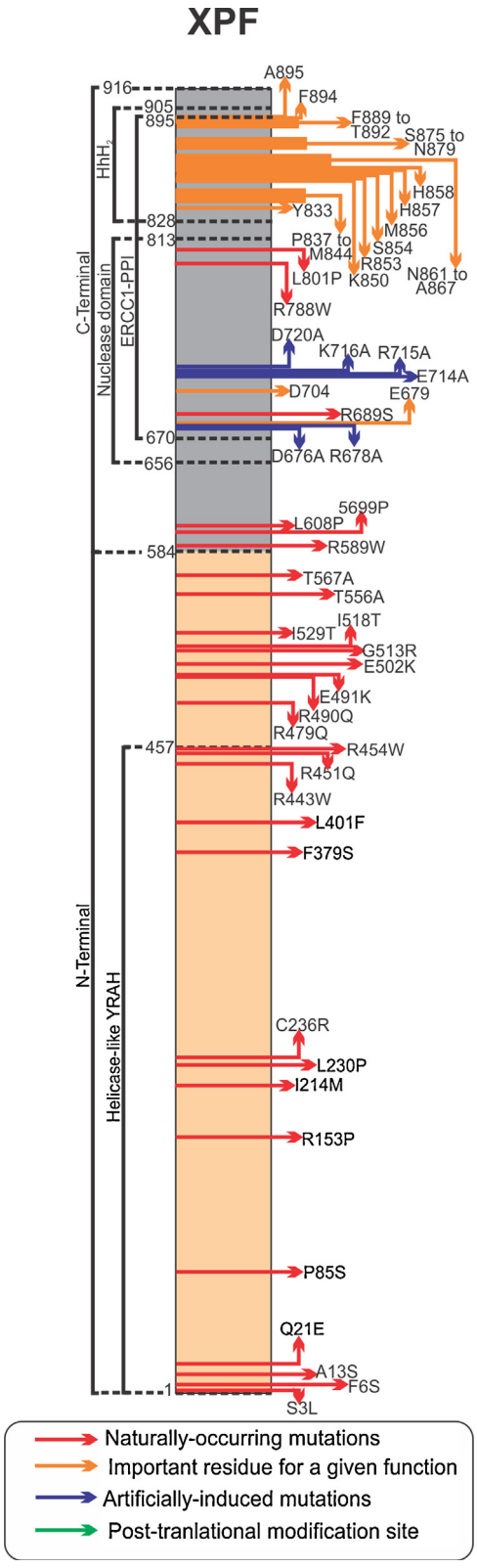


Fig. 5. Schematic model showing regions necessary for protein interaction, important amino acids residues, post-translational sites, known naturally occurring and artificially-induced mutations, and protein domains of XPF. It is necessary to observe that "important amino acid residues" are all those needed for the maintenance of XP proteins-associated function and structure. Substitution of those amino acids by natural-occurring mutations or by artificial mutagenesis protocols is indicated in the figure. For information about the role of each residue or region, please address Tables 1-3 and the main text of the manuscript.

represented as GDX_nERKX_3D, which is essential for catalytic activity [92,95,96] (for a comprehensive review of the evolutionarily conserved XPF/Rad1/Mus81 signature sequence, see [95]). As observed for XPG, XPF also contains a metal binding site for Mg^{2+} and Mn^{2+} ions, with a preference for Mn^{2+}, in which four residues coordinate metal ion binding [97] (Table 1; Fig. 5). Specific residues are also involved in the nuclease activity of XPF, and mutations of these residues impaired XPF binding of both dsDNA and ssDNA (Table 2; Fig. 5) [97].

A detailed description of the human XPF HhH motif in the C-terminal domain (residues 863–905) has indicated that several XPF residues are required for interaction with ERCC1 (Table 1; Fig. 5) [98,99], which is important for heterodimer formation.

ERCC1 has been shown to be able to correctly fold only when associated with XPF [100,101]. In contrast, the homologous XPF protein from the crenarchaeon *Aeropyrum pernix* is able to bind both ssDNA and dsDNA in the absence of additional proteins [96]. This study also developed a model from which the HhH motif of XPF appears to exhibit a preference for ssDNA; human XPF can also bind ssDNA [101].

Biophysical and biochemical assays have revealed that XPF primarily interacts with the phosphate backbone of DNA *via* residues Lys850, Arg853, Ser854, His857 and His858, and with the nucleotides *via* residues Ser875, Ile876, Leu877, Gly878 and Asn879 [101]. Moreover, RPA is necessary for the recruitment of XPF to UV-damaged sites, which is mediated by Pro85 in the N-terminal region of XPF [102] (Table 1; Fig. 5). This study demonstrated that CHO cells expressing XPF harboring a Pro85Ser mutation resulted in XPF localization in the cytoplasm instead of the nucleus, preventing resistance of cells to UV radiation [102].

Numerous XPF mutations have been described in XP and CS patients, with the majority of the mutations located in the helicase-like domain and in the surrounding regions, particularly the N-terminus (Table 2; Fig. 5). Other mutations have also been found in patients with Fanconi anemia (Table 2).

The role of post-translational modifications of XPF has not yet been demonstrated. One acetylation site has been predicted in XPF [36]; however, a function for this modification has not been described.

6. Final remarks

In this review, we have discussed the structure and known post-translational modifications of each XP protein. Overall, the structures of the human XP proteins have not been fully characterized. For example, an *in silico* survey of three-dimensional XP proteins using human sequences as input parameters indicated that complete protein crystal data are available for only DDB1 and DDB2. Some XP proteins have been modeled *in silico* or peptide fragments of these proteins have been mapped onto determined crystal structures of XP protein homologs. Although homologous XP proteins from unicellular organisms can be used as a model to understand the structure and function of the human XP proteins, it is imperative to obtain the complete structures of all HsaXP proteins to clarify their role in NER and understand how mutations in these proteins lead to disease. Computational analysis may be beneficial in this regard, with the availability of numerous software dedicated to the prediction of protein dynamics, secondary structure and protein folding. The information provided in Tables 1–3 and Figs. 1–5 may be beneficial for these predictions and may serve as a foundation for structural modeling.

Funding

This work was supported by research grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq;

grant no. 301149/2012-7), the Programa Institutos Nacionais de Ciência e Tecnologia (INCT de Processos Redox em Biomedicina-REDOXOMA; grant no. 573530/2008-4), Fundação de Amparo a Pesquisa do Rio Grande do Sul FAPERGS (PRONEM grant no. 11/2072-2), and the Programa Binacional de Terapia Celular – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PROBITEC-CAPES; grant no. 004/12).

Conflict of interest statement

The authors declare that there are no conflicts of interest.

References

[1] O.D. Schärer, Nucleotide excision repair in eukaryotes, Cold Spring Harb. Perspect. Biol. 5 (2013), a012609.
[2] H. Naegeli, K. Sugasawa, The xeroderma pigmentosum pathway: decision tree analysis of DNA quality, DNA Repair (Amst) 10 (2011) 673–683.
[3] J.A. Marteijn, H. Lans, W. Vermeulen, J.H.J. Hoeijmakers, Understanding nucleotide excision repair and its roles in cancer and ageing, Nat. Rev. Mol. Cell Biol. 15 (2014) 465–481.
[4] S.M. Shell, Y. Zou, Other proteins interacting with XP proteins, Adv. Exp. Med. Biol. 637 (2008) 103–112.
[5] D.L. Croteau, Y. Peng, B. Van Houten, DNA repair gets physical: mapping a XPA binding site on ERCC1, DNA Repair 7 (2008) 819–826.
[6] L.C.J. Gillet, O.D. Schärer, Molecular mechanisms of mammalian global genome nucleotide excision repair, Chem. Rev. 106 (2006) 253–276.
[7] A.R. Lehmann, D. McGibbon, M. Stefanini, Xeroderma pigmentosum, Orphanet J. Rare Dis. 6 (2011) 70.
[8] J.-M. Egly, F. Coin, A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor, DNA Repair (Amst) 10 (2011) 714–721.
[9] C.F. Menck, V. Munford, DNA repair diseases: what do they tell us about cancer and aging? Genet. Mol. Biol. 37 (2014) 220–233.
[10] S.G. Khan, A. Metin, E. Gozukara, H. Inui, T. Shahlavi, V. Muniz-Medina, et al., Two essential splice lariat branchpoint sequences in one intron in a xeroderma pigmentosum DNA repair gene: mutations result in reduced XPC mRNA levels that correlate with cancer risk, Hum. Mol. Genet. 13 (2004) 343–352.
[11] K. Sugasawa, XPC: its product and biological roles, in: Mol. Mech. Xeroderma Pigment, 2008, 47–56.
[12] M. Araki, C. Masutani, M. Takemura, A. Uchida, K. Sugasawa, J. Kondoh, et al., Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair, J. Biol. Chem. 276 (2001) 18665–18672.
[13] J.-B. Charbonnier, E. Renaud, S. Miron, M.H. Le Du, Y. Blouquit, P. Duchambon, et al., Structural, thermodynamic, and cellular characterization of human centrin 2 interaction with xeroderma pigmentosum group C protein, J. Mol. Biol. 373 (2007) 1032–1046.
[14] J.R. Thompson, Z.C. Ryan, J.L. Salisbury, R. Kumar, The structure of the human centrin 2-xeroderma pigmentosum group C protein complex, J. Biol. Chem. 281 (2006) 18746–18752.
[15] C. Bunick, M. Miller, B. Fuller, Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein, Biochemistry 45 (2006) 14965–14979.
[16] J.-H. Min, N.P. Pavletich, Recognition of DNA damage by the Rad4 nucleotide excision repair protein, Nature 449 (2007) 570–575.
[17] F.C. Clement, N. Kaczmarek, N. Mathieu, M. Tomas, A. Leitenstorfer, E. Ferrando-May, et al., Dissection of the xeroderma pigmentosum group C protein function by site-directed mutagenesis, Antioxid. Redox Signal. 14 (2011) 2479–2490.
[18] U. Camenisch, D. Träutlein, F.C. Clement, J. Fei, A. Leitenstorfer, E. Ferrando-May, et al., Two-stage dynamic DNA quality check by xeroderma pigmentosum group C protein, EMBO J. 28 (2009) 2387–2399.
[19] J. Slyskova, A. Naccarati, B. Pardini, V. Polakova, L. Vodickova, Z. Smerhovsky, et al., Differences in nucleotide excision repair capacity between newly diagnosed colorectal cancer patients and healthy controls, Mutagenesis 27 (2012) 225–232.
[20] A. Matakidou, T. Eisen, C. Fleischmann, H. Bridle, R.S. Houlston, Evaluation of xeroderma pigmentosum XPA, XPC, XPD, XPF, XPB, XPG and DDB2 genes in familial early-onset lung cancer predisposition, Int. J. Cancer 119 (2006) 964–967.
[21] L. Li, E. Bales, C. Peterson, R. Legerski, Characterization of molecular defects in xeroderma pigmentosum group C, Nat. Genet. 5 (1993) 413–417.
[22] K. Sugasawa, Y. Okuda, M. Saijo, R. Nishi, N. Matsuda, G. Chu, et al., UV-induced ubiquitylation of XPC protein mediated by UV–DDB–ubiquitin ligase complex, Cell 121 (2005) 387–400.
[23] Q.-E. Wang, Q. Zhu, G. Wani, M.A. El-Mahdy, J. Li, A.A. Wani, DNA repair factor XPC is modified by SUMO-1 and ubiquitin following UV irradiation, Nucleic Acids Res. 33 (2005) 4023–4034.
[24] Q.-E. Wang, M. Praetorius-Ibba, Q. Zhu, M.A. El-Mahdy, G. Wani, Q. Zhao, et al., Ubiquitylation-independent degradation of xeroderma pigmentosum group C protein is required for efficient nucleotide excision repair, Nucleic Acids Res. 35 (2007) 5338–5350.
[25] T.A. Nguyen, S.D. Slattery, S.H. Moon, Y.F. Darlington, X. Lu, L.A. Donehower, The oncogenic phosphatase WIP1 negatively regulates nucleotide excision repair, DNA Repair 9 (2011) 813–823.
[26] A. Scrima, E.S. Fischer, G.M. Lingaraju, K. Böhm, S. Cavadini, N.H. Thomä, Detecting UV-lesions in the genome: the modular CRL4 ubiquitin ligase does it best! FEBS Lett. 585 (2011) 2818–2825.
[27] G. Chu, W. Yang, Here comes the sun: recognition of UV-damaged DNA, Cell 135 (2008) 1172–1174.
[28] A. Scrima, R. Konícková, B.K. Czyzewski, Y. Kawasaki, P.D. Jeffrey, R. Groisman, et al., Structural basis of UV DNA-damage recognition by the DDB1–DDB2 complex, Cell 135 (2008) 1213–1223.
[29] O.D. Schärer, A.J. Campbell, Wedging out DNA damage, Nat. Struct. Mol. Biol. 16 (2009) 102–104.
[30] E.S. Fischer, A. Scrima, K. Böhm, S. Matsumoto, G.M. Lingaraju, M. Faty, et al., The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation, Cell 147 (2011) 1024–1039.
[31] J.I. Yeh, A.S. Levine, S. Du, U. Chinte, H. Ghodke, H. Wang, et al., Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) E2737–E2746.
[32] C.K.-M. Chen, N.-L. Chan, A.H.-J. Wang, The many blades of the β-propeller proteins: conserved but versatile, Trends Biochem. Sci. 36 (2011) 553–561.
[33] C. Xu, J. Min, Structure and function of WD40 domain proteins, Protein Cell 2 (2011) 202–214.
[34] Y.J. He, C.M. McCall, J. Hu, Y. Zeng, Y. Xiong, DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases, Genes Dev. 20 (2006) 2949–2954.
[35] A. Nag, T. Bondar, S. Shiv, The xeroderma pigmentosum group E gene product DDB2 is a specific target of cullin 4A in mammalian cells, Mol. Cell. Biol. 21 (2001) 6738–6747.
[36] C. Choudhary, C. Kumar, F. Gnad, M.L. Nielsen, M. Rehman, T.C. Walther, et al., Lysine acetylation targets protein complexes and co-regulates major cellular functions, Science 325 (2009) 834–840.
[37] L. Fan, A.S. Arvai, P.K. Cooper, S. Iwai, F. Hanaoka, J.A. Tainer, Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair, Mol. Cell 22 (2006) 27–37.
[38] E. Hilario, Y. Li, Y. Nobumori, X. Liu, L. Fan, Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE, Acta Crystallogr. D: Biol. Crystallogr. 69 (2013) 237–246.
[39] V. Oksenych, B. Bernardes de Jesus, A. Zhovmer, J.-M. Egly, F. Coin, Molecular insights into the recruitment of TFIIH to sites of DNA damage, EMBO J. 28 (2009) 2971–2980.
[40] K. Oh, S.G. Khan, N.G.J. Jaspers, A. Raams, T. Ueda, A. Lehmann, et al., Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with cockayne syndrome, Hum. Mutat. 27 (2006) 1092–1103.
[41] G. Weeda, E. Eveno, I. Donker, W. Vermeulen, O. Chevallier-Lagente, A. Taïeb, et al., A mutation in the XPB/ERCC3 DNA repair transcription gene, associated with trichothiodystrophy, Am. J. Hum. Genet. 60 (1997) 320–329.
[42] F. Coin, V. Oksenych, J.-M. Egly, Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair, Mol. Cell 26 (2007) 245–256.
[43] H. Hall, J. Gurský, A. Nicodemou, I. Rybanská, E. Kimlícková, M. Pirsel, Characterization of ERCC3 mutations in the Chinese hamster ovary 27-1, UV24 and MMC-2 cell lines, Mutat. Res. 593 (2006) 177–186.
[44] F. Coin, J. Auriol, A. Tapias, P. Clivio, W. Vermeulen, J.-M. Egly, Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity, EMBO J. 23 (2004) 4835–4846.
[45] B.C. Broughton, M. Berneburg, H. Fawcett, E.M. Taylor, C.F. Arlett, T. Nardo, et al., Two individuals with features of both xeroderma pigmentosum and trichothiodystrophy highlight the complexity of the clinical outcomes of mutations in the XPD gene, Hum. Mol. Genet. 10 (2001) 2539–2547.
[46] W.J. Kleijer, F.A. Beemer, B.W. Boom, Intermittent hair loss in a child with PIBI(D)S syndrome and trichothiodystrophy with defective DNA repair-xeroderma pigmentosum group D, Am. J. Med. Genet. 52 (1994) 227–230.
[47] A.R. Lehmann, XPD structure reveals its secrets, DNA Repair (Amst) 7 (2008) 1912–1915.
[48] H. Liu, J. Rudolf, K.A. Johnson, S.A. Mcmahon, M. Oke, A. Mcrobbie, et al., Structure of the DNA repair helicase XPD, Cell 133 (2012) 801–812.
[49] R.J. Bienstock, M. Skorvaga, B.S. Mandavilli, B. Van Houten, Structural and functional characterization of the human DNA repair helicase XPD by comparative molecular modeling and site-directed mutagenesis of the bacterial repair protein UvrB, J. Biol. Chem. 278 (2003) 5309–5316.
[50] J.O. Fuss1, J.A. Tainer, XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase, DNA Repair 10 (2012) 697–713.
[51] L. Fan, J.O. Fuss, Q.J. Cheng, A.S. Arvai, M. Hammel, V.A. Roberts, et al., XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations, Cell 133 (2008) 789–800.
[52] S.C. Wolski, J. Kuper, P. Hänzelmann, J.J. Truglio, D.L. Croteau, B. Van Houten, et al., Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD, PLoS Biol. 6 (2008) e149.
[53] J. Rudolf, V. Makrantoni, W.J. Ingledew, M.J.R. Stark, M.F. White, The DNA repair helicases XPD and FancJ have essential iron–sulfur domains, Mol. Cell 23 (2006) 801–808.
[54] E. Bergmann, J.M. Egly, Trichothiodystrophy, a transcription syndrome, Trends Genet. 17 (2001) 279–286.

[55] J.M. Graham, K. Anyane-Yeboa, A. Raams, E. Appeldoorn, W.J. Kleijer, V.H. Garritsen, et al., Cerebro-oculo-facio-skeletal syndrome with a nucleotide excision-repair defect and a mutated XPD gene, with prenatal diagnosis in a triplet pregnancy, Am. J. Hum. Genet. 69 (2001) 291–300.
[56] D.S. Kim, Y. Hahn, Gains of ubiquitylation sites in highly conserved proteins in the human lineage, BMC Bioinform. 13 (2012) 306.
[57] J. Kuper, S.C. Wolski, G. Michels, C. Kisker, Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation, EMBO J. 31 (2012) 494–502.
[58] S. Ito, I. Kuraoka, P. Chymkowitch, E. Compe, A. Takedachi, C. Ishigami, et al., XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients, Mol. Cell 26 (2007) 231–243.
[59] C. De Sanctis, A. Cacchione, L'idiozia Xerodermica, Riv. Sper. Freniat. 56 (1932) 269–292.
[60] G.W. Buchko, G.W. Daughdrill, R. de Lorimier, K. Rao B, N.G. Isern, J.M. Lingbeck, et al., Interactions of human nucleotide excision repair protein XPA with DNA and RPA70 Delta C327: chemical shift mapping and 15N NMR relaxation studies, Biochemistry 38 (1999) 15116–15128.
[61] C.L. Bartels, M.W. Lambert, Domains in the XPA protein important in its role as a processivity factor, Biochem. Biophys. Res. Commun. 356 (2007) 219–225.
[62] T. Ikegami, I. Kuraoka, M. Saijo, Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA, Nat. Struct. Mol. Biol. 5 (1998) 701–706.
[63] S. Ahmad, F. Hanaoka, Molecular Mechanisms of Xeroderma Pigmentosum, 1st ed., Landes Bioscience and Springer Science, Austin, TX, 2008.
[64] C.H. Park, D. Mu, J.T. Reardon, A. Sancar, The general transcription-repair factor TFIIH is recruited to the excision repair complex by the XPA protein independent of the TFIIE transcription factor, J. Biol. Chem. 270 (1995) 4896–4902.
[65] U. Camenisch, R. Dip, S.B. Schumacher, B. Schuler, H. Naegeli, Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair, Nat. Struct. Mol. Biol. 13 (2006) 278–284.
[66] N. Sugitani, S.M. Shell, S.E. Soss, W.J. Chazin, Redefining the DNA-binding domain of human XPA, J. Am. Chem. Soc. 136 (2014) 10830–10833.
[67] L. Li, X. Lu, C. Peterson, R. Legerski, An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair, Mol. Cell. Biol. 15 (1995) 5396–5402.
[68] O.V. Tsodikov, D. Ivanov, B. Orelli, L. Staresincic, I. Shoshani, R. Oberman, et al., Structural basis for the recruitment of ERCC1–XPF to nucleotide excision repair complexes by XPA, EMBO J. 26 (2007) 4768–4776.
[69] L. Li, C.A. Peterson, R.X. Lu, J. Legerski, Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair, Mol. Cell. Biol. 15 (1995) 1993–1998.
[70] M. Wakasugi, H. Kasashima, Y. Fukase, M. Imura, R. Imai, S. Yamada, et al., Physical and functional interaction between DDB and XPA in nucleotide excision repair, Nucleic Acids Res. 37 (2009) 516–525.
[71] K.M. Gilljam, R. Müller, N.B. Liabakk, M. Otterlei, Nucleotide excision repair is associated with the replisome and its efficiency depends on a direct interaction between XPA and PCNA, PLOS ONE 7 (2012) e49199.
[72] K.M. Gilljam, E. Feyzi, P.A. Aas, M.M.L. Sousa, R. Müller, C.B. Vågbø, et al., Identification of a novel, widespread, and functionally important PCNA-binding motif, J. Cell Biol. 186 (2009) 645–654.
[73] Y. Pan, E. Lee, UV-dependent interaction between Cep164 and XPA mediates localization of Cep164 at sites of DNA damage and UV sensitivity, Cell Cycle 8 (2009) 655–664.
[74] W. Fan, J. Luo, SIRT1 regulates UV-induced DNA repair through deacetylating XPA, Mol. Cell 39 (2010) 247–258.
[75] X. Wu, S.M. Shell, Z. Yang, Y. Zou, Phosphorylation of nucleotide excision repair factor xeroderma pigmentosum group A by ataxia telangiectasia mutated and Rad3-related-dependent checkpoint pathway promotes cell survival in response to UV irradiation, Cancer Res. 66 (2006) 2997–3005.
[76] T.-H. Kang, J.T. Reardon, A. Sancar, Regulation of nucleotide excision repair activity by transcriptional and post-transcriptional control of the XPA protein, Nucleic Acids Res. 39 (2011) 3176–3187.
[77] M.J. Cheesbrough, P.D.S. Kinmont, Xeroderma pigmentosum: a unique variant with neurologic involvement, Br. J. Dermatol. 99 (1978) 61.
[78] W. Keijzer, N.G. Jaspers, P.J. Abrahams, A.M. Taylor, C.F. Arlett, B. Zelle, et al., A seventh complementation group in excision-deficient xeroderma pigmentosum, Mutat. Res. 62 (1979) 183–190.
[79] A. Constantinou, Conserved residues of human XPG protein important for nuclease activity and function in nucleotide excision repair, J. Biol. Chem. 274 (1999) 5637–5648.
[80] M. Hohl, I. Dunand-Sauthier, L. Staresincic, P. Jaquier-Gubler, F. Thorel, M. Modesti, et al., Domain swapping between FEN-1 and XPG defines regions in XPG that mediate nucleotide excision repair activity and substrate specificity, Nucleic Acids Res. 35 (2007) 3053–3063.
[81] I. Dunand-Sauthier, M. Hohl, F. Thorel, P. Jaquier-Gubler, S.G. Clarkson, O.D. Schärer, The spacer region of XPG mediates recruitment to nucleotide excision repair complexes and determines substrate specificity, J. Biol. Chem. 280 (2005) 7030–7037.
[82] J.F. Houle, E.C. Friedberg, The *Drosophila* ortholog of the human XPG gene, Gene 234 (1999) 353–360.
[83] F. Thorel, Definition of a short region of XPG necessary for TFIIH interaction and stable recruitment to sites of UV damage, Mol. Cell. Biol. 24 (2004) 10670–10680.
[84] Z. He, L.A. Henricksen, M.S. Wold, C.J. Ingles, RPA involvement in the damage-recognition and incision steps of nucleotide excision repair, Nature 374 (1995) 566–569.
[85] N. Iyer, M.S. Reagan, K.J. Wu, B. Canagarajah, E.C. Friedberg, Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein, Biochemistry 35 (1996) 2157–2167.
[86] E. Warbrick, PCNA binding through a conserved motif, Bioessays 20 (1998) 195–199.
[87] R. Gary, D. Ludwig, H. Cornelius, The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21, J. Biol. Chem. 272 (1997) 24522–24529.
[88] V. Mocquet, J.P. Lainé, T. Riedl, Z. Yajin, M.Y. Lee, J.M. Egly, Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step, EMBO J. 27 (2008) 155–167.
[89] S.-L. Yu, M.-S. Kang, H.-Y. Kim, C.M. Gorospe, T.-S. Kim, S.-K. Lee, The PCNA binding domain of Rad2p plays a role in mutagenesis by modulating the cell cycle in response to DNA damage, DNA Repair (Amst) 16C (2014) 1–10.
[90] M. Tillhon, O. Cazzalini, T. Nardo, D. Necchi, S. Sommatis, L.A. Stivala, et al., p300/CBP acetyl transferases interact with and acetylate the nucleotide excision repair factor XPG, DNA Repair (Amst) 11 (2012) 844–852.
[91] S. Moriwaki, C. Nishigori, S. Imamura, T. Yagi, C. Takahashi, N. Fujimoto, et al., A case of xeroderma pigmentosum complementation group F with neurological abnormalities, Br. J. Dermatol. 128 (1993) 91–94.
[92] G.S. Salvesen, F.L. Scott, Pruning DNA: structure-specific endonucleases (XPF/Rad1/Mus81), Structure 11 (2002) 365–366.
[93] T. Nishino, K. Komori, Y. Ishino, K. Morikawa, X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease similarity between its endonuclease domain and restriction enzymes, Structure 11 (2003) 445–457.
[94] J.A. Roberts, M.F. White, An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81, J. Biol. Chem. 280 (2005) 5924–5928.
[95] L. Aravind, D.R. Walker, E.V. Koonin, Conserved domains in DNA repair proteins and evolution of repair systems, Nucleic Acids Res. 27 (1999) 1223–1242.
[96] M. Newman, J. Murray-Rust, J. Lally, J. Rudolf, A. Fadden, P.P. Knowles, et al., Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition, EMBO J. 24 (2005) 895–905.
[97] J.H. Enzlin, O.D. Schärer, The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif, EMBO J. 21 (2002) 2045–2053.
[98] Y.-J. Choi, K.-S. Ryu, Y.-M. Ko, Y.-K. Chae, J.G. Pelton, D.E. Wemmer, et al., Biophysical characterization of the interaction domains and mapping of the contact residues in the XPF–ERCC1 complex, J. Biol. Chem. 280 (2005) 28644–28652.
[99] K. Tripsianes, G. Folkers, E. Ab, D. Das, H. Odijk, N.G.J. Jaspers, et al., The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair, Structure 13 (2005) 1849–1858.
[100] D. Das, K. Tripsianes, N.G.J. Jaspers, J.H.J. Hoeijmakers, R. Kaptein, R. Boelens, et al., The HhH domain of the human DNA repair protein XPF forms stable homodimers, Proteins 70 (2007) 1551–1563.
[101] D. Das, G.E. Folkers, M. van Dijk, N.G.J. Jaspers, J.H.J. Hoeijmakers, R. Kaptein, et al., The structure of the XPF–ssDNA complex underscores the distinct roles of the XPF and ERCC1 helix- hairpin-helix domains in ss/ds DNA recognition, Structure 20 (2012) 667–675.
[102] R. Of, I. Of, X.P.F. With, R.P.A. In, E. Repair, Role of interaction of XPF with RPA in nucleotide, J. Mol. Biol. 413 (2012) 337–346.
[103] I. Kuraoka, E.H. Morita, M. Saijo, T. Matsuda, K. Morikawa, M. Shirakawa, et al., Identification of a damaged-DNA binding domain of the XPA protein, Mutat. Res. 362 (1996) 87–95.
[104] J.C. States, E.R. McDuffie, S.P. Myrand, M. McDowell, J.E. Cleaver, Distribution of mutations in the human xeroderma pigmentosum group A gene and their relationships to the functional regions of the DNA damage recognition protein, Hum. Mutat. 12 (1998) 103–113.
[105] I. Satokata, K. Tanaka, S. Yuba, Y. Okada, Identification of splicing mutations of the last nucleotides of exons, a nonsense mutation, and a missense mutation of the XPAC gene as causes of group A xeroderma pigmentosum, Mutat. Res. 273 (1992) 203–212.
[106] G. Yasuda, R. Nishi, E. Watanabe, T. Mori, S. Iwai, D. Orioli, et al., In vivo destabilization and functional defects of the xeroderma pigmentosum C protein caused by a pathogenic missense mutation, Mol. Cell. Biol. 27 (2007) 6606–6614.
[107] E.M. Taylor, B.C. Broughton, E. Botta, M. Stefanini, A. Sarasin, N.G. Jaspers, et al., Xeroderma pigmentosum and trichothiodystrophy are associated with different mutations in the XPD (ERCC2) repair/transcription gene, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 8658–8663.
[108] F. Coin, J.C. Marinoni, C. Rodolfo, S. Fribourg, A.M. Pedrini, J.M. Egly, Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH, Nat. Genet. 20 (1998) 184–188.
[109] K. Paszkowska-Szczur, R.J. Scott, P. Serrano-Fernandez, A. Mirecka, P. Gapska, B. Górski, et al., Xeroderma pigmentosum genes and melanoma risk, Int. J. Cancer 133 (2013) 1094–1100.
[110] K. Takayama, E.P. Salazar, B.C. Broughton, A.R. Lehmann, A. Sarasin, L.H. Thompson, et al., Defects in the DNA repair and transcription gene ERCC2(XPD) in trichothiodystrophy, Am. J. Hum. Genet. 58 (1996) 263–270.
[111] P. Shiyanov, S.A. Hayes, M. Donepudi, F. Nichols, S. Linn, B.L. Slagle, et al., The naturally occurring mutants of DDB are impaired in stimulating nuclear import of the p125 subunit and E2F1-activated transcription, Mol. Cell. Biol. 19 (1999) 4935–4943.

[112] V. Rapic-Otrin, True XP group E patients have a defective UV-damaged DNA binding protein complex and mutations in DDB2 which reveal the functional domains of its p48 product, Hum. Mol. Genet. 12 (2003) 1507–1522.
[113] M. Bogliolo, B. Schuster, C. Stoepker, B. Derkunt, Y. Su, A. Raams, et al., Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia, Am. J. Hum. Genet. 92 (2013) 800–806.
[114] K. Kashiyama, Y. Nakazawa, D.T. Pilz, C. Guo, M. Shimada, K. Sasaki, et al., Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia, Am. J. Hum. Genet. 92 (2013) 807–819.
[115] Y. Matsumura, C. Nishigori, T. Yagi, S. Imamura, H. Takebe, Characterization of molecular defects in xeroderma pigmentosum group F in relation to its clinically mild symptoms, Hum. Mol. Genet. 7 (1998) 969–974.
[116] A. Sijbers, P. van, V. Vader, Homozygous R788W point mutation in the XPF gene of a patient with xeroderma pigmentosum and late-onset neurologic disease, J. Investig. 110 (1998) 832–836.
[117] D.I. Zafeiriou, F. Thorel, A. Andreou, W.J. Kleijer, A. Raams, V.H. Garritsen, et al., Xeroderma pigmentosum group G with severe neurological involvement and features of Cockayne syndrome in infancy, Pediatr. Res. 49 (2001) 407–412.
[118] P. Lalle, T. Nouspikel, A. Constantinou, F. Thorel, S.G. Clarkson, The founding members of xeroderma pigmentosum group G produce XPG protein with severely impaired endonuclease activity, J. Invest. Dermatol. 118 (2002) 344–351.
[119] S. Emmert, H. Slor, D.B. Busch, S. Batko, R.B. Albert, D. Coleman, et al., Relationship of neurologic degeneration to genotype in three xeroderma pigmentosum group G patients, J. Invest. Dermatol. 118 (2002) 972–982.
[120] M. Wakasugi, The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair, J. Biol. Chem. 272 (1997) 16030–16034.
[121] S. Moriwaki, M. Takigawa, N. Igarashi, Y. Nagai, H. Amano, O. Ishikawa, et al., Xeroderma pigmentosum complementation group G patient with a novel homozygous missense mutation and no neurological abnormalities, Exp. Dermatol. 21 (2012) 304–307.

CAPÍTULO II

Dynamics of the DDB2-DDB1 protein complex and different mutants in Xeroderma Pigmentosum disease

Artigo a ser submetido para a revista *Nucleic Acid Research*

Nota sobre o Capítulo II

Esse capítulo é redigido em forma de artigo científico, já organizado nas normas da revista em que ele será submetido posteriormente. Contudo, diferente da versão de submissão, as figuras e tabelas foram inseridas no corpo do artigo para facilitar a leitura. Contudo, o Material Suplementar, contendo figuras e tabelas com informações adicionais, ainda se encontra no final do manuscrito, após o item de referências. O material digital será encaminhado por e-mail a cada membro avaliador e, posteriormente, eles poderão ser acessados juntamente com o artigo no futuro repositório vinculado a revista na qual ele será publicado.

Dynamics of DDB2-DDB1 complex and different mutants in Xeroderma Pigmentosum disease

Bruno César Feltes[1#], Conrado Pedebos[1], Diego Bonatto[1] and Hugo Verli[1*]

[1]Biotechnology Center of the Federal University of Rio Grande do Sul, Department of Molecular Biology and Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, RS – Brazil.

Short title: Dynamics of the DDB2-DDB1 complex

[*] **Corresponding author:**
Hugo Verli
Centro de Biotecnologia da UFRGS
Departamento de Biologia Molecular e Biotecnologia
Universidade Federal do Rio Grande do Sul - UFRGS
Avenida Bento Gonçalves 9500 - Prédio 43431
Caixa Postal 15005
Porto Alegre – Rio Grande do Sul
BRAZIL
91500-970
Phone: (+55 51) 3308-6060
Fax: (+55 51) 3308-7309
Contract/grant sponsor: CNPq, FAPERGS, CAPES

[#] **Co-corresponding author:**
Bruno César Feltes
Centro de Biotecnologia da UFRGS
Departamento de Biologia Molecular e Biotecnologia
Universidade Federal do Rio Grande do Sul - UFRGS
Avenida Bento Gonçalves 9500 - Prédio 43431
Caixa Postal 15005
Porto Alegre – Rio Grande do Sul
BRAZIL
91500-970
Contract/grant sponsor: CAPES

Abstract: Xeroderma Pigmentosum (XP) is a disease caused by mutations in the nucleotide excision repair (NER) pathway, a molecular mechanism responsible for removing UV-induced DNA lesions. Patients with XP exhibit high propensity to skin cancers, and can present neurological impairments and premature aging. During NER, the DNA lesion recognition is performed by the DDB-Complex, composed by DDB2 (XPE) and DDB1. However, not much is known about how those mutations affect XP proteins structure and complex assembly. Thus, we searched for structural evidences associated to the role of three naturally occurring mutations found on DDB2 on XPE patients on the DDB-complex: R273H, K244E, and L350P. Through a series of molecular dynamics simulations, DDB2 mutation promoted loss of flexibility in the overall protein structure, producing a different conformational behavior in comparison to the wild-type, especially in a region comprising residues 354 to 371, which appeared to impact on protein molecular behavior. Furthermore, the DDB complex containing the mutated forms of DDB2 showed distinct behaviors for each mutant: R273H displayed higher instability when complexed; L350P affected DDB1 protein-protein binding with DDB2; and K244E, similar to the L350P mutant, loss all interacting residues when compared to the WT. The structural data gathered throughout the analyzes helps in the understanding of how naturally-occurring mutations found in XPE patients impact on DDB2 and DDB1 structural behavior. The information can also aid to delineate future experiments aiming to explain this complex and neglected disease.

Key-words: DDB-Complex, DDB2, DDB1, Xeroderma Pigmentosum, Molecular Dynamics, DNA repair, Protein Network

1. Introduction

UV-induced mutations can lead to tumor formation and the development of diseases, such as Cockayne Syndrome (CS), Trichothiodystrophy (TTD) and Xeroderma Pigmentosum (XP) (1). In eukaryotes, the nucleotide excision repair (NER) pathway is responsible for removing such DNA lesions and restoring the DNA molecule (1, 2). The NER pathway is majorly coordinated by the XP proteins (XPPs), that range from XPA to XPG (1, 2). Each XPP is engaged to a given role in NER, which are: (i) lesion recognition, promoted by XPC and DDB2, the latter being the product of the XPE gene; (ii) DNA unraveling and opening, promoted by XPB and XPD; (iii) assembly of the excision machinery, coordinated by XPA and; (iv) excision of the damaged site, promoted by XPF an XPG (2). Mutation in any XP gene can potentially cause CS, TTD and XP development (1, 3).

Although there is a considerable amount of data regarding the structure of XPP from humans and model organisms, such as *Danio rerio*, *Sulfolobus acidocaldarius* and *Thermoplasma acidophilum,* little is known about how the mutations found in XP patients impact on human XPPs molecular behavior, as well as on their complexes assembly (3). One particular case is the DDB-Complex, which is composed by DDB2, a DNA-binding protein of 427 amino acid residues, the gene product of XPE and characterized by a single WD β-propeller region; and DDB1, an 1140 residues DNA-binding protein, composed by three WD β-propeller domains (3, 4), and both DDB2 and DDB1 are necessarily for proper damage recognition during NER. DDB2 is the active damage recognizer, having a higher affinity for cyclobutane pyrimidine dimers (CPD) adducts, and DDB1, although not related to XP development, is necessary for anchoring DDB2 on the DNA molecule for proper recognition (3, 4). In this sense, XP, type E patients, (individuals caring mutations on the XPE gene) produce a defective DDB2 protein, which compromise DNA damage recognition by the DDB-complex. Additionally, These patients have 1000× more probability to develop skin cancers among the XP types, making its understand a relevant subject for the biomedical sciences (5).

Thus, to understand how common naturally-occurring mutations found in XPE patients affects the DDB-Complex, three of the four known missense mutations in

DDB2 were modeled in the current work: L350P and K244E, related to instability of the DDB-Complex (6, 7), and R273H, described to impair DDB2-DDB1 interaction (7, 8). Multiple molecular dynamics and dynamic residues interaction networks (DRIN) were performed in triplicates for each system containing DDB2 and its mutant variants, as well for triplicates of systems containing DDB2 complexed with DDB1, and the same complex composed by DDB2 mutants.

2. Experimental Procedures

2.1. *Mutants modeling and validation*

The human DDB-Complex, composed by DDB2 and DDB1, is available under the PDB code 3EI4 (9). Both DDB2 and DDB1 have missing portions of their structures; therefore, the proteins were separated, and these gaps were reconstructed using different software. The *automodel* class from Modeller 9.10 (10), using the native protein sequence as target and the available crystallographic structure as template was employed to fill the missing segments of DDB1, whereas the missing N-terminal portion of DDB2 (residues 1 to 74) was reconstructed using PHYRE 2 [http://www.sbg.bio.ic.ac.uk/phyre2] (11). The final structures, henceforth known as $DDB2^{WT}$ and $DDB1^{WT}$, were submitted to the server SWISS-MODEL [http://swissmodel.expasy.org/] (12) for validation, using PROCHECK for stereochemical assessment (13), Anolea for non-local atomic interaction energy evaluation (14) and Qmean for combined estimation of local and per-residue quality of the model (15). The mutant variants of DDB2 ($DDB2^{R273H}$, $DDB2^{K244E}$, and $DDB2^{L350P}$) were individually designed employing PyMOL [https://www.pymol.org/], using $DDB2^{WT}$ as a template. Side-chain rotamers were selected based on similarity to the conformation already available for the native residue in the mutated position. Finally, the reconstructed proteins were reassembled using DeepView (16) and divided into four different complexes: (i) $DDB1^{WT}$ - $DDB2^{WT}$, named Co-DDB^{WT}; (ii) $DDB1^{WT}$ - $DDB2^{R273H}$, termed Co-DDB^{R273H}; (iii) $DDB1^{WT}$ - $DDB2^{K244E}$, termed Co-DDB^{K244E} and; (iv) $DDB1^{WT}$ - $DDB2^{L350P}$, titled DDB-Co^{L350P}.

2.2. *Molecular dynamics simulations*

For the molecular dynamics simulations, the GROMACS 4.5.1 and GROMACS 4.6.4 (17, 18) simulation suites were employed. In this sense, all previously mentioned proteins were then inserted in dodecahedral boxes and solvated with SPC (19) water model under periodic boundary conditions. The force field GROMOS54a7 (20) was employed in the simulations, and Na^+ or Cl^- counterions were added to neutralize the systems when necessary. Covalent bonds were constrained using the LINCS algorithm (21), and an integration step of 2 fs was applied. The Particle Mesh Ewald (PME) (22) method was employed for calculation of electrostatic interactions. The Parrinello-Rahman barostat (23, 24) was employed, with a 2.0 ps coupling constant, while the V-rescale (25), was used with a coupling constant of $\tau = 0.1$. Steepest Descent algorithm was used in the energy minimization step, prior to the simulations. Three independent simulations of 200ns each were performed for each of the studied systems, generating new velocities at the beginning of each run, in an effort to analyze multiple results to exclude low probability phenomena, and the results are presented as a mean result of the triplicates of each system.

2.3. *Dynamic Residue Interaction Networks*

The final trajectory of each independent system was then submitted to a Dynamic Residue Interaction Network (DRIN) analysis, where the trajectory files are concatenated by the program CatDCD [http://www.ks.uiuc.edu/Development/ MDTools/catdcd/] and latter transformed into networks and analyzed in the VMD 1.9.1 platform (26) using the program Carma 0.8 (27). In this method, each alpha carbon atom of each residue is represented by a node, and the motion correlation between the nodes is depicted by an edge. The weight of the motion correlation between nodes during the simulated time is represented by the thickness of the edges, thus, the stronger the motion correlation between residues, the thicker the edge. The analysis also takes into consideration the number of communities (independent groups of residues with strong motion correlation among themselves). The molecular behavior and flexibility of a protein is directly linked to the motion of individual, pairs, and groups of residues (28), and DRIN can be a powerful approach to understand the

molecular behavior and impact of mutations on protein structure. Other works have employed similar analysis, please refer to (29) and (30).

3. Results and Discussion

3.1. Loss of stability, flexibility, and molecular motion of DDB2 mutants

The final model for $DDB2^{WT}$ showed 85% residues in most favored regions and 10.6% in additional allowed regions of the Ramachandran Plot(**S-Table.1,** see **Supplementary Material**). In addition, the great majority DDB2 regions were in in favorable energy environments and good overall quality (**S-Fig.1,** see **Supplementary Material**).

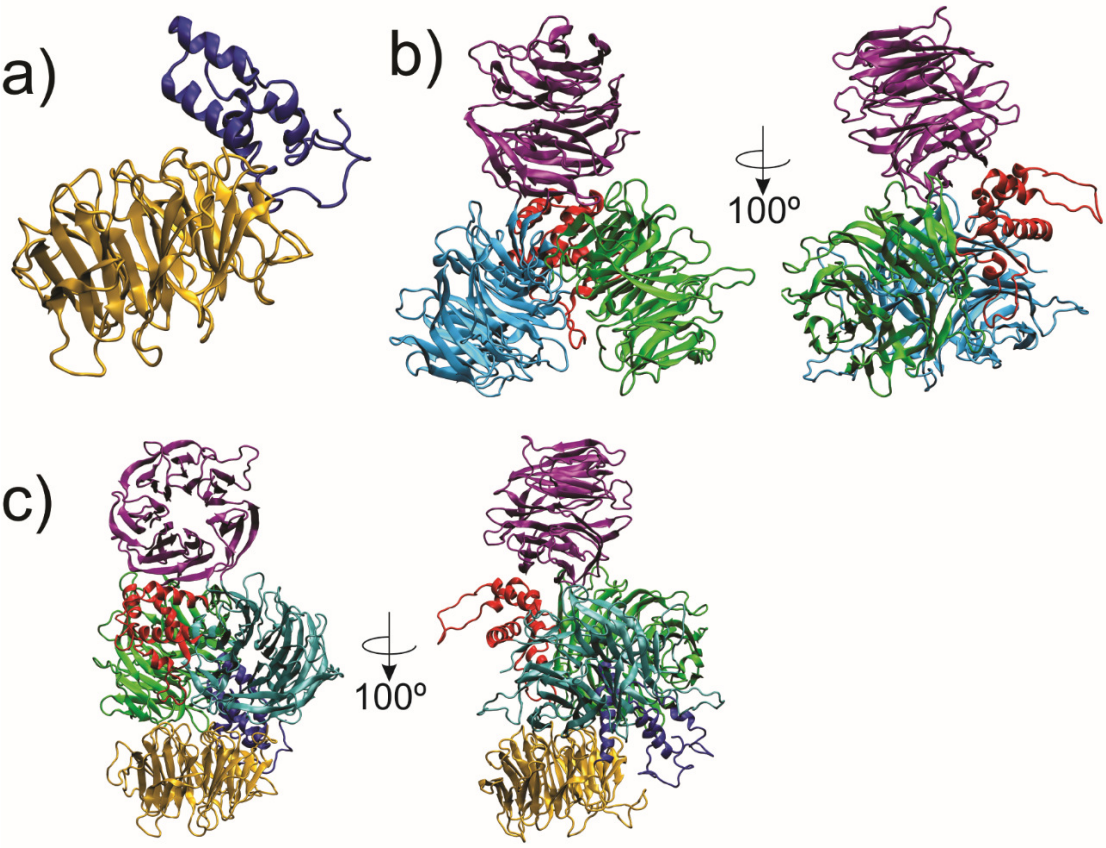


Fig.1. DDB1 and DDB2 with their domain divisions, and the DDB-Complex. **A)** Division of the two major portions of DDB2. Colored in blue and located in the N-terminus, this flexible domain ranges from residue 1 to 101. Residues 1 to 74 were modeled (*ab initio*) by the prediction tool PHYRE 2. From residue 102 to 427 lies the β-propeller domain of DDB2 (colored in gold). **B)** Division of the four distinct domains of

DDB1: (i) β-propeller A (in green); (ii) BPB (in purple) and; (iii) BPC (cyan). In red lies the α-helix region that lies in the back of DDB1, between the three propellers. **C)** DDB2 complex. Note that DDB2 is inserted between the BPA and C.

Afterwards, DDB2 was divided in two domains for examination: (i) a flexible domain located in its N-terminus, ranging from residue 1 to 101 (**Fig.1A**) and; (ii) the β-propeller, which ranges from residue 102 to 427 (**Fig.1A**). Accordingly, mutations on DDB2 were able to modulate the dynamics of the flexible domain (**Fig.2A-F**). It comes as no surprise that the β-propeller displayed lower stability in comparison to the flexible domain, even in the mutated variants. Evolutionarily, β-propellers structures are extremely stable, especially due to their nature of forming a funnel-shaped structure with numerous molecular interactions that strongly stabilizes the β-sheets blades inside the propeller (31, 32). However, a region comprising residues 354 to 371, in the β-propeller, had an increased rigidity upon mutation (**Fig.2D-F**). This is a β-turn region, which is near the flexible domain (**S-Fig.3,** see **Supplementary Material)**, and such location implies that it might affect the motion of this domain and the conformation of DDB2 due to the role of turns in molecular structure. In this sense, β-turns are crucial structures that change the direction of a given polypeptide chain and that could play a role in protein folding and, more important, in the stability of the native state, and intra/inter molecular interactions (33).

Hence, to investigate the connection between this region and the flexible domain, as well as the inter-domain connection, an interaction energy analysis was employed. Overall, there was a slight change in the mean interaction energy between the flexible domain and the β-propeller, because of DDB2 R273H and L350P mutations (**Fig.2H**), suggesting a loss of inter-domain interaction for those two mutations. The $DDB2^{R273H}$ and $DDB2^{L350P}$ mutations also promoted a stronger interaction between the 354-371 region and the flexible domain (**Fig.2G**). $DDB2^{WT}$ displayed almost no interaction energy between the two portions for the first 150ns, whereas $DDB2^{R273H}$ and $DDB2^{L350P}$, were the opposite. However, $DDB2^{K244E}$ showed no interaction energy for the entire simulation.

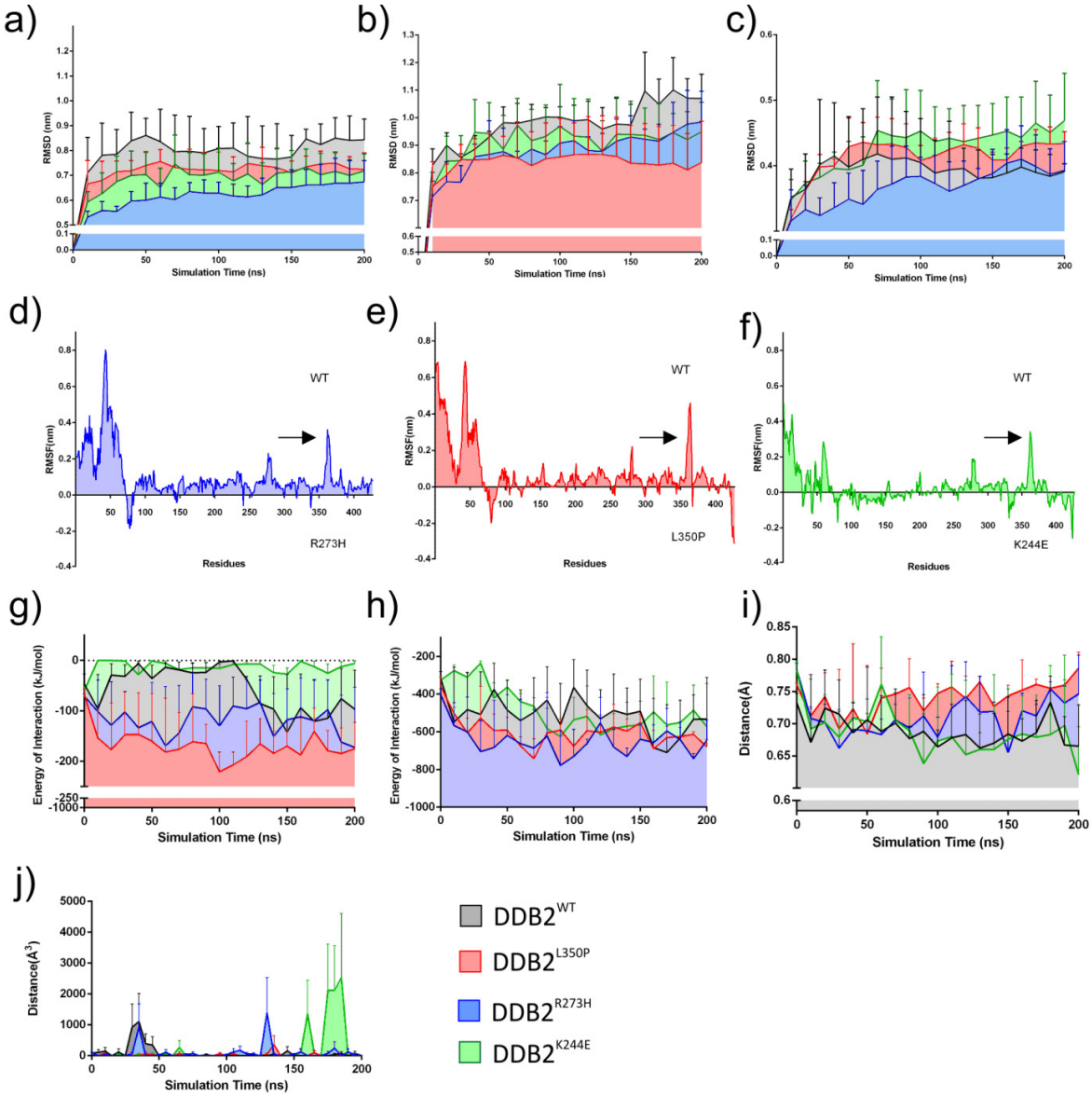


Fig.2. Mean RMSD, RMSF, and interaction energy, cavity volume and atomic distance between critical pair of residues for DDB2^{WT} and DDB2 mutant variants. **A)** Mean RMSD for the flexible domain (res 1 to 101). **B)** Mean RMSD for the β-propeller domain (res 102 to 427). **C)** Mean RMSD for the full DDB2 structure. **D-F)** Mean RMSF for the triplicates of DDB2^{WT} vs DDB2^{R273H}, DDB2^{L350P} and DDB2^{K244E}, respectively. The upper values represent higher variation in the WT, whereas the bottom values represent residues that displayed higher variation in the mutants. The arrows marks a region in the β-propeller that displayed similar variation in all mutants. **G-H)** Mean energy of interaction between the flexible domain and the 354-371 region, and the flexible domain and the β-propeller, respectively. **I)** Mean interaction distance

between Val291 and Thr306 for DDB2^{WT} and its mutants. **J)** Mean cavity volume for DDB2 β-propeller and its variants.

The results imply that DDB2^{WT} experience an interaction between the flexible domain and the 354-371 region 150ns. However, the DDB2^{R273H} and DDB2^{L350P} mutations undergo such interaction since the beginning, and that could not only affect the mobility of the flexible domain but this phenomenon can also play a role in the conformational change that DDB2 experience before its complexation with DDB1 in later moments. However, the DDB2^{K244E} mutant does not sustain such behavior and does not change conformation, denoting a loss of this event.

In agreement with the previous results, DDB2^{WT} presented a different behavior than its mutant variants in all triplicates for each systems (**S-Video 1,** see **Supplementary Material**). In this sense: (i) the flexible domain of DDB2^{WT} manifested a left-to-right swipe-like movement and a compacted conformation; (ii) DDB2^{R273H} loss the compacted conformation and expanded the flexible domain down, losing the swipe-like movement; (iii) in DDB2^{L350P}, the domain was not compacted and appeared to close on itself, also losing the movement seen in the WT, and; (iv) in DDB2^{K244E} although similarly compacted as DDB2^{WT}, it expanded the domain up and down. Moreover, the movement of the β-propeller was more difficult to perceive. However, the cavity of the β-propeller for DDB2^{K244E} was clearly higher, denoting a change in the β-propeller cavity behavior for this variant (**Fig.2J**). A similar result was observed for DDB2^{R273H}. This indicates that the DDB-Complex might be affected by the aberrant movement of the flexible domain, which could impact its insertion between the BPA and BPC of DDB1. The loss of the swipe-like movements it is also a noteworthy result since the flexible domain is the major interacting domain of DDB1. These observations aggregate to the idea that the connection between the 354-371 region and the flexible domain could lead to a loss of flexibility of the domain. In addition, the opening of DDB2 β-propeller in two variants in the cavity analysis (**Fig.2J**), especially in DDB2^{K244E}, indicates a motion change in the β-propeller.

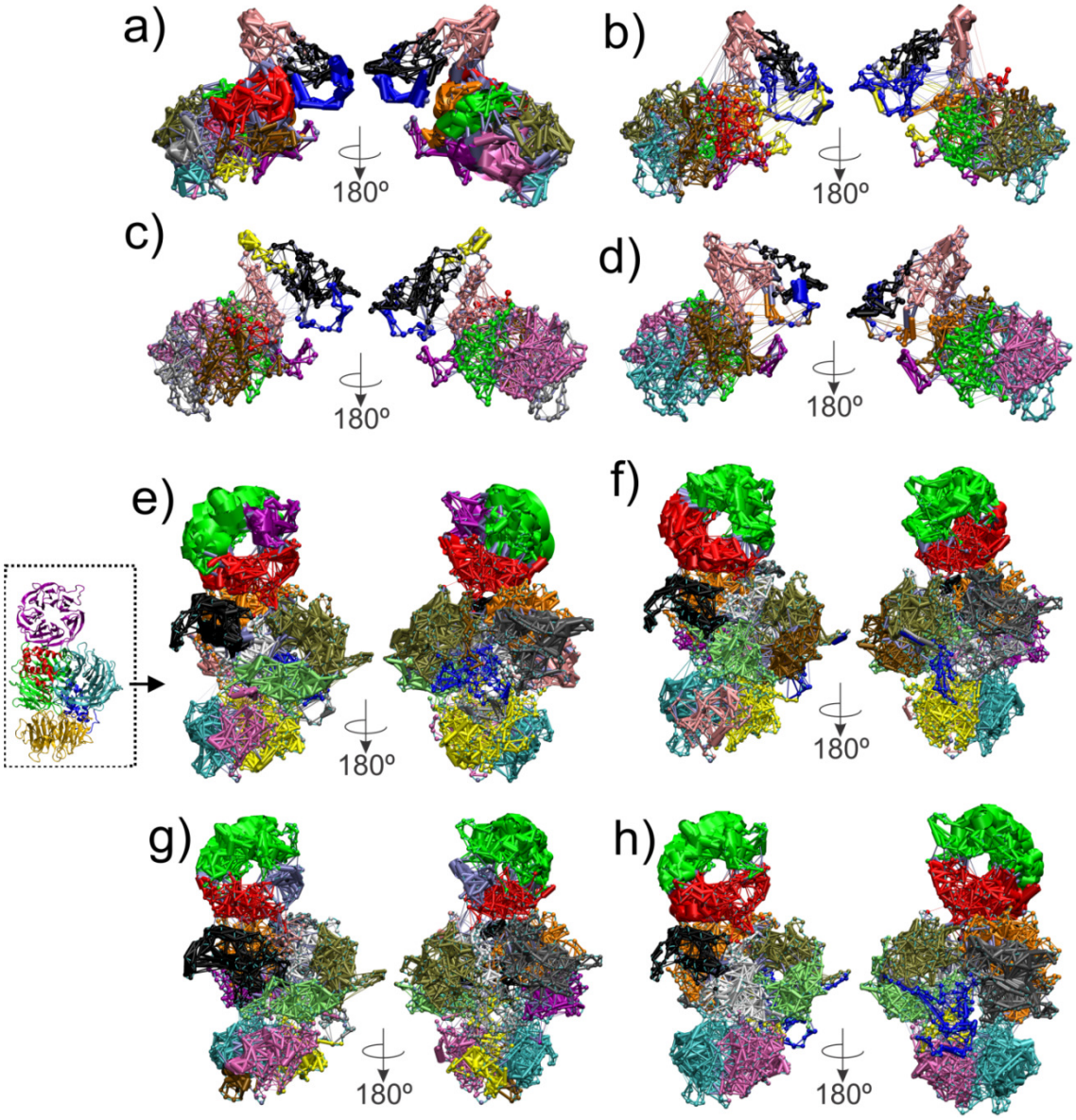


Fig.3. The dynamic residues interaction networks (DRIN) for DDB2^{WT} and its mutant variants, as well for the DDB-Complex, in the first molecular dynamic analysis, depicting the communities and motion correlation between residues. **A-D)** DDB2^{WT}, DDB2^{R273H}, DDB2^{L350P} and DDB2^{K244E}, respectively. Note that the mutants displayed thinner edges than the WT. **E-H**) Co-DDB2^{WT}, Co-DDB2^{R273H}, Co-DDB2^{L350P} and Co-DDB2^{K244E}, respectively

Correspondingly, mutations on DDB2 do not produced major modifications in its dynamics, considering the transmission of conformational information along the protein and the formation of communities of residues with correlated movement (**Fig.3A-D**). The mean number of communities between the three independent systems

displayed no apparent change when compared to the WT (**S-Table 2,** see **Supplementary Material**), with the exception of DDB2^{R273H}. Moreover, one difference between the WT and the mutants was the reduced transmission of conformational information between residues, which were seen in all systems (**Fig.3A-D**). In this sense, the mutants showed thinner edges, which denotes that the general motion of the protein is affected in a mutated variant, indicating a possible loss of compromise of conformational transmission, especially in DDB2^{R273H}, which was the only mutant that showed loss of communities, thus, losing independent groups of residues motion.

In order to observe how the conformational information is transmitted along between residues communities and along the entire protein, the critical residues for the propagation of motion between groups, constraining conformational information and affecting the molecular behavior of the protein, were evaluated (**S-Fig.4,** see **Supplementary Material**). One pair of residues was common for all systems, Val291-Thr306, indicates that those residues could be essential for DDB2 behavior. In fact, the distance between Val 291 and Thr306 is higher in DDB2^{R273H} and DDB2^{L350P}, when compared to the WT (**Fig.2I**). On the other hand, three pairs were in common only among the mutants: Phe296-Trp344, Lys163-Asp208, and Trp344-Pro346.

The change of motion seen in the flexible domain raised the question if it was a result of a change of secondary structure. The β-propeller showed no distinct changes observed in the secondary structure analysis (**Fig.4B**). However, in the mutants the flexible domain presented a gain in α-helix and a loss of coils and bends (**Fig.4A**). The flexible domain is inserted between the BPA and BPC of DDB1 in the DDB-Complex (**Fig.1C**) and the loss of disordered portions, such as coils and bends could impact on the proper protein-protein interactions (PPI) binding (34).

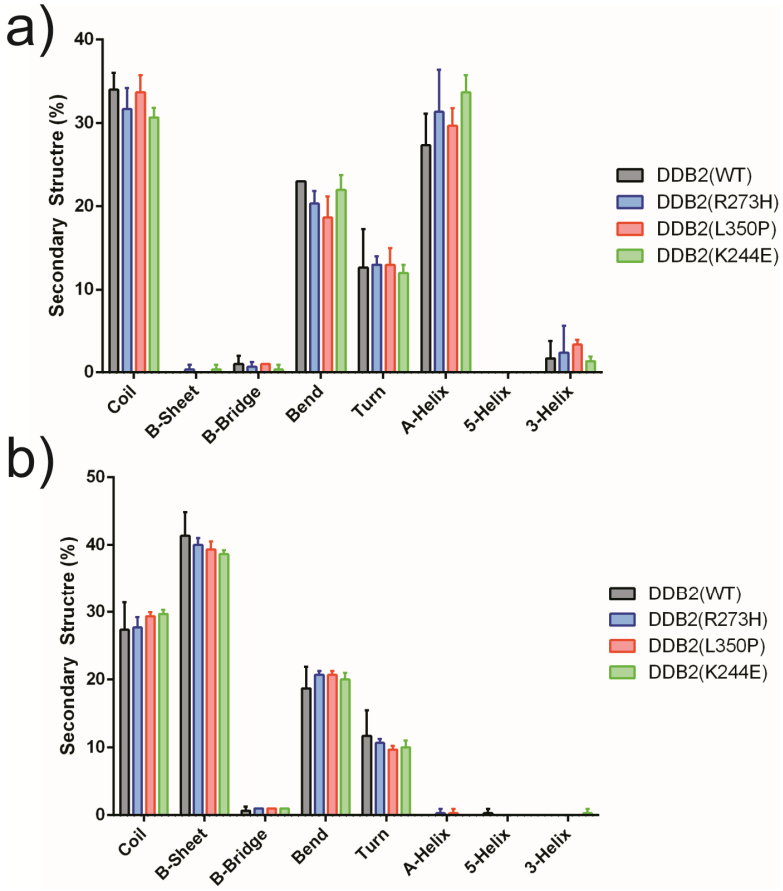


Fig.4. Mean percentage of secondary structures count for each domain along the simulation time of 200ns. **A)** Mean percentage of secondary structures count for the flexible domain. **B)** Mean percentage of secondary structures count for the β-propeller.

Afterward, we investigated if the mutations could affect the electrostatic potential of DDB2 in its initial state (1ns) and in the end of the simulation (200ns). $DDB2^{WT}$ was highly positive with only dispersed narrow negative and neutral pockets in the β-propeller (**Fig.5**), an expected result for a DNA-binding protein. In contrast, distinct changes were observed in $DDB2^{L350P}$ and $DDB2^{K244E}$ in comparison to $DDB2^{WT}$. In this sense, in these mutants, the inner cavity of the β-propeller in 1 and 200ns displayed negative and neutral charges (**Fig.5**). The same was not seen in $DDB2^{R273H}$, although in 1ns this mutation exhibit neutral charge in a considered portion of the β-propeller. Since the major function of DDB is to recognize DNA-damage, the loss of positively charged regions could affect its DNA-binding capability. In addition, for the

K244E mutants, the change of electrostatic potential in the inner cavity of the propeller could be a result of its opening, previous stated in the cavity analysis (**Fig. 2J**).

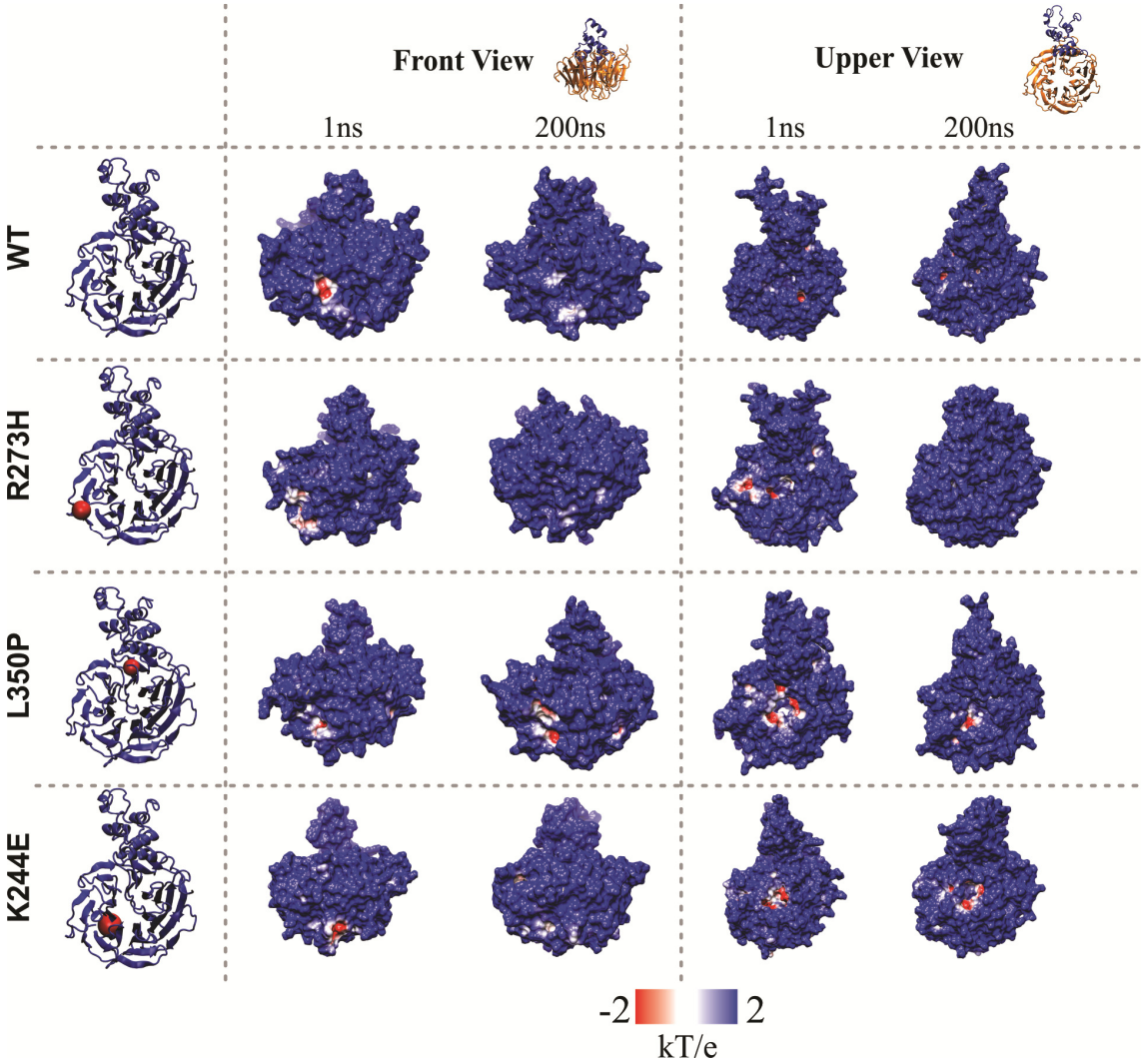


Fig.5. Representative images of the first and last nanoseconds (ns) of the triplicates for each system. On the left, in cartoonish view, the structure of DDB2 is depicted and the location of each mutation is indicated. On the top, the small images specify the position of the surface view of each protein. The electrostatic potential analysis was performed by the program UFSC Chimera, using APBS tool.

3.2. DDB1 displays a clamp-like movement

The final model of $DDB1^{WT}$ displayed 87.6% residues in most favored regions and 9.7% in additional allowed regions, and similar to DDB2, most regions were in

favorable energy environments (**S.Fig.2,** see **Supplementary Material**). Next, DDB1^{WT} was divided into its three β-propellers for better comprehension (**Fig.1B**). All propellers were notably stable, but the BPC displayed larger variance in comparison to the BPA and BPB, which showed similar RMSD values (**S-Fig.5A,** see **Supplementary Material**). Moreover, DDB1^{WT} exhibit distinct flexibility in the RMSF (**S-Fig.3B,** see **Supplementary Material**), which is expected since DDB1 is a PPI platform not only for DDB2 and its interaction with DNA but also the ubiquitin-ligase complex composed by CUL4A and ROC1/RBX1 (35, 36).

In terms of DRINs DDB1^{WT} showed high transmission of conformational information between residues and numerous communities (**S-Table 2; S-Fig.6,** see **Supplementary Material**). These large number individual groups of high motion correlation among residues are in agreement with an idea that DDB1 is a particularly flexible protein. In the PCA analysis, DDB1 manifested a distinct clamp-like movement between the BPA and BPC (**S-Video 3;** see **Supplementary Material**). This is the first time this phenomenon is seen for DDB1. The open-close movement between these two domains implicated that DDB1 clamps DDB2 to form the complex. Thus, the loss of the movement observed for the DDB2 flexible domain, discussed previously, could impact on DDB2 clamping by the BPA and BPC of DDB1, forming an unstable complex that does not sustain its binding long enough for proper DNA-damage recognition and recruitment of the whole NER complex. In this scenario, the BPC appears to be the most important domain, since it was the higher motion.

3.3. DDB-Co^{R273H} shows abnormal behavior after complexed

The impact of each mutation in a complexed DDB1-DDB2 (Co-DDB1 and Co-DDB2) was assessed. In this sense, Co-DDB2^{R273H} displayed higher variation a slight change in flexibility in the flexible domain when compared to the WT (**Fig.6A-C**). However, DDB2 β-propeller was the most affected domain (**Fig.6B-C**). It presented the highest variation in the RMSD among all variants, and Co-DDB^{R273H} was the only mutation in which the 354-371 region displayed higher motion, even after the complexation of the two proteins (**Fig.6C**). By analyzing the Co-DDB1 β-propellers, we observed that there was no major change in structure variation of Co-DDB1$^{DDB2(R273H)}$ BPB when compared to the Co-DDB1^{WT}, but a higher variance in BPA and a lower

alteration in the BPC (**Fig.7A-D**). No loss of flexibility was seen (**Fig.7D**). These results point that the higher motion detected in Co-DDB2^{R273H} in this system was might not be a consequence of a propagation of structure disturbance that could affect BPA or BPC of DDB1, but due to an aberrant behavior of the complexed DDB2^{R273H} alone.

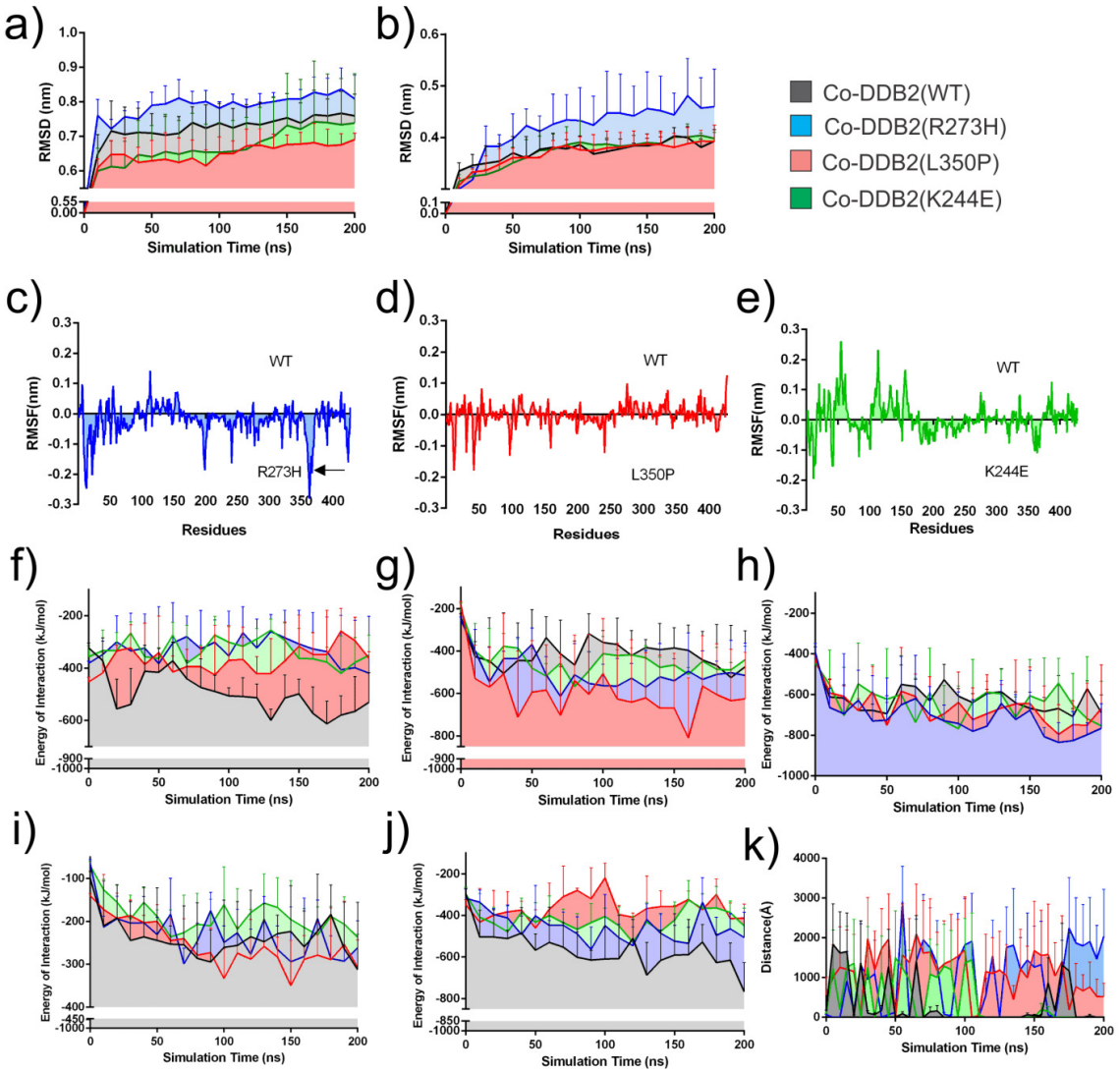


Fig.6. Mean RMSD, RMSF, and interaction energy, and cavity volume for the triplicates of the complexed DDB2 WT (Co-DDB2^{WT}) and Co-DDB2 mutant variants. **A)** Mean RMSD for the flexible domain (res 1 to 101). **B)** Mean RMSD for the β-propeller domain (res 102 to 427). **C-E)** Mean RMSF for the triplicates of Co-DDB2^{WT} vs Co-DDB2^{R273H}, Co-DDB2^{L350P} and Co- DDB2^{K244E}, respectively. The upper values represent higher variation in the WT, whereas the bottom values represent residues that displayed higher variation in the mutants. The arrow marks the 354-375 region. **F)** Mean interaction energy between Co-DDB2 β-propeller and flexible domain in the four

variants. **G-H)** Mean interaction energy between Co-DDB2 flexible domain and the BPA and BPC of DDB1, respectively. **I-J)** Mean interaction energy between Co-DDB2 β-propeller and the BPA and BPC of DDB1, respectively. **K)** Mean cavity analysis for Co-DDB2 in the four variants.

The PCA analysis revealed distinct changes in Co-DDB^{R273H} when compared to the WT: (i) in Co-DDB2^{WT} the β-propeller motion was similar to DDB2^{WT}, whereas in Co-DDB2^{R273H} it behaved like it was compressed from the right side-to-center (**S-Videos4 and 5, ,** see **Supplementary Material**); (ii) two coil portions of the flexible domain in Co-DDB2^{WT} approached the β-propeller, however, in Co-DDB2^{R273H}, the flexible domain showed no such movement; (iii) The BPB of Co-DDB1^{WT} showed a strong side-to-side motion in direction to the BPC, but in Co-DDB1$^{DDB2(R273H)}$ the BPB twisted in its own axis; (iv) the BPA and BPC of Co-DDB1^{WT} loss the clamp-like movement, but showed a small open-close motion. Co-DDB1$^{DDB2(R273H)}$ exhibit similar changes, but slightly less (**S-Videos6 and 7; ,** see **Supplementary Material**).

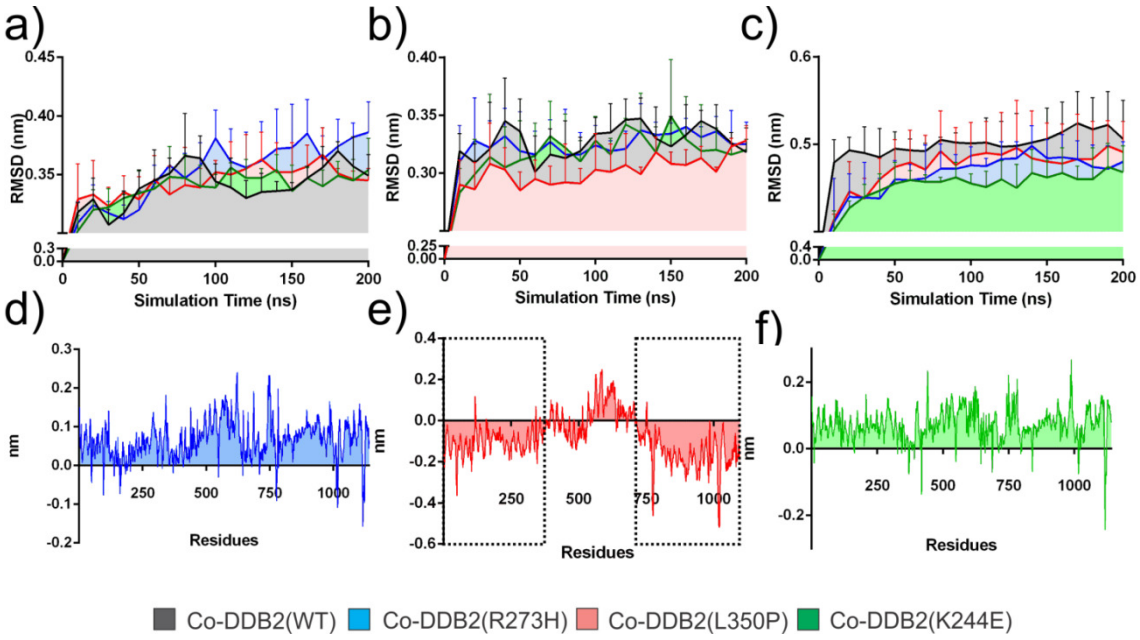


Fig.7. Mean RMSD and RMSF for the triplicates of the complexed DDB1 WT (Co-DDB1^{WT}) and Co-DDB1^{WT} complexed with the different DDB2 mutant variants (Co-DDB1$^{DDB2)}$. **A-C)** Mean RMSD for the BPA, BPB and BPC, respectively**. D-F)** Mean RMSF for Co-DDB1^{WT} vs Co-DDB1$^{DDB2(R273H)}$, Co-DDB1$^{DDB2(L350P)}$ and Co-DDB1$^{DDB2(K244E)}$, respectively. The doted squares correspond to the altered regions of BPA and BPC of Co-DDB1$^{DDB2(L350P)}$, respectively.

Moreover, in agreement with the isolated systems, the DRIN analysis showed that Co-DDB^{R273H} exhibit lower transmission of conformational information between residues (**Fig.4F**), revealing that the mutants not only impact on the overall molecular behavior DDB2 but in the complex as well. Nevertheless, no change was detected in the number of communities of Co-DDB2^{R273H}, although Co-DDB1$^{DDB2(R273H)}$ showed loss of communities (**S-Table 2,** see **Supplementary Material**). We further analyzed the DRINs for the interacting residues between DDB1 and DDB2. All interacting residues were obtained and only the common pair of residues shared among the triplicates of each system was taken into consideration. When comparing the interacting residues between DDB2 and DDB1 for Co-DDB^{R273H} vs Co-DDB^{WT}, a change of interaction of 11 pairs of residues was perceived (**S-Fig.7,** see **Supplementary Material; Table 1**). These results indicate that the instability caused by this mutation not only impact on the general transmission of conformational information but also their PPI connection. Nonetheless, Co-DDB^{R273H} was the only variant to show some similarity between the WT complex, strengthening the hypothesis that the R273H mutation does not destabilize DDB1, and thus, can still maintain interacting residues. For example, four pairs of interaction residues are near the 354-371 region in DDB2 and the BPC of DDB1, and those residues are sustained in this mutation (**Table 1**). In addition, for DDB2, in the WT complex, 56% of the residues that interact with DDB1 reside in the flexible domain; this is basically sustained in Co-DDB^{R273H}. No critical residues, shared between the triplicates of each system, were found in common for the interaction of DDB1 and DDB2 (**S-Table 3,** see **Supplementary Material**).

Interaction energy revealed that Co-DDB2^{R273H}, such as all mutant variants, displayed less interaction between the flexible domain and the β-propeller of DDB2, depicting a loss of inter-domain connection in DDB2 (**Fig.6F**). Moreover, the interaction between the flexible domain and the BPA of DDB1 appears to be altered in Co-DDB2^{R273H} (**Fig.6G**). The loss of inter-domain interaction change seen in DDB2, as well for loss of the PPI interaction between the flexible domain and the BPA and BPC, observed in this mutation, could explain the movement alteration seen in the PCA analysis, such as the loss of the clamp-like movement. The fact that we see that DDB2 slightly moves toward the BPC supports the loss of interaction seen for the flexible domain and the BPA, but not as much for the BPC (**Fig.6G-H**). Analysis also revealed

that the interaction between the Co-DDB2^{R273H} β-propeller and the BPC of DDB1 is altered after 50ns, where the WT is strongly connected to the BPC, but the same is not noticed for the BPA (**Fig.6I-J**). Additionally, Co-DDB2^{R273H} cavity was greatly expanded, when compared to the WT (**Fig.6K**). This results corroborates with the high variation for this variant, and suggests a destabilization of the β-propeller. Cavity changes could also impact on the DNA-binding of DDB2.

Table 1. Table listing all interacting residues between DDB2 and DDB1 found in common among the triplicates of each system.

Complex Variant	Interacting residues*
Co-DDB^{WT}	122-950; 318-139; 319-159; 319-161; **347-928**; **347-952**; **347-954; 348-928**; 402-929; 403-929; 403-948; 63-839; 70-815; 71-813; 73-913; 74-813; 76-1004; 77-1006; 78-329; 78-359; 78-361; 78-381; 80-329; 83-954; 83-971
Co-DDB^{R273H}	122-950; 253-1080; 256-112; 257-113; 318-140; 319-159; 319-160; 324-163; **347-928; 347-952; 347-954; 348-928**; 73-913; 74-723; 74-813; 77-1006; 78-329; 78-359; 78-361; 78-382; 83-954; 83-971; 87-954; 90-927; 96-840
Co-DDB^{L350P}	77-1003; *78-1005*; *258-112*; 301-112; 300-114; *325-162*; 79-328; *80-328*; 81-328; *79-358*; *79-360*; 79-380; 81-382; *72-812*; *71-814*; 64-838; 427-906; 425-907; 406-909; 425-909; 74-912; 78-913; 74-926; ***348-927; 349-927; 348-951***; 84-953; **348-953**; *84-970*
Co-DDB^{K244E}	*78-1005*; 78-1033; 254-1079; *258-112*; 318-112; 320-158; *325-162*; *80-328*; *79-358*; *79-360*; 71-812; *72-812*; *71-814*; 404-909; 91-910; 91-926; ***348-927; 349-927***; 403-928; 123-949; 348-951; 88-953; 77-970; *84-970*

* The first residue in each given pair, always correspond to DDB2, whereas the second correspond to a residue in DDB1. The underlined residues are common between Co-DDB^{WT} and CO-DDB^{R273H}, and the pair of residues in italic corresponds to residues in common between Co-DDB^{L350P} and Co-DDB^{K244E}. The bold residues correspond to the DDB2 residues near the 354-371 region and the BPC of DDB1. Interacting residues were obtained from the DRIN analysis.

The central finding in this system is that the 354-371 region of Co-DDB2^{R273H} displayed higher motion after complexation, which directs to the idea that it could not be stabilized after the complexation, a phenomenon that is not observed in the other mutations or in the WT (**Fig.6D-E**). In addition, the most affected connection appears to the between Co-DDB2^{R273H} flexible domain and the BPA of DDB1, as well for the inter-domain interaction of Co-DDB2^{R273H}. This mutation is described to disrupt the PPI of DDB2 and DDB1 (7, 8). While we did not witness the disruption in the time scale simulated, the results indicate that DDB2^{R273H} is less stable compared to DDB2^{WT}, and the other mutations, and without a direct complexation, such as we did in this work, it probably would not bind to DDB1. Co-DDB2^{R273H} also displayed loss of motion correlation, critical pair of residues and point to a destabilization of Co-DDB2^{R273H} β-propeller.

3.4. Co-DDB^{L350P} generates anomalous binding to DDB1

Next, we sought to understand the effect of the DDB2^{L350P} mutation in the DDB complex. There was no change in the structure variation of in the flexibility for Co-DDB2^{L350P}, except for a small lower variance in the flexible domain RMSD (**Fig.6A-B and D**). The 354-371 region also did not display any variance. Therefore, we aimed to analyze DDB1 for answers. The BPB appeared to be slightly affected, exhibiting a distinct lower variance (**Fig.7B**). The BPC also had a greater standard deviation when compared to the other variants (**Fig.7C**). However, the true change was detected in the flexibility, where distinct alterations for Co-DDB1$^{DDB2(L350P)}$ can be observed in the BPA and, especially, in the BPC region (**Fig.7E)**. This indicates that the L350P mutation, in contrast to the R273H, causes disturbances in DDB1.

When analyzing the PCA for this mutation, a different scenario was observed from the previous mutation: (i) in Co-DDB2^{L350P}, the flexible domain has a higher movement than the WT, showing a lack of stabilization; (ii) the β-propeller of Co-DDB2^{L350P} is compressed on the left side and pushes the BPA of Co-DDB1$^{DDB2(L350P)}$ up, whereas in the WT the BPA is dragged down, and the contrary was perceived for the BPC; (iv) the BPC hurls a coil region upon the BPA, something not seen in the WT; (v) the BPA and BPC are closer to each other than the WT, probably due to the aberrant

conformation of the flexible domain; and (vi) the BPB moved side-to-side, but less than the WT (**S-Videos4 and 6,** see **Supplementary Material**). These observations explain the previous results, showing how the BPA and BPC are so distinctively altered in the RMSF, and the changes in the interaction energy as seen bellow.

In agreement with the previous data, DRIN analysis for Co-DDB^{L350P} presented lower transmission of conformational information among residues and fewer communities for DDB1 (**Fig.4G**), but the most noteworthy data for this system was the fact that Co-DDB^{L350P} shared no interacting residues between DDB1-DDB2 with the other systems (**S-Fig.7,** see **Supplementary Material; Table 1**). This indicates that, when compared to the WT, the residues interface between the two proteins is completely disrupted. Additionally, Co-DDB^{L350P} was the mutant variant that presented the highest values of interaction energy between the flexible domain and BPA, and the lowest between the β-propeller and the BPC, reinforcing the idea that, in this mutation, the major problem resides in an aberrant PPI connection, probably due to a propagation of aberrant motion in the interface between DDB1 and DDB2 (**Fig.6G and J**; **S-Fig.7,** see **Supplementary Material**).

Furthermore, Co-DDB1$^{DDB2(L350P)}$ was the only complexed DDB1 that exhibited a loss of flexibility and lower variation for DDB1 BPB than the other mutations (**Fig.7B and E**). This could indicate that the ubiquitin-ligase complex composed by CUL4A and ROC1/RBX1 that binds to the BPB of DDB1 can be impaired in this mutation. The ubiquitination of seven residues in the flexible domain of DDB2, by CRL4, increases its stability against degradation (3, 37), and our results point that this phenomenon might be affected by this mutation, by changing BPB movement and, therefore, the proper ubiquitination of DDB2.

It is important to highlight that among the interacting residues shared between DDB1-DDB2, four connection are near the 354-371 region of DDB2 and the BPC in DDB1, and are absent in Co-DDB^{L350P} (**Table 1**). All variants contain residues that are near the 354-371 region and interact with BPC, however, in Co-DDB^{L350P}, as well for Co-DDB^{K244E}, the interacting residues suffer a displacement of ± 1 residue in both DDB2 and DDB1. Co-DDB2^{L350P}, similarly to Co-DDB2^{R273H} displayed the second greatest alterations in DDB2 β-propeller cavity (**Fig.6K**). This expansion of the inner cavity could explain the lower interaction of DDB2 β-propeller with the BPC.

Thus, the L350P mutation in DDB2 generates an unstable complex due to abnormal PPI connection between DDB2 and the BPA and BPC of DDB1, caused by loss of interaction between the β-propeller and BPC and the higher connection between the flexible domain and the BPA. Moreover, the complete loss of similarity in the pairs of interacting residues when compared to the WT complex and aberrant DDB1 stability that could affect not only its proper binding to DDB1, but also the complex assembly of the ubiquitin-ligase complex in the BPB of DDB1, appears to be crucial to understand this mutation.

3.5. Co-DDB^{K244E} generates unstable PPI connection

Finally, Co-DDB^{K244E} revealed to be the least affected variant for both DDB2 and DDB1 (**Figs. 6A-B** and **E** and **Fig. 7A-C** and **F**). Nevertheless, the true abnormal behavior of this mutant was a lower interaction energy between the flexible domain and the β-propeller of Co-DDB2^{K244E} and the lowest between the β-propeller and the BPA (**Fig.6F and I**). PCA for this mutation showed slight changes when compared to the WT. The BPA and BPC had less motion, probably due to the abnormal conformation of the Co-DDB2^{K244E} flexible domain, which elongates between the two β-propellers and exhibited a coil region approaching the BPC (**S-Video 7,** see **Supplementary Material**). In addition, Co-DDB2^{K244E} β-propeller is also affected, displaying a similar of motion as Co-DDB2^{L350P}. Finally, the BPB had a similar motion than Co-DDB1$^{DDB2(R273H)}$, where it twisted in its own axis, but in this mutation, the twisting motion was reduced. Taking these results, together with the fact that Co-DDB^{K244E} was the only variant to lose Co-DDB2^{K244E} communities in the DRIN analyzes (**S-Table 2,** see **Supplementary Material**) and, identical to Co-DDB^{L350P}, showed no interacting residues in common to the WT complex (**S-Fig.7,** see **Supplementary Material**), it becomes clear that this mutation affects the PPI connection. However, in contrast to Co-DDB^{L350P}, it does not appear to be a propagation of aberrant motion between DDB2 and DDB1, nor a change in the cavity of Co-DDB^{K244E} (**Fig.6K**). In the previously discussed mutation, there was a clear abnormal motion of DDB1, whereas here, DDB1 appears to be more stable (**Fig.7A-C and F**). In agreement, Co-DDB^{K244E} also showed transmission of conformational information correlation among residues for DDB2, but not as much of DDB1.

Hence, this mutation especially affects the BPB of DDB1, and does not affect DDB1, as seen in Co-DDB^{L350P}, nor does it affect Co-DDB2 in the same way seen in Co-DDB^{R273H}. In contrast to those mutations, Co-DDB^{K244E} behavior was a mix of the two mutations, indicating both instability and abnormal PPI connection.

4. Conclusions

Little is known of how the naturally occurring mutations found in XPE patients are related to the disease at protein structure level. Experimental data for all XP proteins are available, but they do not explain protein behavior or the conformational changes that each mutation can generate on the intrinsic PPI relationship that permeates the NER complex. This work is the first computational study to characterize DDB2 molecular behavior before and after its complexation to DDB1 and explore how three naturally occurring mutations influence protein conformation and complexation at the atomic-level.

In summary, in DDB2 mutant variants, the β-propeller changed its conformational state, showing loss of the flexibility seen in the WT. The inner cavity electrostatic potential of the β-propeller was clearly affected in DDB2^{L350P} and DDB2^{K244E}, a result that, when combined with the other observations, could indicate that the mutations incapacitate DDB2 DNA-damage recognition. In all mutants, a region comprising residues 354 to 371 featured loss of flexibility in comparison to the WT. In DDB2^{R273H} and DDB2^{L350P}, the loss of flexibility seen in the analysis might be because this region has strong connections to the flexible domain. In DDB2^{K244E} the absence of interaction energy with the 354-371 region implies change in structure behavior. All mutants displayed loss of transmission of conformational information among residues that influence the molecular behavior, possibly impairing PPI and domain motion. This hypothesis was confirmed by PCA analysis, which showed that, in the mutant variants, the flexible domain displayed aberrant behavior, completely losing the left-to-right movements seen in the WT, probably influencing its PPI with DDB1. Finally, the mutants do not seem to cause loss of secondary structure on isolated systems. In addition, we characterized and analyzed the DDB-Complex containing each mutant. The results indicated that each mutation affects the complex in a different way. The R273H mutation generates a DDB2 that is not stabilized by the

complex, where K244E disrupts all PPI residues seen in the WT. However, L350P was the most aggressive mutation, which not only affects DDB2, but also generates an aberrant behavior in all DDB1 β-propellers, including the BPB, which is not related to PPI with DDB2 but is related to the ubiquitin-ligase complex of the DDB-Complex. The employment of DRIN was beneficial to understand and corroborates the gathered data, giving additional information and supporting possible explanations for the abnormal behavior of both DDB2 alone and the DDB-Complex.

4. References

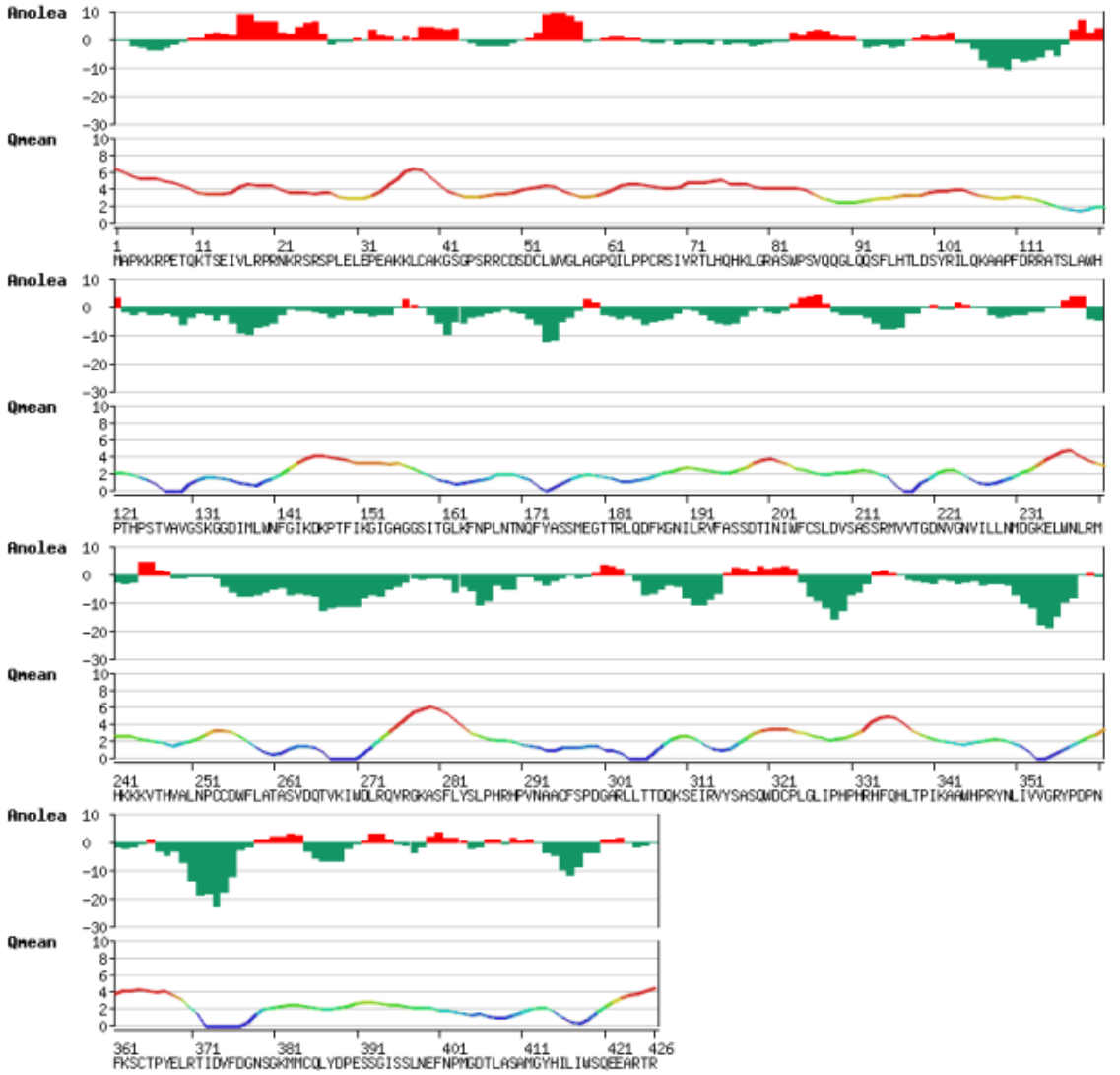
1. Gillet,L.C.J. and Schärer,O.D. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.*, **106**, 253–76.
2. Schärer,O.D. (2013) Nucleotide excision repair in eukaryotes. *Cold Spring Harb. Perspect. Biol.*, **5**, a012609.
3. Feltes,B.C. and Bonatto,D. (2015) Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications. *Mutat. Res. - Rev. Mutat. Res.*, **763**, 306–320.
4. Wittschieben,B.Ø., Iwai,S. and Wood,R.D. (2005) DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *J. Biol. Chem.*, **280**, 39982–9.
5. Marteijn,J. a, Lans,H., Vermeulen,W. and Hoeijmakers,J.H.J. (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.*, **15**, 465–81.
6. Rapic-Otrin,V. (2003) True XP group E patients have a defective UV-damaged DNA binding protein complex and mutations in DDB2 which reveal the functional domains of its p48 product. *Hum. Mol. Genet.*, **12**, 1507–1522.
7. Shiyanov,P., Hayes,S.A., Donepudi,M., Nichols,F., Linn,S., Slagle,B.L., Nichols,A.F. and Raychaudhuri,P. (1999) The Naturally Occurring Mutants of DDB Are Impaired in Stimulating Nuclear Import of the p125 Subunit and E2F1-Activated Transcription. *Mol Cell Biol*, **19**, 4935–4943.
8. Hwang,B.J., Toering,S., Francke,U. and Chu,G. (1998) p48 Activates a UV-damaged-DNA binding factor and is defective in xeroderma pigmentosum group E cells that lack binding activity. *Mol Cell Biol*, **18**, 4391–4399.
9. Scrima,A., Konícková,R., Czyzewski,B.K., Kawasaki,Y., Jeffrey,P.D., Groisman,R., Nakatani,Y., Iwai,S., Pavletich,N.P. and Thomä,N.H. (2008) Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell*, **135**, 1213–23.
10. Webb,B. and Sali,A. (2014) Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinforma.*, **47**, 5.6.1-5.6.32.

11. Kelley,L.A. and Sternberg,M.J.E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.
12. Kiefer,F., Arnold,K., Künzli,M., Bordoli,L. and Schwede,T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.*, **37**, 387–392.
13. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
14. Melo,F. and Feytmans,E. (1998) Assessing protein structures with a non-local atomic interaction energy. *J. Mol. Biol.*, **277**, 1141–1152.
15. Benkert,P., Schwede,T. and Tosatto,S.C. (2009) QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. *BMC Struct. Biol.*, **9**, 35.
16. Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
17. Pronk,S., Páll,S., Schulz,R., Larsson,P., Bjelkmar,P., Apostolov,R., Shirts,M.R., Smith,J.C., Kasson,P.M., Van Der Spoel,D., *et al.* (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.
18. Páll,S., Abraham,M.J., Kutzner,C., Hess,B. and Lindahl,E. (2015) Tackling exascale software challenges in molecular dynamics simulations with GROMACS.
19. Berendsen,H.J.C., Grigera,J.R. and Straatsma,T.P. (1987) The Missing Term in Effective Pair Potentialst. *J. Phys. Chem.*, **91**, 6269–6271.
20. Schmid,N., Eichenberger,A.P., Choutko,A., Riniker,S., Winger,M., Mark,A.E. and van Gunsteren,W.F. (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J*, **40**, 843–56.
21. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
22. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089.
23. Nose,S. and Klein,M.. (1983) Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, **50**, 1055–1076.

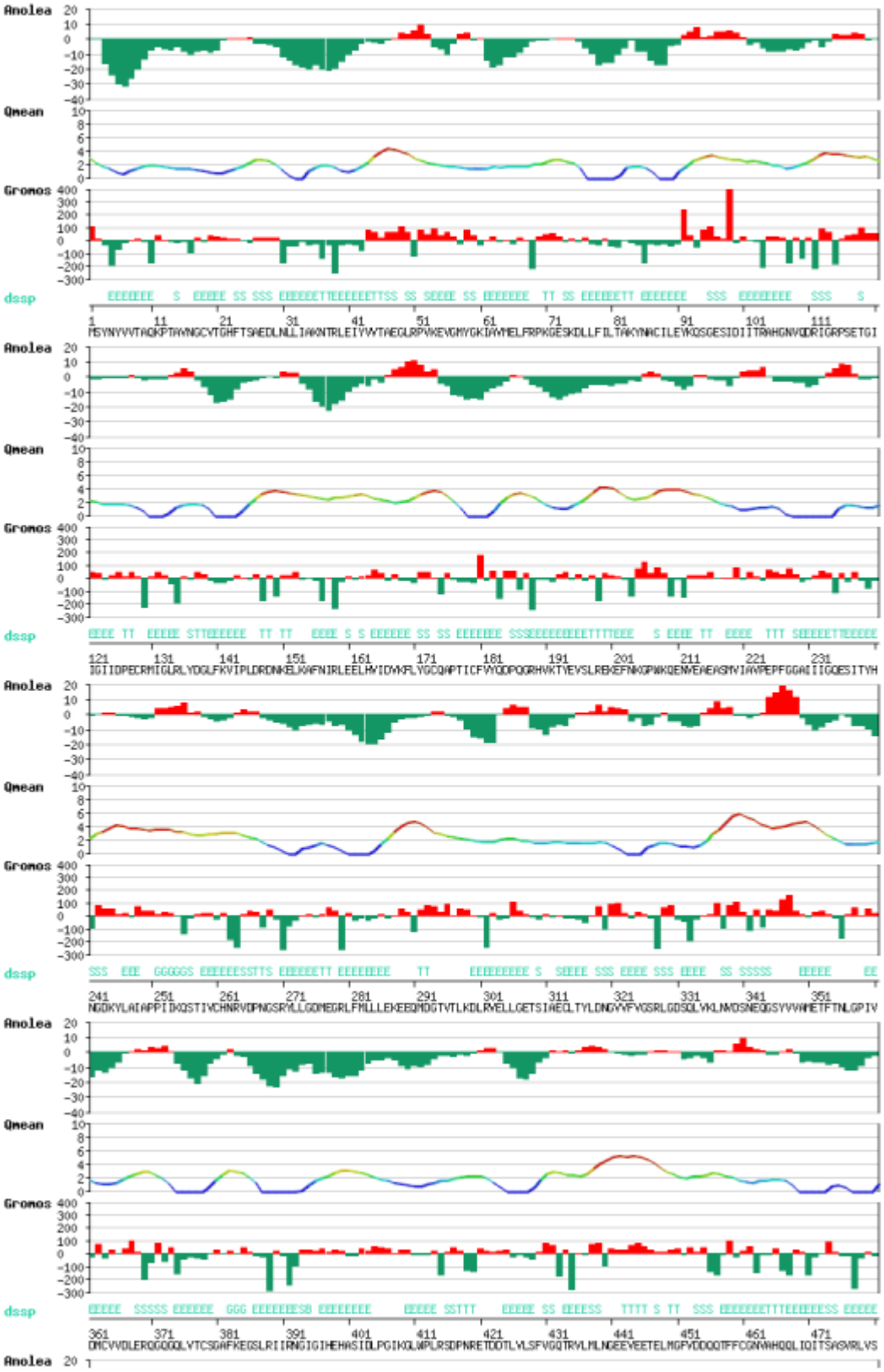
24. Parrinello,M. and Rahman,A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–90.
25. Bussi,G., Donadio,D. and Parrinello,M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**.
26. Eargle,J. and Luthey-Schulten,Z. (2012) NetworkView: 3D display and analysis of protein·RNA interaction networks. *Bioinformatics*, **28**, 3000–1.
27. Glykos,N.M. (2006) Carma: a molecular dynamics analysis program. *J Comput Chem*, **27**, 1765–1768.
28. Teilum,K., Olsen,J.G. and Kragelund,B.B. (2009) Functional aspects of protein flexibility. *Cell Mol Life Sci*, **66**, 2231–2247.
29. Pol-Fachin,L. and Verli,H. (2014) Structural glycobiology of heparin dynamics on the exosite 2 of coagulation cascade proteases: Implications for glycosaminoglycans antithrombotic activity. *Glycobiology*, **24**, 97–105.
30. Sethi,A., Eargle,J., Black,A.A. and Luthey-Schulten,Z. (2009) Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A*, **106**, 6620–5.
31. Chen,C.K.-M., Chan,N.-L. and Wang,A.H.-J. (2011) The many blades of the β-propeller proteins: conserved but versatile. *Trends Biochem. Sci.*, **36**, 553–61.
32. Xu,C. and Min,J. (2011) Structure and function of WD40 domain proteins. *Protein Cell*, **2**, 202–14.
33. Marcelino,A.M.C. and Gierasch,L.M. (2008) Roles of B-turns in protein folding: From peptide models to protein engineering. *Biopolymers*, **89**, 380–391.
34. Oldfield,C.J. and Dunker,K.A. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem*, **83**, 553–84.
35. Shell,S.M. and Zou,Y. (2008) Other Proteins Interacting with XP Proteins. *Adv Exp Med Biol*.
36. Scrima,A., Koníčková,R., Czyzewski,B.K., Kawasaki,Y., Jeffrey,P.D., Groisman,R., Nakatani,Y., Iwai,S., Pavletich,N.P. and Thomä,N.H. (2008) Structural Basis of UV DNA-Damage Recognition by the DDB1-DDB2 Complex. *Cell*, **135**, 1213–1223.
37. Fischer,E.S., Scrima,A., Böhm,K., Matsumoto,S., Lingaraju,G.M., Faty,M., Yasuda,T., Cavadini,S., Wakasugi,M., Hanaoka,F., *et al.* (2011) The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell*, **147**, 1024–39.

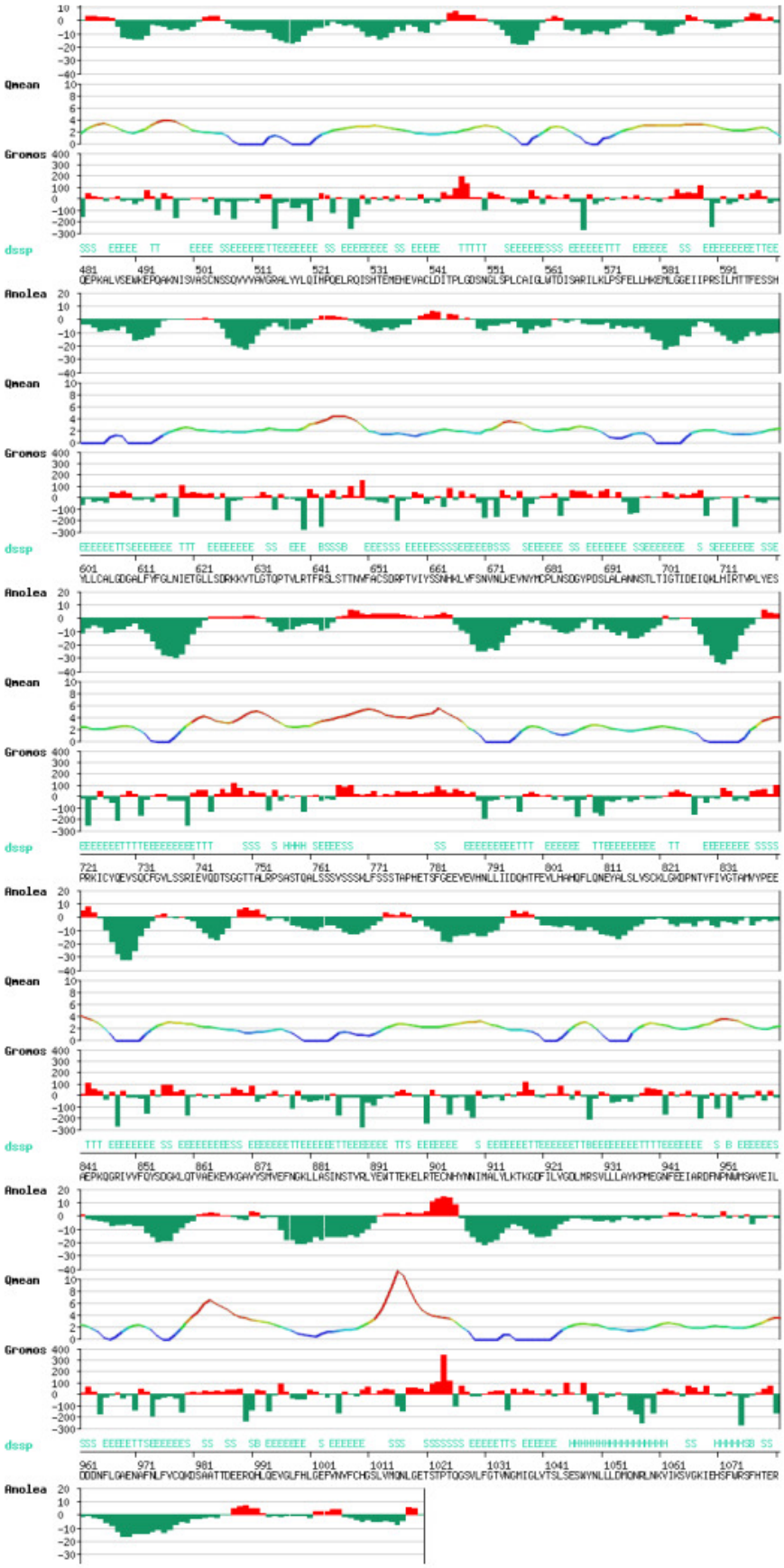
S-Table 1. Table summarizing the Ramachandran plot results for the final DDB1 structure, named $DDB1^{WT}$, after the Modeller reconstruction, and $DDB2^{WT}$, after PHYRE 2 structure prediction.

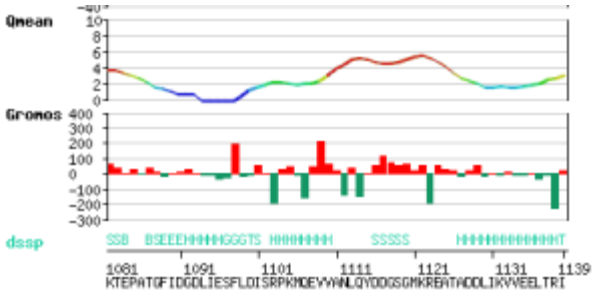
Protein	Res in most favored regions	Res in additional allowed regions	Res in generously allowed regions	Res in disallowed regions
DDB2	312 (85%)	39 (10.6%)	9 (2.5%)	7 (1.9%)
DDB1	889 (87.6%)	98 (9.7%)	18 (1.8%)	10 (1%)



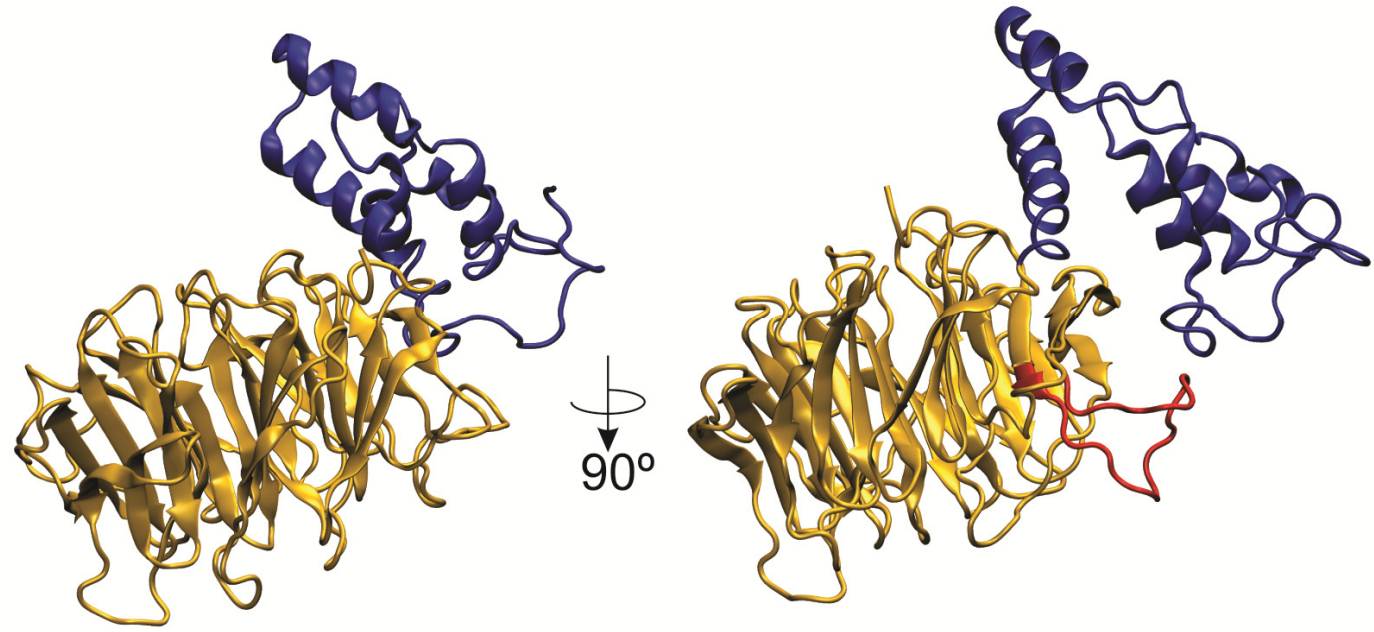
S-Fig.1. Results derived from the Anolea and Qmean analyzes performed in the SwissModel online tool [http://swissmodel.expasy.org] for DDB2^{WT}. Anolea revealed that the great majority of DDB2^{WT} model was in favored energy environments (green areas), with only the first segments from residues 11 to 60 and sparse regions in non-favorable energy environments (red areas). Qmean showed similar results for the overall quality of the model, combining local and per-residue quality estimation. Regions mostly composed of coils and bends were considered of lower quality, as expected.


Anolea
Qmean
Gromos
dssp


Qmean
Gromos
dssp
Anolea



S-Fig.2. Results derived from the Anolea and Qmean analyzes performed in the SwissModel online tool [http://swissmodel.expasy.org] for DDB1^{WT}. Anolea revealed that the great majority of DDB1^{WT} model was in favored energy environments (green areas), with sparse regions in non-favorable energy environments (red areas). Qmean showed similar results for the overall quality of the model, combining local and per-residue quality estimation.



S-Fig.3. Region 354-371 in the β-propeller of DDB2. Highlighted in red, this region showed a loss of flexibility in the RMSF of the mutant variants (**Fig.3D-E**).

S-Table 2. Communities and motion correlation between residues for DDB2^{WT} and its mutant variants, both simulated alone and complexed.

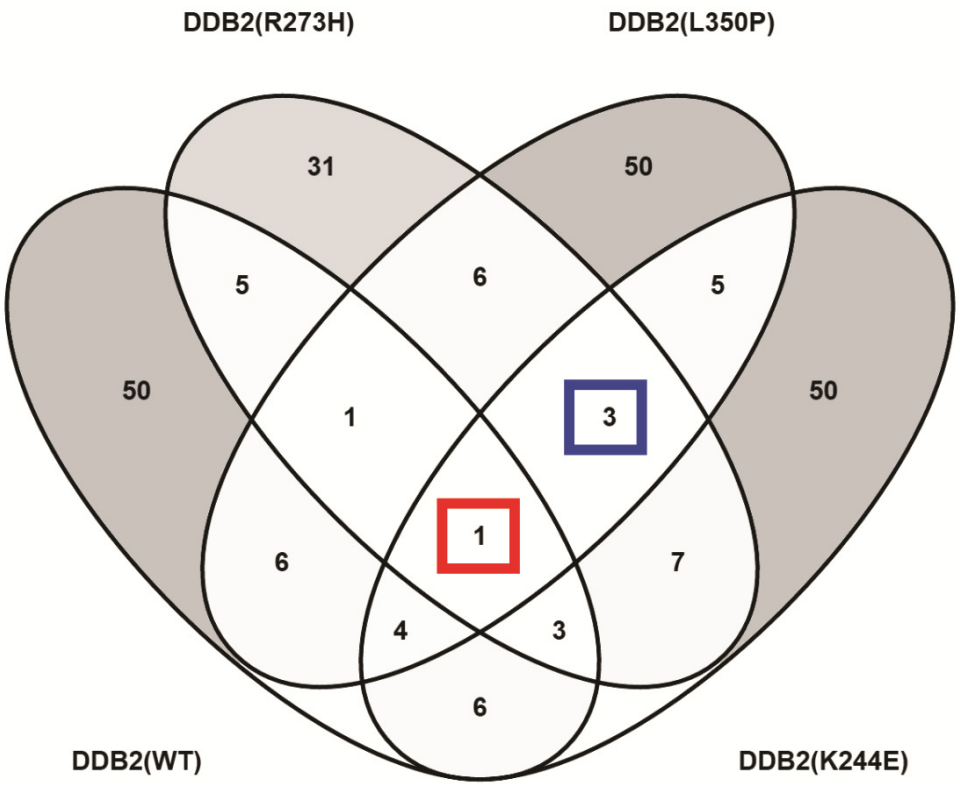
Protein	Mean nº of communities*	Motion correlation (vs WT)#
DDB2^{WT}	10	NA
DDB2^{R273H}	8	Lower
DDB2^{L350P}	10.6	Lower
DDB2^{K244E}	10	Lower
Co-DDB2^{WT}	6.3	NA
Co-DDB2^{R273H}	6.6	Lower
Co-DDB2^{L350P}	6	Lower
Co-DDB2^{K244E}	5	Lower
DDB1^{WT}	12.3	NA
Co-DDB1^{WT}	13.6	NA
Co-DDB1$^{DDB2(R273H)}$	12.6	Lower
Co-DDB1$^{DDB2(L350P)}$	11	Lower
Co-DDB1$^{DDB2(K244E)}$	12	NOD

* Mean value of the triplicates for each system. Communities composed of 1, 2 or 3 residues, either alone or complexed, were not considered. In the complexes, communities containing DDB1 and DDB2 were considered.

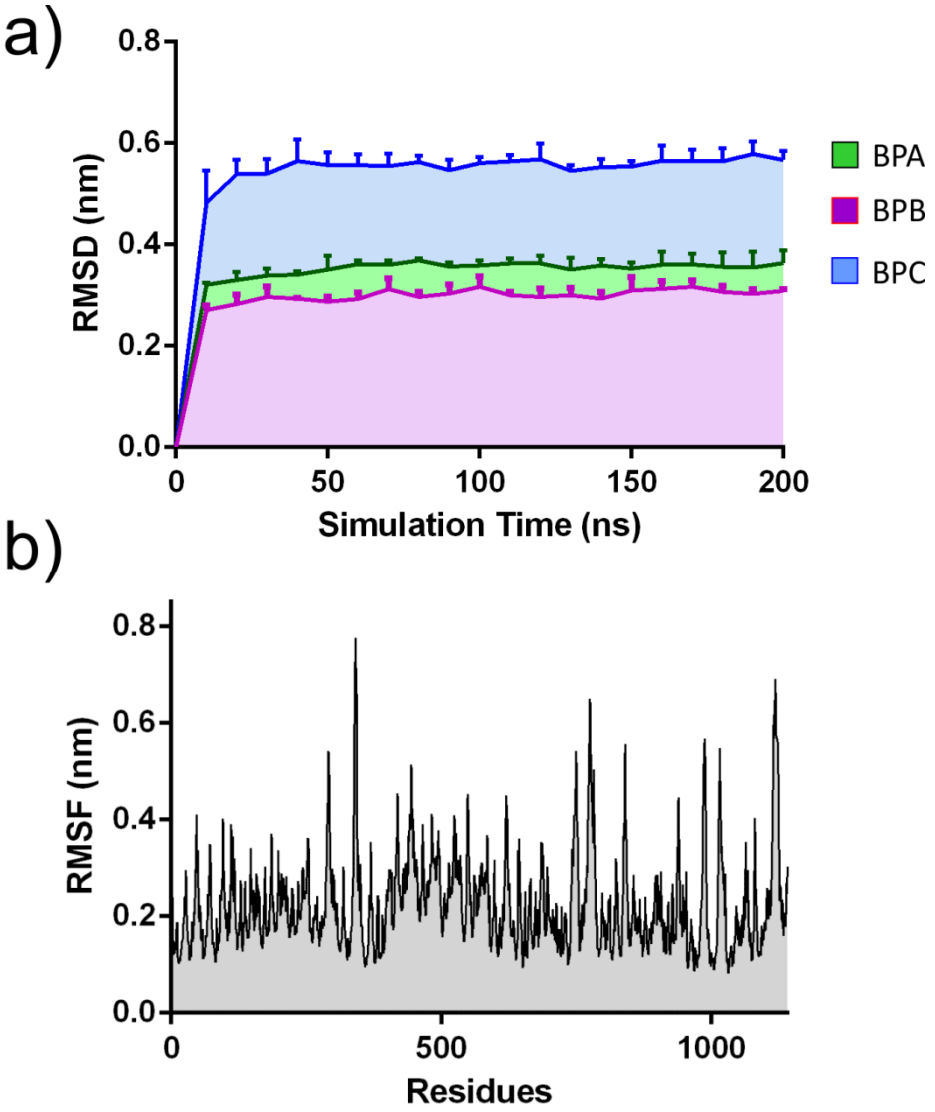
Global motion correlation among all residues of the protein.

NA = Not Applied

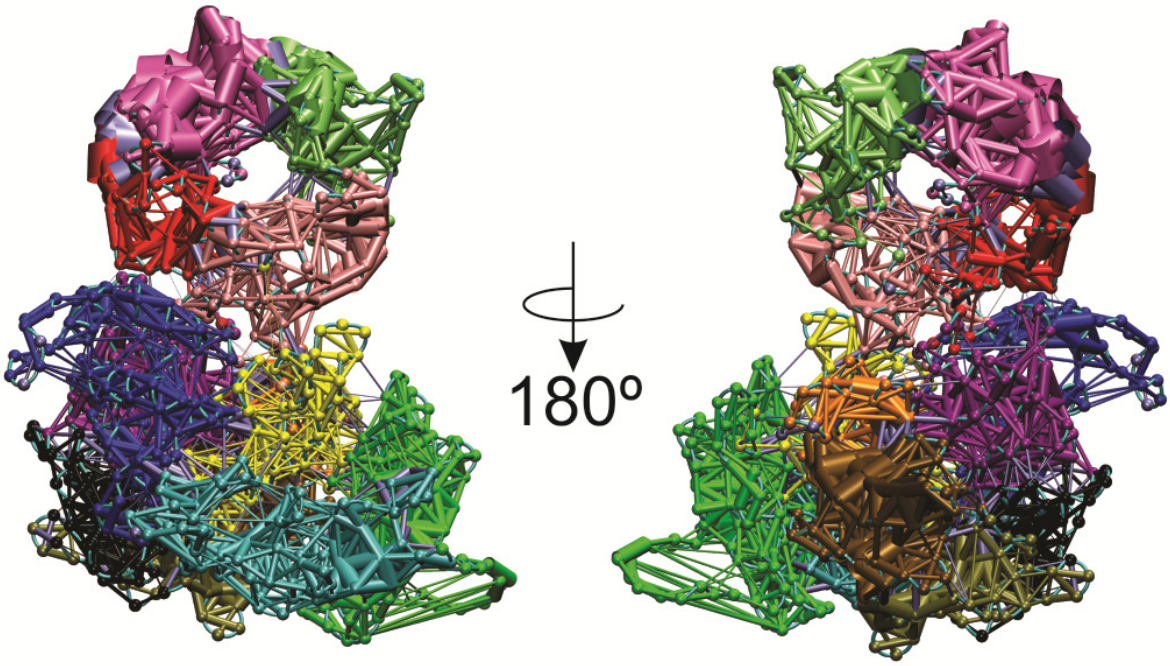
NOD = No Observable Difference



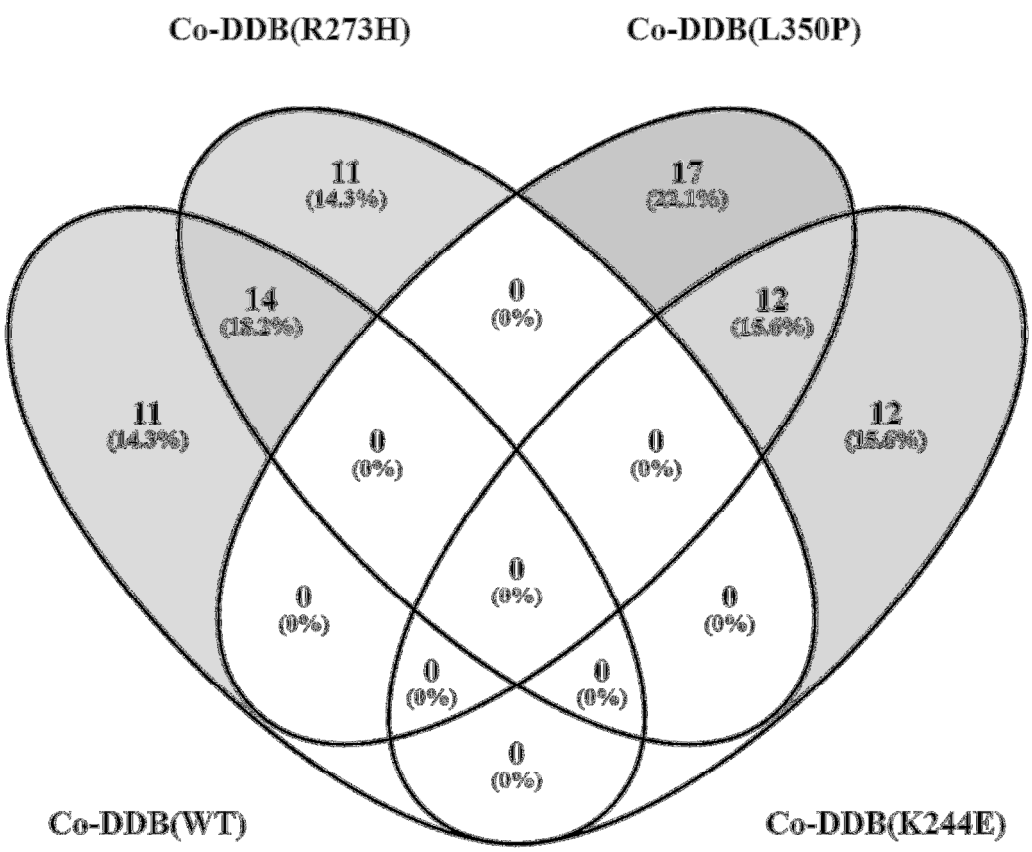
S-Fig.4. Venn diagram depicting the critical residues (residues with the highest motion correlation between communities) among all simulates systems for DDB2 and DDB2 mutant variants. Only one pair of residues was common for the fours systems: Val291-Thr306 (red square). In addition, three pairs were in common only among the mutants: Phe296-Trp344, Lys163-Asp208, and Trp344-Pro346 (blue square).



S-Fig.5. Mean RMSD for DDB1 β-propellers and RMSF for the whole protein. **A)** DDB1 β-propellers. The domains were colored according to **Fig.1** (see main text): (i) β-propeller A, in green; (ii) β-propeller B, in purple, and; (iii) β-propeller C, in blue. **B)** Mean RMSF for the whole protein.



S-Fig.6. The dynamic residues interaction networks (DRIN) depicting the communities and motion correlation between residues in $DDB1^{WT}$. Each color represents a given community.



S-Fig.7. Venn diagram depicting the interacting residues shared among each variant of the DDB-Complex. Only the residues in common among the triplicates of each system were considered. No pair of residues was shared between the four variants.

S-Table 3. Table listing all critical residues between DDB2 and DDB1 in common among the triplicates of each system.

Complex Variant	Critical residues*
Co-DDB^{WT}	699-742 (DDB1)
Co-DDB^{R273H}	31-33 (DDB2)
Co-DDB^{L350P}	355-388 (DDB2)
Co-DDB^{K244E}	NA

* No critical residues shared between the triplicates of each system were found between DDB1 and DDB2.

S-Video1. Upper view of the PCA analysis of DDB2 protein variants. In the upper left is $DDB2^{WT}$; $DDB2^{R273H}$ is seen in the upper right, whereas $DDB2^{K244E}$ and $DDB2^{L350P}$ can be observed in the bottom left and right, respectively. The flexible domain corresponds to the blue portion and the yellow region corresponds to the β-propeller. See the main text for explanations about the protein motion.

S-Video2. Back view of the PCA analysis of DDB2 protein variants. In the upper left is $DDB2^{WT}$; $DDB2^{R273H}$ is seen in the upper right, whereas $DDB2^{K244E}$ and $DDB2^{L350P}$ can be observed in the bottom left and right, respectively. The flexible domain corresponds to the blue portion and the yellow region corresponds to the β-propeller. See the main text for explanations about the protein motion.

S-Video3. PCA analysis of $DDB1^{WT}$. On the left is the front view of $DDB1^{WT}$ and on the right, the bottom view can be observed. The BPA, BPB, and BPC are colored in green, purple, and cyan, respectively. The gray portion corresponds to a α-helix/coil region. See the main text for explanations about the protein motion.

S-Video4. PCA analysis of $Co\text{-}DDB^{WT}$. Front, back, left and right views of the complex are depicted, respectively. The flexible domain of DDB2 corresponds to the blue portion and the yellow region corresponds to DDB2 β-propeller. The BPA, BPB, and BPC of DDB1 are colored in green, purple, and cyan, respectively. The gray portion corresponds to a α-helix/coil region in DDB1. See the main text for explanations about the protein motion.

S-Video5. PCA analysis of $Co\text{-}DDB^{R273H}$. Front, back, left and right views of the complex are depicted, respectively. The flexible domain of DDB2 corresponds to the blue portion and the yellow region corresponds to DDB2 β-propeller. The BPA, BPB, and BPC of DDB1 are colored in green, purple, and cyan, respectively. The gray portion corresponds to a α-helix/coil region in DDB1. See the main text for explanations about the protein motion.

S-Video6. PCA analysis of $Co\text{-}DDB^{L350P}$. Front, back, left and right views of the complex are depicted, respectively. The flexible domain of DDB2 corresponds to the blue portion and the yellow region corresponds to DDB2 β-propeller. The BPA, BPB, and BPC of DDB1 are colored in green, purple, and cyan, respectively. The gray portion corresponds to a α-helix/coil region in DDB1. See the main text for explanations about the protein motion.

S-Video7. PCA analysis of $Co\text{-}DDB^{K244E}$. Front, back, left and right views of the complex are depicted, respectively. The flexible domain of DDB2 corresponds to the blue portion and the yellow region corresponds to DDB2 β-propeller. The BPA, BPB, and BPC of

DDB1 are colored in green, purple, and cyan, respectively. The gray portion corresponds to a α-helix/coil region in DDB1. See the main text for explanations about the protein motion.

4. DISCUSSÃO GERAL

Múltiplos aspectos devem ser considerados no estudo de uma doença, como seu fenótipo, seus diferentes sintomas, quesitos epidemiológicos e os seus mecanismos moleculares subjacentes. Nesse sentido, o atual trabalho apresentou um estudo sobre o impacto de três mutações encontradas em pacientes com XP, tipo E, e como elas afetam o comportamento molecular da proteína DDB2 e do complexo DDB2-DDB1, além de caracterizar o comportamento molecular da proteína DDB1. Aplicando simulações por DM e RD, os dados apresentados no Capítulo II indicaram que as três mutações desestabilizam especialmente uma região específica da proteína DDB2, além de mudar o comportamento de seus dois domínios, quando comparado à sua forma selvagem. O estudo também mostrou que o comportamento do complexo constituído por DDB1 e DDB2 possui uma potencial instabilidade de ligação, quando formado pelas variantes mutantes de DDB2. Nesse sentido, a variante contendo a mutação L350P manifestou forte alteração no comportamento de DDB1, enquanto R273H e K244E ocasionaram enrijecimento de DDB2. Na análise de RD, tanto as proteínas estudadas separadamente, quanto complexadas, diminuíram a correlação de movimento entre seus aminoácidos e, em certos casos, perda de comunidades.

Contudo, outras observações podem ser feitas levando em consideração a estrutura da proteína DDB1 e DDB2, assim como sua relação com a doença XP.

4.1. Os β-*propellers* da família WD40: características e questões evolutivas

Tanto a proteína alvo do estudo apresentado, DDB2, quanto sua principal parceira no processo de NER, DDB1, são compostas por domínios denominados de β-*propellers*, onde dos 427 resíduos que compõe DDB2, 326 (76.3%) correspondem ao β-*propeller* e dos 1140 resíduos totais da proteína DDB1, 1031 (90.4%) são referentes aos três β-*propellers*, conectados por um agrupamento de α-hélices centrais (Feltes & Bonatto, 2015). Os dados apresentados demonstraram que os domínios sofrem alteração no comportamento molecular nas diferentes mutações estudadas, contudo uma discussão mais aprofundada sobre sua natureza e características evolutivas é potencialmente importante para o estudo de XP, tipo E.

Esses domínios são encontrados de procariotos a eucariotos, com estruturas altamente conservadas, caracterizados por uma estrutura simétrica, afunilada, em formato de disco, constituídos por lâminas de quatro fitas β antiparalelas (Chen, et al., 2011, Xu & Min, 2011) (**Fig. 10**). Os β-*propellers* são estruturas com uma alta estabilidade, dada a grande quantidade de interações intramoleculares, formando uma plataforma favorável para múltiplas IPP, ligações com diferentes ligantes (i.e. carboidratos, íons) e plasticidade necessária para funções diversificadas (Chen, et al., 2011). Existem diversas famílias correspondentes a esses domínios, como: (i) a família reguladora de condensação de cromatina (do inglês *regulator of chromatin condensation* 1 – RCC1), constituídas de sete lâminas com motivos conservados de resíduos de Asp e Gli; (ii) a família kelch, formada por domínios β-*propeller* de seis ou sete lâminas, com motivos repetidos de Tir, Trp e Arg; (iii) a família YWTD é principalmente composta de domínios com seis lâminas com repetições de Tir, Trp, Tre e Asp (YWTD) no final da segunda fita de cada lâmina; (iv) a outra família é a NIH, composta de β-*propellers* de seis lâminas que, igual à família YWTD, possuem, no final da segunda fita de cada lâmina, um motivo repetido, nesse caso de Tir, Val, Tre e Asp (YVTD); (v) por fim, a família mais vasta é a WD40, composta por seis, sete ou oito lâminas. Seu nome é resultado de um motivo repetido de Trp-Asp (WD), que existe em todos os membros do grupo e de uma repetição de 44-60 resíduos em uma única lâmina (Chen, et al., 2011, Xu & Min, 2011). Em sua maioria, os membros de todas as famílias de β-*propellers* não possuem atividade enzimática (Bergamin, et al., 2014).

Sendo assim, os β-*propellers* são, de certa forma, altamente estáveis e flexíveis. Isso foi visto nos resultados apresentados previamente no Capítulo II, especialmente com DDB1, que demonstrou alta flexibilidade, e tanto DDB1 quando DDB2 mostraram estabilidade em suas formas selvagens. Esse fato levanta uma questão intrigante: como uma única mutação de troca de aminoácidos pode afetar uma estrutura evolutivamente conservada e estável como o β-*propeller*?

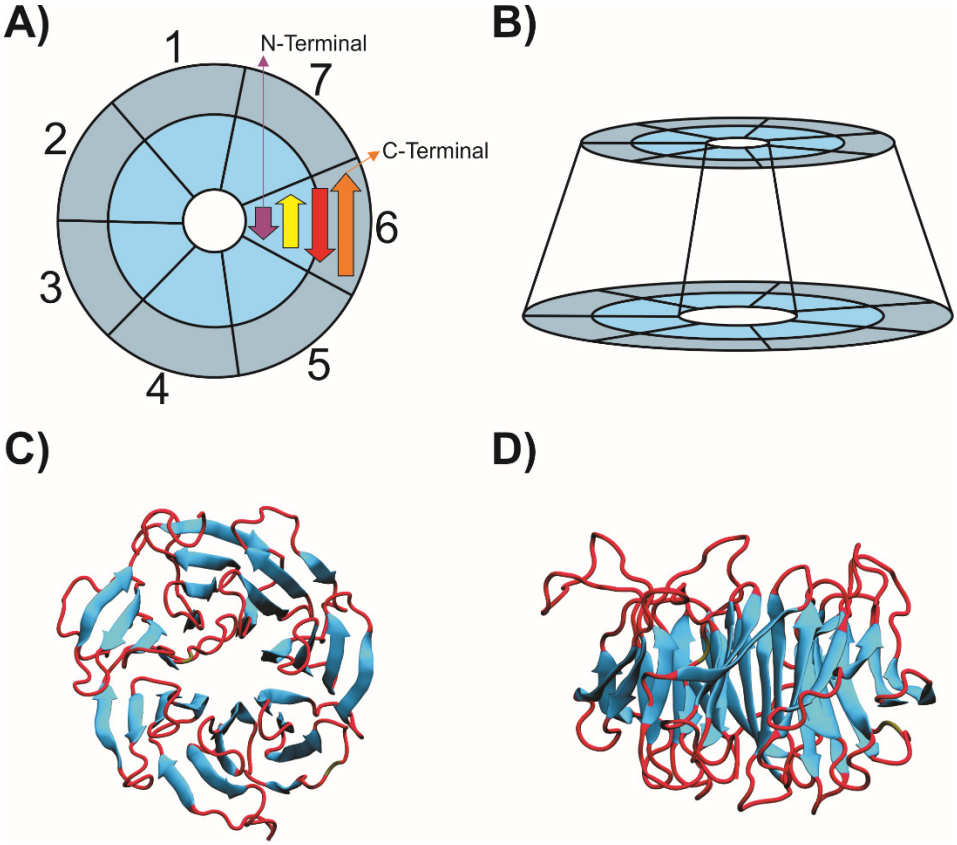


Fig.10. Características gerais de uma estrutura β-*propeller*. **A)** Visão de um *propeller* hipotético visto de cima. Cada lâmina da estrutura se encontra numerada de 1 a 7. As quatro fitas β que compõem cada lâmina são indicadas por diferentes cores, meramente ilustrativas, na lâmina de número 6. Cada lâmina possui um C-terminal na fita de maior tamanho da lâmina, enquanto o N-terminal localiza-se na menor fita β. **B)** Visão lateral do mesmo *propeller*, mostrando sua estrutura afunilada. **C)** Visão de cima da estrutura β-*propeller* (resíduos 102 a 427) da proteína DDB2 para exemplificar o que foi visto nos itens anteriores. Em azul encontram-se as lâminas compostas por fitas β e, em vermelho, os *coils* e *turns*. **D)** Visão lateral de DDB2, mostrando sua estrutura afunilada.

Uma possível explicação pode ser a própria conservação evolutiva do β-*propeller*. Ao contrário de proteínas com maior variabilidade estrutural, os domínios que possuem características evolutivamente conservadas podem ser mais sensíveis a mutações. A conservação evolutiva garante a preservação, ou similaridade, da função bioquímica da proteína ou parceiros de IPP, porém não assegura resiliência perante mutações. Outra explicação pode ser que as mutações encontradas em DDB2 situam-

se em regiões críticas para sua função. O estado nativo de uma proteína é considerado seu estado de maior equilíbrio dinâmico dentre todas suas possíveis conformações em um espaço conformacional (Saldaño, et al., 2016). Sendo assim, mutações em regiões críticas desse estado favorável de equilíbrio tendem a perturbar funções específicas, como IPP, especialmente em casos de conservação evolutiva (Saldaño, et al., 2016). Essa pode ser uma explicação plausível para os resultados observados anteriormente: mutações em DDB2 não só causaram um comportamento molecular aberrante na proteína, como também afetaram sua ligação com DDB1.

Por fim, a explicação pode não estar relacionada apenas com a estrutura da proteína, mas sim com a via bioquímica a qual ela afeta. A via de NER é um mecanismo de reparo de DNA evolutivamente conservado em mamíferos e essencial para a manutenção genômica (Gillet & Schärer, 2006). Sendo que cada passo da via de NER é necessário para o recrutamento da próxima etapa, uma perturbação em qualquer momento afetará o resto da via. DDB2 atua nos primeiros passos da via de NER e sua regulação, assim como a de XPC caracteriza a etapa mais crucial de todas: o reconhecimento do dano, tanto que pacientes com XP, tipo E e tipo C, possuem a maior incidência de câncer de pele, dentre todos os subtipos da doença (Fassihi, 2013).

4.2. Possíveis explicações para a falta de dados estruturais das proteínas XP: regiões intrinsicamente desordenadas

Outro ponto crucial na discussão do estudo estrutural das proteínas XP é a quantidade limitada de informações sobre suas estruturas, e esse fato não só justifica vários pontos dos dados discutidos, como também a urgência de mais informações sobre esse assunto.

Como visto no Capítulo I, as informações sobre a estrutura tridimensional das proteínas XP são escassas: apenas DDB2 possui mais de 80% da sua estrutura disponível em humanos. A estrutura de XPA é a segunda mais completa, com aproximadamente 45% da sua estrutura obtida; todas as outras possuem pequenos fragmentos elucidados ou nenhuma informação em humanos, como é o caso de XPG (Feltes & Bonatto, 2015). Indaga-se: o que justificaria a falta de estruturas tridimensionais disponíveis, se, como apresenta o Capítulo I, informações estruturais estão disponíveis?

O principal motivo pode ser a natureza das estruturais tridimensionais das outras proteínas XP, que, provavelmente, devem possuir regiões de alta desordem, consequentemente dificultando sua cristalização. Como discutido anteriormente, DDB2 é composta por um β-*propeller*, que se caracteriza por uma estrutura altamente estável e conservada. Graças a esse fato, a cristalização de sua estrutura tridimensional é facilitada. Contudo, o mesmo não ocorre em regiões de desordem cuja falta de formação de estruturas secundárias sem um substrato, permanecem sem uma estrutura definida e podem se dobrar de forma diferente de acordo com o substrato (Babu, 2016).

Através de uma análise pelo programa PSIPRED [http://bioinf.cs.ucl.ac.uk/psipred/] (McGuffin, et al., 2000; Buchan, et al., 2013), é possível observar que este é o caso das proteínas XP, onde as regiões que não correspondem a suas porções cristalizadas possuem alto grau de desordem (**Fig. 11**).

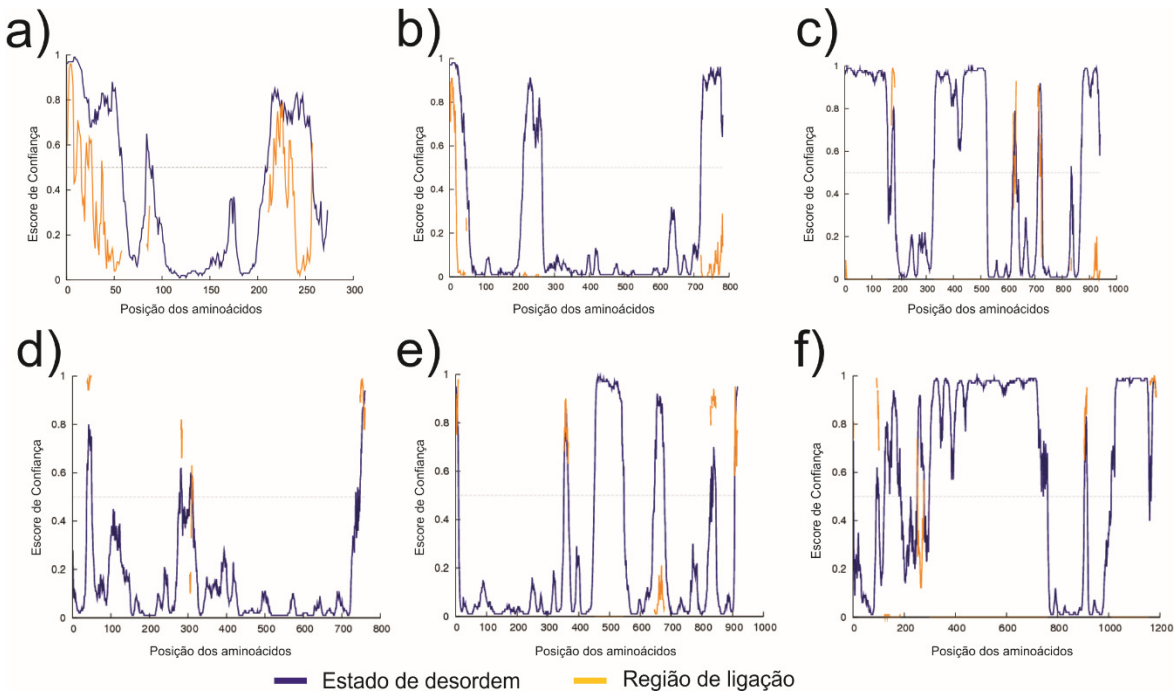


Fig. 11. Gráfico de desordem pelo programa PSIPRED. No eixo *y*, se encontra o escore de confiança da análise, enquanto no eixo *x* é determinada a posição dos aminoácidos. **A)** Gráfico de regiões de desordem e IPP de XPA. O fragmento cristalizado de XPA corresponde ao domínio central da proteína, que compreende os resíduos 98 a 225 (PDB: 1XPA). É possível notar que essa região é a com menor grau de desordem, enquanto o restante da proteína (mais que 50% do seu tamanho) encontra-

se em um alto estado desordem. **B)** Gráfico de regiões de desordem e IPP de XPB. O fragmento cristalizado de XPB (PDB: 4ERN) abrange os resíduos 502 a 730 (36.9% do seu tamanho). É possível observar no gráfico que essa região corresponde a uma porção com baixa desordem. **C)** Gráfico de regiões de desordem e IPP de XPC. XPC é a proteína XP com segundo maior grau de desordem. **D)** Gráfico de regiões de desordem e IPP de XPD. **E)** Gráfico de regiões de desordem e IPP de XPF. XPF também se encontra entre as proteínas XP com maior grau de desordem. **F)** Gráfico de regiões de desordem e IPP de XPG. XPG é a proteína com maior grau de desordem observada na análise.

Como discutido no Capítulo I, XPA possui várias IPP, além de se ligar ao DNA (Feltes & Bonatto, 2015). Sendo assim, é necessário que sua estrutura seja flexível o suficiente para permanecer ligada à fita simples de DNA, proteínas da via de NER e outros ligantes como CEP-164 (Shell & Zou, 2008). No gráfico, é possível observar que as regiões de desordem estão altamente relacionadas com as regiões de IPP. Sendo XPA de natureza flexível, e devido à quantidade significativa de IPP que ela possui, existe uma alta probabilidade de que essas porções não possuam estrutura definida e se dobrem de formas diferentes dependendo do ligante. O mesmo é observado para XPC. XPC é responsável por reconhecer o dano de DNA e recrutar o complexo XPB e XPD, tornando-se uma proteína que exige uma estrutura flexível (Feltes & Bonatto, 2015). A análise de desordem também aponta para um ponto interessante: as únicas proteínas com menor quantidade de regiões desordenadas são as helicases XPB e XPD, as helicases do multiproteico TFIIH (Schultz, et al., 2000). É interessante observar que, para ambas as proteínas, as regiões de menor desordem não são referentes a regiões de IPP, mas sim ao domínio de helicases II (Feltes & Bonatto, 2015). O mesmo é visto para XPF, cujas regiões de menor desordem são referentes aos domínios helicase-*like* e de nuclease (Feltes & Bonatto, 2015). Em resumo, a análise aponta que, sem ser em regiões evolutivamente conservadas dos domínios funcionais das proteínas XP (Bienstock, et al., 2003, Fan et al., 2006, Lehmann, 2008, Nishino, et al., 2003, Roth, et al., 2012), o restante de suas estruturas não são definidas e de alta flexibilidade, em parte explicando a falta de dados sobre as estruturas tridimensionais das proteínas XP.

5.CONCLUSÕES

- A proteína DDB1, quando simulada sozinha, mostrou um movimento de pinça entre o BPA e BPC que poderia ser sua forma de se complexar à DDB2.
- Os dados indicam que todas as mutações afetam uma região específica, entre os resíduos 354 até 371, e que esta região parece estar envolvida no estado conformacional de DDB2.
- As redes dinâmicas permitiram observar que todos os mutantes perdem correlação de movimento entre aminoácidos, indicando uma perda de flexibilidade da proteína, assim como de domínios independentes.
- Quando complexada, todos os mutantes geram uma maior interação entre os domínios de DDB2, enrijecendo-a.
- Em todos os mutantes complexados houve um aumento da cavidade de DDB2.
- Em todos os mutantes, a região 354-371 mostrou menor interação com o domínio flexível.
- Quando complexada, a mutação R273H gera uma abertura do *β-propeller* e uma maior flexibilidade na região 354-371, em DDB2, afetando seu comportamento molecular e indicando uma possível desestabilização.
- A mutação L350P, em uma DDB2 complexada, alterou completamente a interface de IPP entre DDB2 e DDB1, não possuindo nenhum par de conexões em comum com a forma selvagem, além de ocasionar uma interação mais fraca entre o domínio flexível de DDB2 e o BPA de DDB1.
- A mutação L350P foi a única que claramente afetou o comportamento de DDB1 diretamente, mostrando um fenômeno de comportamento molecular aberrante em todo o complexo.
- A mutação K244E não possui nenhum par de interações em comum com a forma selvagem, e menor interação entre o *β-propeller* e o BPAe BPC de DDB1, indicando um comportamento que poderia impactar no reconhecimento ao DNA, assim como a interação entre as duas proteínas.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AHN, Y.Y., AHNERT, S.E., BAGROW, J.P. et al. Flavor network and the principles of food pairing. Sci Rep, 1, 196, 2011.

AMITAI, G., SHEMESH, A., SITBON, E., et al. Network analysis of protein structures identifies functional residues. J Mol Biol, 344(4), 1135-1146, 2004.

BABU, M.M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem Soc Trans*, 44(5), 1185-1200, 2016.

BARABÁSI, A.L. & OLTVAI, Z.N. Network biology: understanding the cell's functional organization. Nat Rev Genet, 5(2), 101-113, 2004.

BERGAMIN, E., BLAIS, A., COUTURE, J.F. Keeping them all together: β-propeller domains in histone methyltransferase complexes. *J Mol Biol*, 426(20), 3363-3375, 2014.

BESSHO, T. Nucleotide excision repair 3' endonuclease XPG stimulates the activity of base excision repair enzyme thymine glycol DNA glycosylase. *Nucleic Acids Res*, 27, 979–983, 1999

BIENSTOCK, R.J., SKORVAGA, M., MANDAVILLI, B.S., et al. Structural and functional characterization of the human DNA repair helicase XPD by comparative molecular modeling and site-directed mutagenesis of the bacterial repair protein UvrB. *J Biol Chem*, 278(7), 5309-5316, 2003.

BOHR, V.A. Gene-specific DNA repair. *Carcinogenesis*, 12(11), 1983-1992, 1991.

BUCHAN, D.W.A, MINNECI, F., NUGENT, T.C.O., et al. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res*, 41 (W1), W340-W348.

BUSCH, D.B., CLEAVER, J.E. & GLASER D.A. Large-scale isolation of UV-sensitive clones of CHO cells. *Somatic Cell Genet*, 6(3), 407-418, 1980.

CHEN, C.K., CHAN, N.L., WANG, A.H. The many blades of the β-propeller proteins - conserved but versatile. *Trends Biochem Sci*, 36(10), 553-561, 2011.

CLEAVER, J.E. Defective repair replication of DNA in xeroderma pigmentosum. *Nature*, 218, 652-656, 1963.

CLEAVER, J.E. Xeroderma pigmentosum: a human disease in which an initial stage of DNA repair is defective. *Proc Natl Acad Sci U S A*, 63(2), 428-435, 1969.

CLEAVER, J.E. Chapter I: Historical Aspects of Xeroderma Pigmentosum and Nucleotide Excision Repair. Em: *Molecular Mechanisms of Xeroderma Pigmentosum*. Ed. por Shamin I. Ahmad & Fumio Hanaoka. Advances in Experimental Medicine and Biology, Landes Biosciences & Springer Science, 637, 1-9, 2008.

CLEAVER, J.E. & BOOTSMA, D. Xeroderma pigmentosum: biochemical and genetic characteristics. *Annu Rev Genet*, 9, 19-38. 1975.

COIN, F., PROIETTI DE SANTIS, L., NARDO, T. et al. p8/TTD-A as a repair-specific TFIIH subunit. *Mol Cell*, 21, 215–226, 2006.

COLTON, S.L., XU, X.S., WANG, Y.A. et al. The involvement of ataxia-telangiectasia mutated protein activation in nucleotide excision repair-facilitated cell survival with cisplatin treatment. *J Biol Chem*, 281, 27117–27125, 2006.

DATTA, A., BAGCHI, S., NAG, A., et al. The p48 subunit of the damaged-DNA binding protein DDB associates with the CBP/p300 family of histone acetyltransferase. *Mut Res*, 486, 89–97, 2001.

DE WEERD-KASTELEIN, E.A., KEIJZER, W. & BOOTSMA, D. Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. *Nat New Biol*, 238(81), 80-83, 1972.

DONCHEVA, N.T., KLEIN, K., DOMINGUES, F.S. et al. Analyzing and visualizing residue networks of protein structures. Trends Biochem Sci, 36(4), 179-182, 2011.

DONCHEVA, N.T., ASSENOV, Y., DOMINGUES, F.S. et al. Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protoc*, 7(4), 670-685, 2012.

FAN, L., ARVAI, A.S., COOPER, P.K., et al. Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol Cell*, 22(1), 27-37, 2006.

FASSIHI, H. Spotlight on 'xeroderma pigmentosum'. *Photochem Photobiol Sci*, 12(1), 78-84, 2013.

FELTES, B.C., POLONI, J.F., MIYAMOTO, K.N. et al. Human diseases associated with genome instability. Em: *Genome Instability: From Virus to Human Application*. Ed. por Igor Kovalchuk & Olga Kovalchuk. Translational Epigenetics Series, Elsevier, 447-462, 2016. **(Incluído no Adendo 1 desta tese).**

FISHER, G.J. & JOHNS, H.E. Chapter 5 – Pyrimidine Photodimers. Em: *Photochemistry and Photobiology of Nucleic Acids*. Ed. por Shih Yi Wang. Chemistry. Vol 1, Elsevier, 225-294, 1976.

FRANKLIN, W.A., LO, K.M. & HASELTINE, W.A. Alkaline lability of fluorescent photoproducts produced in ultraviolet light-irradiated DNA. *J Biol Chem*, 257(22), 13535-13543, 1982.

FRIEDBERG, E.C. & MEIRA, L.B. Database of mouse strains carrying targeted mutations in genes affecting cellular responses to DNA damage. *Mutat Res*, 459(4), 243-274, 2000.

GARY, R., LUDWIG, D.L., CORNELIUS, H.L., et al. The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21. *J Biol Chem*, 272, 24522–24529, 1997.

GIANNELLI, F. & PAWSEY, S.A. DNA Repair Synthesis in Human Heterokaryons. II. A test for heterozygosity in xeroderma pigmentosum and some insight into the structure of the defective enzyme. *J Cell Sci*, 15(1), 163-176, 1974.

GILLET, L.C. & SCHÄRER, O.D. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev*, 106(2), 253-276, 2006.

GOH, C.S., MILBURN, D., & GERSTEIN, M. Conformational changes associated with protein-protein interactions. *Curr Opin Struct Biol*, 14(1), 104 – 109, 2004.

GOH, K.I., CUSICK, M.E, VALLE, D. et al. The human disease network. *Proc Natl Acad Sci U S A*, 104(21), 8685-8690, 2007.

GREENE, L.H. & HIGMAN, V.A. Uncovering network systems within protein structures. J Mol Biol, 334(4), 781-791, 2003.

GUNSTEREN, W.F., BAKOWIES, D., BARON, R. et al. Biomolecular Modeling: Goals, Problems, Perspectives. *Angew Chem Int Ed Engl*. 45(25), 4064-4092, 2006.

GUSTAFSSON, M., NESTOR, C.E., ZHANG, H. et al. Modules, networks and systems medicine for understanding disease and aiding diagnosis. *Genome Med*, 6(10), 82, 2014.

GUVENCH, O. & MACKERELL, A.D. Comparison of Protein Force Fields for Molecular Dynamics Simulations. Em: *Molecular Modeling of Proteins*. Ed. por Andreas

Kukol. Methods Molecular Biology 2 ed., Springer New York Heidelberg Dordrecht London, 443, 63-88, 2008.

HALL, H., GURSKY, J., NICODEMOU, A. et al. Characterization of ERCC3 mutations in the Chinese hamster ovary 27–1, UV24 and MMC-2 cell lines. Mut Res, 593, 177–186, 2006.

HAYES, S., SHIYANOV, P., CHEN, X. et al. DDB, a putative DNA repair protein, can function as a transcriptional partner of E2F1. *Mol Cell Biol*, 18, 240–249, 1998.

HE, Y.J., MCCALL, C.M., HU, J. et al. DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4–ROC1 ubiquitin ligases. Genes Dev, 20(21), 2949-2954, 2006.

HENGGE, U.R. & EMMERT, S. Chapter II: Clinical Features of Xeroderma Pigmentosum. Em: *Molecular Mechanisms of Xeroderma Pigmentosum*. Ed. por Shamin I. Ahmad & Fumio Hanaoka. Advances in Experimental Medicine and Biology, Landes Biosciences & Springer Science, 637, 10-18, 2008.

IOVINE, B., IANNELLA, M.L. & BEVILACQUA, M.A. Damage-specific DNA binding protein 1 (DDB1): a protein with a wide range of functions. Int J Biochem Cell Biol, 43(12), 1664-1667, 2011.

KARPLUS, M. & MCCAMMON, J.A. Dynamics of proteins: elements and function. *Annu Rev Biochem*, 52, 263-300, 1983.

KULAKSIZ, G., REARDON, J.T. & SANCAR, A. Xeroderma pigmentosum complementation group E protein (XPE/DDB2): purification of various complexes of XPE and analyses of their damaged DNA binding and putative DNA repair properties. *Mol Cell Biol*, 25, 9784–9792, 2005.

LEACH, A.R. Chapter 7 – Molecular Dynamics Simulation Methods. Em: Molecular Modelling: Principles and Applications. Ed. por Andrew R. Leach. Pearson, vol. 2, 351-409, 2001a.

LEACH, A.R. Chapter 4 – Empirical Force Field Models: Molecular Mechanics. Em: Molecular Modelling: Principles and Applications. Ed. por Andrew R. Leach. Pearson, vol. 2, 165-252, 2001b.

LEHMANN, A.R. XPD structure reveals its secrets. DNA Repair (Amst), 7(11), 1912-1915, 2008.

LEHMANN, A.R., MCGIBBON, D. & STEFANINI, M. Xeroderma Pigmentosum. *Orphanet J Rare Dis*, 6, 70, 2011.

LEVEILLARD, T., ANDERA, L., BISSONNETTE, N. et al. Functional interactions between p53 and the TFIIH complex are affected by tumour-associated mutations. *EMBO J*,15, 1615–1624, 1996.

LIU, J., MENG, X. & SHEN, Z. Association of human RAD52 protein with transcription factors. Biochemical and Biophysical Research Communications, 297,1191–1196, 2002.

MANSBRIDGE, J.N. & HANAWALT, P.C. Cellular responses to DNA damage. Em: *UCLA Symposium on molecular and cellular biology*. Ed. por Errol C. Friedberg e B.A. Bridges. New series, A.R. Liss, 195-207, 1983.

MARTEIJN, J.A., LAND, H., VERMEULEN, W. et al. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol*, 15(7), 465-81, 2014.

MCGUFFIN, L.J., BRYSON, K., JONES, D.T. The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404-405, 2000.

MENCK, C.F.M. & MUNFORD, V. DNA repair diseases - What do they tell us about cancer and aging. *Genet Mol Biol* 37(1 Suppl), 220-233, 2014.

MONTICELLI, L. & TIELEMAN, D.P. Force fields for classical molecular dynamics. Em: Biomolecular Simulations. Ed. por Luca Monticelli & Emppu Salonen. *Methods in Molecular Biology*, Springer New York, 924, 197-213, 2013.

MOTYCKA, T.A., BESSHO, T., POST, S.M. et al. Physical and functional interaction between the XPF/ ERCC1 endonuclease and hRad52. *J Biol Chem*, 279, 13634–13639, 2004.

NAKATSU, Y., ASAHINA, H., CITTERIO, E. et al. XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. *J Biol Chem*, 275, 34931–34937, 2000.

NEWMAN, M.E.J. The Structure and Function of Complex Networks. *SIAM Review*, 45, 167-256, 2003.

NEWMAN, M.E.J. Modularity and community structure in networks. *Proc Natl Acad Sci U S A*, 103(23), 8577-8582, 2006.

NISHINO, T., KOMORI, K., ISHINO, Y., et al. X-ray and biochemical anatomy of an archaeal XPF-Rad1-Mus81 family nuclease - similarity between its endonuclease domain and restriction enzymes. Structure, 11(4), 445-457, 2003.

NITTA, M., SAIJO, M., KODO, N. et al. A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein XPA. *Nucleic Acids Res*, 28, 4212–4218, 2000.

PALLA, G., BARABÁSI, A.L. & VICSEK, T. Quantifying social group evolution. *Nature*, 446(7136), 664-667, 2007.

ROTH, H.M., RÖMER, J., GRUNDLER, V., et al. XPB helicase regulates DNA incision by the Thermoplasma acidophilum endonuclease Bax1. *DNA Repair (Amst)*, 11(3), 286-293, 2012.

SALDAÑO, T.E., MONZON, A.M., PARISI, G. et al. Evolutionary Conserved Positions Define Protein Conformational Diversity. *PLoS Comput Biol*, 12(3), e1004775, 2016.

SANDROCK, B. & EGLY, J.M. A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH. *J Biol Chem*, 276, 35328–35333, 2001.

SCHÄRER, O.D. Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol*, 5(10), a012609, 2013.

SCHLICK, T. Biomolecular Structure and Modeling: Historical Perspective. Em: Molecular Modeling and Simulation: An Interdisciplinary Guide. Ed. por Tamar Schlick. Springer New York Heidelberg Dordrecht London, vol 2, 1-40, 2010

SCHULTZ, P., FRIBOURG, S., POTERSZMAN, A., et al. Molecular structure of human TFIIH. *Cell*, 102(5), 599-607, 2000.

SETHI, A., EARGLE, J., BLACK, A.A. et al. Dynamical networks in tRNA: protein complexes. *Proc Natl Acad Sci U S A*, 106(16), 6620-6625, 2009.

SETLOW, R.B., REGAN, J.D., GERMAN, J. et al. Evidence that xeroderma pigmentosum cells do not perform the first step in the repair of ultraviolet damage to their DNA. *Proc Natl Acad Sci U S A*, 64(3), 1035-1041, 1969.

SHELL, S.M. & ZOU, Y. Other proteins interacting with XP proteins. *Adv Exp Med Biol*, 637, 103-112, 2008.

SHIMIZU, Y., IWAI, S., HANAOKA, F. et al. Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase. *EMBO J*, 22, 164173, 2003.

SNEHA, P. & DOSS, C.G.P. Molecular Dynamics: New Frontier in Personalized Medicine. Em: *Advances in Protein Chemistry and Structural Biology*. Ed. por Rossen Donev. Elsevier, 102, 181-224, 2016.

del SOL, A., FUJIHASHI, H., AMOROS, D. et al. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol Syst Biol*, 2, 2006.0019, 2006.

del SOL, A. & O'MEARA, P. Small-World Network Approach to Identify Key Residues in Protein–Protein Interaction. *Proteins*, 58(3), 672-682, 2005.

SRIDHARAN, D., BROWN, M., LAMBERT, W.C. et al. Nonerythroid alphaII spectrin is required for recruitment of FANCA and XPF to nuclear foci induced by DNA interstrand cross-links. *J Cell Sci*, 116, 823–835, 2003.

TAVERNA, D.M. & GOLDSTEIN, R.A. Why are proteins so robust to site mutations? J Mol Biol, 315(3), 479-784, 2002.

THOMPSON, L.H., RUBIN, J.S., CLEAVER, J.E. et al. A screening method for isolating DNA repair-deficient mutants of CHO cells. *Somatic Cell Genet*, 6(3), 391-405, 1980.

VENDRUSCOLO, M., DOKHOLYAN, N.V., PACI, E. et al. Small-world view of the amino acids that play a key role in protein folding. *Phys Rev E Stat Nonlin Soft Matter Phys*, 65(6 Pt 1), 061910, 2002.

VERLI, H. Capítulo 8: Dinâmica Molecular. Em: *Bioinformática: da Biologia a Flexibilidade Molecular*. Ed. por Hugo Verli. Sociedade Brasileira de Bioqímica, 173-187, 2014.

VLACHAKIS, D., BENCUROVA, E., PAPANGELOPOULOS, N. et al. Current State-of-the-Art Molecular Dynamics Methods and Applications. Em: *Advances in Protein Chemistry and Structural Biology*. Ed. por Rossen Donev. Elsevier, 269-313, 2014.

WAGNER, G.P., PAVLICEV, M. & CHEVERUD, G.M. The road to modularity. *Nat Rev Genet.* 8(12), 921-931, 2007.

WITTSCHIEBEN, B.O. & WOOD, R.D. DDB complexities. DNA Repair (Amst), 2(9):1065-1069, 2003.

WOOD, R.D., ROBINS, P. & LINDAHL, T. Complementation of the xeroderma pigmentosum DNA repair defect in cell-free extracts. *Cell*, 53(1), 97-106, 1988.

WU, X., SHELL, S.M., LIU, Y. et al. ATR-dependent checkpoint modulates XPA nuclear import in response to UV irradiation. *Oncogene*, 26(5), 757-764, 2007.

WUCHTY, S., BARABÁSI, A.L. & FERDIG, M.T. Stable evolutionary signal in a yeast protein interaction network. *BMC Evol Biol*, 6, 8, 2006.

YILDIRIM, M.A., GOH, K.I., CUSICK, M.E. et al. Drug-target network. *Nat Biotechnol*, 25(10), 1119-1126, 2007.

YOOK, S.H., JEONG, H. & BARABÁSI, A.L. Modeling the Internet's large-scale topology. *Proc Natl Acad Sci U S A*, 99(21), 13382-13386, 2002.

XU, C., MIN, J. Structure and function of WD40 domain proteins. *Protein Cell*, 2(3), 202-214, 2011.

ZHANG, N., KAUR, R., LU, X. et al. The Pso4 mRNA splicing and DNA repair complex interacts with WRN for processing of DNA interstrand cross-links. *J Biol Chem*, 280(49), 40559-40567, 2005.

7. APÊNDICES

APÊNDICE I

Human Diseases Associated With Genome Instability

Capítulo de livro publicado no livro *Genome Instability: From Virus to Human Application*

Chapter 26

Human Diseases Associated With Genome Instability

B.C. Feltes[a], J. de Faria Poloni[a], K.N. Miyamoto[a], D. Bonatto[a]
Federal University of Rio Grande do Sul, Porto Alegre, Brazil

Chapter Outline

1. INTRODUCTION

The main focus of the following section (Section 2) of this review is to introduce rare genetic diseases associated with different aspects and pathways of DNA repair. Molecular aspects regarding the connection among different aspects of the repair pathways are also considered. In Section 3, we address the main genetic alterations that drive cells to genome instability resulting in acquiring cancerous phenotypes. In addition, we discuss why understanding these phenomena are useful in oncological clinical care. Finally, in Section 4 we discuss epigenetic mechanisms that influence the cell-cycle regulation and the DNA-repair response. Furthermore, the topic also contemplates the most common abnormalities in the epigenetic-regulation mechanisms and their impact on the cell-fate acquisition.

2. RARE GENETIC DISEASES ASSOCIATED WITH DNA REPAIR

The DNA molecule is constantly threatened by a wide range of exogenous and endogenous mutagenic agents, such as reactive oxygen species (ROS), chemical pollutants, drugs, and radiation such as ultraviolet (UV) light [1,2]. However, during evolution, cells have established molecular mechanisms to protect and repair the DNA molecule. These include, but are not limited to, compacting the DNA in the form of chromatin, lowering its contact with the cellular environment, and developing repair mechanisms like the nucleotide excision repair (NER), base excision repair (BER), homologous recombination (HR), DNA interstrand cross-link repair (ICLR), double-strand break (DSB), and mismatch repair (MMR) [1,3]. However, if the damage is not repaired, the cell can undergo apoptosis, senescence, or can lose control of its mitosis and can start an abnormal proliferation and become a tumor [1].

Different consequences can arise from nonrepaired DNA mutations caused by defects in the repair mechanisms, the so-called genetic diseases. In the next two sections, we focus on rare diseases related to defects in the repair machinery: xeroderma pigmentosum (XP), Cockayne syndrome (CS), trichothiodystrophy (TTD), and Fanconi anemia (FA). All four diseases are associated with defects in genes that encode proteins related to DNA repair—XP, TTD, and CS phenotypes

a. All authors contributed equally to this work.

Genome Stability. http://dx.doi.org/10.1016/B978-0-12-803309-8.00026-4

are derived from mutations in genes that act on the NER pathway, whereas FA is a result from mutations in genes on the ICLR pathway [1,4].

We then focus on the diseases affected by mutations in RECQ family genes, which are Bloom syndrome (BS), Rothmund–Thomson syndrome (RTS), and Werner syndrome (WS).

In the last section, we discuss genetic diseases that are not specific to a single pathway, such as ataxia telangiectasia (AT) and Hutchinson–Gilford progeria syndrome (HGPS). Explanation of the whole spectrum of outcomes and molecular pathways of the listed diseases are complex, hence we focus on the major differences and how they are associated.

2.1 NER-Related Diseases: Xeroderma Pigmentosum, Trichothiodystrophy, and Cockayne Syndrome

In order to comprehend the complexity of CS, TTD, and XP, it is required to understand the functionality of the NER pathway. NER is specialized in removing UV-induced DNA damage, where 6,4-photoproducts (6,4-PP) and cyclobutane pyrimidine dimers (CPDs) are the most common lesions, although there are other types of UV-induced lesions [1]. NER is divided in two sub-pathways: the global genome NER (GG-NER), which is responsible for the removal of DNA lesion in nonactive genes, heterochromatin, and transcribed strands of active genes, and the transcription-coupled NER (TC-NER), responsible for removing DNA damage only from transcribed strands of active genes [5,6]. Molecularly, the main difference in both sub-pathways is that in TC-NER RNA polymerase is hindered in the lesion site with the aid of specific factors, the DNA-dependent ATPases (CSA, CSB) and the pre-mRNA splicing factor XAB2 that bind to the lesion where RNA polymerase is stalled, whereas in GG-NER the lesion is recognized by the *xeroderma pigmentosum*, *complementation group C* (XPC)-HR23B (RAD23B) heterodimer or the DDB-complex (composed by the DNA-binding proteins DDB1 and DDB2, and the ubiquitin–ligase complex CUL4A and ROC1). XPC-HR23B have a high affinity for 6,4-PP lesions and the DDB complex for CPD lesions, but it is known that the DDB complex recruits XPC–HR23B to the site once the damage is recognized [5]. After the damage recognition, both pathways follow a core NER reaction of damage excision as follows: (1) the recruitment of the TFIIH helicase complex to open the damaged site; (2) recruitment of XPA–RPA heterodimer to form a platform of protein–protein interaction; (3) DNA-damage excision by the endonucleases XPF and XPG; and (4) synthesis of a new DNA strand [5] (Fig. 26.1). XPF forms a heterodimer with the ERCC1 protein, and it is still debatable whether XPG is recruited to the excision complex or it is a subunit of the TFIIH complex [7].

CS, TTD, and XP are autosomal diseases characterized by hypersensitivity to sunlight, premature aging, and a shorter life span, but differ in the extension of other symptoms. XP was the first NER-related discovered disease, described in 1874 by Moriz Kaposi [1]. XP affects 1:250,000 individuals in Western countries and 1:45,000 in Japan and North Africa, where individuals show severe risk to develop skin cancer and sunburns, in which skin neoplasms can appear during childhood [1,2]. They also present ocular degeneration in the lids, cornea, and conjunctiva [8]. Neurological symptoms are less common, but can appear in some cases [1,2].

Moreover, CS was the second NER-related disease, discovered 62 years later, in 1936, by Edward Alfred Cockayne, and 44 years later, in 1980, TTD was described by Price [1]. In contrast to XP, CS, and TTD individuals commonly present cognitive impairments and neurological degeneration, cachectic dwarfism, skeletal and muscular defects as well for a facial characteristic called "bird-like" face, defined by deep sunken eyes and preeminent ears [2,9,10]. Some cases of CS can develop cerebro-oculo-facial-skeletal (COFS) syndrome, a disorder that can cause neurological and visual deficiencies, whereas TTD patients can present decreased fertility and osteosclerosis, combined with more aggressive neurological symptoms, such as tremors, low IQ, and incomplete myelination of nervous fibers [2,5,9]. It is also interesting to highlight that CS individuals do not exhibit skin cancer predisposition, although they are hypersensitive to sunlight, indicating that the TC-NER pathway is not required to prevent skin cancer.

The differences in each syndrome are related to the genes affected. As seen in Fig. 26.1, three proteins are specific for TC-NER, CSA, CSB, and XAB2, where mutations in CSA and CSB are responsible for the CS phenotypes. Mutations in CSA are related to Type I (classical) form of CS, where manifestations occur around the first years of life, and to Type III (mild), where individuals show a greater life span than other types and retain basic cognitive function such as walking and speaking [10]. On the other hand, mutations in CSB can manifest themselves as any type of CS, including Type II (severe), where individuals have a maximum life span of 7 years and display strong mental retardation and loss of basic cognitive functions [10].

In XP, the differences lie on which XP gene was compromised. There are seven XP genes in NER (XPA–XPG) (Fig. 26.1); mutations in any of those genes provoke an XP phenotype and in case of XPB, XPD, XPF, and XPG, some manifestations show a CS-like characteristic, such as neurological abnormalities [2,10].

Different from XP and CS, the core origin of the TTD phenotype lies in mutations in the helicases XPB and XPD, with XPD mutations being the major cause [9]. Thus, TTD phenotype appears to be related to the TFIIH complex activity more

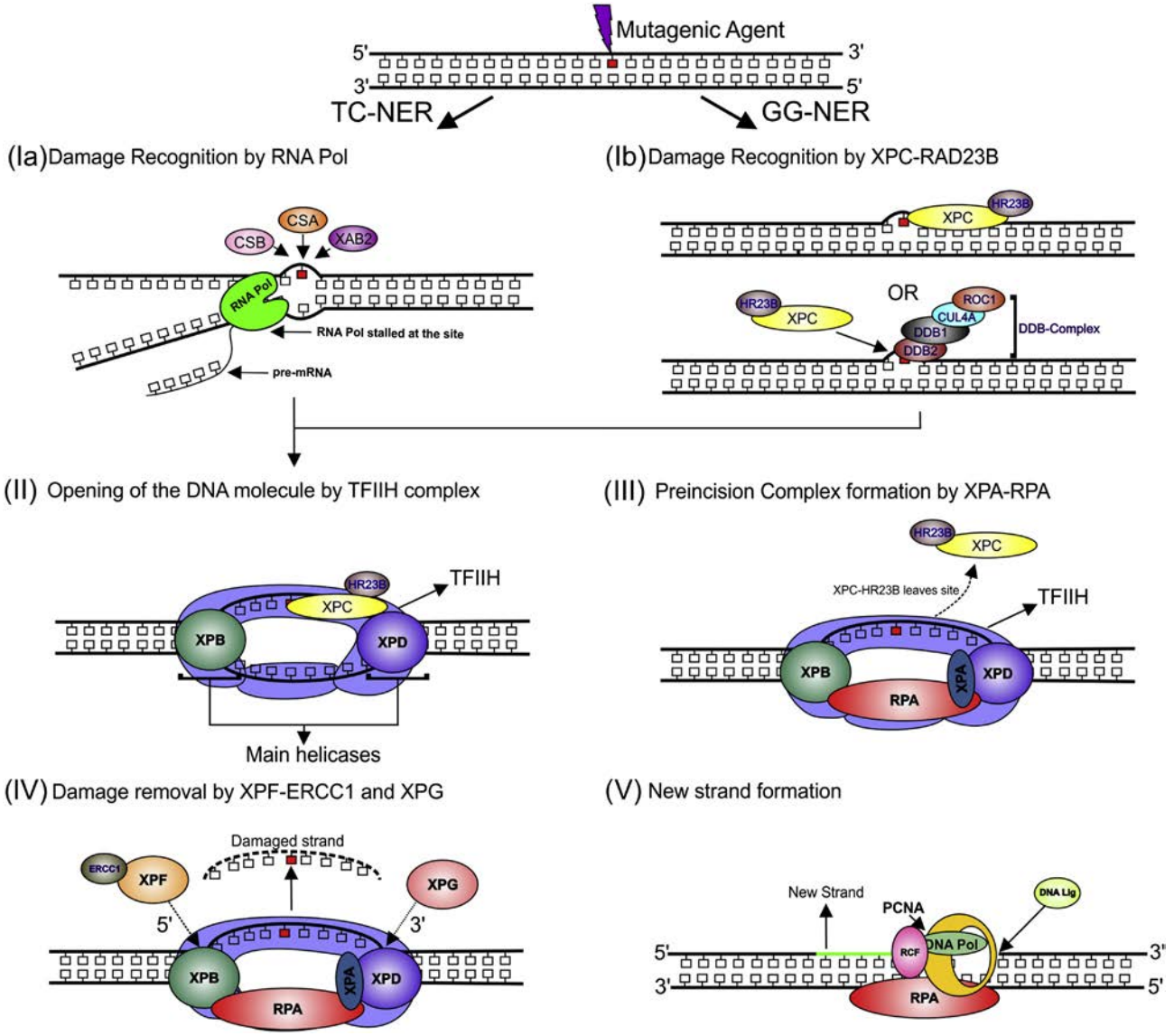


FIGURE 26.1 **The NER pathway divided by its fundamental steps.** It begins after DNA damage and one of the two sub-pathways is triggered. (Ia) TC–NER pathway is triggered if the damage site is in an active gene that is currently being transcribed. RNA Pol II is stalled at the lesion with the aid of CSA and CSB, that bind to the lesion and help in the recruitment of other factors, such as XAB2. (Ib) GG–NER can be triggered in any case, since it can act on heterochromatin, nonactive genes, and euchromatin. In this case, the damage is recognized by the XPC–HR23B heterodimer or the DDB complex. This separation is required to understand that it is a case of affinity: both complexes can recognize different damages, but they have higher specificity for a given substrate. XPC–HR23B have higher affinity for 6,4-PP while the DDB complex has it for CPDs. Nonetheless, XPC–HR23B is recruited by the DDB complex after it recognized the damage. After the initial recognition, both pathways converge in the core NER steps. (II) XPC–HR23B recruits the helicase complex TFIIH, where the 3′-helicase XPB and the 5′-helicase XPD act on opening the damage site. (III) XPC–HR23B leaves the site and the XPA–RPA heterodimer binds to unwounded DNA to allow a protein–protein platform. (IV) The 3′-endonuclease XPF and the 5′-endonuclease XPG are recruited to excise the damaged strand. (V) Finally, all proteins leave the site except for RPA which is required for the final polymerization step that recruits PCNA, DNA Polδ, and RCF to create a new strand of DNA. DNA ligase connects the new strand with the ends of the old strand.

than any other molecular aspect. The severity of TTD manifestation depends on what residue was mutated in the XPD protein, where the R112H, R592P, D673G, and R722W mutations are the cause of a severe phenotype, and R658C, R658H, and A725P result in a mild form of TTD [9].

It is interesting to observe that XP, TTD, and CS share the core NER reaction of damage excision, and the mutations in the XP genes related to this step can provoke similar phenotypes. Moreover, although all three diseases are associated with NER, the regular phenotypes are distinct, indicating that the proteins derived from the mutated genes are probably acting on other pathways beside DNA repair.

One explanation is that XP proteins, shared in both pathways, are interacting with a broad range of other proteins [11]. Taking the XP genes related to XP/CS/TTD phenotypes for example: XPB and XPD helicases are part of the TFIIH complex, where both exert structural role in interconnecting other subunits of the complex [11]. The TFIIH helicase complex is also required for regular transcription, and its deregulation can affect a broad range of cellular processes. In addition, XPB and XPD interact with proteins related to DNA repair such as p53, and to RAD52, where they would have a role in HR pathway [11]. Additionally, XPG is associated with the BER pathway by interacting with the protein NTH1. NTH1 plays a role in repairing thymine glycol mutations, and its affinity for the lesion is increased by XPG [11]. Finally, XPF interacts with RAD51 and RAD52 of the HR-repair pathway and with TRF2, a telomere elongation factor and with the Fanconi anemia, complementation group A (FANCA) protein, which is discussed later in more detail [11]. These relations clearly indicate that affecting XP proteins may disrupt a variety of mechanisms besides NER.

2.2 Fanconi Anemia

FA was discovered by Guido Fanconi in 1927. FA is distinct from XP, TTD, and CS, since the genes involved in the FA are mainly connected to the ICLR pathway instead of NER. ICLs are DNA lesions that covalently links paired strands of DNA, preventing the separation of the strands and the formation of the replication/transcription fork [12].

FA is mainly a hematopoietic disease that ultimately causes bone-marrow failure [12]. Individuals do not show most of the symptoms that appear in XP or CS, although they can present short stature and facial deformities [13]. Although the whole ICLR pathway consists of more than 30 genes, there are 16 FA genes related to ICLR, where 8 of them compose a multisubunit ubiquitin E3 ligase complex (FA core), and mutation in any of those 16 genes leads to FA [4,12–14]. A summary of the ICLR pathway can be found in Fig. 26.2. Broadly, the ICLR pathway can be divided into five stages: (1) damage recognition, (2) FA core recruitment, (3) complex assembly, (4) translesion polymerase activation, and (5) HR pathway triggering [4,12–16] (Fig. 26.2).

One interesting aspect of this pathway is its link to NER, since XPF is one of the main proteins that act on the pathway, and one of the responsible proteins for the XP/CS phenotype [2,10,15]. Remarkably, a 2013 work from Kashiyama et al. [17] described a patient who showed phenotypes associated with XP, CS, and FA [17]. The XPF–ERCC1 heterodimer is extremely important to proper removal of damaged sites, and it is known that ERCC1 mutant mice display neurodegeneration, whereas mutation in XPF results in a genetic disease called XPE progeroid syndrome, which is characterized by premature aging and aging related-diseases [15,17]. These studies indicate that there is a connection between ICLR and NER that may result in a combined phenotype of the three diseases, although this association is yet to be established. It is possible that this association was evolutionarily selected to enhance the response to DNA damage.

2.3 RECQ-Related Diseases: Rothmund–Thomson Syndrome, Werner Syndrome, and Bloom Syndrome

Another syndrome associated with the DNA-repair pathways, like BS, RTS, and WS arise from mutations in the RECQ helicase family. In this sense, before we discuss each syndrome individually, a broad view of the RECQ role on DNA repair is necessary [18].

The human RECQ helicases family consists of five proteins, RECQL1, WRN, BLM, RECQL4, and RECQL5, all of which play a crucial role in DNA damage–sensing and –repair pathways, either for helping other proteins to assemble the repair machinery or to recognize and unwind specific DNA rearrangements (Fig. 26.3) [18]. They interact with DSB repair–pathway proteins at different stages, and when DSB repair is initiated by HR, they are important at the initial DSB recognition, further disassembly of RAD51–ssDNA nucleoprotein filaments during recombination, and in resolving double Holliday junctions (DHJ) (Fig. 26.3-IV) at the branch migration phase [18]. On the other hand, in nonhomologous end joining (NHEJ) pathway, RECQ helicases act by modulating protein complexes like the DNA-dependent protein kinase catalytic subunit (DNA-PKCS), the Ku70/80 heterodimer, which detect DNA damage, and the XRCC4/ligase IV, responsible for DNA end ligation [18]. RECQ family, especially WRN, also mediates base lesion targeting in BER pathways [19], which is discussed later. Additionally, the role of RECQ helicases in the NER pathway remains poorly understood. Finally, RECQ proteins are also important to replication events; they are recruited at stalled or collapsed replication forks, interact with replication repair–machinery proteins, mainly RPA, guiding the DNA-damage fixing, and further replication restart [19]. This is an interesting fact, since RPA is also a close partner of XPA during NER (Fig. 26.1-III) [1,2].

Since the RECQ family is essential for genome maintenance, mutations in these genes could cause defects in many repair pathways, leading to genome instability. For this reason, diseases like WS, BS, and RTS, all caused by mutations in members of the RECQ family, are characterized by a wide range of symptoms and cancer development. We address each syndrome individually.

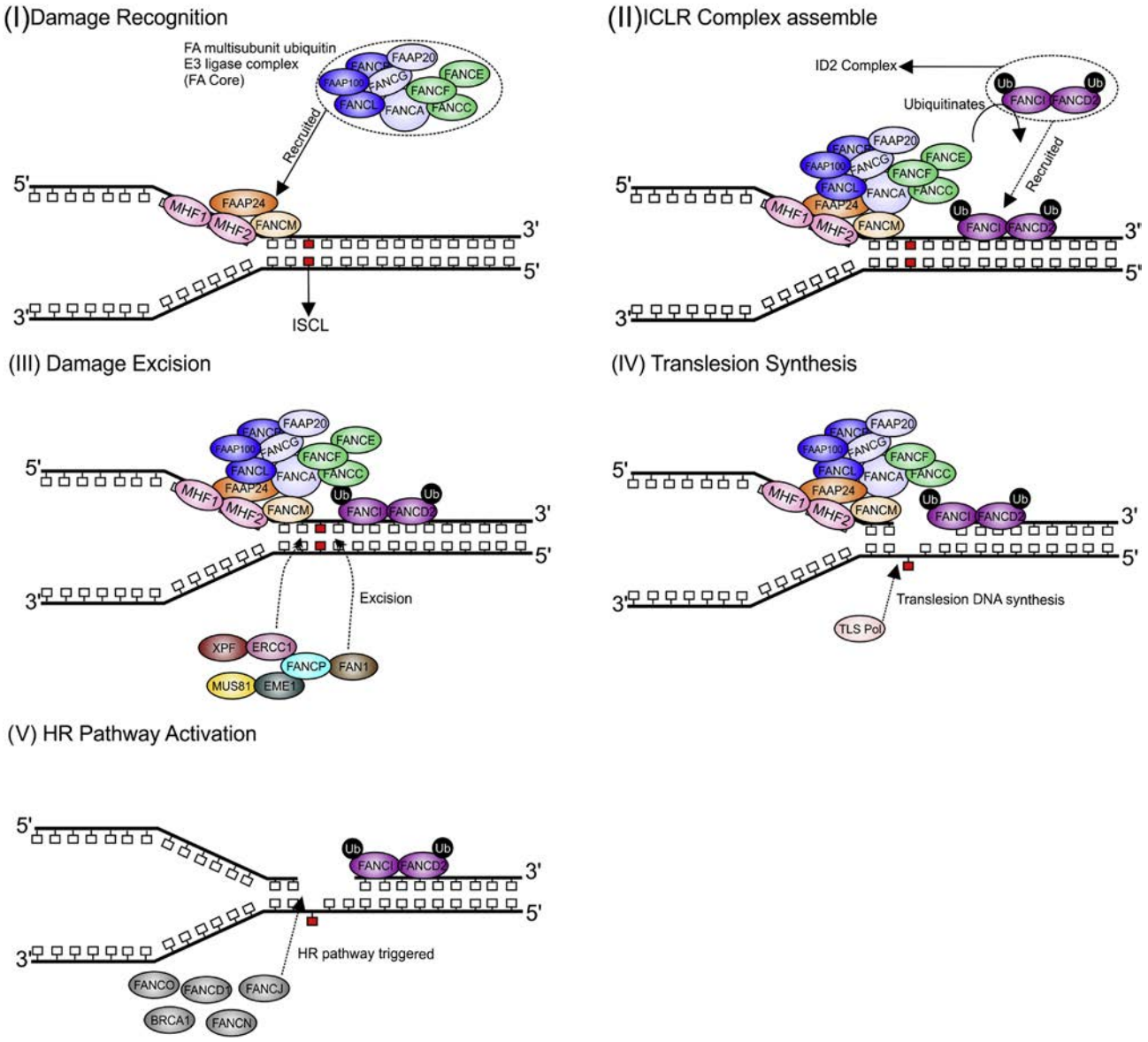


FIGURE 26.2 **The ICLR pathway divided by its fundamental steps.** (I) It begins with the ICL recognition by the proteins FANCM, MHF1–MHF2, and FAAP24 that bind to unwounded DNA and recruit the FA core. The FA core is composed by three subcomplexes, one composed of FANCL, FAAP100, and FANCB (blue); the other composed of FANCG, FANCA, and FAAP20 (light gray); and the third composed of FANCF, FANCC, and FANCE (green). (II) The FA core then ubiquitinates the ID2 complex, which is composed of FANCI and FANCD2 and binds to the unwounded DNA. With the complex formed, the excision machinery composed of XPF–ERCC1, MUS81–EME1, FANCP, and FAN1 excise the damaged region. (III) TLS polymerase than adds nucleotides to the removed strand. (IV) Finally, the HR machinery is triggered by FANCO, BRCA1, BRCA2, FANCD1, and FANCJ.

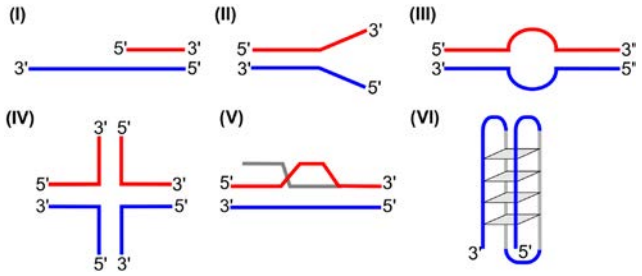


FIGURE 26.3 **Types of DNA arrangements that are substrates of the RECQ family.** (I) 3′-tailed DNA. (II) Forked DNA. (III) "Bubble" structured DNA. (IV) Holliday junction. (V) D-loop. (VI) G-quadruplex.

RTS is an autosomal recessive disease first described in 1868 by August von Rothmund, and then redescribed, with addition of phenotype variances by Matthew Sydney Thomson in 1926 [20]. However, the term Rothmund–Thompson was just coined in 1967 by William Taylor [20]. RTS major symptoms are epidermis-related tissue malformation (eg, hair, skin, and nails), short stature derived from skeletal malformations, cataracts, and cancer predisposition, especially osteosarcomas and spinocellular carcinomas [20].

RTS is caused by mutations in the RECQL4 protein, and ATP-dependent helicase, that is the only RECQ helicase present in both nucleus and mitochondria [21]. RECQL4 is necessary for the initiation of DNA replication and studies with RTS patients have shown that RECQL4 is also associated with sister chromatid separation, DSB, and BER repair pathways and telomere replication [18,20,21]. This protein can unwind forked duplexes, Holliday junctions, G-quadruplex structures, "bubble" structures, and D-loops, but cannot unwind normal duplex DNA (Fig. 26.3) [22]. Interestingly, RECQL4 shows little DNA-unwinding activity, when compared to other RECQ proteins, and seems to be more prone to anneal DNA [21]. Another fact to be observed is that mutations in RECQL4 are also related to two other syndromes: (1) Baler–Gerold syndrome (BGS), characterized by craniosynostosis and radial hypoplasia, together with short stature, and (2) the radial hypoplasia, patella hypoplasia and cleft or arched palate, diarrhea and dislocated joints, little size and limb malformation, slender nose and normal intelligence (RAPADILINO) syndrome [20].

Since RECQL4 is the only of the RECQ family to be present in the nucleus and the mitochondria, further studies focusing on understanding of RTS relationship with mitochondrial function are necessary. Additionally, the fact that RECQL4 appears to be colocalized with XPA after UV irradiation in the nucleus suggests a possible role for RECQL4 in NER that might show promising explanations for the RTS phenotype and possibly for XP [22].

Another autosomal recessive disease related to defects in genes of the RECQ family is Bloom syndrome, which was discovered by David Bloom in 1954 [23]. BS individuals present morphological abnormalities, such as long limbs, short stature, and the bird-like features similar to CS and TTD, as well as low subcutaneous fat content and dermatological conditions like photosensitivity and poikilorderma [24]. In addition, patients show high predisposition to cancers (eg, breast, larynx, skin, and colorectal cancers), lymphoma, and leukemia [24].

BS is another disease caused by a mutation in a gene of the RECQ family—in this case, the ATP-dependent RECQL2 known as BLM [18,25]. BLM can unwind the same DNA structures as RECQL4 plus 3′-end of the DNA, but shows preference for unwinding G-quadruplex DNA (Fig. 26.3), [25]. Similarly to RECQL4, BLM also has an ss-DNA annealing capacity, although the full mechanism by which it can promote strand annealing is not fully understood [25]. BLM also repairs centers of collapsed or stalled replication forks, where it appears to promote fork regression [18]. Thus, the loss of BLM function is related to a broad range of chromosomal aberrations and cancer formation [24,25]. It is not a surprise that BS individuals show high levels of sister chromatids exchanges, chromosomal breakage, translocation, and chromosomal quadri-radials [24,25].

One interesting molecular aspect of BLM is the fact that it appears to interact with FANCM, one of the proteins that comprises the FANC complex [24], showing that there might be a connection between the molecular pathways that lead to BS and FA. FANCM-deficient cells show high levels of sister chromatids exchange, similar to BS-derived cells [26]. A study made by Hoadly et al. [26] suggests that FANCM binds to the damage site and recruits the BLM complex (composed of BLM, topoisomerase IIIα, and the RMI1/RMI2 heterodimer) through its interaction with the RMI1/RMI2 heterodimer [26].

Chromosomal maintenance is a complex process that involves a wide variety of proteins and pathways, and understanding the interplay between BLM and these other mechanisms might help improve the knowledge about BS.

Finally, the last disease related to mutations in the RECQ family is WS, an autosomal recessive disease originated from mutations in WRN gene, which encodes for RECQL3/WRN protein. WS was described for the first time in 1904 by Carl Wilhelm Otto Werner. Its clinical manifestations are extensive, although most of the presented symptoms are aging related, such as atherosclerosis, diabetes mellitus type 2, osteoporosis, and cataracts, among others [27]. Nonaging-like symptoms include hypogonadism, reduced fertility, low height, and others. In addition, WS patients are also susceptible to development of tumors, especially sarcomas [28]. However, even though this could represent an important risk factor, most WS-affected individuals decease by a myocardial infarction between their fourth and fifth decades of life [28].

WRN protein has a helicase domain which has specificity for certain DNA structures (Fig. 26.3), especially G-quadruplex (Fig. 26.3-VI.) and Holliday junction (Fig. 26.3-IV.), DNA structures found mainly in telomeric DNA and in the recombination process, respectively [19,29]. In addition, WRN has a unique 3′–5′ exonuclease activity that digests 3′-recessed termini or blunt DNA duplexes that contain structures like bubbles (Fig. 26.3-III), forked duplexes (Fig. 26.3-II), Holliday junctions, and DNA–RNA heteroduplexes. This exonuclease acts coordinately with a helicase domain through the DNA duplexes size reduction, allowing proper helicase-unwinding role, although both domains have also independent functions [28].

WRN interacts with many regulating proteins related to DNA-repair pathways. For example, in the long-patch BER (LP-BER) pathway, WRN interacts with NEIL1, a formamidopyrimidine lesion glycosylase, DNA polymerase β, responsible for base replacement at the lesion site, and with FEN1, an endonuclease that removes the 5′-overhanging flap [19]. In DSB repair, WRN plays an active role in NHEJ by interacting with KU70/80, DNA-PKCS and XRCC4/ligase IV complexes. During HR, WRN interacts with proteins like RAD51 and RAD52, which are fundamental during strand invasion and annealing [28]. Moreover, WRN is important in telomere replication and maintenance pathways, recognizing D-loops (Fig. 26.3-V), G-quadruplex structures in telomeric DNA, and interacting with the shelterin complex proteins, such as TRF1, TRF2, and POT1 [29]. Conversely, WS cells present characteristics that are directly associated with impaired DNA repair, including telomeric erosions, oxidative DNA damage, and a defective DNA interstrand cross-link removal, which can contribute to develop aging-related symptoms and tumorigenic processes like sarcomas in WS patients [29].

2.4 Ataxia Telangiectasia

Different from all the previously discussed diseases, AT is not caused by defects in a specific DNA-repair pathway. AT is a rare autosomal recessive disorder described originally in 1926, but the term "ataxia telangiectasia" was suggested by Elena Boder and Robert P. Sedgewick in 1957 [30]. The clinical characteristics of AT disorder are progressive neurological dysfunctions, which includes oculocutaneous telangiectasia and cerebellar ataxia. In addition, individuals with this disorder show cancer predisposition, susceptibility to bronchopulmonary disease, and multisystem abnormalities, such as immunodeficiency, radiosensitivity, infertility, and endocrine dysfunctions [30,31].

AT is caused by mutations in ataxia telangiectasia–mutated gene (ATM), located at chromosome 11q22–23, which encodes to an ATM serine/threonine kinase [30]. The ATM protein is a member of the phosphoinositide 3-kinase (PI3K)-related protein kinase (PIKK) family, which is able to induce a DNA-damage response [30,31].

ATM is activated by different biological processes, such as cell-cycle checkpoint and DNA damage [30]. Indeed, many ATM substrates are cell-cycle regulators with important roles in DNA-damage response, such as p53, CHK2, and BRCA1 [31,32]. During DNA-damage repair, a complex composed of MRE11–RAD50–NBS1 (MRN complex) recognizes DSB and leads to ATM activation, as well as performs an adaptor role to subsequent phosphorylation of downstream ATM substrates [30–32]. In response to the MRN-complex signaling, ATM undergoes autophosphorylation at serine 1981, and is converted from inactive multimeric to an active monomeric kinase [30,31]. Once activated, ATM orchestrates a signaling cascade in response to DSB that coordinates cell-cycle arrest, DNA repair, or the cell-apoptosis process [30,31]. The response capacity of ATM to DNA damage is the primary in vivo function of this kinase and is intrinsically related to phenotypes of AT disorder [31].

In this sense, the role of ATM in DNA repair involves the phosphorylation of specific repair factors, like KRAB-associated protein 1 (KAP-1), which relaxes chromatin structure and allow the accessibility of repair proteins (the chromatin role in DNA repair is discussed later) [31]. Another target of ATM phosphorylation is the FANCD2, a protein whose defects leads to FA; in response to DNA damage, FANCD2 is phosphorylated at Ser222 [32].

2.5 Hutchinson–Gilford Progeria Syndrome

HGPS is a progeroid syndrome described at first time by Jonathan Hutchinson in 1886 and by Hastings Gilford in 1897 [33]. However, this syndrome was described in greater detail in 2003, when the molecular basis of the disease was discovered [33,34].

Despite HGPS patients born with normal appearance and weight, the clinical symptoms appear within 12 months and progress rapidly [35]. The best described characteristic of HGPS patients is the development of age-related diseases, such as cardiovascular pathologies, prominent superficial veins, skin complications, and alopecia. In addition, these individuals show disturbed growth, lipodystrophy, joint abnormalities, and osteolysis [35]. This disease is caused by a single nucleotide substitution on the gene *LMNA* that encodes the A-type nuclear lamin proteins [33,34]. Differential alternative splicing generates A-type lamin proteins, with the most abundant being lamins A and C; however, mutated *LMNA* leads to aberrant splicing that results in the deletion of 50 amino acid residues from C-terminal region of prelamin A [33,34,36]. This aberrant splicing produces a mutant protein called "progerin," which accumulates in a farnesylated form, affecting the nuclear organization, chromatin dynamics, epigenetic regulation, and gene expression, causing genomic instability, premature senescence, and telomeres disruptions [34,36].

It was observed that progerin modifies the composition and mechanical properties of nuclear lamina, which are related to abnormal nuclear morphology [34,36]. This occurs due to the high affinity of progerin to nuclear envelope and by immobilization of A-type lamins in the nuclear lamina induced by progerin. In addition, high levels of γH2AX phosphorylation are an

indicative of activated DNA-damage response and cellular accumulation of DNA damage [34,36]. Furthermore, HPGS cells show reduced survival and proliferation, besides sensitivity to DSB and delayed recruitment of repair proteins [34,36].

Among the proteins involved in this impaired recruitment are the components of MRN complex, which are crucial for HR [36]. Furthermore, another DNA-repair defect observed in HPGS cells is the mislocalization of XPA to DSB that can be associated with the delay activation of DNA-repair proteins, such as the MRN complex [36,37]. Thus, HGTS molecular pathways have an interplay with the ATM related–molecular mechanisms that may lead to AT, as well as with NER-related proteins, such as XPA.

2.6 Rare Genetic Diseases: Summary

In conclusion, the understanding of DNA-repair diseases is crucial to boost the knowledge of the DNA-repair machinery and the consequences of its defects. Moreover, the phenotypes that result from diseases described earlier go from a broad range of anatomical abnormalities to neurodegeneration and cancer development, indicating that they regulate much more than DNA repair. For example, XP proteins are also targets of different post-translational modifications and have different protein–protein interaction sites that may answer how they are regulated and to what proteins they may be connected [5]. Other examples are CSA and CSB, that are also known to regulate cellular redox balancing, where cells lacking CSA and CSB show increased level of ROS [38]. These proteins are also connected to BER and the maintenance of the stability of mitochondrial DNA [38]. CSB is associated with biological processes, such as cell growth, angiogenesis, proliferation, and cell death [39]. The same logic goes for the proteins related to FA, BS, RTS, WS, AT, and HGTS which comprise large complexes and are connected to multiple proteins.

3. CANCER AND GENOME INSTABILITY

It is impossible to discuss genetic diseases without mentioning cancer, since it is intimately associated with DNA-repair defects and is a common outcome of the rare genetic diseases described before. Cancer onset begins when precancerous cells acquire uncontrollable growth, sustain angiogenesis, and become able to invade different tissues [40]. These are the main factors that contribute to the extent of cancer malignancy. However, since the human organism has redundant and self-regulating pathways to maintain homeostatic conditions, an extensive set of genes must be affected to reach conditions necessary to carcinogenesis. This change in genome profile can be achieved if DNA damage–repair systems and/or replicating machineries work improperly, or when cells are exposed to mutagenic or genotoxic agents, such as tobacco smoke, UV light, and ionizing radiations, among others. The increased mutation rate that changes drastically the genome landscape deregulates basal expression and surpasses genome integrity, and cell-cycle surveillance promotes cells transformation. It is important to reinforce the idea that genome instability makes cells more susceptible to carcinogenesis [41]. Since there are many pathways associated with cancer development, the focus of this section is to describe the basis of the different types of genome instability, as well as its relationship with cancer development and how understanding these phenomena can be useful in clinical practice and therapy.

Genome instabilities can vary from a single nucleotide mutation to a whole chromosome structure modification (clastogenesis). These changes alter cell homeostasis in many ways, depending on which genes are affected. Single nucleotide mutations arise usually due to the high cell exposure to DNA-damaging agents or when DNA-repair genes involved—for example in NER, BER, and MMR—are mutated, although a nucleotide-deficient environment could also promote such imbalance (Fig. 26.4-I) [42,43]. For example, XP patients are highly susceptible to sunlight UV-induced carcinogenesis (about 2000–10,000 times higher than a healthy person) due to the accumulation of nucleotide mutations that are caused by a deficient NER from one or more mutated XP proteins [8]. Some regions of the genome, containing repeated nucleotide sequences—one to six nucleotides repeated multiple times—called microsatellites, suffer more extensive modifications caused mainly by an inefficient MMR system, leading to insertions and/or deletions (indels) in these regions during the S phase of cell cycle (Fig. 26.4-II). This microsatellite instability (MSI) promotes frameshifts in coding sequences of genes resulting in truncated or nonfunctional proteins. MSI is present in some cancers, mainly in colorectal tumors, which corresponds to 15% of these cases [44].

Chromosomal instability (CIN), especially chimeric chromosomes and/or aneuploidies, on the other hand, is a common genome instability in most cancers. Since the discovery of Philadelphia chromosome, formed by the chromosome 9 and 22 translocation, much has been done to understand the importance of CIN events in tumor progression. In this sense, defective DNA-repair mechanisms can induce multiple chromosomal fragmentations. For example, DSB-repair failures contribute to CIN generation, creating a chromothripsis phenomenon characterized by multiple chromosome breaks and rearrangements (Fig. 26.5-I) [45]. In addition, other factors also contribute to the acquisition of an abnormal karyotype

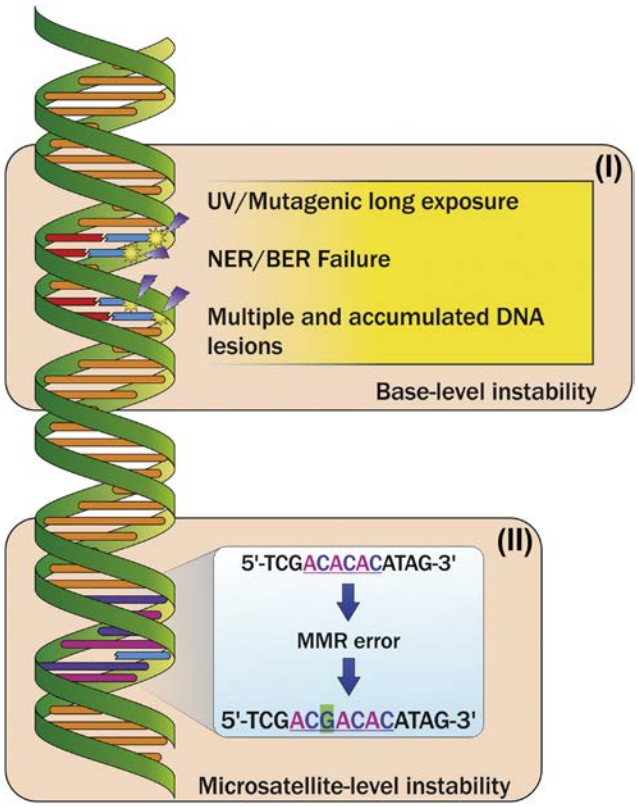


FIGURE 26.4 **Cancer-related DNA mutations that cause genome instability.** (I) Base-level accumulated mutations, caused by high exposure to mutagenic agents and/or defective NER/BER pathways. (II) Indel events in microsatellite regions caused by defective MMR systems.

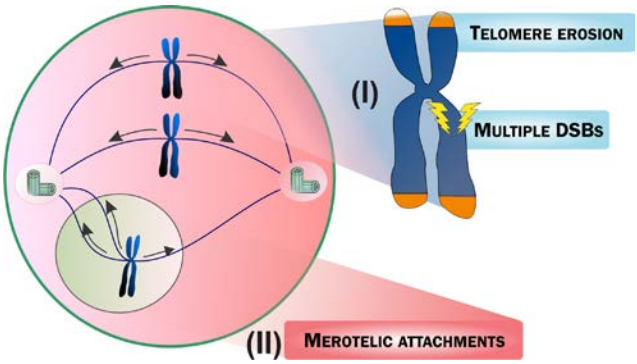


FIGURE 26.5 **Cancer-related events that lead to chromosome instability and clastogenic phenomenon.** (I) Chromosomal breakage and rearrangements by the progressive telomere shortening and/or double-strand breaks. (II) Multiple microtubules attached to one kinetochore (merotelic attachments) that causes spindle pole asymmetry and incorrect chromosome segregation.

in cancer cells. Telomere shortening or loss is one of the main driving forces of chromosome fragmentation. These six tandem repeated sequences at the end of chromosomes, along the protein complex called shelterins, are crucial to protect chromosomes from fusing to each other during cell cycle and generating aberrant chromosomes (Fig. 26.5-I). When cells naturally cease to express telomerase, they die due to a telomeric erosion condition in which cells either enter a replicative senescence state or begin to generate multiple chromosome fusions [46]. This process is tightly controlled in the cell and when they are impaired, the cells are able to proliferate, leading to uncontrolled cell growth. Curiously, tumor cells can even restabilize their genome through the reexpression of telomerase or via a homologous recombination alternative lengthening of telomere (ALT) mechanism, although the triggering mechanisms remain unclear, especially for ALT [47–49].

Finally, another CIN-inducing event is the incorrect segregation of chromosomes during cell cycle. Mutated proteins responsible for chromosome organization and cell structure can compromise the sister chromatids separation (karyokinesis)

or cells separation (cytokinesis) resulting in aneuploidies [50]. Centrosome dynamics during mitosis, for example, is a very coordinated process that requires a myriad of signaling proteins. Thus, any change in the velocity of chromatids separation during the M phase of the cell cycle can induce CIN. This could make the chromosome's kinetochore attach to microtubules coming from both spindle poles, instead of just one of them, forming merotelic attachments (Fig. 26.5-II). This creates a spindle asymmetry that compromises the correct chromosome segregation and creates a lagging chromosome that will result in aneuploidic daughter cells [51].

One of many hypotheses that tried to explain the origins of genome instability is the mutator phenotype, which suggests that mutations in genes involved in the genome-maintenance pathways, known as caretaker genes, makes cells more prone to DNA lesions or replication failures that facilitate the cancer development [52]. In hereditary cancers, mutations in caretaker genes mostly drive cells to genome instability, as observed in the Lynch syndrome (in which MSI caused by MMR failure triggers oncogenesis) as well as other inherited DNA-repair gene mutations, such as mutations in the FANC family and in the BRCA1 gene. However, as demonstrated in many high throughput–sequencing studies, these types of mutations are unlikely to occur in sporadic cancers, and even if they do, it requires that both alleles must be affected to drive cells to genome instability [40]. In these nonhereditary cancers, the most accepted hypothesis is that the same altered pathways that activated oncogenes also drives deacceleration or stalling of replication fork progression, especially in regions denominated as common fragile sites, creating chromosome breakage at these locations. This favors the selection of defective tumor-suppressor genes, such as TP53, and genome amplification of other oncogenes, leading to cancer development through the escape from apoptosis and senescence [40,53–55].

Understanding how complex patterns of genome instability events contribute to cancer has many implications in the clinical health care. In cancer diagnosis, detection of chromosome instabilities is important to determine tumor aggressiveness and patient's prognosis. For example, MSI-containing colorectal cancers are considered to have more favorable prognosis as compared to stable microsatellite colorectal cancer types. Some authors hypothesized that translation frameshifts caused by MSI generate novel peptides at C-terminus region that are immunogenic and stimulate an inflammatory response against tumor cells [44].

Also, the comprehension of how cancer begins and develops is crucial for new chemotherapeutics drug design. One strategy is inducing mitotic catastrophe by small molecules that act on the kinetochore and spindle poles assembly proteins, such as aurora kinase inhibitors [56]. However, since many of these potential chemotherapeutics are toxic to the bone marrow, many types of DNA damage–response inhibitors were tested as adjuvants to maximize genomic instability in cancer cells, promoting mitotic catastrophe and apoptosis and avoiding potential drug resistance [57,58].

Therefore, cancer development can be accelerated by genome instability. High-proliferative capacity, sustained angiogenesis, cell cycle–checkpoint evasion, for example, are part of so-called "Cancer Hallmarks," and most of these features are acquired by multiple events of genome instability [59,60]. Although more studies are required to understand the complex relationship between cancer and genome rearrangements, especially in sporadic tumors, there is still plenty of information available that can help oncologists in clinical care to establish patient prognosis and to search for potential anticancer targets.

4. EPIGENETIC REGULATION OF CELL CYCLE AND DNA REPAIR IN CANCER

The transformation of healthy cell toward a cancerous cell occurs gradually by a series of factors, including genetic and epigenetic modifications. Proper maintenance of epigenetic marks is essential to healthy cells and is associated with cell-fate acquisition [61–63]. Epigenetic alterations can change chromatin structure to loose state, which is transcriptionally active (called euchromatin) or to compact state, resulting in a transcriptionally inactive configuration (called heterochromatin) (Fig. 26.6-I) [64]. In this sense, chromatin alterations changes DNA accessibility and are responsible for modulation of the gene expression by affecting the interaction of DNA with transcriptional complexes, resulting in activation or inhibition of different signaling pathways (Fig. 26.6-I) [61,63]. Furthermore, histone modifications may affect DNA–histone or histone–histone interactions, or recruit nonhistone proteins to chromatin, creating a binding site for specific proteins that can act as regulatory factors [62,65]. In addition, different biological processes such as transcription, cell cycle, DNA repair, and replication are regulated by posttranslational histone modifications [62,65].

During cell cycle, checkpoints are surveillance systems that have the capacity to interrupt cell-cycle progression [66]; however, abnormalities in checkpoints and signaling pathways associated with proliferation are commonly observed in cancerous cells [66]. An example of signaling pathway disturbed in cancerous cells that is addressed in this chapter involves the retinoblastoma tumor-suppressor protein (RB) [66]. RB is a tumor suppressor whose activity is associated with different biological processes, such as differentiation, apoptosis, DNA-damage response and repair, DNA replication, and cell cycle [67]. During cell cycle, RB binds to the transcription factor E2F and prevents the transcriptional activation of E2F target

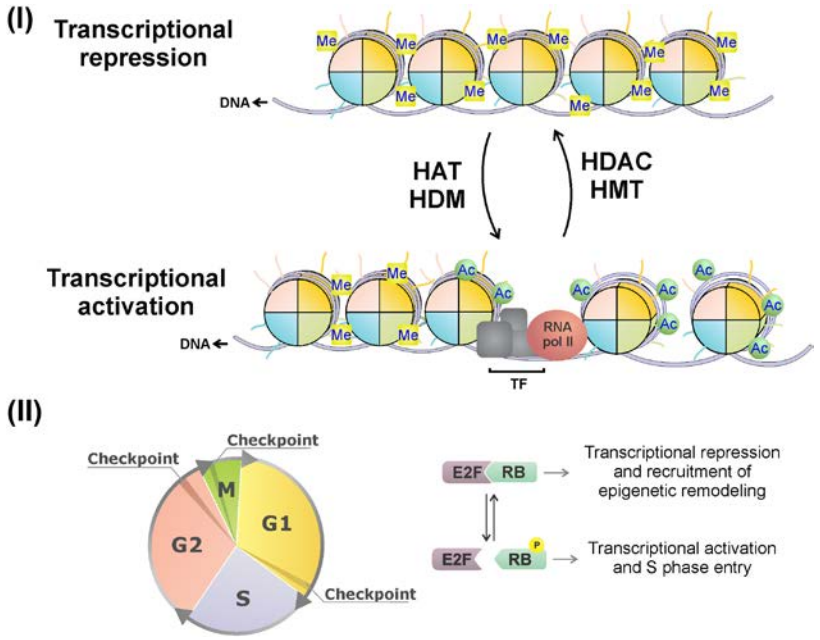


FIGURE 26.6 **(I) Simplified schematic of a transcriptionally inactive configuration of chromatin changing to an active chromatin configuration.** This transition is regulated by modifications in histone tails, where methylation and acetylation are most commonly observed. The proteins responsible for this transition include histone acetyltransferases (HAT), histone demethylases (HDM), histone methyltransferases (HMT), and histone deacetylases (HDAC). (II) Cell-cycle representation and RB-associated transcriptional-regulation mechanism. RB interacts with E2F, preventing the transcription of E2F target genes and recruiting chromatin remodelers to lead the transcriptional repression of genes associated with the cell-cycle progression. Phosphorylated RB is unable to interact with E2F, allowing the transcription and cell-cycle progression.

genes (Fig. 26.6-II) [67]. E2F is a transcription factor that regulates the expression of different genes associated with DNA synthesis and cell-cycle progression from G1 to S phase [66,68]. The RB activity is linked to the inhibition of cell cycle by interacting with E2F, leading to the down-regulation of specific cell cycle–related genes (Fig. 26.6-II) [66,68]. To make this inhibition more efficient, RB also recruits chromatin remodelers (Fig. 26.6-II), such as the co-repressor SIN3 transcription regulator family member B (SIN3B) that promotes lysine deacetylation from histone tails by recruiting histones deacetylases 1 and 2 (HDAC1 and 2) [68]. Lysine acetylation is correlated to transcriptional activation, and its deacetylation leads to a more compacted chromatin structure and consequently, transcriptional repression of E2F-target gene promoters [61,68]. Furthermore, histone methyltransferases (HMT), DNA methyltransferase 1 (DNMT1), and heterochromatin protein 1 (HP1) also are chromatin remodelers recruited by RB stimulation, promoting methylation in promoter region of genes regulated by E2F, contributing to its transcriptional repression [67]. Accordingly, RB is a crucial tumor suppressor, and it is necessary to ensure proper cell-cycle progression, one that promotes the silencing of genes that regulate cell-cycle progression and DNA replication [67].

RB mutations have been associated with reduced H4K20 trimethylation on the N-terminal tail [67,68]. Methylation state of H4K20 has an important role in cell cycle, and has been associated with cell-cycle progression, transcription, chromosome condensation, and origin firing for DNA replication [68]. It is observed that mono- and dimethylation are related to DNA repair and DNA replication, whereas trimethylation of H4K20 is associated with silenced heterochromatin formation and with cell-cycle arrest [68,69]. In addition, aberrant methylation in RB promoter leads to decrease in RB gene expression, and these abnormalities have been observed in different cancers, such as retinoblastoma, bladder cancer, neuroblastoma, and gastric carcinoma, among others [70].

Cell growth and division are both processes regulated tightly by a set of coordinated proteins that monitor cell-cycle progression and DNA integrity. The loss of cell-cycle control and DNA-damage propagation has emerged as the main inducer of cancer and other diseases. In this sense, modifications in the DNA sequence may alter gene products or lead to a loss of gene function. To prevent such genetic deregulation, DNA-damage response is activated, leading subsequently to cell-cycle arrest, recruitment of DNA-repair machinery, and damage correction or apoptosis [71]. Defective activity of epigenetic regulators can also lead to gene expression deregulation and, consequently, cell transformation [72]. As a result, proto-oncogenes expression can be activated by promoter hypomethylation, whereas the expression of tumor suppressors may be silenced by its promoter hypermethylation [72].

Thus, proper DNA repair is necessary for the coordination between chromatin modifications, cell cycle, and DNA-repair machineries [65]. In this sense, different proteins are able to mediate the communication among chromatin and repair, such as ATM, whose activation occurs in response to chromatin structure changes, like formation of DSB [65]. In addition to ATM, DNA-dependent protein kinase (DNA-PK) and/or Rad3-related protein (ATR) mediate the phosphorylation on the variant histone H2AX, which creates a binding motif to mediator of DNA damage–checkpoint protein (MDC1), that recruits other DNA-repair proteins, such as E3 ubiquitin–protein ligase (RNF8) and Nijmegen breakage syndrome 1 (NBS1) [65]. H2AX phosphorylation is the histone modification in response to a DNA break, but it is viewed to act as a broad signal in response to DNA damage, although primarily in the form of DSBs, as well as a triggering pathway in response to stalled replication forks (Fig. 26.7-I) [73].

At DNA-damage sites, the ubiquitin ligase Rnf20/Rnf40 mediates the ubiquitylation of H2B [65]. H2BK123 ubiquitylation is necessary to H3K4 and H3K79 methylation, being these modifications are required to alter the chromatin structure and allow the access of proteins involved in DNA repair (Fig. 26.7-I) [65,73].

Nonetheless, in DNA-damage region, methylated H4K20 (Fig. 26.7-I) acts as a binding platform to the P53-binding protein (53BP1), providing a stable 53BP1–chromatin association [69]. In human cancer–derived cells, the decreasing H4K20 trimethylation has been proposed as a common hallmark related to cell transformation [74]. This transformation can be associated with the fact that low H4K20 methylation avoids the repression of genes that regulate cell-cycle progression [74].

Another histone modification is the acetylation of H3K56 (Fig. 26.7-I) that occurs in response to replication fork damage [73]. During DSB repair, H3K56 acetylation is necessary for Rad52-dependent repair and promotes sister chromatids recombination [73].

Furthermore, in response to UV irradiation, H3K9 (Fig. 26.7-I) is acetylated during the NER process [73]. H3K9 acetylation regulates two pathways, the recruitment of histone acetyltransferase GCN5 to DNA lesions and the coordination of the activity of tumor suppressor p53 and acetyltransferase p300 [73]. In addition, ING2 activity is required to enhance the p53 and p300 interaction to induce the histone acetylation and to mediate a relaxation in chromatin structure; it is also required to recruit the XPA protein to the DNA lesions caused by UV irradiation [62,75]. In this sense, ING2 promotes the chromatin remodeling, which is adequate to DNA-damage repair and to the proper NER [75]. Many aspects of tumor biology, including cancer invasion and metastasis, are associated with deficiency in ING activity [75].

The events described earlier illustrate the importance of chromatin modifications and remodeling, acting in the DNA damage–response signaling and allowing formation of the complexes that mediate the DNA repair. In this sense, according to histone modifications, different chromatin–protein interactions are allowed, interfering with the propagation of repair signaling and turnover of factors involved in repair signaling. In addition, once repair process finishes, histone modifications are frequently reversed, allowing to establish a "prior to DNA damage" chromatin state.

Chromatin remodeling during cell cycle and DNA repair is a mechanistic step that allows or impairs the access to the specific DNA regions. Different chromatin remodelers and histone modifications act as signaling messengers, promoting the recruitment of proteins responsible for different DNA-repair pathways or cell-cycle progression. In addition, epigenetic inactivation of different genes is associated with increased genetic instability and with abnormal cell growth. In this sense, interference with the establishment of histone modifications, changes in chromatin accessibility, or silencing of chromatin are intrinsically associated with tumorigenesis. Thus, epigenetics has become an area of increased interest for the development of therapy and clinical strategies against cancer.

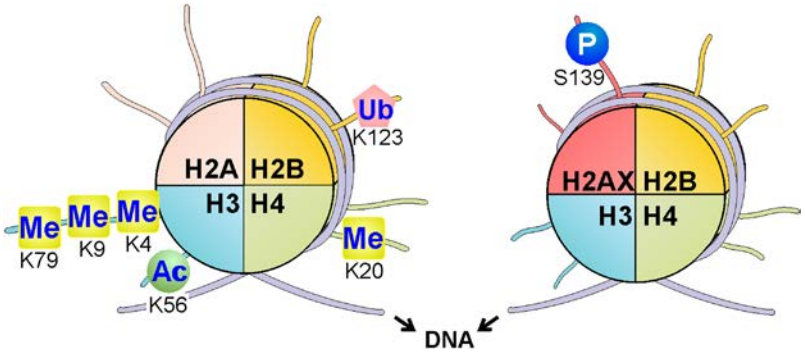


FIGURE 26.7 **Nucleosome representation with the most described modifications in histone tails related to DNA-repair response.**

GLOSSARY

Ataxia telangiectasia (AT) Is a disorder caused by mutations in the ATM gene. This disease is characterized by neurological dysfunctions, cancer predisposition, immunodeficiency, radiosensitivity, infertility and endocrine dysfunctions, among others.

Bloom syndrome (BS) Autosomal recessive genetic disease characterized by the lack of the BLM protein activity. Individuals affected by this disease show morphological abnormalities, photosensitivity, poikiloderma, and high predisposition to cancers.

Cerebro-oculo-facial-skeletal (COFS) syndrome Autosomal recessive genetic disease characterized by an intrinsic inability of the global genome nucleotide excision repair machinery to remove DNA lesions. Individuals affected by this disease present developmental delay, facial abnormalities, microcephaly, cataracts, and microphthalmia, among other symptoms.

Chromothripsis An event of multiple chromosomal rearrangements in a single event.

Clastogenesis A process defined by the loss, addition, or any rearrangement of chromosomes.

Cockayne syndrome (CS) Autosomal recessive genetic disease characterized by an intrinsic inability of the transcription-coupled nucleotide excision repair machinery to remove DNA lesions. Individuals affected by this disease present impaired neurodevelopment photosensitivity progeria, among other symptoms. However, unlike XP, individuals bearing the CS phenotype do not display high predisposition to skin cancers.

Common fragile sites Encountered in the majority of individuals, they are specific locations that are prone to chromosomal rearrangements.

Fanconi anemia Inherited blood disorder caused by intrinsic defects on the interstrand cross-link–repair machinery. Individuals affected by this disease shows severe predisposition to develop myelogenous leukemia and bone-marrow failure, in addition to numerous morphological abnormalities.

Hutchinson–Gilford progeria syndrome (HGTS) Genetic disease caused by a mutation in the LMNA gene. It is a progeroid syndrome characterized by age-related diseases, such as cardiovascular pathologies, skin complications, alopecia, lipodystrophy, joint abnormalities, and osteolysis.

Indels Base insertion and/or deletion.

Lynch syndrome (formerly known as hereditary nonpolyposis colorectal cancer) An inherited condition that makes carriers susceptible to develop certain types of cancer (especially colorectal cancers). They are characterized by mutations in mismatch repair genes.

Merotelic attachments Characterized when a centromere is attached to microtubules coming from both spindle poles.

Mutator phenotype A carcinogenesis hypothesis which postulates that cells acquire cancerous features (see Cancer Hallmarks) due to defective genes responsible for the maintenance of the genome stability.

Philadelphia chromosome A translocation of chromosomes 9 and 22, resulting in a shorter 22 chromosome and a BCR–ABL gene fusion. This phenomenon is present in some hematological malignancies such as acute lymphoblastic leukemia and chronic myelogenous leukemia.

Rothmund–Thomson syndrome (RTS) Autosomal recessive genetic disease characterized by the lack of the RECQL4 protein activity. Individuals affected by this disease show epidermis-related tissue malformation, morphological abnormalities, and cancer predisposition.

Trichothiodystrophy Autosomal recessive genetic disease characterized by an intrinsic inability of the global-genome nucleotide excision repair machinery to remove DNA lesions. Individuals affected by this disease present neurological impairments, brittle hair, and short stature, but do not show photosensitivity.

Werner syndrome (WS) A rare progeroid autosomal recessive disease defined by mutations in WRN gene. Clinical manifestations are aging-like symptoms such as diabetes mellitus type-2, osteoporosis, and cataracts, among others. The affected individuals are also more susceptible to develop cancers (especially sarcomas) and cardiovascular diseases.

Xeroderma pigmentosum (XP) Autosomal recessive genetic disease characterized by an intrinsic inability of the global-genome nucleotide excision repair machinery to remove DNA lesions. Individuals affected by this disease present with high predisposition of skin cancers, skin hyperpigmentation, and, in some cases, can develop neurological impairments, progeria, and cataracts.

LIST OF ACRONYMS AND ABBREVIATIONS

53BP1 P53-binding protein
6,4-PP 6,4-Photoproducts
ALT Alternative lengthening of telomeres
AT Ataxia telangiectasia
ATM Ataxia telangiectasia mutated
ATR Rad3-related protein
BER Base excision repair
BGS Baler–Gerold syndrome
BLM/RECQL3 Bloom syndrome, RecQ helicase-like
BRCA1 Breast cancer gene 1
BS Bloom syndrome
CHK2 Checkpoint kinase 2
CIN Chromosome instability
CPD Cyclobutane pyrimidine dimmers
COFS Cerebro-oculo-facial-skeletal syndrome
CSA Cockayne syndrome WD repeat protein CSA
CSB Cockayne syndrome protein CSB

CS Cockayne syndrome
CUL4A Cullin 4A
DDB1 and DDB2 Damage-specific DNA-binding protein 1 and 2
DNA-PK DNA-dependent protein kinase
DNA Lig DNA ligase IV
DNMT1 DNA methyltransferase
DNA Pol DNA polymerase (delta)
DSB Double-strand break
E2F E2F transcription factor
ERCC1 Excision repair cross-complementation group 1
FA Fanconi anemia
FANC family Fanconi anemia complementation group (composed by many proteins)
FEN1 Flap structure-specific endonuclease 1
GCN5 Histone acetyltransferase GCN5
GG-NER Global genome-nucleotide excision repair
H2AFX H2A histone family, member X
H2BK123 Lysine 123 of histone H2B
H3K4 Lysine 4 of histone H3
H3K9 Lysine 20 of histone H3
H3K79 Lysine 79 of histone H3
H4K20 Lysine 20 of histone H4
HAT Histone acetyltransferase
HDAC1 and 2 Histones deacetylase 1 and 2
HDM Histone demethylase
HGPS Hutchinson–Gilford progeria syndrome
HMT Histone methyltransferase
HP1 Heterochromatin protein 1
HR Homologous recombination
HR23B/RAD23B XP-C repair complementing protein
ICLR Interstrand cross-link repair
ING2 Inhibitor of growth family, member 2
KAP1 KRAB-associated protein 1
KU70 and KU80 X-ray repair cross-complementing protein 6 and 5
LP-BER Long-patch base-excision repair
LMNA Lamin
MDC1 Mediator of DNA-damage checkpoint
MMR Mismatch repair
MRE11 Meiotic recombination 11 homolog 1
MSI Microsatellite instability
NBS1 Nijmegen breakage syndrome 1
NEIL1 Nei endonuclease VIII-like 1 (*Escherichia coli*)
NER Nucleotide excision repair
NHEJ Nonhomologous end joining
NTH1 Nth endonuclease III-like 1
P53 Tumor protein p53
P300 E1A-binding protein P300
PCNA Proliferating cell nuclear antigen
PIKK Phosphoinositide 3-kinase(PI3K)-related protein kinase family
POT1 Protection of telomere 1
PR-Set7 Lysine N-methyltransferase
RAD51 RAD51 recombinase
RAD52 RAD52 homolog (*Saccharomyces cerevisiae*)
RAD53 Serine/threonine-protein kinase RAD53
RAPADILINO *RA*dial hypoplasia, *P*atella hypoplasia and cleft or *A*rched palate, *DI*arrhea and dislocated joints, *LI*ttle size and limb malformation, slender *N*ose and n*O*rmal intelligence syndrome
RB Retinoblastoma tumor suppressor protein
RCF Replication Factor C
RECQL4 RecQ protein-like 4
RMI1/RMI2 RecQ-mediated genome instability 1 and 2
RNA Pol RNA polymerase II

Rnf20/Rnf40 Ring finger protein 20/40 complex
RNF8 Ring finger protein 8
ROC1 Regulator of cullins 1
ROS Reactive oxygen species
RPA Replication protein A
RTS Rothmund–Thomson syndrome
SIN3A SIN3 transcription regulator family member A
SIN3B SIN3 transcription regulator family member B
ssDNA Single-stranded DNA
TC-NER Transcription-coupled-nucleotide excision repair
TFIIH Transcription factor II human
TRF1 and 2 Telomeric repeat binding factor 1 and 2
TTD Trichothiodystrophy
UV Ultraviolet light
WRN/RECQL2 Werner syndrome, RecQ helicase like
WS Werner syndrome
XP Xeroderma pigmentosum
XPA to XPG Xeroderma pigmentosum complementation group A to G
XAB2 XPA-binding protein 2
XRCC4 X-ray repair cross-complementing protein 4

REFERENCES

[1] Gillet LC, Schärer OD. Molecular mechanisms of mammalian global genome nucleotide excision repair. Chem Rev 2006;106(2):253–76.
[2] Menck CF, Munford V. DNA repair diseases: what do they tell us about cancer and aging? Genet Mol Biol 2014;37:220–33.
[3] Palomera-Sanchez Z, Zurita M. Open, repair and close again: chromatin dynamics and the response to UV-induced DNA damage. DNA Repair (Amst) 2011;10(2):119–25.
[4] Kim H, D'Andrea AD. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. Genes Dev 2012;26(13):1393–408.
[5] Feltes BC, Bonatto D. Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications. Mutat Res Rev Mutat Res 2015;763:306–20.
[6] Schärer OD. Nucleotide excision repair in eukaryotes. Cold Spring Harb Perspect Biol 2013;5(10):a012609.
[7] Egly JM, Coin F. A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor. DNA Repair (Amst) 2011;10(7):714–21.
[8] Lehmann AR, et al. Xeroderma pigmentosum. Orphanet J Rare Dis 2011;6:70.
[9] Bergmann E, Egly JM. Trichothiodystrophy, a transcription syndrome. Trends Genet 2011;17(5):279–86.
[10] Jaarsma D, et al. Cockayne syndrome pathogenesis: lessons from mouse models. Mech Ageing Dev 2013;134(5–6):180–95.
[11] Shell SM, Zou Y. Other proteins interacting with XP proteins. Adv Exp Med Biol 2008;637:103–12.
[12] Clauson C, et al. Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. Cold Spring Harb Perspect Biol 2013;5(10):a012732.
[13] Longerich S, et al. Stress and DNA repair biology of the Fanconi anemia pathway. Blood 2014;124(18):2812–9.
[14] Zhang J, Walter JC. Mechanism and regulation of incisions during DNA interstrand cross-link repair. DNA Repair (Amst) 2014;19:135–42.
[15] Mouw KW, D'Andrea AD. Crosstalk between the nucleotide excision repair and Fanconi anemia/BRCA pathways. DNA Repair (Amst) 2014;19:130–4.
[16] Su X, Huang J. The Fanconi anemia pathway and DNA interstrand cross-link repair. Protein Cell 2011;(9):704–11.
[17] Kashiyama K, et al. Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. Am J Hum Genet 2013;92(5):807–19.
[18] Singh DK, et al. RecQ helicases in DNA double strand break repair and telomere maintenance. Mutat Res 2012;736(1–2):15–24.
[19] Bohr VA. Rising from the RECQ-age: the role of human RECQ helicases in genome maintenance. Trends Biochem Sci 2008;33(12):609–20.
[20] Larizza L, et al. Rothmund-Thomson syndrome. Orphanet J Rare Dis 2010;5:1–16.
[21] Croteau DL, et al. RECQL4 in genomic instability and aging. Trends Genet 2012;28(12):624–31.
[22] Fan W, Luo J. RecQ4 facilitates UV light-induced DNA damage repair through interaction with nucleotide excision repair factor xeroderma pigmentosum group A (XPA). J Biol Chem 2008;283(43):29037–44.
[23] Bloom D. Congenital telangiectatic erythema resembling lupus erythematosus in dwarfs; probably a syndrome entity. AMA Am J Dis Child 1954;88(6):754–8.
[24] Arora H, et al. Bloom syndrome. Int J Dermatol 2014;53(7):798–802.
[25] Cheok CF, et al. Roles of the Bloom's syndrome helicase in the maintenance of genome stability. Biochem Soc Trans 2005;33(Pt 6):1456–9.
[26] Hoadley KA, et al. Defining the molecular interface that connects the Fanconi anemia protein FANCM to the Bloom syndrome dissolvasome. Proc Natl Acad Sci USA 2012;109(12):4437–42.
[27] Ozgenc A, Loeb LA. Current advances in unraveling the function of the Werner syndrome protein. Mutat Res Fund Mol M 2005;577(1–2 Spec. Iss.):237–51.

[28] Opresko PL, et al. Werner syndrome and the function of the Werner protein; what they can teach us about the molecular aging process. Carcinogenesis 2003;24(5):791–802.
[29] Rossi ML, et al. Roles of Werner syndrome protein in protection of genome integrity. DNA Repair 2010;9(3):331–44.
[30] Lavin MF. Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer. Nat Rev Mol Cell Biol 2008;9(10):759–69.
[31] McKinnon PJ. ATM and the molecular pathogenesis of ataxia telangiectasia. Annu Rev Pathol 2012;7:303–21.
[32] Shiloh Y. ATM and related protein kinases: safeguarding genome integrity. Nat Rev Cancer 2003;3(3):155–68.
[33] Ghosh S, Zhou Z. Genetics of aging, progeria and lamin disorders. Curr Opin Genet Dev 2014;26:41–6.
[34] Kudlow BA, et al. Werner and Hutchinson-Gilford progeria syndromes: mechanistic basis of human progeroid diseases. Nat Rev Mol Cell Biol 2007;8(5):394–404.
[35] Parreno J, Cruz AV. Accelerated aging in patients with Hutchinson-Gilford progeria syndrome: clinical signs, molecular causes, treatments, and insights into the aging process. UBC Med J 2011;3:8–12.
[36] Gonzalo S, Kreienkamp R. DNA repair defects and genome instability in Hutchinson-Gilford progeria syndrome. Curr Opin Cell Biol 2015;34:75–83.
[37] Liu Y, et al. Involvement of xeroderma pigmentosum group A (XPA) in progeria arising from defective maturation of prelamin A. FASEB J 2008;22(2):603–11.
[38] D'Errico M, et al. The role of CSA and CSB protein in the oxidative stress response. Mech Ageing Dev 2013;134(5–6):261–9.
[39] Vélez-Cruz R, Egly JM. Cockayne syndrome group B (CSB) protein: at the crossroads of transcriptional networks. Mech Ageing Dev 2013;134(5–6):234–42.
[40] Negrini S, et al. Genomic instability – an evolving hallmark of cancer. Nat Rev Mol Cell Bio 2010;11(3):220–8.
[41] Lengauer C, et al. Genetic instabilities in human cancers. Nature 1998;396(6712):643–9.
[42] Abbas T, et al. Genomic instability in cancer development. Cold Spring Harb Perspect Biol 2013;5:1–18.
[43] Bester AC, et al. Nucleotide deficiency promotes genomic instability in early stages of cancer development. Cell 2011;145(3):435–46.
[44] Kloor M, et al. Clinical significance of microsatellite instability in colorectal cancer. Langenbeck Arch Surg 2014;399(1):23–31.
[45] Forment JV, et al. Chromothripsis and cancer: causes and consequences of chromosome shattering. Nat Rev Cancer 2012;12(10):663–70.
[46] Desmaze C, et al. Telomere-driven genomic instability in cancer cells. Cancer Lett 2003;194(2):173–82.
[47] Günes C, Rudolph KL. The role of telomeres in stem cells and cancer. Cell 2013;152(3):390–3.
[48] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144(5):646–74.
[49] Murnane JP. Telomere dysfunction and chromosome instability. Mutat Res-Fund Mol M 2012;730(1–2):28–36.
[50] Ricke RM, et al. Whole chromosome instability and cancer: a complex relationship. Trends Genet 2008;24(9):457–66.
[51] Nam HJ, et al. Centrosome dynamics as a source of chromosomal instability. Trends Cell Biol 2015;25(2):65–73.
[52] Loeb LA. A mutator phenotype in cancer. Cancer Res 2001;61(8):3230–9.
[53] Halazonetis TD, et al. An oncogene-induced DNA damage model for cancer development. Science 2008;319(5868):1352–5.
[54] Mazouzi A, et al. DNA replication stress: causes, resolution and disease. Exp Cell Res 2014;329(1):85–93.
[55] Ozeri-Galai E, et al. The complex basis underlying common fragile site instability in cancer. Trends Genet 2012;28(6):295–302.
[56] Kollareddy M, et al. Aurora kinase inhibitors: progress towards the clinic. Invest New Drug 2012;30(6):2411–32.
[57] Curtin NJ. DNA repair dysregulation from cancer driver to therapeutic target. Nat Rev Cancer 2012;12(12):801–17.
[58] Vitale I, et al. Mitotic catastrophe: a mechanism for avoiding genomic instability. Nat Rev Mol Cell Bio 2011;12(6):385–92.
[59] Geigl JB, et al. Defining "chromosomal instability". Trends Genet 2008;24(2):64–9.
[60] Nowak MA, et al. The role of chromosomal instability in tumor initiation. P Natl Acad Sci USA 2002;99(25):16226–31.
[61] Sharma S, et al. Epigenetics in cancer. Carcinogenesis 2010;31(1):27–36.
[62] Li S. Implication of posttranslational histone modifications in nucleotide excision repair. Int J Mol Sci 2012;13(10):12461–86.
[63] Ma Y, et al. How the cell cycle impacts chromatin architecture and influences cell fate. Front Genet 2015;6:19.
[64] Lahtz C, Pfeifer GP. Epigenetic changes of DNA repair genes in cancer. J Mol Cell Biol 2011;3(1):51–8.
[65] Johnson DG, Dent SY. Chromatin: receiver and quarterback for cellular signals. Cell 2013;152(4):685–9.
[66] McLaughlin F, et al. The cell cycle, chromatin and cancer: mechanism-based therapeutics come of age. Drug Discov Today 2003;8(17):793–802.
[67] Gonzalo S, Blasco MA. Role of Rb family in the epigenetic definition of chromatin. Cell Cycle 2005;4(6):752–5.
[68] Corney DC, Coller HA. On form and function: does chromatin packing regulate the cell cycle? Physiol Genomics 2014;46(6):191–4.
[69] Jorgensen S, et al. Histone H4 Lysine 20 methylation: key player in epigenetic regulation of genomic integrity. Nucleic Acids Res 2013;41(5):2797–806.
[70] Mastrangelo D, et al. Retinoblastoma as an epigenetic disease: a proposal. J Cancer Ther 2011;2(3):362–71.
[71] Cao J, Yan Q. Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. Front Oncol 2012;2:26.
[72] Henrique R, et al. The epigenetics of renal cell tumors: from biology to biomarkers. Front Genet 2012;3:94.
[73] House NCM, et al. Chromatin modifications and DNA repair: beyond double-strand breaks. Front Genet 2014;5:296.
[74] Evertts AG, et al. H4K20 methylation regulates quiescence and chromatin compaction. Mol Biol Cell 2013;24(19):3025–37.
[75] Jafarnejad SM, Li G. Regulation of p53 by ING family members in suppression of tumor initiation and progression. Cancer Metastasis Rev 2012;31(1–2):55–73.

APÊNDICE II

Impact of *xpa-1* mutation in *Caenorhabditis elegans* on cell cycle and reproductive development

Artigo submetido para a revista *Functional and Integrative Genomics*

Accessing the impact of *xpa-1* mutation in *Caenorhabditis elegans* through systems biology and transcriptomic approaches

Itamar José Guimarães Nunes[1] and Bruno César Feltes*[1]

[1]Biotechnology Center of the Federal University of Rio Grande do Sul, Department of Molecular Biology and Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, RS – Brazil

Short title: Impact of *xpa-1* mutation in *C. elegans*

* **To whom correspondence should be sent:**

Bruno César Feltes

Centro de Biotecnologia da UFRGS - Sala 219

Departamento de Biologia Molecular e Biotecnologia

Universidade Federal do Rio Grande do Sul - UFRGS

Avenida Bento Gonçalves 9500 - Prédio 43421

Caixa Postal 15005

Porto Alegre – Rio Grande do Sul

BRAZIL

91509-900

Phone: (+55 51) 3308-6080

Fax: (+55 51) 3308-7309

Contract/grant sponsor: CAPES

Abstract

Individuals carrying mutations in *Xeroderma pigmentosum* type A (XPA) gene manifest higher propensity to skin cancer and neurodegenerative disorders. One possible approach to study XPA deficiency is using *Caenorhabditis elegans*, which is broadly applied in studies with DNA repair, including mutations in *xpa-1*, the homolog to human XPA. In this sense, we conducted a rigorous statistically filtered transcriptomic analysis in order to evaluate expression differences in *xpa-1* mutants, and used multiple systems biology approaches to study the impact of XPA deficiency and the molecular mechanisms underlying XPA disease. Our results indicated that genes participating in mitotic and meiotic processes, including cyclins, MCM family members and centrosome assembly genes, are overexpressed, ratifying that *xpa-1* deficiency is associated to increased cell proliferation. Higher expression of PUF genes, specially *puf-8*, *fbf-1* and *fbf-2*, can also contribute to higher rates of oocyte differentiation and survival, but on the other hand, are suggested to play an inhibitory role in *C. elegans'* developmental process. PUF overexpression can also favor oocyte differentiation and suppress spermatogenesis, and lower spermatocyte production is observed in mice and humans with impaired XPA. In addition, oocyte apoptosis can be attenuated by the enhanced expression of *mrg-1*, *cgh-1*, and *mcm-6*, pointing that fertility in both human and *C. elegans* can be impacted in a similar manner. Finally, we discussed other human homologs already evidenced in different tumor types and phenotypic outcomes in XPA patients and suggested a new molecular model to explain the similarity between the worm *xpa-1* deficiency and the human tumorigenesis.

Key-words: Xeroderma pigmentosum; XPA; *xpa-1*; Systems biology; Development; Cell replication; Cancer.

1. Introduction

DNA molecule lesions can occur due to environmental and intrinsic factors, such as UV radiation exposure, oxidative stress and genotoxic substances, being associated to aging, cancer and other pathologies [1,2]. The resulting damage can be recognized and corrected by DNA repair mechanisms, preserving genome integrity [2]. In this sense, the nucleotide excision repair (NER) pathway acts on the bulky single-stranded DNA lesions (e.g., cyclobuthane pyrimidine dimers and 6,4-photoproducts), removing the lesion and repairing the DNA molecule [3].

NER regards on xeroderma pigmentosum protein complex recruitment, that ranges from XPA to XPG, that performs the recognition, the double-strand unwinding and the damaged site excision [1,4]. The protein complex loss of function, ensued from a mutation in one of the XP genes, culminates in NER failure to repair UV damage, leading to the Xeroderma Pigmentosum (XP) disease [5]. This pathology is related to a higher propensity of skin cancer, and patients carrying mutations in the XPA gene can also present premature aging and neurodegenerative disorders [6]. However, the role of XPA mutation in organism development still isn't clear, and the molecular patterns underlying disease progression are poorly understood.

In this sense, the nematode *Caenorhabditis elegans* has been employed for studies involving XPA mutations, in addition to be widely used in aging and DNA repair studies [7,8]. Different transcriptomic datasets for *xpa-1* (homolog of human XPA) mutants in *C. elegans* are also publically available [9,10]. Thus, through DNA microarray transcriptome analyses and multiple systems biology approaches, we explored the *C. elegans xpa-1* expression loss at molecular level using multiplexes networks, topological, and enrichment analyses to predict major biological pathways related to changes in the nematodes carrying the *xpa-1* mutations. The data gathered from this work was employed to propose a novel and comprehensive molecular model that suggests a connection between XPA, cell cycle, and reproductive development, concomitantly pointing associations with tumorigenesis in humans.

2. Materials and methods

2.1. Microarray gene expression analysis

Two different public available transcriptomic datasets were gathered from the Gene Expression Omnibus (GEO) database [http://www.ncbi.nlm.nih.gov/gds] to establish the differential expression in *C. elegans xpa-1* mutants. The datasets are identified as GSE16405, as described by Fensgård et al. [9], and GSE39252, as described by Arczewska *et al*. [10]. All samples were separated between two groups: experimental group ("EXP"), which is represented by *C. elegans xpa-1* mutant strain, and the control group ("CTRL"), that corresponds to the N2 wild-type (WT) strain.

Statistical analyses on the transcriptomic data were performed using the R platform [https://www.r-project.org] together with the following packages and their respective parameters: (i) GEOquery for matrix importing and data parsing [11]; (ii) *expresso* function from the affy package [12] with Affymetrix Microarray Suite algorithm (*MAS*) for background correction, *quantiles* for normalization, perfect match-only option (*pmonly*) for perfect match correction and average difference method (*avgdiff*) for summary expression value computation; (iii) ArrayQualityMetrics for array quality analyses [13]; and (iv) limma for differential expression analysis [14]. The results from the quality analysis validating the reliability for the expression values of the two datasets, which showed no outliers, are described in **Fig. 1.**

A normalization value ($\log_2$) for gene expression fold change was calculated for both "EXP" and "CTRL" groups, where "EXPvalue" and "CTRLvalue" were defined for the gene expression values in the experiment and control sample groups, respectively. The computation of the normalized values was performed by *limma* according to the equation (1) for EXPvalue greater than CTRLvalue:

$$log_2 \frac{EXPvalue}{CTRLvalue} \quad (1)$$

If CTRLvalue is greater than EXPvalue, the equation (2) was applied:

$$-\left(log_2 \frac{CTRLvalue}{EXPvalue}\right) \quad (2)$$

The corrected *p*-value also was calculated applying Benjamini & Hochberg False Discovery Rate (FDR) algorithm, provided inside the limma package [14,15]. Genes with corrected *p*-value less than 0.05 and $\log_2$ of fold-change (Log2FC) greater than 1.00 or less than -1.00 were selected for further analyses.

2.2. Interatomic data prospection and network design

In order to obtain the networks related to *xpa-1* mutation for *C. elegans*, the metasearch engine STRING 10 [http://string.embl.de] was applied using the differentially expressed genes (DEGs) from transcriptomic analysis as an initial input (**S-material 1**) [16]. The selected parameters in STRING were: all prediction methods active, except for the "text mining" and "gene fusion" options. The networks derived from the obtained DEGs were expanded until maximum expansion of additional nodes for until saturation using medium confidence score (0.400) and no more than 10 interactions per node. The resulting networks were analyzed by the software Cytoscape, using versions 2.8.3 and 3.2.0 [17,18]. The network composed by the overlap between DEGs from the GSE16405 dataset and from the GSE39252 microarray data, named "Expression Network" (EN), displays the predicted connections between each DEG (**Fig. 2A**). The maximum expansion of additional nodes in STRING, using the DEGs that compose the EN as input, resulted in a single large network, named "United Network" (UN) (**Fig. 2B**).

2.3. Centrality analysis

Through centrality analysis, the topological importance of all nodes within the network can be evaluated to predict topologically relevant DEGs that could play crucial role in bioprocesses affected by the *xpa-1* mutation. In this sense, the Cytoscape plug-in CentiScaPe 1.2 [19] was employed for the identification of two centralities within the network: node degree and betweenness. Node degree corresponds to the number of neighbor nodes connected to a specific node [20]. Thus, if the number of connections is above the average network node degree value, the node is characterized as a "hub" [20]. Moreover, the betweenness parameter calculates the number of shortest pathways that passes through each node [21]. A node with high betweenness degree compared to the average network betweenness degree value is called "bottleneck", and usually

connects two or more groups of highly connected nodes (clusters) in an interatomic network. Bottlenecks are characterized by constraining the information flow within a network, and their removal results in error propagation and/or cluster-to-cluster communication failure [20,21]. In topological terms, nodes with both centrality properties are termed "hub-bottlenecks" (HB) and are the most relevant within a network [20]. A centrality sub-network for *C. elegans,* named "Centralities Network" (CN), was generated containing the main HB derived from the UN network (**Fig. 2C**).

2.4. Clustering and modularity analysis

To identify the major clusters inside the UN network, the Cytoscape plug-in Molecular Complex Detection (MCODE) was applied [22]. The cluster analysis allowed the recognition high degree connectivity regions in the UN network, attributing a clustering coefficient (C*i*) named "cliquishness" for each cluster [22]. In biological terms, clustering can be used to identify groups of proteins that cooperate in specific biological processes. Hence, clusters that include DEGs could point processes more likely affected by *xpa-1* deficiency. The parameters used in MCODE were as follows: "include loops" option activated; degree cutoff 2; single connected nodes removal from clusters (haircut option activated); "fluff" option enabled; node score cutoff 0.2; K-Core 2; maximum depth 100. The clusters identified by MCODE were determined as "MClusters".

The network modules were also identified using the default Modularity algorithm from Gephi 0.9 [23], as proposed by Blodel *et al.* [24,25]. This algorithm is implemented to subdivide the overall network in different modules that do not overlap themselves, in contrast to MCODE clustering analysis. The networks segments calculated by Gephi's Modularity analysis can vary from: (i) the densest module, which include nodes that are almost exclusively relevant within the same module, but not for rest of the network; (ii) to the most dispersed module, whose nodes have more topological relevance to nodes from other modules than the same module [25]. Hence, this analysis allowed distinguishing the relevance of each module and their impact within the network. Thus, the ranking clustering method allows us to predict which group of DEGs could have a higher impact on the global network topology or within a cluster. In this sense, the parameter that measures this property, termed "modularity

class ranking", organizes the network segments on hierarchically-divided groups where the first group (e.g., Class 1 module) is the most isolated and densely connected module, whereas the Class 19 module, the last one, is broadly interconnected to other modules, but is the less connected among itself. On our analysis, the selected resolution in the Modularity algorithm was 1.0, the software's default parameter. The modules found through Gephi's Modularity algorithm were named as "GModules".

2.5. Gene ontology analysis

The UN, CN, as well for the MCODE clusters were analyzed regarding their main biological processes, using the Cytoscape plug-in Biological Network Gene Ontology (BiNGO) 2.44 [26]. The gene ontology annotations for *C. elegans* were obtained from the Gene Ontology Consortium website [http://geneontology.org/page/download-annotations] [27]. For every network, the *p-value* functional enrichment value was calculated using the hypergeometric distribution test, and the multiple test correction was evaluated by the Benjamini & Hochberg FDR correction algorithm, which is fully implemented inside BiNGO, at a significance level of *p* less than 0.05 [15]. The most statistically relevant processes, according to these parameters, were listed in **S-material 2** and were taken into account for the design of the molecular model.

The detailed description of the gene ontologies annotated for *C. elegans* were collected from Wormbase WS251 database [http://www.wormbase.org]. The workflow employed in this work can be found in **Fig. S1** in the **S-material 3**.

3. Results

3.1. XPA network design, clustering and centrality analysis

The DEGs obtained from the transcriptomic analysis of both GSE16405 and GSE39252 were overlapped, resulting in a total of 1,059 DEGs (633 overexpressed and 426 underexpressed) identified from the two *xpa-1* mutants datasets. Only the DEGs obtained from this overlap were taken into consideration for the systems biology analyses. The complete list is available in **S-material 1**. It was observed 1057 DEGs (632 overexpressed and 425 underexpressed) from GSE16405 and six DEGs (2 overexpressed and 4 underexpressed) from GSE39252, whereas three underexpressed

genes (F57F4.4, *sym-2* and *xpa-1*) and one overexpressed gene (*gfi-1*) were identified in both datasets. Since the comparison was performed using only two samples per group in GSE16405 and five samples per group in GSE39252, as mentioned above in our methodology, the number of DEGs obtained by GSE39252 was significantly smaller than GSE16405. However, the loss of expression of *xpa-1* was observed in the comparisons from both datasets, validating the expected loss of *xpa-1* expression.

The DEGs from both GEO datasets were overlapped, and this union was used as input in STRING to generate a preliminary network, the EN (**Fig. 2A**). Multiple networks were generated and expanded using the genes from EN network in the STRING database (**Fig. 2** and **Fig. S1** in **S-material 3**). The obtained networks were merged in Cytoscape by using the Advanced Network Merge tool, resulting in a single large network, the UN. In this context, UN corresponds to the union of all generated networks in STRING, and is composed by a total of 5,101 nodes and 242,827 edges (**Fig. 2B**).

The three topological analyses allowed the identification of relevant node groups in the UN. Firstly, a number of clusters were selected by using MCODE (MClusters), but only two clusters were chosen for further discussion, named MCluster 1 and MCluster 2 (**Figs. S2** and **S3** in **S-material 3**). MCluster 1 has a C*i* score of 48,413 and includes 1,640 nodes, and MCluster 2, showed C*i* of 16,724 and contains 1,564 nodes. Secondly, through Gephi's modularity analysis, it was identified a total of 19 Gephi Modules (GModules) ordered hierarchically, which can be visualized in **Fig. 3**. The GModule 1 is the most clustered UN segment, according to the Gephi algorithm, thus, the most isolated from the rest of network, whereas GModule 19 is the opposite. Thirdly, the results derived from the centrality analysis were used to identify HB nodes obtained from the UN network, generating a new sub-network named "Centralities Network" (CN) (**Fig. 2C**). This network is composed by 1,592 nodes and 187,878 edges, and includes 156 HB DEGs (141 overexpressed and 15 underexpressed) (**S-material 1**). The inclusion of DEGs in the CN and in the clusters is listed in **S-material 1**.

3.2. Identification of bioprocesses and their association to the HBs, clusters and DEGs

Based on the BiNGO results, the ontologies were organized in the following categories: (i) Aging; (ii) Cell cycle and replication; (iii) Embryonic development; (iv) Gene expression; (v) General bioprocesses; (vi) Larval development; and (vii) Reproduction. The bioprocesses included in these categories were listed in **S-material 2**. Moreover, we selected the HB-DEGs (a total of 95 nodes) that are included in more than one bioprocess group in order to further represent correlations between the expression changes in *xpa-1* mutants and bioprocesses from different categories, and listed these genes in **Table 1**.

On MCluster 1, 213 nodes were annotated in the "cell cycle" gene ontology, where 169 (79.4%) are DEGs (156 overexpressed and 13 underexpressed) (**S-material 1** and **2**). Among the overexpressed genes, it was identified three cyclins (*cyb-1*, *cyb-3* and *cye-1*), four ATPase's (*atp-2*, H28O16.1, *vha-5* and *vha-12*), five members from MCM family (*mcm-2* to *4*, *mcm-6* and *mcm-7*) and six members from PUF family (*puf-3*, *puf-6* to *8*, *fbf-1* and *fbf-2*), whose functions will be discussed below (**Table 1**). Moreover, 24 upregulated genes annotated for "cell cycle" are also HB, as described in **Table 1**.

On MCluster 2, a total of 162 nodes were annotated in the cell cycle biological process (**S-material 1 and 2**), which 60 (37%) are HB and 36 (22.2%) are DEGs (34 overexpressed and 2 underexpressed) (**Table 1**; **S-material 1** and **2**). This DEG group includes two ATPase's, two PUF genes and one gene from MCM family (**S-material 1**). A total of 14 (8.6%) overexpressed genes associated to cell cycle in MCluster 2 are also HB (**Table 1; S-material 1**). Among the HB annotated in the cell cycle progression-related bioprocesses within the MCluster 2, there are 5 overexpressed genes annotated for translation, 10 for M phase, two for spermatogenesis, and one for response to unfolded protein (**Table 1; S-material 1**). MCluster 2 also included 5 DEGs from heat shock proteins (HSP) family, where two are overexpressed and 3 are underexpressed (**Table 1; S-material 1**).

Additionally, combining the results from Gephi's modularity analysis with gene ontology analysis, we observed that, from the total of 308 genes annotated for cell cycle biological process in the UN, 162 (52,6%) genes belong to the GModule 2 (the

second most isolated) and 127 (41,2%) are from the GModule 7 (the seventh most isolated) (**S-material 1**). In GModule 2, 24 nodes are DEGs (21 overexpressed and 3 underexpressed) and 27 are HB (**S-material 1**). In GModule 7, it is included 46 DEGs (44 overexpressed and 2 underexpressed) and 99 genes are HB nodes (**S-material 1**).

Taking together transcriptome, Gephi's modularity and MCODE clustering analyses, we found that GModule 2 and MCluster 1 shared 39 overexpressed HB genes and one underexpressed HB gene (F55B11.2). GModule 2 also overlaps 20 overexpressed HB genes with MCluster 2 (**S-material 1**). GModule 7, in turn, overlaps 23 overexpressed HB genes with MCluster 1. GModule 7 also has 35 overexpressed HB genes in common with MCluster 2 (**S-material 1**). The DEGs in common between the two distinct modularity analyzes were taken into consideration in the new proposed molecular model.

4. **Discussion**

4.1. Changes in germ cell proliferation and fertility in C. elegans *and its similarity to humans*

Through transcriptome, topological and gene ontology analyses, it was demonstrated that members from PUF and MCM families could have an impact in the fertility of *xpa-1* mutants. Part of these genes are overlapped in both MClusters and GModule 2 (**S-material 1**), especially *cgh-1*, *cyb-1*, *cyb-3*, *hcp-1*, *mcm-2* to *4*, *mcm-6*, *mcm-7*, *mrg-1*, *puf-3*, *puf-6* to *8*, which are the main candidates for our molecular model. It is expected that GModule 2, due to its high density and isolation within UN, embraces a higher number of proteins acting in more specific biological processes that do not necessarily require nodes from other modules. Since the impact of the differential expression can be more restricted within this module, we took into consideration DEGs included in GModule 2 and annotated in similar gene ontologies.

Three of the overexpressed genes overlapping MCluster 1 and 2, and GModule 2, are involved in cell proliferation and are annotated in ontologies related to reproductive development: *mrg-1*, *cgh-1* and *mcm-6* (**Table 1**). Following this context, *mrg-1* is expressed primarily in oocytes at the initial stage of development (0–6 hours) and it is required for primordial germ cells to initiate mitotic proliferation in *C. elegans*

[28,29]. Curiously, while RNAi silencing of *mrg-1* causes sterility and, in some cases, developmental abnormalities (like vulval protrusion and tail shortening in *C. elegans*), it was demonstrated that this gene overexpression in HeLa cells by transfection causes cell death to almost all cultured cells, likely due to the molecular competition with human homologs, pointing that the worm counterpart acts through similar pathways that in humans [30]. Another DEG, *cgh-1*, is a RNA helicase required for protection from physiological germline apoptosis and gametogenesis [31]. *Cgh-1* silencing also results in non-functional sperm and cell death of almost all developing oocytes through the germline apoptotic pathway [31]. Similarly, another HB-DEG, *mcm-6*, can also be related to oogenesis (**Table 1**), since RNAi silencing of this gene is demonstrated to reduce the number of oocytes [32]. This suggests that *cgh-1*, *mrg-1* and *mcm-6*, which are overexpressed with the loss of *xpa-1* expression, enhance or contribute for oocyte proliferation and survival.

Several members from the Pumilio/FBF (PUF) family of RNA binding genes are also related to cell cycle and reproductive development, especially the targets *puf-6* to *8* included in **Table 1**. Previous studies suggests that two overexpressed PUF genes groups, *puf-3/11* and *puf-5 to 7*, controls gene expression by repressing mRNA translation through common and distinct sets of 3' untranslated regions to ensure proper cell differentiation [33,34]. RNAi silencing of *puf-3* also results in defective internal membrane compartments, less occurrence of apoptosis and reduced oocyte number [32]. The *puf-8* silencing contributes to masculinization of germline [35]. Additionally, there is a synergic role of *puf-8* and the PUF genes (*fbf-1* and *fbf-2*) in the germline cell fate, whereas *puf-8*/*fbf-1* and *fbf-1*/*fbf-2* double mutants fail to switch the germline differentiation from spermatocytes to oocytes, a process that occurs normally in WT hermaphrodites [35–38]. On HeLa cells, siRNA knockdown of the human homologs of *cgh-1 and puf-8*, respectively DDX6 and PUM1, resulted in increased apoptosis through CASP-3 cleavage [39], also supporting the protective anti-apoptotic role of *cgh-1* on oocytes, as discussed above. Although the role of *puf-8* in *C. elegans* germ cell differentiation is well established [40], these observations indicate that PUF genes, when overexpressed in *xpa-1*-deficient worms, positively contribute for oocyte differentiation and survival, while causing defects in spermatocyte development.

Curiously, two other HB-DEGs are associated to reproductive development: *cct-4* (identified by sequence K01C8.10) and *fib-1* (**Table 1; S-material 1**). *Cct-4* sub-expression is found to be related to *gst-4* (identified by sequence K08F4.7) activation induction upon K01C8.10 RNAi silencing [41]. The protein codified by K08F4.7 is a putative glutathione-requiring prostaglandin D synthase whose silencing in L4 stage hermaphrodites can result in premature depletion of self-derived sperm [42]. Another gene, *fib-1*, is involved in gonad development (**Table 1**). A study using *fib-1* RNAi demonstrated that *fib-1* silencing can result in degenerated gonads in worms [32]. This observation suggests that, in terms of *xpa-1* mutation, the overexpression of *cct-4* could interfere in *gst-4* activity, thereby changing sperm production, and the overexpression of *fib-1* can play a role in gonad development.

There are conflicting results about the presence of typical phenotypes from XPA patients and NER-deficient mice in *C. elegans xpa-1* mutants, such as lifespan shortening [43]. Nonetheless, there are clear evidences of similar consequences within these organisms regarding reproductive and cell cycle processes, the latter being discussed in the next section. It was implied that spermatogenesis is impaired in Xpa-deficient mice [44] and on humans carrying the XPA(-4) G/A allele polymorphisms [45], which results in lower XPA expression. It was also evidenced a significantly decreased expression of NER genes, including XPA, in human tumor testis cell lines [46–48]. These results suggest that: (i) the mechanisms favoring oogenesis can be stimulated, whereas the spermatogenesis process from germline cells could be impaired in response to *xpa-1* mutation in worms; and (ii) part of these mechanisms could be present in mammals, where the fertility, particularly spermatogenesis, is negatively impacted. We describe this proposal in our new molecular model depicted in **Fig. 4**.

4.2. Cell replication and development abnormalities associated to xpa-1

Studies related to mutations of *xpa-1* in *C. elegans* have described a number of phenotypes, including UV hypersensitivity, high spontaneous mutation rate, increased size, defects in molting and D-neuron development (in UV-irradiated worms), and vulval defects, likely due background mutations [49,50]. However, it is not clear how this process occurs at the molecular level.

In this sense, two overexpressed HB genes in our analysis are the cyclins *cyb-1* and *cyb-3,* are directly related to cell cycle progression. These cyclins are also related to oogenesis, genitalia development and cytoskeleton organization (**Table 1**). Both *cyb-1* and *cyb-3* demonstrated similar spatial-temporal expression in *C. elegans* in addition to their interaction with cyclin-dependent kinases for cell cycle triggering. However, they can have distinct roles in mitotic chromosome condensation, meiotic division of oocytes and embryonic development [51–53]. Similarly, the two overexpressed HBs *plk-1* and *plk-2* (**S-material 1**) are homologs for the human PLK2 and PLK1, respectively, whereas the latter is a protein that targets cyclin B1 (homolog of worm *cyb-1*) during the prophase stage in vertebrates [54]. The *plk-1* protein plays a similar role to its human homolog PLK2, whereas the worm *plk-1* determines the size and density of centrosome and the human PLK2 induces the centrosome amplification and duplication during mitosis [55–58]. These observations suggest that, in worms, cell cycle progression is stimulated by cyclins in oocytes and in somatic cells in response to the loss of *xpa-1* expression, possibly involving a synergic activity of *plk-1* or *plk-2* (**Fig. 4**).

Other overexpressed genes involved in cell cycle include Kinetochore NuLl (*knl-1*) and two HoloCentric chromosome binding Proteins, *hcp-1* and *hcp-2*. The protein coded by *knl-1* plays a role in cytoskeleton organization (**Table 1**), required for proper kinetochore assembly to connect the mitotic chromosomes to spindle microtubules during cell division [59,60]. The centromere proteins *hcp-1* and *hcp-2*, both related to cytoskeleton organization activity, are necessary for proper chromosome alignment and segregation by targeting CLASP, a microtubule-specific protein [61–63]. In addition, the proper kinetochore assembly also is conducted by the recruitment of *hcp-1*, *hcp-2* and other proteins through a *knl-1*-dependent pathway during metaphase in order to initiate the central spindle assembly during anaphase [64]. Interestingly, *knl-1* and *hcp-2* genes are homologs to human Girdin and MYH10, respectively, where the first one plays an important role in the cell division and migration by interacting with microtubules and actin filaments in the centromere and the latter is necessary for cytokinesis process during male meiotic cell divisions [65,66]. These interpretations indicate that the upregulated *knl-1*, *hcp-1* and *hcp-2* genes play a role in the cell replication machinery through chromatin remodeling during *xpa-1* deficiency (**Fig. 4**).

Furthermore, it was expected that changes in cell replication would affect organism development. In this sense, our analyses identified three overexpressed PUF genes *puf-8*, *fbf-1* and *fbf-2* (**S-material 1**), implied as vulval development repressors through translation [67], and loss of function by genetic mutations could follow multi-vulva phenotypes [67] (**Fig. 4**). This observation proposes that PUF genes could affect worm reproduction through negative regulation of vulval development. Curiously, another overexpressed gene that could be involved in reproductive behavior defects is *atp-2* (**Table 1**). The protein codified by *atp-2* gene can play a inhibitory role in reproductive activity by physically interacting to *lov-1* and causing a defect in the male mating behavior through male-specific sensory neurons [68] (**Fig. 4**). It also can be connected to oocyte development alterations, as proposed in our model, since the *atp-2* silencing is related to delayed oocyte budding and defective membrane partitions [32] (**Fig. 4**). These observations indicated that *xpa-1* deficiency can contribute to defects in reproductive development and behavior associated to the upregulation of PUF genes and *atp-2*, respectively (**Fig. 4**).

The present study pointed out DEGs playing a direct role in the cell cycle, including the cyclins *cyb-1* and *cyb-3*, the gene cluster from MCM family and *hcp- 2*. Although it remains unclear how the deregulation of these DEGs in worms reflects on the phenotypes observed in XPA patients, there is evidence that human homologs of these genes are differentially expressed during tumor formation.

Accordingly, the results from the transcriptome analysis pointed that the overexpression of six genes from MCM family, *mcm-2 to 4*, *mcm-6* and *mcm-7*, shared homology with the human MCM complex and act during cell cycle as an ATPase complex, a DNA-helicase complex and a ssDNA-binding oligomer in different conditions [69] (**Fig. 4**). This complex is related to the tumor and precancerous cells progression due to its replicative and proliferative properties [70]. The MCM3 gene is overexpressed in lung, colon and kidney cancers, as well in melanoma, the most common cancer type in XP patients [71]. MCM2 and MCM6 genes, the two human homologs of *C. elegans mcm-2* and *mcm-6*, respectively, are part of DNA replication machinery and are upregulated in premalignant and tumorigenic cells relative to bronchial epithelial cells [72].

The cyclins are also overexpressed in a number of tumors, including breast cancer, skin cell carcinomas and HeLa cell lineages [73–75] (**Fig. 4**). A previous study described that the human cyclin B1 (homolog of worm *cyb-1*) is overexpressed in melanoma cell lines and can contribute to p53 inhibition via iASPP phosphorylation [76], corroborating the *cyb-1* role in the cell proliferation of *xpa-1* deficient worm. On non-melanoma skin cancers, cyclin B1 is implied to be upregulated together with PLK1 (homolog of worm *plk-2*, also overexpressed), which is suggested to promote mitosis in absence of p53 activity [74]. A similar case is the overexpression of a human homolog of *hcp-2* involved in chromosome segregation, MYH10, which is associated to lower survival in patients with melanoma [77].

Since the mitotic cell cycle is highly stimulated during tumorigenesis, that is a recurrent pathological condition in XPA patients, these results point that *xpa-1* deficiency in *C. elegans* not only can result in increased cell proliferation through mechanisms involving cyclins, *hcp-2*, PUF and MCM genes, but could also correspond to an analogue process to cancer in XPA patients. (**Fig. 4**). Although evidences that *xpa-1* deficiency is related to spontaneous tumorigenesis and impaired spermatogenesis already exist for mice [44], this study presents a novel model suggesting the connection between reproduction, cell proliferation and tumor progression as a result of loss of *xpa-1* in molecular level.

5. **Conclusions**

The information extracted from the biological networks, topological analyzes and microarray data allowed us to identify new targets for studies with XP type A. The main findings lead us to propose a molecular model explaining the phenotypes related to *xpa-1* gene mutation in *C. elegans* (**Fig. 4**). The model suggests that the lack of *xpa-1* expression can contribute to enhance oogenesis activity rather than the reduction of spermatogenesis, an outcome that could be derived from: (i) the higher recurrence of molecules that favor oocyte maturation and survival such as *atp-2*, *mrg-1*, *mcm-6*, *cgh-1* and PUF genes; and (ii) overexpression of *cct-4*, which could suppress *gst-4*, a protein that prevent the depletion of sperms production. Similar roles of these proteins are shown in humans, including the homologs of *puf-8* and *cgh-1*, PUM1 and DDX6

respectively, which are necessary to control and prevent apoptosis. Corroborating with our observations, other studies observed that the XPA deficiency resulted in spermatogenesis impairment in mice and humans. In addition to the positive regulation of oogenesis, the overexpression of PUF genes also could repress vulva development though translational suppression, while other up-regulated genes (like *fib-1* and *atp-2*) could play a positive role in gonad development and in the formation of internal membrane partitions, respectively. Analogously, in humans, reduced XPA levels are highlighted in testis cancer, suggesting that the impaired NER activity could be associated to cell replication modifications in human gonads. Similarly, the model indicates that cell cycle inductors (cyclins, PUF genes, MCM genes and centrosome activity modulators, including *knl-1*, *plk-2* and *hcp-2*) amplify the cell replication machinery in response to loss of *xpa-1* expression through processes that, in humans, could be associated to tumor proliferation. As demonstrated on our model, this process could be directly associated to the upregulation of the cell cycle inducers, including PLK1, Cyclin B1 and MCM family genes in XPA patients (**Fig. 4**).

6. Acknowledgements

This work was sponsored by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES). We thank Dr. Diego Bonatto and MSC. Kendi Nishino Miyamoto for the important suggestions made in this work.

7. Conflict of interest statement

The authors declare that they have no conflict of interest.

8. References

1 A. Barzilai and K.-I. Yamamoto, *DNA Repair (Amst).*, 2004, **3**, 1109–15.

2 D. Branzei and M. Foiani, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 297–308.

3 J. H. Lee, C. J. Park, A. I. Arunkumar, W. J. Chazin and B. S. Choi, *Nucleic Acids Res.*, 2003, **31**, 4747–4754.

4 A. Sancar, L. a Lindsey-Boltz, K. Unsal-Kaçmaz and S. Linn, *Annu. Rev. Biochem.*, 2004, **73**, 39–85.

5 K. H. Kraemer and J. J. DiGiovanna, *GeneReviews®*, 2014.

6 S. I. Ahmad and F. Hanaoka, *Molecular Mechanisms of Xeroderma Pigmentosum*, 2009, vol. 637.

7 T. R. Golden and S. Melov, *WormBook*, 2007, 1–12.

8 S. J. Boulton, A. Gartner, J. Reboul, P. Vaglio, N. Dyson, D. E. Hill and M. Vidal, *Sci. (New York, NY)*, 2002, **295**, 127–131.

9 Ø. Fensgård, H. Kassahun, I. Bombik, T. Rognes, J. M. Lindvall and H. Nilsen, *Aging (Albany. NY).*, 2010, **2**, 133–159.

10 K. D. Arczewska, G. G. Tomazella, J. M. Lindvall, H. Kassahun, S. Maglioni, A. Torgovnick, J. Henriksson, O. Matilainen, B. J. Marquis, B. C. Nelson, P. Jaruga, E. Babaie, C. I. Holmberg, T. R. Bürglin, N. Ventura, B. Thiede and H. Nilsen, *Nucleic Acids Res.*, 2013, **41**, 5368–81.

11 D. Sean and P. S. Meltzer, *Bioinformatics*, 2007, **23**, 1846–1847.

12 L. Gautier, L. Cope, B. M. Bolstad and R. A. Irizarry, *Bioinformatics*, 2004, **20**, 307–315.

13 A. Kauffmann, R. Gentleman and W. Huber, *Bioinformatics*, 2009, **25**, 415–416.

14 M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi and G. K. Smyth, *Nucleic Acids Res.*, 2015, **43**, e47.

15 Y. Benjamini and Y. Hochberg, *J. R. Stat. Soc. Ser. B*, 1995, **57**, 289–300.

16 C. von Mering, L. J. Jensen, B. Snel, S. D. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M. A. Huynen and P. Bork, *Nucleic Acids Res.*, 2005, **33**, D433–D437.

17 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.*, 2003, **13**, 2498–2504.

18 M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R.

Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker and G. D. Bader, *Nat. Protoc.*, 2007, **2**, 2366–82.

19 G. Scardoni, M. Petterlini and C. Laudanna, *Bioinformatics*, 2009, **25**, 2857–2859.

20 H. Yu, P. M. Kim, E. Sprecher, V. Trifonov and M. Gerstein, *PLoS Comput. Biol.*, 2007, **3**, e59.

21 M. E. J. Newman, *Soc. Networks*, 2005, **27**, 39–54.

22 G. Bader and C. Hogue, *BMC Bioinformatics*, 2003, **4**, 2.

23 M. Bastian, S. Heymann and M. Jacomy, *Third Int. AAAI Conf. Weblogs Soc. Media*, 2009, 361–362.

24 V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, *J. Stat. Mech. Theory Exp.*, 2008, **10008**, 6.

25 P. J. Mcsweeney, *Google Summer Code*, 2009, 1–8.

26 S. Maere, K. Heymans and M. Kuiper, *Bioinformatics*, 2005, **21**, 3448–3449.

27 The Gene Ontology Consortium, *Nucleic Acids Res.*, 2015, **43**, D1049–D1056.

28 M. Fujita, T. Takasaki, N. Nakajima, T. Kawano, Y. Shimura and H. Sakamoto, *Mech. Dev.*, 2002, **114**, 61–9.

29 A. A. Hill, C. P. Hunter, B. T. Tsung, G. Tucker-Kellogg and E. L. Brown, *Science (80-.).*, 2000, **290**, 809–12.

30 A. Olgun, T. Aleksenko, O. M. Pereira-Smith and D. K. Vassilatis, *J. Gerontol. A. Biol. Sci. Med. Sci.*, 2005, **60**, 543–548.

31 R. E. Navarro, E. Y. Shim, Y. Kohara, a Singson and T. K. Blackwell, *Development*, 2001, **128**, 3221–32.

32 R. A. Green, H. L. Kao, A. Audhya, S. Arur, J. R. Mayers, H. N. Fridolfsson, M. Schulman, S. Schloissnig, S. Niessen, K. Laband, S. Wang, D. A. Starr, A. A. Hyman, T. Schedl, A. Desai, F. Piano, K. C. Gunsalus and K. Oegema, *Cell*, 2011, **145**, 470–482.

33 A. Hubstenberger, C. Cameron, R. Shtofman, S. Gutman and T. C. Evans, *Dev. Biol.*, 2012, **366**, 218–231.

34 A. L. Lublin and T. C. Evans, *Dev. Biol.*, 2007, **303**, 635–49.

35 J. L. Bachorik and J. Kimble, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10893–

10897.

36 S. Zanetti, S. Grinschgl, M. Meola, M. Belfiore, S. Rey, P. Bianchi and A. Puoti, *RNA*, 2012, **18**, 1385–94.

37 U. S. Datla, N. C. Scovill, A. J. Brokamp, E. Kim, A. S. Asch and M. H. Lee, *J. Cell. Physiol.*, 2014, **229**, 1306–1311.

38 S. L. Crittenden, D. S. Bernstein, J. L. Bachorik, B. E. Thompson, M. Gallegos, A. G. Petcherski, G. Moulder, R. Barstead, M. Wickens and J. Kimble, *Nature*, 2002, **417**, 660–663.

39 D. Subasic, T. Stoeger, S. Eisenring, A. M. Matia-González, J. Imig, X. Zheng, L. Xiong, P. Gisler, R. Eberhard, R. Holtackers, A. P. Gerber, L. Pelkmans and M. O. Hengartner, *Genes Dev.*, 2016, **30**, 2213–2225.

40 S. Vaid, M. Ariz, A. Chaturbedi, G. A. Kumar and K. Subramaniam, *Development*, 2013, **140**, 1645–1654.

41 N. W. Kahn, S. L. Rea, S. Moyle, A. Kell and T. E. Johnson, *Biochem. J.*, 2008, **409**, 205–213.

42 H. M. Kubagawa, J. L. Watts, C. Corrigan, J. W. Edmonds, E. Sztul, J. Browse and M. A. Miller, *Nat Cell Biol*, 2006, **8**, 1143–1148.

43 H. Lans and W. Vermeulen, *Mol. Biol. Int.*, 2011, **2011**, 1–12.

44 H. Nakane, S. Hirota, P. J. Brooks, Y. Nakabeppu, Y. Nakatsu, Y. Nishimune, A. Iino and K. Tanaka, *DNA Repair (Amst).*, 2008, **7**, 1938–1950.

45 A. Gu, G. Ji, P. Zhu, Y. Zhou, G. Fu, Y. Xia, L. Song, S. Wang and X. Wang, *Fertil. Steril.*, 2010, **94**, 2620–2625.

46 C. Welsh, R. Day, C. McGurk, J. R. W. Masters, R. D. Wood and B. Köberle, *Int. J. Cancer*, 2004, **110**, 352–361.

47 B. Köberle, J. R. W. Masters, J. A. Hartley and R. D. Wood, *Defective repair of cisplatin-induced DNA damage caused by reduced XPA protein in testicular germ cell tumours*, 1999, vol. 9.

48 C. J. McGurk, M. Cummings, B. Köberle, J. A. Hartley, R. T. Oliver and J. R. Masters, *J. Cell. Biochem.*, 2006, **97**, 1121–1136.

49 J. W. Astin, N. J. O'Neil and P. E. Kuwabara, *DNA Repair (Amst).*, 2008, **7**, 267–80.

50 W. A. Boyd, T. L. Crocker, A. M. Rodriguez, M. C. K. Leung, D. W. Lehmann, J. H. Freedman, B. Van Houten and J. N. Meyer, *Mutat. Res.*, 2010, **683**, 57–67.

51 M. Van Der Voet, M. A. Lorson, D. G. Srinivasan, K. L. Bennett and S. Van Den Heuvel, *Cell Cycle*, 2009, **8**, 4091–4102.

52 G. M. R. Deyter, T. Furuta, Y. Kurasawa and J. M. Schumacher, *PLoS Genet.*, 2010, **6**.

53 R. Sonneville and P. Gönczy, *Development*, 2004, **131**, 3527–43.

54 F. Toyoshima-Morimoto, E. Taniguchi, N. Shinya, A. Iwamatsu and E. Nishida, *Nature*, 2001, **410**, 215–220.

55 M. Decker, S. Jaensch, A. Pozniakovsky, A. Zinke, K. F. O'Connell, W. Zachariae, E. Myers and A. A. Hyman, *Curr. Biol.*, 2011, **21**, 1259–1267.

56 O. Wueseke, D. Zwicker, A. Schwager, Y. L. Wong, K. Oegema, F. Julicher, A. A. Hyman and J. Woodruff, *bioRxiv*, 2016, 67223.

57 O. Cizmecioglu, S. Warnke, M. Arnold, S. Duensing and I. Hoffmann, *Cell Cycle*, 2008, **7**, 3548–3555.

58 H. Ling, K. Hanashiro, T. H. Luong, L. Benavides and K. Fukasawa, *Cell Cycle*, 2015, **14**, 544–553.

59 A. Desai, S. Rybina, T. Müller-Reichert, A. Shevchenko, A. Shevchenko, A. Hyman and K. Oegema, *Genes Dev.*, 2003, **17**, 2421–35.

60 J. P. I. Welburn, M. Vleugel, D. Liu, J. R. Yates, M. A. Lampson, T. Fukagawa and I. M. Cheeseman, *Mol. Cell*, 2010, **38**, 383–392.

61 L. L. Moore, M. Morrison and M. B. Roth, *J. Cell Biol.*, 1999, **147**, 471–479.

62 J. H. Stear and M. B. Roth, *Mol Biol Cell*, 2004, **15**, 5187–5196.

63 I. M. Cheeseman, I. MacLeod, J. R. Yates, K. Oegema and A. Desai, *Curr. Biol.*, 2005, **15**, 771–777.

64 G. Maton, F. Edwards, B. Lacroix, M. Stefanutti, K. Laband, T. Lieury, T. Kim, J. Espeut, J. C. Canman and J. Dumont, *Nat. Cell Biol.*, 2015, **17**, 697–705.

65 J. Z. Mao, P. Jiang, S. P. Cui, Y. L. Ren, J. Zhao, X. H. Yin, A. Enomoto, H. J. Liu, L. Hou, M. Takahashi and B. Zhang, *Cancer Sci.*, 2012, **103**, 1780–1787.

66 F. Yang, Q. Wei, R. S. Adelstein and P. J. Wang, *Dev. Biol.*, 2012, **369**, 356–361.

67 C. B. Walser, G. Battu, E. F. Hoier and A. Hajnal, *Development*, 2006, **133**, 3461–3471.

68 J. Hu and M. M. Barr, *Mol. Biol. Cell*, 2005, **16**, 458–69.

69 C. Schaffitzel, *http://dx.doi.org/10.1080/15384101.2016.1214046*, 2016, **15**,

2391–2392.

70 M. Lei, *Curr Cancer Drug Targets*, 2005, **5**, 365–380.

71 S.-A. Ha, S. M. Shin, H. Namkoong, H. Lee, G. W. Cho, S. Y. Hur, T. E. Kim and J. W. Kim, *Clin. Cancer Res.*, 2004, **10**, 8386–95.

72 H. Kadara, L. Lacroix, C. Behrens, L. Solis, X. Gu, J. J. Lee, E. Tahara, D. Lotan, W. K. Hong, I. I. Wistuba and R. Lotan, *Cancer Prev. Res. (Phila).*, 2009, **2**, 702–11.

73 R. Agarwal, A.-M. Gonzalez-Angulo, S. Myhre, M. Carey, J.-S. Lee, J. Overgaard, J. Alsner, K. Stemke-Hale, A. Lluch, R. M. Neve, W. L. Kuo, T. Sorlie, A. Sahin, V. Valero, K. Keyomarsi, J. W. Gray, A.-L. Borresen-Dale, G. B. Mills and B. T. Hennessy, *Clin. Cancer Res.*, 2009, **15**, 3654–62.

74 T. L. Schmit, W. X. Zhong, M. Nihal and N. Ahmad, *Cell Cycle*, 2009, **8**, 2697–2702.

75 S. Sherwood, D. Rush, A. Kung and R. Schimke, *Exp Cell Res*, 1994, **211**, 275–281.

76 M. Lu, H. Breyssens, V. Salter, S. Zhong, Y. Hu, C. Baer, I. Ratnayaka, A. Sullivan, N. R. Brown, J. Endicott, S. Knapp, B. M. Kessler, M. R. Middleton, C. Siebold, E. Y. Jones, E. V. Sviderskaya, J. Cebon, T. John, O. L. Caballero, C. R. Goding and X. Lu, *Cancer Cell*, 2013, **23**, 618–633.

77 S. Mandruzzato, A. Callegaro, G. Turcatel, S. Francescato, M. C. Montesco, V. Chiarion-Sileni, S. Mocellin, C. R. Rossi, S. Bicciato, E. Wang, F. M. Marincola and P. Zanovello, *J. Transl. Med.*, 2006, **4**, 50.

Tables

Table 1: Gene ontology groups and the HB-DEGs. This table contains all HB-DEGs included in at least two of the bioprocess groups present in **S-material 2**, and point the main potential targets selected for the final molecular model. These genes are also listed together to other DEGs described in **S-table 1**.

Gene ontology	Bioprocess group	Genes
Aging	Aging	ATP-2, CCT-2, CCT-4, CCT-6, DAO-5, DRP-1, H28O16.1, HSP-1, INF-1, MCM-2, MDH-2, PDHB-1, POD-2, RHA-2, RPN-1, RPN-2, SMK-1, TUFM-1, UCR-1, Y45F10D.7, Y47D3A.29
Cell cycle	Cell cycle and replication	CDC-48.2, CGH-1, CHK-1, CIN-4, CPG-1, CYB-1, CYB-3, DAO-5, F17C11.10, HCP-1, HIM-17, HPR-17, HSP-1, IFY-1, KNL-1, MAT-1, MCM-6, MRG-1, PLK-1, PUF-6, PUF-7, PUF-8, SMK-1, XND-1
Cytoskeleton organization	Cell cycle and replication	CGH-1, CYB-3, HCP-1, IFY-1, KNL-1, MDH-2, PLK-1, RHA-2
DNA duplex unwinding	Cell cycle and replication	CIN-4, MCM-2, MCM-3, MCM-4, MCM-6, MCM-7
DNA repair	Cell cycle and replication	F55A3.3, HPR-17, Y47D3A.29
DNA replication	Cell cycle and replication	CIN-4, F55A3.3, K04C2.2, MCM-2, MCM-3, MCM-4, MCM-6, MCM-7, RNR-1, T24C4.5, Y47D3A.29
Germline cell cycle switching, mitotic to meiotic cell cycle	Cell cycle and replication	CGH-1, HSP-1, MRG-1
M phase	Cell cycle and replication	CDC-48.2, CGH-1, CIN-4, CYB-1, CYB-3, DAO-5, F17C11.10, HCP-1, HIM-17, IFY-1, KNL-1, MAT-1, MRG-1, PLK-1, PUF-8, SMK-1, XND-1
M phase of meiotic cell cycle	Cell cycle and replication	CDC-48.2, CIN-4, DAO-5, F17C11.10, HIM-17, IFY-1, KNL-1, MRG-1, PLK-1, PUF-8, SMK-1, XND-1, CYB-1, CYB-3, HCP-1, KNL-1, MAT-1, PLK-1
Meiotic cell cycle	Cell cycle and replication	CDC-48.2, CIN-4, CYB-3, DAO-5, F17C11.10, HIM-17, IFY-1, KNL-1, MRG-1, PLK-1, PUF-8, SMK-1, XND-1
Mitotic cell cycle	Cell cycle and replication	CGH-1, CYB-1, CYB-3, HCP-1, IFY-1, KNL-1, MAT-1, PLK-1

Anatomical structure morphogenesis	Embryonic development	ATP-2, C24H12.5, C27D9.1, CCT-4, CHK-1, CPG-1, D1043.1, DAO-5, DRP-1, HSP-1, MCM-7, MDH-2, MRG-1, RHA-2, SQT-1, T02G5.7
Embryonic development ending in birth or egg hatching	Embryonic development	ACS-1, ACS-4, ATP-2, C16A3.3, C24H12.5, C27D9.1, CBD-1, CCT-2, CCT-4, CCT-6, CDC-48.2, CGH-1, CHK-1, CIN-4, CPG-1, CYB-1, CYB-3, D1043.1, DAO-5, DEPS-1, DRP-1, EGG-6, EIF-3.D, EIF-6, F17C11.10, F55A3.3, F55B11.2, FIB-1, GFM-1, GLA-3, H28O16.1, HCP-1, HIM-17, HSP-1, IFY-1, INF-1, K04C2.2, K07C5.4, K12H4.4, KNL-1, MAT-1, MCM-2, MCM-3, MCM-4, MCM-6, MCM-7, MDH-2, MEX-5, MRG-1, NCBP-1, NOL-5, NPP-10, PDHB-1, PLK-1, POD-2, PRP-21, PRP-8, PUF-3, RHA-2, RME-2, RNR-1, RPC-1, RPN-1, RPN-2, SGT-1, SKR-17, SMK-1, SNRP-200, SQT-1, SWSN-2.2, T20B12.3, T24C4.5, TUFM-1, UCR-1, W05F2.6, WRT-4, Y45F10D.7, Y47D3A.29, Y48B6A.1, Y57G11C.15, Y61A9LA.10, ZK792.5
Embryonic morphogenesis	Embryonic development	ATP-2
Muscle structure development	Embryonic development	MDH-2, RHA-2
Formation of translation preinitiation complex	Gene expression	EIF-3.D
Gene silencing by RNA	Gene expression	CIN-4, NCBP-1, PRP-21, RHA-2, Y61A9LA.10
Negative regulation of gene expression	Gene expression	CIN-4, MRG-1, NCBP-1, PRP-21, RHA-2, Y61A9LA.10
Regulation of gene expression, epigenetic	Gene expression	CIN-4, PRP-21, RHA-2, Y61A9LA.10
Regulation of transcription from RNA polymerase II promoter	Gene expression	MRG-1, SWSN-2.2
RNA interference	Gene expression	CIN-4, PRP-21, RHA-2, Y61A9LA.10
RNA processing	Gene expression	C16A3.3, FIB-1, K01G5.5, NCBP-1, NOL-6, PRO-1, PRP-21, PRP-8, Y45F10D.7, Y48B6A.1
ATP synthesis coupled proton transport	General	ATP-2, H28O16.1

Glucose metabolic process	General	PDHB-1
Locomotion	General	ACS-1, ACS-4, ATP-2, C24H12.5, CCT-6, CDC-48.2, D1043.1, DAO-5, EGG-6, EIF-3.D, FIB-1, HSP-1, INF-1, K04H4.2, K12H4.4, LST-3, MCM-7, NCBP-1, NPP-10, PCCB-1, POD-2, PRO-1, PRP-21, PRP-8, RNR-1, SMK-1, SNRP-200, T02G5.7, T20B12.3, WRT-4
Negative regulation of programmed cell death	General	CGH-1
Organelle fission	General	CYB-1, CYB-3, DRP-1, HCP-1, KNL-1, MAT-1, PLK-1
Response to heat	General	HSP-1
Response to unfolded protein	General	CDC-48.2
Ubiquitin-dependent protein catabolic process	General	CDC-48.2, SKR-17
Genitalia development	Larval development	C24H12.5, CCT-2, CCT-6, CGH-1, CYB-3, DAO-5, EIF-3.D, HSR-9, IFY-1, K04C2.2, K07C5.4, KNL-1, MAT-1, MCM-2, MCM-3, MCM-6, MCM-7, MRG-1, NCBP-1, NOL-5, PLK-1, PRO-1, PRP-21, PRP-8, RNR-1, RPC-1, SKR-17, SMK-1, SWSN-2.2, Y48B6A.1
Gonad development	Larval development	ATP-2, C16A3.3, CDC-48.2, FIB-1, NPP-10, PRO-1, PRP-21, PRP-8, SNRP-200
Larval development	Larval development	ACS-1, ATP-2, C16A3.3, C24H12.5, CCT-2, CCT-4, CCT-6, CDC-48.2, CGH-1, CHK-1, CYB-1, D1043.1, DAO-5, DEPS-1, EIF-6, FIB-1, GFM-1, H28O16.1, HSP-1, HSR-9, INF-1, K01G5.5, K04C2.2, K04H4.2, K07C5.4, K12H4.4, MAT-1, MCM-2, MCM-4, MDH-2, MEX-5, MRG-1, NCBP-1, NOL-5, NPP-10, PLK-1, POD-2, PRO-1, PRP-8, RHA-2, RNR-1, RPC-1, RPN-1, RPN-2, SGT-1, SKR-17, SMK-1, SNRP-200, SWSN-2.2, T20B12.3, TUFM-1, UCR-1, W05F2.6, Y45F10D.7, Y48B6A.1, Y61A9LA.10
Molting cycle, collagen and cuticulin-based cuticle	Larval development	F55A3.3, K12H4.4, PCCB-1, POD-2, PRP-8, WRT-4

Negative regulation of vulval development	Larval development	LST-3, MRG-1, NOL-5, PUF-8, WRT-4
Reproductive developmental process	Larval development	ATP-2, C16A3.3, C24H12.5, CCT-2, CCT-6, CDC-48.2, CGH-1, CPG-1, CYB-1, CYB-3, DAO-5, EIF-3.D, FIB-1, HSR-9, IFY-1, K04C2.2, K07C5.4, KNL-1, MAT-1, MCM-2, MCM-3, MCM-6, MCM-7, MRG-1, MSP-38, MSP-77, MSP-79, NCBP-1, NOL-5, NPP-10, PLK-1, PRO-1, PRP-21, PRP-8, PUF-6, PUF-7, PUF-8, RME-2, RNR-1, RPC-1, SKR-17, SMK-1, SNRP-200, SWSN-2.2, Y48B6A.1
Sex differentiation	Larval development	ATP-2, C16A3.3, C24H12.5, CCT-2, CCT-6, CDC-48.2, CGH-1, CYB-3, DAO-5, EIF-3.D, FIB-1, HSR-9, IFY-1, K04C2.2, K07C5.4, KNL-1, MAT-1, MCM-2, MCM-3, MCM-6, MCM-7, MRG-1, NCBP-1, NOL-5, NPP-10, PLK-1, PRO-1, PRP-21, PRP-8, RNR-1, RPC-1, SKR-17, SMK-1, SNRP-200, SWSN-2.2, Y48B6A.1
Female gamete generation	Reproduction and sexual maturation	CGH-1, CPG-1, CYB-1, CYB-3, EGG-6, EIF-3.D, F55B11.2, MCM-6, MSP-38, MSP-77, MSP-79, NPP-10, PLK-1, PUF-6, PUF-7, RME-2, SMK-1, SQT-1, UCR-1
Oogenesis	Reproduction and sexual maturation	CGH-1, CPG-1, CYB-1, CYB-3, EGG-6, EIF-3.D, F55B11.2, MCM-6, MSP-38, MSP-77, MSP-79, NPP-10, PLK-1, PUF-6, PUF-7, RME-2, SMK-1, SQT-1, UCR-1
Ovulation	Reproduction and sexual maturation	CYB-3, RME-2
Reproduction	Reproduction and sexual maturation	ACS-1, ACS-4, ATP-2, C16A3.3, C24H12.5, C27D9.1, CBD-1, CCT-2, CCT-4, CCT-6, CDC-48.2, CGH-1, CHK-1, CIN-4, CPG-1, CYB-1, CYB-3, DAO-5, DEPS-1, EGG-6, EIF-3.D, EIF-6, F55B11.2, FIB-1, GFM-1, GLA-3, H28O16.1, HSP-1, HSR-9, IFY-1, INF-1, K01G5.5, K04C2.2, K07C5.4, K12H4.4, KNL-1, MAT-1, MCM-2, MCM-3, MCM-6, MCM-7, MRG-1, MSP-38, MSP-77, MSP-79, NCBP-1, NOL-5, NOL-6, NPP-10, PLK-1, PRO-1, PRP-21, PRP-8, PUF-3, PUF-6, PUF-7, PUF-8, RHA-2, RME-2, RNR-1, RPC-1, RPN-1, RPN-2, SGT-1, SKR-17, SMK-1, SNRP-200, SQT-1, SWSN-2.2, T20B12.3, TUFM-1, UCR-1, XND-1, Y45F10D.7, Y47D3A.29, Y48B6A.1, Y57G11C.15, Y61A9LA.10, ZK792.5

Reproductive behavior	Reproduction and sexual maturation	ATP-2, C16A3.3, SMK-1, SNRP-200
Sexual reproduction	Reproduction and sexual maturation	C16A3.3, CGH-1, CPG-1, CYB-1, CYB-3, EGG-6, EIF-3.D, F55B11.2, MCM-6, MRG-1, MSP-38, MSP-77, MSP-79, NPP-10, PLK-1, PRO-1, PRP-21, PUF-3, PUF-6, PUF-7, PUF-8, RME-2, SMK-1, SNRP-200, SQT-1, UCR-1, Y47D3A.29
Single fertilization	Reproduction and sexual maturation	CYB-3, PUF-3, Y47D3A.29
Spermatogenesis	Reproduction and sexual maturation	CGH-1, PUF-8

Figures and Legends

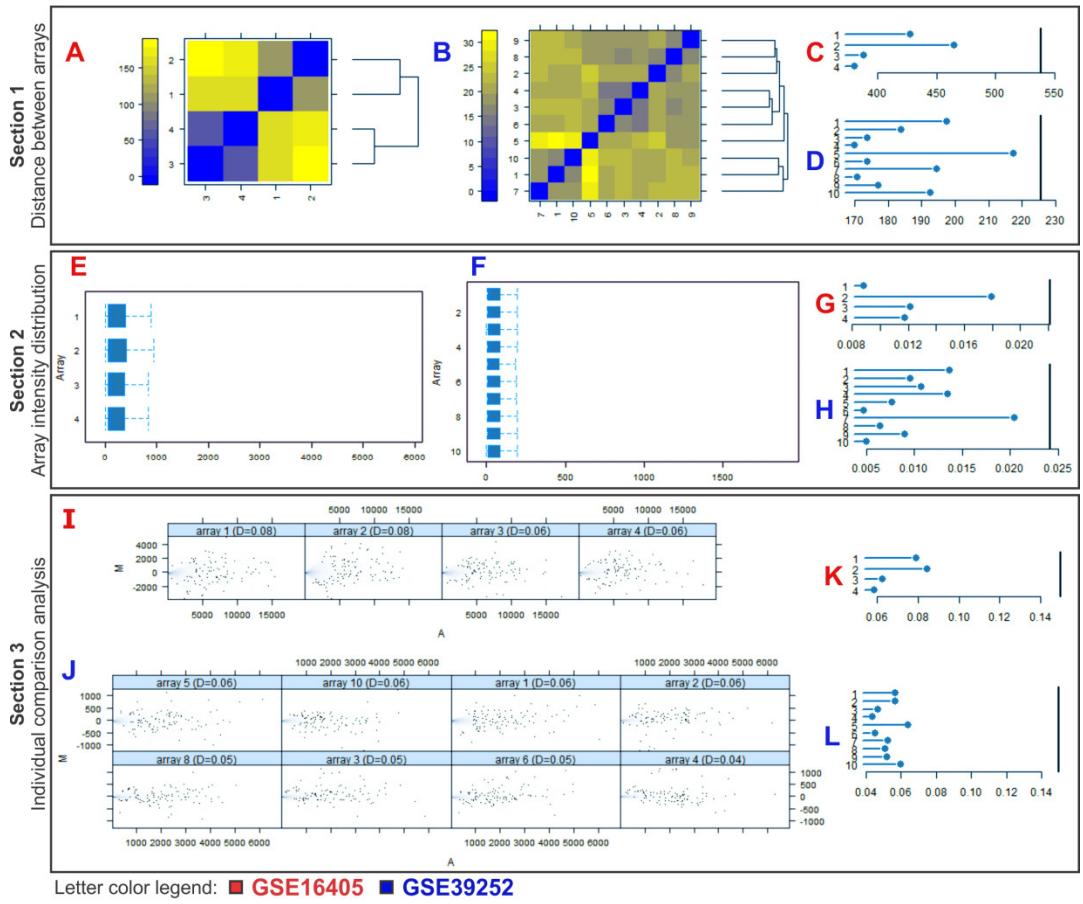


Fig. 1: Transcriptome quality analysis. Through the R platform package ArrayQualityMetrics, the quality analysis was performed and showed no outliers. The analysis is divided in three main sections: (i) comparison between arrays (**Figs. 1A-D**); (ii) array intensity distributions (**Figs. 1E-H**); and (iii) individual array quality (**Figs. 1I-L**). Each section included two outlier detection plots (**Figs. 1C-D, 1G-H**, and **1K-L**) indicating that expression values are below the outlier threshold (vertical bolded line). In the section 1, the quality is validated based on the distances between the sample arrays, and heatmaps (**Figs. 1A-B**) present these distances depicted in blue (for the shortest) and yellow (for the longest) colors. In the section 2, the boxplots (**Figs. 1E-F**) summarize the global expression intensity of each sample, where the contrasted deviations are a consequence of background noise derived from the experiment. In the section 3, the MA plots (**Figs. 1I-J**) represent the comparison between the probes of each sample, where it is expected that the data points are similarly distributed within the axis in similar experimental conditions. The red letters indicate the quality analysis from GSE16405, whose samples are enumerated as 1 and 2 for *xpa-1*-mutated worms, and 3 and 4 for wild-type (N2) worms. The blue letters describe the quality analysis from GSE39252, enumerated from 1 to 5 for N2 worms and from 6 to 10 for *xpa-1*-mutated worms.

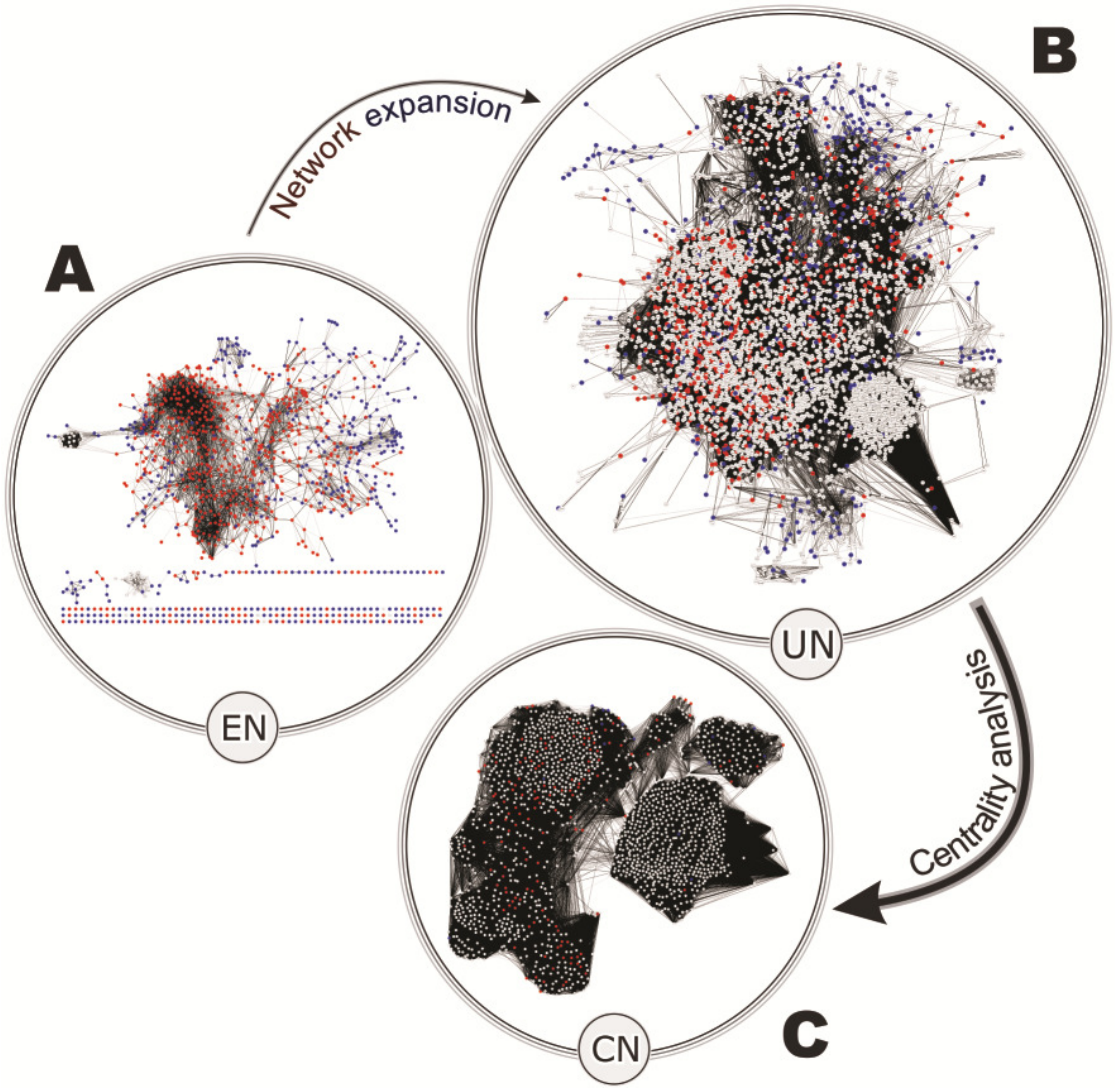


Fig. 2: Generated networks for topological analysis. (A) Expression sub-network (EN), which contains over- and underexpressed genes used as an initial input for creating the United Network (UN). The red nodes represent the overexpressed genes, and the blue nodes represent the underexpressed genes. The DEGs list can be found at **S-material 1**. (B) The UN network, composed by 5,101 nodes and 242,827 edges. (C) Centrality analysis-derived sub-network, named Centrality Network (CN), which represents the most topologically relevant nodes calculated in the UN network. This network is composed by 1,592 nodes and 187,878 edges.

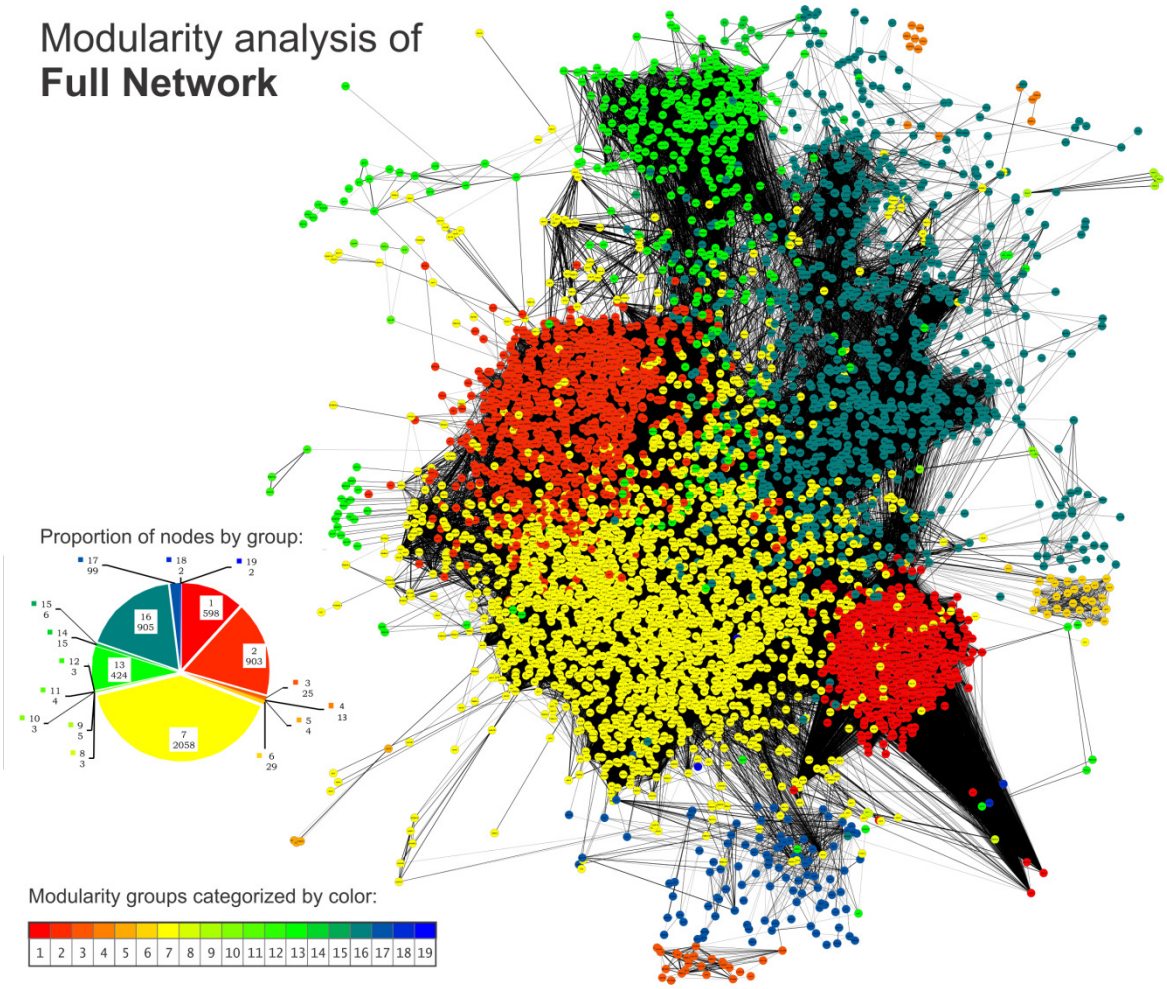


Fig. 3: Modularity analysis for the UN. Using Gephi for modularity analysis, 19 modules were identified. The nodes were highlighted according to its modularity class from the GModule 1 (red), which correspond to the most connected cluster, to GModule 19 (in blue), which correspond to the less connected cluster. The pie chart depicts the number of nodes within each module, pointing that the GModules 1 (in red), 2 (in orange-red), 7 (in yellow), 13 (in green) and 16 (in dark teal) are the UN segments with the higher number of included nodes.

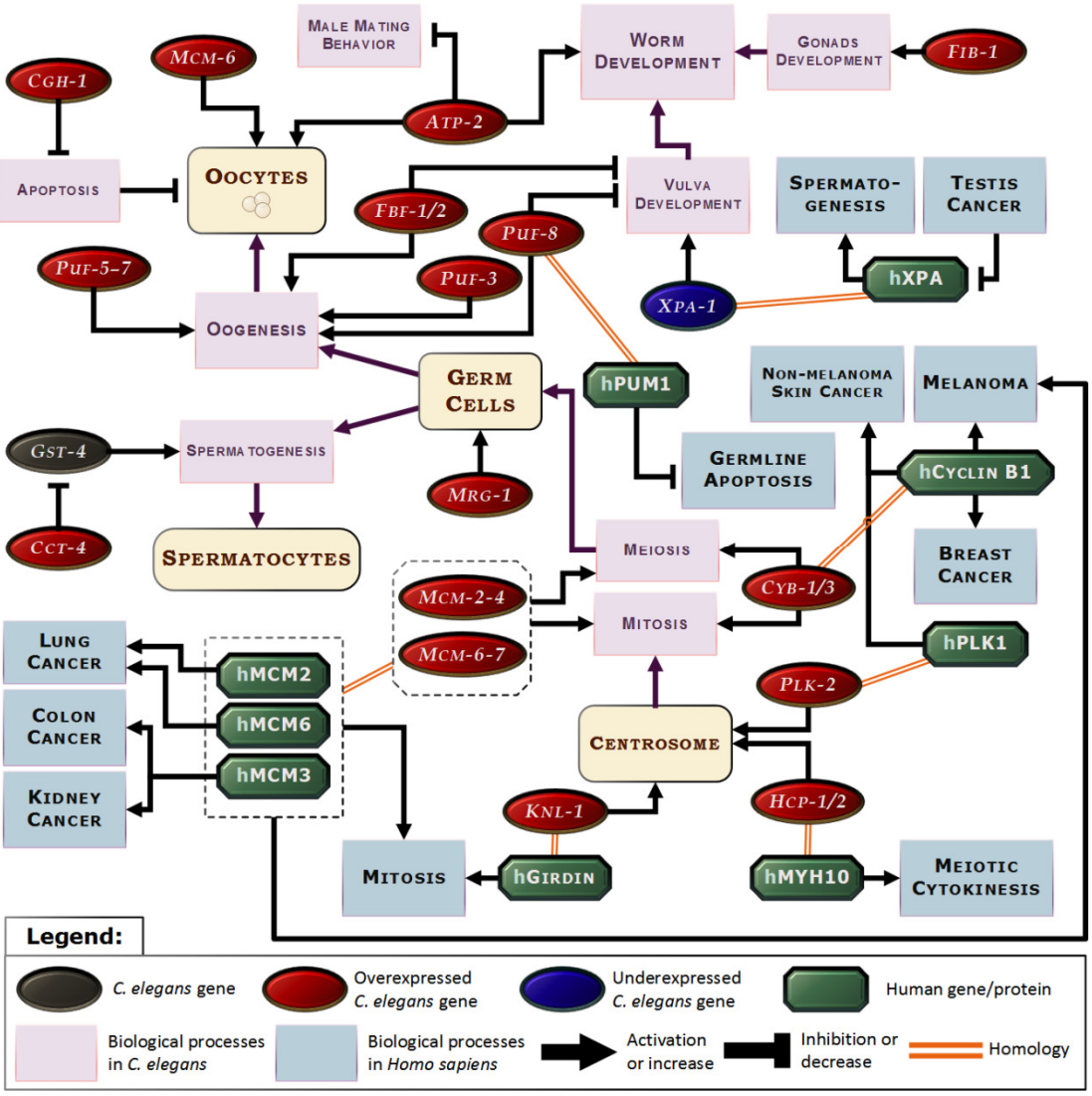


Fig. 4: Final molecular interaction model. This figure summarizes the major connections found between fertility, development, replication and cancer throughout the targets selected from our system biology analyses. The red ellipses represent the overexpressed genes, most of them included in **Table 1**, and the edges denominated as "homology" illustrate the homology relationship between *C. elegans* and human genes. This model suggests that the germ cell differentiation process is altered in favor to oocytes, rather than spermatocytes. In addition, the organs development, such as vulva, appears to be impaired in our model. Likewise, the increase of cell replication process in *xpa-1* mutants would be analogue to the development of malignant tumors in humans.

Curriculum Vitae
Bruno César Feltes

APRESENTAÇÃO

INFORMAÇÕES PESSOAIS

Nome: Bruno César Feltes
Nome em citações bibliográficas: FELTES, B.C.; Feltes, Bruno César.
Sexo: Masculino
Nascimento: Junho, 1988
Cidadania: Brasileiro

ENDEREÇO

Universidade Federal do Rio grande do Sul (UFRGS) – Laboratório Biologia Molecular e Computacional – Sala 219
Departamento de Biologia Molecular e Biotecnologia
Universidade Federal do Rio Grande do Sul – UFRGS
Avenida Bento Goncalves 9500 - Predio 43421
Caixa Postal 15005
Porto Alegre – Rio Grande do Sul
BRASIL
Fone: (+55 51) 3308-6080

CV On-line:
Curriculum Lattes: http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4270137E6
ResearchGate: https://www.researchgate.net/profile/Bruno_Cesar_Feltes/info?ev=prf_info
LinkedIn: https://br.linkedin.com/pub/bruno-c%C3%A9sar-feltes/42/255/97b
E-mails: bcfeltes@gmail.com ou brunofeltes@hotmail.com

FORMAÇÃO ACADÊMICA

- **2007- 2011:** Bacharel em Ciências Biológicas. Universidade de Caxias do Sul, UCS, Caxias do Sul, Brasil. *Título do trabalho de conclusão*: Influência da nicotina na diferenciação celular durante o desenvolvimento embrionário. Orientador: Daniel Luis Notari.
- **2012 – 2013:** Mestrado em Biologia Celular e Molecular. Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil. *Título da dissertação*: Estudo da atuação de substâncias de abuso durante o desenvolvimento embrionário por meio da químio-biologia de sistemas. Orientador: Diego Bonatto.
- **2013 – Atual**: Doutorado em andamento em Biologia Celular e Molecular. Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil. *Título provisório da tese:* Estudo conformacional do complexo proteico DDB2-DDB1 e diferentes variantes mutantes na doença Xeroderma Pigmentosum. Orientador: Diego Bonatto. Co-orientador: Hugo Verli.

ATIVIDADADES PROFISSINAIS AO LONGO DA FORMAÇÃO ACADÊMICA

- Voluntário no Projeto Bioetanol: Isolamento de fenol oxidases em meio sólido e liquido de *Pleurotus sajor-caju* PS-200 e seu uso para degradação de compostos polifenólicos e tintura têxtil. Universidade de Caxias do Sul (UCS), Caxias do Sul – RS, Brasil, 2007-2007.
- Bolsista (BIC-UCS), Projeto Celulases: Isolamento de linhagens recombinantes de *Penicillium echinulatum* por fusão de protoplastos, visando a produção de celulases. Universidade de Caxias do Sul (UCS), Caxias do Sul – RS, Brasil, 2007-2009.
- Bolsista (CNPq), Modulação Epigenética do desenvolvimento embrionário em *Homo sapiens*, *Mus musculus* e *Gallus gallus* como observado pela biologia de sistemas. Universidade de Caxias do Sul (UCS), Caxias do Sul – RS, Brasil, 2009-2011.
- Bolsista de mestrado (CNPq): Estudo da Atuação de Pequenas Moléculas Durante o Desenvolvimento Embrionário Por Meio de Químio-Biologia de Sistemas. Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre – RS, Brasil, 2012 – 2013.
- Bolsista de doutorado (CAPES): Caracterização de regiões interação das proteínas da família Xeroderma Pigmentosa. Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre – RS, Brasil, 2013 - hoje.

ÁREAS DE ATUAÇÃO

- Biologia de Sistemas;
- Redes Biológicas de livre-escala;
- Mecanismos epigenéticos em modelos biológicos;
- Mecanismos de longevidade em modelos biológicos;
- Toxicologia;
- Biologia do Desenvolvimento;
- Biologia Estrutural;
- Análises transcritômicas de microarranjo;
- Dinâmica Molecular;
- Organismos Modelo;
- Reparo de DNA;
- Bioinformática.

IDIOMAS

- Compreende: Inglês (bem) Alemão (pouco) Espanhol (moderado)
- Fala: Inglês (bem) Alemão (pouco) Espanhol (pouco)
- Lê: Inglês (bem) Alemão (pouco) (moderado)
- Escreve: Inglês (bem) Alemão (pouco) (pouco)

INFORMAÇÕES PROFISSIONAIS

PRÊMIOS E MENÇÕES HONROSAS

- Melhor trabalho apresentado no IX Congresso Brasileiro da Sociedade Brasileira de Mutagênese, Carcinogênse e Teratogênese Ambiental, (SBMCTA), Ouro Preto – MG, Brasil, 2009.
- Destaque no XXVI Salão de Iniciação Científica da UFRGS, por orientação de trabalho do aluno Itamar José Guimarães Nunes. Universidade Federal do Rio Grande do Sul – Porto Alegre – RS, Brasil, 2014.

BIBLIOGRAFIA

1. ARTIGOS PUBLICADOS:

- BONATTO, D., **FELTES, B.C**., POLONI, J.F. *Aging as a consequence of intracellular water volume and density*. *Medical Hypotheses*, 77: 982-984, 2011.

- **FELTES, B.C.,** POLONI, J.F., BONATTO, D. *The developmental aging and origins of health and disease hypotheses explained by different protein networks*. *Biogerontology*, 12: 293-308, 2011.

- POLONI, J.F., **FELTES, B.C.,** BONATTO, D. *Melatonin as a central molecule connecting neural development and calcium signaling. Functional & Integrative Genomics*, 11: 383-388, 2011.

- VARGAS, J., **FELTES, B.C.**, POLONI, J.F., LENZ, G., BONATTO, D. *Senescence; an endogenous anticancer mechanism. Frontiers in Bioscience*, 17: 2616-2643, 2012.

- **FELTES, B.C.,** POLONI, J.F., NOTARI, D.L., BONATTO, D. *Toxicological effects of the different substances in tobacco smoke on human embryonic development by a systems chemo-biology approach*. PLoS ONE, 8(4):e61743, 2013.

- **FELTES, B.C.,** BONATTO, D. *Combining small molecules for cell reprogramming through an interatomic analysis. Molecular BioSystems*, 9(11): 2741-2763, 2013.

- **FELTES, B.C.,** POLONI, J.F., NUNES, I.J.G., BONATTO, D. *Fetal Alcohol Syndrome, Chemo-Biology and OMICS: Ethanol Effects on Vitamin Metabolism During*

Neurodevelopment as Measured by Systems Biology Analysis. OMICS: *A Journal of Integrative Biology*, 18: 344-363, 2014.

- POLONI, J. F., CHAPOLA, H., **FELTES, B. C.**, BONATTO, D. *The importance of sphingolipids and reactive oxygen species in cardiovascular development. Biology of the Cell*, 106: 167-181, 2014.

- **FELTES, B. C.**, BONATTO, D. *Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications*. *Mutation Research. Reviews in Mutation Research* (Print), 306-320, 2014.

2. CAPÍTULOS DE LIVROS PUBLICADOS:

- POLONI, J. F., **FELTES, B. C.**, SILVA, F. R., BONATTO, D. Biologia de Sistemas. In: Hugo Verli. (Org.). Bioinformática: da Biologia à Flexibilidade Moleculares. 1ed.: SBBq, 116-146, 2014.

- **FELTES, B.C.**, POLONI, J.F., BONATTO, D. Development and Aging: Two Opposite but Complementary Phenomena. In: Yashin AI; Jazwinski SM. (Org.). Interdisciplinary Topics in Gerontology. 1ed.: S. KARGER AG, 40: 74-84, 2014.

- **FELTES, B. C.**, POLONI, J. F., MIYAMOTO, K. N., BONATTO, D. Human diseases associated with genome instability. In: Igor Kovalchuk, Olga Kovalchuk. (Org.). Genome Stability: From Virus to Human Application. 1ºed.Cambridge: Elsevier, 447-462, 2016.

REVISOR DE PERIÓDICOS:

- **2013 – Atual**: *Current Bioinformatics*.
- **2014 – Atual**: *Journal of Molecular Pathophysiology*.

APRESENTAÇÕES DE TRABALHOS:

1. PALESTRAS

- **FELTES, B. C.** Efeitos Toxicológicos do Tabaco no Desenvolvimento Embrionário Humano. 2014. Saúde em Debate: Métodos Atuais de Biologia Molecular, ULBRA – Canoas, RS, 2014.

- **FELTES, B. C.** *Conformational strudy of a mutated protein complex in Xeroderma Pigmentosum disease using molecular dynamiics and residue interaction networks. 4th International Conference of Integrative Biology*, Berlim - Alemanha, 2016.

2. SEMINÁRIOS

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. . Atuação da Proteína p53 em Resposta a Danos de DNA Induzidos por Espécies Reativas de Oxigênio no Desenvolvimento Embrionário. 2010. (Apresentação de Trabalho/Comunicação).

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. . A influência da melatonina no desenvolvimento neural e na sinalização de cálcio durante a embriogênese. 2010. (Apresentação de Trabalho/Comunicação).

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Efeito modulatório da nicotina na diferenciação celular durante o desenvolvimento embrionário. 2010. (Apresentação de Trabalho/Comunicação).

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Modulação epigenética do desenvolvimento embrionário em *Homo sapiens*, *Mus musculus* e *Gallus gallus* como observado pela biologia de sistemas.. 2009. (Apresentação de Trabalho/Comunicação).

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. A Influência do Ritmo Circadiano no desenvolvimento Embrionário de Diferentes Modelos de Vertebrados por Meio de Análise de Biologia de Sistemas. 2009. (Apresentação de Trabalho/Comunicação).

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Modulação Epigenética do Desenvolvimento Embrionário em Homo sapiens, Mus musculus e Gallus gallus como Observado pela Biologia de Sistemas. 2009. (Apresentação de Trabalho/Comunicação).

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. A Influência da Melatonina Materna no Ritmo Circadiano Fetal Durante o Desenvolvimento Embrionário de Diferentes Modelos de Vertebrados Por Meio de Análise de Biologia de Sistemas. 2009. (Apresentação de Trabalho/Comunicação).

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Epigenetic Modulation of Embryonic Development and Senescence in Homo sapiens, Mus musculus and Gallus gallus as observed by the Systems Biology. 2009. (Apresentação de Trabalho/Congresso).

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. A systems biology analysis of the major proteins associated to vasculogenesis and enzymatic antioxidant mechanisms during the embryonic development of H. sapiens and M. Musculus. 2009. (Apresentação de Trabalho/Congresso).

- **FELTES, B. C.** ; DILLON, A. ; POLONI, J. F. Obtenção de recombinantes entre mutantes de penicillium echinulatum para a produção de celulases por fusão de protoplastos. 2008. (Apresentação de Trabalho/Comunicação).

TEXTOS EM JORNAIS DE NOTÍCIAIS/REVISTAS:

- **FELTES, B. C.** ; BONATTO, D. Acadêmicos destacados em eventos científicos.. Atos & Fatos, Caxias do Sul, p. 4 - 4, 14 dez. 2009.

RESUMOS EXPANDIDOS EM ANAIS DE CONGRESSOS:

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. *Developmental aging in mammals: a systems biology analysis of epigenetic, developmet and senescence mechanisms*. No: *International Society for Computational Biology Latin America Conference*, 2010, Montevideo. ISCB Latin America 2010.

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. *A systems biology view about the role of nuclear melatonin receptors in neurogenesis.* No*: International Society for Computational Biology Latin America Conference*, 2010, MOntevideo. ISCB Latin America, 2010.

RESUMOS PUBLICADOS EM ANAIS DE CONGRESSOS:

- **FELTES, B. C.**; PEDEBOS, C. ; VERLI, H. ; BONATTO, D. *Conformational strudy of a mutated protein complex in Xeroderma Pigmentosum disease using molecular dynamiics and residue interaction networks*. No: *4th International Conference on Integrative Biology, Berlin. Proceedings of 4th International Conference on Integrative Biology*, 2016.

- NUNES, I. J. G., **FELTES, B. C.** , BONATTO, D. *The role in neuroinflammation during brain aging through an interatomic analysis*. No: XVII *Meeting of the Brazilian Society for Cell Biology*, Cataratas do Iguaçu. XVII *Meeting of the Brazilian Society for Cell Biology*, 2014.

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. *Inhibitory Effect of Nicotine in the Retinoic Acid Signalization and Bone Development During Embryogenesis*. No: X Congresso Brasileiro da Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental, 2011, São Pedro. Anais do X Congresso Brasileiro da Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental, 2011.

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. *Exploring the Molecular Effects of Waste Metallurgy Components From a Systemic Perspective by Appluing Systems Chemo-Biology Tools*. No: X Congresso Brasileiro da Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental, 2011, São

- Pedro. Anais do X Congresso Brasileiro da Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental, 2011.

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Atuação da Proteína p53 em Resposta a Danos de DNA Induzidos por Espécies Reativas de Oxigênio no Desenvolvimento Embrionário. In: XVIII Encontro de Jovens Pesquisadores da UCS, 2010, Caxias do Sul. Anais do XVIII Encontro de Jovens Pesquisadores da UCS, 2010.

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. A influência da melatonina no desenvolvimento neural e na sinalização de cálcio durante a embriogênese. In: XVIII Encontro de Jovens Pesquisadores da UCS, 2010, Caxias do Sul. Anais do XVIII Encontro de Jovens Pesquisadores da UCS, 2010.

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Efeito modulatório da nicotina na diferenciação celular durante o desenvolvimento embrionário. In: XXII Salão de Iniciação Científica, 2010, Porto Alegre. Anais do XXII Salão de Iniciação Científica, 2010.

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Epigenetic modulation of embryonic development in Homo sapiens, Mus musculus and Gallus gallus as observed by the systems biology. In: 55º Congresso Brasileiro de Genética, 2009, Águas de Lindóia. 55º Congresso Brasileiro de Genética, 2009. p. 292-292.

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. The influence of circadian rhythm in embryonic development in different types of vertebrates by the analysis of systems biology. In: 55º Congresso Brasileiro de Genética, 2009, Águas de Lindóia. 55º Congresso Brasileiro de Genética, 2009.

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Epigenetic Modulation of Embryonic Development and Senescence in Homo sapiens, Mus musculus and Gallus gallus as observed by the Systems Biology. In: IX Congresso da Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental, 2009, Ouro Preto. Anais do IX Congresso da Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental, 2009.

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. A systems biology analysis of the major proteins associated to vasculogenesis and enzymatic antioxidant mechanisms during the embryonic development of H. sapiens and M. Musculus. In: IX Congresso da Sociedade Brasileira de Mutagênse, Carcinogênese e Teratogênese Ambiental, 2009, Ouro Petro. Anais do IX Congresso da Sociedade Brasileira de Mutagênse, Carcinogênese e Teratogênese Ambienta, 2009.

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Modulação Epigenética do Desenvolvimento Embrionário em Homo sapiens, Mus musculus e Gallus gallus

- como Observado pela Biologia de Sistemas. In: XXI Salão de Iniciação Científica, 2009, Porto Alegre. XXI Salão de Iniciação Científica, 2009.

- **FELTES, B. C.** ; POLONI, J. F. ; BONATTO, D. Modulação Epigenética do Desenvolvimento Embrionário em Homo sapiens, Mus musculus e Gallus gallus como Observado pela Biologia de Sistemas. In: XVII Encontro de Jovens Pesquisadores, 2009, Caxias do Sul. XVII Encontro de Jovens Pesquisadores, 2009.

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. A Influência da Melatonina Materna no Ritmo Circadiano Fetal Durante o Desenvolvimento Embrionário de Diferentes Modelos de Vertebrados Por Meio de Análise de Biologia de Sistemas. In: XXI Salão de Iniciação Científica, 2009, Porto Alegre. XXI Salão de Iniciação Científica, 2009.

- POLONI, J. F. ; **FELTES, B. C.** ; BONATTO, D. A Influência do Ritmo Circadiano no desenvolvimento Embrionário de Diferentes Modelos de Vertebrados por Meio de Análise de Biologia de Sistemas. In: XVII Encontro de Jovens Pesquisadores, 2009, Caxias do Sul. XVII Encontro de Jovens Pesquisadores, 2009.

- CAMASSOLA, Marli ; **FELTES, B. C.** ; DILLON, A. Avaliação de termoestabilidade de xilanases de Penicillium echinulatum em diferentes processos.. In: 5º Congresso Brasileiro de Micologia, 2007, Recife. 5º Congresso Brasileiro de Micologia. Recife: Editora Universitária da UFPE, 2007. v. 1. p. 231-231.

PARTICIPAÇÃO EM EVENTOS, COGRESSOS, EXPOSIÇÕE E FEIRAS

- 4th International Conference of Integrative Biology. Conformational strudy of a mutated protein complex in Xeroderma Pigmentosum disease using molecular dynamics and residue interaction networks. 2016. (Congresso).

- Saúde em Debate: Métodos Atuais de Biologia Molecular.Saúde em Debate: Métodos Atuais de Biologia Molecular. 2014. (Seminário).

- X Congresso Brasileiro da Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental. Inhibitory Effect of Nicotine in the Retinoic Acid Signalization and Bone Development During Embryogenesis. 2011. (Congresso).

- International Society for Computational Biology Latin America Conference. Developmental aging in mammals: a systems biology analysis of epigenetic, developmet and senescence mechanisms. 2010. (Congresso).

- XVIII Encontro de Jovens Pesquisadores da UCS.Atuação da Proteína p53 em Resposta a Danos de DNA Induzidos por Espécies Reativas de Oxigênio no Desenvolvimento Embrionário. 2010. (Encontro).

- XXII Salão de Iniciação Científica.Efeito modulatório da nicotina na diferenciação celular durante o desenvolvimento embrionário. 2010. (Encontro).

- 55º Congresso Brasileiro de Genética. Epigenetic modulation of embryonic development in Homo sapiens, Mus musculus and Gallus gallus as observed by the systems biology. 2009. (Congresso).

- IX Congresso da Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental. Epigenetic Modulation of Embryonic Development and Senescence in Homo sapiens, Mus musculus and Gallus gallus as observed by the Systems Biology. 2009. (Congresso).

- XVII Encontro de Jovens Pesquisadores da UCS..Modulação epigenética do desenvolvimento embrionário em Homo sapiens, Mus musculus e Gallus gallus como observado pela biologia de sistemas.. 2009. (Encontro).

- XXI Salão de Iniciação Científica.Modulação Epigenética do Desenvolvimento Embrionário em Homo sapiens, Mus musculus e Gallus gallus Como Observado Pela Biologia de Sistemas. 2009. (Encontro).

- 1º Simpósio Internacional de Microbiologia Clínica. 2008. (Simpósio).

- IV Simpósio Internacional sobre Cogumelos no Brasil e III Simpósio Nacional sobre Cogumelos Comestíveis. 2008. (Simpósio).

- XVI encontro de Jovens Pesquisadores da UCS.Obtenção de recombinantes entre mutantes de penicillium echinulatum para produção de celulase por fusão de protoplastos. 2008. (Encontro).

- Semana Acadêmica de Biologia - O meio ambiente começa no meio da gente. 2008. (Encontro).

- XVI Simpósio Nacional de Bioprocessos (SINAFERM). 2007. (Simpósio).

ORGANIZAÇÃO DE EVENTOS

- 2008, Simpósio Internacional de Cogumelos no Brasil e III Simpósio nacional de Cogumelos Comestíveis, 27-30 de Outubro, Caxias do Sul, Brasil.
- 2009, Curso Prático e Teórico: Introdução a Primatologia, 11-18 de Maio, Caxias do Sul, Brasil.
- 2009, Semana Acadêmica de Biologia - A Evolução da Consciência a Favor do Meio Ambiente. Caxias do Sul, Brasil.
- 2010, Semana Acadêmica de Biologia - "Uma Mente que se Abre a uma Nova Idéia Nunca Volta ao Seu tamanho Original". Caxias do Sul, Brasil.

- 2013, 8º Edição do Curso de Férias - "ALIMENTOS, O que você gostaria de saber a ciência pode explicar".
- 2017, II Escola Gaúcha de Bioinformática, 17-21 de Julho Porto Alegre, Brasil.

CURSOS MINISTRADOS

- 2017, Introdução a Biologia de Sistemas e Análise de Redes – II Escola Gaúcha de Bioinformática, 17-21 de Julho, Porto Alegre, Brasil.

ORIENTAÇÕES EM ANDAMENTO

- Itamar José Guimarães Nunes. Análise Interatômica Interligando os Mecanismos de Inflamação e Envelhecimento Utilizando Ferramentas da Biologia de Sistemas. Início: 2012. Iniciação científica (Graduando em Biotecnologia) - Universidade Federal do Rio Grande do Sul. (Orientador).