

Building a molecular machine: heterologous
expression and biochemical characterisation of
transcription factor IIH

Reyes Sanles-Falagan



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of
Doctor of Philosophy (PhD)
at the University of St Andrews

October 2018

TABLE OF CONTENTS

TABLE OF CONTENTS	I
DECLARATIONS	VII
ACKNOWLEDGEMENTS.....	XI
ABBREVIATIONS	XIII
LIST OF FIGURES AND TABLES.....	XVII
ABSTRACT	XXI
1. Introduction	1
1.1. DNA damage and repair pathways	1
1.2. The nucleotide excision repair pathway	5
1.2.1. Nucleotide excision repair in Bacteria	5
1.2.2. Nucleotide excision repair in Eukarya	7
1.2.2.1. Damage recognition and verification	8
1.2.2.2. Opening of the repair bubble.....	10
1.2.2.3. Removal of the damaged fragment	14
1.2.2.4. Synthesis and ligation of the new fragment.....	14
1.2.2.5. Recessive disorders associated with NER	16
1.2.2.6. The NER pathway as a drug target	18
1.2.3. Nucleotide excision repair in Archaea	19
1.3. TFIIH: a dual transcription and DNA repair factor	21
1.3.1. Structural study of TFIIH	22
1.3.1.1. XPD takes part in three different complexes.....	24
1.3.2. Role of TFIIH in transcription	25
1.3.3. Evolutionary conservation	27
1.3.4. TFIIH as a viral target	28
1.4. The MultiBac™ system	29
1.4.1. Impact of baculovirus expression systems	29
1.4.2. The MultiBac™ system: a powerful cloning tool	30
1.4.2.1. Heterologous protein expression using MultiBac™.....	30
1.4.2.2. MultiBac™ allows expression of multi-protein complexes.....	32
1.5. Aims and objectives.....	33
2. Materials and methods	35
2.1. Cloning using the MultiBac™ system.....	35

TABLE OF CONTENTS

2.1.1. Engineered genes. 2A Sequences.....	35
2.1.2. Purification tags	36
2.1.3. Tetracysteine (TCP) tag.....	39
2.1.4. Cloning using the MultiBac™ multiplication module.....	40
2.1.5. Tn7 Transposition and bacmid purification.....	42
2.2. Insect cell culture techniques.....	44
2.2.1. Routine maintenance of insect cell cultures	44
2.2.2. Insect cell transfection	45
2.2.3. Insect cell infection	45
2.2.4. Insect cell co-infection.....	45
2.3. Purification of the TFIIH complex.....	46
2.3.1. TFIIH Extraction and BioSprint expression tests	46
2.3.1.1. BioSprint expression test in anaerobic conditions	46
2.3.1.2. SDS-PAGE analysis.....	47
2.3.1.3. Western blotting analysis	47
2.3.2. Immunoprecipitation using a V5 antibody.....	48
2.3.3. Immobilised metal ion affinity chromatography (IMAC)	48
2.3.3.1. HisTrap™ Fast Flow (FF) column.....	48
2.3.3.2. XPD purification using detergent ANAPOE C ₁₂ E ₈	49
2.3.3.3. Talon® Superflow™ resin.....	49
2.3.4. Size exclusion chromatography	50
2.3.5. Cation exchange chromatography	50
2.3.6. Affinity chromatography	50
2.4. Purification of XPC constructs	51
2.4.1. XPC Extraction and BioSprint expression tests.....	51
2.4.2. Immobilised metal ion affinity chromatography.....	51
2.4.2.1. HisTrap™ FF column.....	51
2.4.2.2. Talon® Superflow™ resin.....	51
2.4.3. Size exclusion chromatography	52
2.4.4. Cation exchange chromatography	52
2.4.5. Heparin chromatography	52
2.5. Biochemical characterization of XPC and the TFIIH Core sub-complex.....	53
2.5.1. ReAsH-EDT ₂ Labelling of TCP-tagged XPC and TFIIH Core	53
2.5.2. Crosslinking assays	54
2.5.3. DNA substrate preparation.....	54

TABLE OF CONTENTS

2.5.4. Electrophoretic mobility shift assays (EMSA)	55
2.5.5. Fluorescence experiments	56
2.5.5.1. Fluorescence anisotropy assays	56
2.5.5.2. Bulk fluorescence resonance energy transfer (FRET) assays	57
2.5.5.3. Fluorescence-based helicase assays.....	59
3. XPC: the protein that starts it all	61
3.1. Introduction.....	61
3.1.1. <i>In vivo</i> XPC is a heterotrimeric complex.....	61
3.1.2. XPC structure: what we know	62
3.1.3. XPC detects mismatched bases in the DNA	63
3.2. Results	66
3.2.1. Cloning, expression and purification of XPC	66
3.2.1.1. Monomeric XPC.....	66
3.2.1.1.1. Cloning and expression tests	66
3.2.1.1.2. Large-scale purification	67
3.2.1.2. TCP-Tagged XPC.....	68
3.2.1.2.1. Cloning and expression tests	68
3.2.1.2.2. Large-scale purification	69
3.2.1.2.3. ReAsH-EDT ₂ Labelling of TCP-tagged XPC	70
3.2.1.3. XPC-HR23B.....	73
3.2.1.3.1. Cloning and expression tests	73
3.2.1.3.2. Large-scale purification	74
3.2.2. Biochemical characterisation of XPC	76
3.2.2.1. Anisotropy assays.....	77
3.2.2.2. Electrophoretic mobility shift assay	78
3.2.2.3. Bulk FRET assays.....	81
3.3. Discussion	85
3.3.1. Heterologous expression of XPC constructs	85
3.3.1.1. Monomeric XPC.....	85
3.3.1.2. TCP-tagged XPC	86
3.3.1.3. Heterodimer XPC-HR23B	88
3.3.2. XPC binds to a range of double-stranded DNA substrates with similar affinity	89
3.3.2.1. Anisotropy experiments.....	90
3.3.2.2. Electrophoretic mobility shift assay	91

TABLE OF CONTENTS

3.3.2.3. Bulk FRET assays.....	92
3.4. Conclusions	93
4. Cloning, expression and purification of the TFIIH complex	97
4.1. Introduction.....	97
4.2. Results	100
4.2.1. TFIIH Core sub-complex.....	100
4.2.1.1. TFIIH Core cloning and expression: first attempts.....	100
4.2.1.2. A redistribution of the purification tags is required	105
4.2.1.3. A different purification tag might be the answer.....	114
4.2.1.4. Tagging the C-terminal end of XPD.....	121
4.2.1.5. Twin-Strep tag®: the solution to our problem	122
4.2.2. 10-Subunit TFIIH complex	124
4.2.2.1. Obtaining the CAK sub-complex	124
4.2.2.2. Co-infecting to obtain the 10-subunit TFIIH complex	126
4.2.2.3. Purification of 10-subunit TFIIH.....	127
4.2.3. Cloning and expression of other TFIIH constructs.....	129
4.2.3.1. TCP-Tagged TFIIH Core.....	129
4.2.3.1.1. Cloning and expression tests	129
4.2.3.1.2. Purification	130
4.2.3.1.3. ReAsH-EDT ₂ Labelling of TCP-tagged TFIIH Core.....	131
4.3. Discussion	132
4.3.1. Cloning and expressing the TFIIH Core sub-complex	132
4.3.2. The elusive XPD helicase	134
4.3.3. Cloning and expressing the 10-subunit TFIIH complex	136
4.3.4. Cloning and expression TCP-tagged TFIIH Core	137
4.4. Conclusions.....	138
5. Studying TFIIH interactions with DNA.....	141
5.1. Introduction.....	141
5.1.1. TFIIH: Early structural knowledge	141
5.1.2. TFIIH's redefined mechanism of action.....	145
5.2. Results	147
5.2.1. Observing TFIIH binding to damaged DNA.....	147
5.2.1.1. Electrophoretic mobility shift assay (EMSA).....	147
5.2.1.2. Anisotropy assays.....	148
5.2.1.2.1. 10-subunit TFIIH complex	148

TABLE OF CONTENTS

5.2.1.2.2. TFIIH Core sub-complex	149
5.2.1.3. Bulk FRET assays.....	155
5.2.2. Characterization of TFIIH’s unwinding activity.....	157
5.2.2.1. Determining the polarity of the TFIIH Core unwinding	158
5.2.2.2. Stablishing the optimal conditions for the assay	161
5.2.2.3. Effect of RPA and XPA on TFIIH Core unwinding	163
5.2.2.4. Unwinding of damaged DNA substrates.....	164
5.3. Discussion.....	166
5.3.1. TFIIH binds dsDNA without the presence of XPC	166
5.3.2. A fluorescence-based helicase assay shows TFIIH unwinding of dsDNA	170
5.4. Conclusions.....	173
6. Conclusions and further work	177
6.1. Summary.....	177
6.2. Conclusions.....	181
6.3. Future work	183
BIBLIOGRAPHY.....	187
APPENDIX A: Synthetic gene sequences	205
TFIIH Core synthetic gene sequences (5' → 3').....	205
<i>XPB</i>	205
<i>XPD</i>	206
<i>p62</i>	207
<i>p52-V5-F2A-p8</i>	208
<i>p44-T2A-p34</i>	209
CAK synthetic gene sequences.....	211
<i>MAT1</i>	211
<i>Cdk7-T2A-Cyclin H</i>	211
<i>XPC</i> Synthetic gene sequence	213
<i>HR23B</i> Synthetic gene sequence	214
APPENDIX B: Plasmids and bacmids	217
APPENDIX C: PCR oligonucleotides.....	221
APPENDIX D: DNA substrates.....	223

DECLARATIONS

Candidate's declaration

I, Reyes Sanles-Falagan, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 49,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in May 2014.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date

Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date

Signature of supervisor

DECLARATIONS

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Reyes Sanles-Falagan, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of 1 year on the following ground(s):

Publication would preclude future publication

Supporting statement for printed embargo request

Publication of the thesis might reduce possibility of publication in a scientific journal

Electronic copy

Embargo on all of electronic copy for a period of 1 year on the following ground(s):

Publication would preclude future publication

Supporting statement for electronic embargo request

Publication of the thesis might reduce possibility of publication in a scientific journal

Title and Abstract

I agree to the title and abstract being published.

Date

Signature of candidate

Date

Signature of supervisor

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Professor Malcolm White for his incredible support, trust, and guidance through my PhD. I would also like to thank Dr Carlos Penedo Esteiro for his invaluable advice, and Mrs Biljana Petrovic-Stojanovska for her hard work on this project.

My gratitude goes to the current and past members of the White and Penedo labs for their help and encouragement, specially to Dr Jose Peregrina, Dr Cibran Perez-Gonzalez and Dr Clare Rollie for their incredible support in and outside the lab. Dr Shirley Graham also deserves a special mention for her help and endless patience (again, in and outside the lab).

I am very grateful to the School of Biology, EastBio and the Biotechnology and Biological Sciences Research Council (BBSRC) for supporting this work [grant number BB/J01446X/1] and having given me this tremendous opportunity.

My heartfelt thanks go to my family for their unwavering support and understanding. And last but not least, my biggest thanks to Jonathan for always being there for me, and to Ollie, because he makes everything a little bit better.

*“Science, my lad, is made up of mistakes,
but they are mistakes which it is useful to make,
because they lead little by little to the truth.”*

Jules Verne, A Journey to the Centre of the Earth

ABBREVIATIONS

AcMNPV	<i>Autographa californica</i> multiple nucleopolyhedrovirus
ANT2	Adenine nucleotide translocase 2
APC/C	Anaphase promoting complex
ATP	Adenosine triphosphate
BER	Base excision repair
BEV	Baculoviral expression vector
BHD	β -Hairpin domain
BME	2-Mercaptoethanol
BSA	Bovine serum albumin
BS3	Bis(sulfosuccinimidyl)suberate
CAK	Cdk-activating kinase
Cdk7	Cyclin-dependent kinase 7
CETN2	Centrin 2
CIA	Cytoplasmic iron-sulfur cluster assembly complex
CIAO1	Cytosolic iron-sulfur assembly component 1
Cisplatin	<i>cis</i> -diaminodichloroplatinum
COFS	Cerebro-oculo-facio-skeletal syndrome
CPD	Cyclobutane pyrimidine dimer
CryoEM	Electron cryomicroscopy
CS	Cockayne syndrome
CSB	Cockayne syndrome complementation group B
CTD	C-terminal domain
DDB1	Damaged DNA-binding protein 1
DMPS	2,3-Dimercaptopropanesulfonate
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded deoxyribonucleic acid
ssDNA	Single-stranded deoxyribonucleic acid
DNApol	Deoxyribonucleic acid polymerase

ABBREVIATIONS

DMP	Dimethyl pimelimidate
DMPS	2,3-dimercaptopropanesulfonate
DRD	Damage recognition domain
DSB	Double-strand break
DSG	Daughter strand gap
DTT	Dithiothreitol
EBNA2	Epstein-Barr virus nuclear antigen 2
EBV	Epstein-Barr virus
EDT ₂	1,2-ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
EndoMS	Endonuclease mismatch specific
ERCC1	Excision repair cross-complementation group 1
FAD	Flavin adenine dinucleotide
FAM	Fluorescein amino modified
FCS	Fluorescence correlation spectroscopy
FMDV	Foot-and-mouth disease virus
FRET	Fluorescence resonance energy transfer
GGR	Global genome repair
HD1	Helicase domain 1
HD2	Helicase domain 2
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HR	Homologous recombination
IDT	Integrated DNA technologies
IMAC	Immobilised metal ion affinity chromatography
IP	Immunoprecipitation
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Luria-Bertani broth
MAT1	Ménage-à-trois 1
MES	2-(N-morpholino)ethanesulfonic acid

ABBREVIATIONS

Mfd	Mutation frequency decline
MIP18	MSS19-interacting protein of 18 kDa
MMR	Mismatch repair
MMXD	MMS19-MIP18-XPB complex
MS	Mass spectrometry
MutH	Mutator H
MutL	Mutator L
MutS	Mutator S
NAD	Nicotinamide adenine dinucleotide
NER	Nucleotide excision repair
NHEJ	Non-homologous end-joining
NSs	Non-structural protein
OCT4	Octamer-binding transcription factor 4
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCNA	Proliferating cell nuclear antigen
PH	Pleckstrin homology domain
PHR	Photoreactivation
PIC	Pre-initiation complex
6-4PP	6–4 pyrimidine–pyrimidone photoproduct
RFC	Replication factor C
RNApol	Ribonucleic acid polymerase
RPA	Replication protein A
RVFV	Rift Valley fever virus
SD	Splayed duplex
SDS	Sodium dodecyl sulphate
Sf	<i>Spodoptera frugiperda</i>
SFM	Serum-free medium
SOX2	SRY-box 2
SRY	Sex- determining region Y
SSB	Single-stranded DNA binding protein

ABBREVIATIONS

TBE	Tris – Borate – EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TCP	Tetracysteine peptide
TCR	Transcription coupled repair
TFIIA	Transcription factor II A
TFIIB	Transcription factor II B
TFIID	Transcription factor II D
TFIIE	Transcription factor II E
TFIIF	Transcription factor II F
TFIIH	Transcription factor II H
TGD	Transglutaminase-homology domain
ThM	Thumb domain
TIRF	Total internal reflection fluorescence
TTD	Trichothiodystrophy
UV	Ultraviolet
UV ^{SS}	UV sensitivity syndrome
UVSSA	UV stimulated scaffold A
VP16	Herpes simplex virion protein 16
VWA	von Willebrand factor A fold
XP	Xeroderma pigmentosum
XPA	Xeroderma pigmentosum complementation group A
XPB	Xeroderma pigmentosum complementation group B
XPC	Xeroderma pigmentosum complementation group C
XPD	Xeroderma pigmentosum complementation group D
XPE	Xeroderma pigmentosum complementation group E
XPF	Xeroderma pigmentosum complementation group F
XPG	Xeroderma pigmentosum complementation group G

LIST OF FIGURES AND TABLES

Figure 1: DNA repair pathways.....	1
Figure 2: Representation of the GGR and TCR pathways in Bacteria.....	7
Figure 3: Damage recognition in the eukaryotic GGR and TCR sub-pathways.....	9
Figure 4: The TFIIH complex opens the DNA around the damaged site.	11
Figure 5: Structure of the TFIIH 5' to 3' helicase XPD.....	12
Figure 6: Damage verification and DNA unwinding in eukaryotic NER.	13
Figure 7: Excision and repair synthesis in NER.....	15
Table 1: Autosomal recessive disorders caused by mutations in NER factors	16
Figure 8: The TFIIH complex as a target for chemotherapeutic drugs.	19
Figure 9: The NER pathway in Archaea.	20
Figure 10: Structure of the 10-subunit complex TFIIH.....	23
Figure 11: Different roles associated to helicase XPD.	24
Figure 12: Sequential assembly of the PIC.....	26
Figure 13: Applications of BEVS in biotechnology.	30
Figure 14: MultiBac™ is a powerful cloning tool.....	31
Table 2: UniProt accession numbers for the proteins included in this study.....	35
Figure 15: TFIIH genes engineered with 2A autocleavable sequences.....	36
Figure 16: 6xHis-V5-TEV purification tag.....	37
Figure 17: 8xHis-V5-spacer-TEV purification tag.....	37
Figure 18: TEV-10xHis-V5 purification tag.....	38
Figure 19: Twin-Strep-tag® purification tag.	38
Figure 20: TCP-tagged XPC and XPB.....	39
Figure 21: pACEBac1 and pACEBac2 vector maps	40
Figure 22: Cloning using MultiBac™'s multiplication module.	41
Table 3: PCR reaction for verification of intermediate TFIIH Core constructs.	42
Figure 23: Transposition of a multi-gene construct into the baculoviral genome.	43
Figure 24: Summary of <i>Sf9</i> culture techniques.....	44
Figure 25: DNA substrates employed in EMSA assays.	55
Figure 26: DNA substrates employed in fluorescence anisotropy assays.....	57
Figure 27: DNA substrates employed in bulk FRET assays.....	59
Figure 28: DNA substrates employed in helicase activity assays.....	60
Figure 29: Cartoon view of XPC's homologue Rad4.....	62
Figure 30: XPC interacts with TFIIH subunit p62.	63

LIST OF FIGURES AND TABLES

Figure 31: XPC detects distortions in the DNA.	64
Figure 32: Model showing the two-step detection of lesions in the DNA by XPC.	65
Figure 33: Purification test of monomeric XPC.	66
Figure 34: Large-scale purification of monomeric XPC.	67
Figure 35: Purification test of TCP-tagged monomeric XPC.	69
Figure 36: Large-scale purification of TCP-tagged monomeric XPC.	70
Figure 37: ReAsH-EDT ₂ labelling of a TCP-tagged protein.	71
Figure 38: Structure of biarsenic ligands FIAsH-EDT ₂ (A) and ReAsH-EDT ₂ (B).	72
Figure 39: ReAsH-EDT ₂ labelling of construct TCP-XPC.	73
Figure 40: Purification test of heterodimer XPC-HR23B.	74
Figure 41: Large-scale purification of heterodimer XPC-HR23B.	75
Figure 42: Substrates used in the study of binding of monomeric XPC, heterodimer XPC-HR23B and complex XPC-HR23B-TFIIH.	76
Figure 43: Fluorescence anisotropy assay showing binding of monomeric XPC to substrates S1, S2, S3 and S4.	78
Figure 44: EMSA assay (10% polyacrylamide gel) showing binding of heterodimer XPC-HR23B to substrates S7 and S8.	79
Figure 45: EMSA assay (6% polyacrylamide gel) showing binding of the XPC-HR23B complex to substrates S2, S5 and S6.	80
Figure 46: Calculation of the RatioA.	81
Figure 47: FRET analysis of binding of XPC-HR23B to substrate S7.	82
Figure 48: FRET analysis of binding of XPC-HR23B to substrate S8.	83
Figure 49: FRET analysis of binding of TFIIH Core to the S7-XPC-HR23B complex. .	84
Figure 50: FRET analysis of binding of TFIIH Core to the S8-XPC-HR23B complex. .	84
Table 4: Dissociation constants for binding of monomeric XPC, heterodimer XPC-HR23B and complex XPC-HR23B-TFIIH to a range of substrates in a fluorescence anisotropy, EMSA or bulk FRET assay.	89
Figure 51: Structure of the 10-subunit complex TFIIH.	98
Figure 52: Overview of the cloning and expression of the TFIIH complex.	99
Figure 53: TFIIH constructs cloned and expressed in this PhD project.	100
Figure 54: Steps in the cloning of TFIIH Core with 6xHis-V5-TEV-tagged p52 and XPD.	101
Figure 55: Cloning of TFIIH Core with 6xHis-V5-TEV-tagged p52 and XPD (I).	102
Figure 56: Cloning of TFIIH Core with 6xHis-V5-TEV-tagged p52 and XPD (II).	103
Figure 57: Purification test of TFIIH Core with 6xHis-V5-TEV-tagged p52 and XPD.	104
Figure 58: Steps in the cloning of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.	105
Figure 59: Cloning of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.	106
Figure 60: Purification test of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52. .	107

LIST OF FIGURES AND TABLES

Figure 61: Purification of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.	108
Figure 62: Subunit XPD is lost in the purification of TFIIH Core.	109
Figure 63: Optimised purification of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.	111
Figure 64: TFIIH complex and sub-complexes obtained by crosslinking.	112
Table 5: Crosslinks connecting TFIIH Core subunits.	112
Figure 65: Map of the crosslinks obtained in our assays.	114
Figure 66: TFIIH constructs carrying an 8xHis-V5-spacer-TEV tag.	115
Figure 67: Purification test of TFIIH Core with 8xHis-V5-spacer-TEV-tagged XPD and p52.	116
Figure 68: Purification of TFIIH Core with 8xHis-V5-spacer-TEV-tagged XPD and p52.	117
Figure 69: Purification test of 6-subunit TFIIH Core with 8xHis-V5-spacer-TEV-tagged p52.	118
Figure 70: Purification of 6-subunit TFIIH Core with 8xHis-V5-spacer-TEV-tagged p52.	118
Figure 71: Purification test and anaerobic purification of the 8xHis-V5-spacer-TEV- tagged XPD helicase.	119
Figure 72: IMAC and IP purification of 8xHis-V5-spacer-TEV -tagged XPD helicase.	120
Figure 73: Purification tests of 8xHis-V5-spacer-TEV-tagged XPD helicase using different detergents.	122
Figure 74: Purification of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52 and helicase XPD carrying a Twin-Strep-tag® at its C-terminal end.	123
Figure 75: Purification test of the CAK sub-complex with 8xHis-V5-spacer-TEV-tagged MAT1, cdk7.	124
Figure 76: Purification test of the CAK sub-complex with 8xHis-V5-spacer-TEV-tagged MAT1 (A) and 8xHis-V5-spacer-TEV-tagged cdk7 (B).	125
Figure 77: Purification test of the 10-subunit TFIIH complex.	126
Figure 78: Purification of the 10-subunit TFIIH complex.	128
Figure 79: Purification test of the TCP-tagged TFIIH Core sub-complex.	129
Figure 80: Purification of the TCP-tagged TFIIH Core sub-complex.	130
Figure 81: ReAsH labelling of construct TFIIH Core.	131
Figure 82: Structure of the XPB homologue from <i>A. fulgidus</i>	142
Figure 83: Crystal structure of different archaeal homologues of helicase XPD.	143
Figure 84: Architecture of the different TFIIH Core subunits.	145
Figure 85: TFIIH opens DNA through the combined action of XPB and XPD.	146
Figure 86: EMSA assay (10% polyacrylamide gel) showing binding of the TFIIH Core to substrate S8.	148
Figure 87: Fluorescence anisotropy assay showing binding of the 10-subunit TFIIH complex to substrate S1.	149

LIST OF FIGURES AND TABLES

Figure 88: Fluorescence anisotropy assay showing binding of the TFIIH Core sub-complex to substrates S1, S2, S3 and S4 in buffer MES.	150
Figure 89: Influence of the reaction buffer in our anisotropy experiments.	152
Figure 90: Fluorescence anisotropy assay showing binding of the TFIIH Core sub-complex to substrates S3 and S4 in buffer 20 mM HEPES, 50 mM NaCl.	153
Figure 91: Fluorescence anisotropy assay showing binding of the TFIIH Core sub-complex to substrates S5 and S6.	154
Figure 92: Fluorescence anisotropy assay showing binding of the XPD-enriched TFIIH Core sub-complex to substrates S5 and S6.	155
Figure 93: Bulk FRET assay showing binding of TFIIH Core to substrates S7 and S8.	156
Figure 94: DNA substrates employed to characterize TFIIH's unwinding activity.	157
Figure 95: Mechanism of action of our fluorescence-based helicase assay.	158
Figure 96: TFIIH Core unwinds a substrate with a 5' overhang.	159
Figure 97: TFIIH Core opens a splayed duplex substrate preferentially.	160
Figure 98: XPD discriminates substrate composition.	161
Figure 99: Unwinding of the SD substrate increases with TFIIH Core concentration.	162
Figure 100: Establishing the ideal ATP concentration and temperature for our assay.	163
Figure 101: Effect of XPA and RPA on TFIIH Core unwinding of substrate SD.	164
Figure 102: TFIIH Core is stalled by damage in the translocating and the non-translocating strands.	165
Figure 103: Kinetics of the TFIIH Core unwinding of substrates SD, SD13 and SD24.	166
Table 6: Dissociation constants for binding of different TFIIH constructs to a range of substrates in EMSA, fluorescence anisotropy or bulk FRET assays.	167
Figure 104: Unravelling the repair bubble.	184
Figure 105: Hypothetical model describing the first stages of a GGR reaction.	185

ABSTRACT

Organisms across Archaea, Bacteria and Eukarya possess a series of mechanisms to protect the integrity of the deoxyribonucleic acid (DNA) from the constant attack of internal and external factors that modify the structure of the double helix, potentially causing an increase in mutagenesis. The Nucleotide Excision Repair (NER) pathway removes bulky lesions in the DNA caused by ultraviolet (UV) radiation and other sources that can introduce a strong distortion in the double helix. A key element in this repair mechanism is the transcription factor IIH (TFIIH), a ten-subunit complex which opens and extends the DNA bubble originated at the damaged site to allow the subsequent repair factors access to the lesion, so the insult can be removed. Mutations in TFIIH subunits xeroderma pigmentosum group B (XPB), xeroderma pigmentosum group D (XPD) or p8 can cause a series of autosomal recessive disorders with symptoms that include mild-to-extreme photosensitivity, progeria, physical and neurological abnormalities, and in some cases an increased susceptibility to cancer.

This thesis describes the processes of cloning, expression and purification followed to obtain damage-detector heterodimer XPC-HR23B, the 7-subunit TFIIH Core sub-complex and the 10-subunit TFIIH complex by means of a powerful baculoviral expression vector (BEV) called MultiBac™, a tool specifically conceived for the obtaining of multi-subunit eukaryotic complexes. The biochemical characterization of TFIIH Core showed that the sub-complex can bind a series of double-stranded DNA (dsDNA) substrates with an affinity in the nM range, and the presence of a bubble or damage at the duplex does not change TFIIH Core's binding affinity significantly. The study of TFIIH Core's unwinding ability confirmed that the sub-complex can open a series of dsDNA substrates only when a 5' overhang end is available, and this activity can be stalled by a lesion located in both the translocating and non-translocating strands.

1. Introduction

1.1. DNA damage and repair pathways

The integrity of the deoxyribonucleic acid (DNA) is constantly threatened by internal and external factors (e.g. metabolic products, ultraviolet (UV) radiation), with a human cell typically suffering more than 10^5 lesions in a single day (Tomkinson *et al.*, 2006). These distortions introduced in the double helix can potentially lead to an increase in mutagenesis, which in turn may result in a higher risk of cancer. As damaged DNA cannot be replaced, cells depend on different repair systems to maintain its integrity.

DNA repair mechanisms (figure 1) can be classified into three different groups: (1) direct repair, (2) recombinational repair and (3) excision repair, with each one including many sub-classes, although some overlapping between mechanisms frequently occurs (Eisen & Hanawalt, 1999).

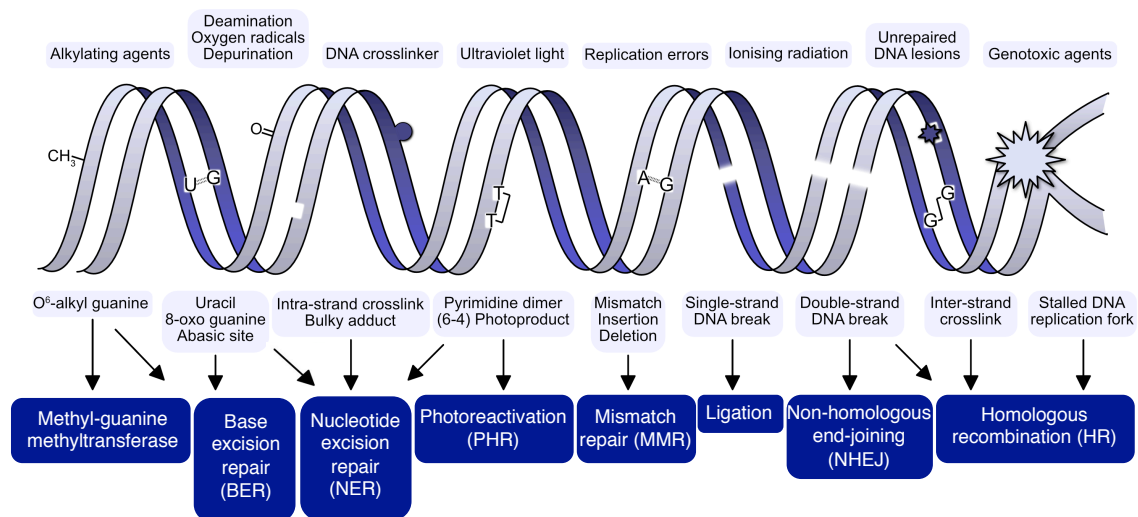


Figure 1: DNA repair pathways.

Summary of the most frequent lesions encountered in the DNA and the different types of repair mechanisms employed to correct them. Adapted from (White & Allers, 2018).

INTRODUCTION

The major pathways included in each group are briefly described below:

- Direct repair:
 - Photoreactivation (PHR): PHR refers to the process of reversing the toxic effects of UV radiation (particularly cyclobutane pyrimidine dimers (CPDs) and 6–4 pyrimidine–pyrimidone photoproducts (6–4PPs)) in a reaction that depends on visible light and a flavin adenine dinucleotide (FAD) cofactor, performed by an enzyme called photolyase. Photolyases can be found in Bacteria, Archaea and Eukarya, although homologues presenting PHR activity have not been found in placental mammals (Weber, 2005).
 - DNA ligation: the joining of separated DNA strands is an essential, universal mechanism required in the replication, excision repair and recombination processes; additionally, a ligation reaction performed to fix a break in the DNA can be considered a form of direct repair. Several classes of adenosine triphosphate (ATP)- and nicotinamide adenine dinucleotide (NAD)-dependent ligases have been identified, frequently involved in repair mechanisms such as excision repair (ligases I and III) (Tomkinson *et al.*, 2001) and recombinational repair (ligase IV) (Grawunder *et al.*, 1998).
 - Alkylation reversal: alkylation of the DNA (both by external and internal factors) leads to lesions that can block replication, cause mutations and even cell death (Sedgwick, 2004). Although guanine bases are the most susceptible to damage by alkylating agents (Mishina *et al.*, 2006), the N⁷-methylguanine lesion is frequently removed spontaneously, leaving an abasic site that will be repaired through a different pathway (Wilson & Barsky, 2001). The vast majority of alkylation-induced lesions can be solved by one of three different enzymatic activities: O⁶-methylguanine-DNA-methyltransferases like the *E. coli* protein Ada (Sedgwick *et al.*, 1988), with no homologues found in eukaryotes so far (Mishina *et al.*, 2006); methyladenil-DNA glycosylases like AlkA (Labahn *et al.*, 1996); and the iron (II)-dependent dioxygenase AlkB (Trewick *et al.*, 2002).

- Recombinational repair
 - Homologous recombination (HR): HR is the preferred pathway to repair double-strand breaks (DSBs) in Archaea, Bacteria, and Eukarya. Intra-strand crosslinks and post-replication daughter strand gaps (DSGs) are also repaired by HR. The process comprises four steps: creation of a 3' single-stranded DNA (ssDNA) substrate (checkpoint for the resolution of the DSB by HR or by non-homologous end-joining (NHEJ) (Kass & Jasin, 2010) capable of invading an undamaged homologous duplex that can act as a template, repair synthesis, branch migration (Jasin & Rothstein, 2013) and resolution of the Holliday junction by the combined action of helicases and structure-specific endonucleases (Mimitou & Symington, 2009).
 - Non-homologous end-joining (NHEJ): NHEJ is mostly a eukaryotic pathway, with only one group of Archaea possessing all the factors required for a complete NHEJ complex (Bartlett *et al.*, 2013). It is also present in some species of Bacteria, where the system is especially helpful at the stationary phase, when only one copy of the genome is available (Matthews & Simmons, 2014). Contrary to HR, NHEJ permits the repair of DSBs without a template strand, and so frequent deletion or addition of nucleotides can happen at the repaired site. Although this flexibility can seem undesired in a repair system, the benefit of NHEJ restoring chromosomal integrity far outweighs the inconvenience of these smaller changes in the DNA (Lieber, 2008).

- Excision repair
 - Base excision repair (BER): the highly conserved BER pathway eliminates single- to a few nucleotide lesions caused by oxidative damage or alkylating agents, among others. The BER reaction occurs in four steps: first, the altered base is detected and removed by a DNA glycosylase. The type of damage will determine which glycosylase carries out the removal of the lesion, which in turn will define the type of BER required (short BER if only one nucleotide has to be resynthesized, long BER if a fragment of 2-6 nucleotides has to be repaired) (Fortini *et al.*,

INTRODUCTION

1999). Second, the glycosylase will cleave the N-glycosylic bond between the base affected by the lesion and the deoxyribose sugar, releasing the damaged base and leaving an apurinic / apyrimidinic site (Krokan *et al.*, 1997). Next, the DNA backbone is nicked by an endonuclease, and a copy to repair the abasic site is subsequently synthesized by a polymerase, with a ligase finally sealing the gap to complete the reaction (Robertson *et al.*, 2009).

- Mismatch repair (MMR): as mismatched bases often occur naturally during DNA replication, MMR is not only a system to repair lesions, but a mechanism to ensure the fidelity of the copy generated. MMR is highly conserved in Bacteria and Eukarya (Eisen & Hanawalt, 1999), with the majority of Archaea lacking an obvious Mutator S - Mutator L (MutS-MutL) system (Kelman & White, 2005). In the MMR pathway, best described in *E. coli*, the MutS homodimer recognizes the lesion in the DNA and interacts with protein MutL, which enhances this damage recognition role and recruits the endonuclease Mutator H (MutH), which will excise the damaged fragment. Finally, the single-stranded gap will be filled by DNA polymerase III, and a ligase will bind the new fragment to the repaired strand to end the reaction (Li, 2008). Recently, an archaeal endonuclease mismatch specific (EndoMS, originally named NucS) has been reported to cleave a range of mismatches on both strands, leaving a DSB and thus connecting the MMR and HR pathways (Ishino *et al.*, 2016).
- Nucleotide excision repair (NER): NER is a complex system that removes bulky adducts in the DNA caused by sources such as UV radiation or genotoxic agents. While the three defining steps of the NER mechanism (damage recognition, unwinding of the DNA duplex around the distortion and removal of the lesion) that lead to the restoration of the double helix are common to Bacteria, Archaea and Eukarya, the systems that perform the reaction are very different in all three (Sugasawa, 2016). This pathway will be discussed at length in the following sections.

1.2. The nucleotide excision repair pathway

The highly conserved NER pathway is a defense mechanism that protects cells from mutagenic and carcinogenic processes by removing an extensive range of bulky adducts and lesions that introduce a significant distortion in the double helix. In Bacteria and Eukarya NER is divided in two sub-pathways, according to the location of the damage: the global genome repair (GGR) pathway removes lesions present anywhere in the genome (Kisker *et al.*, 2013) (Compe & Egly, 2012), while the transcription-coupled repair (TCR) pathway removes alterations occurring in an actively-transcribed DNA strand (Hanawalt & Spivak, 2008), typically in a faster manner than removal of damage in a strand that is not being transcribed (Mellon *et al.*, 1987).

1.2.1. Nucleotide excision repair in Bacteria

The NER process in Bacteria requires only a handful of proteins, as opposed to eukaryotic NER, where the repairosome includes more than twenty different proteins (de Laat *et al.*, 1999). Moreover, although bacteria and eukaryotes share the basic steps comprising the repair mechanism (damage detection, DNA unwinding and removal of the lesion), the genes that perform these steps show no homology whatsoever, suggesting different origins for the two systems (Eisen & Hanawalt, 1999). However, the high degree of conservation observed for the Uvr group of proteins suggests a common NER mechanism for most bacterial species (Peng *et al.*, 2011).

The UvrABC system responsible for the NER pathway in Bacteria (figure 2) was first described in 1983 (Sancar & Rupp, 1983), and its efficiency seems to be directly related to its ability to detect alterations in the double helix rather than specific lesions, as demonstrated by the wide range of substrates repaired by it (Van Houten *et al.*, 2005).

The homodimer UvrA is the first element to detect and bind to the damaged site in the GGR pathway, forming a short-lived complex before passing the DNA to UvrB in an ATP-dependent manner (Stracy *et al.*, 2016). UvrB is a central player in the bacterial NER process, as it interacts with all Uvr proteins. Association with UvrA and the damaged DNA activates its ATPase function, required for later recruitment of UvrC. This interaction

INTRODUCTION

also induces a conformational change in UvrB that will facilitate the insertion of a β -hairpin between the two strands to open the DNA (Theis *et al.*, 2000). A model in which UvrA₂-bound UvrB is actually a dimer has been proposed: this transient UvrA₂B₂ complex now translocates along the undamaged strand, opening the duplex in this way rather than unwinding it. At the same time, it searches for the cause of the distortion: if no actual damage is present, the complex is disbanded, but upon finding the lesion, UvrB promotes the removal of UvrA₂ and a stable UvrB₂-DNA complex is formed (Kisker *et al.*, 2013).

In the next step in the repair process, UvrB recruits the dual endonuclease UvrC to the damaged site, where it will make an incision on both sides of the lesion, removing a fragment of about 13 nucleotides (Verhoeven *et al.*, 2000). The recruitment of UvrC is accompanied by the release of one of the UvrB molecules (Verhoeven *et al.*, 2002).

UvrB also recruits the 3' to 5' helicase UvrD to the incised site. The helicase will unwind the incised fragment, causing the excision of the fragment containing the lesion and the release of UvrC from the UvrBC complex, while UvrB remains attached to the undamaged strand (Ahn, 2000). Next, DNA polymerase (DNApol) I synthesises a new copy of the DNA, using the undamaged strand as a template, and finally causing the removal of UvrB from the now repaired site (Theis *et al.*, 2000). The reaction is finalized when a ligase seals the newly synthesized and the old strands together (Tomkinson *et al.*, 2006).

Bacteria also possess a TCR-NER pathway (figure 2). Here, the mutation frequency decline (Mfd) protein promotes the ATP-dependent release of the RNApol after it becomes stalled when encountering the lesion (Selby & Sancar, 1993). Mfd subsequently recruits UvrA₂ to the damaged site, which will recruit UvrB. Mfd is released together with UvrA₂, leaving the UvrB-DNA complex to complete the following steps as previously described for the GGR pathway (Selby & Sancar, 1993). More recently an alternative, Mfd-independent TCR pathway has been reported, in which helicase UvrD forces the backtracking of the lesion-stalled RNApol, thus exposing the alteration in the DNA and allowing access of the subsequent NER factors to the damaged site (Epshtein *et al.*, 2014).

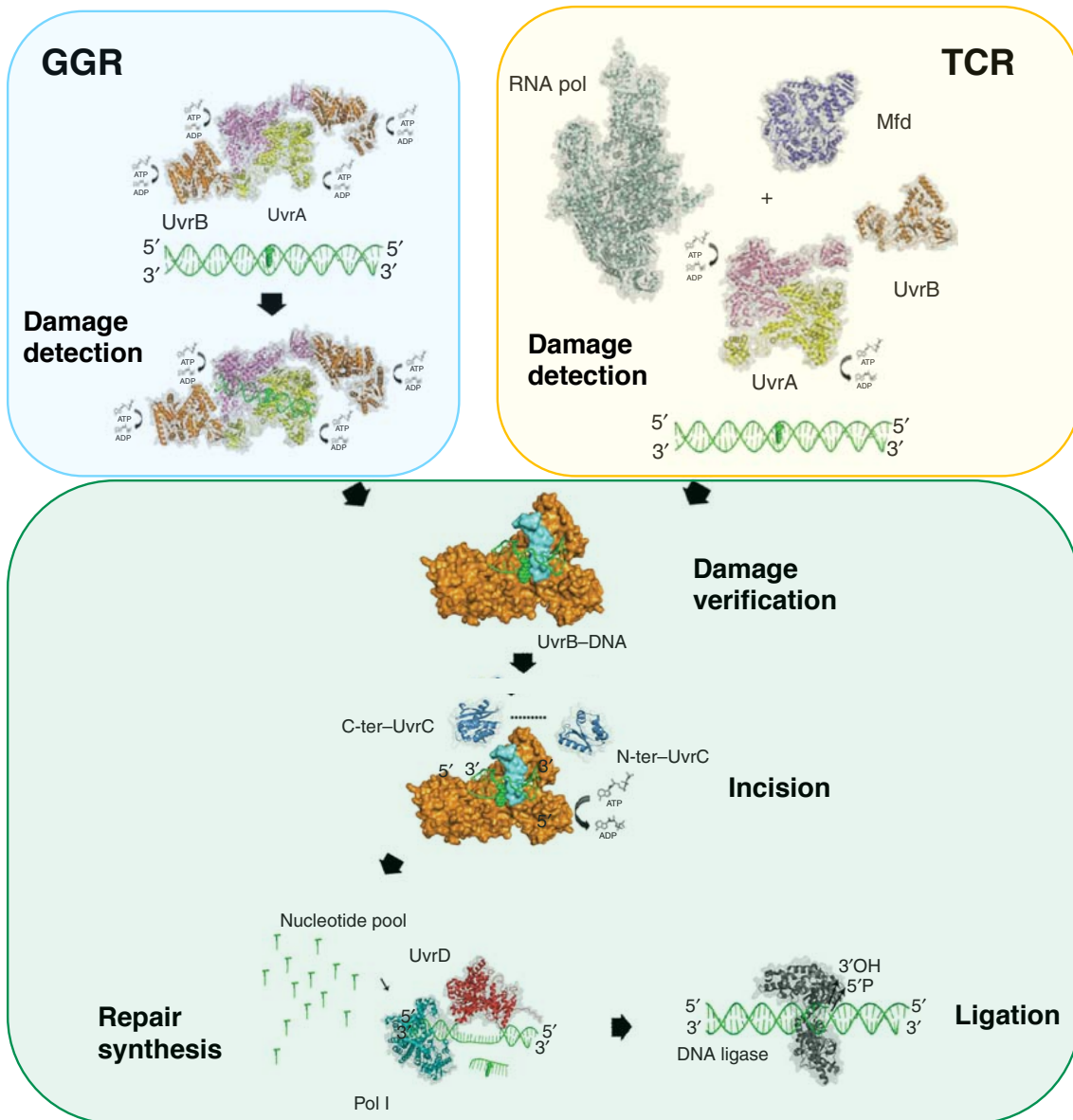


Figure 2: Representation of the GGR and TCR pathways in Bacteria.

Damage is identified by homodimer UvrA (GGR) or by an RNA polymerase (RNAPol) (TCR), with the subsequent steps (damage verification, incision, removal of the damaged fragment, repair synthesis and ligation) shared by both pathways. Adapted from (Kisker *et al.*, 2013).

1.2.2. Nucleotide excision repair in Eukarya

The eukaryotic NER pathway can effectively remove an impressive range of lesions in the DNA, such as damage caused by the UV-light to the pyrimidine bases (CPDs, 6-4PPs), or damage caused by chemically reactive agents (*cis*-diaminodichloroplatinum (cisplatin)) (Gillet & Schärer, 2006). This efficiency to remove such a wide selection of

INTRODUCTION

lesions can be explained by the ability of damage-recognition protein xeroderma pigmentosum complementation group C (XPC) to identify a distortion in the double helix, rather than the actual lesion (Dip *et al.*, 2004) (Maillard *et al.*, 2007) (Naegeli & Sugasawa, 2011).

1.2.2.1. Damage recognition and verification

The eukaryotic GGR pathway is initiated with the identification of the lesion by damage-detector XPC, which recognizes a structural perturbation in the double helix (Sugasawa *et al.*, 2001) and introduces a further distortion in the DNA at the impaired site (Janićijević *et al.*, 2003) (figure 3). Although it was first suggested that XPC is stabilized and its binding capacity stimulated by protein HR23B upon recognition of the lesion (Xie *et al.*, 2004), recent studies propose a model in which the heterodimer XPC-HR23B is the actual damage sensor, with HR23B safely delivering XPC to the lesion and then quickly dissociating from it as the protein is stabilized by binding to the damaged site (Bergink *et al.*, 2012).

It has been observed that CPDs are very poorly recognized by XPC, as the distortion they introduce in the double helix is minimal (Kusumoto *et al.*, 2001). The identification of this particular lesion is additionally supported by the heterodimer DNA damage-binding protein 1 / xeroderma pigmentosum complementation group E (DDB1/XPE) (Scrima *et al.*, 2008), which stably binds to CPDs and provides a platform at the lesion for XPC to access the damaged site so the NER reaction can progress normally (Moser *et al.*, 2005).

This recognition mechanism makes evident the need for a lesion-verification step further down the repair process. Xeroderma pigmentosum complementation group A (XPA) and transcription factor IIH (TFIIH) subunit xeroderma pigmentosum complementation group D (XPD) have both been proposed as candidates to carry out this function, based on the observations that XPD is stalled by bulky adducts in both the translocating and, to a lesser extent, the non-translocating strands (Mathieu *et al.*, 2010) (Buechner *et al.*, 2014). A model in which a ternary XPC-TFIIH-XPA complex translocates along the DNA until it becomes stalled by the lesion was proposed (Sugasawa *et al.*, 2009) and later reinforced by the work of the Yang group, which confirmed that the presence of XPA

enhances the inhibitory nature of these bulky lesions (Li *et al.*, 2015).

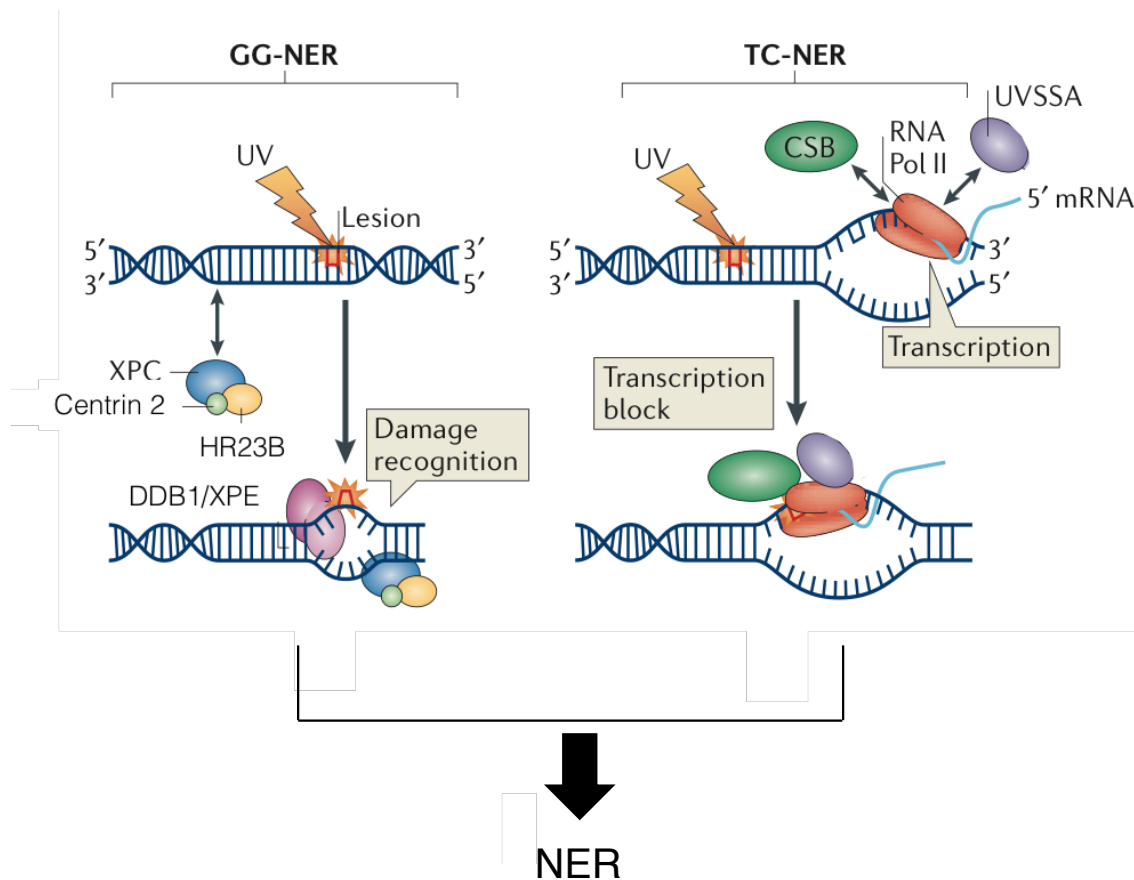


Figure 3: Damage recognition in the eukaryotic GGR and TCR sub-pathways.

The lesion is identified by the XPC-HHR23B-Centrin 2 complex (GGR) or by the stalled RNAPol (TCR). The subsequent steps until completion of NER (damage verification, incision, removal of the damaged fragment, repair synthesis and ligation) are common to both GGR and TCR. Protein Centrin 2 (CETN2) will be discussed in chapter 5. Adapted from (Marteijn *et al.*, 2014).

The correct positioning of XPC at the site of the distortion will determine the recruitment and positioning of all subsequent NER factors, and by extension, the success of the repair reaction itself. The next repair factor, the 10-subunit TFIIH complex, joins the damaged site through interactions between XPC's C-terminal module and subunits p62 and xeroderma pigmentosum complementation group B (XPB) within the TFIIH complex (Yokoi *et al.*, 2000) (Uchida *et al.*, 2002) (Bernardes de Jesus *et al.*, 2008).

By contrast, the TCR pathway removes adducts blocking the translocation of RNAPol II

INTRODUCTION

during the transcription of a gene (Hanawalt & Spivak, 2008). A link between transcription and DNA repair was proposed by the Hanawalt group based on their observations that damage in an actively-transcribed strand was repaired considerably faster than a lesion located in a non-transcribed strand (Hanawalt, 1987). Years later, the Egly group determined that that link was the TFIIH complex, which is essential for both the transcription and DNA repair processes (Schaeffer *et al.*, 1993). In this sub-pathway of NER, it's the stalled RNAPol II itself that acts as a damage sensor with the aid of the Cockayne Syndrome complementation group B (CSB) protein, a DNA-dependent ATPase with a loose interaction with the RNAPol (Selby & Sancar, 1997) that will be strengthened following transcription arrest (Van Gool *et al.*, 1997). The next factor to join the repair bubble is the TFIIH complex, recruited to the damaged site by the UV stimulated scaffold A (UVSSA) coupling factor, which interacts with subunit p62 in a remarkably similar manner as GGR damage-detection protein XPC (Okuda *et al.*, 2017) (figure 3). The subsequent steps ensuing the verification of the distortion in the DNA and removal of the damage following recruitment of TFIIH are common to both GGR and TCR (Compe & Egly, 2012) (Compe & Egly, 2016) (Hanawalt & Spivak, 2008).

1.2.2.2. Opening of the repair bubble

The 10-subunit TFIIH complex (composed of sub-complexes TFIIH Core and cyclin-dependent kinase (CDK)-activating kinase (CAK)) is recruited to the damaged site to open the double helix around the lesion. Our understanding of how this mechanism occurs has been aided by the studies of the structures and mechanisms of action of several archaeal homologues of helicases XPB (Fan *et al.*, 2006) and XPD (Liu *et al.*, 2008) (Wolski *et al.*, 2008) (Fan *et al.*, 2008).

The coordinated actions of XPB and XPD will open and extend the repair bubble to allow the following repair factors access to the damaged site (figure 4). An initial model for the opening was proposed in which a DNA-stimulated, ATP-dependent conformational change in XPB would anchor TFIIH to the DNA and facilitate the separation of the two strands, allowing XPD to unwind the duplex (Oksenysh *et al.*, 2009). The fact that only XPB's ATPase activity is essential for NER (Coin *et al.*, 2007) (Oksenysh *et al.*, 2009) and that archaeal XPB is able to open DNA substrates by translocating along the undamaged strand rather than unwinding it (Rouillon & White, 2010) support the more

recent proposal that eukaryal XPB also acts as a translocase rather than displaying a traditional helicase activity (He *et al.*, 2016) (Schilbach *et al.*, 2017). The ATPase activity of XPB is regulated by damage sensor XPC (Bernardes de Jesus *et al.*, 2008), and TFIIH subunits p52 (Coin *et al.*, 2007) and p8 (Coin *et al.*, 2006).

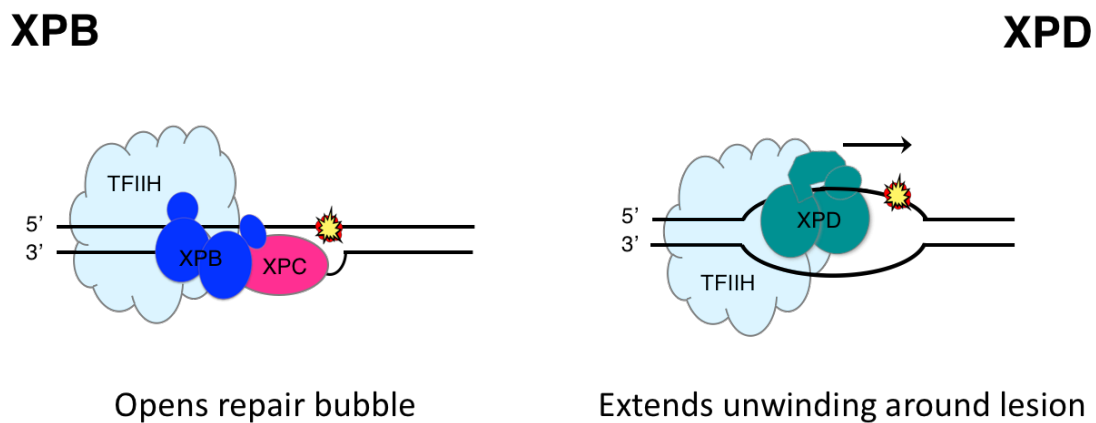


Figure 4: The TFIIH complex opens the DNA around the damaged site.

ATP-dependent 3' to 5' helicase XPB will open the DNA as it translocates through the duplex, allowing ATP-dependent 5' to 3' helicase XPD to unwind the DNA, thus extending the repair bubble.

As opposed to XPB, XPD's helicase activity is fundamental for NER, as evidenced by the fact that mutations affecting its helicase motifs or its region of interaction with regulatory subunit p44 impair the ability of TFIIH to unwind the DNA (Coin *et al.*, 1998) (Dubaele *et al.*, 2003). The initial mechanism proposed for unwinding of the DNA by XPD postulated that the helicase domain 1 (HD1), the Arch domain and the 4FeS cluster in the helicase form a ring that translocates ssDNA in an ATP-dependent manner through the channel left by the structure (Liu *et al.*, 2008) (Wolski *et al.*, 2008) (Fan *et al.*, 2008) (Kuper *et al.*, 2012). A more recent study proposes an update on this mechanism, with ssDNA stably interacting with the helicase domain 2 (HD2) and the Arch domain and 4FeS cluster opening transitorily at their interface, to allow passage of the ssDNA through the pore and posterior binding to HD1 (Constantinescu-Aruxandei *et al.*, 2016) (figure 5). XPD also plays a structural role within TFIIH, as it anchors the CAK sub-complex to the TFIIH Core sub-complex to form the complete 10-subunit TFIIH

INTRODUCTION

(Adamczewski *et al.*, 1996) (Schultz *et al.*, 2000) (Abdulrahman *et al.*, 2013). A third role for the helicase as a damage-verification factor has also been suggested (Sugasawa *et al.*, 2009) (Kuper *et al.*, 2012) (Li *et al.*, 2015), but not verified up to date. XPD's helicase activity is stimulated by interaction with subunits p34 and p44 (Coin *et al.*, 1998) (Schmitt *et al.*, 2014).

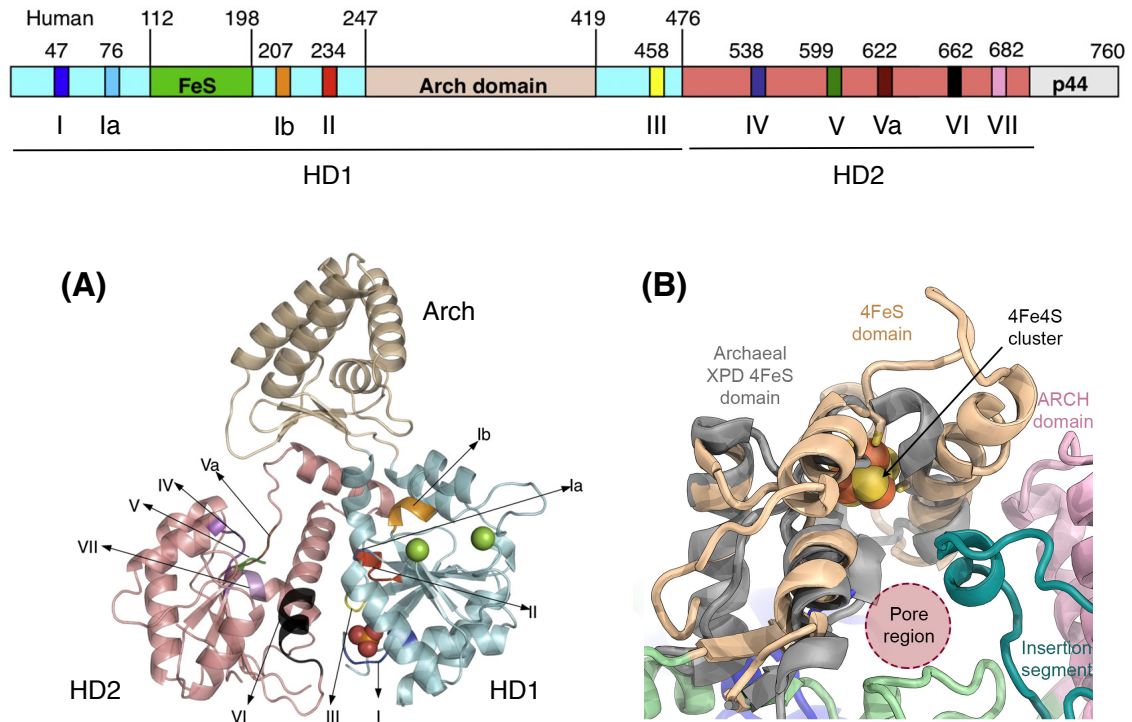


Figure 5: Structure of the TFIIF 5' to 3' helicase XPD.

Structure of the *Sulfolobus tokodaii* archaeal XPD, showing helicase domains 1 and 2 and the Arch domain. The 4FeS domain area is delimited by the green spheres (A) (adapted from (Liu *et al.*, 2008)). The Arch domain, 4Fe4S cluster and HD1 (not shown in the figure) form a channel that translocates ssDNA after opening of the interface between the Arch and the 4FeS domains (B) (Constantinescu-Aruxandei *et al.*, 2016) (XPD structure from (Greber *et al.*, 2017)).

Subunit p62 was thought to have a merely structural role, and although a recent study further supports this role by describing a new region in the protein that interacts with XPD and binds the helicase to the Core sub-complex (Luo *et al.*, 2015), p62 is also responsible for an extended network of contacts with other transcription factors and proteins, particularly with damage-detectors XPC in GGR (Okuda *et al.*, 2015) and

UVSSA in TCR (Okuda *et al.*, 2017), as this binding will recruit complex TFIIH to the repair bubble.

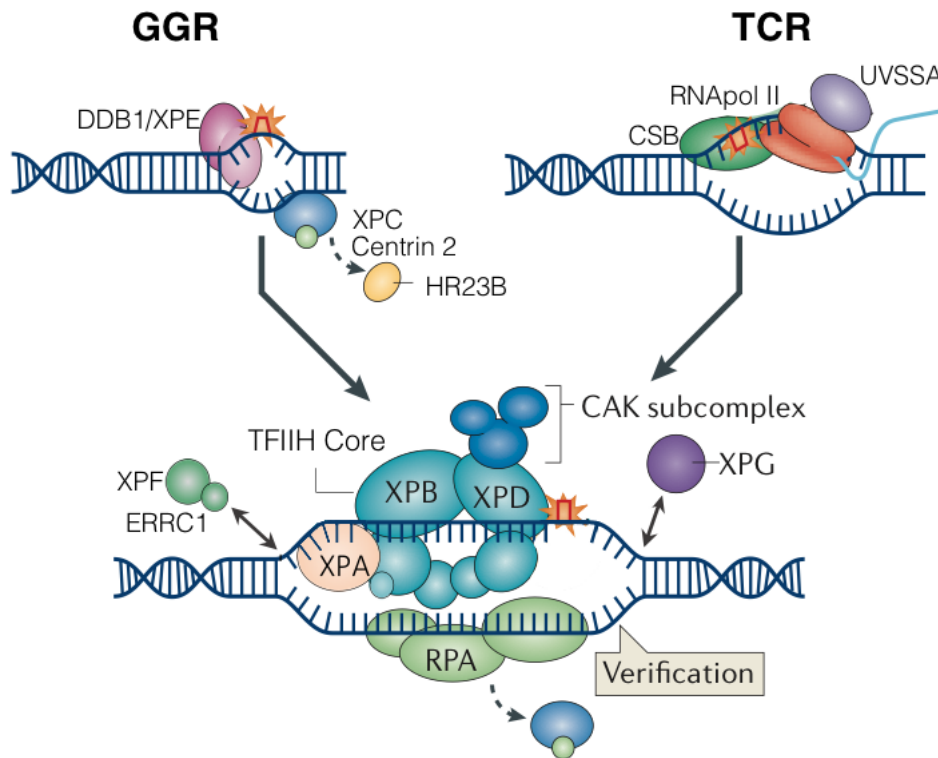


Figure 6: Damage verification and DNA unwinding in eukaryotic NER.

The GGR and TCR sub-pathways converge after the detection of the lesion. The TFIIH complex is recruited to open the DNA duplex; next, factors RPA and XPA are recruited to protect the ssDNA from degradation, and to promote the release of the CAK sub-complex and recruit endonucleases XPG and XPF/ERRC1, respectively. Factors in this figure are not drawn to scale. Adapted from (Marteijn *et al.*, 2014).

The heterotrimeric replication protein A (RPA) and XPA proteins are now recruited together to the XPC-TFIIH-DNA complex (Krasikova *et al.*, 2010), promoted by TFIIH subunit p8 through stimulation of the ATPase activity of XPB (Coin *et al.*, 2006), which will displace XPC and reorganize the bubble for the arrival of subsequent factors (Tapias *et al.*, 2004). RPA binds to ssDNA to protect it from degradation (Krasikova *et al.*, 2010), while XPA promotes the ATP-dependent removal of the CAK sub-complex from TFIIH in a reaction that is essential for the successful progress of NER (Svejstrup *et al.*, 1995) (Coin *et al.*, 2008), as the sub-complex reportedly regulates negatively the helicase activity of XPD (Sandrock & Egly, 2001) (Abdulrahman *et al.*, 2013). CAK is thought to

be released but not degraded, making the sub-complex available to bind to TFIIH Core again once transcription is resumed (Coin *et al.*, 2008). The asymmetrical binding of XPA to the 5' end of the repair bubble will also determine the recruitment of endonucleases xeroderma pigmentosum complementation group G (XPG) and xeroderma pigmentosum complementation group F / excision repair cross-complementation group 1 (XPF/ERCC1) (Krasikova *et al.*, 2010) (figure 6).

1.2.2.3. Removal of the damaged fragment

The recruited endonucleases will carry out nearly simultaneous but independent incisions at the 3' (XPG) and the 5' (XPF/ERCC1) sides of the lesion (Evans *et al.*, 1997). These incisions are asymmetrical, as XPG cuts the DNA 4-8 nucleotides away from the damage site while the XPF/ERCC1 incision occurs 15-24 nucleotides away from the lesion (Peng *et al.*, 2011) (figure 7). Incision by XPF/ERCC1 requires the recruitment (but not the catalytic activity) of XPG (Tapias *et al.*, 2004), whose binding promotes the release and recycling of heterodimer XPC/HR23B (Riedl *et al.*, 2003). Repair factors TFIIH and XPA will also be released from the repair complex soon after XPF/ERCC1 excises the DNA at the 5' end, while RPA remains to protect the ssDNA portion of the bubble (Riedl *et al.*, 2003). XPG and RPA now promote the recruitment of repair factors proliferating cell nuclear antigen (PCNA) (Gary *et al.*, 1997) and replication factor C (RFC) (Yuzhakov *et al.*, 1999) to support and ensure the correct synthesis of the new DNA fragment; this recruitment will also trigger the release of XPF/ERCC1 (Staresincic *et al.*, 2009).

1.2.2.4. Synthesis and ligation of the new fragment

Repair synthesis can be detected as soon as XPF/ERCC1 excises the DNA at the 5' end (Staresincic *et al.*, 2009). At this stage the repair complex provides a platform for a DNApol ϵ , or a DNApol δ and κ acting together, to join the repair process and start the synthesis of a new fragment, using the undamaged strand as a template (Ogi *et al.*, 2010) (figure 7). Finally, endonuclease XPG excises the 3' of the damaged strand, and as a result a damaged fragment of about 30 nucleotides (Gillet & Schärer, 2006) and repair factors XPG and RPA are released from the complex (Mocquet *et al.*, 2008). Once the DNA synthesis is completed, a DNA ligase III (for DNApol δ) or a ligase I (for DNApol ϵ) is recruited to seal the new fragment (Moser *et al.*, 2007) (figure 7) as RFC and the

DNApol are released (Mocquet *et al.*, 2008). Repair factor PCNA still remains at the repair bubble, acting as a bridge between the synthesis and ligation processes (Gary *et al.*, 1997). Finally, as soon as the NER reaction is complete the chromatin structure is restored to its original packed state (Gaillard *et al.*, 1996) (Polo *et al.*, 2006).

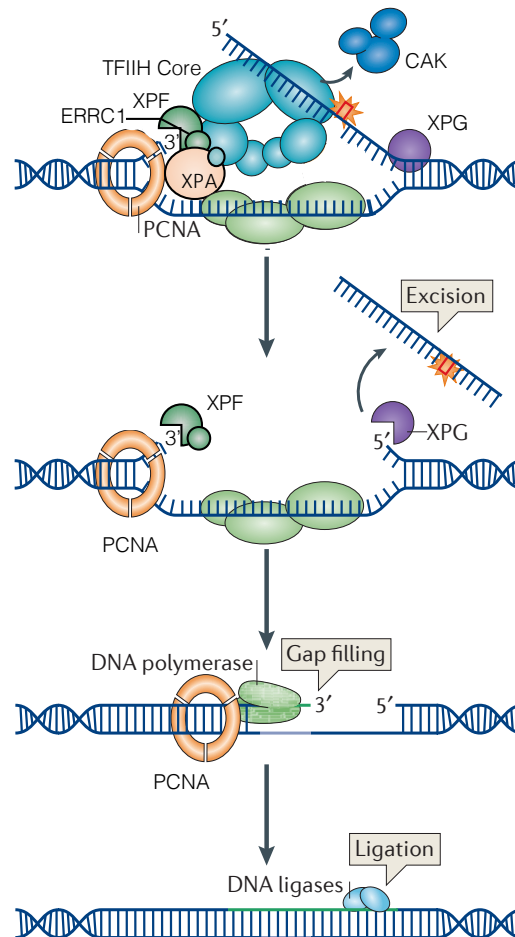


Figure 7: Excision and repair synthesis in NER.

XPA promotes the release of the CAK sub-complex so the combined action of the XPB and XPD helicases can open the DNA duplex at the damaged site. Endonucleases XPF/ERCC1 and XPG will excise both sides of the damaged fragment, releasing a fragment of about 30 nucleotides as a DNApol synthesizes a new fragment and a ligase seals the nicks. The reaction concludes with the restoration of the chromatin structure. Factors in this figure are not drawn to scale. Adapted from (Marteijn *et al.*, 2014).

INTRODUCTION

1.2.2.5. Recessive disorders associated with NER

Mutations affecting the different factors involved in NER are responsible for a variety of serious, sometimes fatal genetic syndromes that present a remarkable heterogeneity in their genetic make-up and clinical manifestations. Table 1 summarises these disorders and their most prominent features. The seven XP complementation groups (XP-A to XP-G) present defects in NER, with all of them showing the characteristic photosensitivity and increased susceptibility to cancer associated to DNA repair defects (Peng *et al.*, 2011). By contrast, CS and TTD patients do not suffer a higher frequency of cancer, but they present an extended variety of mental and physical developmental defects (Lehmann, 2003) that match defects in transcription rather than impairment of the repair process (Dubaele *et al.*, 2003) (Bootsma & Hoeijmakers, 1993).

Table 1: Autosomal recessive disorders caused by mutations in NER factors

(Lehmann, 2003) (Peng *et al.*, 2011) (Diderich *et al.*, 2011).

Disorder	Symptoms	Elevated cancer	Protein(s) affected
UV Sensitivity Syndrome (UV ^{SS})	Photosensitivity Mild skin abnormalities	No	CSB
Cockayne syndrome (CS)	Severe growth retardation Neurological dysfunction Photosensitivity Ocular abnormalities Microcephaly Dwarfism	No	CSB CSA XPB XPD XPG
Xeroderma pigmentosum (XP)	Extreme photosensitivity Photophobia Skin abnormalities Dry, parchment-like skin Neurological abnormalities	Yes	XPA - G

Trichothiodystrophy (TTD)	Brittle hair	No	XPB
	Scaly skin		XPD
	Severe mental retardation		p8
	Developmental retardation		
	Abnormal facies		
	Photosensitivity		
Cerebro-oculo-facio- skeletal syndrome (COFS)	Microcephaly	No	CSB
	Ocular abnormalities		XPB
	Developmental retardation		XPD
	Severe growth failure		XPG
	Facial dysmorphism		
De Sanctis-Cacchione syndrome	Neurological degeneration	No	CSB
	Developmental retardation		XPA – G
	Dwarfism		
	Hypogonadism		
	Facial freckling		

A wide variety of non-lethal mutations have been described for helicase XPD, while only a few have been reported in relation to XPB, consistent with its essential role in transcription (Tirode *et al.*, 1999). The great majority of the pathological mutations affecting XPD are localized in its C-terminal end, affecting either XPD's binding to p44 or the helicase motifs, thus reducing its helicase activity dramatically (Dubaele *et al.*, 2003).

The majority of the mutations affecting TFIIH subunits XPB, XPD and p8 that cause TTD result in a drastic reduction of TFIIH's cellular levels (Botta *et al.*, 2002), which paired with the enormous variability in the severity of the pathological characteristics presented by these patients led to conclude that these mutations had to not only affect the steady-state levels of TFIIH, but also the complex's stability (Giglia-Mari *et al.*, 2006) and its role in transcription (Botta *et al.*, 2002).

1.2.2.6. The NER pathway as a drug target

The ability of repair pathways to remove DNA damage caused by chemotherapeutic agents have turned them into a target for novel cancer treatments, as the use of repair inhibitors would potentially improve the efficiency of the current strategies, burdened by the growing resistance to anti-cancer drugs (Helleday *et al.*, 2008).

These novel strategies have placed particular focus on NER, and more specifically on the TFIIH complex (figure 8), as it is involved in the removal of bulky adducts on the DNA caused not only by UV radiation, but also by chemicals like the widely-used chemotherapeutic drug cisplatin (Rosenberg *et al.*, 1969) (Fuentes *et al.*, 2003) (Rabik & Dolan, 2007). The fact that cisplatin-induced DNA damage is processed and eliminated by NER was confirmed by a study showing that increased levels of the TFIIH subunit XPB are connected to resistance to cisplatin in primary ovarian cancer (Dabholkar *et al.*, 2000). The involvement of NER in resistance to cisplatin treatments was further confirmed by the presence of low levels of both XPA and XPF-ERCC1 in testicular cancer, a form of the disease particularly responsive to treatment with cisplatin (Welsh *et al.*, 2004).

Although the recent publication of a study that successfully overcame resistance to the anti-cancer drug melphalan by inhibiting TFIIH helicase XPB in multiple myeloma cancerous cells (Szalat *et al.*, 2017) confirms that the use of DNA repair inhibitors (in combination with classical approaches or as monotherapy alternatives) as agents in cancer treatment is a real possibility (Helleday *et al.*, 2008) (Zurita & Cruz-Becerra, 2016) (Berico & Coin, 2017) (Desai *et al.*, 2018), the need for drugs that selectively target tumorous cells is a setback that cannot be ignored. Although there's an argument that overlapping between repair pathways could compensate the activity of the inhibited pathway in non-cancerous cells, that same crosstalk could result in resistance to the new treatment. For now, alternative cancer therapies using DNA repair inhibitors remains a challenging field.

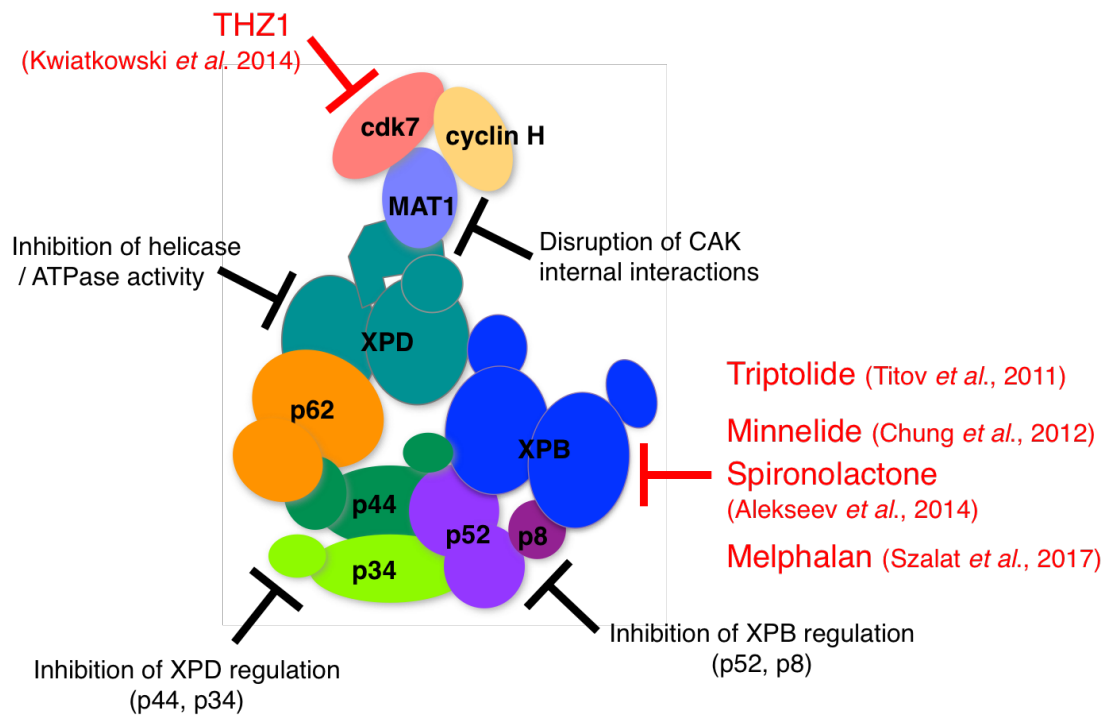


Figure 8: The TFIIH complex as a target for chemotherapeutic drugs.

Inhibitors whose effectiveness has already been described are indicated in red, while potential targets for alternative therapeutic drugs are indicated in black. Adapted from (Zurita & Cruz-Becerra, 2016) (Berico & Coin, 2017) (Titov *et al.*, 2011) (Chung *et al.*, 2012) (Alekseev *et al.*, 2014) (Szalat *et al.*, 2017) (Kwiatkowski *et al.*, 2014).

1.2.3. Nucleotide excision repair in Archaea

The NER pathway in archaea is often confusing, with species presenting bacterial-like genes, eukaryal-like ones or a mixture of both (Rouillon & White, 2011).

Although initial studies proposed a bacterial-like repair system for Archaea (Ögrünç *et al.*, 1998), only a small group (mesophilic methanogens and halophiles) actually possess the UvrABC genes, while most of them present clear orthologues for the eukaryotic NER proteins XPB, XPD, XPG and XPF (figure 9) (White, 2003). Furthermore, the recent description of the ASGARD archaeal lineages (Eme *et al.*, 2017), which present the eukaryotic DNA repair orthologues one would expect in an ancestor, support an evolutionary pattern in which Eukarya emerged from this branch of the Archaea (White & Allers, 2018).

INTRODUCTION

Although no orthologues of XPC or XPA have been described for Archaea, the single-stranded DNA binding protein (SSB) from *Sulfolobus solfataricus* was reportedly able to identify destabilizations in the double helix caused by a series of lesions, including substrates repaired by NER (figure 9), and a model in which *Sso*SSB could potentially detect the damage and recruit the subsequent repair factors through interaction with its C-terminal tail has been proposed (Cubeddu & White, 2005).

Archaeal XPD is a monomer in solution, with no reported partners. The helicase presents a 4FeS cluster (Rudolf *et al.*, 2006), coordinated by four highly conserved cysteine residues, that has an essential structural and functional role, as evidenced by the pathological phenotypes resulting from mutations in the enzyme (White & Dillingham, 2012). The available structures of different archaeal XPDs (Liu *et al.*, 2008) (Wolski *et al.*, 2008) (Fan *et al.*, 2008) have helped enormously in the understanding of the eukaryotic XP, CS and TTD phenotypes, as they have allowed the mapping of these mutations to regions that are essential for the function of the helicase.

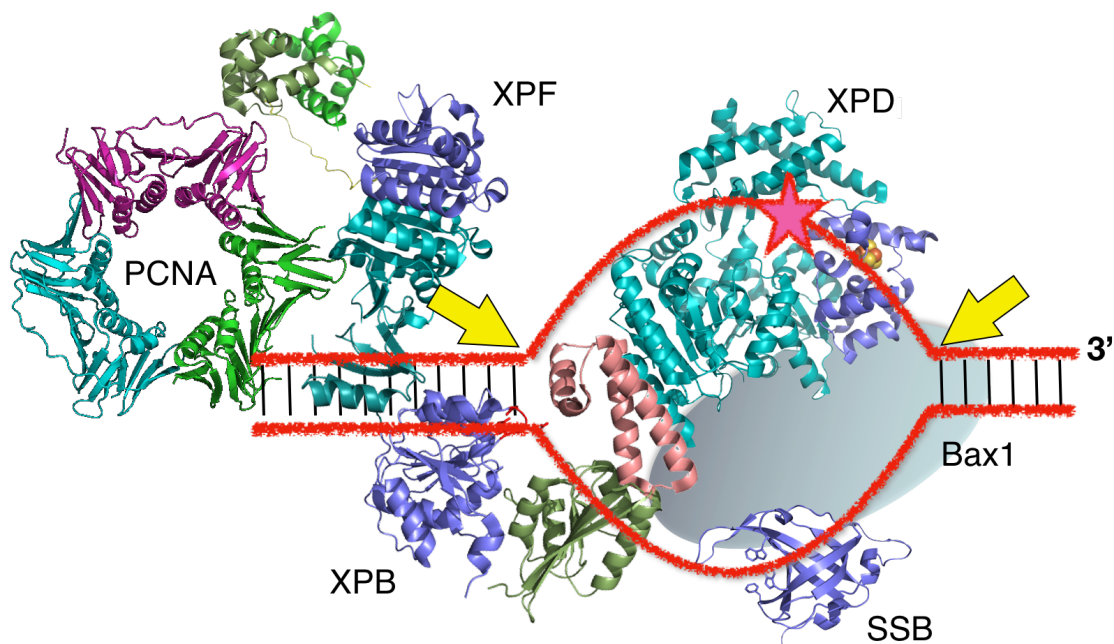


Figure 9: The NER pathway in Archaea.

Structures of putative archaeal NER orthologues PCNA (*Sulfolobus solfataricus*), XPF (*Aeropyrum pernix*), XPD (*Thermoplasma acidophilum*), SSB (*Sulfolobus solfataricus*) and XPB (*Archaeoglobus fulgidus*) (structure of Bax1 yet unknown). The star at the bubble represents a lesion in the DNA. Adapted from (Rouillon & White, 2011).

The archaeal monomeric, weak helicase XPB (Fan *et al.*, 2006) is often found in a complex with endonuclease Bax1, a potential equivalent to eukaryotic XPG, working together towards the extending of the repair bubble and the excision of the damaged fragment (figure 9) (Rouillon & White, 2010). The mechanism of action of the archaeal XPB supports recent findings that report a role of the helicase in NER as a double-strand DNA (dsDNA) translocase rather than the helicase function first assumed (He *et al.*, 2016).

Several studies have shown that deletion of the XPB and XPD genes have little to no effect on the phenotype of the archaea, suggesting that, contrary to what happens in eukaryotes, these genes are not essential for archaeal NER, raising the question of which alternative function they might have in this group, or whether there are multiple, overlapping pathways for NER (White & Allers, 2018).

Although an RNAPol has been reported to be stalled by damage in the DNA, the lack of a Mfd orthologue or any other factors coupling the transcription and DNA repair processes make difficult to define a TCR pathway in archaea (White & Allers, 2018).

1.3. TFIIH: a dual transcription and DNA repair factor

TFIIH is a ten-subunit complex with a dual role in NER and RNAPol II-mediated transcription initiation (Flores *et al.*, 1992) (Feaver *et al.*, 1993). Despite their differences, TFIIH carries out the same basic role in the two processes: unwinding the DNA to allow other transcription or NER factors access to the double helix (Holstege *et al.*, 1996) (Kuper *et al.*, 2014). The central role of TFIIH in both processes is evidenced by the pathological phenotypes caused by mutations in the XPB, XPD or p8 subunits, with symptoms that range from mild photosensitivity to a 2000-fold increased susceptibility to cancer (Bootsma & Hoeijmakers, 1993). The role of TFIIH as a DNA repair factor has been extensively discussed in section 1.2.2.2, so this analysis will focus on its function as a transcription factor.

1.3.1. Structural study of TFIIH

The first published structure of TFIIH revealed a complex with two distinctive regions: a Core sub-complex that included subunits XPB, p62, p52, p44 and p34, and a CAK sub-complex comprising subunits cyclin-dependent kinase 7 (cdk7), cyclin H and ménage-à-trois 1 (MAT1), with subunit XPD positioned at the interface of both sub-complexes (Schultz *et al.*, 2000) (Abdulrahman *et al.*, 2013) (figure 10). This structure lacked any mention to subunit p8, which was not described until 2004 (Giglia-Mari *et al.*, 2004). The recent publication of the electron cryomicroscopy (cryoEM) structure of human TFIIH confirmed the previously described ring-like disposition of the Core sub-complex with the CAK sub-complex protruding from this Core, allowing the unambiguous allocation of most subunits (except p62) and XPD within the ring (Greber *et al.*, 2017).

TFIIH adopts different dispositions for its distinct roles: while the complete ten-subunit complex participates in transcription, only the Core sub-complex is required to carry out the repair reaction (Svejstrup *et al.*, 1995) (Tirode *et al.*, 1999). The Core sub-complex presents two enzymatic subunits. The activity of both ATP-dependent helicases XPB (3' to 5' polarity) and XPD (5' to 3' polarity) is essential for NER (Coin *et al.*, 1998) (Oksenysh *et al.*, 2009). Interestingly, the helicase activity of XPD is not required during transcription, although its role in anchoring the CAK sub-complex to TFIIH Core is necessary for an optimal reaction (Tirode *et al.*, 1999) (Kuper *et al.*, 2014). The third enzymatic activity of TFIIH is in the CAK sub-complex: subunit cdk7 phosphorylates the C-terminal domain (CTD) of Rpb1, the largest subunit of RNAPol II, to facilitate promoter escape in a reaction that is fundamental to move onto the elongation phase of transcription (Lu *et al.*, 1992) (Roy *et al.*, 1994) (Wong *et al.*, 2014). CAK also acts as a down-regulator of XPD's helicase activity (Sandrock & Egly, 2001) (Abdulrahman *et al.*, 2013); because of this, the sub-complex is removed from TFIIH at an early stage when the complex participates in NER (Svejstrup *et al.*, 1995) (Coin *et al.*, 2008). The remaining TFIIH subunits seem to have a structural or regulatory function.

Subunit p8 was described as recently as 2004. Due to its small size, p8 was typically run off the bottom of SDS-PAGE gels, and it was only "discovered" in response to the search for a TFIIH component responsible for a form of TTD not caused by mutations in the XPB or XPD genes (Giglia-Mari *et al.*, 2004). The role of p8 within TFIIH is two-fold: it

regulates the cellular levels of TFIIH (either by establishing the complex or by protecting it from degradation) (Giglia-Mari *et al.*, 2004) and it stimulates XPB's ATPase activity, probably through interaction with subunit p52 (Coin *et al.*, 2006). XPB's ATPase activity is also directly stimulated by subunit p52 (Coin *et al.*, 2007).

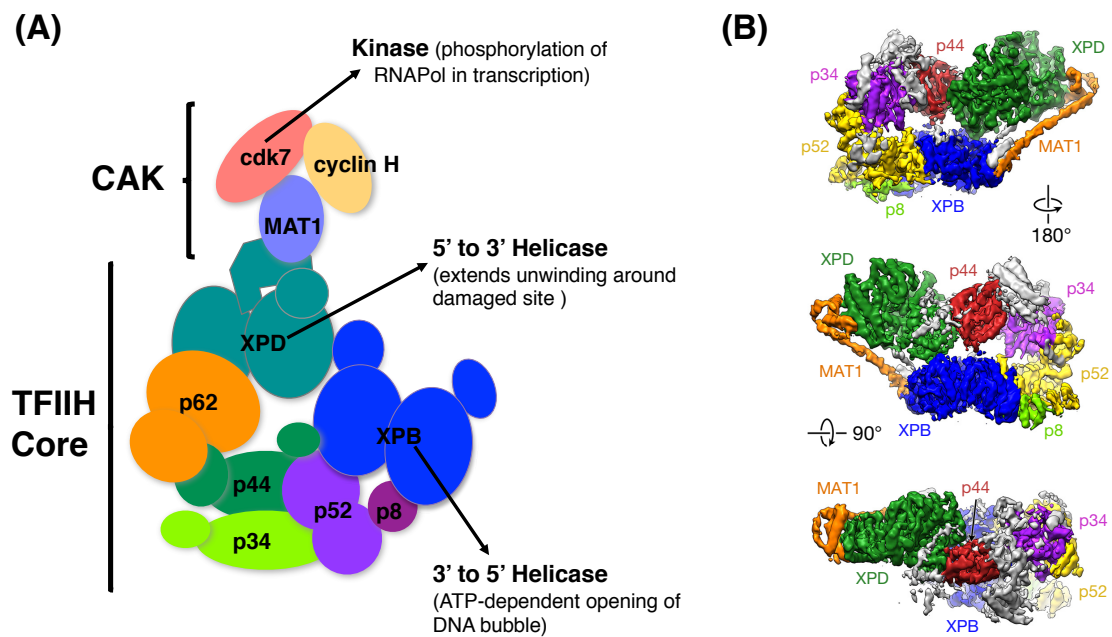


Figure 10: Structure of the 10-subunit complex TFIIH.

Cartoon view showing the CAK and Core sub-complexes that integrate TFIIH (A), highlighting its three enzymatic activities: cdk7 (kinase), XPB (3' to 5' helicase) and XPD (5' to 3' helicase). CryoEM structure of the TFIIH Core sub-complex plus CAK subunit MAT1 (B). The grey density has not been assigned, although it is thought to represent subunit p62 (Greber *et al.*, 2017).

Subunit p44 has been shown to stimulate the helicase activity of XPD (Coin *et al.*, 1998), possibly in collaboration with subunit p34, as hinted by their tight interaction (Schmitt *et al.*, 2014). Protein Ssl1 (the yeast homologue of p44) has been identified as an E3 ubiquitin ligase, reportedly stimulated by protein Tfb4 (the yeast homologue of human p34), but this function has not been confirmed for their human counterparts to date (Takagi *et al.*, 2005).

Subunit p62 is another multi-function component of TFIIH. During NER, p62 is responsible for the recruitment of the TFIIH complex to the damaged site through

INTRODUCTION

interaction with lesion-recognition protein XPC (Okuda *et al.*, 2015). The work of Luo *et al.* described a new region in p62 that extensively interacts with XPD, which led them to propose a further role of the subunit as an extra anchor point between the helicase and the Core sub-complex (Luo *et al.*, 2015). Moreover, p62 is responsible for the interaction between TFIIH and transcription factor IIE (TFIIE) (Okuda *et al.*, 2008), and it's also the target of several viral proteins (this role will be further discussed in section 1.3.4).

1.3.1.1. XPD takes part in three different complexes

XPD is an extremely versatile protein, as proven by its multiple roles (figure 11). Bound to the TFIIH Core sub-complex, it has an essential role in NER (Kuper *et al.*, 2014), and also possesses a structural role as the element that anchors sub-complex CAK to TFIIH Core, providing a platform for the cdk7 subunit to play its fundamental role in transcription together with subunit XPB (Abdulrahman *et al.*, 2013).

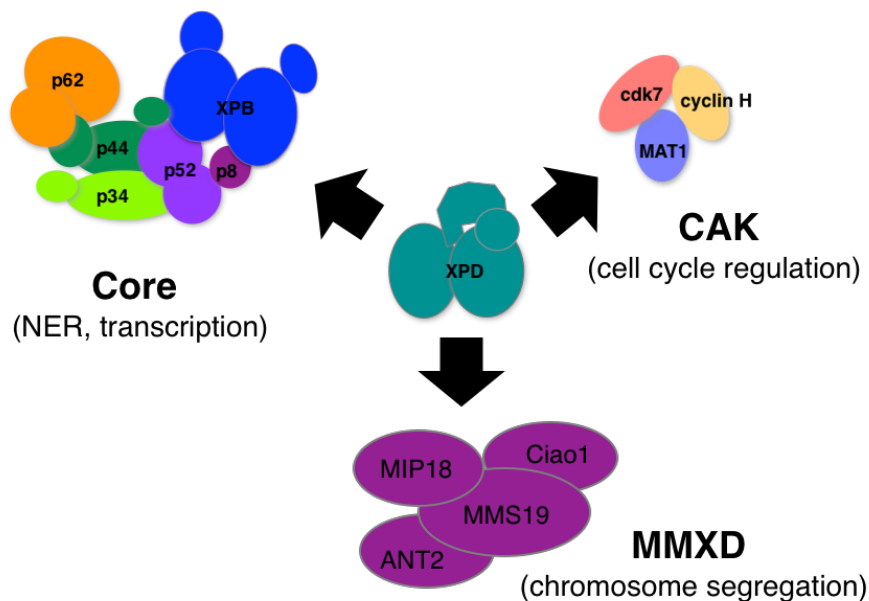


Figure 11: Different roles associated to helicase XPD.

XPD participates in NER and transcription as part of the TFIIH complex, in cell cycle regulation when bound to the TFIIH-independent CAK complex and in chromosome segregation as part of the MMXD complex. Adapted from (Ito *et al.*, 2010).

The CAK sub-complex is not a mere partner of TFIIH Core. As an independent complex, its cdk7 subunit phosphorylates a plethora of other kinases, triggering vital changes for the progression of the cell cycle (Morgan, 1995). In this context, XPD has been reported to regulate both the cell's entrance into mitosis and basal transcription by sequestering or releasing the CAK sub-complex from TFIIH, thus determining the role that cdk7 will play (Chen *et al.*, 2003).

XPD also has an additional function as part of the mitotic spindle-associated MMS19-MIP18-XPD (MMXD) complex (Van Houten *et al.*, 2016). The human homologue of the *Saccharomyces cerevisiae* protein MMS19 reportedly interacts with helicases XPB and XPD, and a tentative regulatory or structural role of the protein in relation to the TFIIH complex was initially proposed (Seroz *et al.*, 2000). A decade later, it was found that XPD actually forms a separate, TFIIH-independent complex that also includes proteins MMS19, MMS19-Interacting Protein of 18 kDa (MIP18), cytosolic iron-sulfur assembly component 1 (CIAO1) and adenine nucleotide translocase 2 (ANT2) (Ito *et al.*, 2010). The MMXD complex has a reported role in chromosome segregation, as knockdown of the MMS19, MIP18 or XPD proteins caused severe alterations of the mitotic spindle (Ito *et al.*, 2010). Interestingly, the MMS19 knocked-down cells present hypersensitivity to UV radiation, suggesting a possible interaction with XPD in the repair process (Ito *et al.*, 2010). The MMS19, MIP18 and CIAO1 proteins also form a cytoplasmic iron-sulfur cluster assembly (CIA) complex, involved in the integration of an 4FeS cluster into XPD prior to its incorporation to the TFIIH complex (Vashisht *et al.*, 2015).

1.3.2. Role of TFIIH in transcription

The transcription process starts with the formation of the pre-initiation complex (PIC) at the promoter (Schilbach *et al.*, 2017). The sequential assembly of the PIC begins with the binding of transcription factor IID (TFIID) to the DNA, stabilized by the subsequent recruitment of transcription factor IIA (TFIIA) and transcription factor IIB (TFIIB). TFIIB promotes now the recruitment of a pre-formed complex RNAPol II - transcription factor IIF (TFIIF). Transcription factor IIE (TFIIE) will be recruited to this stable complex by direct interactions with RNAPol II. Lastly, TFIIE will mediate the recruitment of TFIIH to complete the PIC (Orphanides *et al.*, 1996) (Roeder, 1996) (figure 12).

INTRODUCTION

As in NER, TFIIH plays a central role in transcription as the factor responsible for melting the promoter: the ATPase activity of XPB will be employed to open the DNA around the promoter upstream of the starting site, allowing the RNAPol access to the template strand (Holstege *et al.*, 1996). Recent studies propose that XPB actually melts the promoter by acting as a translocase, separating the duplex as it threads the DNA towards RNAPol II in an ATP-dependent manner (He *et al.*, 2016) (Schilbach *et al.*, 2017).

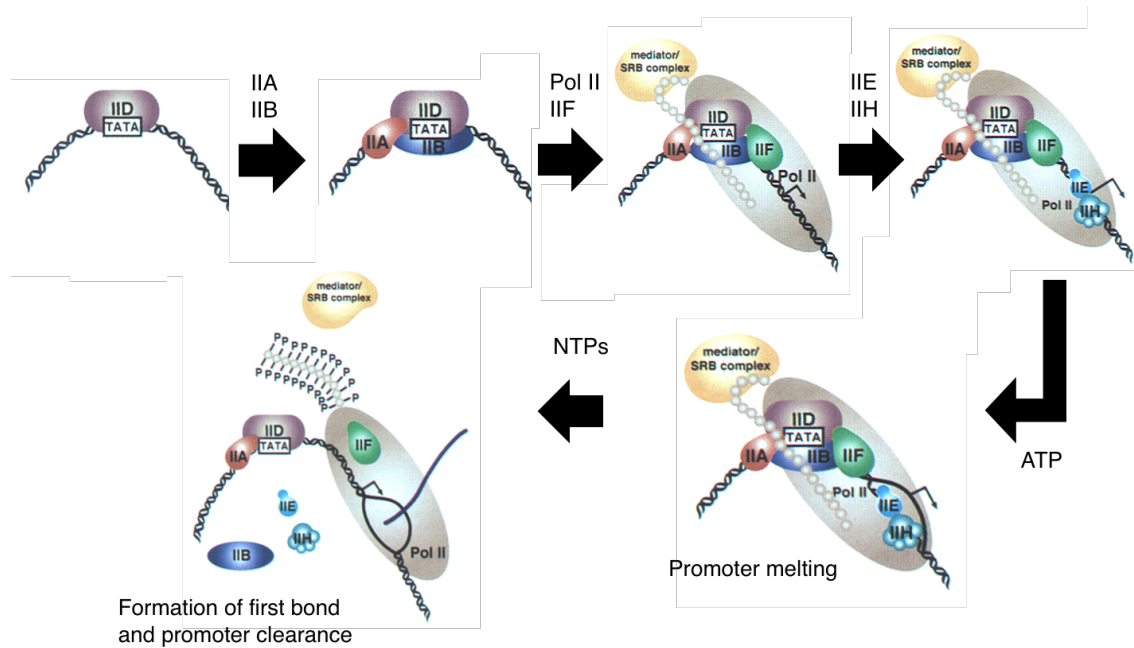


Figure 12: Sequential assembly of the PIC.

TFIID is the first factor to bind the promoter, stabilized by TFIIA and TFIIB. TFIIB recruits a pre-formed RNAPol II – TFIIF complex, and the RNAPol will interact with TFIIE, which finally will mediate the recruitment of TFIIH to complete the PIC. Adapted from (Orphanides *et al.*, 1996).

TFIIH subunit cdk7 also has a fundamental role in the process as the kinase is responsible for the phosphorylation of the CTD of Rbp1, the largest subunit of RNAPol II, thus enabling promoter escape (Lu *et al.*, 1992) (Roy *et al.*, 1994) (Wong *et al.*, 2014). The CTD of RNAPol II comprises a number of repetitions (26 in yeast, 52 in mammals) of the peptide YSPTSPS, reported to be a major target for several kinases (Ossipow *et al.*, 1995), including cdk7, which phosphorylates the CTD at the Ser5 position exclusively (Roy *et al.*, 1994) (Wong *et al.*, 2014). The RNAPol II enzyme can be found *in vivo* in a non-phosphorylated state (IIA), which is the form recruited to the PIC, or a highly

phosphorylated one (IIO), which carries out the elongation phase (Lu *et al.*, 1992).

The regulation mechanisms of the helicase activity of XPB and the kinase activity of cdk7 can potentially act as control checkpoints for transcription. TFIIIE stimulates the ATPase activity of XPB up to 5-fold, promoting the start of the transcription reaction (Ohkuma & Roeder, 1994). In the case of cdk7, the cdk8 kinase, together with the cyclin C protein, phosphorylate cdk7's partner cyclin H, hindering cdk7's kinase activity and thus inhibiting transcription initiation (Akoulitchev *et al.*, 2000). The cdk8/cyclin C complex can also stop transcription initiation through phosphorylation of the CTD of Rpb1 before being recruited to the PIC, thus preventing its assembly (Akoulitchev *et al.*, 2000).

1.3.3. Evolutionary conservation

The three defining steps of the NER process (damage recognition, DNA unwinding and removal of the lesion) are essentially conserved in all domains of life, but the systems that are responsible for the actual repair reaction are completely different, as previously covered in this chapter. Most of the proteins involved in eukaryotic NER can only be found in this group, and the genes involved in bacterial and eukaryotic DNA repair show no homology, hinting at different origins for the two systems (Eisen & Hanawalt, 1999). On the contrary, many archaeal species present clear homologs of the TFIIH helicases XPB and XPD (White, 2003), although these proteins always appear as independent subunits, without forming a complex like their eukaryotic counterparts. Moreover, no homologs of the remaining TFIIH subunits have been found in archaea (Rouillon & White, 2011).

An exhaustive analysis carried out by Bedez and co-workers showed that the TFIIH Core sub-complex is universally present and highly conserved in Eukarya: helicases XPB and XPD and subunit p44 were present in all 63 species included in their study. Subunit p52 is absent in only one species, while p34 was not detected in Euglenozoa. Finally, p8 and p62 were absent only in a few unicellular species (Bedež *et al.*, 2013). A second study in teleost fish found orthologues of all 10 TFIIH subunits in all the species analysed, showing an impressive 80-90% sequence identity, particularly for subunits XPB, XPD and cdk7 (Silva *et al.*, 2014). This study also reported a higher conservation of the Core subunits when compared to the CAK genes, suggesting an evolutionarily-speaking more

relevant role of the 7-subunit sub-complex to functioning TFIIH (Hirsh & Fraser, 2001).

1.3.4. TFIIH as a viral target

As mentioned before, subunit p62 provides the TFIIH complex with an extensive network of contacts with many different proteins. Surprisingly, some of these interactions involve viral proteins that target p62 as a means to affect the transcription process (Lyles, 2000).

The non-structural (NSs) protein of the Rift Valley fever virus (RVFV) (responsible for a type of haemorrhagic fever in humans and high rates of miscarriage in livestock (Boshra *et al.*, 2011) was first proposed to reduce the host's transcription levels by sequestering free subunits p44 and XPB, making them unavailable for incorporation into TFIIH and thus reducing the cellular concentration of the *de novo* synthesized complex (Le May *et al.*, 2010). Later, a more efficient mechanism was observed in which protein NSs promotes the degradation of TFIIH subunit p62 (Kalveram *et al.*, 2011) soon after infection by directing it to the E3 ubiquitin ligase system (Kainulainen *et al.*, 2014), drastically reducing TFIIH levels in the cell.

Subunit p62 is also a target of the transactivator herpes simplex virion protein 16 (VP16) (Xiao *et al.*, 1994). The herpes virus is a common human pathogen that presents an enormous variety of clinical manifestations, from mild lesions of the mucosal membranes to life-threatening symptoms such as encephalitis (Whitley & Roizman, 2001). VP16 stimulates the expression of immediate-early genes in infected cells, triggering a lytic cycle rather than a latent infection (Wysocka & Herr, 2003). In this case, the herpes virus hijacks the host's transcriptional machinery for its own purposes, and binding of VP16 to TFIIH subunit p62 will stimulate both the initiation and elongation stages of transcription (Blau *et al.*, 1996) (Langlois *et al.*, 2008).

Another transactivation factor that stimulates both host and viral transcription through interaction with p62 is the Epstein-Barr virus (EBV) nuclear antigen 2 (EBNA2) (Chabot *et al.*, 2014). EBV is a widely spread virus affecting humans with a predominantly latent cycle (Mckenzie & El-guindy, 2015). The reactivation of the lytic cycle is of particular importance in immunocompromised patients and it has been associated to several types

of cancer, in particular Hodgkin's and Burkitt's lymphoma and nasopharyngeal carcinoma (Mckenzie & El-guindy, 2015).

1.4. The MultiBac™ system

1.4.1. Impact of baculovirus expression systems

The first paper describing the expression of a heterologous protein in insect cells was published by the Summers group in 1983 (Smith *et al.*, 1983): they exploited the polyhedrin promoter in the *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV), a baculovirus with a circular, double-stranded DNA genome of about 134 kb that infects species of the order Lepidoptera (Hawtin *et al.*, 1997), to overexpress human interferon- β .

The development of baculoviral expression vectors (BEVs) was potentiated by several helpful characteristics of the AcMNPV: it's easy to manipulate, and the rod-shape of the baculovirus permits the encapsidation of large DNA fragments besides the viral genome. As well as having a strong promoter, the polyhedrin gene encodes for a protein that is not required for viral replication in cell culture, and its absence in infected cells can be used as a marker for selection of recombinant viruses. Lastly, baculoviruses cannot replicate in vertebrates, so its use as a protein expression system is completely safe (Smith *et al.*, 1983) (Van Oers *et al.*, 2015).

Protein expression using BEVs offers a series of features not available in the more traditional *E. coli* expression system. Frequently, eukaryotic complexes contain subunits that are too large for *E. coli*'s protein production machinery. Insect cells not only don't present this problem, but they also provide post-translational modifications and a folding machinery similar to mammalian cells, and while expression of heterologous protein in insect cells takes a longer time compared to *E. coli*, it also offers a higher yield and better sample quality (Bieniossek *et al.*, 2012) (Assenberg *et al.*, 2013).

Despite its many advantages, it was not until the system was widely available commercially that BEVs became a routinely used tool in laboratories. Although they are still mainly employed as a system for overexpression of recombinant proteins, the current

INTRODUCTION

applications for BEVs are numerous (figure 13), and their efficacy as gene-transfer agents in gene therapy and as a system to produce human vaccine components has already been reported (Airenne *et al.*, 2013).

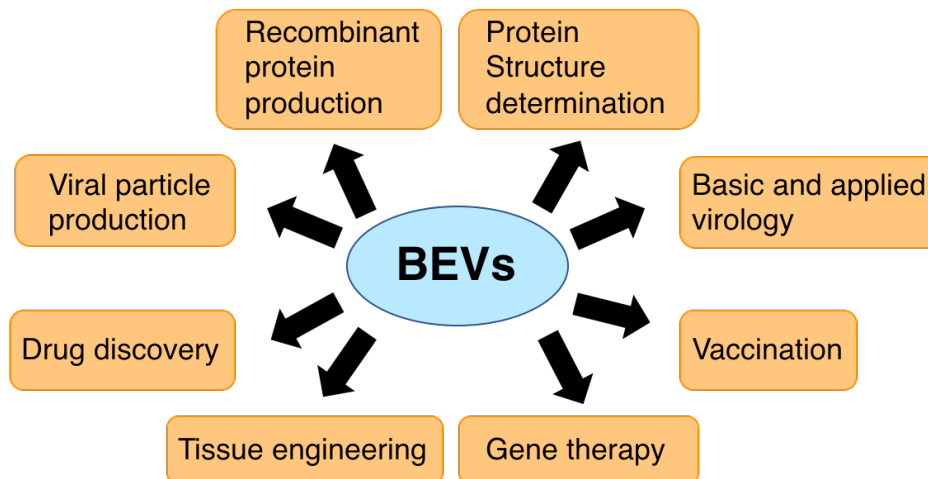


Figure 13: Applications of BEVs in biotechnology.

Adapted from (Airenne *et al.*, 2013).

1.4.2. The MultiBac™ system: a powerful cloning tool

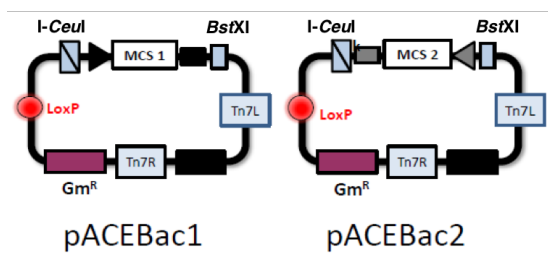
1.4.2.1. Heterologous protein expression using MultiBac™

Three features make MultiBac™ stand out among the many BEVs options commercially available today: (1) a set of small, easy to manipulate vectors that contain a multiplication module that allows the creation of multi-gene expression cassettes, (2) a LoxP site for creation of multi-gene constructs via recombination in a single, quick step, and (3) a baculoviral genome engineered for improved recombinant protein expression (Bieniossek *et al.*, 2013).

MultiBac™ includes a set of five small plasmids (figure 14): acceptors pACEBac1 and pACEBac2, carrying a gentamycin resistance marker, and donors pIDC, pIDS and pIDK, carrying resistance markers for chloramphenicol, spectinomycin and kanamycin, respectively. All five plasmids present a multiplication module, consisting of a restriction site for a homing endonuclease (*I-CeuI* for acceptors and *PI-SceI* for donors) and a second restriction site for the *BstXI* enzyme, that allows the cloning of multiple genes

into a single plasmid. Both acceptors and donors also carry a *LoxP* site that permits the obtaining of multi-gene constructs in a single step via recombination with a *Cre* recombinase. Only acceptors present the transposition elements *Tn7R* and *Tn7L* that will allow transposition into the baculoviral genome at a later stage in the process of expressing an heterologous protein, so while any combination of donors can be employed, the use of at least one acceptor plasmid is indispensable (Bieniossek *et al.*, 2013).

MultiBac™ Acceptor Plasmids



MultiBac™ Donor Plasmids

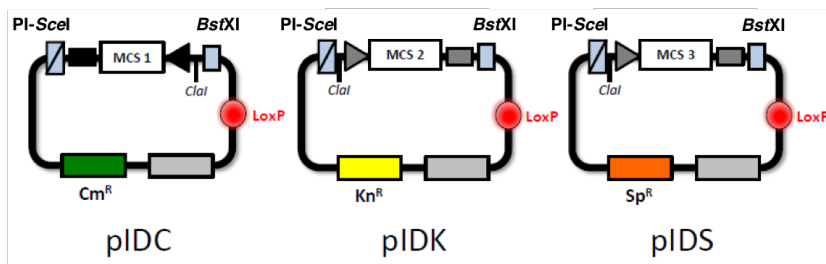


Figure 14: MultiBac™ is a powerful cloning tool.

Cartoon representation of the acceptor (top panel) and donor (bottom panel) plasmids composing the MultiBac™ system. The representation highlights their most important features: homing endonuclease / *BstXI* multiplication module, *LoxP* recombination site, resistance markers, and transposition elements that permit the cloning and expression of multi-subunit eukaryotic complexes. Adapted from (Bieniossek *et al.*, 2013).

The multi-gene construct, previously obtained in a regular *E. coli* strain, will be transposed into the baculoviral genome present in strain DH10 MultiBac™, which also contains a transposase encoded in a helper plasmid (Bieniossek *et al.*, 2013). This baculoviral genome was originally obtained from the AcMNPV virus, and has been

INTRODUCTION

modified for enhanced production of heterologous protein by disruption of the genes V-CATH and chiA (Bieniossek *et al.*, 2012), which are responsible for the liquefaction of the infected larvae to facilitate the dissemination of the new viral particles (Hawtin *et al.*, 1997). This modification results in minimization of the proteolytic activity of the virus, thus reducing cell lysis and improving protein production (Berger *et al.*, 2004).

Finally, the obtained bacmid – the name given to the recombinant baculovirus - will be transfected into insect cells, and the recovered P0 virus will be used for infection of a scaled-up insect cell culture for high-yield expression of heterologous protein (Bieniossek *et al.*, 2012).

1.4.2.2. MultiBac™ allows expression of multi-protein complexes

Very rarely proteins carry out their function on their own: they tend to act as complexes with other proteins and molecules such as nucleic acids, performing essential reactions in fundamental processes like DNA replication or transcription. The research of these complexes has been very limited due to the high quality and quantity of the sample required for their study, rarely achievable with the cloning and expression tools available until recently (Bieniossek *et al.*, 2012), and integrative methods were the best option to approach the study of these macromolecular machines (Robinson *et al.*, 2007). With this in mind, the MultiBac™ system was specifically conceived as a production tool for eukaryotic multi-protein complexes (Berger *et al.*, 2004).

This expression tool solves two problems typically associated with the production of multi-protein complexes: it efficiently handles large, multiple genes, and it allows the quick modification of the complex if a subunit needs to be altered (e. g. a new purification tag needs to be introduced) by replacing the original plasmid with one containing the desired modification, without needing to clone the whole complex again (Bieniossek *et al.*, 2012).

Structures as big as the anaphase promoting complex (APC/C) (a 1.1 MDa, 13-subunit complex that participates in cell cycle regulation) and transcription factor's Mediator head module (223 kDa, 7 subunits), among many others, have been reported after expression using the MultiBac™ system. However, the system still presents room for improvement,

and issues as undesired post-translational modifications and instability of the infecting viral stocks remain to be tackled (Bieniossek *et al.*, 2012).

1.5. Aims and objectives

The TFIIH complex has been the subject of extensive research since its description in 1992 (Flores *et al.*, 1992) due to its essential functions in both transcription and DNA repair. This PhD project aimed to study the role of the complex, and particularly that of the TFIIH Core sub-complex (subunits XPB, XPD, p62, p52, p44, p34 and p8) as a NER factor. The objective of our research was to try to understand the reactions taking place at the repair bubble that lead to the removal and repair of the damaged fragment, as a lot of questions still remain unanswered despite the intensive focus placed on TFIIH:

- How does the complex fit into the repair bubble? Does it bind to the damaged strand, the undamaged strand, or both?
- Does XPD really verify the lesion in the DNA? Does the 4FeS cluster play a role in this verification?
- How do XPB and XPD carry out their roles inside the bubble? And the other subunits? How do they interact with DNA, and with each other? Do they undergo any conformational changes to bind to the damaged site?

Chapter 3 characterises the binding of XPC (responsible for identifying a lesion in the DNA and the first repair factor to bind the damaged site (Sugasawa *et al.*, 1998)) to a series of double stranded substrates, confirming its preference for mispaired sites rather than actual lesions. Chapter 4 describes the process of cloning and expressing both the 10-subunit complex and the 7-subunit TFIIH Core sub-complex using the MultiBac™ cloning system and expressing the complex in *Sf9* insect cells, which allowed us to obtain up to 1 mg of pure complex from a 2 L culture (such a high yield has only been achieved by the Yang group up to date (Li *et al.*, 2015)). Chapter 5 characterizes the binding of TFIIH to a series of double-stranded substrates and the unwinding of both damaged and undamaged substrates, confirming the ATP-dependent 5' to 3' helicase activity of XPD.

2. Materials and methods

2.1. Cloning using the MultiBac™ system

2.1.1. Engineered genes. 2A Sequences

Synthetic genes (not codon-optimized) for all ten TFIIH subunits, XPC and HR23B were ordered from Integrated DNA Technologies (IDT) (see appendix A for full sequences). Uniprot accession codes for the protein isoforms employed are indicated in table 2.

Table 2: UniProt accession numbers for the proteins included in this study.

Subunit	UniProt accession number
XPD	P18074
XPB	P19447
p62	P32780
p52	Q92759
p44	Q13888
p34	Q13889
p8	Q6ZYL4
cdk7	P50613
Cyclin H	P51946
MAT1	P51948
XPC	Q01831
HR23B	P54727

Genes for Core subunits p44-p34, p52-p8 and CAK subunits cdk7-cyclin H were designed as a single unit, with each pair of genes separated by an autocleavable 2A-like encoding sequence (figure 15).

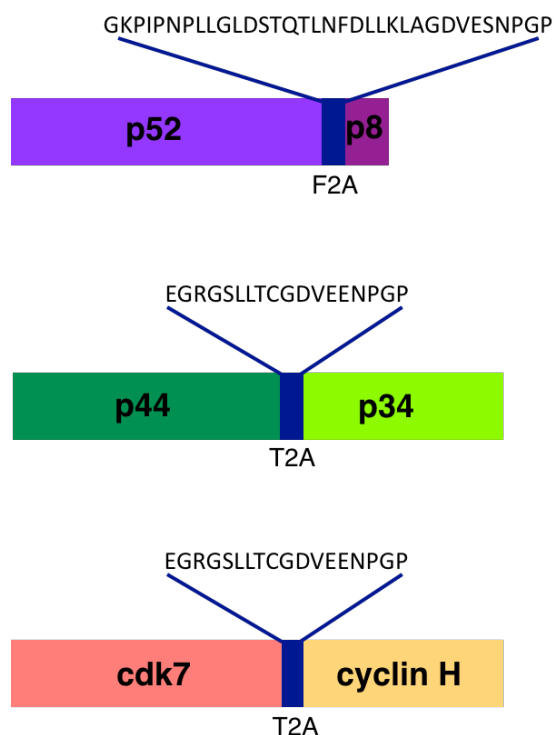


Figure 15: TFIIH genes engineered with 2A autocleavable sequences.

Cartoon representation showing constructs p52-V5-F2A-p8 and p44-T2A-p34 from the Core TFIIH sub-complex, and cdk7-T2A-cyclin H from the CAK sub-complex.

These F2A and T2A peptides are variations of the 2A auto-cleavable sequence encoded by the genome of the Foot-and-Mouth Disease Virus (FMDV). The 2A sequence causes the ribosome to stall during translation, releasing the nascent protein by self-cleaving at the NPG'P site and remaining attached to the C-terminal end of the newly translated protein without affecting its function or expression (Luke *et al.*, 2009). The ribosome will now translocate to the next in-frame starting position, where it will resume translation.

2.1.2. Purification tags

A six-residue polyhistidine tag was initially designed to be attached to one or several of the TFIIH subunits for purification purposes. The tag also included a TEV cleavage site (for removal of the tag) and a V5 epitope for protein detection in a Western blot (figure 16). These two elements are common to all the tags designed for this project.

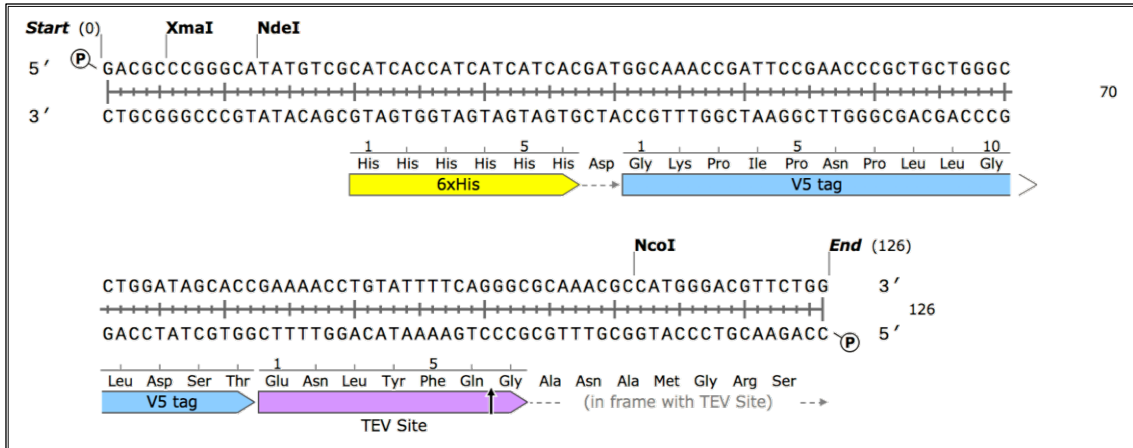


Figure 16: 6xHis-V5-TEV purification tag.

This tag was later modified by introducing a nine-residue spacer between the V5 epitope and TEV cleavage sites (figure 17) and the new tag-encoding sequence was cloned not only into several TFIIF Core subunits, but also into the XPC and HR23B proteins. Both tag-encoding sequences were ordered from IDT as a *gBlock* (small synthetic DNA fragment designed for easy modification of genes) and cloned at the N-terminal end of the chosen subunits using restriction enzymes *NcoI* and *XmaI* (New England Biolabs) following the manufacturer's instructions.



Figure 17: 8xHis-V5-spacer-TEV purification tag.

MATERIALS AND METHODS

Two additional tags were specifically designed for the purification of subunit XPD. First, a ten-residue polyhistidine tag including the TEV cleavage site and the V5 epitope was attached to the C-terminal end of the helicase (figure 18):

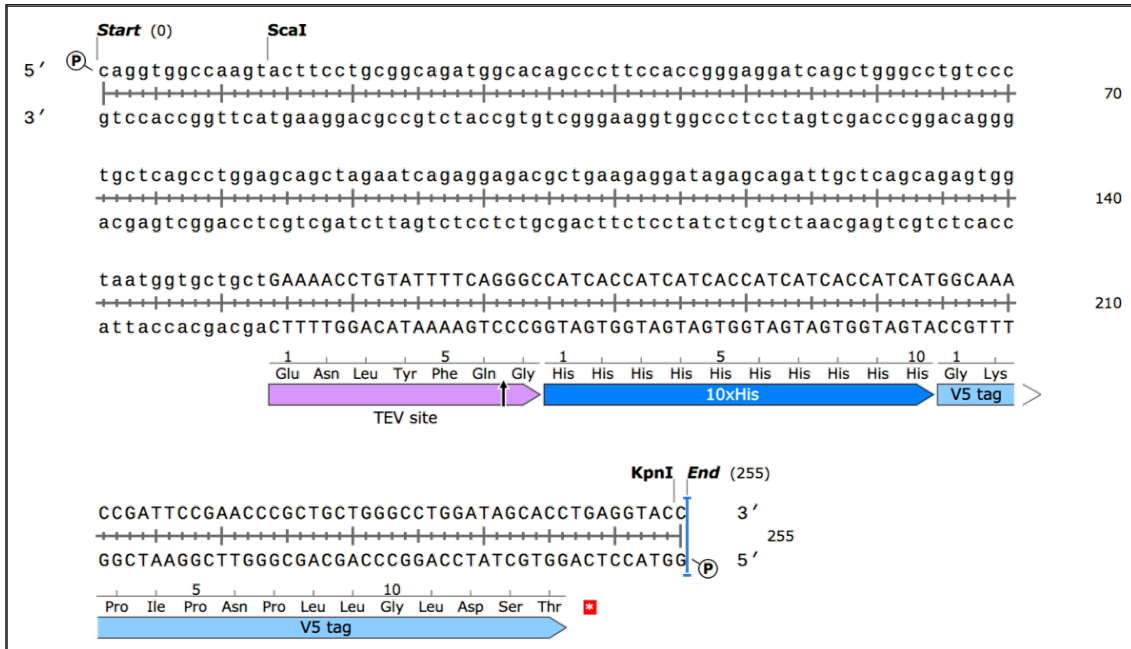


Figure 18: TEV-10xHis-V5 purification tag.

A sequence encoding a Twin-Strep-tag® streptavidin tag was later cloned also at the C-terminal end of XPD (figure 19):

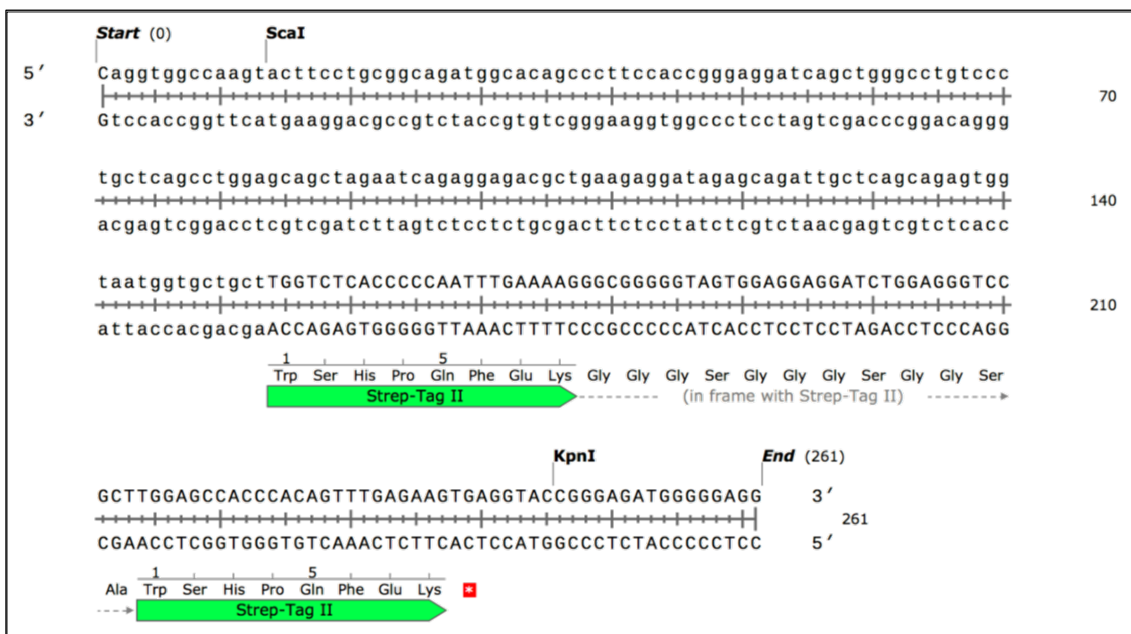


Figure 19: Twin-Strep-tag® purification tag.

Both tags were ordered from IDT as *gBlocks* and cloned using restriction enzymes *Scal* and *KpnI* (New England Biolabs) as indicated by the manufacturer.

2.1.3. Tetracysteine (TCP) tag

A sequence encoding a tetracysteine peptide (TCP) (CCGPCC) was cloned at the N-terminal end of XPC (figure 20 (A)). The tag was ordered from IDT as an oligonucleotide and cloned into XPC using the *NcoI* restriction enzyme (Thermo Fisher Scientific). The XPC synthetic gene had previously been cloned into a pACEBac2 plasmid containing an 8xHis-V5-spacer-TEV purification tag. The presence of the TCP tag was verified by sequencing (GATC Biotech).

The same TCP tag-encoding sequence was also cloned at the N-terminal end of TFIIH Core subunit XPB, whose gene had been previously cloned into a pACEBac2 plasmid using restriction enzymes *NcoI* and *KpnI* (figure 20 (B)). The successful cloning of the tag was also verified by sequencing. The sequence encoding the TCP-tagged XPB was later cloned into the TFIIH Core sub-complex.

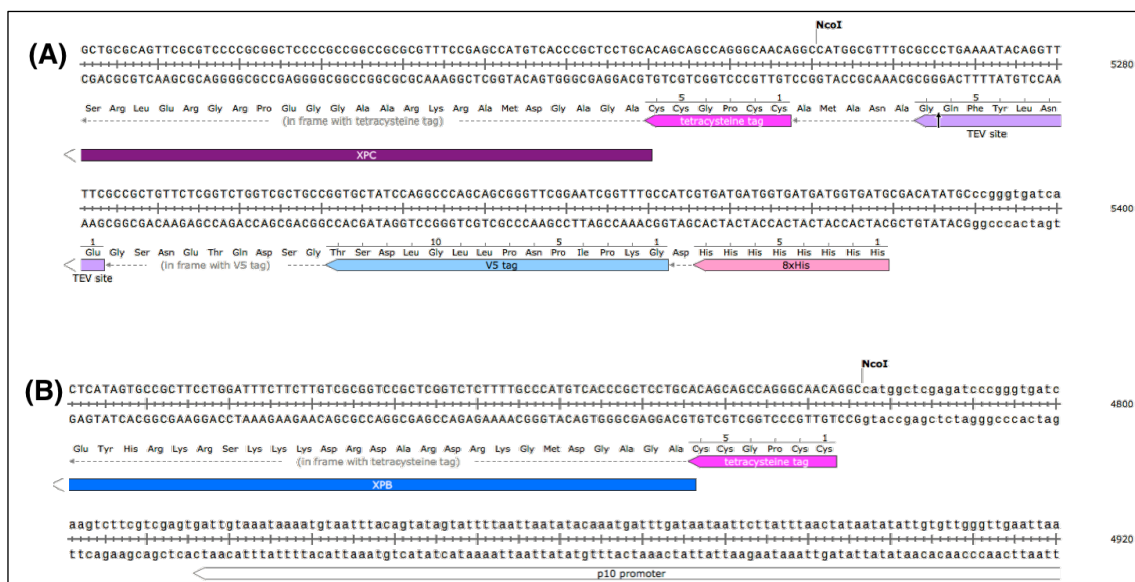


Figure 20: TCP-tagged XPC and XPB.

A TCP tag was cloned at the N-terminal end of both XPC (A) and TFIIH Core subunit XPB (B). In the case of XPC, the full tag included 8xHis-V5-spacer-TEV-TCP.

2.1.4. Cloning using the MultiBac™ multiplication module

The ten-subunit TFIIH complex and the heterodimer XPC-HR23B were cloned using the MultiBac™ baculovirus expression system (Geneva Biotech). As previously mentioned in chapter 1, section 1.4.2.1, the MultiBac™ system includes a set of five plasmids: two acceptors and three donors, but as we cloned all of our genes using the homing endonuclease / *BstXI* multiplication module, only the acceptor plasmids pACEBac1 (2904bp) and pACEBac2 (2761bp) (figure 21) were used. Both plasmids carry a gentamycin resistance marker, the Tn7R / Tn7L elements necessary for later transposition of our TFIIH or XPC-HR23B construct, and a multiplication module defined here by the homing endonuclease I-*CeuI* and the restriction enzyme *BstXI*. This module also encompasses a promoter and a polyadenylation sequence (Bieniossek *et al.*, 2013).

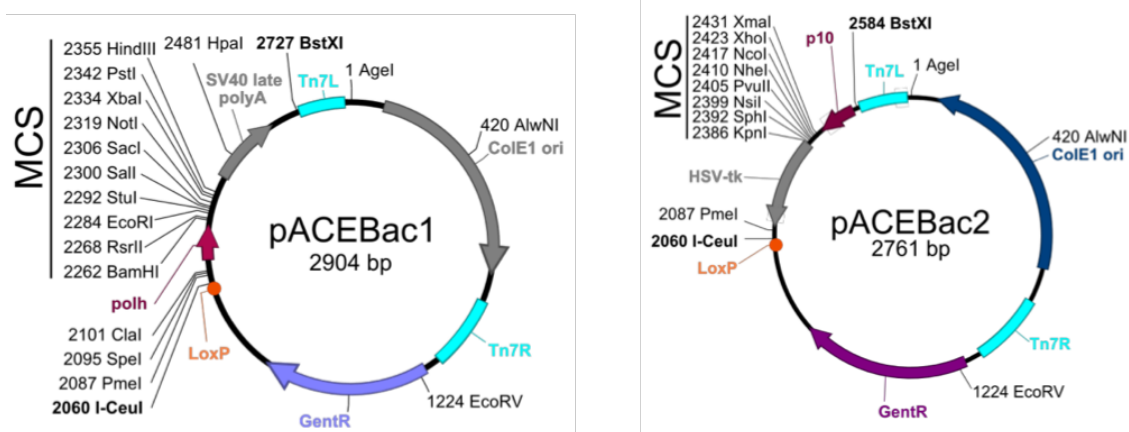


Figure 21: pACEBac1 and pACEBac2 vector maps

(Bieniossek *et al.*, 2013).

In a first step each TFIIH gene, plus XPC and HR23B, was cloned into either pACEBac1 or pACEBac2 using restriction enzymes *BamHI/Sall* and *NcoI/KpnI* (Fermentas), respectively, following the manufacturer's instructions. Afterwards genes were cloned, one at a time, into a single plasmid using the multiplication module I-*CeuI*/*BstXI* (New England Biolabs), as shown in figure 22, until the desired combination was accomplished. Several recombinant multi-gene plasmids were created following this method: one encoding all seven Core subunits (XPD, XPB, p62, p52-F2A-p8 and p44-T2A-p34), another one encoding the three CAK subunits (cdk7-T2A-cyclin H and MAT1), and a final multi-gene plasmid encoding the heterodimer XPC- HR23B. A list containing

all the plasmids generated in the course of this study can be found in appendix B, table B1. All constructs were maintained in DH5α cells.

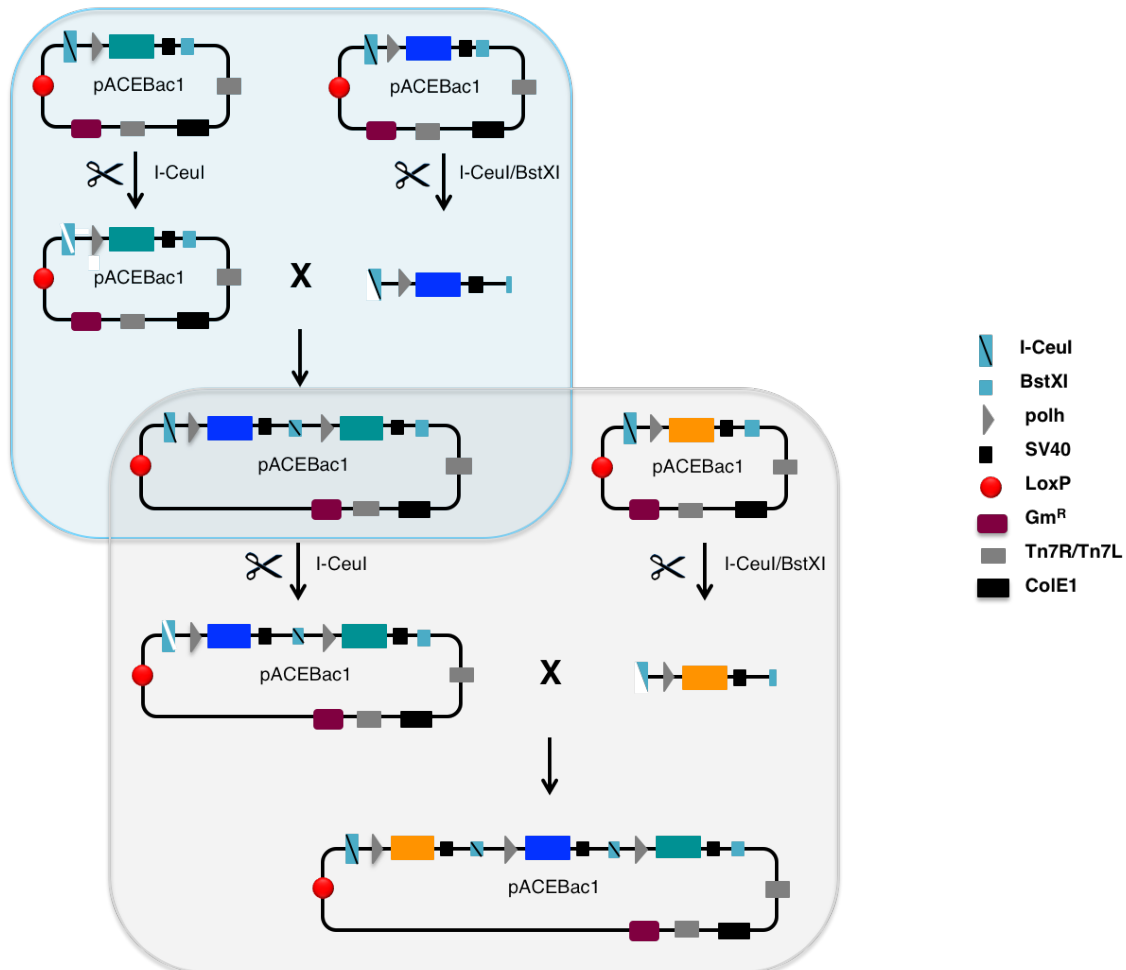


Figure 22: Cloning using MultiBac™'s multiplication module.

In the first reaction (blue box) a plasmid containing one of our genes is linearized with the homing endonuclease *I-CeuI*, which leaves a cohesive end compatible with the one created by *BstXI* when a second gene, located in a different plasmid, is excised with both enzymes. Both the linearized vector and excised insert are now ligated as in a regular cloning procedure. This means that every time a new gene is cloned into the plasmid a hybrid, non-functional *I-CeuI/BstXI* site is created, and the original *BstXI* site is restored. In a second reaction (grey box) the plasmid that now contains two genes is linearized with *I-CeuI* as before, and a third gene, located again in a different plasmid, is excised with the homing endonuclease and *BstXI*, etc. This method theoretically allows the cloning of as many expression cassettes as desired (Bieniossek *et al.*, 2009).

MATERIALS AND METHODS

Intermediate constructs were verified by restriction with the appropriate combinations of different enzymes, such as *NcoI*, *KpnI*, *BamHI*, *HindIII*, etc. (Fermentas) following the manufacturer's instructions. Final TFIIH Core constructs were further analyzed by PCR using the MyTaq™ DNA Polymerase (Bioline) and primers listed in appendix C, according to the reaction shown in table 3. All final constructs (TFIIH Core, CAK sub-complex and heterodimer XPC-HR23B) were ultimately verified by sequencing (GATC Biotech) each gene.

Table 3: PCR reaction for verification of intermediate TFIIH Core constructs.

Cycles	1	30			1
Temperature	95 °C	95 °C	45 – 50 °C	72 °C	72 °C
Time	5 minutes	30 seconds	30 seconds	1 minute	10 minutes

2.1.5. Tn7 Transposition and bacmid purification

The recombinant plasmids carrying the TFIIH Core, CAK or XPC-HR23B genes were transposed into chemically competent DH10 MultiBac™ cells containing an engineered baculoviral genome derived from the AcMNPV virus (Bieniossek *et al.*, 2012) and a helper plasmid containing a Tn7 transposase, provided with the MultiBac™ kit. A full list of the bacmids generated for this project can be found in appendix B, table B2. The multi-gene construct is transposed into a Tn7 site, causing the disruption of the *lacZ* gene, which will allow for blue/white colony selection (figure 23).

The transposition protocol was adapted for each construct according to its size:

- Incubation on ice varied from 30 minutes for smaller constructs to 2 hours for the seven-gene Core construct.
- Heat shock went from 60 to 90 seconds for the bigger constructs.
- Chilling on ice was extended from 5 minutes up to 20 minutes.
- Finally, the recovery step varied from 6 to 24 hours at 37 °C with shaking as the construct grew bigger.

Each culture was then plated into LB agar containing 50 $\mu\text{g/ml}$ kanamycin, 7 $\mu\text{g/ml}$ gentamycin, 10 $\mu\text{g/ml}$ tetracycline, 100 $\mu\text{g/ml}$ ampicillin, 40 $\mu\text{g/ml}$ isopropyl β -D-1-thiogalactopyranoside (IPTG) and 20 $\mu\text{g/ml}$ X-gal and incubated at 37 °C for 2-3 days.

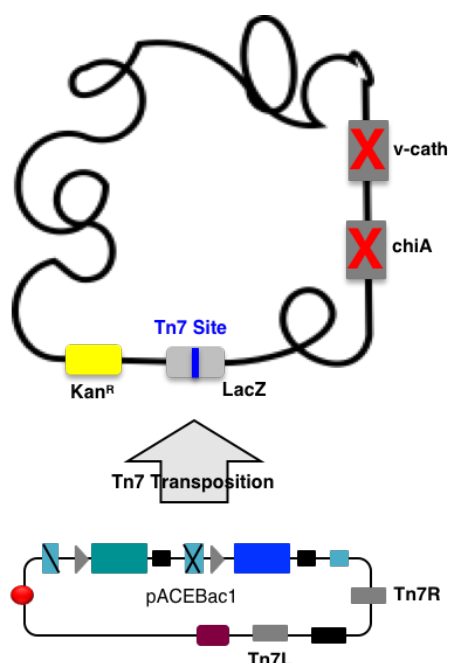


Figure 23: Transposition of a multi-gene construct into the baculoviral genome.

Positive transpositions will appear as white colonies amid a background of blue unsuccessful ones when plated in LB agar containing 50 $\mu\text{g/ml}$ kanamycin, 7 $\mu\text{g/ml}$ gentamycin, 10 $\mu\text{g/ml}$ tetracycline, 100 $\mu\text{g/ml}$ ampicillin, 40 $\mu\text{g/ml}$ IPTG and 20 $\mu\text{g/ml}$ X-gal (Bieniossek *et al.*, 2013).

A minimum of two recombinant white colonies per construct were picked and re-streaked to confirm their white phenotype. These colonies were also used to start 3 ml Luria-Bertani (LB) broth cultures containing 50 $\mu\text{g/ml}$ kanamycin and 7 $\mu\text{g/ml}$ gentamycin. The cultures were incubated at 37 °C, 200 rpm overnight, then centrifuged at 15000 rpm for 1 minute at room temperature (centrifuge 5424, Eppendorf). Pellet was resuspended in buffer P1, then buffer P2 and finally buffer N3 from a QIAprep Spin Miniprep Kit (Qiagen). The suspension was centrifuged at 15000 rpm for 10 minutes at room temperature, and supernatant was transferred to a fresh tube. The baculoviral DNA present in the supernatant was precipitated with isopropanol and washed twice with ethanol 70%. All work following the second washing step was performed in a sterile S@felow 1.2 hood (Bioair Instruments). Ethanol was removed from the tube, and sterile bacmids were finally resuspended in 30 μl deionized water and kept at 4 °C until transfected.

2.2. Insect cell culture techniques

2.2.1. Routine maintenance of insect cell cultures

A running stock of *Spodoptera frugiperda* (Sf9) cells (Thermo Scientific) is kept in the form of monolayers cultured in T75 flasks (Greiner Bio-One) (figure 24), incubated at 27 °C in a MaxQ 6000 incubator (Thermo Scientific). Monolayers are sub-cultured every five to six days in sterile conditions in a S@felow 1.2 hood (Bioair Instruments), with Sf900™ III serum-free medium (SFM) (Gibco).

The cells were adapted from monolayer to suspension conditions before every infection. A starter 25 ml Sf900™ III SFM medium suspension culture with an initial cell density of 1×10^6 cells/ml was initiated in 125 ml disposable shaker flasks (Corning) and incubated at 27 °C with 110 rpm orbital shaking. The suspension was sub-cultured back to 1×10^6 cells/ml every time it reached a density of $2\text{-}3 \times 10^6$ cells/ml (1-2 days), until infected.

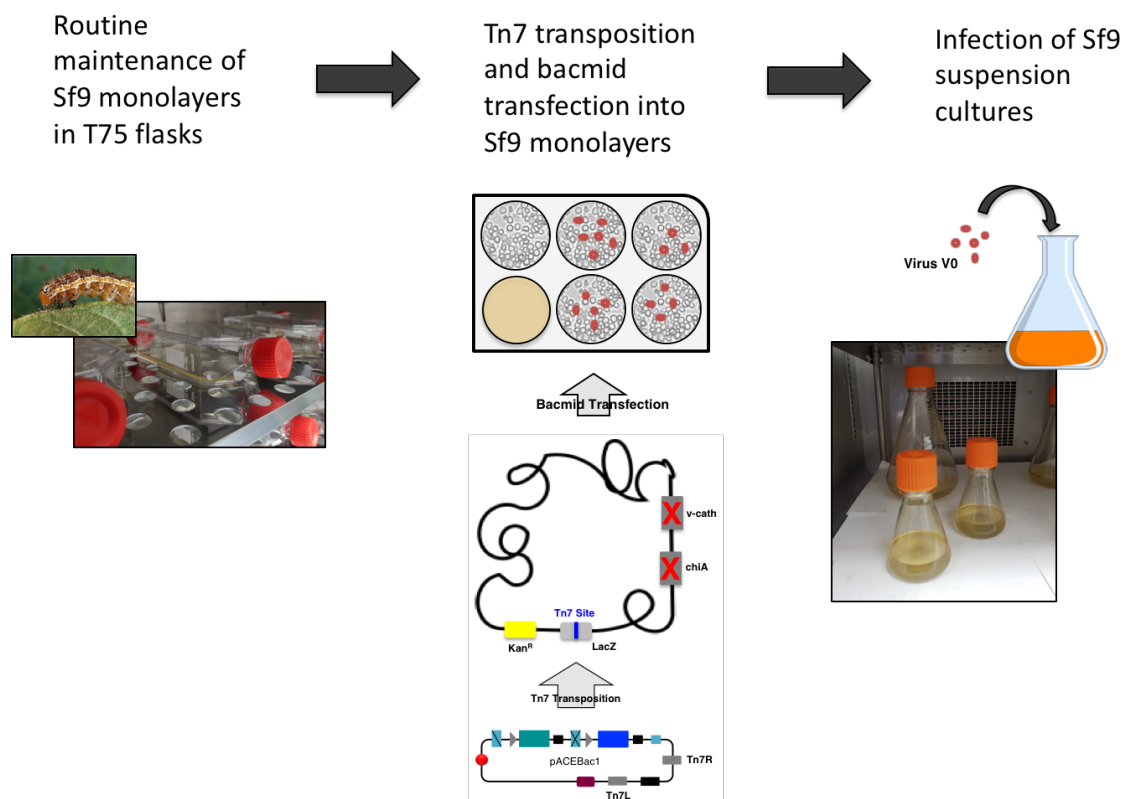


Figure 24: Summary of Sf9 culture techniques.

2.2.2. Insect cell transfection

Recombinant bacmids were transfected into a monolayer of *Sf9* cells using XtremeGENE HP DNA Transfection Reagent (Roche). 1×10^6 Cells were seeded to each well in a 6-well plate (Greiner Bio-One), and volume was topped up to 3 ml with fresh *Sf900*TM III SFM medium. The plate was incubated at 27 °C for a minimum of 2 hours to allow cells to attach to the surface of the well. Meanwhile, each bacmid to be transfected was resuspended in 200 μ l *Sf900*TM III SFM medium. A mixture of 100 μ l *Sf900*TM III SFM medium and 10 μ l transfection reagent was added to each bacmid, and tubes were incubated at room temperature for 30 minutes. Afterwards, the bacmid/transfection reagent mixture was added to the *Sf9* cell monolayer (2 wells for every bacmid to be transfected), and plates were incubated at 27 °C for 60 hours before recovering a P0 viral stock (figure 24).

2.2.3. Insect cell infection

Infections were carried out in a *Sf9* suspension culture with a starting cell density of 1×10^6 cells/ml in a total volume of 25 ml *Sf900*TM III SFM medium. This suspension was infected with 3 ml of a passage 0 (P0) viral stock and incubated for 60 hours in a shaker flask at 27 °C with 110 rpm orbital shaking (figure 24). Afterwards, cell pellet and medium containing viral stock P1 were collected by centrifugation (CF20 centrifuge, Awel) at 1000 rpm for 5 minutes and kept at -80 °C until tested. Once expression levels had been evaluated, the volume of the *Sf9* suspension was scaled up appropriately for TFIIH Core, CAK or XPC-HR23B overexpression. Infections are always performed using 3 ml virus for every 25 ml cell suspension, generally utilizing viral stocks obtained from cultures that have been passaged one (P1) or two (P2) times.

2.2.4. Insect cell co-infection

A co-infection experiment was carried out to obtain the 10-subunit TFIIH complex. For this test, a 25 ml *Sf9* suspension with a starting cell density of 1×10^6 cells/ml was infected with 1 ml of a P3 TFIIH Core viral stock and 3 ml of a P0 CAK viral stock. The suspension was incubated for 60 hours at 27 °C with 110 rpm orbital shaking, and after that time the pellet and medium containing a viral stock P1 that included all ten TFIIH subunits were harvested as described for our regular infection experiments. After evaluation of the

expression levels, the volume of the *Sf9* suspension was scaled up, and co-infections were performed with 3 ml of a P1, P2 or P3 10-subunit TFIIH viral stock for every 25 ml cell suspension.

2.3. Purification of the TFIIH complex

2.3.1. TFIIH Extraction and BioSprint expression tests

Expression levels for the different TFIIH complexes obtained were checked from a 25 ml *Sf9* suspension culture infected with 3 ml of a P0 viral stock as described in section 2.2.3. Cell pellet was harvested by centrifugation and resuspended in 5 ml lysis buffer (phosphate-buffered saline (PBS), 300 mM NaCl, 0.1% Triton, plus 10 units/ml DNaseI (Sigma) and a protease inhibitor cocktail (Roche)) per gram of pellet. The resuspended cells were subsequently lysed with a dounce homogenizer (Fisher Scientific) (20 strokes with a tight pestle, performed on ice) and centrifuged at 13000 rpm for 30 minutes at 4 °C (centrifuge 5415, Eppendorf). The cleared supernatant was then transferred to a fresh tube and analyzed in a BioSprint station (Qiagen): 500 μ l supernatant and 10 μ l magnetic Ni-NTA beads (Promega) were loaded into the first column of the BioSprint sample tray. After binding, the beads were washed twice with 500 μ l buffer PBS, 300 mM NaCl, 30 mM imidazole, and bound proteins were eluted in 50 μ l buffer PBS, 300 mM NaCl, 300 mM imidazole.

2.3.1.1. BioSprint expression test in anaerobic conditions

Pellet was harvested by centrifugation from a 25 ml *Sf9* suspension culture infected with 3 ml of a P0 viral stock and resuspended in 5 ml lysis buffer per gram of pellet, previously degassed with nitrogen for 20 minutes. Extraction of the recombinant TFIIH Core sub-complex was performed in a glove box in an anaerobic atmosphere following the method described in the previous section. The cleared supernatant was transferred to a fresh tube and the sub-complex was purified by performing the same steps included in the BioSprint process manually: 1 ml supernatant was added to a tube containing 30 μ l magnetic Ni-NTA beads (Promega) and incubated for 15 minutes, mixing frequently. The supernatant was separated with the help of a magnetic rack, and the beads were washed twice with 1 ml degassed wash buffer. The TFIIH Core sub-complex was finally eluted

from the beads with 100 μ l degassed elution buffer after a 10-minute incubation period. Beads were finally resuspended in 100 μ l wash buffer and a sample was loaded into a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel to verify that the complex had been successfully eluted.

2.3.1.2. SDS-PAGE analysis

The different fractions obtained from the BioSprint purification test were verified by SDS-PAGE analysis. Loading buffer (NuPAGE™ LDS Sample Buffer (4X) 1 mM dithiothreitol (DTT) (Thermo Scientific)) was added to samples prior to heating them at 90 °C for 2 minutes. Afterwards, samples were loaded into a NuPAGE™ 4-12% Bis-Tris Protein Gel (Thermo Scientific) and run at 200 V for 35 minutes in 1X NuPAGE™ MES SDS Running Buffer (Thermo Scientific). Gels were finally stained with InstantBlue™ Protein Stain (Expedeon) for 20-30 minutes approximately and rinsed with water before analysis.

2.3.1.3. Western blotting analysis

Sometimes expression levels for certain constructs were too low to be observed in an SDS-PAGE gel, and a Western blot was required for the analysis. Samples of interest would be first run in an SDS-PAGE gel as described in the previous section, using a PageRuler™ Prestained Protein Ladder (Thermo Scientific) as a marker. The gel was later blotted using the iBlot® nitrocellulose mini Gel Transfer Stacks required for the iBlot® system (Thermo Scientific), following the manufacturer's indications. Blots were blocked in blocking buffer (500 ml PBS with 25 g milk powder and 250 μ l Tween®20 (Sigma)) for 60 minutes. Afterwards, they were incubated in a blocking buffer solution containing 1:20000 V5 primary antibody (kindly provided by Prof. Richard Randall) for 60 minutes and washed three times with fresh blocking buffer for 10 minutes each time. The blot was incubated then in a solution of fresh blocking buffer containing 1:10000 IRDye 800CW Goat anti-Mouse IgG (H + L) (LI-COR) fluorescent secondary antibody for 60 minutes, protected from light, and washed again three times with PBS for 10 minutes each time, always protected from light. All incubation and washing steps were performed at room temperature with mild rocking. Finally, blots were scanned using an Odyssey CLx instrument (LI-COR) with 778 nm excitation and 795 nm emission wavelengths.

2.3.2. Immunoprecipitation using a V5 antibody

Immunoprecipitation (IP) experiments were performed to determine if the cloned His tag was sufficient to allow the whole TFIID Core sub-complex to be pulled down. 50 μ l Protein G magnetic beads (Invitrogen) and 200 μ l PBS, 0.02% Tween® 20 containing 10 μ g V5 antibody were incubated in a rotator at room temperature for 10 minutes. The protein G beads/V5 antibody complex was crosslinked afterwards with 20 mM dimethyl pimelimidate (DMP), 0.2 M triethanolamine, pH 8.2 following the manufacturer's indications. The supernatant containing the TFIID Core subunits was added to the crosslinked complex, and the mixture was incubated in a rotator at room temperature for 15 minutes. This supernatant was obtained by rupture of the Sf9 cells with a dounce homogenizer, following the same method described in section 2.3.1. TFIID Core was excised from the beads/V5 antibody complex by cleaving with TEV, performed overnight both at room temperature and 4 °C, and the different fractions were analyzed by Western blotting.

2.3.3. Immobilised metal ion affinity chromatography (IMAC)

2.3.3.1. HisTrap™ Fast Flow (FF) column

Cell pellet for the different TFIID Core sub-complexes was harvested by centrifugation at 1000 rpm for 5 minutes at 4 °C (Allegra 21R centrifuge, Beckman Coulter) and resuspended in 5 ml lysis buffer per gram of pellet (PBS, 300 mM NaCl, 0.1% Triton, plus 10 units/ml DNaseI (Sigma) and a protease inhibitor cocktail (Roche), degassed with argon). Suspensions were then passed through a dounce homogeneizer (at least 20 strokes, keeping suspension on ice at all times) and centrifuged at 40000 rpm for 45 minutes at 4 °C (Optima L-90K ultracentrifuge, Beckman Coulter). The cleared supernatant was filtered (0.45 μ m) afterwards and loaded into a 1 ml HisTrap™ FF column (GE Healthcare), and purification was performed using an Äkta system (GE Healthcare). After washing the column with buffer A (PBS pH 7.5, 300 mM NaCl, 30 mM imidazole, 10% glycerol), bound proteins were eluted in an imidazole gradient created by buffer B (PBS pH 7.5, 300 mM NaCl, 500 mM imidazole, 10% glycerol). Purified fractions containing the TFIID core complex were pooled together and concentrated using an Amicon® concentrator with a 30K cutoff (Millipore).

2.3.3.2. XPD purification using detergent ANAPOE C₁₂E₈

Pellet was harvested as previously described and resuspended in lysis buffer 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 300 mM potassium acetate, 2 mM 2-mercaptoethanol (BME), 5 mM ANAPOE C₁₂E₈, plus a protease inhibitor cocktail (Roche). The suspension was sonicated at 12 microns (10 seconds on and 20 seconds off, repeated 10 times) and centrifuged at 15000 rpm at 4 °C for 30 minutes (centrifuge 5424, Eppendorf). The cleared supernatant was subsequently diluted to a final concentration of 1 mM ANAPOE C₁₂E₈ to facilitate binding to the HisTrap™ column, then filtered (0.45 μm) and loaded into a 1 ml HisTrap™ FF column (GE Healthcare). The following steps in the purification were as described in the previous section, with buffer A containing 50 mM HEPES pH 7.5, 300 mM potassium acetate, 2 mM BME, 1 mM ANAPOE C₁₂E₈, 30 mM imidazole, and buffer B consisting of 50 mM HEPES pH 7.5, 300 mM potassium acetate, 2 mM BME, 1 mM ANAPOE C₁₂E₈, 500 mM imidazole. Fractions of interest were analyzed by SDS-PAGE and Western blotting.

2.3.3.3. Talon® Superflow™ resin

In an attempt to increase the yield of subunit XPD the IMAC step was later modified, using a Talon® Superflow™ resin (GE Healthcare) instead of a HisTrap™ FF column to maximize the contact between the helicase and the resin. The resin was prepared for purification following the manufacturer's indications. Cleared supernatant obtained as previously described was added to 1-2 ml resin and incubated at 4 °C for 1 hour with rotation. Resin and supernatant were then transferred into an empty 10 ml gravity-flow column previously prepared as indicated by the manufacturer (Thermo Scientific). Resin was allowed to settle for 15-30 minutes, and after that time the supernatant was eluted. This step was repeated until all of the resin had been packed and all the supernatant had been eluted. Ten times resin bed volume of buffer A (PBS pH 7.5, 300 mM NaCl, 30 mM imidazole, 10% glycerol) were added to the column and incubated at room temperature for 10 minutes, then slowly eluted to wash the resin. The TFIIH Core sub-complex was eluted with five times resin bed volume of buffer B (PBS pH 7.5, 300 mM NaCl, 500 mM imidazole, 10% glycerol): first 1 ml buffer was eluted and collected. The rest of the volume was incubated with the resin at room temperature for 10 minutes, then eluted and collected in 1 ml fractions. The different fractions were analyzed by SDS-PAGE.

2.3.4. Size exclusion chromatography

The concentrated sample obtained in the metal ion affinity chromatography previously described was loaded into a HiPrep™ 16/60 Sephacryl™ S300 HR column (GE Healthcare) and eluted with 20 mM HEPES pH 7.0, 200 mM NaCl, 1 mM DTT, 10% glycerol. Fractions containing the complex were pooled together and concentrated as described in section 2.3.3.1.

2.3.5. Cation exchange chromatography

In a third purification step, the selected fractions from the size exclusion chromatography were loaded into a MonoS™ 4.6/100 PE column (GE Healthcare) after being concentrated. After a washing step with buffer A (20 mM HEPES pH 7.0, 50 mM NaCl, 1 mM DTT, 10% glycerol), bound proteins were eluted in a gradient of salt created by a combination of buffer A and buffer B (20 mM HEPES pH 7.0, 500 mM NaCl, 1 mM DTT, 10% glycerol). The sample was finally concentrated, aliquoted and stored at -80 °C.

2.3.6. Affinity chromatography

Sub-complex TFIIH Core carrying two 6xHis-V5-TEV tags at the N-terminal end of subunits p52 and p62 and a Twin-Strep-tag® at the C-terminal end of subunit XPD was purified through a Talon® resin first, then through a size exclusion column, and finally through a Strep-Tactin®XT resin (IBA GmbH). A Bio-Rad 1.5 x 20 cm gravity flow Econo-column® packed with the Strep-Tactin®XT matrix was first equilibrated with two column volumes of wash buffer PBS pH 7.5, 300 mM NaCl. Fractions of interest collected after size exclusion were concentrated, added to the column and incubated with the resin for 15 – 30 minutes at room temperature, and after that time the supernatant was eluted. The resin was then washed with one column volume of wash buffer. The TFIIH Core sub-complex was finally eluted with three column volumes of elution buffer (PBS pH 7.5, 300 mM NaCl, 50 mM biotin). The different fractions were analyzed by SDS-PAGE, and the purified sub-complex was finally concentrated, aliquoted and stored at -80 °C.

2.4. Purification of XPC constructs

2.4.1. XPC Extraction and BioSprint expression tests

All XPC proteins (monomeric XPC, TCP-tagged XPC and heterodimer XPC- HR23B) were pelleted, extracted and tested for expression in a BioSprint station following the same protocol employed to obtain the different TFIIH constructs (section 2.3.1).

2.4.2. Immobilised metal ion affinity chromatography

2.4.2.1. HisTrapTM FF column

Cell pellet for the different XPC proteins was harvested by centrifugation at 1000 rpm for 5 minutes at 4 °C (Allegra 21R centrifuge, Beckman Coulter) and resuspended in 5 ml lysis buffer per gram of pellet (PBS, 300 mM NaCl, 0.1% Triton, plus 10 units/ml DNaseI (Sigma) and a protease inhibitor cocktail (Roche)). Suspensions were then passed through a dounce homogenizer (at least 20 strokes, keeping suspension on ice at all times) and centrifuged at 40000 rpm for 45 minutes at 4 °C (Optima L-90K ultracentrifuge, Beckman Coulter). The cleared supernatant was filtered (0.45 µm) afterwards and loaded into a 5 ml HisTrapTM FF column (GE Healthcare). Purification was performed using an Äkta system (GE Healthcare). After washing the column with buffer A (PBS pH 7.5, 363 mM NaCl, 30 mM imidazole, 0.5 mM DTT, 10% glycerol), bound proteins were eluted in an imidazole gradient created by buffer B (PBS pH 7.5, 363 mM NaCl, 300 mM imidazole, 0.5 mM DTT, 10% glycerol). Purified fractions containing the XPC construct were pooled together and concentrated using an Amicon[®] concentrator with a 30K cutoff (Millipore).

2.4.2.2. Talon[®] SuperflowTM resin

The HisTrapTM FF column was replaced by a Talon[®] SuperflowTM resin (GE Healthcare) in later purifications to maximize the contact between the protein and the resin, thus increasing protein recovery. Resin preparation, incubation with the cleared supernatant, and packing into a gravity-flow column was performed as previously described in section 2.3.3.3. The resin was then washed with ten times resin bed volume of buffer A (PBS pH 7.5, 363 mM NaCl, 30 mM imidazole, 0.5 mM DTT, 10% glycerol) and incubated at room temperature for 10 minutes, then slowly eluted to wash the resin. The XPC-HR23B

heterodimer was eluted with five times resin bed volume of buffer B (PBS pH 7.5, 363 mM NaCl, 300 mM imidazole, 0.5 mM DTT, 10% glycerol). 1 ml Buffer was first eluted and collected; the remainder volume was incubated with the resin at room temperature for 10 minutes, then it was eluted and collected in 1 ml fractions that were finally analyzed by SDS-PAGE.

2.4.3. Size exclusion chromatography

The concentrated sample obtained from the metal ion affinity chromatography previously described was loaded into a HiPrep™ 16/60 Superdex™ S200 prep grade column (GE Healthcare) and eluted with PBS pH 7.5, 63 mM NaCl, 0.5 mM DTT, 10% glycerol. Fractions containing the XPC construct were pooled together and concentrated as described in section 2.3.3.1. Purified heterodimer XPC-HR23B was concentrated, aliquoted and stored at -80 °C after this chromatographic step.

2.4.4. Cation exchange chromatography

Fractions containing the XPC protein eluted in the previous size exclusion chromatography were loaded into a MonoS™ 4.6/100 PE column (GE Healthcare) after being concentrated. The column-bound protein was then washed with buffer A (PBS pH 7.5, 0.5 mM DTT, 10% glycerol), and finally eluted in a gradient of salt created by a combination of buffer A and buffer B (PBS pH 7.5, 863 mM NaCl, 0.5 mM DTT, 10% glycerol). TCP-tagged XPC was concentrated, aliquoted and stored at -80 °C after this step.

2.4.5. Heparin chromatography

The purification of monomeric XPC and TCP-tagged XPC required an additional chromatography step (in the case of TCP-XPC, this step replaces the size exclusion chromatography usually performed after IMAC). Fractions of interest collected after cation exchange (XPC) or IMAC (TCP-XPC) were concentrated and loaded into a HiTrap™ Heparin HP column (GE Healthcare). The column was washed with buffer A (PBS pH 7.5), before eluting bound proteins in a gradient of salt created by combining buffer A and buffer B (PBS pH 7.5, 1 M (total) NaCl) in different proportions. The purified protein was then concentrated, aliquoted and stored at -80 °C in the case of monomeric

XPC, while fractions containing TCP-XPC were pooled together, concentrated and loaded into a MonoS™ 4.6/100 PE column for a final cation exchange chromatographic step.

2.5. Biochemical characterization of XPC and the TFIIH

Core sub-complex

2.5.1. ReAsH-EDT₂ Labelling of TCP-tagged XPC and TFIIH Core

Both TCP-tagged XPC and TFIIH Core (with subunit XPB carrying the TCP tag) were labelled with the resorufin-based biarsenical reagent ReAsH carrying two 1,2-ethanedithiol (EDT) molecules (ReAsH-EDT₂) (Adams *et al.*, 2002) (Thermo Fisher Scientific) that becomes fluorescent when bound to the tetracysteine motif present in the TCP tag. Additionally, the rigid attachment of the ReAsH ligand to the tetracysteine motif offers a stronger signal compared to conventional fluorescent reagents that present more flexible bonds (Adams *et al.*, 2002).

A 100 μ l reaction containing 1 mM tris(2-carboxyethyl)phosphine (TCEP), 1 mM BME, 10 μ M ReAsH-EDT₂ and TCP-XPC or TCP-TFIIH Core to a final concentration of 2.2 μ M and 0.57 μ M, respectively, was incubated at 4 °C for 30 minutes (protocol adapted from (Park *et al.*, 2004)). The TCEP and BME added to the reaction will break the disulphide bonds in the ReAsH-EDT₂ reagent so it can bind to the tetracysteine motif present in the TCP tag cloned into N-terminal XPC and N-terminal XPB (within the TFIIH Core sub-complex), becoming fluorescent as a result of the binding.

After that time, ReAsH-labeled TCP-XPC and TCP-TFIIH Core were titrated in a Varian Cary Eclipse fluorescence spectrophotometer (Agilent) using the Scan application included in the software package provided with the instrument. The reagent was excited with a wavelength of 593 nm and the emitted fluorescence was collected at a wavelength of 608 nm. To perform the titration both reactions were diluted 100-fold, and 1 μ l of the corresponding dilution was added to a cuvette containing 119 μ l buffer 20 mM HEPES pH 7.0, 500 mM NaCl, 1 mM DTT in each titration step. A reaction in which buffer 20 mM HEPES pH 7.0, 500 mM NaCl, 1 mM DTT was added instead of TCP-XPC or TCP-TFIIH

MATERIALS AND METHODS

was used as a control to measure the intensity of the non-bound ReAsH reagent. The obtained fluorescence emission intensity values were plotted against monomeric XPC / TFIIH Core concentration using the RStudio software (RStudio, Inc.).

2.5.2. Crosslinking assays

Crosslinking assays were performed in a 20 μ l reaction containing buffer 20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM DTT, 10% glycerol, 10 μ g TFIIH Core sub-complex with subunits p52 and p62 tagged at their N-terminal end with 6xHis-V5-TEV, and different concentrations (0.25 μ g/ μ l, 0.5 μ g/ μ l, 0.75 μ g/ μ l, and 1 μ g/ μ l) of crosslinker bis(sulfosuccinimidyl)suberate (BS3) d0/d4 (Thermo Scientific). The different reactions were incubated at 25 °C with 350 rpm shaking for 1 hour, and then loaded into an SDS-PAGE gel. Bands of interests were cut from the gel and analyzed by mass spectrometry (MS), and the resulting peptides were checked for crosslinks using the MassMatrix software.

2.5.3. DNA substrate preparation

Single strands of DNA were purchased from IDT (strands A to E) or Primetech ALC (strands F, G) and resuspended in 50 mM Tris pH 7.5 to a final concentration of 500 μ M (see appendix D for sequences and substrate composition).

Double stranded substrates employed in the experiments described in sections 2.5.4 and 2.5.5 were obtained by annealing in a 50 μ l reaction containing buffer 50 mM Tris pH 7.5, 100 mM NaCl and the appropriate single strands to a final concentration of 20 μ M. The mixture was heated in a water bath at 90 °C for 5 minutes and left to slowly cool down overnight. A final concentration of 3% Ficoll was added to the reaction now containing our double-stranded substrates and they were purified in a 10% polyacrylamide native gel, run in Tris – Borate – Ethylenediaminetetraacetic acid (EDTA) (TBE) buffer, 10 mM NaCl at 4 °C, 120 V for 3-4 hours, protected from light. Bands of interest were cut, placed in an Eppendorf tube and slightly crushed, then covered with buffer 50 mM Tris pH 7.5. The tubes were incubated at 4 °C overnight with rocking to promote diffusion of the substrates from the acrylamide towards the buffer. Afterwards, the buffer was separated from the acrylamide and the double-stranded substrates were

precipitated with 10% volume 3 M sodium acetate pH 5.2 and 250% volume chilled EtOH 100% and incubated at -20 °C overnight. Finally, the reactions were centrifuged at 15000 rpm at 4 °C for 30 minutes (centrifuge 5424, Eppendorf), the supernatant was immediately removed, and the pelleted DNA substrates were resuspended in an appropriate volume of 50 mM Tris pH 7.5 buffer.

2.5.4. Electrophoretic mobility shift assays (EMSA)

Band shifting assays were performed in a 10 μ l reaction containing buffer 20 mM HEPES pH 7.5, 1 mM DTT, 50 mM NaCl, 1 mM EDTA, 4% glycerol, a 90 nM DNA substrate with a fluorescein label (figure 25, full sequence in appendix D) and either sub-complex TFIIH Core or heterodimer XPC-HR23B at several different concentrations. Reaction mixtures were incubated for 30 minutes at room temperature, protected from light. After that time, 10 μ l Ficoll 15% (Sigma) were added to each sample and they were loaded into a 5 - 10% polyacrylamide native gel which was run at 4 °C and 180 V, protected from light, for 3 hours approximately; afterwards, the gel was scanned in a Typhoon FLA 7000 laser scanner (GE Healthcare). A control reaction that included the DNA substrate but not the TFIIH Core or XPC-HR23B complexes was also performed and loaded into every gel.

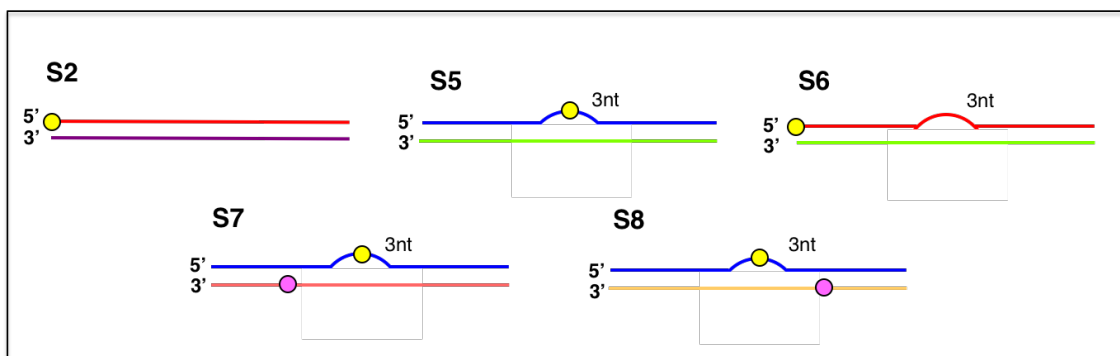


Figure 25: DNA substrates employed in EMSA assays.

Schematic representation of the different DNA substrates carrying a fluorescein label (yellow sphere) used in the band shifting assays performed with heterodimer XPC-HR23B and the TFIIH Core sub-complex. Substrates S7 and S8 also carry a Cy3 dye (pink sphere) at the left (S7) or right (S8) of the three-nucleotide bubble.

2.5.5. Fluorescence experiments

2.5.5.1. Fluorescence anisotropy assays

DNA binding of monomeric XPC, 10-subunit TFIIH and the TFIIH Core sub-complex was further analyzed by measuring fluorescence anisotropy (r). In these experiments, samples were excited with a polarized light, and emitted fluorescence was collected through a perpendicularly oriented polarizer with respect to the excitation light, allowing the calculation of the emission anisotropy (r) based on the equation

$$r = (I_{\parallel} - GI_{\perp}) / (I_{\parallel} + 2GI_{\perp})$$

where G is a geometric factor calculated to correct the polarization bias of the instrument. The fast rotation of our fluorescein-labelled substrates in the reaction solution leads to an efficient depolarization of the light (low r value), while binding to a large molecule such as XPC or the TFIIH sub-complex will cause a restriction in its movement and a less efficient depolarization (high r values). The emission anisotropy can now be plotted against protein concentration, and a K_D value can finally be obtained.

Reactions were carried out in a 150 μ l solution containing 25 nM of a DNA substrate carrying a fluorescein label in buffer 20 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.5 (figure 26, full sequence in appendix D). Later experiments with construct TFIIH Core were performed in buffer 20 mM HEPES pH 7.0, 50 mM NaCl. The solution was loaded into a quartz cuvette and our protein / complex of interest was titrated using the Eclipse ADL application included in the software package provided with the Varian Cary Eclipse fluorescence spectrophotometer (Agilent Technologies), exciting the fluorescein molecule at 480 nm and collecting emitted fluorescence at 525 nm. Geometry factor G was calculated by the program at the beginning of every experiment, and each r value obtained was the resulting average of five measurements taken for every titration point. All experiments were carried out with automatic polarizers at 25 °C. The obtained r values were plotted against XPC-HR23B/ TFIIH Core concentration and the resulting data were fitted to a binding curve with the KaleidaGraph program (Synergy Software), using equation

$$A = A_{min} + [(D + E + K_D) - \{(D + E + K_D)^2 - (4DE)\}^{1/2}] (A_{max} - A_{min}) / (2D)$$

where A is anisotropy, A_{min} is anisotropy of free DNA, D is total DNA concentration, E is total protein concentration, K_D is the disassociation constant and A_{max} is maximum anisotropy of the DNA-XPC-HR23B or DNA-TFIIH Core complex. The equation assumes a DNA:protein binding of 1:1 (Reid *et al.*, 2001).

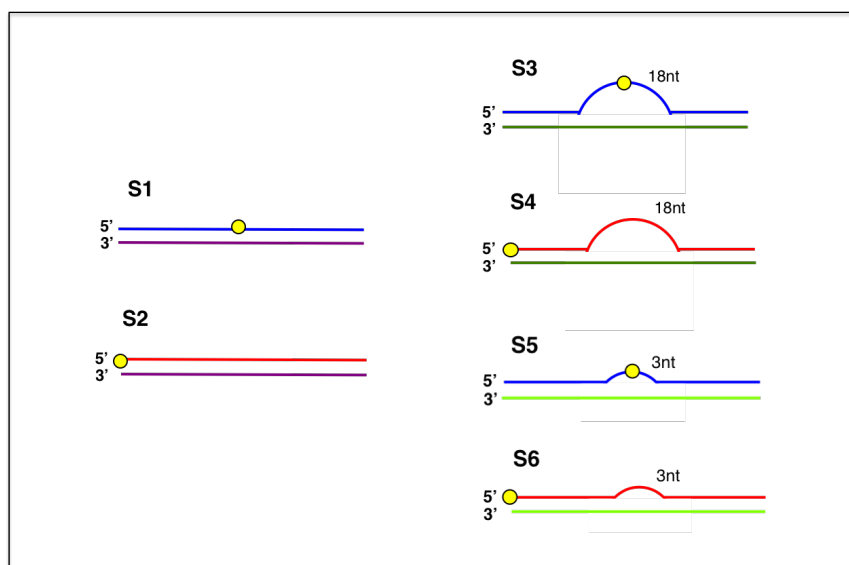


Figure 26: DNA substrates employed in fluorescence anisotropy assays.

Schematic representation of the different DNA substrates tested for binding to monomeric XPC, 10-subunit TFIIH or TFIIH Core by measuring their fluorescence anisotropy. Each substrate carries a fluorescein label (yellow sphere). Substrates S3 and S4 also present an 18-nucleotide bubble, and S5 and S6 carry a 3-nucleotide bubble.

2.5.5.2. Bulk fluorescence resonance energy transfer (FRET) assays

Binding of XPC-HR23B, TFIIH Core and XPC-HR23B-TFIIH Core was also analyzed by bulk FRET, a process based on the non-radiative transfer of energy from a donor molecule to an acceptor one, located in close distance. The XPC-HR23B heterodimer and TFIIH Core sub-complex were titrated in a quartz cuvette containing 150 μ l of a solution of 20 mM HEPES pH 7.0, 50 mM NaCl and a 20 nM DNA substrate S7 or S8 (figure 27, full sequence in appendix D). To analyze the binding of the complex XPC-HR23B-TFIIH Core, the TFIIH Core sub-complex was titrated in a quartz cuvette

MATERIALS AND METHODS

containing 150 μ l of a solution of 20 mM HEPES pH 7.0, 50 mM NaCl, 0.5 mM DTT, a 20 nM DNA substrate S7 or S8 and 100 nM XPC-HR23B heterodimer.

All experiments were carried out at 25 °C in a Varian Cary Eclipse fluorescence spectrophotometer using the Scan application included in the Varian Eclipse software package. The fluorescein label (FRET donor) was excited at 490 nm, while the Cy3 molecule (FRET acceptor) was excited at 542 nm. Ratio A values (defined as the calculated fluorescence signal of the acceptor, in this case Cy3, normalized by the acceptor fluorescence spectrum of the sample (Lorenz & Diekmann, 2006)) were calculated for each data point and plotted against the appropriate complex concentration. In the case of XPC-HR23B and XPC-HR23B-TFIIH Core the resulting data were fitted to a binding curve with the SigmaPlot program (Systat Software Inc.), using the Hill equation (Hill, 1910)

$$f = f_0 + ((b(x)^n) / ((K_D)^n + (x)^n))$$

where f is FRET, f₀ is minimum FRET, b is the amplitude of FRET, x is total protein concentration, K_D is the disassociation constant and n is the Hill coefficient (a measure of the cooperativity of the binding process).

Data corresponding to the binding of XPC-HR23B to substrate S7 was fit to a Hill equation comprising two components

$$f = f_0 + ((b(x)^{n_1}) / ((K_{D1})^{n_1} + (x)^{n_1})) + ((d(x)^{n_2}) / ((K_{D2})^{n_2} + (x)^{n_2}))$$

In the case of TFIIH Core data were fitted to a binding curve, also with the SigmaPlot program, but using equation

$$A = A_{min} + [(D + E + K_D) - \{(D + E + K_D)^2 - (4DE)\}^{1/2}] (A_{max} - A_{min}) / (2D)$$

which assumes a DNA:protein binding of 1:1 (Reid *et al.*, 2001). In this equation A is FRET, A_{min} is minimum FRET, D is total DNA concentration, E is total protein

concentration, K_D is the disassociation constant and A_{max} is maximum FRET.

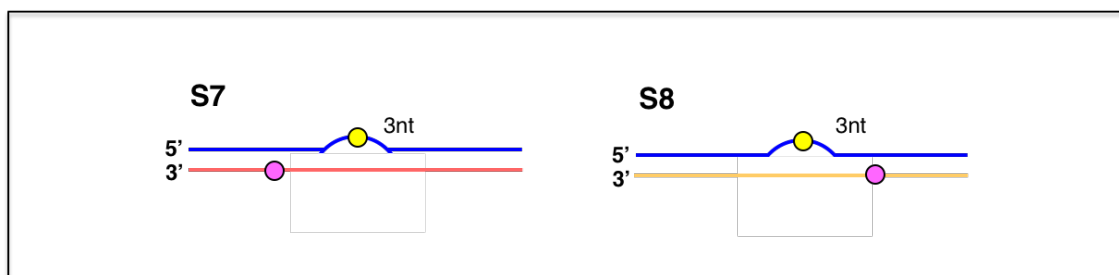


Figure 27: DNA substrates employed in bulk FRET assays.

Cartoon representation of the different DNA substrates employed in bulk FRET experiments to study binding of XPC-HR23B, XPC-HR23B-TFIIH Core and TFIIH Core to double-stranded DNA. Both substrates carry a fluorescein label (yellow sphere) and a Cy3 dye (pink sphere) at the left (S7) or right (S8) of the 3-nucleotide bubble.

2.5.5.3. Fluorescence-based helicase assays

The unwinding activity of TFIIH Core was evaluated in a fluorescence helicase assay employing a series of substrates in which a quencher, placed to suppress a Cy3 dye, will be removed upon unwinding by the sub-complex, thus permitting the detection of the fluorescence emitted by Cy3. Initial helicase activity assays were carried out at 25 - 37 °C in a 150 μ l reaction containing buffer 20 mM HEPES pH 7.0, 50 mM NaCl, 0.1 mg/ml bovine serum albumin (BSA), 1 mM $MgCl_2$, 25 - 50 nM DNA substrate containing a Cy3 dye (figure 28) and 50 nM – 1 μ M TFIIH Core sub-complex, with 0.5 – 1mM ATP added to the cuvette containing the full reaction or to the empty cuvette, to have the rest of the reaction loaded in a second step. Optimal conditions for the assay were finally established as 20 mM HEPES pH 7.0, 50 mM NaCl, 0.1 mg/ml BSA, 1 mM $MgCl_2$, 25 nM DNA substrate 50 nM TFIIH Core sub-complex, performed at 25 °C. The reaction mixture was then incubated at room temperature for 10 minutes before being added to a cuvette containing 1 mM ATP, previously incubated at room temperature for 10 minutes, too. In the assays evaluating the effect of proteins XPA, RPA or both on TFIIH Core unwinding the reaction just described (containing the TFIIH Core sub-complex) was incubated at room temperature for 9 minutes approximately, then 50 nM XPA, RPA or both (as required) was added to the reaction and the final mixture was added to the quartz cuvette containing 1 mM ATP after a total incubation time of 10 minutes. Activity

MATERIALS AND METHODS

was measured for 15 minutes in a Varian Cary Eclipse fluorescence spectrophotometer using the Kinetics application included in the software package provided with the instrument. The Cy3 dye was excited at a wavelength of 547 nm, and emitted fluorescence was collected at a wavelength of 565 nm. The obtained fluorescence emission intensity values were plotted against TFIIH Core concentration and data were fitted to a bi-exponential model with the SigmaPlot program (Systat Software Inc.), using equation:

$$F = F_0 + A \cdot \exp(-k_1 \cdot t) + B \cdot \exp(-k_2 \cdot t)$$

The equation describes a model with a burst phase ($A \cdot \exp(-k_1 \cdot t)$, with A being the amplitude and k_1 the rate for the burst phase) and a slow phase ($B \cdot \exp(-k_2 \cdot t)$, with B being the amplitude and k_2 the rate for the slow phase).

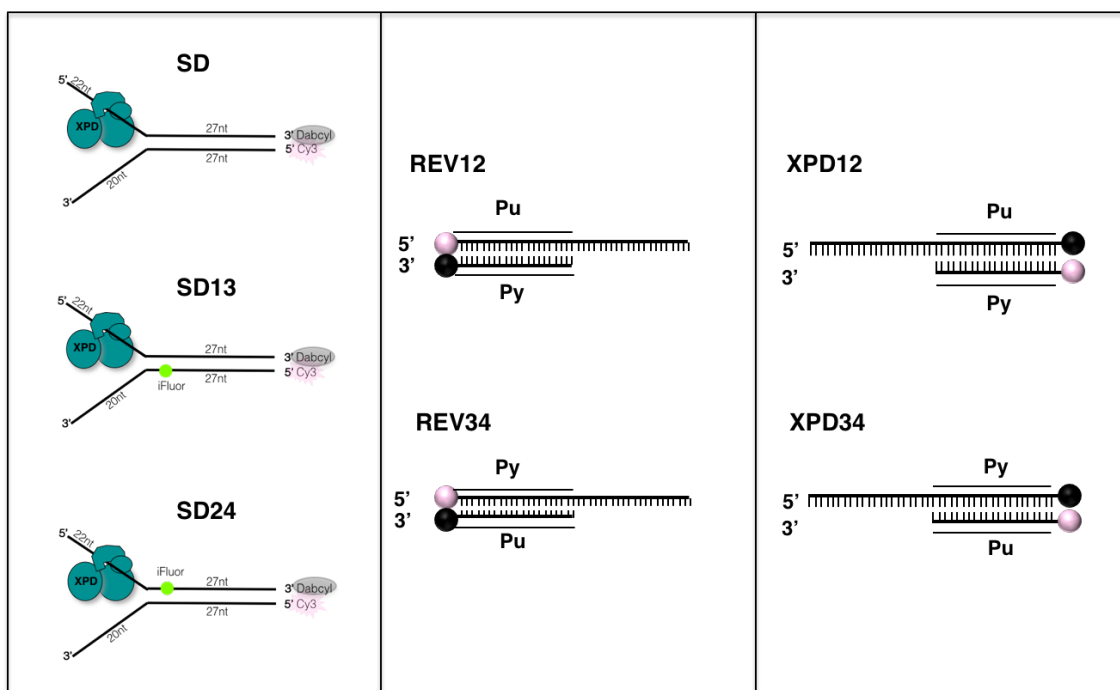


Figure 28: DNA substrates employed in helicase activity assays.

Schematic representation of the different DNA substrates used in the helicase assays performed to study opening of a DNA duplex by the TFIIH Core sub-complex. All substrates carry a Cy3 dye (pink spheres) suppressed by a quencher (black spheres).

3. XPC: the protein that starts it all

3.1. Introduction

3.1.1. *In vivo* XPC is a heterotrimeric complex

Human XPC is a 125 kDa protein, flexible and intrinsically unstable, that has an essential role in NER as a damage detector (Sugasawa *et al.*, 1998). XPC also participates in the BER DNA repair pathway (previously described in chapter 1, section 1.1), and has an additional role in transcription regulation as a co-activator of pluripotency genes, supposedly acting as a bridge between transcription factors octamer-binding transcription factor 4 (OCT4) and sex- determining region Y (SRY)-box 2 (SOX2) and the transcription apparatus to increase the expression of said genes (Zhang *et al.*, 2015).

As it was already explained in chapter 1, section 1.2.2.1, the GGR reaction starts with protein XPC detecting a distortion in the double helix anywhere in the genome (Sugasawa *et al.*, 1998). *In vivo* XPC is part of a heterotrimeric complex consisting of XPC itself and additional proteins HR23 (paralogs A or B) and CETN2 (Sugasawa *et al.*, 1996) (Nishi *et al.*, 2005).

HR23A and HR23B are the mammalian homologs of *Saccharomyces cerevisiae* protein RAD23, with a potentially redundant function in DNA repair (Ng *et al.*, 2003). XPC has been reported to bind preferentially to HR23B, but this appears to be caused by a greater abundance in the cell of HR23B over HR23A, with HR23A binding more to XPC in the absence of HR23B (Okuda *et al.*, 2004). Only a small part of HR23A / HR23B is bound to XPC, which points to an additional function of these two proteins. In the context of NER, they appear to stabilise XPC by inhibiting its polyubiquitination, thus preventing the degradation of XPC by the 26S proteasome (Ng *et al.*, 2003). A further role has been proposed in which the XPC-HR23 complex is the real damage detector, with HR23 acting as a stabilising partner for XPC to help it reach the site of the lesion properly folded, and then quickly dissociating from it (potentially induced by a conformational change in XPC or by influence of the subsequently recruited NER factors) as XPC is now stabilised by its binding to the damaged site (Bergink *et al.*, 2012).

XPC: THE PROTEIN THAT STARTS IT ALL

CETN2 is a small calcium-binding protein that augments XPC's binding activity to damaged DNA (Nishi *et al.*, 2005), although it is reportedly not required for the *in vitro* NER reaction (Araki *et al.*, 2000). The N-terminal domain of CETN2 is thought to enhance the recruitment of XPA to the XPC-TFIIH complex already positioned at the lesion, while its C-terminal domain alone has been shown to increase the binding affinity of XPC to damaged DNA (Nishi *et al.*, 2013). As with HR23B, the largest fraction of the protein is not bound to XPC, suggesting additional functions of CETN2 outside DNA repair (Nishi *et al.*, 2005)

3.1.2. XPC structure: what we know

Our structural knowledge of human XPC is so far limited to a few domains (Okuda *et al.*, 2015), since a high-resolution structure of the protein has still not been obtained. XPC shares some conserved domains with its *S. cerevisiae* homologue Rad4, whose X-ray structure was published in 2007 (Min & Pavletich, 2007). Both XPC and Rad4 present an N-terminal α/β domain that contains the transglutaminase-homology domain (TGD), and three α/β domains next to it, named β -hairpin domain 1 (BHD1), 2 (BHD2) and 3 (BHD3) because of the presence of a characteristic long β -hairpin in all three of them (figure 29).

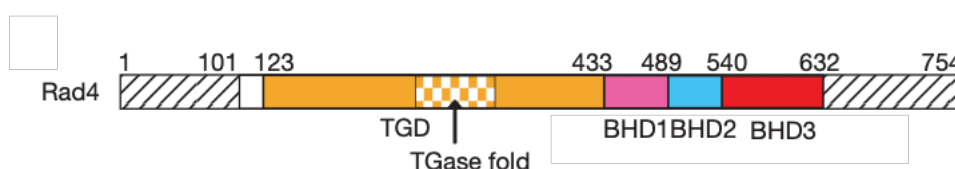


Figure 29: Cartoon view of XPC's homologue Rad4.

The representation includes conserved domains TGD (orange), BHD1 (pink), BHD2 (blue), and BHD3 (red) (Min & Pavletich, 2007).

A structured region towards the otherwise disordered N-terminal domain of XPC was reported to interact with repair factor XPA, while the C-terminal domain interacts physically with both HR23B and CETN2 (Bunick *et al.*, 2006).

XPC recruits TFIIH through interactions with the p62 and, to a lesser extent, XPB

subunits within the complex (Yokoi *et al.*, 2000) (Uchida *et al.*, 2002) (Bernardes de Jesus *et al.*, 2008). An unstructured region towards the N-terminus of XPC becomes an organized tertiary structure when bound to the pleckstrin homology (PH) domain in subunit p62, forming a string-like structure that broadly covers one side of said domain (figure 30). This binding is characterized by extensive electrostatic interactions with basic residues and insertion of a tryptophan and a valine residue in XPC into a binding pocket in p62. Mutation of these residues severely compromises damage removal from the DNA because of XPC's impaired interaction with TFIIH (Okuda *et al.*, 2015).

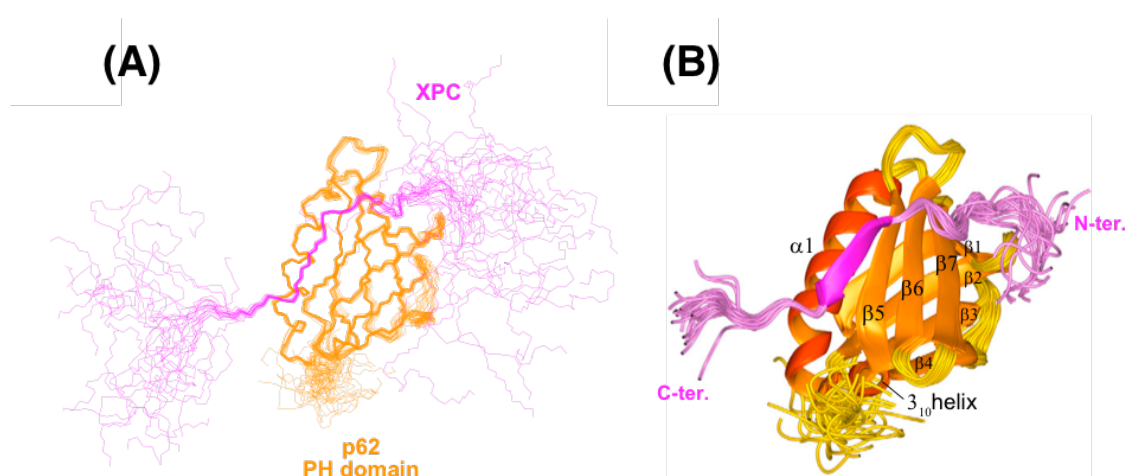


Figure 30: XPC interacts with TFIIH subunit p62.

Representation of the backbone structure of XPC (pink) bound to the PH domain of TFIIH subunit p62 (orange) (A) and cartoon view (B) of the binding (Okuda *et al.*, 2015)

Furthermore, the C-terminal domain of XPC interacts with the N-terminal domain of XPB, and this interaction seems to reinforce the correct positioning of TFIIH at the damaged site, while also stimulating the ATPase activity of the helicase, launching the opening of the double helix at the bubble (Bernardes de Jesus *et al.*, 2008).

3.1.3. XPC detects mismatched bases in the DNA

The binding of Rad4 to damaged DNA occurs in two places: the TGD and BHD1 domains bind to the undamaged portion of the DNA duplex 3' to the lesion, while the BHD2 and BHD3 domains bind to the small segment containing the lesion so BHD3 can introduce

XPC: THE PROTEIN THAT STARTS IT ALL

its β -hairpin through the duplex and make the two bases affected by the lesion flip out of the structure, resulting in a greater distortion of the double helix. BHD2 and BHD3 form a groove that now interacts with the two residues flipped out in the undamaged strand – therefore, there is no actual contact between Rad4 and the lesion (figure 31) (Min & Pavletich, 2007) (Sugasawa, 2016). This recognition of mismatched normal bases rather than identification of the actual damage (Maillard *et al.*, 2007) explains the vast range of lesions detected and repaired by NER, but it also underlines the need for a mechanism to verify the presence of damage in the double helix before triggering the repair reaction.

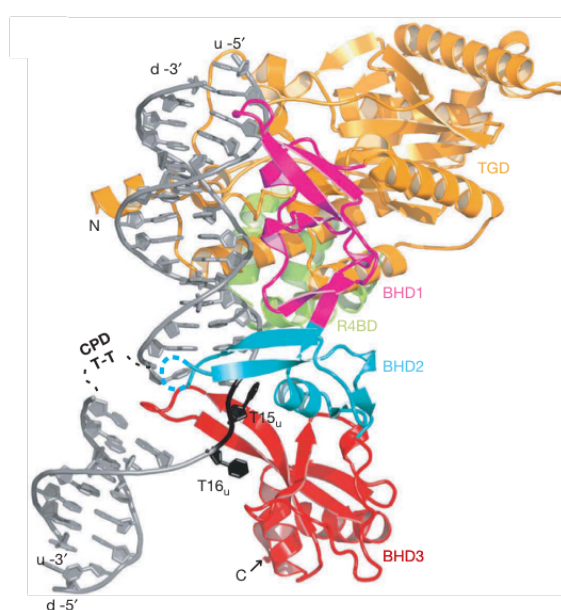


Figure 31: XPC detects distortions in the DNA.

Ribbon diagram showing Rad4 bound to a DNA substrate (depicted by the grey double helix) carrying a CPD lesion. The TGD domain is represented in orange, BHD1 in pink, BHD2 in blue and BHD3 in red (Min & Pavletich, 2007).

Although the conservation between the overall sequences of XPC and Rad4 is not very high, the sequence identity between the BHD1, BHD2 and BHD3 domains and the similarity of their predicted structures suggest that binding of XPC might occur in a similar manner as that of Rad4. Their differences seem to be concentrated in areas associated with the role of XPC as a co-activator, consistent with Rad4 not being active in an *in vitro* transcriptional assay (Zhang *et al.*, 2015).

Experiments carried out with a wide range of DNA substrates suggest that XPC acts as a general sensor for DNA instability, rather than being a specific NER damage detector (Shell *et al.*, 2013). The current model of damage detection proposes that XPC-HR23B scans the DNA in search of mismatches in the double helix by binding to it transiently (Hoogstraten *et al.*, 2008), and upon finding one, a β -hairpin is inserted between the two strands. Only the BHD1 and BHD2 domains and a β -turn structure actually participate in the rapid scanning of the DNA in search of mismatches, and only when one is found the BHD3 domain proceeds to insert its β -hairpin (figures 31 and 32), causing a considerable rearranging of the DNA structure. This two-step mechanism makes the damage-sensing process more energetically favorable (Camenisch *et al.*, 2009). It's the formation of this more stable, longer-lived DNA-XPC complex that serves as a recruitment signal for downstream repair factors TFIIH, XPA and RPA. A further step to verify the lesion is now required before progressing with the repair reaction (Min & Pavletich, 2007)(Shell *et al.*, 2013).

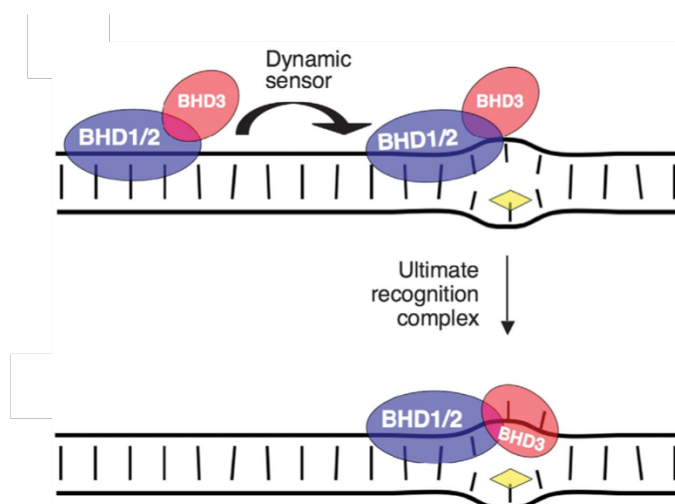


Figure 32: Model showing the two-step detection of lesions in the DNA by XPC.

The BHD1 and BHD2 domains, in conjunction with a β -turn structure, quickly scan the double helix in search of mismatches, and upon finding one the insertion of a β -hairpin between the two strands, carried out by domain BHD3, promotes the formation of a more stable DNA-XPC complex (Camenisch *et al.*, 2009).

As the first component to bind the damaged site in the double helix, thus initiating the repair bubble, we took an interest in how XPC binds to and alters the structure of the

DNA, specifically in the way DNA-bound XPC recruits and interacts with TFIIH at the repair bubble.

In order to study this interaction, we cloned and produced XPC using *Sf9* insect cells as an expression system. We have been able to obtain both XPC and its partner HR23B, allowing us to perform DNA binding assays demonstrating XPC's binding to a variety of damaged substrates, pointing us in the right direction to continue with a more in-depth investigation regarding the interactions between XPC and TFIIH.

3.2. Results

3.2.1. Cloning, expression and purification of XPC

3.2.1.1. Monomeric XPC

3.2.1.1.1. Cloning and expression tests

We cloned the *XPC* gene into a pACEBac2 plasmid containing the N-terminal tag 8xHis-V5-spacer-TEV using enzymes *NcoI* and *KpnI* (Fermentas). The recombinant plasmid was transposed into the baculoviral genome, and the resultant bacmid was purified and later transfected into a monolayer of *Sf9* cells. A 25 ml *Sf9* suspension culture was infected with 3 ml of the recovered P0 viral stock and incubated at 27 °C for 60 hours. After that time the pellet was harvested and analyzed by small-scale affinity purification with magnetic Ni-NTA beads (BioSprint), as described in chapter 2, section 2.3.1. Analysis of the eluted fraction in an SDS-PAGE gel showed that expression levels for this 8xHis-V5-spacer-TEV-tagged XPC were high (figure 33).

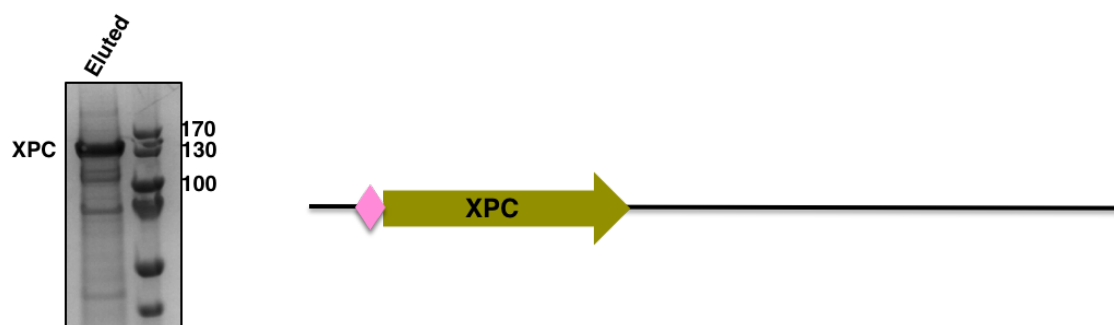


Figure 33: Purification test of monomeric XPC.

Coomassie-stained gel showing a BioSprint purification test of monomeric XPC tagged at its N-terminal end with 8xHis-V5-spacer-TEV (represented by a pink diamond in the cartoon view). Ladder sizes are indicated in kDa.

3.2.1.1.2. Large-scale purification

Scaling-up and optimization of the large-scale infections and purification of monomeric XPC was carried out by Mrs Biljana Petrovic-Stojanovska. We initially followed a protocol comprising the following steps: immobilised metal ion affinity (figure 34, blue box), followed by size exclusion (figure 34, grey box) and finally cation exchange chromatography (figure 34, yellow box), as described in chapter 2, sections 2.4.2.1, 2.4.3 and 2.4.4. While this approach worked well enough, the need of an extra step to further purify XPC was evidenced by the presence of persistent contaminants after the cation exchange chromatography.

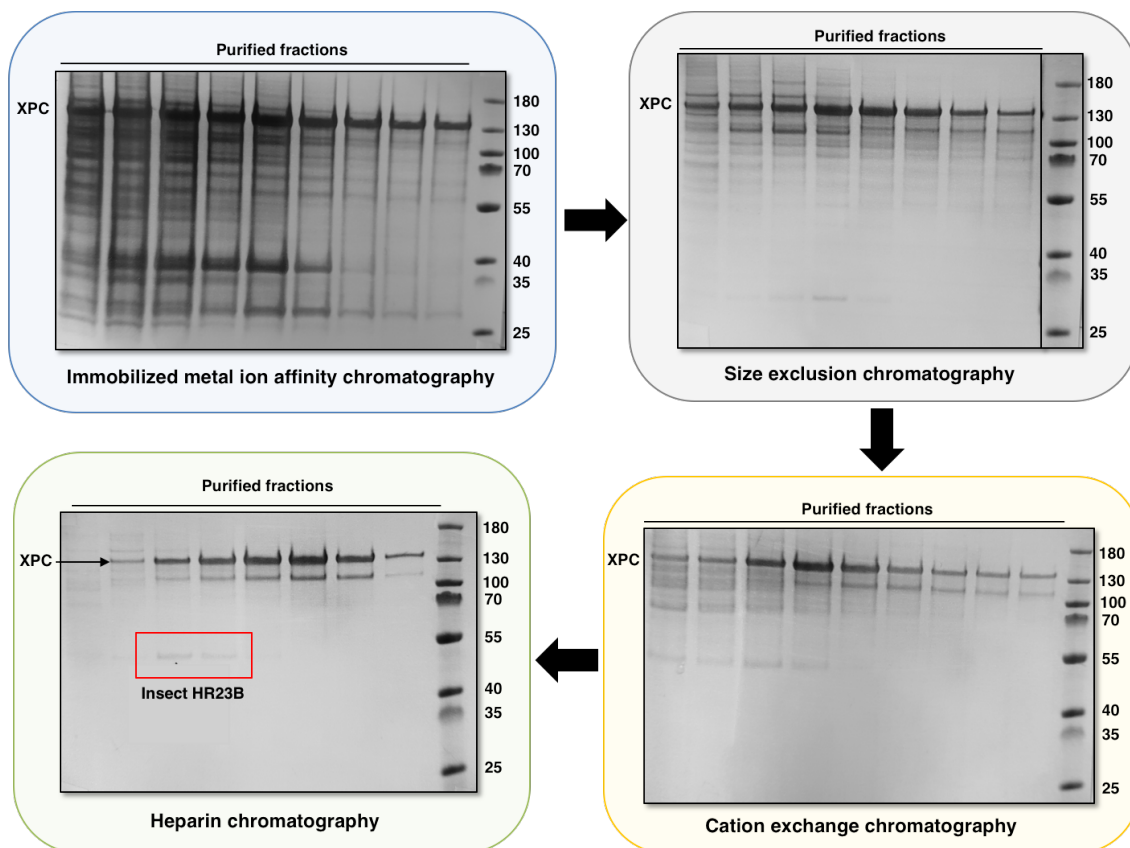


Figure 34: Large-scale purification of monomeric XPC.

The purification of monomeric XPC was carried out in four steps, shown in these Coomassie-

XPC: THE PROTEIN THAT STARTS IT ALL

stained gels. Ladder sizes are indicated in kDa. The blue box shows an SDS-PAGE gel containing eluted fractions from a HisTrap™ FF column (IMAC); this step was followed by a size exclusion chromatography (HiPrep™ 16/60 Superdex™ S200 prep grade column, grey box), a cation exchange chromatography (MonoS™ 4.6/100 PE column, yellow box) and finally a heparin chromatography (HiTrap™ Heparin HP column, green box). The two unspecific bands accompanying pure XPC in the SDS-PAGE gel showing fractions eluted from the heparin column were identified by MS as XPC degradation (approximately 110 kDa), and the insect cell homologue of HR23B (approximately 50 kDa). Data supplied by Mrs Biljana Petrovic-Stojanovska.

As XPC binds DNA, fractions of interest were pooled together and run through a HiTrap™ Heparin HP column (figure 34, green box) following the method described in chapter 2, section 2.4.5. Fractions containing XPC were analyzed in an SDS-PAGE gel that showed that two unspecific bands persisted even after this fourth chromatographic step. MS analysis of both bands showed that the upper band of about 110 kDa was a degraded form of XPC. Unexpectedly, a second band of about 50 kDa was identified as the *Spodoptera frugiperda* homologue of HR23B, the human partner of XPC (figure 34, green box).

3.2.1.2. TCP-Tagged XPC

3.2.1.2.1. Cloning and expression tests

The cloning of a sequence encoding the 6-residue Cys-Cys-Pro-Gly-Cys-Cys tag into monomeric XPC and labelling of the tagged-protein with the red-emitting (608 nm) biarsenic ligand ReAsH-EDT₂ will allow us to study XPC binding to a series of DNA substrates in a fluorescence anisotropy assay.

The gene encoding monomeric protein XPC had previously been cloned into a pACEBac2 plasmid containing the N-terminal tag 8xHis-V5-spacer-TEV. A sequence encoding the CCPGCC tetracysteine tag was later cloned into this construct using restriction enzyme *NcoI*, and the presence of the tag was verified by sequencing. The recombinant plasmid was transposed into the baculoviral genome, and afterwards the purified bacmid was transfected into *Sf9* cells. Later, a 25 ml *Sf9* suspension culture was infected with 3 ml of a P0 viral stock, and after a 60-hour incubation the pellet was harvested, processed and analyzed in a BioSprint station as previously described. SDS-

PAGE analysis of the eluted fraction showed that expression levels for the TCP-tagged XPC are considerably lower than previous results obtained with monomeric XPC when the *S9* cells are infected with a P0 viral stock (figure 35).

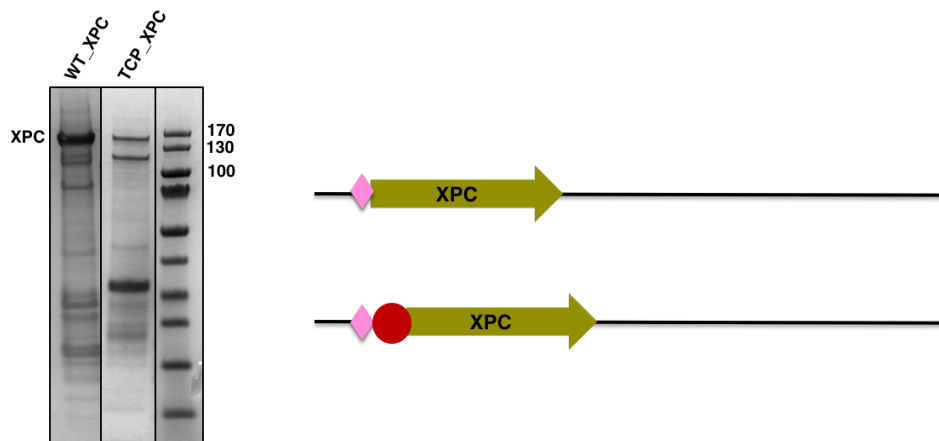


Figure 35: Purification test of TCP-tagged monomeric XPC.

SDS-PAGE gel showing BioSprint tests for monomeric XPC and TCP-XPC, both of them tagged at N-terminal with 8xHis-V5-spacer-TEV (pink diamond in the cartoon representation). TCP-XPC also includes a tetracysteine CCGPCC tag (red sphere) at N-terminal, right after the TEV cleavage site. Ladder sizes are indicated in kDa.

3.2.1.2.2. Large-scale purification

Scaling-up and optimization of the large-scale infections and purification of TCP-tagged XPC was carried out by Mrs Biljana Petrovic-Stojanovska. This work was carried out in parallel with the cloning and expression of monomeric XPC, hence the absence of human HR23B in this construct. A very faint band matching the one corresponding to the *S9* HR23B protein identified by MS in our monomeric XPC purification can also be seen here (figure 36, yellow box).

The low expression levels of TCP-XPC shown in the SDS-PAGE gel containing fractions eluted from the IMAC step (figure 36, blue box) led us to reconsider the protocol employed in the purification of monomeric XPC, finally opting for removing the size exclusion chromatography and simplifying our protocol to three chromatographic steps in an attempt to reduce the loss of TCP-XPC (figure 36).

XPC: THE PROTEIN THAT STARTS IT ALL

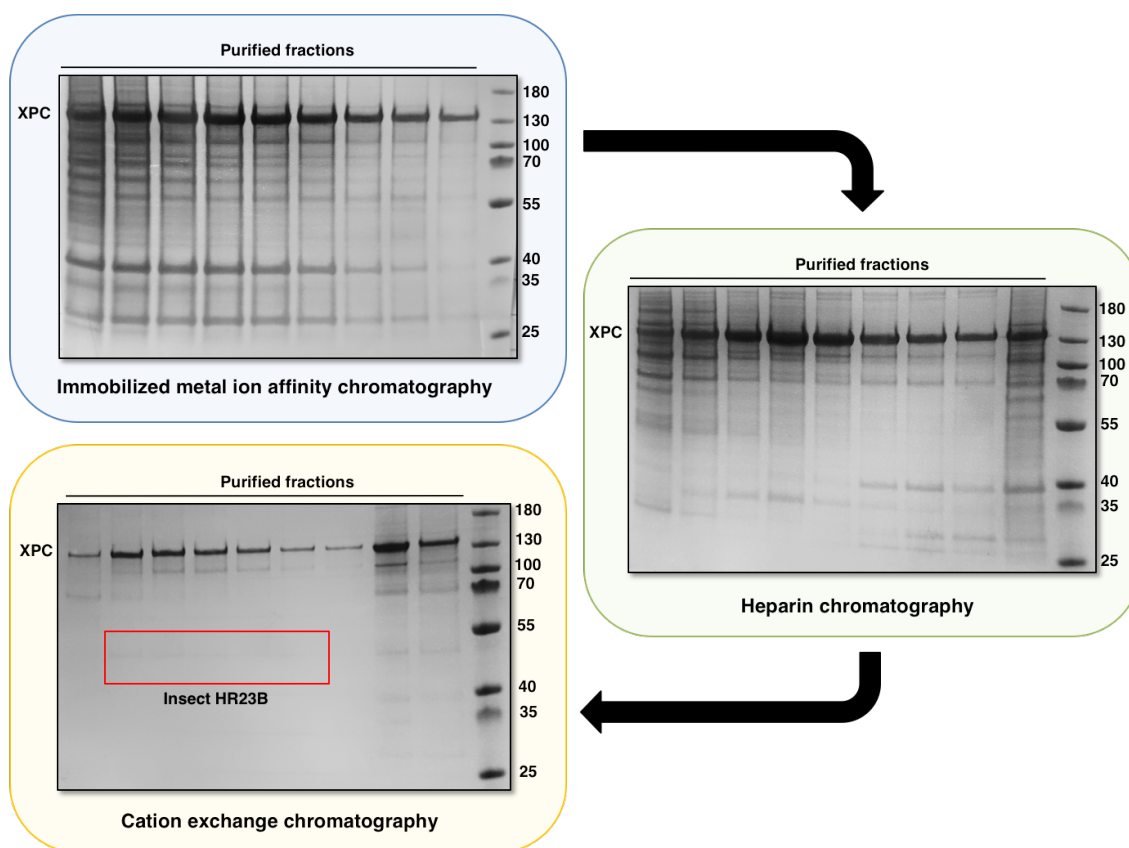


Figure 36: Large-scale purification of TCP-tagged monomeric XPC.

The purification protocol for TCP-tagged XPC was simplified to three chromatographic steps, depicted in these SDS-PAGE gels: the protein was first loaded into a HisTrap™ FF column (IMAC, blue box), followed by a HiTrap™ Heparin HP column (heparin chromatography, green box), and finally a MonoS™ 4.6/100 PE column (cation exchange chromatography, yellow box). Ladder sizes are indicated in kDa. Data supplied by Mrs Biljana Petrovic-Stojanovska.

A faint band for a degraded form of XPC can still be found in the gel accompanying our pure TCP-tagged XPC, and the *Sf9* homologue of HR23B, although barely seen, is still present, too (figure 36, yellow box).

3.2.1.2.3. ReAsH-EDT₂ Labelling of TCP-tagged XPC

The attachment of a label to a protein to determine its cellular localization or as an aide to perform a biochemical assay is a very useful and widely spread technique. The development of the small Cys-Cys-XXX-XXX-Cys-Cys tag meant an enormous advantage over other technologies involving huge inserts that can potentially affect the

folding or even the function of our protein of study. This short motif works in conjunction with a small, membrane-permeable biarsenical ligand whose binding affinity for the target sequence is in the nanomolar range, and which becomes fluorescent upon binding to the thiol groups of the four cysteine residues present in the TCP sequence (Griffin *et al.*, 1998) (figure 37). Optimization experiments reduced the length of the tag from the original 17 amino acids to the current six-residue Cys-Cys-XXX-XXX-Cys-Cys sequence, with XXX being any amino acid other than cysteine, although a stronger binding has been observed when these two amino acids are proline and glycine (Adams *et al.*, 2002).

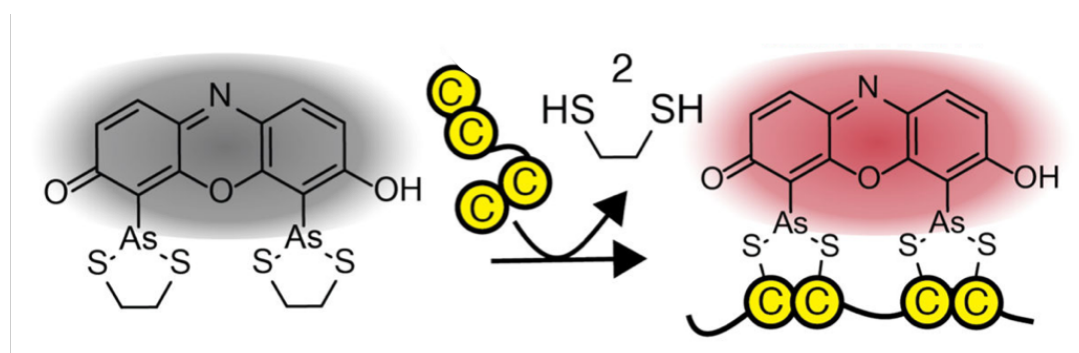


Figure 37: ReAsH-EDT₂ labelling of a TCP-tagged protein.

Adapted from (Walker *et al.*, 2016).

The original tag was initially developed to be used in conjunction with the green-emitting FIAsh-EDT₂ ligand (emission at 528 nm) (figure 38 (A)), obtained by introducing two arsenic (III) substituents in positions 4' and 5' into a fluorescein molecule, to which two molecules of EDT were also added to avoid binding of the ligand to endogenous cysteine residues (Griffin *et al.*, 1998). The position of the two arsenic atoms permits binding to the four cysteine residues in the tag with significant specificity.

The later developed, red-emitting (608 nm) phenoxazine analogue of FIAsh-EDT₂, the resorufin-based ReAsH-EDT₂ reagent (figure 38 (B)) additionally gives the advantage of a reduced cellular absorbance, scattering and autofluorescence compared to the original green-emitting molecule, due to its longer wavelength (Adams *et al.*, 2002).

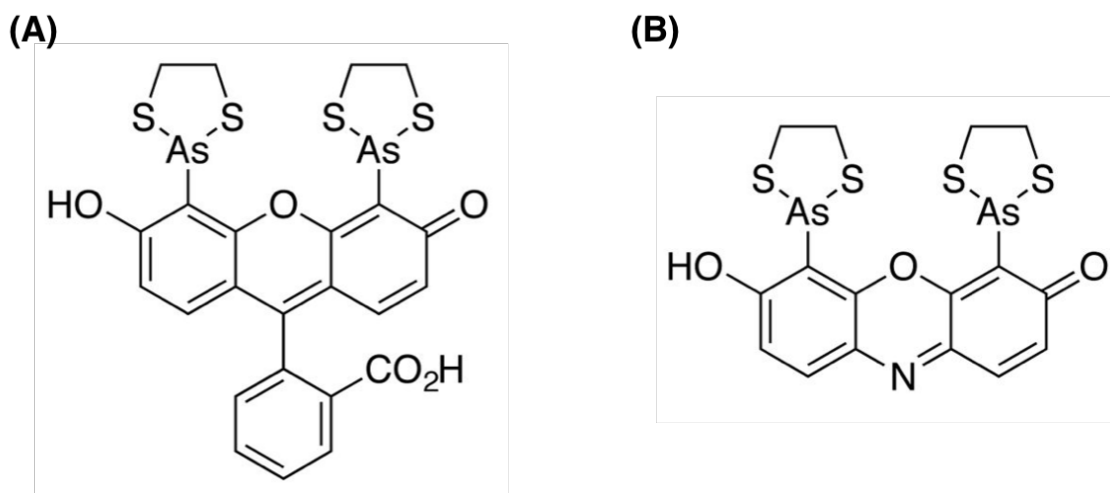


Figure 38: Structure of biarsenic ligands FIAsh-EDT₂ (A) and ReAsH-EDT₂ (B)

(adapted from (Walker et al., 2016)).

This new technology not only permits the localization of tetracysteine-containing proteins by means of labelling, but the high affinity of the binding between the biarsenical ligands and the tetracysteine motif means that the immobilised ligand could potentially be employed in the affinity purification of proteins containing this tag, as the ligand-bound protein can easily be eluted with mM concentrations of a competing ligand such as 2,3-dimercaptopropanesulfonate (DMPS). Our specific interest focuses on another useful application of this small tag: as opposed to other fluorescent ligands with more flexible bonds that offer a reduced signal, the rigid attachment of the biarsenical ligand to the tetracysteine motif now present in proteins XPC and XPB is perfect to perform anisotropy assays (Adams *et al.*, 2002).

ReAsH-EDT₂ labelling of TCP-tagged XPC was carried out as described in chapter 2, section 2.5.1.

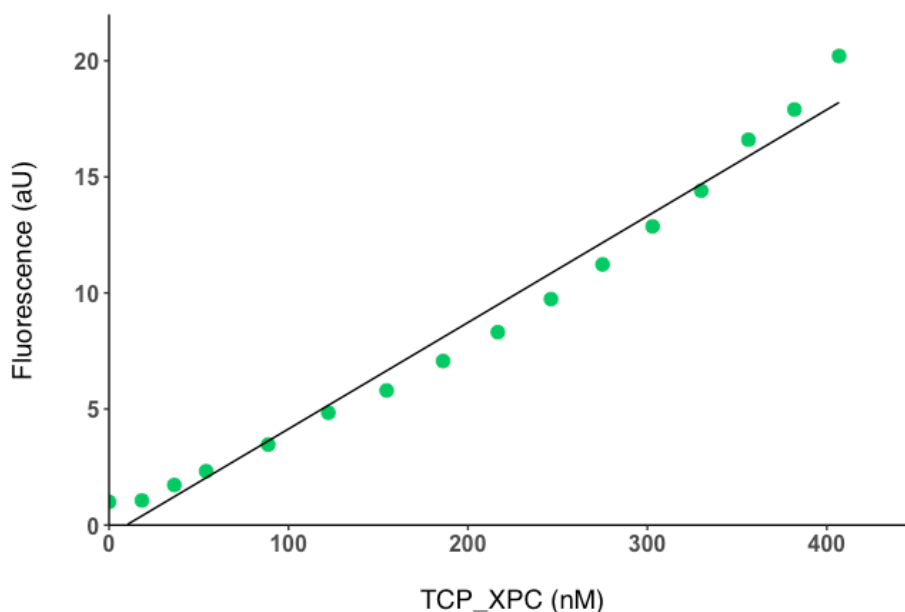


Figure 39: ReAsH-EDT₂ labelling of construct TCP-XPC.

The ligand was excited at a wavelength of 593 nm and emitted fluorescence was collected at a wavelength of 608 nm.

The first measured point shown in figure 39 corresponds to a control reaction in which no TCP-XPC was added, allowing us to determine the intensity corresponding to the free ReAsH-EDT₂ reagent. As the intensity of fluorescence emission (collected at 608 nm) increases with the concentration of TCP-XPC, we conclude that this increment is due to the binding of the ReAsH-EDT₂ ligand to the tetracysteine tag present in XPC, and so our protein can be successfully labelled.

3.2.1.3. XPC-HR23B

3.2.1.3.1. Cloning and expression tests

The human *HR23B* gene was cloned by Mrs Biljana Petrovic-Stojanovska, who also tested the protein for expression in *Sf9* insect cells. The *HR23B* gene was cloned into a pACEBac2 plasmid containing the N-terminal tag 8xHis-V5-spacer-TEV using the enzymes *NcoI* and *KpnI*, and it was later introduced into our pACEBac2 plasmid already

XPC: THE PROTEIN THAT STARTS IT ALL

carrying the *XPC* gene using the multiplication module I-*CeuI* / *BstXI* present in both plasmids, following the method described in chapter 2, section 2.1.4. The bacmid resulting from the successful transposition of the plasmid containing both genes into the baculoviral genome was purified, transfected into a monolayer of *Sf9* cells, and tested for expression in a 25 ml *Sf9* infection, carried out with 3 ml of a P0 viral stock and incubated at 27 °C for 60 hours. Afterwards, the pellet was harvested and analysed as previously described. As already observed in the expression tests for monomeric XPC, expression levels were excellent, and proteins were produced in stoichiometric amounts when extracted from a small-scale infection (figure 40).

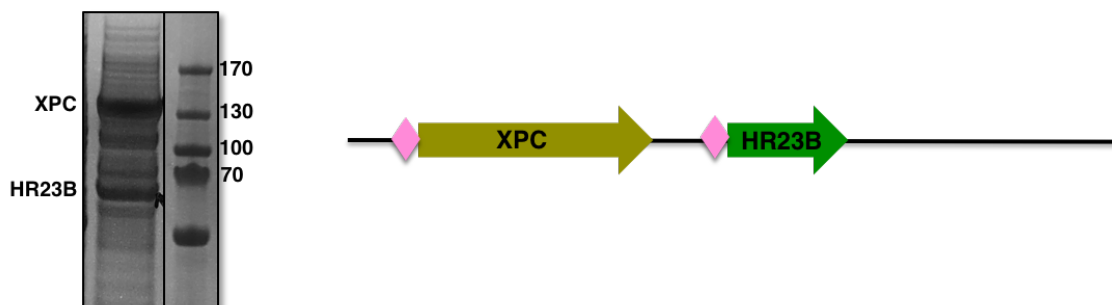


Figure 40: Purification test of heterodimer XPC-HR23B.

The SDS-PAGE gel shows the fraction eluted in a BioSprint purification test, containing both XPC and its partner HR23B, both carrying an 8xHis-V5-spacer-TEVtag at their N-terminal end (pink diamond in the cartoon view). Ladder sizes are indicated in kDa. Data supplied by Mrs Biljana Petrovic-Stojanovska.

3.2.1.3.2. Large-scale purification

The scaling-up of *Sf9* suspension cultures infected with the XPC-HR23B heterodimer presented a new problem: although expression levels of both proteins were good when extracted from a 25 ml infection, they were considerably reduced when extracted from a larger infection (500 ml – 1 L infections).

Optimization of our purification protocol lead us to employ a similar approach as to that used later in the optimization of the TFIID Core sub-complex, using a cobalt Talon® Superflow™ resin (IMAC (figure 41, blue box)) first, followed by a HiPrep™ S200 column

(size exclusion chromatography (figure 41, grey box)). This change in the protocol was forced by the amount of protein obtained after the IMAC step, compelling us to consider which of the columns previously tested for XPC would be more appropriate in this case. We finally opted for a single size exclusion step instead of our previous two-step heparin and cation exchange protocol to minimise protein loss.

The chromatograph corresponding to the size exclusion column showed two peaks: a first peak containing the heterodimer (figure 41, grey box, peak 1) and second, smaller peak that contained faint traces of XPC, but not its partner HR23B (figure 41, grey box, peak 2).

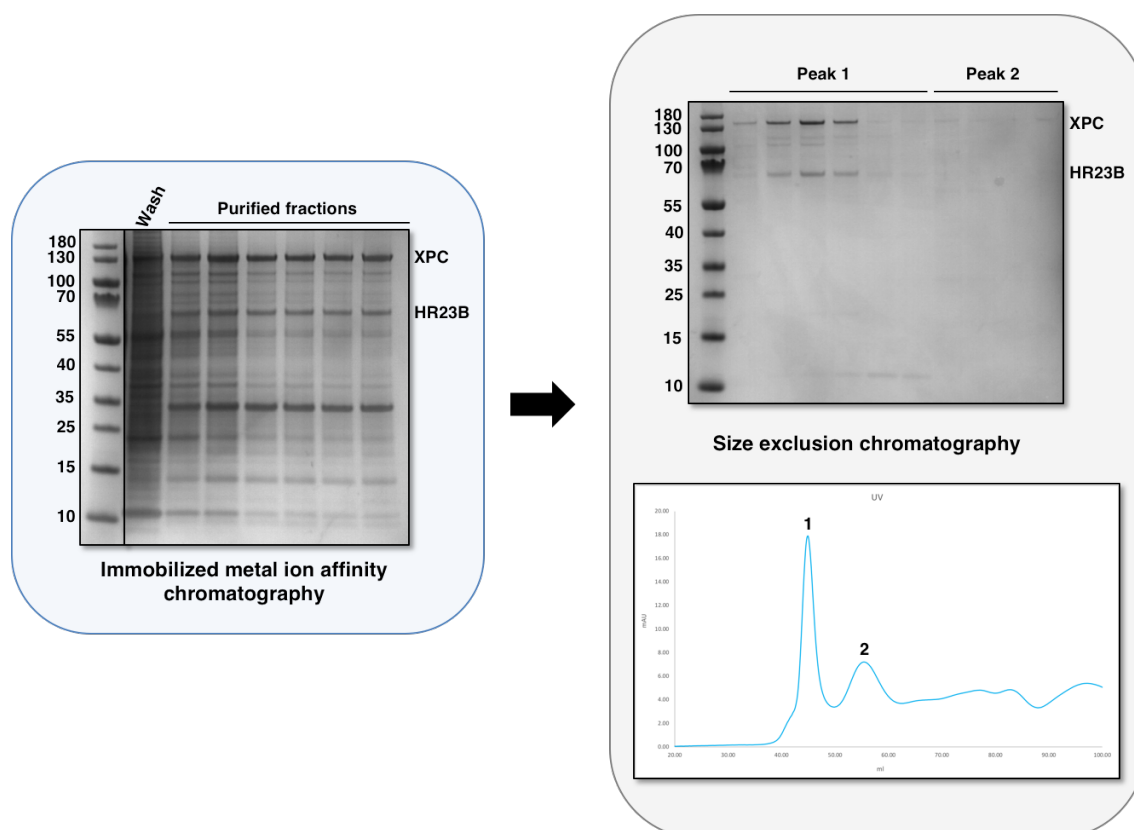


Figure 41: Large-scale purification of heterodimer XPC-HR23B.

Coomassie-stained gels showing the different chromatographic steps followed in the purification of heterodimer XPC-HR23B, both carrying an 8xHis-V5-spacer-TEV tag at their N-terminal end. The purification protocol includes two steps: ion affinity (Talon® Superflow™ resin, blue box), and size exclusion (HiPrep™ 16/60 Superdex™ S200 prep grade column, grey box). Ladder sizes are indicated indicated in kDa.

3.2.2. Biochemical characterisation of XPC

Once we obtained the purified XPC (monomer or HR23B-bound) we tested it for binding to a 44 bp DNA duplex substrate (S1, S2), a duplex presenting an 18-nucleotide bubble (S3, S4) or a duplex carrying a 3-nucleotide bubble (S5, S6, S7, S8) (figure 42).

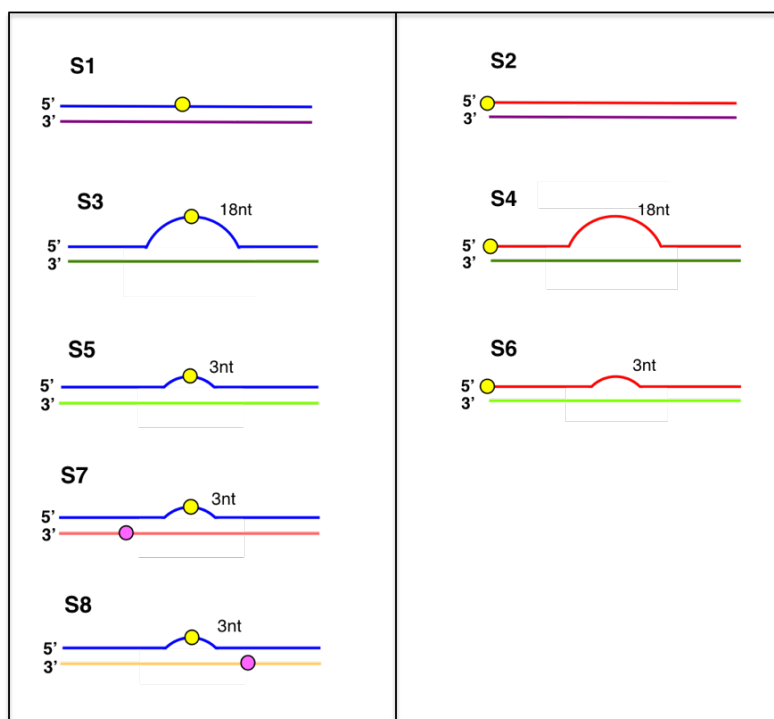


Figure 42: Substrates used in the study of binding of monomeric XPC, heterodimer XPC-HR23B and complex XPC-HR23B-TFIIH.

Substrates include two 44 bp DNA duplexes, one carrying a fluoro-dT modification inside its sequence (S1) and another one carrying a 5'-fluorescein amino modified (FAM)-C6 modification (S2). 44 bp duplexes carrying bubbles include S3 (18-nucleotide mismatch with a fluoro-dT at the bubble), S4 (18-nucleotide bubble with a 5'-FAM-C6 modification), S5 (3-nucleotide mismatch with a fluoro-dT at the bubble) and S6 (3-nucleotide bubble with a 5'-FAM-C6 modification). Finally, substrates S7 and S8 are 44 bp duplexes with a 3-nucleotide bubble carrying a fluoro-dT nucleotide at the mismatch and a Cy3 dye left (S7) or right (S8) to the bubble.

We tested monomeric XPC, heterodimeric XPC-HR23B and complex XPC-HR23B-TFIIH binding using three different methods: fluorescence anisotropy, band shifting, and bulk FRET assays. The low amount of XPC we had available for these experiments (purification yields were always lower than 0.1 mg protein for 1 L culture) meant that

assays could not be carried out in triplicates, and so these results can only be considered as preliminary.

3.2.2.1. Anisotropy assays

We studied the binding of monomer XPC to several 44 bp DNA duplex species: one carrying a fluorescein-dT nucleotide at the centre of the sequence (S1); one with a 5'-FAM-C6 modification on its 5' end (S2); one with an 18-nucleotide mismatch and a fluoro-dT nucleotide at the bubble (S3) and finally one with an 18-nucleotide bubble and a 5'-FAM-C6 modification at the 5' end of the substrate (S4). These were studied using a fluorescence anisotropy experiment (figure 43) as previously described in chapter 2, section 2.5.5.1. The titrations showed a progressive increase in the anisotropy values as we incremented the concentration of monomeric XPC. Anisotropy monitors variations in the rotational time of the biomolecule, which is a function of the mass of the complex. Therefore, an increase in anisotropy reflects a slower rotational time that is consistent with the formation of a higher mass complex induced by XPC binding. The initial and final anisotropy values vary between the different substrates investigated, although the relative change (Δr) is similar (approximately 0.2 units) across all of them. Different initial anisotropy values can arise from variations in the lifetime of the fluorescent probe due to the different locations (5'-end or internal) in the DNA and also due to the single or duplex configuration of the substrate (Nazarenko *et al.*, 2002). The fluorescein probe used in our studies has a fluorescence lifetime of approximately 4 ns in solution. A decrease in this lifetime as a result of quenching by adjacent bases will result in an increase in the initial anisotropy value. The relative changes in anisotropy as a function of the concentration of monomeric XPC were fitted to a quadratic equation reflecting a 1:1 binding model.

The K_D values obtained in these four experiments (table 4) were similar for S2, S3 and S4 (39 nM, 31 nM and 32 nM, respectively) but lower for S1 (5 nM), whose apparent binding affinity was 6-fold higher than the other substrates. This result was expected when comparing substrates S1 and S2, as the fluoro-dT would have introduced a point of weakness into our duplex substrate that would be interpreted by XPC as a mismatch, but it's unclear why the same difference is not observed for substrates S3 and S4. It's possible that the 18-nucleotide bubble introduced in these two substrates was just too

XPC: THE PROTEIN THAT STARTS IT ALL

big to be interpreted as a mismatch. Lastly, the low yields obtained in every XPC purification hindered the biochemical characterization of the protein by limiting the number of data points taken in every measurement. This probably affected the fitting of the data and the K_D values obtained, making the data indicative but not reliable.

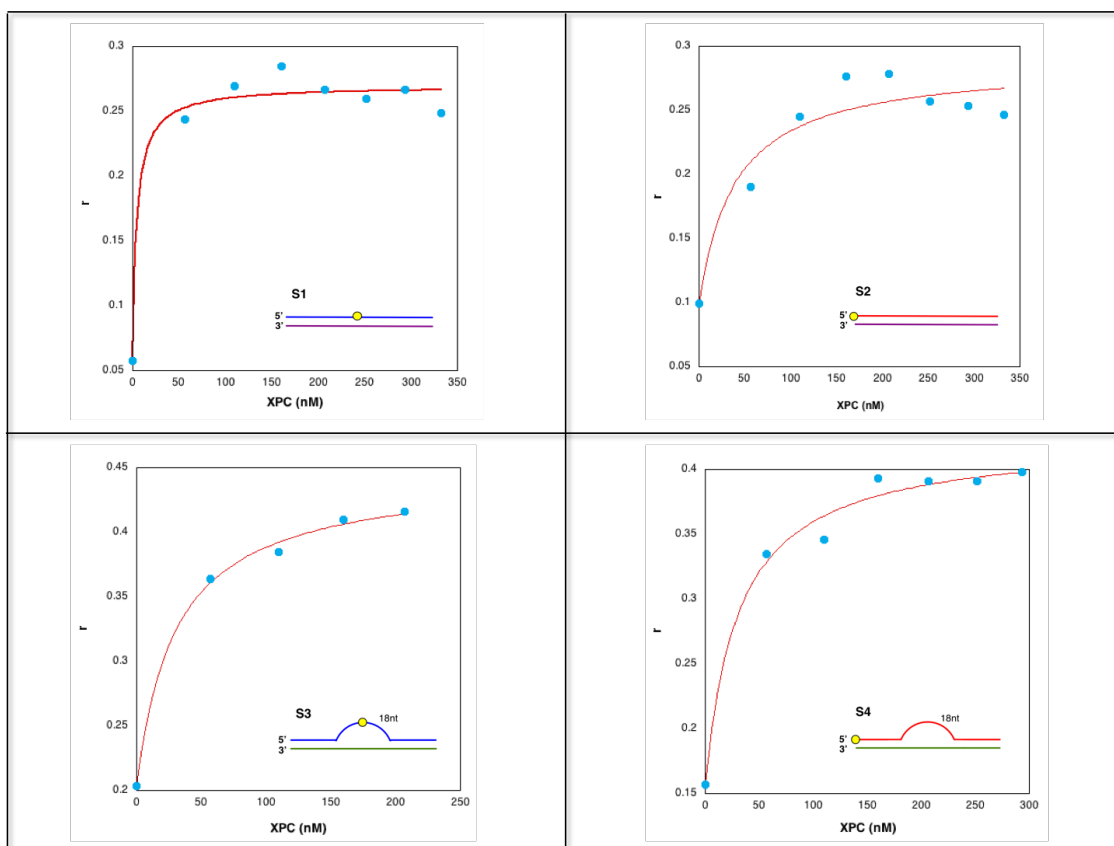


Figure 43: Fluorescence anisotropy assay showing binding of monomeric XPC to substrates S1, S2, S3 and S4.

S1 and S2 are 44-mer duplexes with a fluorescein molecule inside the duplex sequence, mimicking a lesion (S1), or at the 5' end of the substrate (S2) and S3 and S4 are 44-mer duplexes with an 18-nucleotide bubble carrying the fluorescein molecule at the bubble (S3) or at the 5' end of the substrate (S4). Final concentration for all substrates was 25 nM. K_D 's obtained for all four isotherms were in the nM range.

3.2.2.2. Electrophoretic mobility shift assay

Heterodimeric XPC-HR23B was tested for binding to substrates S7 and S8 first (figure 44) and then to substrates S2, S5 and S6 (figure 45) in two different band shifting assays

according to the method described in chapter 2, section 2.5.4. All substrates tested were 44 bp long: S2 is a duplex, while S5, S6, S7 and S8 carry a 3-nucleotide mismatch. A fluoro-dT nucleotide also served as a bulky adduct in the DNA, placed at the centre of the bubble in substrates S5, S7 and S8. Substrates S7 and S8 are essentially a modification of substrate S5, but in addition to the fluoro-dT nucleotide they carry a Cy3 dye at the left (S7) or right (S8) side of the 3-nucleotide bubble, making them suitable for their use in bulk FRET experiments.

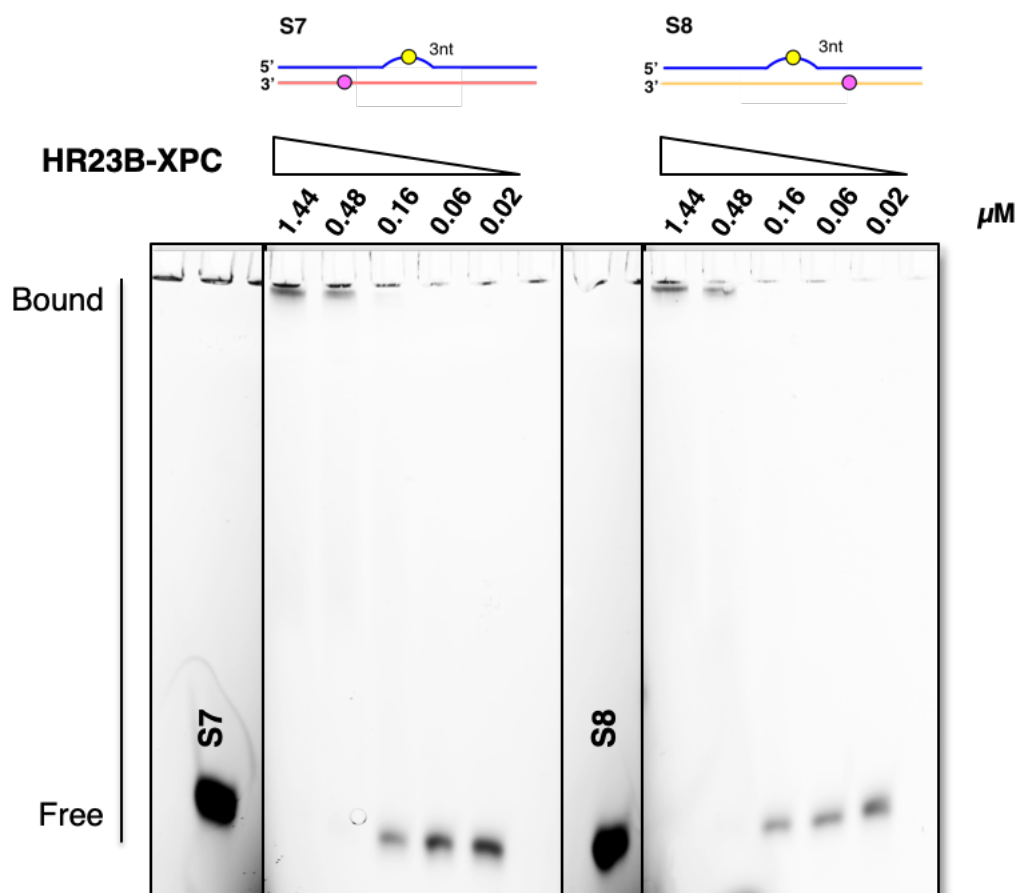


Figure 44: EMSA assay (10% polyacrylamide gel) showing binding of heterodimer XPC-HR23B to substrates S7 and S8.

S7 and S8 are 44-mer duplex substrates with a 3-nucleotide bubble carrying a fluoro-dT nucleotide mimicking a bulky adduct and a Cy3 dye left (S7) or right (S8) to the bubble, shown here in cartoon view. K_D for the binding of the heterodimer seems to be between 0.48 μM (all of the substrate bound) and 0.16 μM (all of the substrate free) for both substrates.

XPC: THE PROTEIN THAT STARTS IT ALL

The limitations of this EMSA assay meant that we could only estimate a range of K_D values for this binding: $0.16 - 0.48 \mu\text{M}$ for both substrates (table 4). In order to determine a more accurate value for this K_D we performed a second band shifting experiment with a finer grid of protein concentrations and substrates S2, S5 (equivalent to the already tested S7 and S8 substrates) and S6.

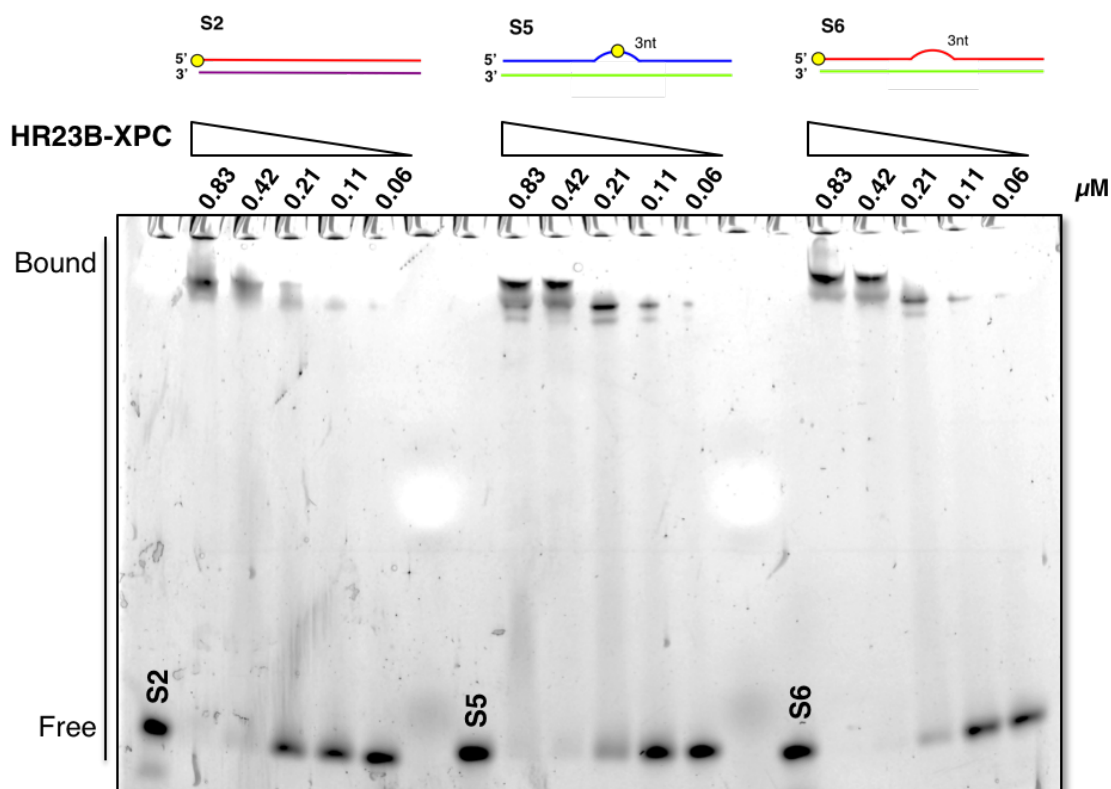


Figure 45: EMSA assay (6% polyacrylamide gel) showing binding of the XPC-HR23B complex to substrates S2, S5 and S6.

S2 is a 44-mer duplex substrate with a fluorescein molecule at the 5' end of the substrate and S5 and S6 are 44-mer duplexes with a 3-nucleotide bubble carrying a fluorescein molecule at the mismatch, mimicking a bulky adduct (S5), or at the 5' end of the substrate (S6). K_D for these bindings seems to be between $0.42 \mu\text{M}$ (substrate mostly bound) and $0.21 \mu\text{M}$ (substrate mostly free) for S2, and between $0.21 \mu\text{M}$ and $0.11 \mu\text{M}$ for S5 and S6. The position of the fluorescent dye didn't seem to affect XPC-HR23B binding to the substrate.

Apparent K_D values for this binding were narrowed down to $0.21 - 0.42 \mu\text{M}$ for S2, and $0.11 - 0.21 \mu\text{M}$ for S5 and S6 (table 4). While the similar results obtained for substrates S5 and S6 were to be expected, a considerably higher K_D for the duplex S2 had been

anticipated. Surprisingly, the gel shows two potentially different species binding to substrates S5 and S6 at a higher heterodimer concentration.

Our EMSA assays confirmed that XPC-HR23B can effectively bind a dsDNA substrate, but the low resolution of our gels and the much wider range of the K_D values obtained (values obtained in our anisotropy experiments were in the nM range) led us to consider a third approach to study XPC-HR23B binding.

3.2.2.3. Bulk FRET assays

We finally tested XPC-HR23B binding to substrates S7 and S8, both a 44 bp duplex carrying a 3-nucleotide bubble with a fluoro-dT nucleotide positioned at the mismatch and a Cy3 dye at the left (S7) or right (S8) side of the bubble, in a bulk FRET assay, following the method previously described in chapter 2, section 2.5.5.2. Fluorescence spectra were measured at an excitation wavelength of 490 nm (for the fluorescein dye) and 542 nm (for the Cy3 dye) for each protein concentration: the addition of XPC-HR23B induced a 2-fold increase in the fluorescein emitted fluorescence for both S7 and S8 substrates, but the change in the case of Cy3 was only noticeable in the S8 substrate (figures 47 (A) and 48 (A)). We used the acceptor-sensitized method (RatioA method, figure 46) to study the effect of XPC-HR23B binding on the structure of our DNA substrates and their relative affinities.

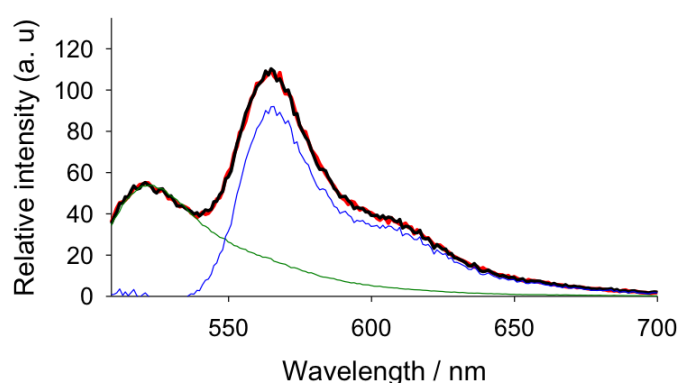


Figure 46: Calculation of the RatioA.

RatioA is the ratio between the Cy3 emitted fluorescence obtained via FRET (blue line) and the spectra of the sample excited at 542 nm (not represented). The Cy3 signal resulting from the excitation of the fluorescein dye is obtained by subtracting the spectra of our pure fluorescein

XPC: THE PROTEIN THAT STARTS IT ALL

fluorophore (green line) to the spectra of the sample excited at 490nm (red line).

The RatioA is the most widely used method to analyze data in FRET-based assays, as it removes potential artefacts resulting from variations in the intensity of the donor due to quenching or enhancement by the protein that are not related to changes in the donor-acceptor distance. Similarly, because it is based on calculating the ratio of acceptor emission obtained by FRET (excitation at 490 nm) with respect to the emission obtained by exciting the acceptor directly, any potential variation in Cy3 emission due to processes other than FRET is removed. We observed a progressive increase in the RatioA as a function of the concentration of the proteins added in all of our experiments, and the relative RatioA variation is similar across all conditions.

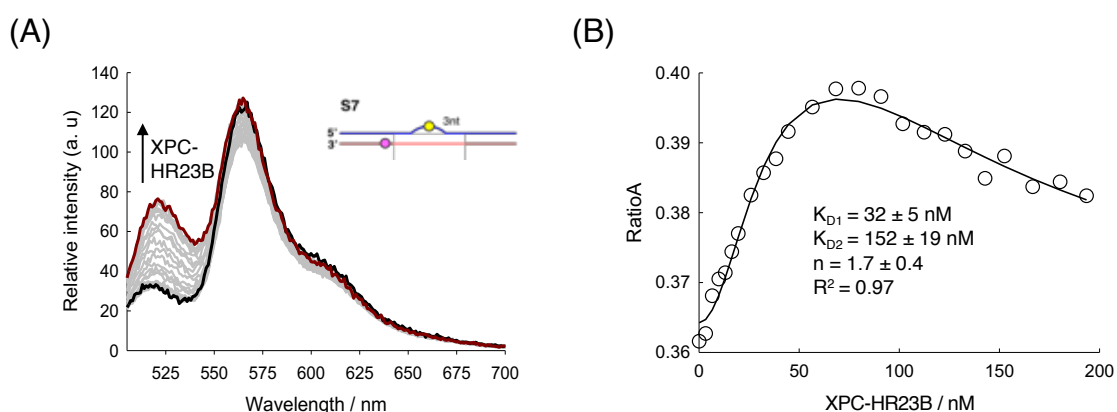


Figure 47: FRET analysis of binding of XPC-HR23B to substrate S7.

Addition of XPC-HR23B caused a 2-fold increase in the fluorescence emitted by the fluorescein dye, but barely affected Cy3 (initial intensity is indicated by a black line, and final intensity after addition of XPC-HR23B is indicated by a red line) (A). Representation of the RatioA as a function of the protein concentration showed two binding events with high cooperativity (B).

RatioA calculations for the binding of the heterodimer to substrate S7 showed a K_{D1} of 32 nM for the first binding event and a K_{D2} of 152 nM for the second event, with high cooperativity ($n \sim 2$), when fitting the data to a Hill model (figure 47 (B)). The FRET efficiencies are directly proportional to RatioA. These results (especially regarding the second binding step, due to the low change) are only preliminary, and experiments including a wider range of concentrations of XPC-HR23B are needed to confirm them.

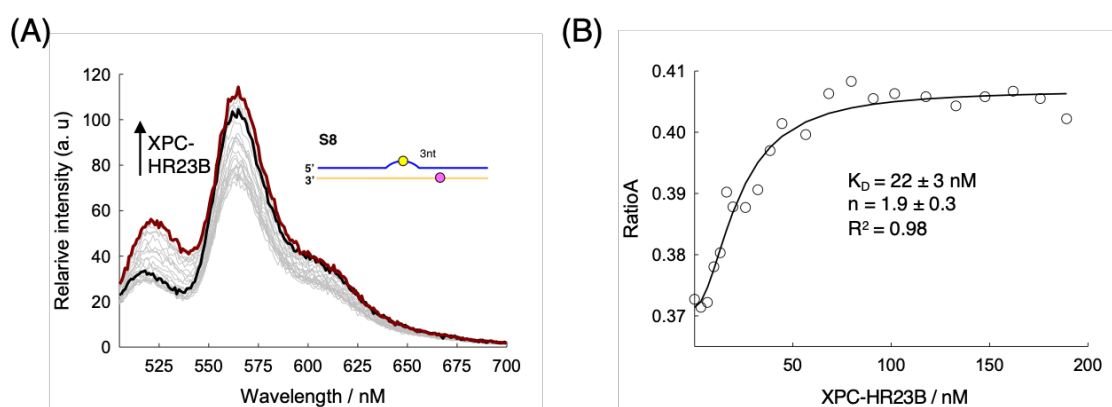


Figure 48: FRET analysis of binding of XPC-HR23B to substrate S8.

As seen for the S7 substrate, addition of XPC-HR23B caused a 2-fold increase in the fluorescence emitted by the fluorescein dye (initial intensity is indicated by a black line, and final intensity after addition of XPC-HR23B is indicated by a red line) (A). Representation of the RatioA as a function of the protein concentration showed a single binding event with high cooperativity (B).

Representation of the RatioA values vs XPC-HR23B concentration showed a single binding step with an affinity of 22 nM and high cooperativity ($n \sim 2$) when the data obtained for the binding of the heterodimer to substrate S8 were fitted to a Hill model (figure 48 (B)). However, analysis of the Cy3 emitted fluorescence as a function of XPC-HR23B concentration showed a biphasic binding isotherm (data not shown), similarly to our observations on XPC-HR23B binding to substrate S7, although the changes in the emitted fluorescence were much higher here than they were for the S7 substrate, suggesting that XPC-HR23B might be positioned closer to the Cy3 in S8 than in S7.

In a second experiment we tried to see the effect of the TFIIH Core sub-complex on XPC-HR23B binding to these substrates (figures 49 and 50). In these assays we titrated the TFIIH Core by adding it to a quartz cuvette containing our binding buffer, DNA substrate and 100 nM of the XPC-HR23B heterodimer (see chapter 2, section 2.5.5.2), as XPC binds first to the damaged site at the DNA. Addition of TFIIH Core to our DNA-bound XPC-HR23B caused an approximately 2-fold decrease of the fluorescence emitted by both the fluorescein and Cy3 dyes for both the S7 and S8 substrates (figures 49 (A) and 50 (A)).

The analysis of the RatioA values obtained for the binding of TFIIH Core to the S7-XPC-

XPC: THE PROTEIN THAT STARTS IT ALL

HR23B DNA-protein complex showed a K_D of 12 nM, with a high cooperativity ($n \sim 3$) (figure 49 (B)), possibly due to the experimental error in the first data points.

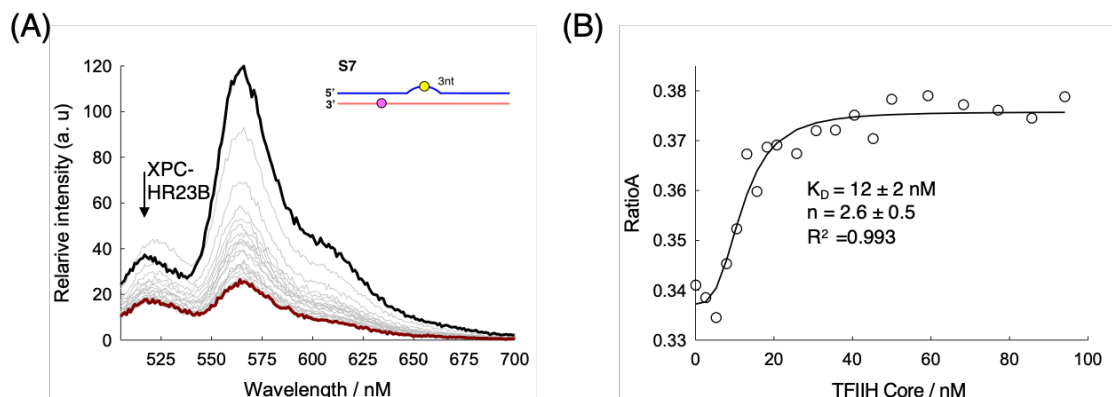


Figure 49: FRET analysis of binding of TFIID Core to the S7-XPC-HR23B complex.

Addition of TFIID Core causes a 2-fold decrease in the fluorescence emitted by both the fluorescein and Cy3 dyes (initial intensity is indicated by a black line, and final intensity after addition of XPC-HR23B is indicated by a red line) (A). A K_D of 12 nM, with a high cooperativity ($n = 2.6$), was obtained in the analysis of the RatioA (B).

The results obtained for the binding of TFIID Core to the S8-XPC-HR23B complex were very similar to the ones obtained with the S7 substrate. Fitting of the data obtained in the analysis of the RatioA values offered a K_D of 9 nM (figure 50 (B)), again with a high cooperativity ($n \sim 3$).

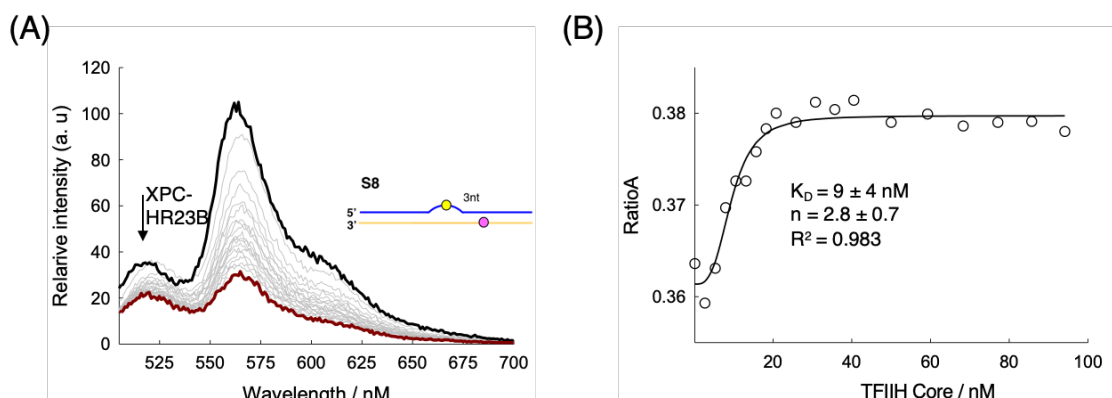


Figure 50: FRET analysis of binding of TFIID Core to the S8-XPC-HR23B complex.

Addition of TFIID Core causes a 2-fold decrease in the fluorescence emitted by both the

fluorescein and the Cy3 dyes, as previously seen with the S7 substrate (initial intensity is indicated by a black line, and final intensity after addition of XPC-HR23B is indicated by a red line) (A). K_D values obtained in the analysis of the RatioA (B) was 9 nM, with a high cooperativity ($n \sim 3$).

The K_D values obtained for the binding of TFIIH Core to the S7-XPC-HR23B and S8-XPC-HR23B complexes were very similar to the ones obtained for the binding of XPC-HR23B alone, indicating that the addition of the TFIIH Core sub-complex didn't have an effect on the binding of XPC. Further experiments will be required to confirm these preliminary results.

3.3. Discussion

3.3.1. Heterologous expression of XPC constructs

Damage-detector XPC is a heterotrimeric complex *in vivo*, composed of proteins XPC, HR23A/HR23B and CETN2 (Sugasawa *et al.*, 1996) (Nishi *et al.*, 2005), although the latter is not required for *in vitro* experiments (Okuda *et al.*, 2015), and monomeric XPC has been proved enough to carry out the repair reaction in an *in vitro* system (Sugasawa *et al.*, 1996) (Reardon *et al.*, 1996). Monomeric XPC performs a quick scan of the DNA in search of mispaired bases (Maillard *et al.*, 2007) (Hoogstraten *et al.*, 2008), and upon locating a mismatch, it introduces a β -hairpin between the two strands, causing a further distortion in the double helix and stabilising XPC at the damaged site (Min & Pavletich, 2007) (Camenisch *et al.*, 2009). It's this DNA-bound XPC structure that will act as a recruitment platform for the subsequent DNA repair factors involved in NER (Min & Pavletich, 2007) (Shell *et al.*, 2013).

3.3.1.1. Monomeric XPC

As previous studies had reported that, despite being a heterotrimeric complex *in vivo*, the activity of XPC alone was sufficient to bind to the DNA and carry out successfully a repair reaction (Sugasawa *et al.*, 1996) (Reardon *et al.*, 1996), we initially aimed to clone the XPC gene and express the monomeric protein on its own.

Expression tests performed in 25 ml infections offered robust expression levels (figure 33), so we proceeded to scale-up our infections. We approached the purification of

monomeric XPC by following a three-step protocol consisting of IMAC, size exclusion and cation exchange chromatography. However, as the purification progressed, we encountered a problem: the sample presented persistent contamination even after three chromatographic steps.

We attempted a further purification of the sample through heparin chromatography (figure 34), as the charge and structure of the polymer makes it an apt replacement for DNA, and so this method is ideal for the purification of DNA binding proteins. Analysis of the fractions eluted from the HiTrap™ Heparin HP column (figure 34, green box) showed a cleaner sample, but two bands of about 110 kDa and 50 kDa stubbornly remained in our purification. MS analysis showed that the higher molecular weight band was a degraded form of XPC. To our surprise, the 50 kDa band was identified as the *Spodoptera frugiperda* homologue of HR23B, the human partner of XPC.

The fact that human XPC was able to pull down the untagged HR23B protein from the Sf9 cells we were using as an expression system pointed to the importance of their partnership, in agreement with more recent publications that suggested a stabilising role of HR23B upon its binding to XPC, helping the damage-sensor to reach the site of lesion properly folded (Bergink *et al.*, 2012). In view of this result, and although our monomeric XPC was able to bind different DNA substrates in anisotropy experiments (figure 43), we decided to clone and express the heterodimeric XPC-HR23B complex for future experiments. The heterodimer would also allow us to compare the effect of HR23B on XPC's activity.

3.3.1.2. TCP-tagged XPC

At the time we were cloning and expressing monomeric XPC, we were also working on a second monomeric construct, with XPC carrying a 6-residue Cys-Cys-Pro-Gly-Cys-Cys tag this time. A BioSprint purification test carried out from a 25 ml suspension showed that expression levels of TCP-tagged XPC were significantly lower compared to the expression obtained with monomeric XPC when both 25 ml cultures were infected with a P0 viral stock (figure 35). This was probably due to a weaker P0 virus being responsible for a less efficient infection, but an interference of the TCP tag with the

expression of XPC or with the pull-down of the protein by the magnetic Ni-NTA beads used in the BioSprint station (as we also observed with the TFIIH Core sub-complex) has also to be considered.

The lower expression levels obtained for this construct led us to simplify our purification protocol in order to minimize protein loss: based on previous results obtained in the purification of monomeric XPC, we decided that the best strategy for this sample was to remove the size exclusion chromatography in favour of a heparin column, effectively removing a chromatographic step from our protocol (figure 36). As we had already seen in the purification of monomeric XPC, a 110 kDa band corresponding to a degraded form of XPC and a faint 50 kDa band corresponding to the *Sf9* homolog of HR23B can still be found accompanying our TCP-tagged XPC construct (figure 36, yellow box). Although it didn't yield a completely pure protein, this simplified purification process represented an improvement over our previous results, as it was much more efficient than our four-step protocol in removing unspecific contaminants from our sample, regardless of the extra chromatographic step.

Our TCP-tagged XPC protein can be successfully labelled with a ReAsH-EDT₂ biarsenic ligand, as shown in figure 39. A control reaction with no added protein established the intensity corresponding to the free ReAsH-EDT₂ reagent when excited at 593 nm, observing afterwards a consistent increase in the intensity of its emitted fluorescence (608 nm) as the concentration of TCP-XPC present in the reactions grows.

Time restrictions have stopped us from performing any experiments with this TCP-tagged protein yet, but future work includes confirmation of our preliminary data regarding DNA binding of XPC to damaged and undamaged substrates in fluorescence anisotropy and FRET assays. Cloning of the HR23B partner of XPC will also be considered as a means to stabilise XPC and potentially improve the protein yield once a successful, standardized protocol for the expression and purification of the heterodimer not carrying the TCP tag is established.

3.3.1.3. *Heterodimer XPC-HR23B*

Expression levels were good for both partners when extracted from a small-scale infection (figure 40), but they dropped dramatically when the heterodimer was extracted from larger infections.

The scaling-up of infections is a delicate, frequently tricky process in which many factors have to be taken into account, with the viability and growth rate of the suspension culture and titre of the viral stock probably being two of the most important ones. Viability and growth rate of a suspension of *Sf9* cells can be easily measured, using that information to decide the suitability of a culture to be infected. In our experience the titre of the virus is more difficult to determine, as plaque assays are very time-consuming and can be difficult to interpret, and we have often observed encouraging results with a P0 viral stock, only to obtain a poor or very poor yield from an infection performed with the next generation of that same virus (P1).

The expression levels observed for the scaled-up infections were quite low, seriously limiting the amount of pure heterodimer we obtained. This forced us to explore a series of combinations of chromatographic steps based on the quality of our samples (IMAC – heparin or IMAC – size exclusion), finally determining that a two-step Talon® Superflow™ resin followed by purification through a HiPrep™ 16/60 Superdex™ S200 prep grade column was the best option for the majority of our infections with the heterodimer. The chromatograph obtained during the purification through the S200 column (figure 41) showed two peaks, but only the first one contained the heterodimer XPC-HR23B. Analysis of the fractions eluted in both peaks in a SDS-PAGE gel showed that the second peak contained very small amounts of XPC, but no HR23B, pointing to a lower expression of the stabilising protein.

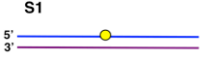

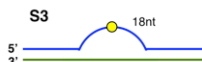
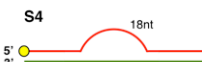
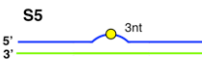
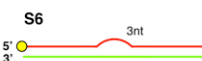
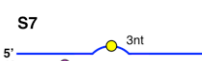
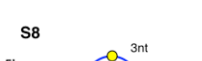
Future work with this construct involves the expression of the heterodimer in High Five™ insect cells (Thermo Fisher Scientific), as they reportedly express 5 to 10 times more heterologous protein than our current expression system. This will require the prior establishment of a standardized work routine and new infection protocol adapted to these High Five™ cells. Optimization of the purification protocol for this construct is also a work in progress.

3.3.2. XPC binds to a range of double-stranded DNA substrates with similar affinity

Early studies characterising the biochemical properties of the damage-detection protein reported that monomeric XPC alone was able to bind a duplex DNA substrate with similar affinity as the heterodimer XPC-HR23B, and was able to even correct a repair defect, albeit less efficiently than the heterodimer (Reardon *et al.*, 1996) (Sugasawa *et al.*, 1996).

Table 4: Dissociation constants for binding of monomeric XPC, heterodimer XPC-HR23B and complex XPC-HR23B-TFIIH to a range of substrates in a fluorescence anisotropy, EMSA or bulk FRET assay.

The limitations of the EMSA experiment meant that K_D could only be determined as an apparent K_D within a range of values. Data from anisotropy experiments were fitted to a binding model that assumes 1:1 DNA:protein (see chapter 2, section 2.5.5.1). Data from bulk FRET experiments were fitted to a Hill model (see chapter 2, section 2.5.5.3).

		K_D (nM)			
		Anisotropy	EMSA		FRET
		XPC	XPC-HR23B	XPC-HR23B	XPC-HR23B-TFIIH
	S1	5 ± 6			
	S2	39 ± 26	210 - 420		
	S3	31 ± 7			
	S4	32 ± 9			
	S5		110 - 210		
	S6		110 - 210		
	S7		160 - 480	(1) 32 ± 5 (2) 152 ± 19	12 ± 2
	S8		160 - 480	22 ± 3	9 ± 4

More recently, an exhaustive study of XPC-HR23B binding to a wide range of DNA substrates was carried out by the Chazin group (Shell *et al.*, 2013). Using a high-throughput fluorescence anisotropy assay, they tested XPC-HR23B binding to DNA substrates of different lengths, with and without mismatches, and presenting a variety of lesions repaired by the NER and BER systems. Their findings showed that length of the substrate did not affect XPC-HR23B binding, as the K_D values obtained were similar for all tested lengths. The presence of a 2-nucleotide or a 6-nucleotide bubble increased binding affinity 9-fold compared to the same-length duplex substrate, but it only increased 2-4-fold when binding to substrates carrying a lesion. Interestingly, XPC-HR23B bound both BER and NER substrates with the same affinity, supporting the current model that considers XPC a sensor of distortion in the double-helix rather than a specific damage detection factor (Hoogstraten *et al.*, 2008) (Min & Pavletich, 2007) (Shell *et al.*, 2013).

Unfortunately, the low and frequently unreliable expression levels of XPC in our infections severely limited our experiments. We opted for testing binding using fluorescence anisotropy, band shifting, and bulk FRET assays to try to determine the best method to observe binding for this protein. Apparent K_D obtained for each construct and each experiment are summarised in table 4.

3.3.2.1. Anisotropy experiments

We obtained similar K_D values when testing binding of monomeric XPC to duplex substrate S2 (39 nM), and to duplex substrates carrying an 18-nucleotide bubble S3 and S4 (31 nM and 32 nM, respectively) (table 4). In the case of duplex substrate S1 the binding affinity was 6-fold higher than that observed for the other substrates. As XPC shows a reportedly higher affinity for substrates carrying a mismatch (Shell *et al.*, 2013), we expected to see this difference in binding affinity when comparing substrates S1 and S2, as the fluoro-dT nucleotide in the middle of the sequence of the S1 substrate introduces a mismatch that would be interpreted by XPC as a lesion, but unexpectedly we don't see the same difference for substrates S3 and S4, both carrying an 18-nucleotide mismatch (S3 additionally carries a fluoro-dT nucleotide at the centre of the bubble, similarly to substrate S1).

It's possible that the length of the bubble introduced in substrates S3 and S4 might be affecting our experiment: Shell *et al.* 2013 reported that in a 60 bp substrate with a 6-nucleotide mismatch the binding equilibrium DNA:XPC changed from 1:1 to 1:2, but it stayed 1:1 when the substrate contained a 2-nucleotide bubble. As our bubble is 3-times bigger, it's entirely possible that the binding we see is not following the same model as binding to substrates S1 and S2.

It can neither be ignored that at this point in our project we hadn't obtained the XPC partner HR23B yet – figure 34 shows that a small amount of the S79 HR23B homologue is pulled down together with XPC in our sample, but it's unclear if it is functional, and in any case the purified amount is small enough that only a minimum of our sample would be the heterodimer XPC-HR23B instead of monomeric XPC, and so the effect of HR23B in binding (if any) could probably not be appreciated.

Our current efforts are focused on the expression and biochemical characterization of heterodimer XPC-HR23B, so future experiments featuring monomeric XPC are unlikely.

3.3.2.2. *Electrophoretic mobility shift assay*

The results obtained in our purification shown in figure 41, coupled with the results for the anisotropy assays summarized in table 4 led us to adopt a different approach in our experiments attempting to characterize XPC binding to DNA: we next studied the binding of heterodimer XPC-HR23B instead of monomeric XPC to mismatched substrates now carrying a 3-nucleotide bubble instead of an 18-nucleotide one.

Apparent K_D values for binding to substrates S7 and S8 (44 bp long, carrying a 3-nucleotide bubble and a fluoro-dT nucleotide at the centre of the bubble as a lesion) were in the range of 0.16 – 0.48 μM for both substrates (figure 44). We tried to determine a more accurate K_D value in our next experiment, performed with a finer grid of protein concentrations and substrates S2, S5 (equivalent to the already tested S7 and S8 substrates) and S6. Apparent K_D values were narrowed down to 0.21 – 0.42 μM for S2 (44 bp duplex), and 0.11 – 0.21 μM for S5 and S6 (44 bp long, carrying a 3-nucleotide bubble – S5 additionally carries a fluoro-dT nucleotide at the centre of the bubble

mimicking a bulky adduct). Figure 45 shows two potentially different species binding to substrates S5 and S6. This could point to two XPC molecules binding to the mismatched substrate at higher concentrations of the heterodimer, as reported by Shell *et al.* 2013 and Hilton and co-workers (Hilton *et al.*, 2016).

The fact that substrates S7 and S8, and S5 and S6 have a similar K_D (table 4) is consistent with XPC detecting a distortion in the double helix rather than an actual lesion in the DNA (Maillard *et al.*, 2007) (Shell *et al.*, 2013). However, we expected a considerably higher K_D for the duplex S2 (a 9-fold difference was reported by Shell *et al.* 2013 for the binding of XPC-HR23B to a 42 bp duplex substrate, both unmodified and with a 6-nucleotide mismatch).

While our EMSA assays demonstrated that heterodimer XPC-HR23B can effectively bind a double-stranded DNA substrate (both unmodified and presenting a mismatch), the low resolution of our gels and the K_D values obtained (in the μM range, as opposed to K_D values reported by Shell *et al.* in their 2013 paper and our own results obtained in our anisotropy experiments, both in the nM range) made us reconsider the suitability of EMSAs to study the binding of XPC-HR23B to DNA. With this results in mind, we next tried to observe binding in bulk FRET assays.

3.3.2.3. Bulk FRET assays

Bulk FRET analysis of binding of the heterodimer XPC-HR23B to substrates S7 and S8 (44 bp duplex carrying a 3-nucleotide bubble with a fluoro-dT nucleotide positioned at the mismatch and a Cy3 dye at the left (S7) or right (S8) side of the bubble) helped us to narrow down the K_D values for a substrate with a mismatch from the μM range seen in our EMSA assays to a nM range (table 4). Surprisingly, these values are similar to the K_D obtained for the binding of monomer XPC to S3 (31 nM), the S7 and S8 equivalent with an 18-nucleotide bubble. This seems to indicate that the size of the mismatch is not relevant to the binding of XPC, as opposed to the report of Shell *et al.* 2013, and it also points to HR23B not having an effect on XPC binding to DNA, confirming the findings of Reardon *et al.* 1996 and Sugasawa *et al.* 1996.

Fitting of our FRET data to a Hill model showed two binding events with high

cooperativity (the binding of the first XPC molecule helps incorporation of a second XPC to the bubble) for both the S7 and S8 substrates, as previously reported in the literature ((Shell *et al.*, 2013), (Hilton *et al.*, 2016)) and our own observations for the S5 and S6 substrates (also carrying a 3-nucleotide bubble) in previous EMSA assays. The binding of a second XPC to such a small bubble might be possible due to a further disruption of the double helix, caused by the binding of the first XPC. Changes in the Cy3 emitted fluorescence are much more pronounced in the S8 substrate than the S7 substrate, suggesting that XPC might be positioned closer to Cy3 in S8 than in S7. This change might be related to the relative orientation of XPC in our bubble, as recently observed by the Penedo group in experiments using FRET to look at distances between a Cy3-Cy5 donor-acceptor pair positioned at the stems of the substrate. Our RatioA data have not been transformed into FRET efficiencies; however, both are directly proportional, so that an increase in RatioA would represent an increase in FRET efficiency, and therefore, a decrease in inter-dye distance. Addition of the TFIIH Core sub-complex to the S7-XPC-HR23B or S8-XPC-HR23B complexes did not alter the binding of XPC, although further experiments will be required to confirm these initial results. The K_D values obtained for these binding experiments were very similar to the ones obtained for binding of XPC-HR23B alone, and similar again to the value obtained for binding of monomeric XPC to substrate S3 (table 4).

Our data suggest that XPC-HR23B and XPC-HR23B-TFIIH distort the structure of the DNA substrate so that the fluorescein and Cy3 dyes are positioned at a closer distance in the protein-substrate complex compared to the DNA alone. We have not characterized in detail the exact structural change induced by XPC or XPC-TFIIH binding, but given the positioning of the FRET pair, we hypothesise that this could arise from a combination of bubble melting and bending. XPC-induced bending of the DNA substrate has been previously reported by Janićijević and co-workers (Janićijević *et al.*, 2003).

3.4. Conclusions

We have analyzed the binding affinity of monomeric XPC, heterodimer XPC-HR23B and complex XPC-HR23B-TFIIH Core to a variety of DNA duplexes, both unmodified and carrying an 18-nucleotide or a 3-nucleotide mismatch, by three different methods:

XPC: THE PROTEIN THAT STARTS IT ALL

fluorescence anisotropy, EMSA, and bulk FRET.

Anisotropy experiments performed with monomeric XPC showed a higher affinity for substrate S1 (duplex with a simulated bulky adduct at the center of the sequence), and we expected the same high affinity for substrates S3 and S4 (both carrying an 18-nucleotide bubble, with S3 also carrying the simulated bulky adduct at the center of the mismatch), but the K_D values obtained were actually similar to the binding affinity observed for substrate S2, a DNA duplex with no internal modifications.

EMSA experiments performed with XPC-HR23B showed a higher affinity of the heterodimer for substrates S5 and S6 (both carrying a 3-nucleotide bubble, with S5 also carrying the simulated bulky adduct at the center of the mismatch) compared to substrate S2, but the presence of a lesion in S5 did not increase XPC's binding affinity. Our EMSA gels also suggest that two XPC molecules might be binding the mismatched substrates (but not a duplex) at higher concentrations of the heterodimer.

Bulk FRET experiments performed with XPC-HR23B helped narrow down the K_D values obtained in our EMSA assays from the μM to the nM range. Surprisingly, the results obtained were similar to the values observed for the binding of monomeric XPC to substrates S3 and S4. They also suggest that two molecules of XPC bind to both our S7 and S8 substrates with high cooperativity at higher concentrations of the heterodimer, and that binding of the protein altered the structure of our bubble. The addition of the TFIIH Core sub-complex didn't seem to have an effect on the binding of XPC once the protein is positioned at the repair bubble.

Taken together, our results seem to indicate that XPC preferentially binds to substrates carrying a mismatch, although the size of that mismatch doesn't seem to be relevant to binding. The presence of a simulated lesion at the mismatch, the addition of partner HR23B to our sample or the addition of the TFIIH Core sub-complex didn't affect XPC's binding affinity either. However, our experiments were severely limited by the amount of available protein, and EMSA experiments in particular offered a low resolution, hence our decision to continue our study performing bulk FRET experiments. For these reasons, these results can only be considered as preliminary, and further experiments are needed to confirm them.

Future work includes:

- Optimization of the XPC-HR23B yield (using High Five® cells as an expression system) and its purification protocol.
- Anisotropy experiments with unmodified and mismatched duplex substrates and a substrate carrying a cisplatin lesion, using our TCP-tagged XPC construct, potentially having included partner HR23B.
- Analysis of TFIIH recruitment by XPC by means of single molecule (Total internal reflection fluorescence (TIRF) and Fluorescence correlation spectroscopy (FCS) assays).
- Analysis of the nature of the alterations caused in the structure of our substrates by binding of XPC in further FRET experiments.

4. Cloning, expression and purification of the TFIIH complex

4.1. Introduction

TFIIH is a 10-subunit complex of about 500 kDa that plays important roles in two very different, yet equally essential processes: transcription and DNA repair (Compe & Egly, 2012). TFIIH was first described in 1992, although only its role in transcription was reported then (Flores *et al.*, 1992), with its role as a DNA repair factor in the NER pathway being described the following year (Feaver *et al.*, 1993).

The first published structure of TFIIH, reported by the Egly group in 2000, revealed two distinctive regions within the complex: a ring surrounding a central cavity that interacts with DNA, corresponding to the Core sub-complex (subunits XPB, p62, p52, p44 and p34), and a structure protruding from the ring, corresponding to the CAK sub-complex (subunits cdk7, MAT1 and cyclin H), with subunit XPD bringing the two of them together (Schultz *et al.*, 2000). The structure reported by Schultz *et al.* only included nine of the ten TFIIH subunits, as subunit p8 was lost in the first obtained samples due to its small size, and it was not until 2004 that it was finally described when a TTD phenotype was observed that did not correspond to any mutations in the XPB or XPD subunits (Gigliamari *et al.*, 2004). Last year a more detailed cryoEM structure of the TFIIH Core sub-complex plus CAK subunit MAT1 was published, finally allowing the positioning of the different subunits unambiguously, and offering more information about the interactions between them. In the reported structure, helicases XPB and XPD occupy the ends of a horseshoe-like structure, with subunits p44, p62, p34, p52 and p8 forming an arch that connects them (Greber *et al.*, 2017) (figure 51).

Purification of eukaryotic multi-protein complexes has remained a challenge until recently due to the limitations of the available cloning and expression tools. The first reported purification of TFIIH included three different fractionation procedures, four columns and four dialysis steps (Flores *et al.*, 1992), making it a very difficult and

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

laborious process that unfortunately remained that way for many years. The development of baculovirus expression systems (BEVs) in the last decades opened new and exciting possibilities that have offered ground-breaking results (Bieniossek *et al.*, 2012). One of the most recent reports regarding the extraction of TFIIH involved the cloning and expression in insect cells of each protein as pair-units, making it a better alternative than the traditional methods, but still a long and arduous process (Sugasawa *et al.*, 2009).

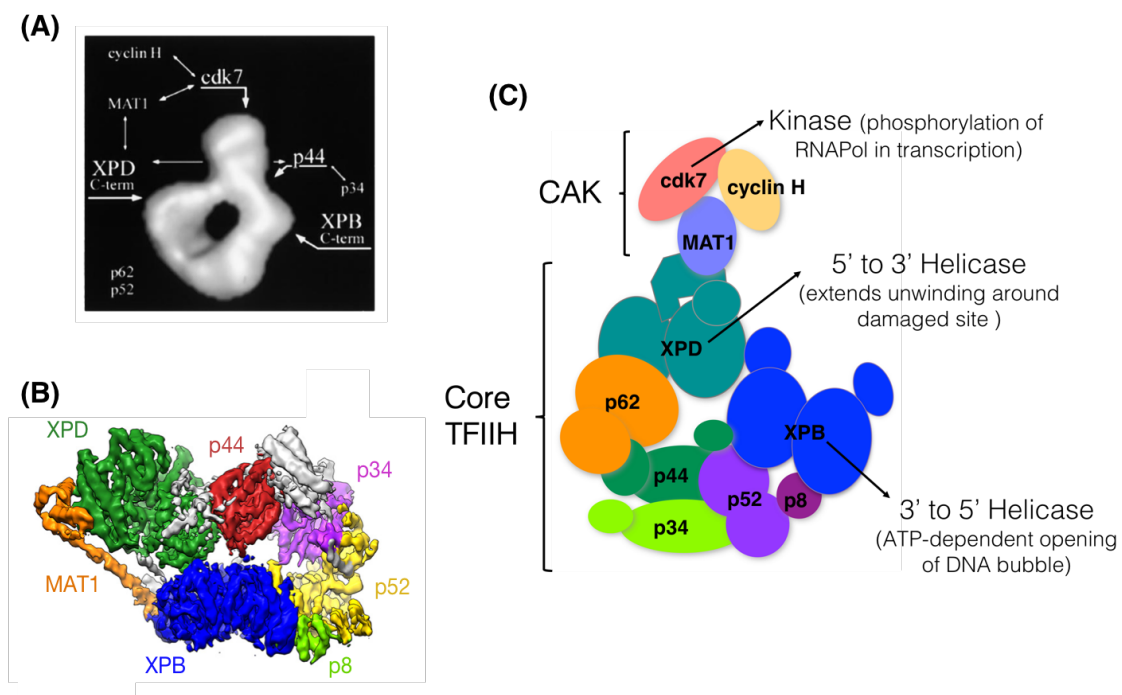


Figure 51: Structure of the 10-subunit complex TFIIH.

The first structure of the human TFIIH complex was reported in 2000, showing two different regions within the complex (A) (adapted from (Schultz *et al.*, 2000)), but it was not until 2017 that a more detailed cryoEM structure was obtained and the position of the different subunits could be assigned unambiguously (B) (adapted from (Greber *et al.*, 2017)). TFIIH presents two different areas, corresponding to its two different sub-complexes: the Core sub-complex shows a ring-like structure surrounding a pore that interacts with DNA, while the CAK sub-complex protrudes from the ring. Subunit XPD acts as a bridge between the two structures (C).

A particularly interesting BEV is the MultiBac™ system, a cloning and expression tool specifically designed for the production of complexes containing several subunits (Berger *et al.*, 2004). All five MultiBac™ plasmids contain a multiplication module that

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

allows the introduction of multiple genes into a single plasmid without requiring unique restriction sites for every new cloning reaction (Bieniossek *et al.*, 2013). The multi-gene construct will be later transposed into a baculoviral genome engineered for improved protein expression (Bieniossek *et al.*, 2012) (Berger *et al.*, 2004), the bacmid will be purified and transfected into a monolayer of insect cells and the recombinant viral stock obtained will be finally used to infect a suspension culture for overexpression of the desired complex (Bieniossek *et al.*, 2012) (figure 52).

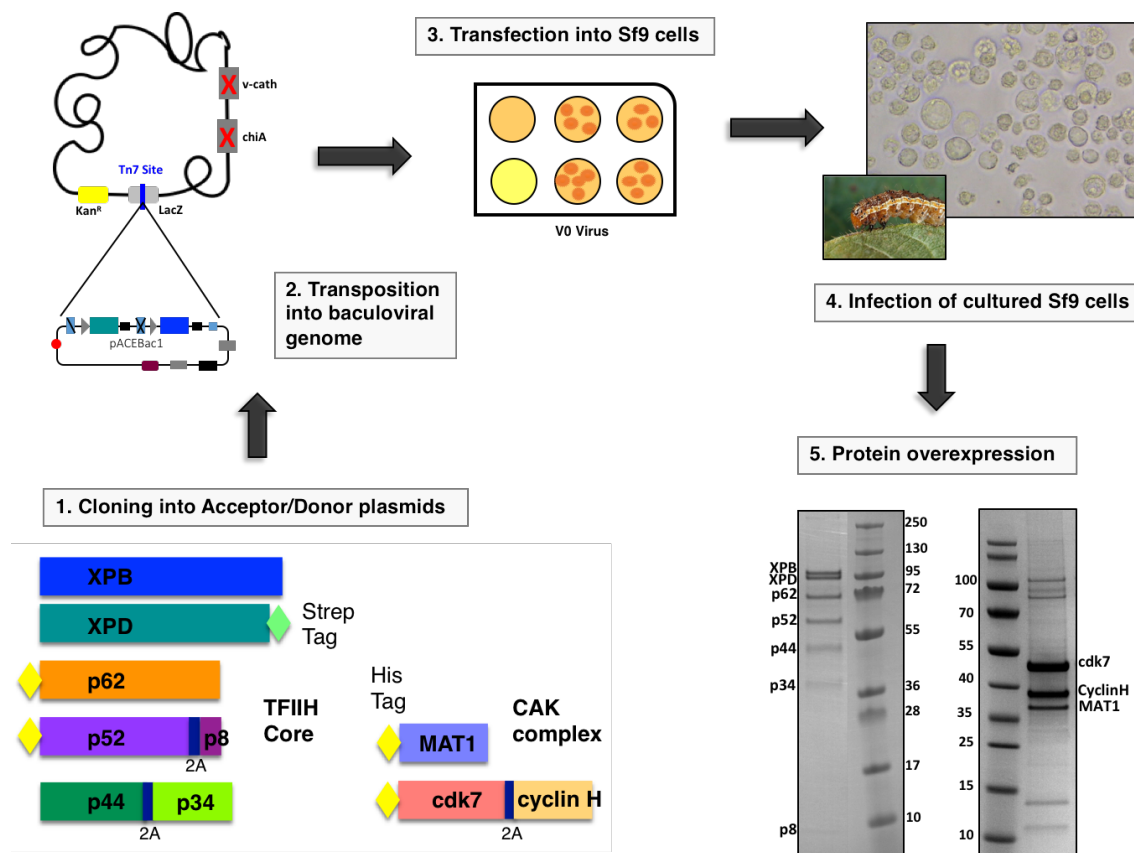


Figure 52: Overview of the cloning and expression of the TFIIH complex.

Summary of all the steps required for the cloning and expression of the TFIIH complex using the MultiBac™ system.

The combined use of the MultiBac™ expression system, together with simple, multi-function purification tags designed by our group and 2A-like auto-cleavable peptides that permit the fusion of two genes into a single unit (Luke *et al.*, 2009) have allowed us to considerably simplify both the cloning and expression processes. Still, designing and

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

optimising a protocol for the cloning, expression and purification of a complex as large as TFIIH has been a challenging process that involved many different constructs, summarised in figure 53 for future reference.

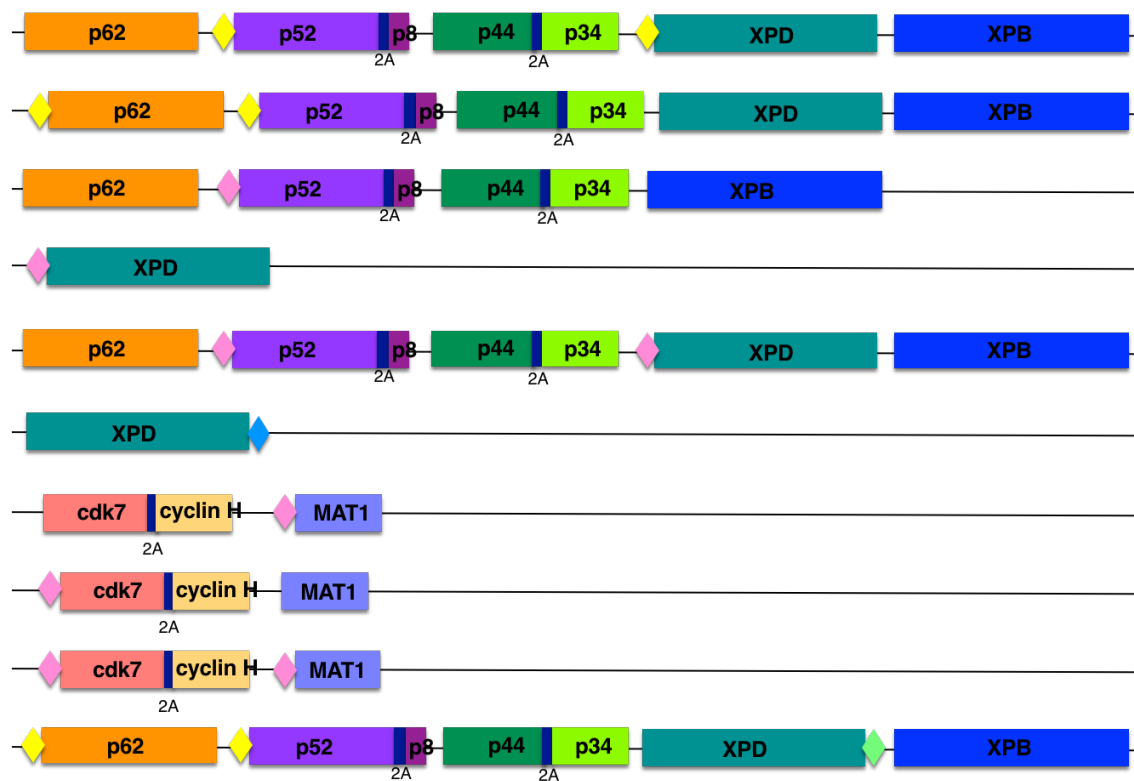


Figure 53: TFIIH constructs cloned and expressed in this PhD project.

Cartoon view of the different TFIIH Core and CAK constructs cloned for this PhD project. 6-residue His tags are indicated with a yellow diamond; 8-residue His tags are represented with a pink diamond; 10-residue His tags are depicted with a blue diamond, and finally a Strep tag is represented with a green diamond.

4.2. Results

4.2.1. TFIIH Core sub-complex

4.2.1.1. TFIIH Core cloning and expression: first attempts

The 10-subunit TFIIH complex is composed of sub-complexes TFIIH Core and CAK, and while the whole complex participates in transcription (Hanawalt & Spivak, 2008), only the Core sub-complex is required for the NER reaction (Compe & Egly, 2012). As our main

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

interest regarding the TFIIH complex focuses on its role as a DNA repair factor, we first aimed to clone and express the TFIIH Core sub-complex on its own.

The first consideration we took into account was the number and position of the purification tags. Observing the position of each TFIIH Core subunit within the complex, we decided that the best approach would be to place a 6-residue His tag at the N-terminal end of subunit p52, due to its strategic position at the base of the ring, and another 6-residue His tag at the N-terminal end of the XPD helicase which, as the bridge between the Core and CAK sub-complexes (Schultz *et al.*, 2000) (Greber *et al.*, 2017), is the most flexible subunit within TFIIH. All seven sequences encoding the different TFIIH Core subunits were first cloned into individual pACEBac2 plasmids, then sequentially cloned into multi-gene constructs using the *I-CeuI/BstXI* multiplication module as described in chapter 2, section 2.1.4, following the steps described in figure 54.

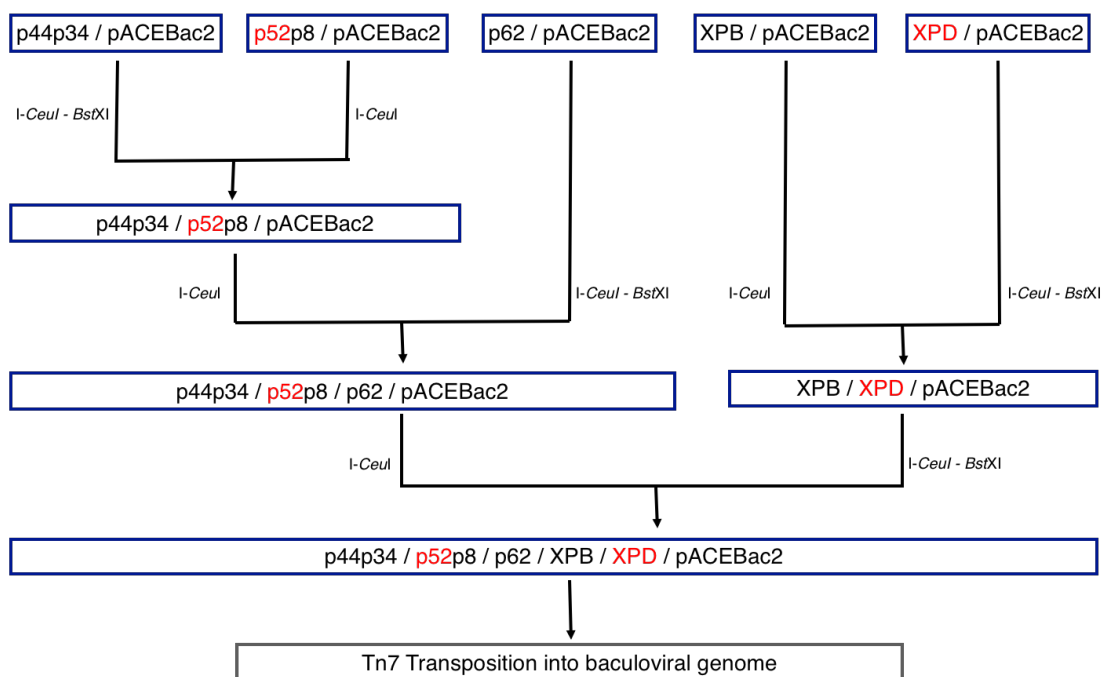


Figure 54: Steps in the cloning of TFIIH Core with 6xHis-V5-TEV-tagged p52 and XPD.

Workflow showing the different reactions performed for the cloning of construct TFIIH Core, with subunits p52 and XPD tagged with 6xHis-V5-TEV (indicated in red) at their N-terminal end.

The integrity of each intermediate construct was checked by restriction with

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

combinations of several different enzymes, comparing the band pattern predicted *in silico* by the NEBcutter® tool (New England Biolabs) with the one we obtained after running our restricted plasmids in a 1% agarose gel. The final seven-gene construct was further analysed by PCR and lastly, by sequencing (GATC Biotech) of the seven genes using the primers listed in table C1, appendix C. Figures 55 and 56 show the restriction gels and PCR results that confirmed that the TFIIH Core construct with subunits p52 and XPD carrying an N-terminal 6-residue His tag had been successfully cloned.

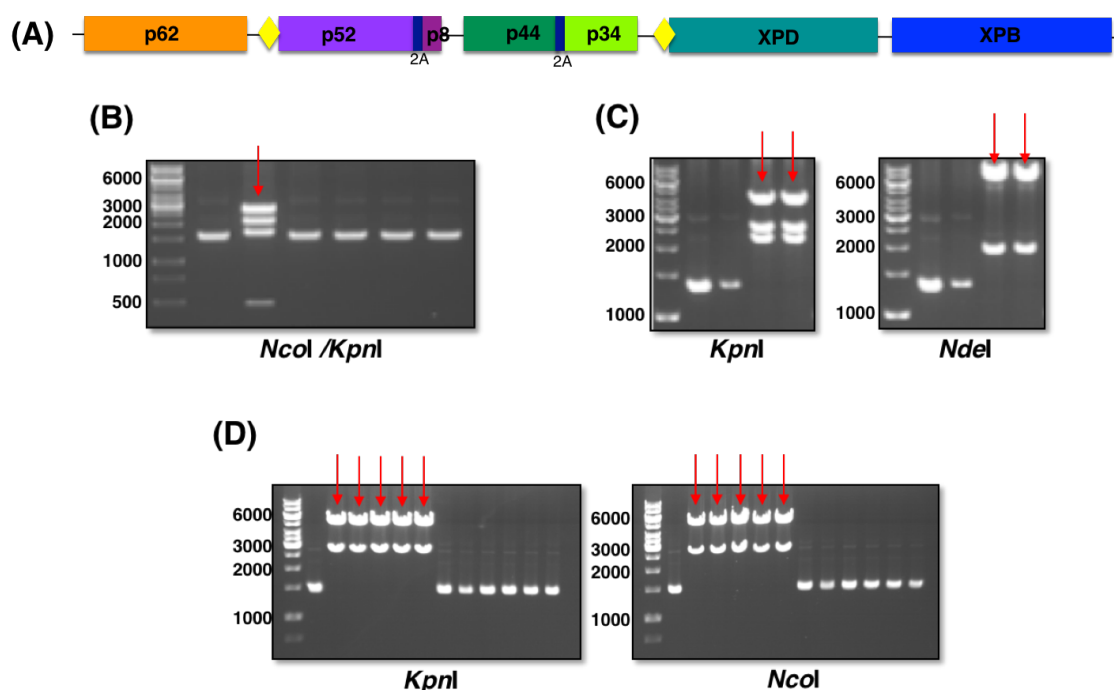


Figure 55: Cloning of TFIIH Core with 6xHis-V5-TEV-tagged p52 and XPD (I).

(A) Cartoon view of construct TFIIH Core, with subunits p52 and XPD tagged with 6xHis-V5-TEV (yellow diamonds) at their N-terminal end. (B) *NcoI*/*KpnI* restriction of construct p44-p34/p52-p8. The red arrow indicates the only clone showing the expected band pattern for plasmid pACEBac2 containing genes p44-p34 and p52-p8 (bands 2826 bp, 2164 bp, 1702 bp and 486 bp). (C) *KpnI* and *NdeI* restrictions of construct p44-p34/p52-p8/p62. Expected band pattern for clones containing all five genes are 4471 bp, 2650 bp and 2188 bp bands for restriction with *KpnI* and 7320 bp and 1989 bp bands for restriction with *NdeI* (D) *KpnI* and *NcoI* restrictions of construct XPB/XPD. The red arrows point to successful clones containing both the XPB and XPD genes. Expected patterns are 5107 bp and 2833 bp bands for restriction with *KpnI* and 5173 bp and 2767 bp bands for restriction with *NcoI*. All ladder sizes are indicated in base pairs.

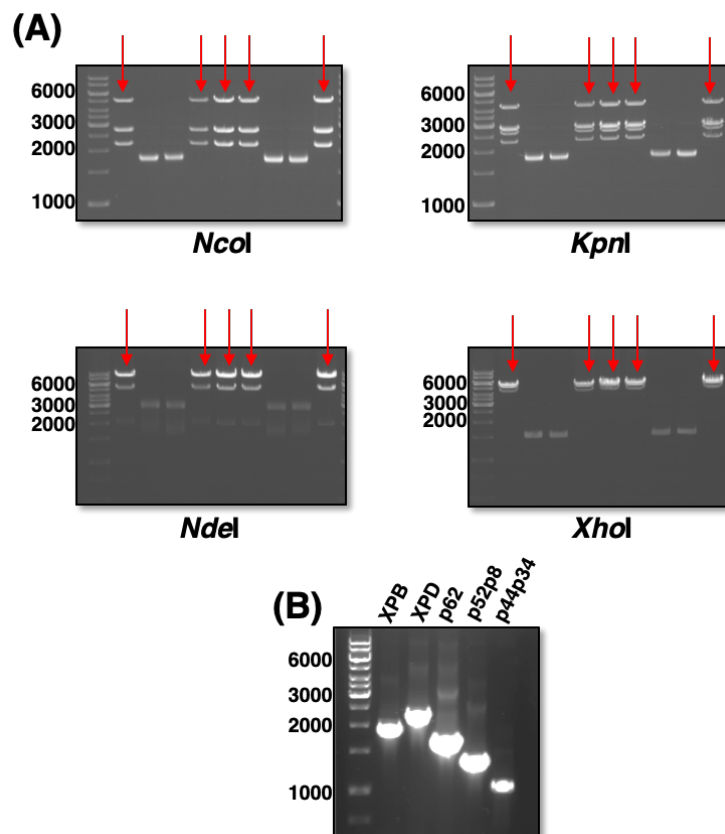


Figure 56: Cloning of TFIIF Core with 6xHis-V5-TEV-tagged p52 and XPD (II).

NcoI, *KpnI*, *NdeI* and *XhoI* restrictions (A) and PCR (B) of construct p44-p34/p52-p8/p62/XPB/XPD. Ladder sizes are indicated in base pairs. The red arrows indicate successful clones showing the band pattern expected for plasmid pACEBac2 containing all seven TFIIF Core genes (bands 5081 bp, 2767 bp, 2742 bp, 2223 bp and 2188 bp for *NcoI*; bands 4379 bp, 2859 bp, 2833 bp, 2650 bp and 2280 bp for *KpnI*; bands 8082 bp, 4930 bp and 1989 bp for *NdeI*; and bands 5509 bp, 5081 bp and 4411 bp for *XhoI*). A set of primers for each gene (see appendix C for sequences) was designed to perform a PCR, targeting the middle of the gene sequence rather than the 3' and 5' ends to avoid overlapping with the repeated pACEBac2 fragments on both ends of every gene. As a result, we expected to obtain fragments of 2197 bp (XPD), 1910 bp (XPB), 1637 bp (p62), 1317 bp (p52-p8) and 1038 bp (p44-p34), as seen in the gel.

The recombinant plasmid containing all seven TFIIF Core genes was next transposed into the baculoviral genome as described in chapter 2, section 2.1.5. The transposition protocol had to be adapted as new genes were introduced into our recombinant plasmid, increasing both the incubation and recovery times. Using freshly-made chemically competent cells also helped to improve the transposition efficiency, although the number

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

of white colonies obtained per reaction was frequently still lower than five.

The successfully transposed recombinant plasmid, called bacmid, was purified and transfected into a monolayer of *Sf9* cells. The P0 virus recovered from this transfection was used to infect a 25 ml suspension culture, which was harvested and tested for TFIIH Core expression after 60 hours of incubation. To verify the Core sub-complex expression levels, we ran a BioSprint purification test as described in chapter 2, section 2.3.1, which easily and quickly allowed us to see that this particular construct was not highly expressed (figure 57).

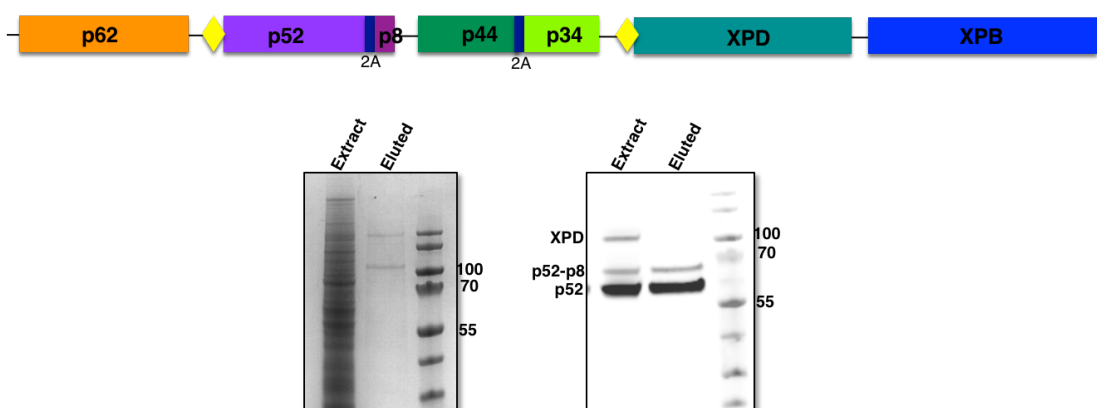


Figure 57: Purification test of TFIIH Core with 6xHis-V5-TEV-tagged p52 and XPD.

Coomassie gel (left) and Western blot (right) showing a BioSprint purification test for an infection with a P0 recombinant virus containing construct TFIIH Core with subunits p52 and XPD tagged at their N-terminal end with 6xHis-V5-TEV. Ladder sizes are indicated in kDa.

A Western blot performed with an anti-V5 primary antibody and IRDye 800CW Goat anti-Mouse IgG (H + L) as a secondary antibody showed a strong signal for p52, but not for XPD, suggesting lower expression levels for the helicase than for the rest of the Core subunits, or a loss of the protein during the purification process. Quantification of the blot bands also showed that although cleaving of the 2A-like sequence joining subunits p52 and p8 was very efficient (87%), a small percentage (13%) remained uncleaved.

4.2.1.2. A redistribution of the purification tags is required

After optimisation of the transfection and infection protocols failed to improve the expression levels of construct TFIIH Core with subunits p52 and XPD carrying an N-terminal 6xHis-V5-TEV tag, we reconsidered our strategy regarding the position of the purification tags. The Western blot on figure 57 shows that the tag placed on N-terminal p52 worked better than the one placed on XPD, so we pondered if placing a second tag in subunit p62 could help to keep the sub-complex together during the purification process, given the more central position of the subunit within the ring-like structure of the Core sub-complex (Schultz *et al.*, 2000).

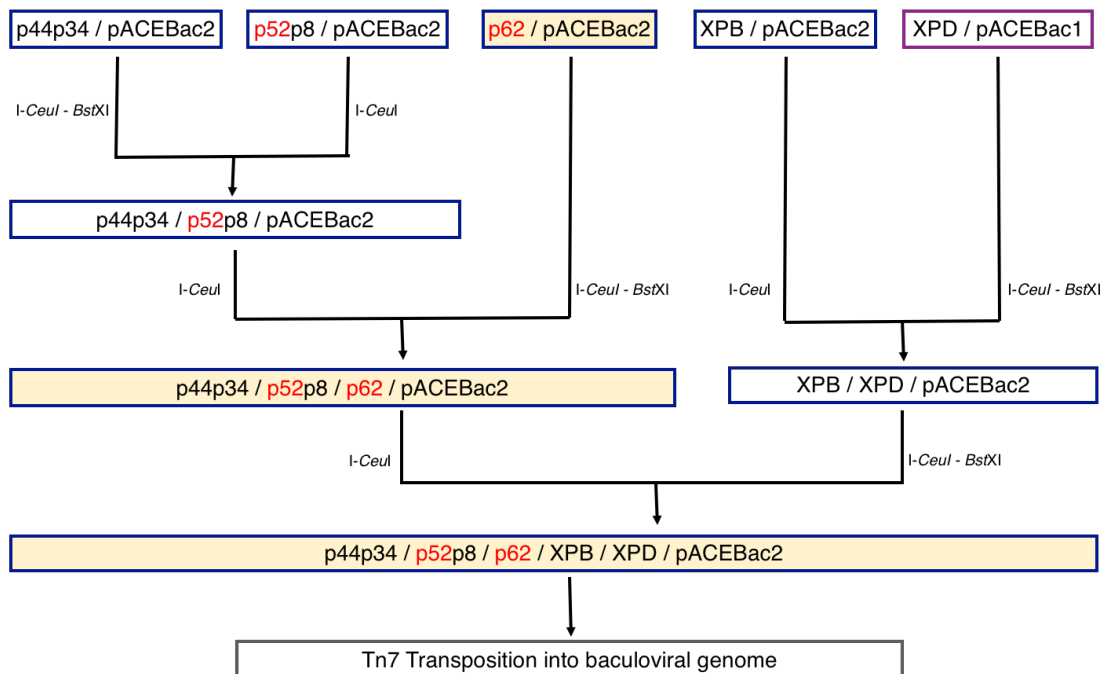


Figure 58: Steps in the cloning of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.

Workflow showing the different steps performed for the cloning of construct TFIIH Core sub-complex, with subunits p52 and p62 tagged with 6xHis-V5-TEV (indicated in red) at their N-terminal end. Only the cloning steps highlighted in yellow had to be carried out to obtain the full seven-gene plasmid, as intermediates (white boxes) had already been obtained in the cloning of a previous construct.

Many of the intermediate constructs required for the cloning of this new TFIIH Core sub-complex with subunits p52 and p62 carrying a 6xHis-V5-TEV tag at their N-terminal end

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

had already been obtained during the cloning process of our previous construct. The characteristics of the MultiBac™ system allow the use of these intermediate multi-gene plasmids (Bieniossek *et al.*, 2012), which meant that the new construct was obtained in only three cloning steps. This advantage considerably reduced the time scale necessary to obtain a new full TFIIH Core sub-complex construct, as can be appreciated in figure 58.

As with the previous construct, the integrity of the final seven-gene TFIIH Core sub-complex was verified by restriction with enzymes *KpnI*, *BamHI / HindIII* and *Sall / XhoI* and PCR (figure 59).

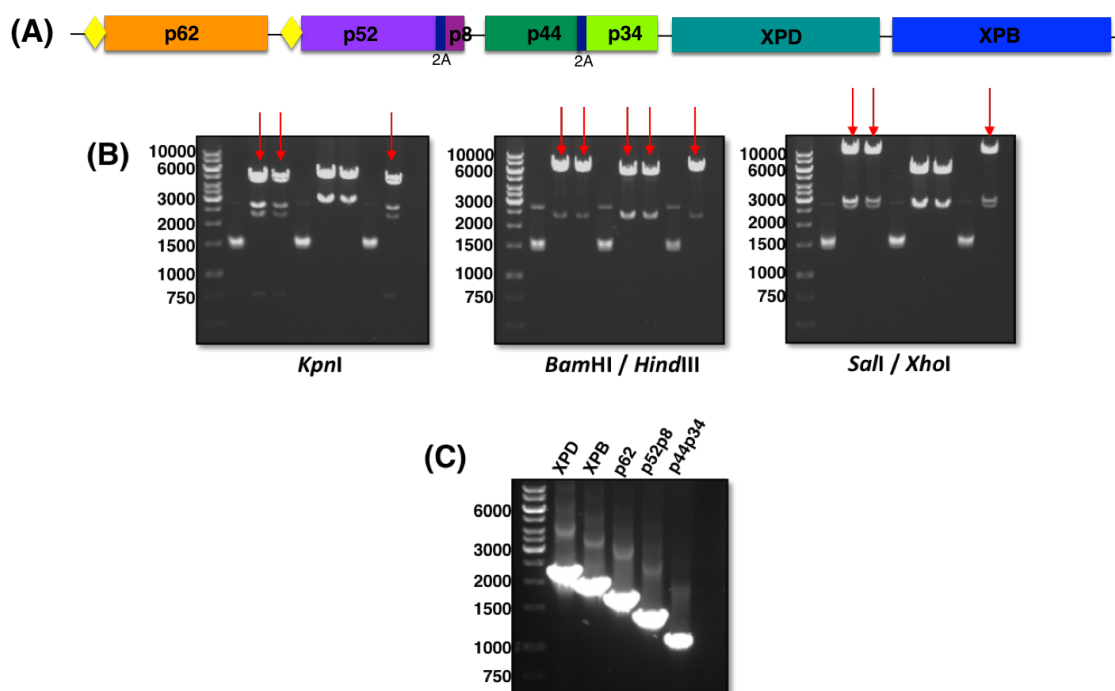


Figure 59: Cloning of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.

(A) Cartoon view of construct TFIIH Core with subunits p52 and p62 tagged with 6xHis-V5-TEV (yellow diamonds) at their N-terminal end. Restriction reactions (B) and a PCR (C) were performed to analyse the integrity of this construct. The red arrows indicate the clones showing the expected band pattern for plasmid pACEBac2 containing all seven TFIIH Core genes (bands 5000 bp, 4471 bp, 2650 bp, 2280 bp and 748 bp for *KpnI*; bands 6767 bp, 6030 bp, and 2352 bp for *BamHI/HindIII*; and bands 9584 bp, 2916 bp and 2649 bp for *Sall/XhoI*). A PCR performed with primers targeting the middle of the gene sequence further confirmed the successful cloning of genes XPD, XPB, p62, p52-p8 and p44-p34. Ladder sizes are indicated in base pairs.

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

The recombinant plasmid containing all seven TFIIH Core genes was transposed into the baculoviral genome and successfully transposed bacmids, selected as described in chapter 2, section 2.1.5, were transfected into a monolayer of *Sf9* cells. The P0 virus recovered was used to infect a 25 ml *Sf9* suspension culture from which the new TFIIH Core construct was extracted and analysed for expression levels in a BioSprint purification test as previously described (figure 60 (A)).

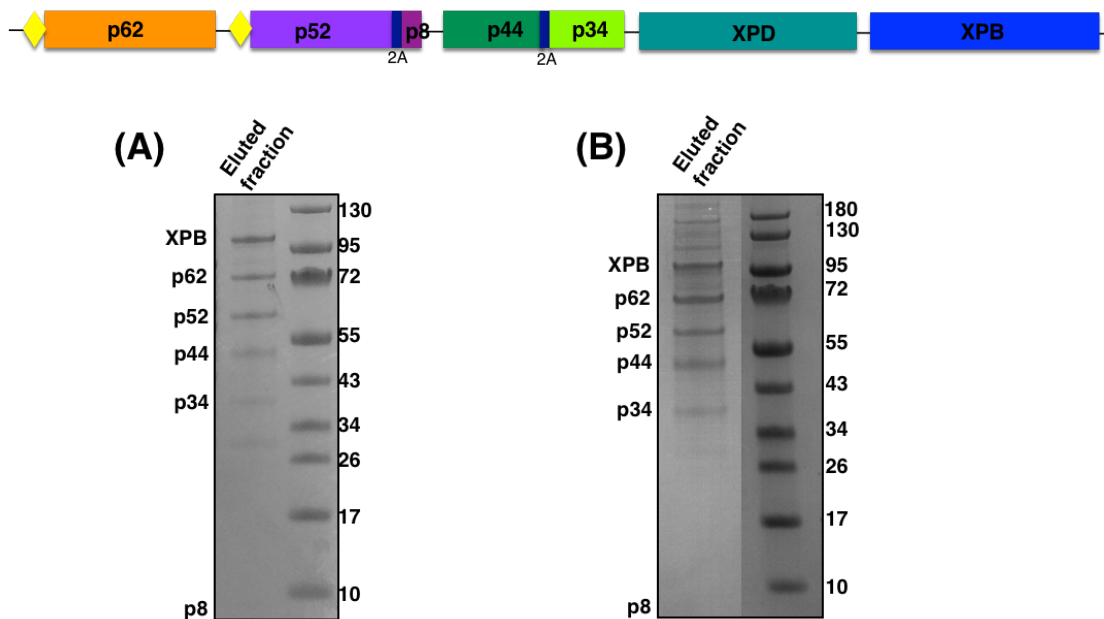


Figure 60: Purification test of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.

Coomassie-stained gel showing a BioSprint expression test (A) and a BioSprint test in anaerobic conditions (B) of an infection with a P0 recombinant virus containing construct TFIIH Core with subunits p52 and p62 carrying a 6-residue His tag at their N-terminal end. Ladder sizes are indicated in kDa. XPD is missing in both gels.

This result was an enormous improvement over our first attempt to purify the Core sub-complex, as subunits XPB, p62, p52, p44, p34 and p8 all clearly appear in the gel in stoichiometric quantities and were confirmed by MS analysis (data not shown), but the achievement was somewhat dampened by the fact that subunit XPD was missing from the gel.

With this in mind, we performed a BioSprint purification test under anaerobic conditions that would prevent the iron-sulfur (4FeS) cluster present in XPD from being oxidized,

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

hopefully avoiding loss of the helicase during the purification of the Core sub-complex. However, a SDS-PAGE gel with the samples extracted in anaerobiosis showed a similar result to that obtained with the aerobic BioSprint (figure 60 (B)).

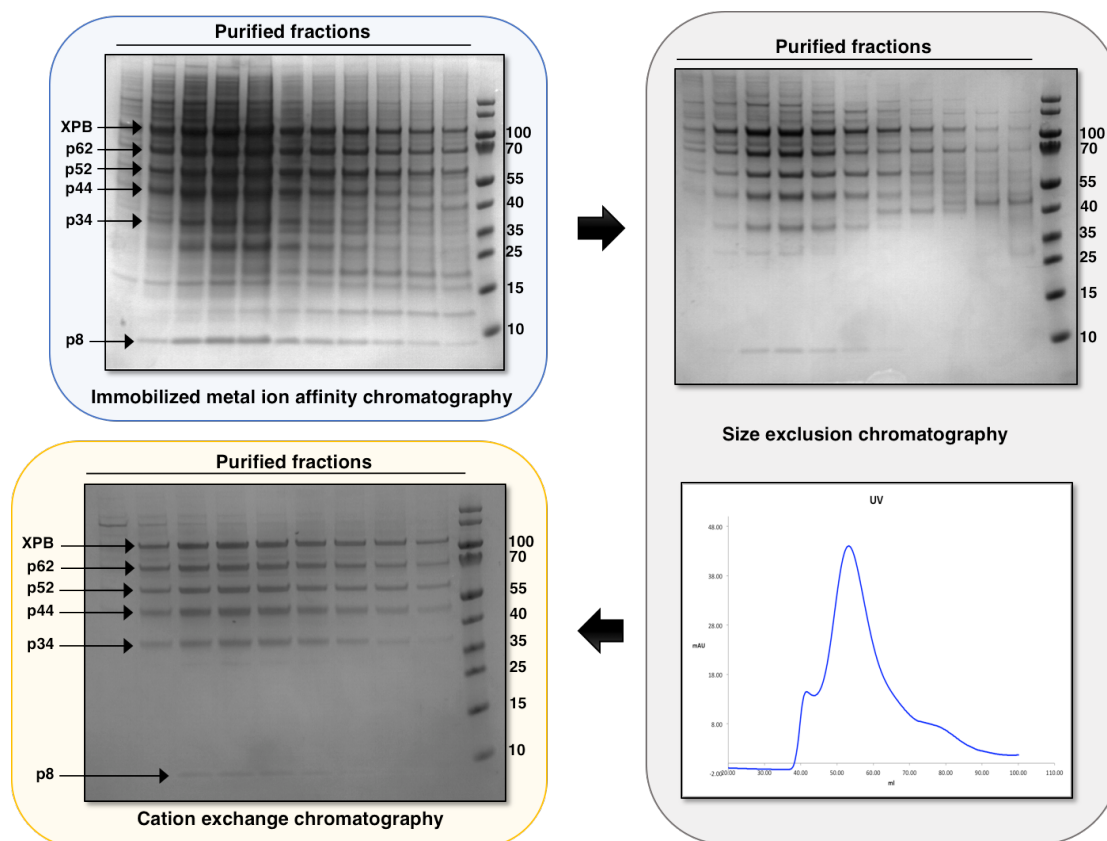


Figure 61: Purification of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.

Coomassie-stained gels showing peak fractions for the different chromatographic steps followed in the purification of the TFIIH Core sub-complex with subunits p52 and p62 tagged at their N-terminal end with 6xHis-V5-TEV: ion affinity (HisTrap™ FF column, blue box), size exclusion (HiPrep™ 16/60 Sephacryl™ S300 HR column, grey box) and cation exchange (MonoS™ 4.6/100 PE column, yellow box). Ladder sizes are indicated in kDa.

Since the purification of the sub-complex in anaerobic conditions didn't represent any improvement regarding helicase XPD or any of the other Core subunits, our efforts focused on scaling up our infections to improve the yield for the sub-complex and troubleshoot our aerobic purification protocol to solve the issues affecting XPD. This involved testing and optimising different buffers and conditions until a 3-step procedure was finally established. In this 3-step protocol, the Core sub-complex was first run through a Nickel column (immobilized metal ion affinity chromatography (IMAC)), as

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIF COMPLEX

described in chapter 2, section 2.3.3.1, followed by a size-exclusion chromatography step (section 2.3.4) and finally a cation exchange chromatography (section 2.3.5). Peak fractions were checked in a SDS-PAGE gel after every purification step, showing that our protocol allows the extraction of a highly pure complex containing subunits XPB, p62, p52, p44, p34 and p8, but no XPD (figure 61).

Scaling up and optimisation of the expression and purification procedures increased yield to 1 mg of pure sub-complex for every 2 L culture. Subunits XPB, p62, p52, p44, p34 and p8 can be efficiently purified with this protocol, but subunit XPD is lost along the process.

To investigate why XPD was not being purified we took samples from every different step along our purification protocol and performed a Western blot that would allow us to confirm if XPD was indeed being expressed, and if so, where exactly it was that the helicase was lost (figure 62).

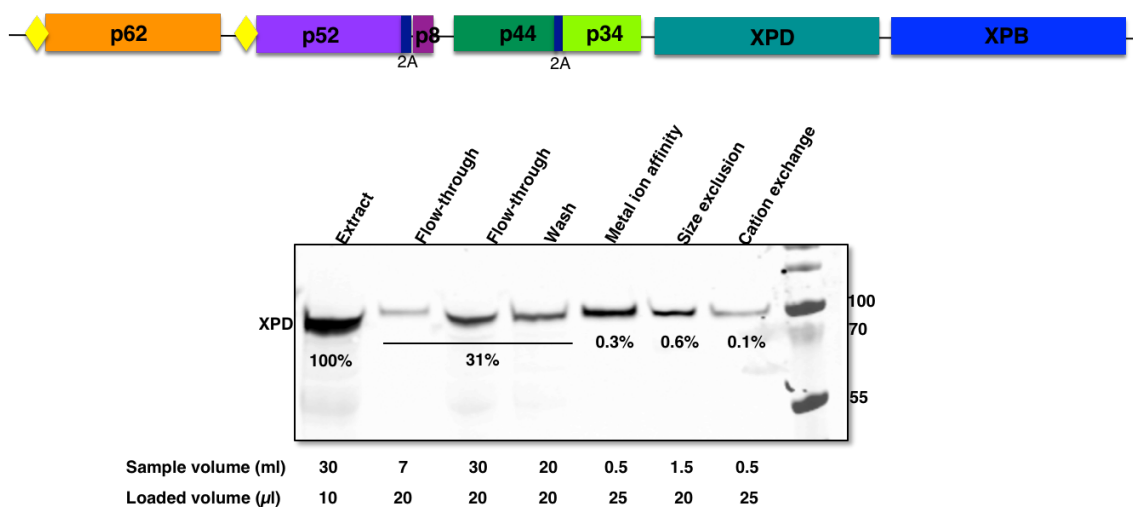


Figure 62: Subunit XPD is lost in the purification of TFIIF Core.

Western blot showing XPD after every purification stage included in our protocol. 31% of the helicase was lost due to poor binding (flow-through, washing step), and only 0.3% of the loaded protein was recovered after IMAC, to finally retrieve only 0.1% of the initial XPD available in our sample after three purification steps. Ladder sizes are indicated in kDa.

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

The Western blot showed that although the helicase is clearly present in the extract, significant losses occur at every step, even before binding to the first chromatography column, to finally end with as little as 0.1% XPD at the end of the process. Only 0.3% of the initial XPD is recovered after its purification through a HisTrap™ FF column, suggesting that most of the helicase present in the extract probably precipitates and/or aggregates as it passes through the column. Taken together, these results pointed to XPD being more susceptible to dissociation from the sub-complex after extraction and binding not being as strong as it was for the rest of the subunits in the Core sub-complex.

Our three-step IMAC – size exclusion – cation exchange protocol was further improved by substituting the nickel column for a Talon® Superflow™ resin and adapting our method as described in chapter 2, section 2.3.3.3. Purification with a loose resin instead of the pre-packed matrix present in the HisTrap™ FF column we had been using so far allowed us to increase the binding time, resulting in a higher recovery of the sub-complex. Loss of the sub-complex was also further reduced by eliminating the cation exchange chromatographic step: after IMAC with the Talon® resin and size exclusion in a HiPrep™ 16/60 Sephacryl™ S300 HR column, the recovered sample had approximately the same quality standards as the sample obtained after our previous three-step protocol, but recovery of the sub-complex was slightly higher, and the purification process was now shorter and simpler. Fractions of interest were analyzed in a SDS-PAGE gel after every purification step, showing again a pure complex containing stoichiometric amounts of subunits XPB, p62, p52, p44, p34 and p8, but no XPD could be detected by the Coomassie-staining of the gel (figure 63).

MS analysis confirmed that this sample contained a considerably reduced amount of XPD compared to the other six TFIIH Core subunits (data not shown), thus making it not suitable for the biochemical characterization of the sub-complex, but this didn't affect a crosslinking assay, which further confirmed the presence of XPD in our purified sample, and also offered additional information regarding the interactions between the different Core subunits.

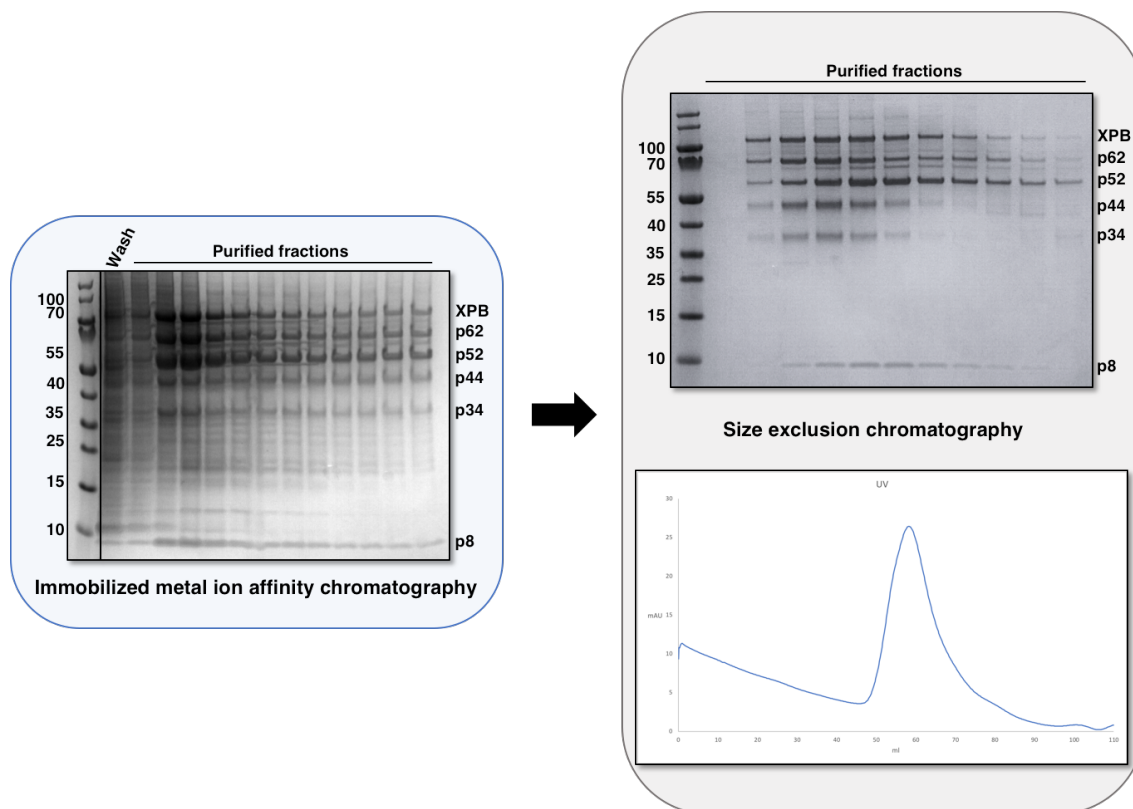


Figure 63: Optimised purification of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52.

Coomassie-stained gels showing a simplified purification protocol for construct TFIIH Core with subunits p52 and p62 tagged at their N-terminal end with 6xHis-V5-TEV. The procedure was reduced to two steps: ion affinity (Talon® Superflow™ resin, blue box), and size exclusion (HiPrep™ 16/60 Sephacryl™ S300 HR column, grey box). Ladder sizes are indicated in kDa.

The TFIIH Core sub-complex was crosslinked with a BS3 amine-reactive crosslinker, able to link lysine residues both within a single subunit (intralinks) or between two different subunits (interlinks) situated as far as 30 Å apart. Crosslinked samples were loaded into a SDS-PAGE gel (figure 64) and bands showing potential crosslinks were analyzed as explained in chapter 2, section 2.5.2, obtaining a total of 25 interlinks and 10 intralinks (table 5). Nine of the 25 interlinks had not been previously reported (Luo *et al.*, 2015), and two of those were obtained more than once (a p52-p62 interlink (positions 440:453 and 544:548), obtained three times, and a XPB-p34 interlink (positions 326:334 and 150:158), obtained twice).

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

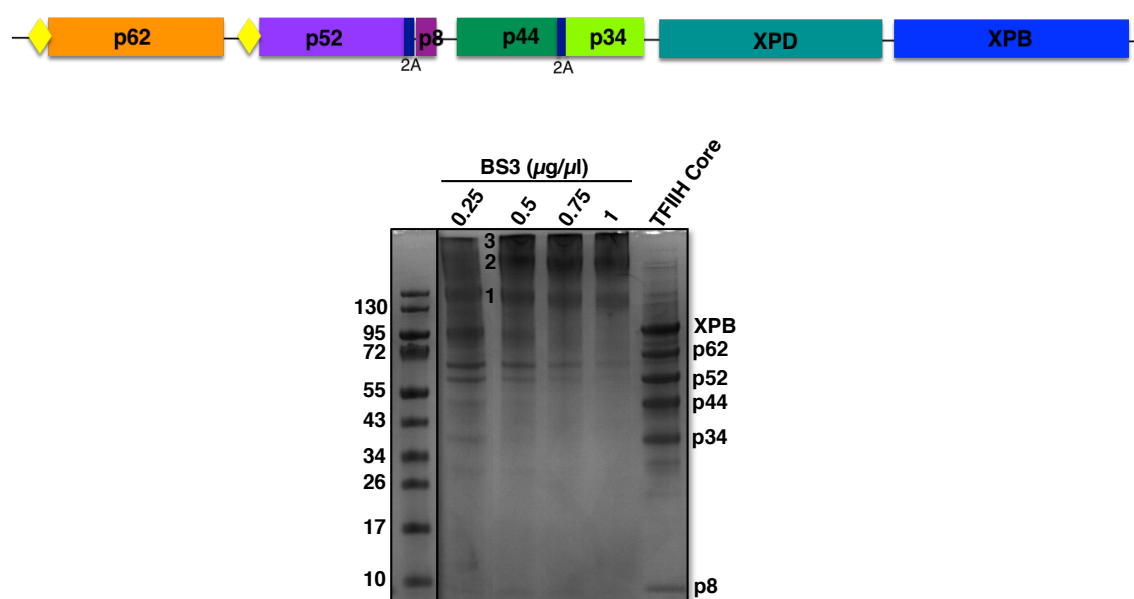


Figure 64: TFIIH complex and sub-complexes obtained by crosslinking.

Coomassie-stained gel showing the different complexes and sub-complexes obtained by crosslinking of TFIIH Core with subunits p52 and p62 carrying an N-terminal 6-residue His tag. Ladder sizes are indicated in kDa. Bands 1, 2 and 3 were analysed by MS and the resulted data was further evaluated with the MassMatrix software.

Table 5: Crosslinks connecting TFIIH Core subunits.

Summary of all the intralinks (top) and interlinks (bottom table) obtained in our experiments. Novel crosslinks not reported in previous works are highlighted in yellow. Crosslinked lysine residues are highlighted in red, with their exact position within the protein chain indicated by symbol \$.

INTRALINKS	POSITION	SUBUNIT	REGION	REPEATED
NSEGEATELITETFTSK(\$217)SAISK K(\$326)MFGNGR	201:222 326:332	XPB	DRD	x1
FGTMSSMSGADDTVMEYHSSRSK(\$767)APSK FLVDQGYFSK(\$688)VITK	744:771 679:692	XPB	C-terminal	x1
NDSVNPIDINLDKPTAVLRPYQEK(\$322)SLR K(\$326)MFGNGR	299:325 326:332	XPB	DRD	x1
NDSVNPIDINLDK(\$311)PTAVLRPYQEK SLRK(\$326)MFGNGR	299:322 323:332	XPB	DRD	x1
LVLK(\$171)HNR MGK(\$3)RDR	168:174 01:06	XPB	N-terminal	x1
SWEAERVMEWLK(\$431)TQEWGLMILDEVHTIPAK EYVAIKK(\$528)K	420:449 521:529	XPB	HD1 HD1-HD2	x1
GK(\$671)TDYGLMVFADKR K(\$370)PLRFAER	670:683 370:378	XPB	HD2 Arch	x1
TYPVAVKMK(\$206)YAENVPHNMTEK VVK(\$492)MKSNLER	199:218 490:499	p62	N-terminal C-terminal	x1
LMVVTAPAGHSDVK(\$452)R K(\$77)EFSK	440:453 77:81	p52	C-terminal N-terminal	x1
LMVVTAPAGHSDVK(\$452)R FWK(\$456)R	440:453 454:457	p52	C-terminal	x1

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIF COMPLEX

INTERLINKS	POSITION	SUBUNIT	REGION	REPEATED
TRAHPVMLK(\$378)QTPVLPPTITDQIR SK(\$112)R	370:392	p52	C-terminal	x3
	111:113	p44	VWA	
VGELMDQNAFSLTK(\$71) TK(\$528)K	57:71	p8	C-terminal	x3
	527:529	XPB	HD2	
LMVVTPAGHSDVK(\$452)R LMKK(\$547)T	440:453	p52	C-terminal	x3
	544:548	p62	C-terminal	
K(\$326)MFGNGRAR EVK(\$152)DNQEMK	326:334	XPB	DRD	x2
	150:158	p34	VWA	
DAVK(\$93)DLLQLLPK IADLVSK(\$223)ELARK	90:102	p62	PH	x1
	217:228	XPB	HD1	
IALLGGGK(\$127)AWSDDTSQLGPK KLSK(\$145)TGVDPGIMQFIK	120:140	p52	N-terminal	x1
	142:157	XPB	N-terminal	
MNK(\$149)EVKDNQEMK VDEYGA(\$59)DYS	147:158	p34	VWA	x1
	53:62	XPB	N-terminal	
GHGVLEMPSTGK(\$48)TVSLLALIMAYQR K(\$326)MFGNGR	36:61	XPB	N-terminal	x1
	326:332	XPB	DRD	
SIK(\$296)ENSNAAIKR TMEDQDLK(\$77)PNR	294:306	p62	N-terminal	x1
	70:80	p44	VWA	
K(\$113)NLCIHPEVTPFRFGK K(\$378)SDR	113:128	XPB	FeS	x1
	378:381	p62	C-terminal	
EVPLPAGIYNLDDLK(\$181)ALGRR ISLPPVLKAK(\$300)K	167:186	XPB	FeS	x1
	291:301	p34	Zn finger	
K(\$142)LSKTVGVPDGIMQFIK DNQEMK(\$158)SRILVIK	142:157	XPB	N-terminal	x1
	153:165	p34	VWA	
LSKTVGVPDGIMQFIK(\$157)LCTVSYGK K(\$378)SDR	143:165	XPB	N-terminal	x1
	378:381	p62	C-terminal	
SAISK(\$222)TAESSGGPSTSR MATSSEEVLLIVK(\$13)K	218:234	XPB	N-terminal	x1
	01:14	p62	PH	
K(\$648)GMVAEEYNFAFFYSLSVSDTQEMAYSTK GK(\$603)VSEGIDFVHHYGR	648:675	XPB	HD2	x1
	602:616	XPB	C-terminal	
YNLTSDIIESIFRTYPAVK(\$204)MK EVK(\$152)DNQEMK	186:206	p62	N-terminal	x1
	150:158	p34	VWA	
MATSSEEVLLIVK(\$13)K TDYGLMVFADK(\$682)R	01:14	p62	PH	x1
	672:683	XPB	HD2	
EFSK(\$81)AQEESTGLLSGLR ADRDK(\$11)K	78:94	p52	N-terminal	x1
	07:12	XPB	N-terminal	

The representation of our crosslinks over a model of the TFIIF Core sub-complex showed a compact, tightly bound structure with multiple interactions (figure 65). These findings were obtained at the same time the Ranish group published an exhaustive and excellent paper on this same line of research (Luo *et al.*, 2015), showing good agreement with our own observations and overlapping with most of the interlinks and intralinks obtained in our experiments, with the exception of nine novel interlinks exclusively seen in our studies.

Although a Western blot confirmed that XPD was indeed expressed (figure 62), and our crosslink assays (table 5 and figure 65) even proved that some XPD was present in our purified sample, we still hadn't managed to purify XPD in the same quantity as the other Core subunits. Since the Western blot seemed to suggest that the most pressing issue was the binding of the helicase to the HisTrap™ FF column, we next focused on improving the extraction and binding of XPD during the IMAC step.

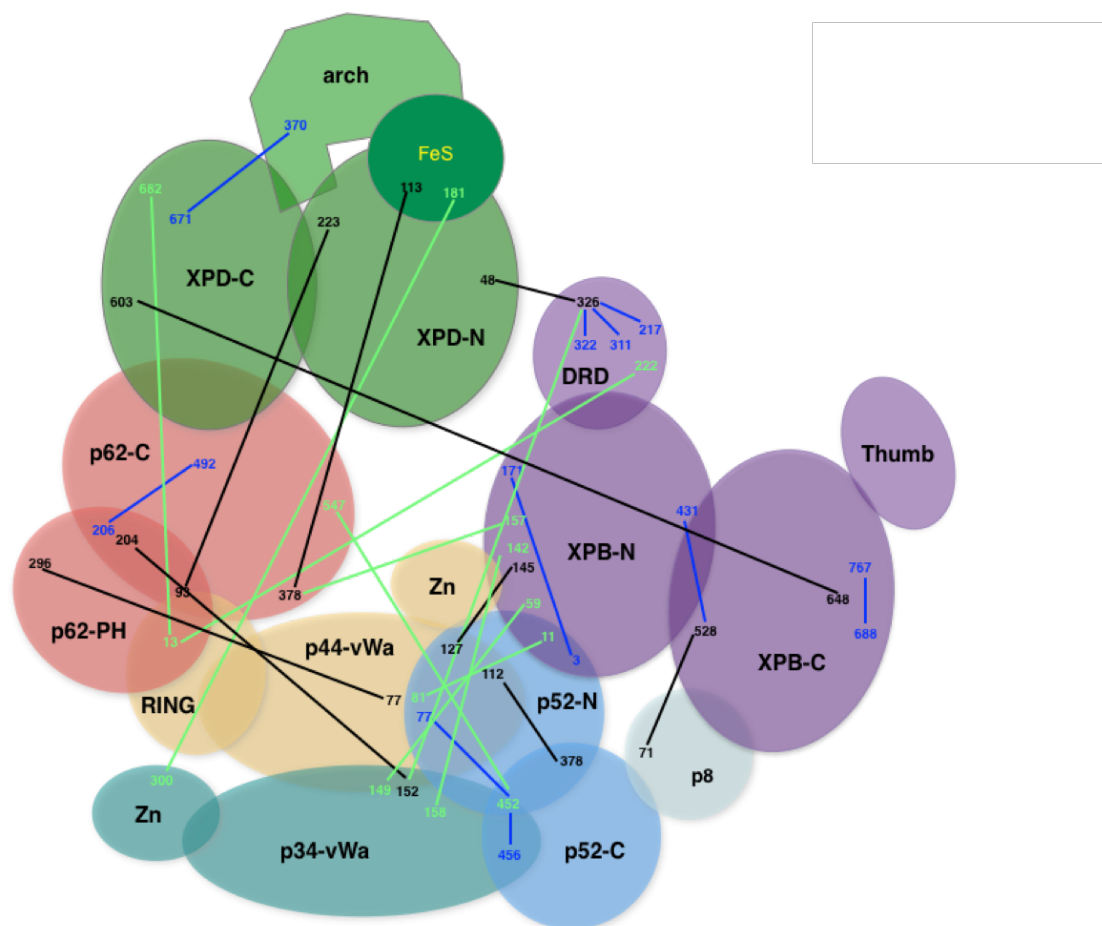


Figure 65: Map of the crosslinks obtained in our assays.

Intralinks are shown as blue lines, previously reported interlinks are represented with black lines, and brand-new interlinks reported in this thesis are shown as green lines.

4.2.1.3. A different purification tag might be the answer

Since the versatility of the MultiBac™ system allowed us to obtain different Core constructs in a relatively short time, we also tried different approaches to obtain the TFIIH Core sub-complex while optimising the purification protocol for the Core construct with subunits p52 and p62 carrying a 6xHis-V5-TEV N-terminal tag.

The 6xHis-V5-TEV tag had been designed with several functions in mind: (1) purification of our proteins in an IMAC by means of the 6-residue His tag, (2) Western blot detection and immunoprecipitation (IP) of the complex through the V5 epitope, and (3) removal of the tag from the purified protein if necessary through cleaving at the TEV site. However, our attempts to purify the Core sub-complex by IP (data not shown) exposed a problem

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

with this design: when release of the protein with the appropriate elution buffer failed, we tried to achieve this by cleaving with TEV protease, but the sub-complex still remained bound to the beads. This was probably due to the closeness between the V5 epitope and the TEV cleavage site in our tag, which meant that a bound V5 antibody would block access to the TEV site. To solve this problem, we introduced a 9-residue spacer between the V5 and TEV sequences (see chapter 2, section 2.1.2), and we cloned the new tag into three different constructs: TFIIH Core with subunits p52 and XPD tagged at their N-terminal end with 8xHis-V5-spacer-TEV, and a 6-subunit (XPB, p62, p52, p44, p34 and p8) Core construct with p52 carrying the N-terminal 8xHis-V5-spacer-TEV tag, with XPD (also tagged with 8xHis-V5-spacer-TEV at its N-terminal end) cloned into a separate plasmid (figure 66).

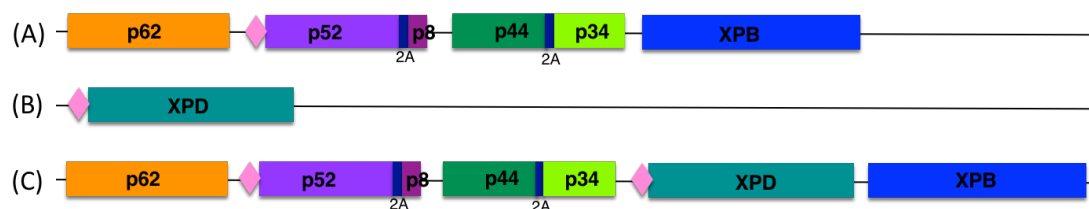


Figure 66: TFIIH constructs carrying an 8xHis-V5-spacer-TEV tag.

Cartoon view of constructs (A) 6-subunit (XPB, p62, p52, p44, p34 and p8) TFIIH Core, (B) XPD and (C) 7-subunit (XPD, XPB, p62, p52, p44, p34 and p8) TFIIH Core. Subunits p52 and XPD carry the N-terminal 8xHis-V5-spacer-TEV tag (represented here by a pink diamond) in all three constructs.

We first cloned and expressed the sub-complex containing all seven Core subunits following the same method explained for previous constructs. Our first BioSprint test showed very poor expression levels (figure 67, (A)), so we tried a second infection and BioSprint test. This second attempt (figure 67, (B)) only offered slightly better results, and subunit XPD still could not be seen in the gel. A Western blot confirmed that XPD is present in our extract, although approximately 5 times less abundant than p52, and it mostly remains in the unbound fraction of our purification. As for subunit p52, we only recover 4% of the total protein present in our extract, but it's still 9-fold the amount of the recovered XPD. The Western blot also confirms that a lot of the sub-complex present in our original sample (78% XPD and 72% p52) is lost, probably due to precipitation or

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

aggregation, before the purification process is finished (figure 67, (C)).

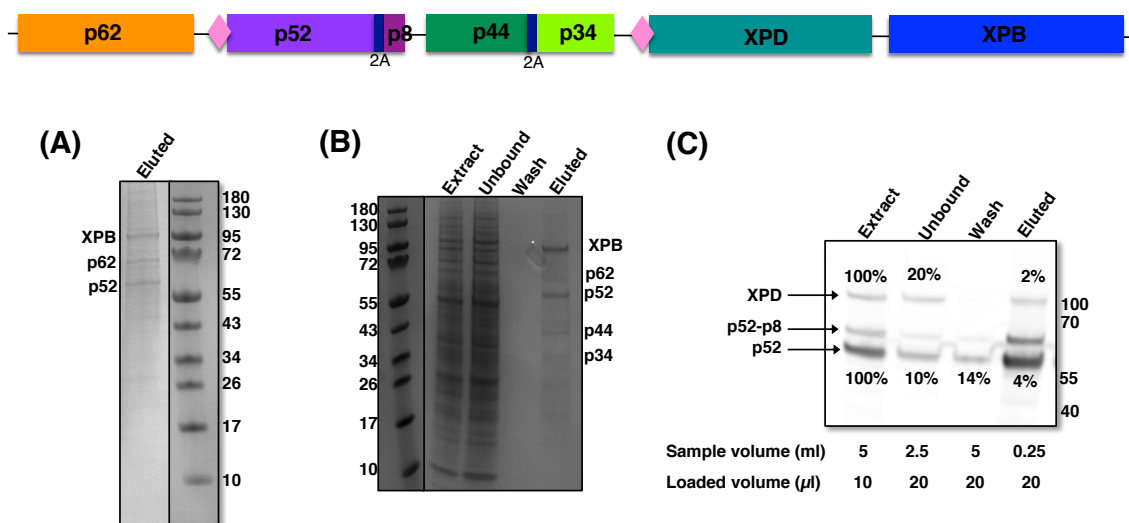


Figure 67: Purification test of TFIIH Core with 8xHis-V5-spacer-TEV-tagged XPD and p52.

Coomassie-stained gel showing our first BioSprint expression test (A) and a second BioSprint test for a new infection (B) for construct TFIIH Core with subunits p52 and XPD carrying an 8-residue His tag at their N-terminal end. A Western blot was also performed for this second infection (C), showing that only 2% of the total XPD available is recovered after purification, with 20% of it lost due to poor binding. In the case of p52 the recovered percentage increases to 4%, with 24% of the protein lost in the process of binding the beads (unbound and wash fractions). However, as the initial amount of p52 available in the sample is considerably higher than that of XPD, the final amount of purified p52 is still 9 times higher than purified XPD. Ladder sizes are indicated in kDa.

Scaling up our infections and purifying the construct following our IMAC - size exclusion - cation exchange chromatography protocol offered much better results with regards to the yield of subunits XPB, p62, p52, p44, p34 and p8, but the SDS-PAGE gels showed no noticeable improvement regarding XPD (figure 68); moreover, the sub-complex with the 6xHis-V5-TEV-tagged p52 and p62 subunits consistently offered a better yield than this construct.

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIDH COMPLEX

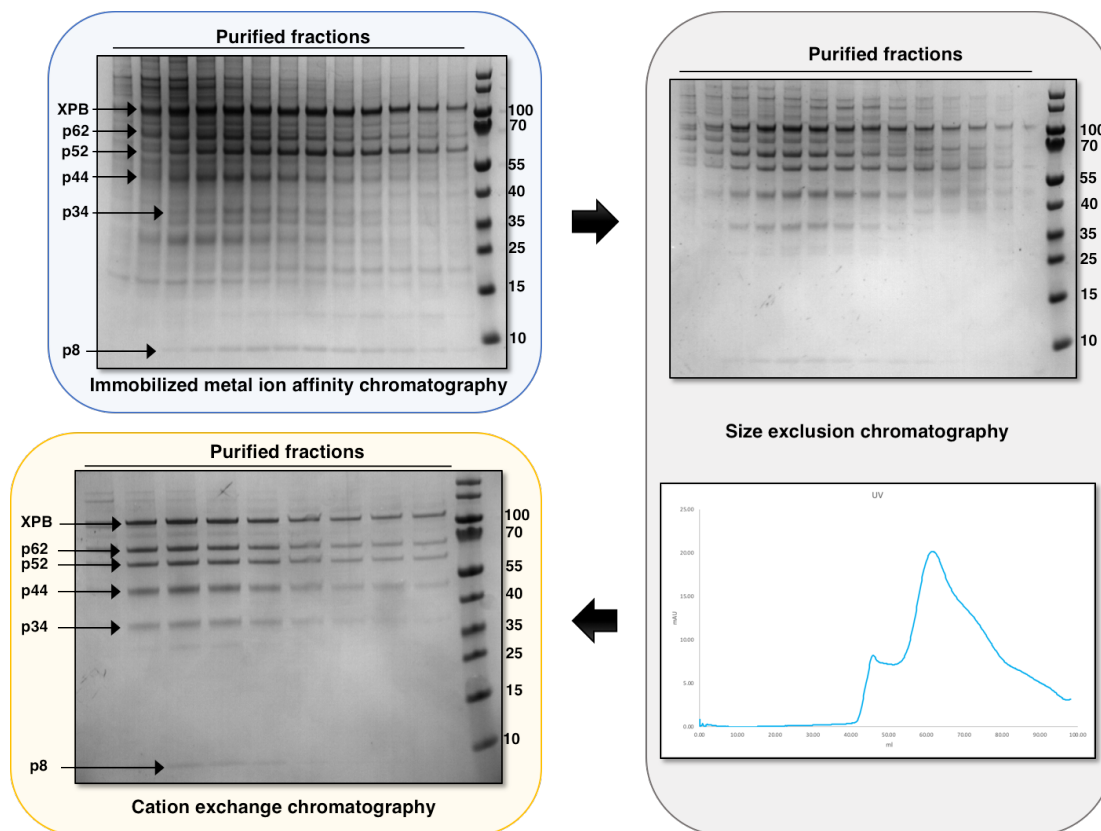


Figure 68: Purification of TFIIDH Core with 8xHis-V5-spacer-TEV-tagged XPD and p52.

Construct TFIIDH Core with subunits p52 and XPD tagged at their N-terminal end with 8xHis-V5-spacer-TEV was purified following the same protocol developed for the sub-complex with subunits p52 and p62 tagged with 6xHis-V5-TEV. The Coomassie-stained gels included here show the sub-complex at the different stages: ion affinity (HisTrap™ FF column, blue box), size exclusion (HiPrep™ 16/60 Sephacryl™ S300 HR column, grey box) and cation exchange (MonoS™ 4.6/100 PE column, yellow box). Ladder sizes are indicated in kDa.

As the 7-subunit, 8xHis-V5-spacer-TEV-tagged Core sub-complex had not offered any improvement over our previous results, we moved onto the expression of the 6-subunit sub-complex, with XPD produced separately. This new approach, with helicase XPD cloned and expressed on its own, offered us two options: we could either co-infect our cultures with two different viruses (one containing XPD and the other containing the 6-subunit Core sub-complex) or we could infect cultures with each virus separately, expressing and purifying the constructs on their own and combining our purified samples at the end. Unfortunately, the BioSprint expression test for the 6-subunit Core sub-complex showed that the subunits were not expressed equally in this particular construct

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

(figure 69), with subunits XPB and p52 clearly more abundant than p62, p44, p34 and p8. A scaled-up infection and purification following the previously described three-step protocol confirmed this result (figure 70).

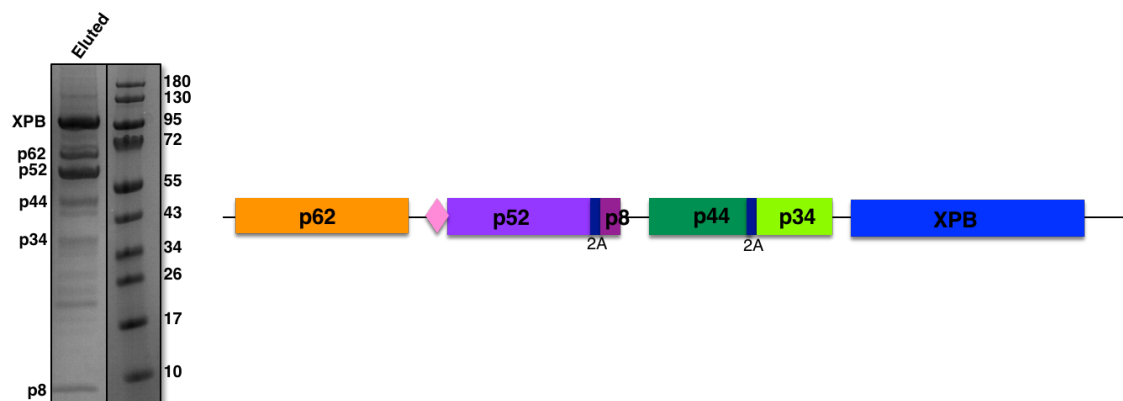


Figure 69: Purification test of 6-subunit TFIIH Core with 8xHis-V5-spacer-TEV-tagged p52.

Coomassie-stained gel showing a BioSprint expression test for construct 6-subunit TFIIH Core with subunit p52 carrying an 8-residue His tag at N-terminal.

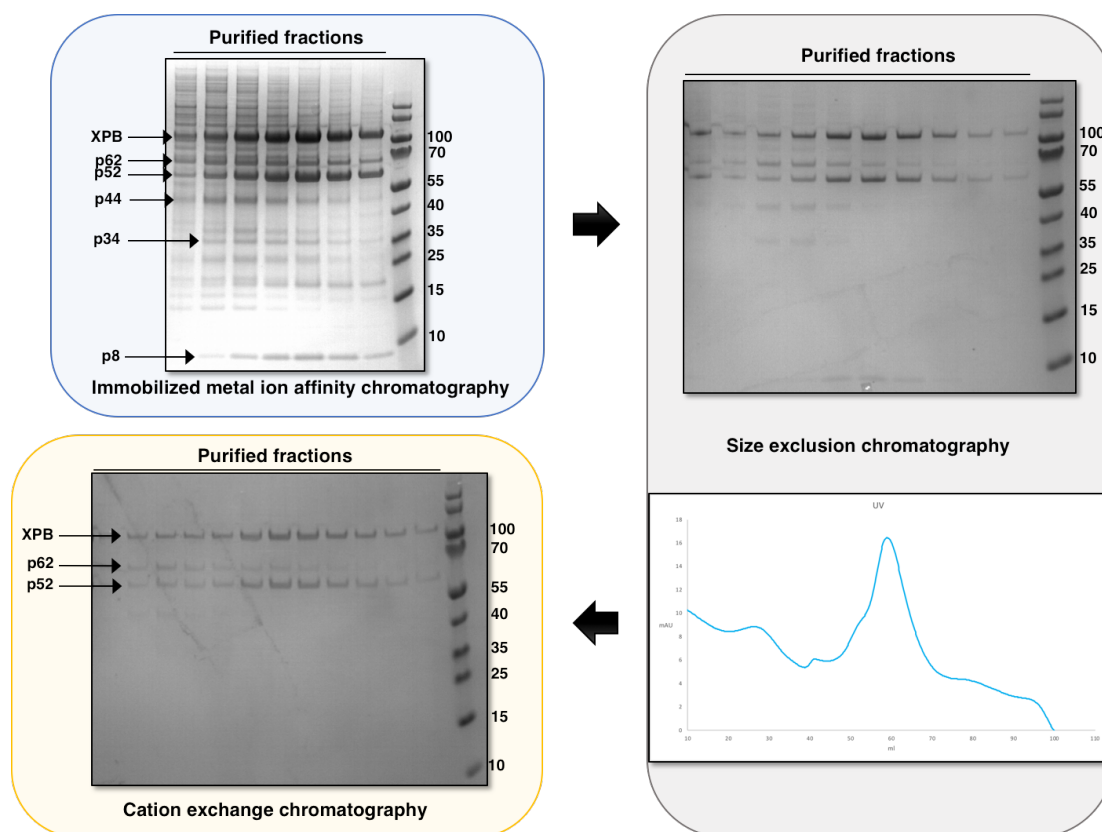


Figure 70: Purification of 6-subunit TFIIH Core with 8xHis-V5-spacer-TEV-tagged p52.

Coomassie-stained gels showing the different chromatographic steps followed to purify construct

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIF COMPLEX

6-subunit TFIIF Core sub-complex with subunit p52 tagged at its N-terminal end with 8xHis-V5-spacer-TEV. The protocol included the same steps followed in the purification of previous TFIIF Core constructs: ion affinity (HisTrap™ FF column, blue box), size exclusion (HiPrep™ 16/60 Sephacryl™ S300 HR column, grey box) and cation exchange (MonoS™ 4.6/100 PE column, yellow box). Ladder sizes are indicated in kDa.

Expressing XPD as an independent subunit didn't offer the results we hoped to obtain either. When our traditional approach (cloning, expression test in *Sf9* cells followed by a BioSprint purification test) once more failed to yield pure XPD (figure 71 (A)), we tried to extract and purify the helicase in an anaerobic atmosphere to prevent the oxidation of the 4FeS cluster, using both Ni-NTA beads (figure 71 (B)) and Protein G beads (figure 71 (C)), following a method adapted from our regular BioSprint test (see chapter 2, section 2.3.1.1).

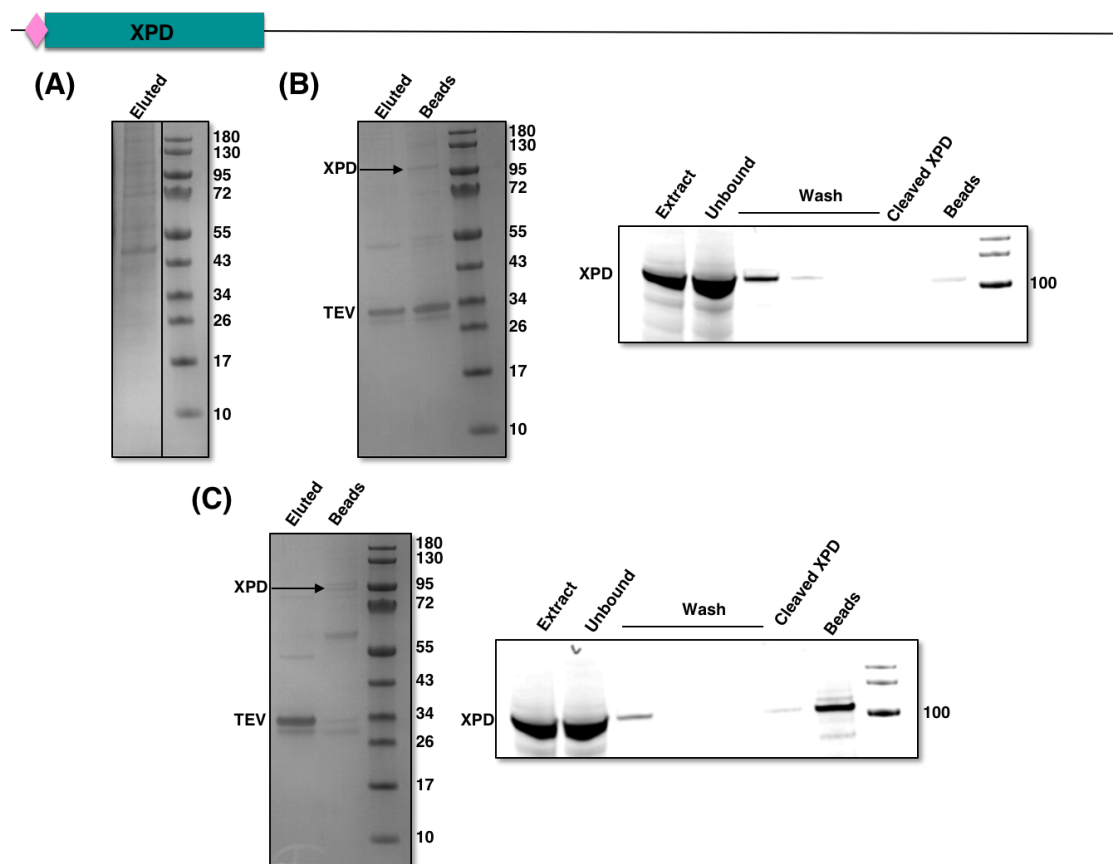


Figure 71: Purification test and anaerobic purification of the 8xHis-V5-spacer-TEV-tagged XPD helicase.

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIF COMPLEX

(A) Coomassie-stained gel showing a BioSprint expression test for subunit XPD carrying an 8xHis-V5-spacer-TEV tag at its N-terminal end. Purification of the helicase in anaerobic conditions using Ni-NTA beads (B) and protein G beads (C). All three methods failed to produce any pure XPD. Ladder sizes are indicated in kDa.

While the Western blot performed after the anaerobic purification of XPD using Ni-NTA beads (figure 71, (B), right panel) demonstrated that the enzyme could only be found in the unbound fraction, the Western blot for the protein G beads (figure 71, (C), right panel) showed that while a small part of XPD could bind to the beads, most of it could not be cleaved by TEV, and the majority of the helicase remained in the unbound fraction as observed in the attempted purification with the Ni-NTA beads.

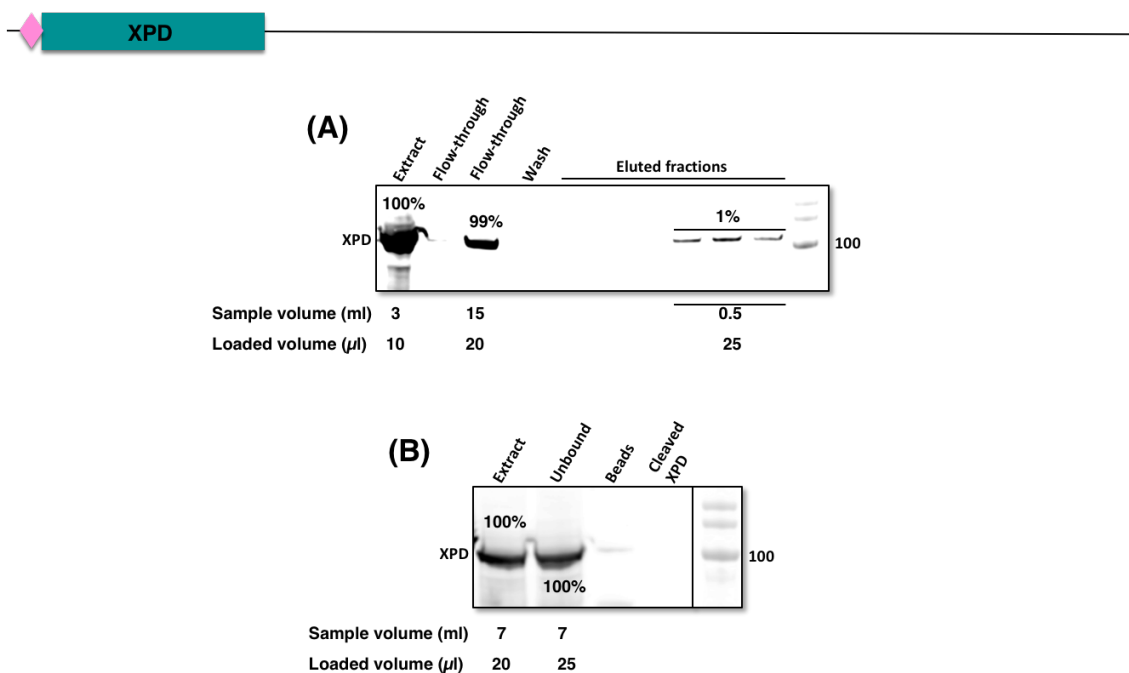


Figure 72: IMAC and IP purification of 8xHis-V5-spacer-TEV -tagged XPD helicase.

(A) Western blot showing the IMAC purification of construct XPD tagged with 8xHis-V5-spacer-TEV at its terminal end. The extraction buffer contained 5 mM ANAPOE C₁₂E₈ detergent, which was diluted to 1 mM in the subsequent buffers employed in the purification process. (B) Purification of XPD by means of IP also failed to produce pure XPD as the helicase was unable to bind the protein G beads, as shown in the Western blot.

We tried to improve the amount of XPD obtained by extracting the protein with a buffer containing 5 mM ANAPOE C₁₂E₈ detergent, as reported in (Pullara *et al.*, 2013), following the procedure described in chapter 2, section 2.3.3.2. After extraction the sample was loaded into a 1 ml HisTrap™ FF column, and fractions of interest were analysed by SDS-PAGE and Western blotting. The expression levels were good, and extraction had indeed improved compared to previous experiments, but binding to the column remained very weak (figure 72 (A)). Finally, an attempt to obtain the helicase by IP completely failed to pull down any protein at all because of poor binding to the protein G beads (figure 72 (B)).

4.2.1.4. Tagging the C-terminal end of XPD

The poor results obtained for the many different methods used to try to purify helicase XPD forced us to once again reconsider our approach to obtain this subunit. Since none of the different tags attached to the N-terminal end of the helicase had made a difference regarding its recovery, we pondered if using a C-terminal tag instead of an N-terminal one might improve our results. To test this, the *XPD* gene was first cloned into plasmid pACEBac2 using enzymes *NcoI* and *KpnI*, then a sequence encoding the TEV-10xHis-V5 C-terminal tag was cloned into the recombinant plasmid using enzymes *Scal* and *KpnI*. The P0 viral stock obtained from the transfection of the recombinant bacmid was used to infect a 25 ml *Sf9* suspension culture, and the pellet harvested was analysed in a BioSprint station as previously described (figure 73 (A)).

We tried to improve the extraction step, which seemed to be even more inefficient than that observed in previous experiments, by performing a BioSprint using different lysis buffers with different detergents as suggested in the 2013 Levine paper (Pullara *et al.*, 2013) (figure 73 (B)). A Western blot performed with samples from the extract and eluted fractions showed a worrying degradation of XPD that we had not seen before.

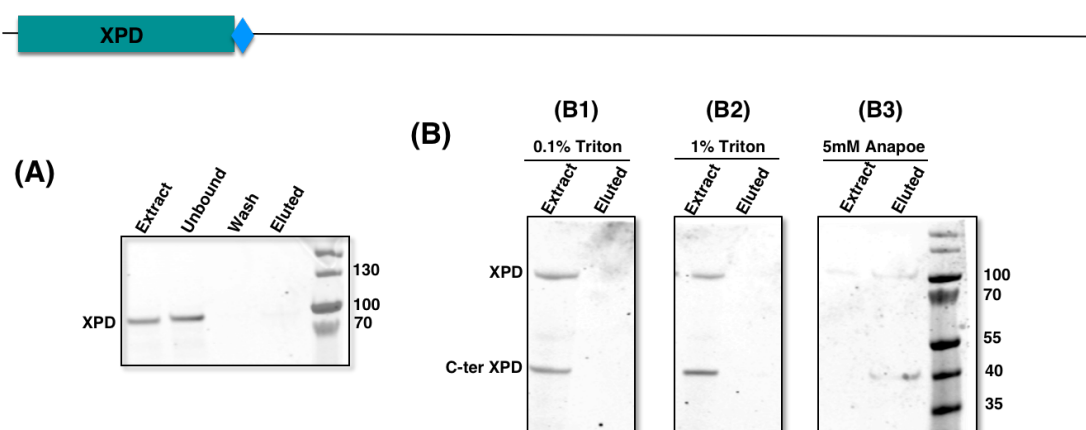


Figure 73: Purification tests of 8xHis-V5-spacer-TEV-tagged XPD helicase using different detergents.

(A) Western blot showing a BioSprint purification test for helicase XPD tagged with TEV-10xHis-V5 at its C-terminal end. The new tag fails to bind to the magnetic Ni-NTA beads, as evidenced by the fact that roughly the same amount of protein is found both in the extract and the unbound fractions. (B) Western blot showing BioSprint tests performed after extraction of XPD using three different cell lysis buffers: PBS, 300 mM NaCl, 0.1% Triton (B1), PBS, 300 mM NaCl, 1% Triton (B2), and 50 mM HEPES pH 7.5, 300 mM potassium acetate, 2 mM BME, 5 mM ANAPOE C₁₂E₈ (B3), all of them containing DNaseI and a protease inhibitor cocktail. All three of them show the same degradation of XPD, and extraction seems to be less efficient rather than improving with the new lysis buffers. The poor signal obtained for the ANAPOE extracted fraction can probably be attributed to the fact that although extraction is performed with 5 mM ANAPOE, the sample must be diluted to 1 mM to allow binding to the Ni-NTA beads. Ladder sizes are indicated in kDa.

4.2.1.5. Twin-Strep tag®: the solution to our problem

The use of a Strep-Tactin®XT Superflow® resin to purify proteins carrying a matching Twin-Strep-tag® is currently very extensive due to a reported affinity in the pM range. As one of our most pressing problems was the low binding of the XPD subunit to any of the means we employed in the purification of the TFIIF Core sub-complex, the cloning of a Twin-Strep-tag® into XPD seemed like a very promising, potential solution to our difficulties in purifying the helicase. The new tag was cloned at the C-terminal end of XPD using enzymes *Scal* and *KpnI*, then the Strep-tagged XPD was cloned into a pACEBac2 plasmid already containing subunits XPB, p62, p52, p44, p34 and p8 as previously described. This new construct was cloned, expressed and purified by Mrs.

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

Biljana Petrovic-Stojanovska, following the method previously described.

The combined use of the Twin-Strep-tag® cloned into XPD and the Strep-Tactin®XT Superflow® resin as a third recovery step (figure 74, orange box) in our purification protocol finally allowed us to obtain helicase XPD in similar amounts as TFIIH Core subunits XPB, p62, p52, p44, p34 and p8 (confirmed by MS), and with a yield very similar to that obtained previously with our simplified two-step procedure, despite the addition of a third purification step.

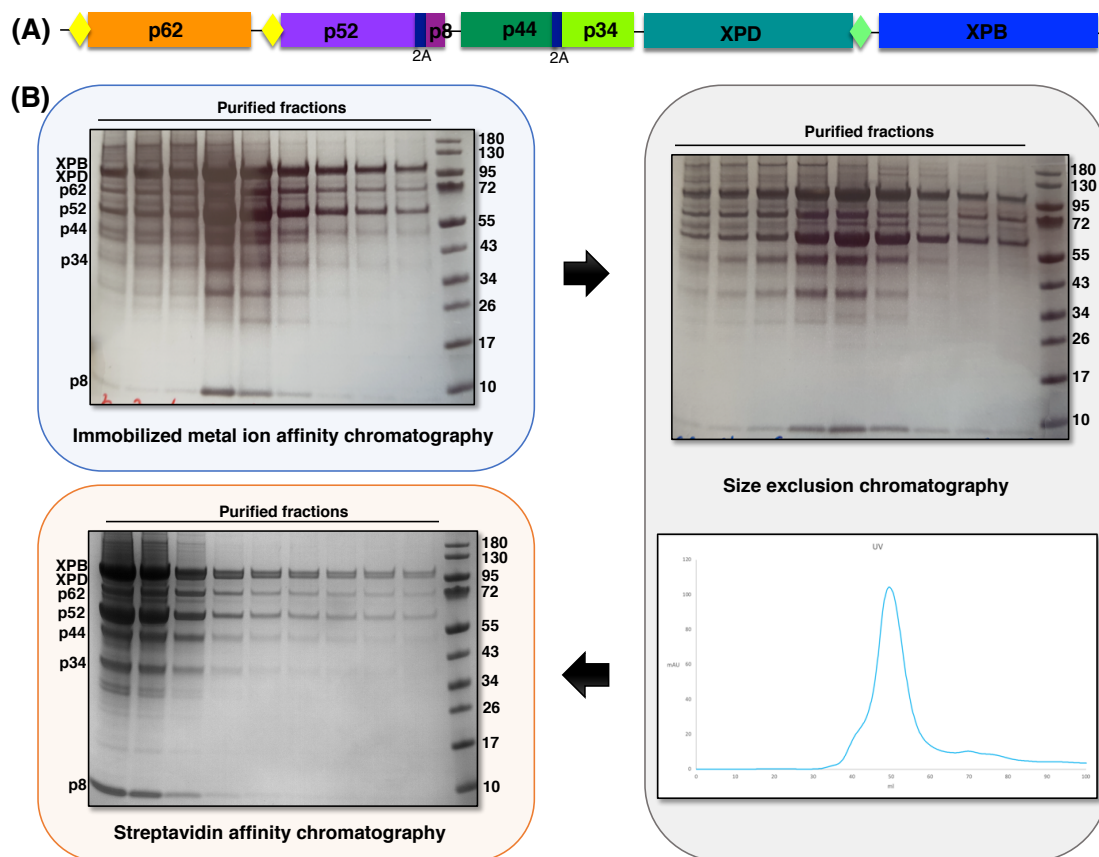


Figure 74: Purification of TFIIH Core with 6xHis-V5-TEV-tagged p62 and p52 and helicase XPD carrying a Twin-Strep-tag® at its C-terminal end.

Cartoon view of construct TFIIH Core with subunits p52 and p62 carrying an N-terminal 6xHis-V5-TEV tag (yellow diamond) and helicase XPD carrying a C-terminal Twin-Strep-tag® (green diamond) (A) and Coomassie-stained gels showing the purification process for construct TFIIH Core with subunits p52 and p62 tagged at their N-terminal end with 6xHis-V5-TEV and subunit XPD carrying a Twin-Strep-tag® at its C-terminal end (B). The protocol includes now three steps:

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIF COMPLEX

ion affinity (Talon® Superflow™ resin, blue box), size exclusion (HiPrep™ 16/60 Superdex™ S200 prep grade column, grey box) and Streptavidin affinity (Strep-Tactin®XT Superflow® resin, orange box). Ladder sizes are indicated in kDa. Subunit XPD is clearly visible in a Coomassie-stained gel together with the other six TFIIF Core subunits.

4.2.2. 10-Subunit TFIIF complex

4.2.2.1. Obtaining the CAK sub-complex

Cloning of the genes encoding the three subunits integrating the CAK sub-complex into our TFIIF Core construct was approached as a way to stabilise XPD and facilitate its purification at a time when we had yet to attempt the cloning of the Twin-Strep-tag® into the helicase.

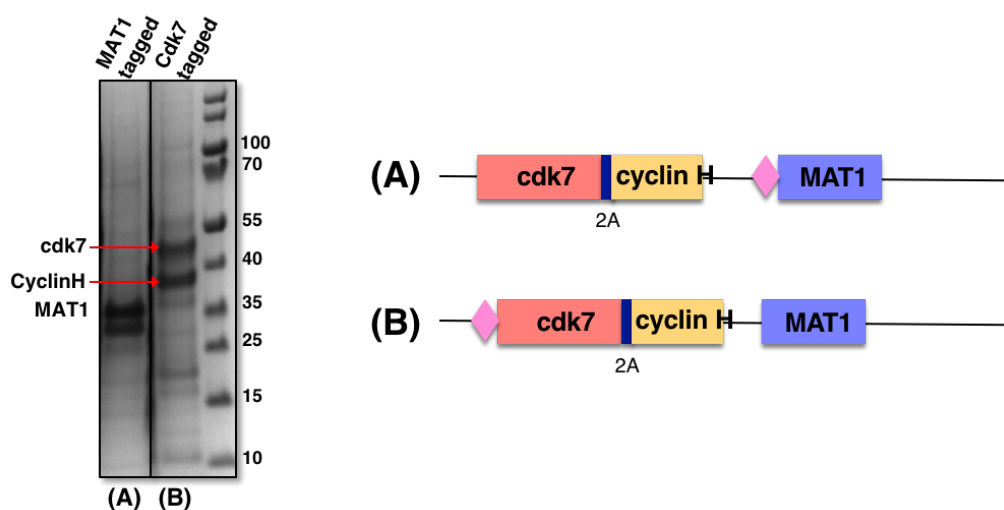


Figure 75: Purification test of the CAK sub-complex with 8xHis-V5-spacer-TEV-tagged MAT1, cdk7.

Cartoon view and SDS-PAGE gels showing BioSprint expression tests for two different CAK constructs: one with subunit MAT1 tagged at N-terminal with 8xHis-V5-spacer-TEV (A) and another one with the same tag attached to the N-terminal end of subunit cdk7 (B). Both gels show that the tag is only able to pull down the subunit it is attached to. Ladder sizes are indicated in kDa.

Synthetic genes encoding the three CAK subunits MAT1, cdk7 and cyclin H were ordered, with these last two designed as a single unit, joined by a 2A-like auto-cleavable

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIF COMPLEX

sequence as previously described for subunits p52-F2A-p8 and p44-T2A-p34. Two different CAK constructs (one with subunit MAT1 carrying an 8xHis-V5-spacer-TEV tag at its N-terminal end (figure 75 (A)) and another one with subunit cdk7 carrying the same tag at its N-terminal end (figure 75 (B))) were cloned as previously described for the TFIIF Core constructs. The two different three-gene plasmids were transposed into the baculoviral genome, which was subsequently purified and transfected into *Sf9* cells and the P0 virus obtained from this transfection was used to infect a 25 ml *Sf9* suspension that was later processed and tested in a BioSprint station as previously described. The eluted fractions were analysed by SDS-PAGE, showing in both cases that one tag is insufficient to pull down the untagged subunits present in the sub-complex. This result hints to a weak interaction between the three CAK subunits.

To solve this problem we cloned, following the same method, a third CAK construct in which both the MAT1 and cdk7 subunits were tagged at their N-terminal end with 8xHis-V5-spacer-TEV. The construct was transposed, transfected, and finally a 25 ml *Sf9* suspension infected with this P0 virus was tested for expression as explained before. This time all three subunits were pulled down, and expression was robust for all of them (figure 76).

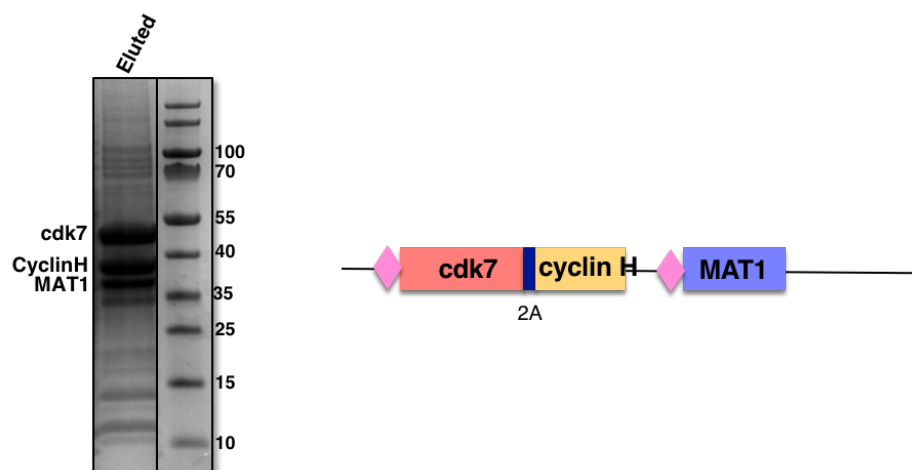


Figure 76: Purification test of the CAK sub-complex with 8xHis-V5-spacer-TEV-tagged MAT1 (A) and 8xHis-V5-spacer-TEV-tagged cdk7 (B).

Cartoon view and SDS-PAGE gel showing a BioSprint expression test for a third CAK construct in which both the MAT1 and cdk7 subunits were tagged at their N-terminal end with 8xHis-V5-spacer-TEV. All three subunits were successfully pulled down this time. Ladder sizes are indicated in kDa.

4.2.2.2. Co-infecting to obtain the 10-subunit TFIIH complex

Cloning of the CAK sub-complex-encoding sequences into our plasmid already carrying the seven Core genes was attempted, but when we didn't get immediate positive results we opted for the more straight-forward alternative of co-infecting a *Sf9* suspension culture using two different viral stocks: one carrying the TFIIH Core sub-complex and another one carrying the CAK sub-complex.

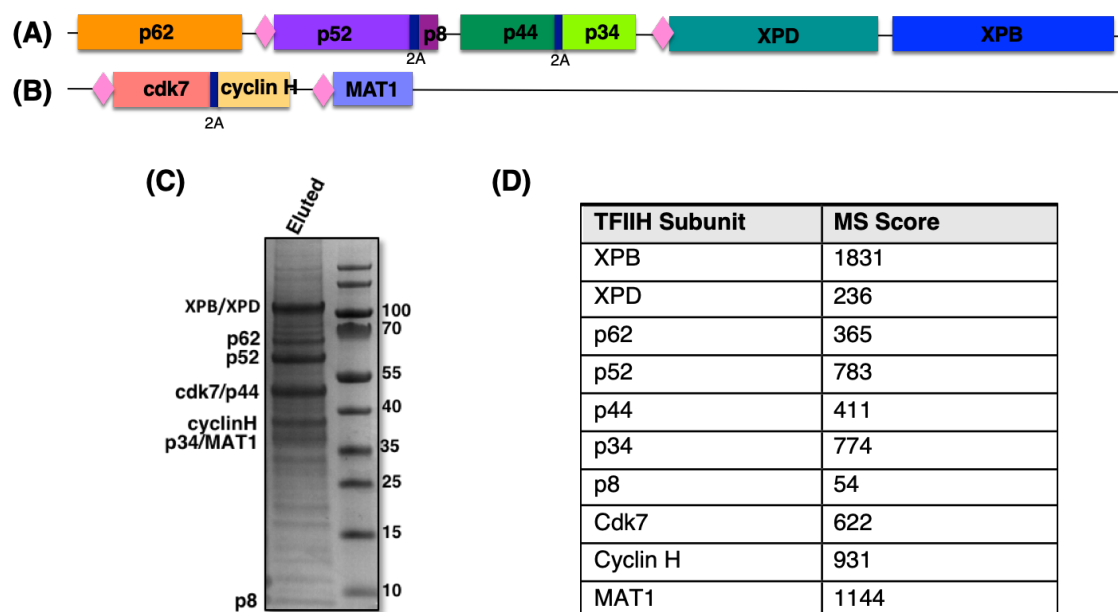


Figure 77: Purification test of the 10-subunit TFIIH complex.

A *Sf9* cell suspension culture co-infected with two different viral stocks (one obtained after infection with the TFIIH Core construct (A) and another obtained with the CAK construct (B), both shown here in cartoon view) was tested in a BioSprint station for 10-subunit TFIIH complex expression. Fractions of interest were analysed in a SDS-PAGE gel (C) which showed clear bands for subunits XPB, p62, p52 and p8, and mixed bands for subunits cdk7/p44, p34/MAT1, and cyclin H, as these subunits are very close in size (41/44 kDa, 34/32 kDa and 38 kDa, respectively). A MS analysis of a liquid sample further confirmed the presence of all 10 TFIIH subunits (D). Ladder sizes are indicated in kDa.

A 25 ml *Sf9* suspension was co-infected as described in chapter 2, section 2.2.4, with constructs TFIIH Core with subunits p52 and XPD tagged at their N-terminal end with 8xHis-V5-spacer-TEV (figure 77 (A)) and CAK with subunits MAT1 and cdk7 carrying

the same tag at their N-terminal end (figure 77 (B)). After the appropriate incubation time, the pellet was harvested and processed as previously described for our TFIIH Core constructs, and the cleared supernatant was tested for complex expression in a BioSprint station. The eluted fraction was subsequently analysed in a SDS-PAGE gel (figure 77 (C)). Subunits XPB, p62 and p52 were obvious in the gel, and a second band could be seen right under XPB, presumably corresponding to XPD. The bands for subunits p44, p34, MAT1, cdk7 and cyclin H were difficult to see individually, as all five subunits have similar sizes, all in a narrow range (34 to 44 kDa). A sample of the eluted fraction was analysed by MS, confirming the presence of all ten subunits (figure 77 (D)).

4.2.2.3. Purification of 10-subunit TFIIH

After the encouraging results obtained for the expression test, we scaled-up our infections and tried to pull-down the 10-subunit TFIIH complex following our updated IMAC – size exclusion – cation exchange chromatography purification protocol, in this case using a Talon® Superflow™ resin for the IMAC step (figure 78, blue box). Fractions containing the 10-subunit TFIIH complex were concentrated and diluted to a 1:3 ratio with buffer 20 mM HEPES pH 7, 1 mM DTT, 10% glycerol before being loaded into a MonoS™ 4.6/100 PE column (figure 78, yellow box), removing the size exclusion chromatography step due to the lower expression of the complex obtained in this infection.

The chromatogram obtained for this purification showed six different peaks, which hinted to a probable break down of the complex into smaller sub-complexes as the purification progresses. This was confirmed by the SDS-PAGE analysis of selected fractions taken from all six peaks, with peak 5 containing the 7-subunit TFIIH Core complex we initially aimed to purify. All fractions of interest from peak 5 were subsequently concentrated and a sample was analysed by SDS-PAGE again, revealing a band directly under the one corresponding to the XPB subunit which was identified as a mixture of both XPB and XPD by MS analysis. This represented a significant improvement over our previous attempts to purify XPD, and definitely was our best result in terms of extraction of the helicase up to that date. Despite this improvement, very little XPD was actually recovered, and the pattern of peaks obtained could not be reproduced in subsequent purifications, so we focused our efforts in other alternatives to finally obtain all seven

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

TFIIH Core subunits stoichiometrically.

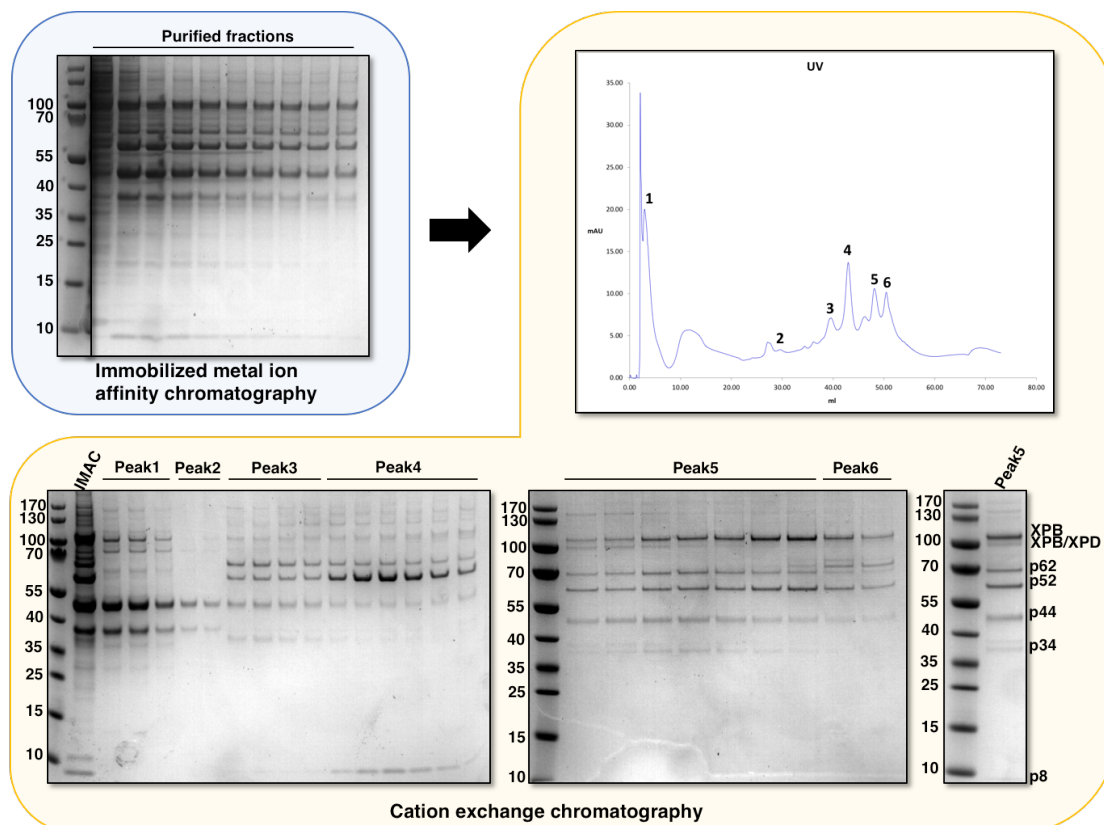


Figure 78: Purification of the 10-subunit TFIIH complex.

Coomassie-stained gels showing the different chromatographic steps followed to purify our 10-subunit TFIIH complex. The protocol was simplified to two steps: ion affinity (Talon® Superflow™ resin, blue box), and cation exchange (MonoS™ 4.6/100 PE column, yellow box). Ladder sizes are indicated in kDa. The six peaks obtained during cation exchange were analysed in a SDS-PAGE gel, confirming that the 10-subunit complex breaks down into several different sub-complexes along its purification. Fractions included in peak 5 were concentrated and further analysed in a gel that showed the 7-subunit TFIIH Core sub-complex, with the band immediately under the XPB subunit identified as a mixture of XPB and XPD by MS.

4.2.3. Cloning and expression of other TFIIH constructs

4.2.3.1. TCP-Tagged TFIIH Core

4.2.3.1.1. Cloning and expression tests

A CCGPCC tetracysteine tag encoding sequence was cloned into TFIIH Core subunit XPB, as we had previously done with damage-detector protein XPC to aid in our fluorescence anisotropy and bulk FRET assays. To build up the seven-gene TFIIH Core construct, the sequence encoding the TCP-tagged XPB was cloned into a pACEBac2 plasmid already containing the *XPD* gene as previously described. The TCP-tagged *XPB/XPD* cassette was subsequently cloned into another pACEBac2 plasmid containing the genes encoding subunits p62, p52, p44, p34 and p8, following the same protocol. Subunits p62 and p52 had previously been tagged at N-terminal with 6xHis-V5-TEV. The TCP tag had previously been confirmed by sequencing, and the seven-gene construct was verified by restriction with enzymes *KpnI*, *BamHI/HindIII* and *Sall/XhoI*.

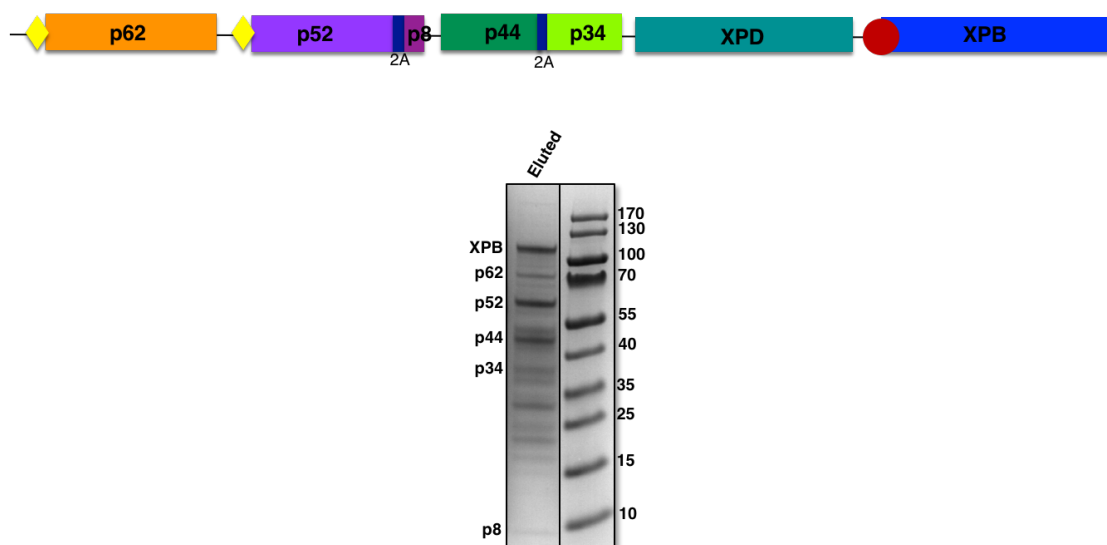


Figure 79: Purification test of the TCP-tagged TFIIH Core sub-complex.

SDS-PAGE gel showing a BioSprint expression test for construct TFIIH Core with subunits p52 and p62 tagged at their N-terminal end with 6xHis-V5-TEV (yellow diamond) and subunit XPB tagged with a CCGPCC tetracysteine motif at its N-terminal end, too (red sphere). The presence of the TCP tag had previously been confirmed by sequencing. Ladder sizes are indicated in kDa.

The TCP-tagged, seven-gene plasmid was transposed into the baculoviral genome, and

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

afterwards the purified bacmid was transfected into a monolayer of *Sf9* cells. An expression test was carried out in a 25 ml *Sf9* suspension culture infected with 3 ml of a P0 viral stock and expression levels were analyzed in a BioSprint purification test as previously described. The eluted fraction was analysed in a SDS-PAGE gel, and as with other TFIIH constructs, the gel showed a good expression for subunits XPB, p62, p52, p44, p34 and p8, with helicase XPD expressed in a much lower concentration, hence not showing in the gel (figure 79).

4.2.3.1.2. Purification

The next step involved the scaling-up of the infected *Sf9* suspension cultures and purification of the TCP-tagged TFIIH Core sub-complex in two steps: IMAC utilising a Talon® Superflow™ resin (figure 80, blue box) and a cation exchange chromatography through a MonoS™ 4.6/100 PE column (figure 80, yellow box). Fractions eluted after IMAC were diluted 1:3 with buffer 20 mM HEPES pH 7, 1 mM DTT, 10% glycerol before loading them into the next chromatographic step.

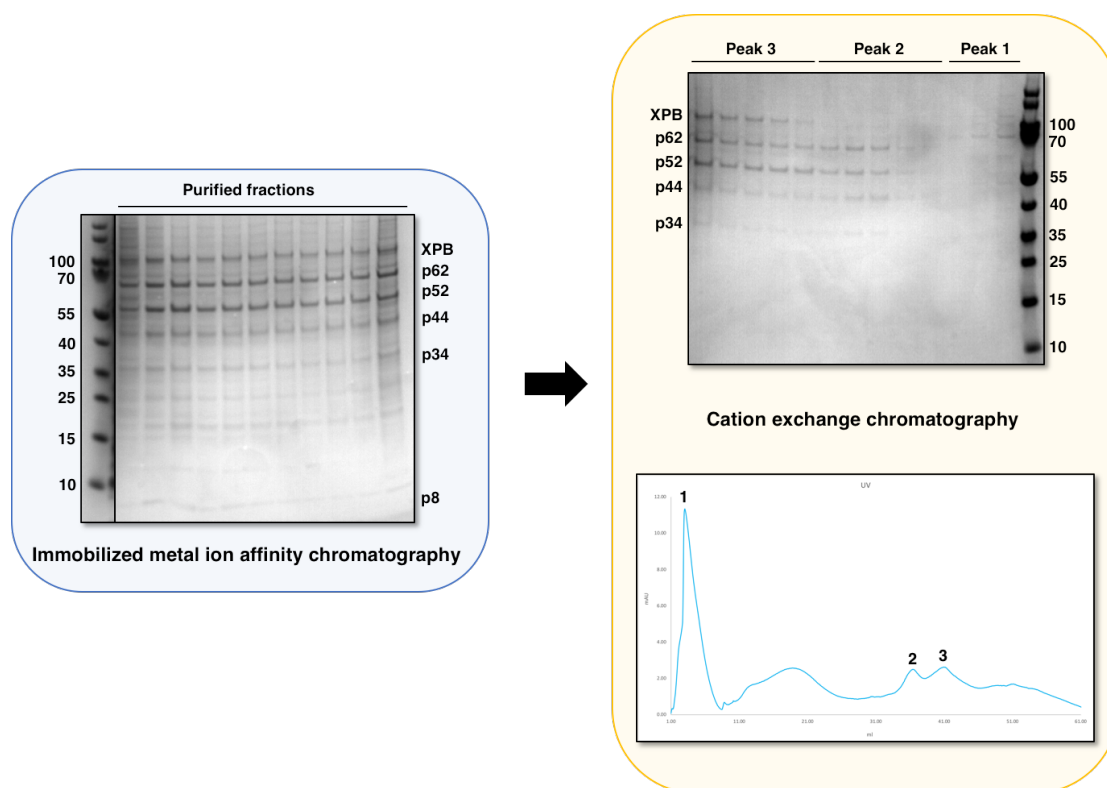


Figure 80: Purification of the TCP-tagged TFIIH Core sub-complex.

Coomassie-stained gels showing the two chromatographic steps followed to purify our TCP-tagged TFIIF Core sub-complex: IMAC (Talon® Superflow™ resin, blue box), and cation exchange (MonoS™ 4.6/100 PE column, yellow box). Ladder sizes indicated in kDa. Fractions eluted in peak 3 contained similar amounts of XPB, p62, p52, p44, p34 and p8, and residual amounts of XPD (not seen in the gel). Surprisingly, fractions eluted in peak 2 also contained the Core sub-complex, but TCP-tagged subunit XPB was barely present here.

4.2.3.1.3. ReAsH-EDT₂ Labelling of TCP-tagged TFIIF Core

The fractions collected in peak 3 were pooled together and concentrated, and TCP-tagged TFIIF Core was labelled with the biarsenic ligand ReAsH-EDT₂ as described in chapter 2, section 2.5.1.

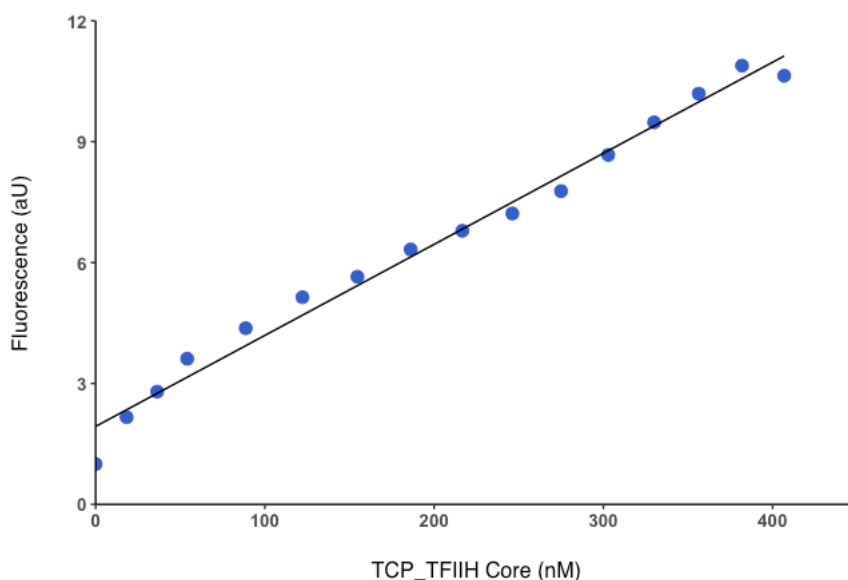


Figure 81: ReAsH labelling of construct TFIIF Core.

Subunit XPB carries the tetracysteine motif at its N-terminal end. The ligand was excited at a wavelength of 593 nm and emitted fluorescence was collected at a wavelength of 608 nm.

No TCP-TFIIF Core sub-complex was added to the reaction corresponding to the first measured point shown in the graph, acting as a control to determine the intensity corresponding to the free ReAsH-EDT₂ reagent. An increment in the intensity of the emitted fluorescence, collected at 608 nm, was observed as the concentration of TCP-TFIIF Core increases (figure 81). This increment was expected as a result of the binding

of the ReAsH-EDT₂ ligand to the tetracysteine tag cloned into subunit XPB, confirming that our Core sub-complex can be labelled successfully.

4.3. Discussion

4.3.1. Cloning and expressing the TFIIH Core sub-complex

The TFIIH complex not only carries out different functions in transcription and in DNA repair: its structure itself is different in the two processes. All ten subunits (holo-TFIIH) participate in the role of the complex in transcription, but only the Core sub-complex is required for NER (Svejstrup *et al.*, 1995). The CAK sub-complex has even been reported to inhibit the NER process by down-regulating the helicase activity of XPD (Sandrock & Egly, 2001), and so it is removed after the holo-complex binds to the damaged site in a reaction reportedly promoted by XPA (Coin *et al.*, 2008). Taking this into account, and as our interest regarding the TFIIH complex focuses on its role as a DNA repair factor, we initially attempted the cloning of the TFIIH Core sub-complex on its own, with subunits XPD and p52 carrying a 6xHis-V5-TEV tag at their N-terminal end.

We obtained a seven-gene construct containing the sequences encoding subunits XPD, XPB, p62, p52, p44, p34 and p8 by cloning every gene, one at a time, into a single plasmid using the multiplication module present in the vectors comprising the MultiBac™ system (Bieniossek *et al.*, 2013), a BEV for the cloning and expression of multi-protein complexes (Berger *et al.*, 2004). To facilitate the cloning of the sub-complex the genes for subunits p52 and p8, and p44 and p34, were designed as a single unit, with each pair of genes separated by a 2A-like auto-cleavable peptide (Luke *et al.*, 2009), meaning that only five cloning reactions were needed to obtain the seven-gene construct. Unfortunately, expression levels in Sf9 cells for this construct were quite disappointing (figure 57), and after optimisation of the transfection and infection processes failed to improve our results a change of strategy concerning the position of the purification tags was approached to try to solve this problem. The Western blot on figure 57 also shows a remarkable cleaving efficiency for the 2A-like peptides, with almost 90% of the p52-F2A-p8 gene actively translated into independent subunits p52 and p8.

As several multi-gene intermediates had already been obtained in the cloning of the

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

previous TFIIH construct, the characteristics of the MultiBac™ plasmids (Bieniossek *et al.*, 2012) allowed the construction of a new seven-gene TFIIH Core sub-complex with subunits p52 and p62 carrying a 6xHis-V5-TEV tag in only three cloning steps. Analysis of the fractions eluted in an expression test for the new construct in a SDS-PAGE gel clearly showed subunits XPB, p62, p52, p44, p34 and p8 in stoichiometric amounts, but no XPD (figure 60 (A)). Furthermore, an expression test performed in anaerobic conditions also failed to improve the extraction and purification of XPD (figure 60 (B)).

Although disappointing, this result was not very surprising, as it was already hinted in our previous purifications of the sub-complex that we might be losing XPD during the process. Our observations were reinforced by the knowledge that XPD is not only the most flexible subunit within the TFIIH Core sub-complex, being able to bind to both the Core and CAK sub-complexes (Greber *et al.*, 2017), but the presence of an 4FeS cluster in its structure (Rudolf *et al.*, 2006) also affects the purification of the sub-complex enormously.

Scaling-up and optimisation of the protocol for the purification of the TFIIH Core sub-complex in aerobic conditions finally resulted in an easily reproducible procedure that included an IMAC step, followed by size-exclusion chromatography and finally a cation exchange chromatography (figure 61). This protocol yielded up to 1 mg of pure sub-complex (containing subunits XPB, p62, p52, p44, p34 and p8, with most of subunit XPD lost along the process as confirmed by the Western blot in figure 62) for every 2 L culture. TFIIH Core yield was further improved by substituting the HisTrap™ FF column in the IMAC step for a Talon® Superflow™ resin that allowed longer binding times, and the removal of the cation exchange chromatography, which translated into a significantly reduced loss of the complex (figure 63).

The presence of a reduced amount of XPD in our purified sample was further confirmed by crosslinking assays. A total of 25 interlinks (nine of which had not been previously reported) and 10 intralinks (table 5) were obtained, depicting a compact structure profusely connected. The publication of an extensive structural study based on crosslinking (Luo *et al.*, 2015) around the time we were carrying out these assays led us to pursue a different line of research, focusing on a biochemical approach rather than a

structural study of the sub-complex. However, after the recent publication of the cryoEM structure of the Core sub-complex plus CAK subunit MAT1 (Greber *et al.*, 2017) we revisited our data to check how our unreported crosslinks fit within the structure. Unfortunately, all of our crosslinks seem to fall on unstructured, poorly defined zones in the published structure, so no information could be extracted from this comparison other than a further confirmation of TFIIH's flexibility. Although the publication of the cryoEM structure obtained by the Nogales group meant a considerable advance in our knowledge of TFIIH, our observations emphasise the need for improved structural information on the complex.

4.3.2. The elusive XPD helicase

IP experiments revealed a weakness in the design of our 6xHis-V5-TEV tag: when antibody V5 was bound to its epitope in the tag, the TEV site was not accessible for cleaving with the protease. To solve this problem, a new tag with a 9-residue spacer between the V5 and TEV sites (8xHis-V5-spacer-TEV) was designed and cloned into three different TFIIH constructs: a 7-subunit Core with subunits p52 and XPD tagged at their N-terminal end, a 6-subunit (XPB, p62, p52, p44, p34 and p8) Core with only subunit p52 carrying the tag, and a tagged XPD subunit cloned on its own. By cloning these three constructs we hoped to solve our ongoing problem of not recovering enough XPD in our purification, either by introducing a stronger tag that would be able to pull down all seven subunits in the Core sub-complex, or by expressing problematic XPD on its own and later adding it to a purified sample containing the other six Core subunits.

Expression and purification of the new 8xHis-V5-spacer-TEV-tagged 7-subunit Core construct revealed no noticeable improvement regarding XPD (figures 67 and 68), and its yield for the remaining six subunits was never as good as levels offered by its 6xHis-V5-TEV-tagged counterpart. Expression and purification of the 6-subunit sub-complex showed that the Core subunits were not expressed stoichiometrically for this construct (figure 69), with only subunits XPB and p52 remaining in our purified sample after the cation exchange chromatography. This could point to subunits XPB and p52, and possibly p8 (as seen in the SDS-PAGE gel corresponding to the size exclusion chromatography in figure 70, grey box) forming a smaller sub-complex within TFIIH Core, supported by the tight relationship between the three subunits, as p52 and p8 act as

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIF COMPLEX

regulators of the helicase (Coin *et al.*, 2006) (Coin *et al.*, 2007) (Luo *et al.*, 2015).

Different approaches to try to obtain helicase XPD as an independent subunit didn't offered the expected results either. We initially tried our previous approach of expression in *Sf9* cells followed by an aerobic BioSprint purification test (figure 71 (A)). When this didn't work, an anaerobic purification using both Ni-NTA beads (figure 71 (B)) and protein G beads (figure 71 (C)) was attempted, but while the helicase failed to bind to the Ni-NTA beads, most the fraction bound to the protein G beads could not be cleaved by TEV, suggesting that the spacer introduced in the tag was insufficient to make the cleaving site available to the protease.

A final attempt to try to improve the amount of XPD obtained was made by adding 5 mM ANAPOE C₁₂E₈ detergent to the extraction buffer, and 1 mM detergent to our purification buffers A and B, as reported in (Pullara *et al.*, 2013). Although the addition of the detergent somewhat improved the amount of extracted XPD compared to previous experiments, it did not enhance the weak binding of the helicase to the column (figure 72 (A)). Trying to obtain the helicase by IP didn't work either, as the protein completely failed to bind to the protein G beads (figure 72 (B)). This might be due to the presence of 1 mM ANAPOE C₁₂E₈ detergent in our buffers, as a previous IP also performed with protein G beads showed that XPD did bind to the beads, and it was actually unable to be cleaved from them (figure 71 (C)).

As none of the different methods approached to obtain XPD offered an improvement on the recovery of the helicase, we pondered if the N-terminal position of the two different tags used in our constructs might be the cause behind these poor results. To test this hypothesis, a sequence encoding a new TEV-10xHis-V5 C-terminal tag was cloned into the helicase, and purification of XPD in a BioSprint test was attempted, adding a variety of detergents to our lysis buffer as a means to improve the extraction of the protein (Pullara *et al.*, 2013). Surprisingly, the results obtained with the C-terminal tag were even worse than those seen for the N-terminal tags. A Western blot performed with fractions obtained after purification of the protein with three different buffers showed that most of XPD was cleaved in two, with one of the fragments showing roughly the size expected for C-terminal XPD, independently of the buffer used for cell lysis (figure 73 (B)).

The optimised purification process we had developed so far meant that we were able to obtain pure TFIIH Core sub-complex in a higher yield than any purification reported until the publication of the Yang paper in 2015 (Li *et al.*, 2015), but the loss of the majority of the XPD helicase present in our sample meant that our sub-complex was not suitable for a biochemical characterization of its unwinding activity. The presence of two His tags in subunits p52 and p62 was sufficient to pull down all six Core subunits (XPB, p62, p52, p44, p34 and p8), but the recovery of XPD remained a challenge for us. Finally, and although previous results seemed to discourage the use of a C-terminal tag on XPD, we were able to obtain the elusive protein by cloning a Twin-Strep-tag® precisely at the C-terminal end of the helicase.

The use of the Strep-Tactin®XT Superflow® resin, which reportedly has an affinity in the pM range for the matching Twin-Strep-tag®-encoding sequence cloned into XPD, proved to be the solution to the low binding affinity the helicase had shown in all of our previous purification attempts. With construct TFIIH Core with subunits p52 and p62 carrying a 6xHis-V5-TEV N-terminal tag and subunit XPD carrying a C-terminal Twin-Strep-tag® we were finally able to obtain all seven Core subunits in stoichiometric amounts (figure 74). Moreover, the introduction of a third chromatographic step only had a minimal impact in our complex yield compared to our previous two-step protocol. The seven-subunit TFIIH Core sub-complex obtained in this manner has been tested for DNA binding and helicase activity and has been proved to be active in a variety of assays described in chapter 5.

4.3.3. Cloning and expressing the 10-subunit TFIIH complex

Obtaining helicase XPD in a similar amount as the other Core subunits when purifying the sub-complex was for a long time the hardest challenge of this PhD project. As XPD binds to both the Core and the CAK sub-complexes (Schultz *et al.*, 2000) (Greber *et al.*, 2017), one of the approaches we pursued (before we introduced the Twin-Strep-tag® into our construct) was cloning the three genes encoding the CAK sub-complex into our 7-subunit construct in an effort to stabilise XPD, hopefully improving its purification this way. A BioSprint purification test showed that a single 8xHis-V5-spacer-TEV tag was not enough to pull down all three CAK subunits (figure 75), hinting to a weak interaction between subunits cdk7, cyclin H and MAT1. This problem was solved by cloning a new

CAK construct in which subunits MAT1 and cdk7 were tagged at their N-terminal end with 8xHis-V5-spacer-TEV, which showed excellent expression levels in a BioSprint purification test (figure 76).

Analysis in a SDS-PAGE gel of a purification test for a co-infection experiment with two different recombinant viruses, one carrying the TFIIH Core sub-complex and another one carrying the CAK sub-complex, showed clear bands for subunits XPB, p62 and p52, with a fainter band right under XPB that was tentatively identified as XPD. The bands for subunits p44, p34, MAT1, cdk7 and cyclin H were more difficult to differentiate, as sizes for these subunits are very close (34 to 44 kDa), but the presence of all ten TFIIH subunits was confirmed by MS analysis of a sample of the eluted fraction (figure 77). The hit list obtained for this analysis showed that expression levels for XPD were much closer to those of the other TFIIH Core subunits than anything obtained in purification attempts carried out up to that date. However, a scaled-up purification showed that the complex broke down into different sub-complexes as the purification progressed (figure 78), with one of these sub-complexes being the 7-subunit Core we aimed to purify initially. A concentrated sample of the fractions containing the Core sub-complex was further analyzed in a SDS-PAGE gel, showing a band that MS identified as a mixture of both XPB and XPD (with the sample containing a slightly higher amount of XPD), thus making this the first time we were able to see XPD in a Coomassie-stained gel (figure 78). Unfortunately, very little XPD was actually recovered, and more importantly this purification was extremely difficult to reproduce, which convinced us to go back to our original plan of cloning the TFIIH Core construct on its own, eventually achieving our goal with the introduction of a Twin-Strep-tag® at the C-terminal end of the helicase.

4.3.4. Cloning and expression TCP-tagged TFIIH Core

A TFIIH Core construct in which subunit XPB carried a 6-residue Cys-Cys-Pro-Gly-Cys-Cys motif was cloned to help us study binding of the sub-complex to a DNA substrate in anisotropy and bulk FRET experiments. Previous purifications of the TFIIH Core sub-complex we used to introduce the TCP tag (TFIIH Core with subunits p52 and p62 tagged at the N-terminal end with 6xHis-V5-TEV and XPD carrying no tag) had showed good expression levels and stoichiometric amounts for the Core subunits, with the exception of XPD, which was obtained in a much lower amount. Since the tetracysteine motif is

only six amino acids long and unlikely to interfere with the folding or the purification of XPB, we expected the same result for our TCP-tagged construct. However, in the course of this purification we obtained not one, but two peaks containing the TFIIH Core sub-complex (figure 80), and while peak 3 contains all seven subunits as we expected, TCP-tagged XPB is noticeably missing in peak 2, suggesting that the tag might be affecting XPB expression.

The concentrated sample obtained from peak 3 was successfully labelled with a ReAsH-EDT₂ biarsenic ligand (figure 81), as a progressive increase in the intensity of fluorescence emitted at 608 nm was observed as the concentration of TCP-TFIIH Core construct grows. Intensity corresponding to the free ReAsH-EDT₂ reagent when excited with wavelength 593 nm had previously been established.

This construct was cloned in parallel with another TCP-tagged XPC construct, and as with the damage detector, no experiments have actually been performed with it yet due to time restrictions. The introduction of a Twin-Strep-tag® at the C-terminal end of XPD to increase the yield of the helicase in our TCP-tagged sample is a priority in our future plans for this construct. The presence of the TCP tag in our construct will potentially improve our anisotropy and FRET assays and will offer confirmation of our preliminary data regarding TFIIH Core binding to DNA and to a DNA-XPC complex.

4.4. Conclusions

We have cloned and expressed the ~ 500 kDa, 10-subunit TFIIH complex using the MultiBac™ system, a BEV specifically designed for the production of eukaryotic multi-protein complexes that allows the expression of every subunit in equimolar amounts (Berger *et al.*, 2004), as opposed to infection with several different viral stocks. Purification of the sub-complex was achieved after establishing a simple and efficient 3-step protocol that allows the extraction of up to 1 mg of pure sub-complex from a 2 L culture.

The cloning of the genes encoding the Core sub-complex was simplified by the introduction of 2A-like sequences (Luke *et al.*, 2009), designing the synthetic genes for

CLONING, EXPRESSION AND PURIFICATION OF THE TFIIH COMPLEX

subunits p52 and p8, p44 and p34 and cdk7 and cyclin H as a single unit, with each pair of genes separated by these autocleavable peptides – hence reducing the number of cloning steps necessary to obtain the multi-gene construct. Cleaving efficiency was shown to be close to 90%.

The purification process was optimised until a 3-step IMAC – size exclusion – streptavidin affinity chromatography was established as a standardized protocol, aided by the introduction of a Twin-Strep-tag® at the C-terminal end of helicase XPD and a 6xHis-V5-TEV tag at the N-terminal end of subunits p52 and p62. This small, versatile tag designed by our lab has a triple purpose: purification of the desired peptide through its His tail, Western blot detection by means of its V5 epitope and removal of the tag by cleaving at its TEV site.

We are currently working on the cloning and expression of the TFIIH Core sub-complex carrying helicase mutants XPD K46R (Coin *et al.*, 1998) and XPB K346R (Tirode *et al.*, 1999) (both affecting the Walker A domain) and XPD Y192A and XPD K196R (Mathieu *et al.*, 2013) (affecting the DNA binding pocket on the enzyme). We intend to study the binding (or lack thereof) of these mutants to damaged and undamaged DNA substrates in fluorescence anisotropy and bulk FRET assays, and their unwinding capacity in fluorescence-based helicase assays. These studies will further reinforce and complement our current observations of the binding of the wild type TFIIH Core sub-complex.

Future work regarding the extraction of the TFIIH Core sub-complex will focus on the expression of the sub-complex on High Five® cells to further improve the yield of our infections.

5. Studying TFIIH interactions with DNA

5.1. Introduction

5.1.1. TFIIH: Early structural knowledge

The publication of several crystal structures of archaeal homologues of the helicases XPB (Fan *et al.*, 2006) and XPD (Liu *et al.*, 2008) (Wolski *et al.*, 2008) (Fan *et al.*, 2008) and cryoEM structures of TFIIH itself (Schultz *et al.*, 2000) (Greber *et al.*, 2017) (He *et al.*, 2016) (Schilbach *et al.*, 2017) have helped us understand the roles the complex plays in transcription and NER and the mechanisms it employs to carry out both functions (both extensively discussed in chapter 1, sections 1.2.2.2 and 1.3.2).

The structure of the XPB helicase from the archaeon *Archaeoglobus fulgidus* was published by Fan and co-workers in 2006 (figures 82 and 84). A comparison between *AfXPB* and *HsXPB* showed that both of them included the two RecA like helicase domains HD1 (containing helicase motifs I, Ia, II and III) and HD2 (containing helicase motifs IV, V and VI) present in all helicases from the SF1 and SF2 families (Fan *et al.*, 2006). Motifs I, II, III and VI are known to participate in ATP-binding, while motifs Ia, IV and V participate in DNA-binding (Fairman-Williams *et al.*, 2010). They also share a series of structural elements: an area for interaction with damaged DNA (the Damage Recognition Domain (DRD), situated at N-terminal) and the R210-E211-D212 (RED) motif in HD1, suggested by Fan and co-workers to be involved in DNA unwinding in a similar manner as the β -hairpin of the UvrB helicase in bacterial NER (Theis *et al.*, 2000). At its C-terminal end, the protein presents a Thumb (ThM) domain, a flexible insertion into HD2 of approximately 60 residues that participates in binding to the DNA (Fan *et al.*, 2006). *HsXPB* presents extended C-terminal and N-terminal domains compared to *AfXPB* that are thought to carry out accessory functions for the human protein (i.e. interaction with other repair factors), as viable mutations have only been described for these regions (Weeda *et al.*, 1997) (Coin *et al.*, 2004), hinting that the core presented in the *AfXPB* is essential for XPB function and so mutations in this area are lethal (Fan *et al.*, 2006).

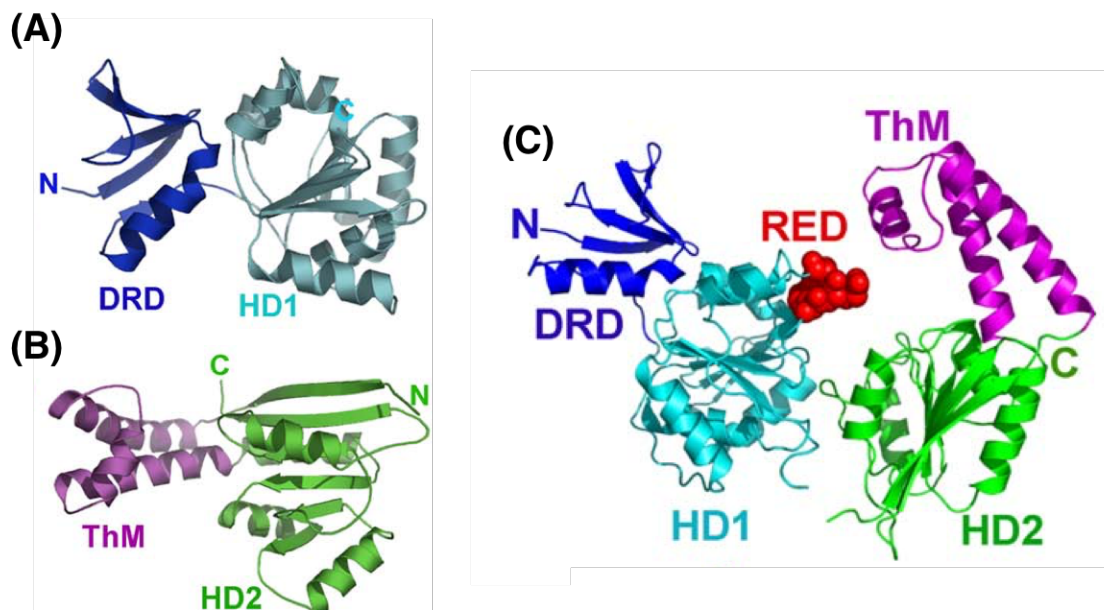


Figure 82: Structure of the XPB homologue from *A. fulgidus*.

Structure of the N-terminal domain (A), C-terminal domain (B) and full-length (C) *A*XPB, showing different orientations of the canonical helicase domains 1 and 2 and XPB's specific domains DRD, Thm and the RED motif. Adapted from (Fan *et al.*, 2006)

Three different crystal structures of the archaeal XPD were published in 2008: an XPD homologue from *Sulfolobus tokodaii* (Liu *et al.*, 2008), one from *Sulfolobus acidocaldarius* (Fan *et al.*, 2008) and a last one from *Thermoplasma acidophilum* (Wolski *et al.*, 2008) (figures 83 and 84). As with XPB, all of them showed the seven canonical helicase motifs distributed in two different helicase domains (I, Ia, II and III in HD1; IV, V and VI in HD2), with the two domains separated by an interface for ATP hydrolysis (Liu *et al.*, 2008). Liu and co-workers additionally identified motifs Ib, Va and VII, thought to interact with ssDNA (Liu *et al.*, 2008). The XPD structure also presents two specific domains as inserts in HD1: an Arch domain and an FeS domain. The Arch domain presents a novel fold and is located between HD1 motifs II and III (Fan *et al.*, 2008). Liu *et al.* proposed a role for this domain as a platform to anchor the CAK sub-complex to TFIIH Core, later confirmed by the work of the Egly and Poterszman groups (Abdulrahman *et al.*, 2013). XPD also presents a 4FeS cluster situated between HD1 motifs Ia and II (Fan *et al.*, 2008) and coordinated by four cysteine residues, three of

which are highly conserved. This 4FeS cluster has been proposed to have an essential role in the function of XPD either as a stabiliser of the overall structure of the helicase (Fan *et al.*, 2008), a structural aid to open the DNA (Rudolf *et al.*, 2006) (Liu *et al.*, 2008) or a damage-sensor component (Wolski *et al.*, 2008). The three groups coincide in proposing that these three elements (HD1, Arch and 4FeS cluster) form a channel for the translocation of DNA during unwinding.

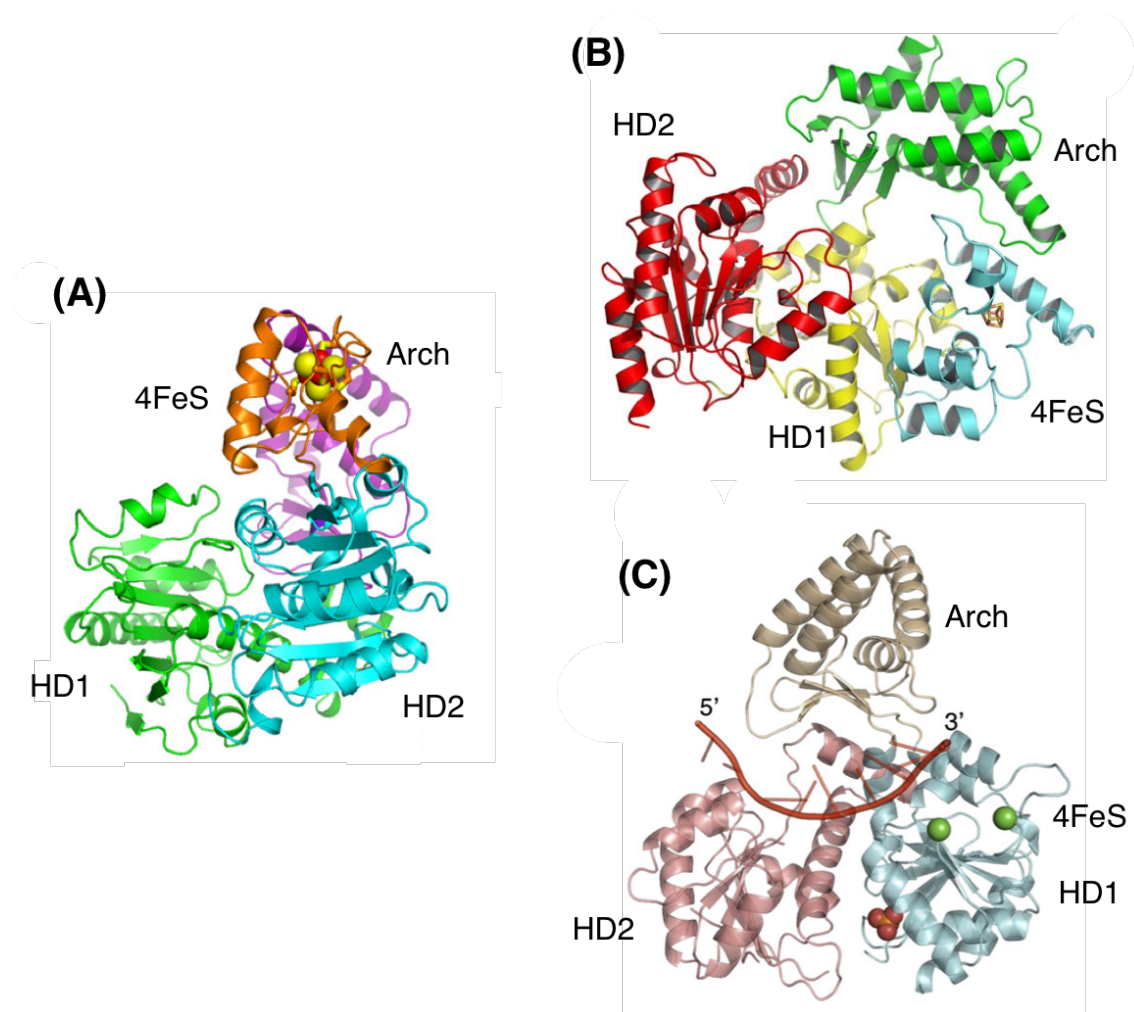


Figure 83: Crystal structure of different archaeal homologues of helicase XPD.

The structures of the XPD homologues from *S. acidocaldarius* (A), *T. acidophilum* (B) and *S. tokodaii* (C) all show the two canonical helicase domains 1 and 2, and the novel Arch domain. (A) and (B) also show the 4FeS domain, delimited in (C) by the green spheres as it was not visible in the electron density map. Figure (A) adapted from (Fan *et al.*, 2008), (B) from (Wolski *et al.*, 2008) and (C) from (Liu *et al.*, 2008).

The little information available regarding the structure of the remaining TFIIH subunits was limited to a few domains. The publication of the structure of the N-terminal domain of subunit p62 revealed a pleckstrin homology (PH)-like fold (figure 84), the first described in a transcription factor, and it was shown to be involved in the recruitment of NER factor XPG (Gervais *et al.*, 2004). Since then several publications have proved the essential role of the PH domain of p62 in establishing interactions between the TFIIH complex and a wide range of different proteins like damage-recognition factor XPC (Okuda *et al.*, 2015), transcription factor TFIIIE (Okuda *et al.*, 2008) or the Epstein-Barr virus protein EBNA2 (Chabot *et al.*, 2014).

The crystal structure of the interaction between Tfb5 and Tfb2 (the yeast orthologues of subunits p8 and p52, respectively) revealed that almost the whole of Tfb5 binds to C-terminal Tfb2, with the two regions sharing a similar fold (figure 84) (Kainov *et al.*, 2008). Deletion of the Tfb5-interacting domain in Tfb2 didn't affect the architecture of the TFIIH complex, but it led to impaired NER (Kainov *et al.*, 2008). As it is known that the absence of p8 drastically reduces the cellular levels of TFIIH (Giglia-Mari *et al.*, 2004), Kainov and co-workers proposed that Tfb5 protects TFIIH from degradation through binding to Tfb2.

The crystal structure of the N-terminal domain of protein Ssl1 (the *Saccharomyces cerevisiae* homologue of subunit p44) showed a von Willebrand factor A (VWA) fold (figure 84) with a loop essential for stimulation of Rad3, the yeast homologue of helicase XPD, as proved by the fact that a double mutant in the β 4- α 5 loop showed both significantly reduced binding to XPD and helicase activity (Kim *et al.*, 2015). A second study produced the structure of the C-terminal domain of p44, proposed to be involved in binding to subunit p34 (Kellenberger *et al.*, 2005). This structure shows a C4C4 RING domain with two zinc atoms coordinated by eight cysteine residues and the overall fold stabilised by hydrophobic residues conserved across all RING folds (Kellenberger *et al.*, 2005). The research of Schmitt and co-workers also showed a VWA-like fold (figure 84) in the structure of the N-terminal domain of the p34 protein from *Chaetomium thermophilum*, corresponding to an area that binds tightly to subunit p44, potentially forming a regulatory dimer for the activity of XPD (Schmitt *et al.*, 2014).

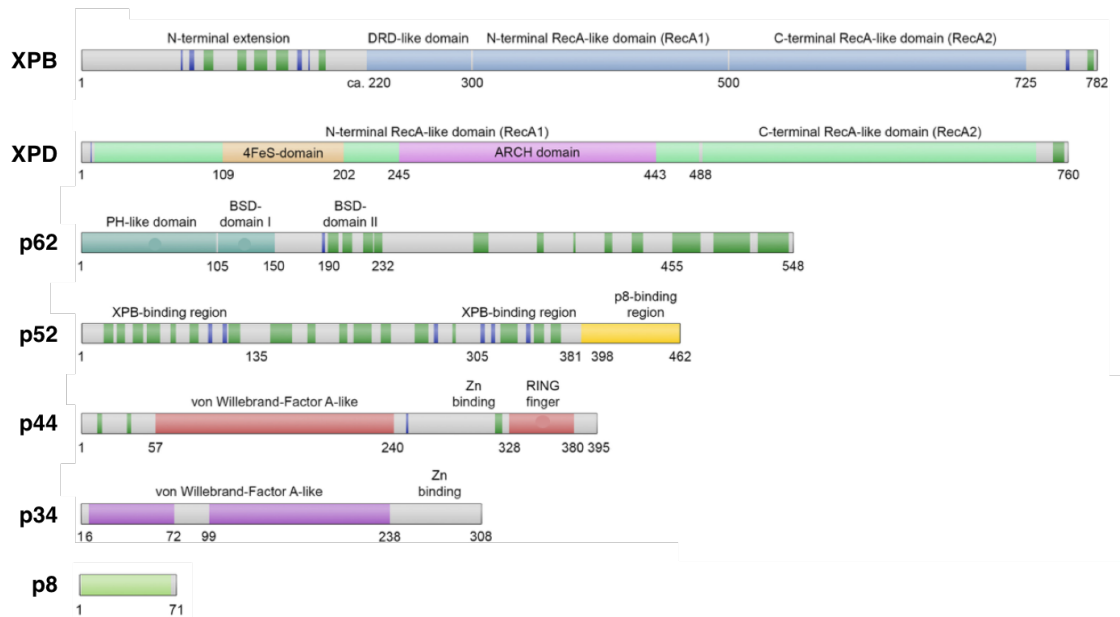


Figure 84: Architecture of the different TFIIH Core subunits.

Representation of the structural characteristics of all TFIIH Core subunits, with specific domains of known structure indicated by coloured bars. Unlabeled green bars represent predicted α -helices, while unlabeled blue bars represent predicted β -sheets. Adapted from (Greber *et al.*, 2017).

5.1.2. TFIIH's redefined mechanism of action

The work carried out by the Ranish group provided additional information about the interaction of the TFIIH subunits within the complex. Based on extensive crosslinking experiments with both the human and the yeast TFIIH, they proposed a novel model in which subunit p62 interacts with subunits p44 and p34, located at the base of the complex, and with XPD, thus anchoring the helicase to the TFIIH Core (Luo *et al.*, 2015). Similarly, the interaction between p52 and p8 not only regulates the ATPase activity of XPB, but also ties the ATPase to the sub-complex (Luo *et al.*, 2015). Luo *et al.* also proposed that the binding of the CAK sub-complex to the Core occurs through interactions not only with XPD, but also with Core subunits XPB and p44.

The Ranish model was later validated with the publication of the cryoEM structure of the human TFIIH Core sub-complex plus CAK-subunit MAT1 (Greber *et al.*, 2017) and the structure of the *S. cerevisiae* homologue of TFIIH as part of the PIC (Schilbach *et al.*,

STUDYING TFIIH INTERACTIONS WITH DNA

2017). Greber and co-workers show a horseshoe-shaped complex with helicases XPB and XPD positioned at the open ends. Importantly, p62 could not be unequivocally placed in their cryoEM map, although observation of the unassigned density in the structure led Greber *et al.* to agree with Luo's previous assessment regarding the importance of p62 as the subunit that holds the TFIIH Core together. The extended contacts and positioning of p62 within TFIIH were confirmed in the structure of PIC-bound TFIIH (Schilbach *et al.*, 2017).

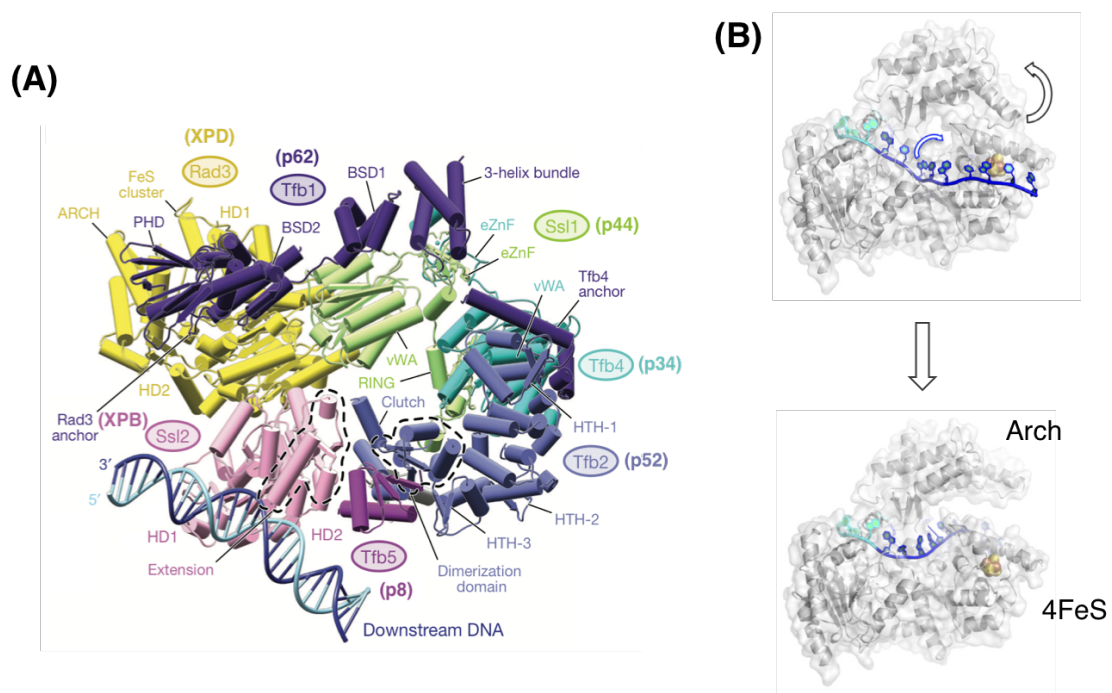


Figure 85: TFIIH opens DNA through the combined action of XPB and XPD.

Proposed mechanisms for the initial opening of DNA by XPB (A) and posterior unwinding and extension of the repair bubble by XPD (B). Figure (A) adapted from (Schilbach *et al.*, 2017) and (B) from (Constantinescu-Aruxandei *et al.*, 2016).

Although an initial model in which XPB experienced a huge ATP-dependent conformational change to anchor TFIIH to the damaged site and open the DNA was first proposed (Fan *et al.*, 2006) (Oksenysh *et al.*, 2009), recent works regarding TFIIH's role in transcription support a model in which XPB opens the duplex not by unwinding it as XPD does, but by translocating along the DNA (Fishburn *et al.*, 2015) (He *et al.*, 2016) (Schilbach *et al.*, 2017) from its position downstream the transcription starting site, threading DNA into the RNAPol (figure 85 (A)) (Schilbach *et al.*, 2017). Hypothetically,

this would happen without the need for substantial conformational changes, as the remaining TFIIH subunits might already provide enough support for XPB to carry out its function without a drastic reorganization (Greber *et al.*, 2017). On the contrary, the new model regarding how XPD unwinds DNA proposes that HD2 engages ssDNA before the transient opening of the interface between the Arch and 4FeS domains, which will allow ssDNA to pass through the channel formed by Arch, 4FeS and HD1 to finally engage with HD1 (figure 85 (B)) (Constantinescu-Aruxandei *et al.*, 2016).

5.2. Results

5.2.1. Observing TFIIH binding to damaged DNA

We tested the binding of three different TFIIH complexes (10-subunit TFIIH complex, TFIIH Core sub-complex with suboptimal amounts of XPD (referred to as TFIIH Core) and TFIIH Core sub-complex with all seven subunits in stoichiometric amounts (referred to as XPD-enriched TFIIH Core) to a series of 44 bp DNA duplex substrates (S1, S2), substrates presenting an 18-nucleotide bubble (S3, S4) or a 3-nucleotide bubble (S5, S6, S7, S8) (chapter 3, section 3.2.2, figure 42). As we previously did with XPC, TFIIH binding was tested by three different methods: band shifting, fluorescence anisotropy and bulk FRET assays.

5.2.1.1. Electrophoretic mobility shift assay (EMSA)

The TFIIH Core sub-complex was tested for binding to substrate S8 in a band shifting assay according to the method described in chapter 2, section 2.5.4. S8 is a 44 bp-long duplex carrying a 3-nucleotide mismatch, with a fluoro-dT nucleotide placed at the centre of the bubble serving as a bulky adduct. Substrates S7 and S8 only differ in the position of the Cy3 dye they carry additionally to the fluoro-dT nucleotide (Cy3 located at the left (S7) or right (S8) side of the 3-nucleotide bubble). Both substrates had been previously tested for binding of damage-detector XPC in an EMSA assay, proving that the position of the Cy3 dye did not affect the binding of the protein in this type of assay. For this reason, only substrate S8 was tested for binding of the sub-complex (figure 86).

The assay showed an apparent K_D of approximately $0.4 \mu\text{M}$ (table 6). While this

STUDYING TFIIH INTERACTIONS WITH DNA

experiment confirmed that TFIIH Core can bind a dsDNA substrate without XPC being present at the bubble, making it a promising starting point in our research, the lack of precision of the assay as shown in our previous results obtained for the binding of XPC to substrate S8 led us to approach the study of TFIIH Core binding to DNA in fluorescence anisotropy assays rather than further exploring additional band shifting assays.

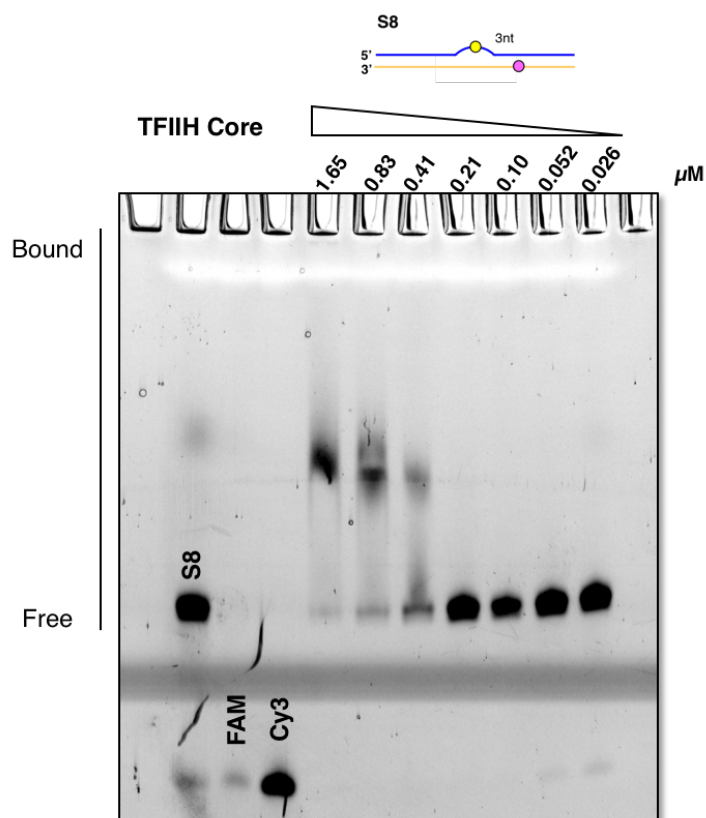


Figure 86: EMSA assay (10% polyacrylamide gel) showing binding of the TFIIH Core to substrate S8.

The TFIIH Core sub-complex binds to substrate S8 (44-mer duplex with a 3-nucleotide bubble carrying a fluorescein molecule mimicking a bulky adduct and a Cy3 dye right to the bubble) with an apparent K_D of approximately $0.4 \mu\text{M}$ (50% substrate bound).

5.2.1.2. Anisotropy assays

5.2.1.2.1. 10-subunit TFIIH complex

We first analysed the binding of the complete 10-subunit TFIIH complex to a 44-mer dsDNA substrate carrying a fluoro-dT modification at the centre of its sequence (S1)

(figure 87) in a fluorescence anisotropy assay following the method described in chapter 2, section 2.5.5.1.

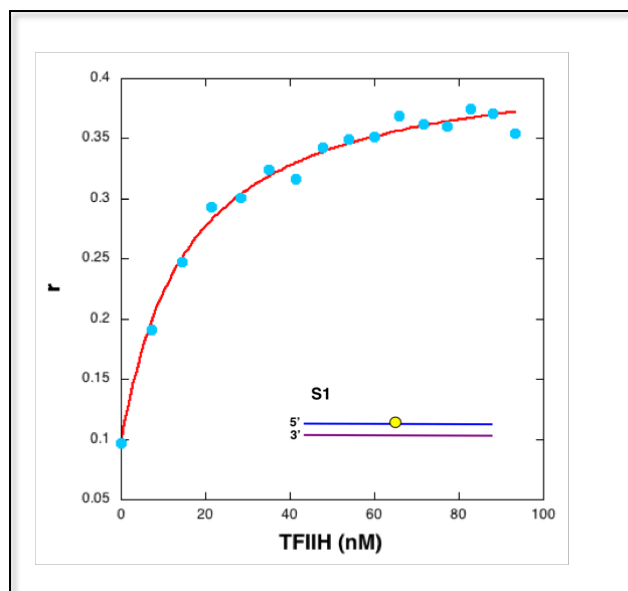


Figure 87: Fluorescence anisotropy assay showing binding of the 10-subunit TFIIH complex to substrate S1.

K_D obtained for the binding isotherm corresponding to substrate S1 (44-mer duplex with a fluorescein molecule mimicking a lesion inside the sequence, final concentration 25 nM) was in the nM range.

Data obtained in this experiment were fitted to a quadratic equation assuming a 1:1 binding model, offering a K_D of 16 nM, considerably lower than the 0.4 μ M value obtained in our EMSA assay performed with TFIIH Core and the mismatched duplex S8. Unfortunately, the difficulties experienced in reproducing the purification of the 10-subunit TFIIH complex meant that the sample available for the biochemical characterization of the full complex was extremely limited, so we focused on the study of the binding of the TFIIH Core sub-complex instead.

5.2.1.2.2. TFIIH Core sub-complex

Initial fluorescence anisotropy experiments to characterize the binding of the TFIIH Core

STUDYING TFIIH INTERACTIONS WITH DNA

sub-complex to DNA were performed using substrates S1, S2, S3 and S4 (44-mer duplexes carrying a fluoro-dT nucleotide (S1 and S3 – the latter additionally carrying an 18-nucleotide bubble) or a 5'FAM-C6 modification (S2 and S4 – as S3, S4 additionally carries an 18-nucleotide mismatch)) (figure 88). All four experiments were performed in buffer 20 mM MES pH 6.5, as we had previously done in the XPC and 10-subunit TFIIH fluorescence anisotropy assays.

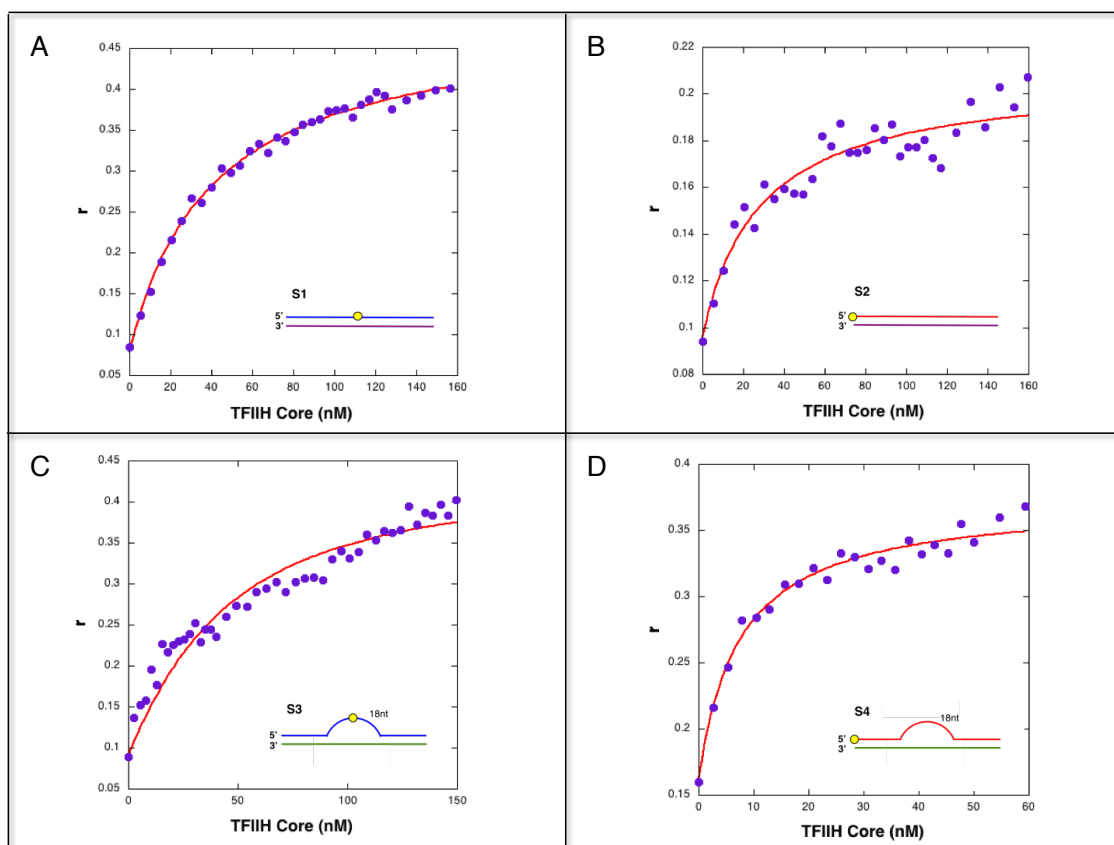


Figure 88: Fluorescence anisotropy assay showing binding of the TFIH Core sub-complex to substrates S1, S2, S3 and S4 in buffer MES.

Substrates S1 and S2 are 44-mer duplexes carrying a fluorescein molecule at the centre of the sequence (S1, panel A) or at the 5' end of the substrate (S2, panel B). Substrates S3 (panel C) and S4 (panel D) are equivalent to S1 and S2, respectively, but presenting an 18-nucleotide mismatch besides the fluorescein molecule (which also mimics a bulky adduct in the case of substrates S1 and S3). Final concentration for all substrates was 25 nM. K_D 's obtained for all four binding isotherms were in the nM range.

As we previously saw in our anisotropy assays with protein XPC, initial and final anisotropy values change for the different substrates tested due to variations in the lifetime of our fluorescein probe as a result of its 5' end or internal location in the substrate (Nazarenko *et al.*, 2002). The relative change is approximately 0.2 units for all four substrates. Data were fitted to a quadratic equation reflecting a 1:1 binding model.

K_D values obtained were 40 nM for substrate S1, 28 nM for S2, 29 nM for S3 and 8 nM for S4 (table 6). All four values are quite similar, suggesting that the presence of a mismatch does not increase TFIIH's affinity for the DNA. The binding affinities observed here are about 10 times lower than the value obtained in the EMSA assay for substrate S8 (similar to S3 but with a 3-nucleotide bubble instead of an 18-nucleotide one), validating our decision to employ fluorescence anisotropy assays in our research as they offer much more accurate information.

We observed that the addition of the Core sub-complex to our substrates caused a considerable enhancement of the emitted fluorescence (figure 89) when the reaction was performed in MES buffer. As we purify the TFIIH Core in HEPES buffer, we decided to perform the binding reactions in 20 mM HEPES pH 7.0, 50 mM NaCl, as an attempt to eliminate this enhancement and minimize the changes in the sub-complex. The fluorescein molecule behaved more stably in these conditions, probably due to the effect of the salt component in the HEPES buffer on the two negative charges present in the dye, so our binding reactions were subsequently carried out in this buffer in our remaining experiments.

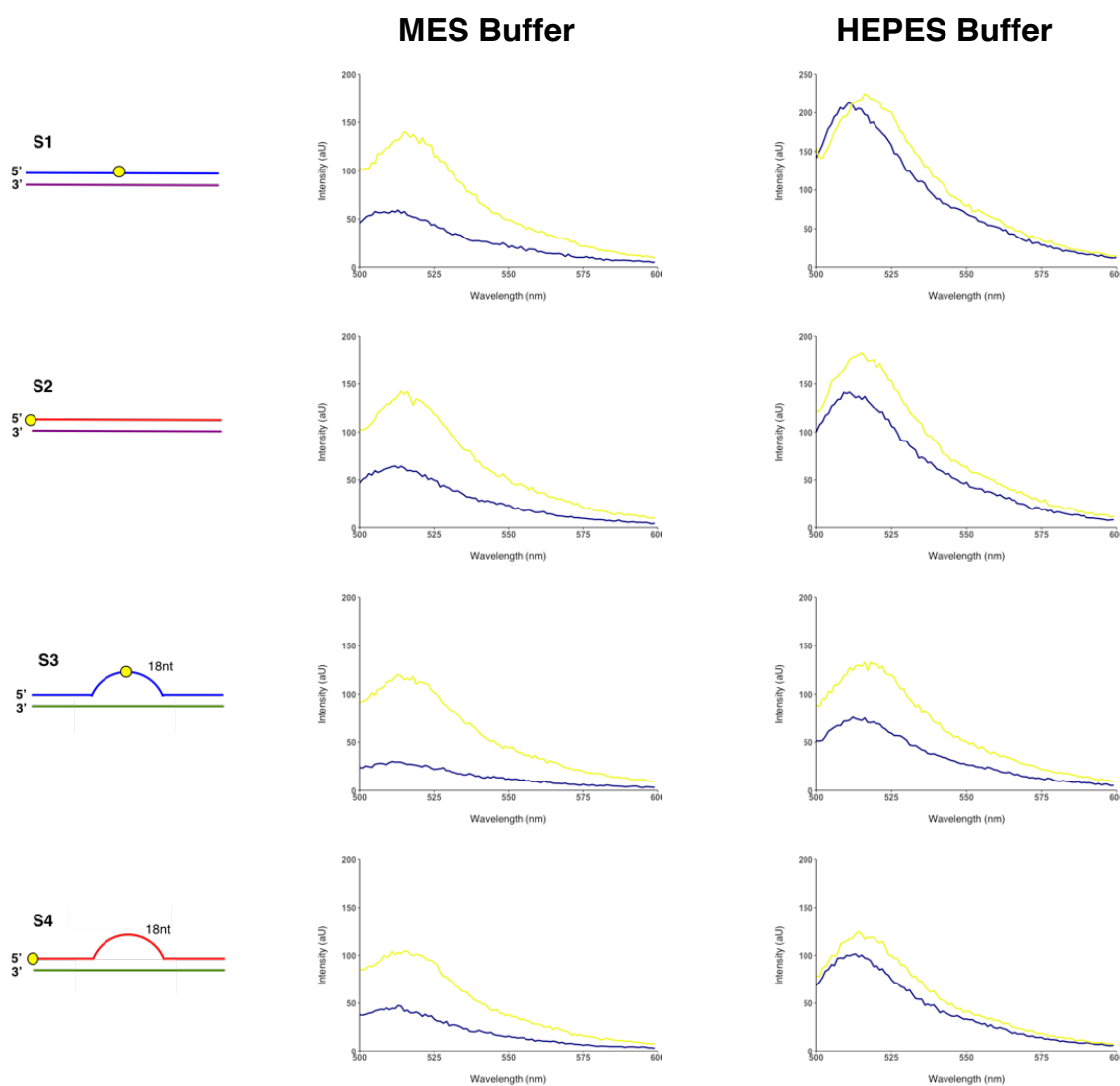


Figure 89: Influence of the reaction buffer in our anisotropy experiments.

Addition of TFIIH Core to substrates S1, S2, S3 and S4 caused a considerable enhancement of the emitted fluorescence (yellow lines) compared to free substrate (dark blue lines) when the reaction is performed in MES buffer. The same reaction performed in HEPES buffer with 50 mM NaCl showed a less pronounced influence of protein binding in the fluorescent response of the dye. Final concentration for all substrates was 25 nM.

To verify that the change of buffer did not alter the binding affinity of the complex we next analyzed the binding of TFIIH Core to substrates S3 and S4 in a reaction carried out in buffer 20 mM HEPES pH 7.0, 50 mM NaCl (figure 90). K_D values obtained were 13 nM for S3 and 42 nM for S4, very close to the values obtained in our previous experiment performed in MES buffer (table 6). The binding affinity for substrate S3 was

approximately 3 times higher than affinity for substrate S4, hinting a possible preference for a damaged substrate over an undamaged one, although the difference between the two values was too small to confirm this point.

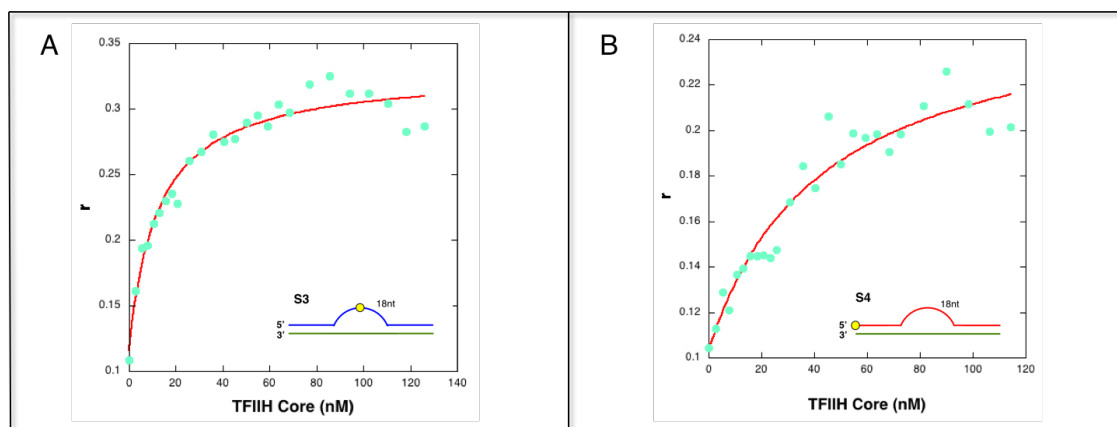


Figure 90: Fluorescence anisotropy assay showing binding of the TFIIH Core sub-complex to substrates S3 and S4 in buffer 20 mM HEPES, 50 mM NaCl.

K_D 's obtained for substrates S3 (A) and S4 (B) were still in the nM range and very close to the values obtained when the reaction was performed in MES buffer. Final concentration for all substrates was 25 nM.

Previous experiments carried out with damage-detector XPC suggested that an 18-nucleotide bubble was too big to be recognised by the protein as a mismatch, and so binding to a similar substrate, carrying a 3-nucleotide bubble (substrates S5 and S6) instead of an 18-nucleotide one (substrates S3 and S4), was further studied. With this result in mind, TFIIH Core was tested for binding to substrates S5 and S6 to determine if the size of the bubble also had an effect on the binding of the sub-complex (figure 91).

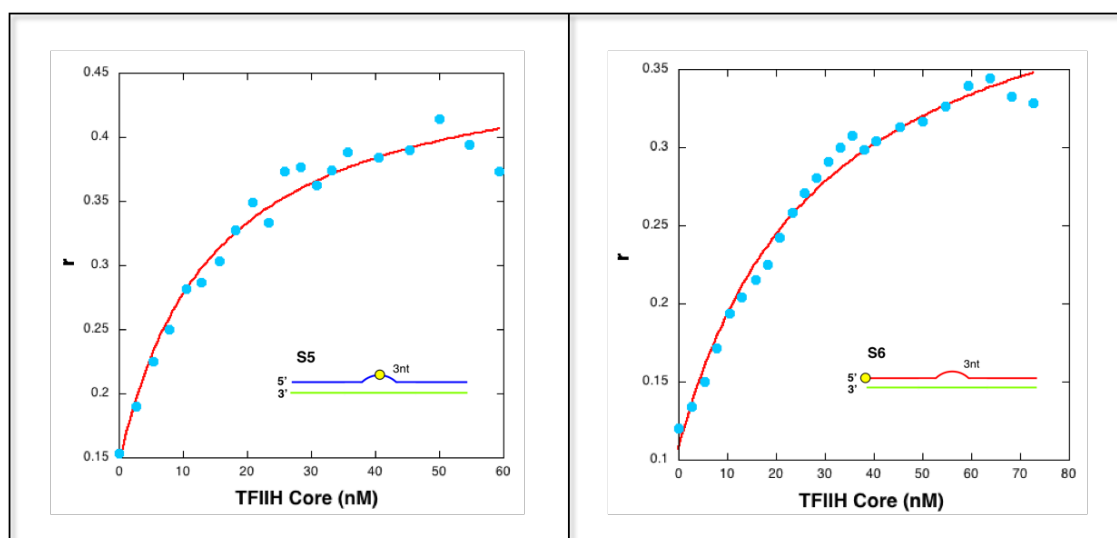


Figure 91: Fluorescence anisotropy assay showing binding of the TFIIH Core sub-complex to substrates S5 and S6.

Substrates S5 and S6 are very similar to S3 and S4, carrying a 3-nucleotide bubble instead of an 18-nucleotide one. A fluorescein molecule is positioned at the centre of the bubble, mimicking a bulky lesion (S5, A), or at the 5' end (S6, B). Final concentration for all substrates was 25 nM. K_D obtained were in the nM range and similar to the values obtained for binding of the sub-complex to the substrates carrying an 18-nucleotide bubble.

K_D obtained for binding to substrate S5 was 15 nM, while K_D for binding to substrate S6 was 29 nM. These values are similar to the ones obtained for the substrates carrying the 18-nucleotide bubble, and the K_D obtained for substrate S5 (equivalent to S3) is approximately two times lower than the K_D obtained for S6 (equivalent to S4). Although this supports our previous observation that the presence of damage at the mismatch might increase the binding affinity of the Core sub-complex, the difference between the two K_D values is again too small to draw a final conclusion. A further investigation will be required to confirm this point.

Once we solved the ongoing issues reducing the amount of helicase XPD available in our samples we tested our new XPD-enriched TFIIH Core sample for binding to substrates S5 and S6 to see if the presence of the helicase in similar amounts as the other six Core subunits had an effect on the binding of the sub-complex (figure 92).

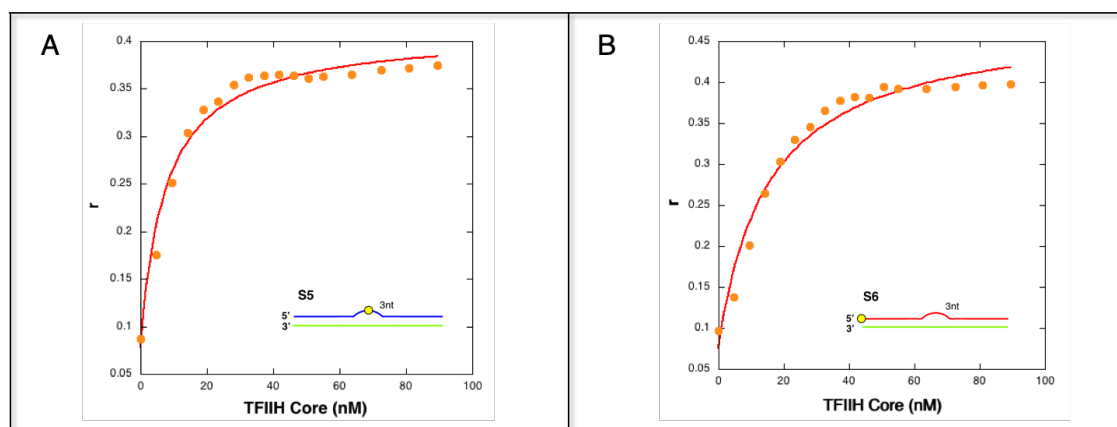


Figure 92: Fluorescence anisotropy assay showing binding of the XPD-enriched TFIIH Core sub-complex to substrates S5 and S6.

Substrates S5 and S6 are 44-mer duplexes with a 3-nucleotide mismatch carrying a fluorescein molecule at the bubble, mimicking a lesion (S5, A), or at the 5' end of the substrate (S6, B). Final concentration for the two substrates was 25 nM. The higher presence of XPD in this sample (now stoichiometrically similar to the other six TFIIH Core subunits) didn't increase the binding affinity of the sub-complex.

K_D values obtained for substrates S5 and S6 were 8 nM and 15 nM, respectively, which is approximately two-fold lower than the values of 15 nM and 29 nM previously obtained with our sample containing suboptimal amounts of XPD (table 6). Although the binding affinity of the sub-complex had been slightly increased by the presence of XPD, this enhancement is not enough to determine if XPD contributes to binding of the TFIIH sub-complex to DNA.

5.2.1.3. Bulk FRET assays

The preliminary results obtained in our fluorescence anisotropy experiments suggested that the TFIIH Core sub-complex can tightly bind a duplex substrate, and a substrate presenting either an 18-nucleotide or a 3-nucleotide mismatch, and the binding affinity is approximately two times higher when a bulky adduct is present at the site of the mismatch.

We further tested the binding capacity of the TFIIH sub-complex in a bulk FRET assay using substrates S7 and S8 (44 bp long duplexes carrying a 3-nucleotide mismatch and

STUDYING TFIIH INTERACTIONS WITH DNA

a fluoro-dT nucleotide at the centre of the bubble, plus a Cy3 dye at the left (S7) or the right (S8) side of the mismatch) (figure 93), following the method described in chapter 2, section 2.5.5.2. Data were fitted to a quadratic equation reflecting a 1:1 binding model.

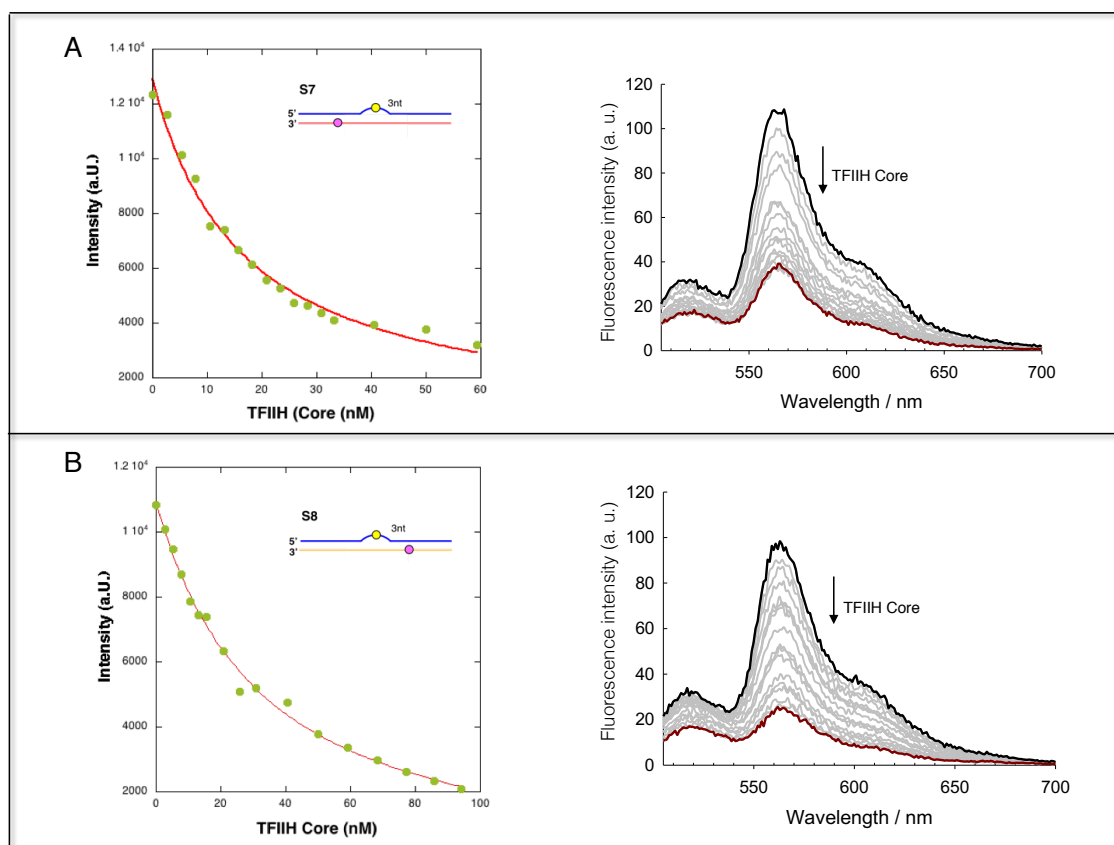


Figure 93: Bulk FRET assay showing binding of TFIH Core to substrates S7 and S8.

Substrates S7 and S8 are 44-mer duplexes with a 3-nucleotide bubble carrying a fluorescein molecule mimicking a bulky adduct and a Cy3 dye left (S7, A) or right (S8, B) to the bubble. K_D obtained for the two of them were in the nM range. Initial intensity for both the fluorescein and the Cy3 dyes are indicated by a black line, while final intensity after addition of the TFIH sub-complex is indicated by a red line.

K_D values obtained were 16 nM for substrate S7 and 32 nM for substrate S8. These values are similar to those obtained for substrates S3, S4, S5 and S6 in our anisotropy assays (table 6), further supporting our previous observation that TFIH Core is able to bind a DNA duplex carrying a mismatch without the presence of damage-detector XPC. The emitted fluorescence of both the fluorescein and Cy3 dyes was collected after every titration point, showing that the addition of the TFIH Core sub-complex causes an

approximate quenching of 80% for both dyes when directly excited, although no change of the ratio *A* was observed (data not shown).

5.2.2. Characterization of TFIIH's unwinding activity

The helicase activity of the XPD-enriched Core sub-complex was tested in a fluorescence-based unwinding assay with several different substrates, as discussed in chapter 2, section 2.5.5.3: an undamaged splayed duplex (SD), a damaged splayed duplex with a fluorescein molecule mimicking a bulky adduct located in either the non-translocating strand (SD13) or the translocating strand (SD24), an asymmetric duplex with a 5' overhang and a translocating strand mainly composed of purine bases (XPD12) or pyrimidine bases (XPD34) and an asymmetric duplex with a 3' overhang and a translocating strand mainly composed of purine bases (REV12) or pyrimidine bases (REV34) (figure 94).

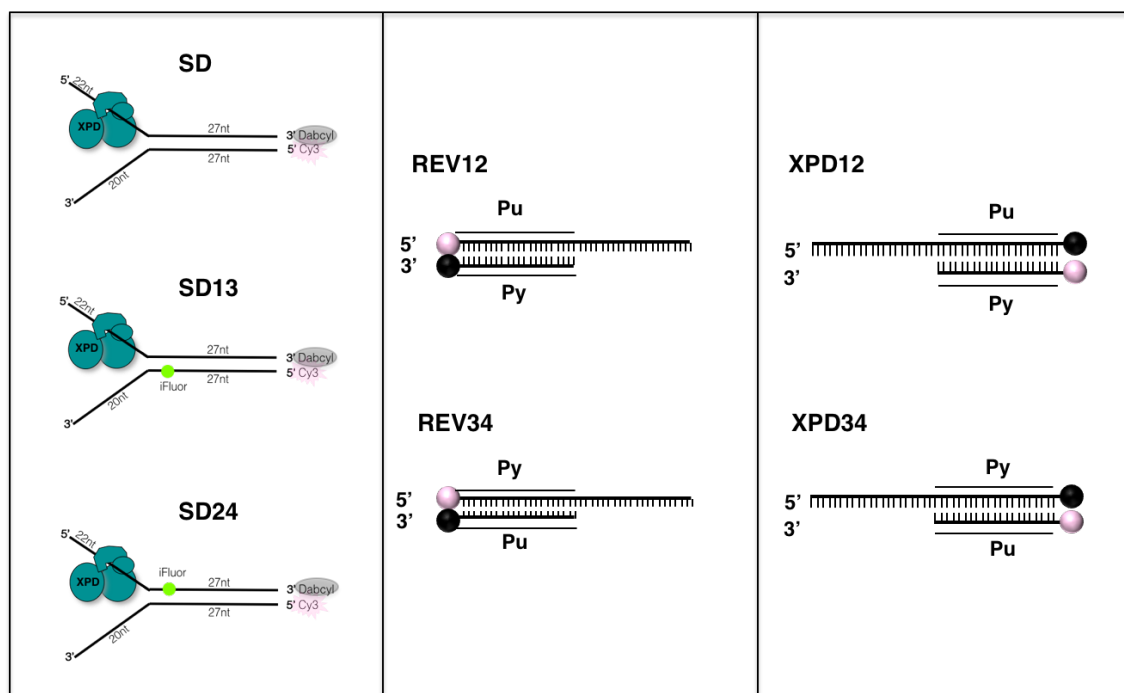


Figure 94: DNA substrates employed to characterize TFIIH's unwinding activity.

Schematic representation of the different DNA substrates used in the helicase assays performed to study the unwinding activity of the TFIIH Core sub-complex. All substrates carry a Cy3 dye (pink spheres) suppressed by a quencher (black spheres).

STUDYING TFIIH INTERACTIONS WITH DNA

All these substrates include a Cy3 dye, whose fluorescence is suppressed by a quencher in the duplex. Upon binding of the TFIIH Core sub-complex, the XPD helicase will unwind the substrate in an ATP-dependent reaction, effectively removing the quencher and allowing the detection of the fluorescence emitted by the Cy3 dye (figure 95).

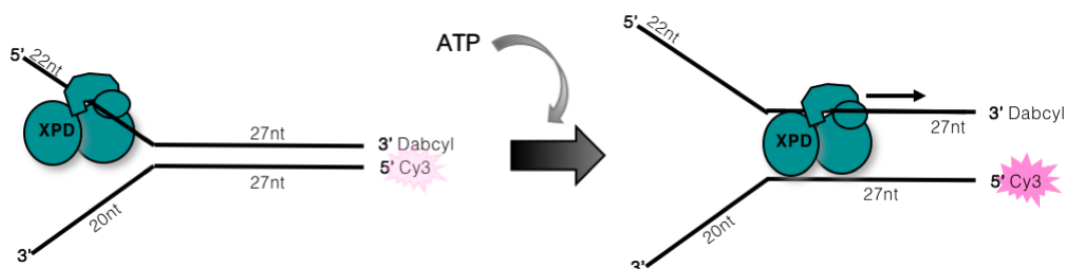


Figure 95: Mechanism of action of our fluorescence-based helicase assay.

The ATP-dependent unwinding activity of XPD will open the splayed duplex substrate, thus removing a quencher that was suppressing the emission of the Cy3 dye, observing now an increment of its emitted fluorescence in time.

5.2.2.1. Determining the polarity of the TFIIH Core unwinding

The TFIIH Core sub-complex possesses two helicases of opposite polarities: XPB (3' to 5') and XPD (5' to 3'). However, XPB has recently been defined as a translocase rather than a helicase in the canonical sense (Fishburn *et al.*, 2015) (He *et al.*, 2016) (Schilbach *et al.*, 2017). To verify the lack of helicase activity of XPB and the polarity of the unwinding by XPD we performed an initial fluorescence helicase experiment with an undamaged splayed duplex (SD) and an asymmetrical duplex with a 3' overhang end (REV34) (figure 96).

The assay showed that unwinding only happened when a 5' overhang end was available, as only the SD substrate was opened by the sub-complex while the REV34 substrate was not unwound at all, thus confirming the 5' to 3' polarity of the unwinding activity carried out by the TFIIH Core sub-complex.

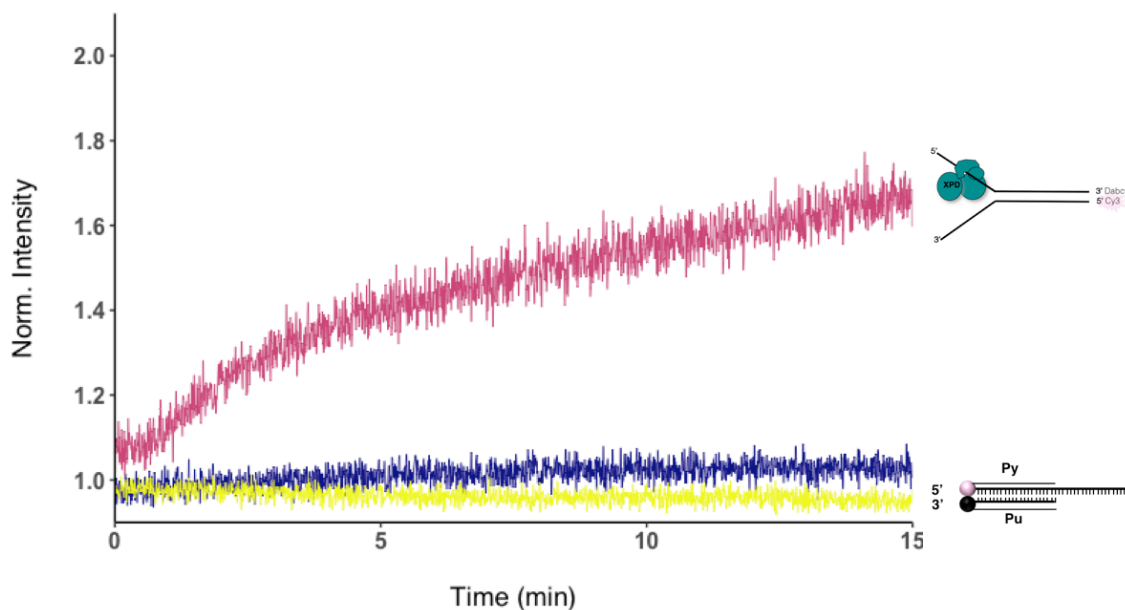


Figure 96: TFIIH Core unwinds a substrate with a 5' overhang.

The helicase activity of the TFIIH Core sub-complex was tested in a fluorescence helicase assay with two different substrates: a splayed duplex (SD) and an asymmetric duplex with a 3' overhang (REV34), both at a final concentration of 50 nM. A control reaction without ATP showed no unwinding (blue line). TFIIH Core was able to efficiently unwind the SD substrate (pink line), but not the REV34 substrate (yellow line).

To further study this result we tested the unwinding capacity of the TFIIH Core sub-complex with four asymmetric duplexes, two with a 5' overhang (XPD12 and XPD34) and another two with a 3' overhang (REV12 and REV34), with each pair differing in the composition of the translocating strand (figure 97).

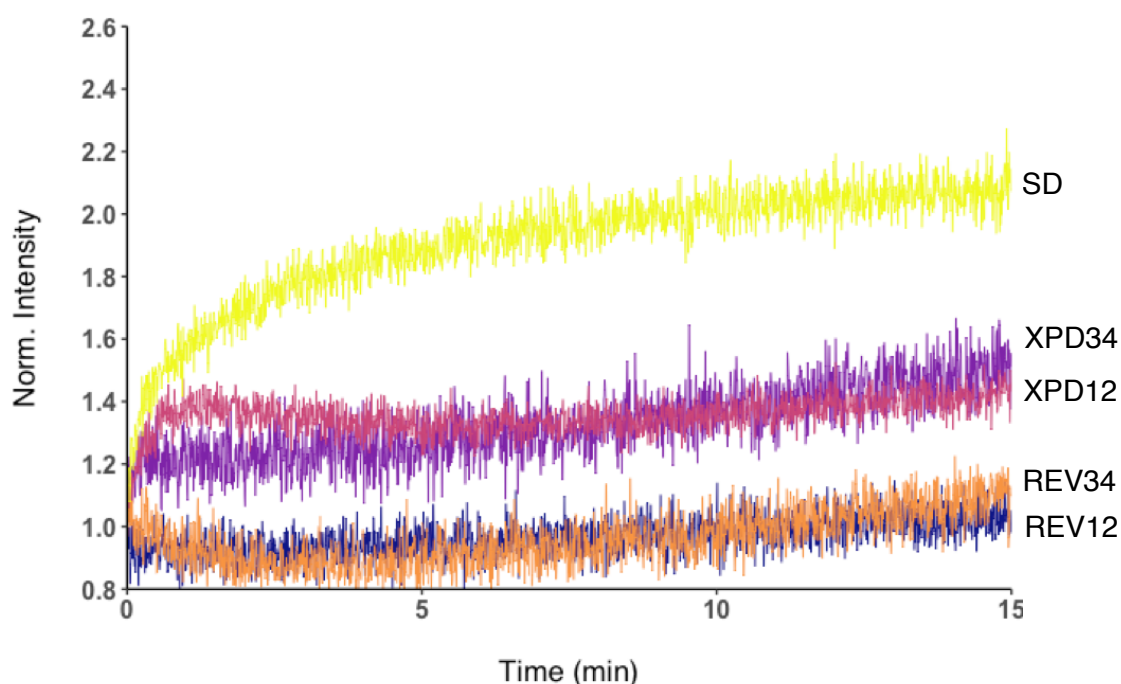


Figure 97: TFIIH Core opens a splayed duplex substrate preferentially.

Fluorescence helicase assay showing unwinding of substrates SD (yellow), XPD12 (pink), XPD34 (purple), REV12 (blue) and REV34 (orange), all at a final concentration of 25 nM. The REV substrates were not opened, as expected, due to their 3' to 5' polarity. The asymmetric duplexes XPD12 and XPD34 can be unwound by XPD, but the sub-complex shows a marked preference for a splayed duplex conformation.

The REV12 and REV34 substrates could not be unwound, reaffirming the lack of apparent 3' to 5' helicase activity of TFIIH Core. Although the sub-complex showed a marked preference for the SD substrate, both the XPD12 and XPD34 substrates were unwound too, and a possible preference for the XPD12 substrate (with a majority of purine bases making up the translocating strand) over the XPD34 substrate (with a majority of pyrimidine bases making up the translocating strand) was hinted at. This was further confirmed by our experiments in a separate assay using substrates XPD12 and XPD34 alone, with XPD12 showing approximately 50% more unwound product than XPD34 (figure 98).

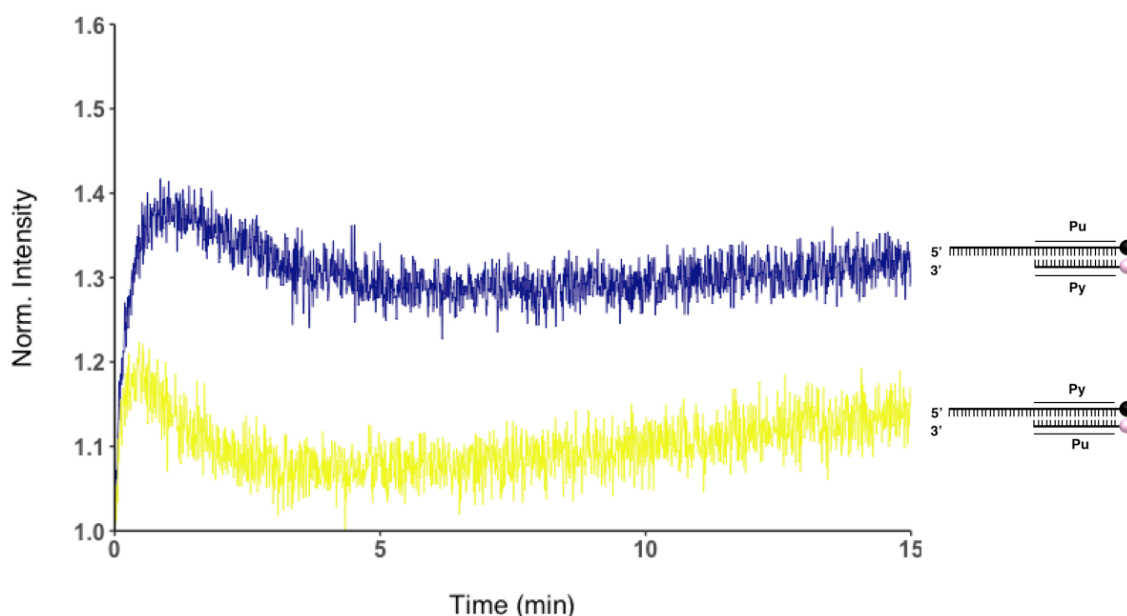


Figure 98: XPD discriminates substrate composition.

Fluorescence helicase assay showing the unwinding of two asymmetric duplexes with a 5' overhang end, one with a majority of purine bases (XPD12) and the other with a majority of pyrimidine bases (XPD34) composing the translocating strand. Unwinding of the XPD12 substrate (blue line) is approximately 2-times higher than unwinding of the XPD34 substrate (yellow line), suggesting that XPD can differentiate and open preferentially a substrate based on its nucleotide composition. Final concentration for both substrates was 25 nM.

5.2.2.2. Establishing the optimal conditions for the assay

Next, we investigated the effect of TFIIH Core concentration on the unwinding of substrate SD (figure 99). Increasing concentrations of the sub-complex (25, 50, 100 and 200 nM) were added to our unwinding assay, showing a progressive increment of about 50% of the unwound product as concentration of the sub-complex increases, which confirms that unwinding is dependent on TFIIH Core concentration. As this assay requires a considerable amount of sub-complex to be performed and there is only a 30% difference in the amplitude of the unwinding between 50 nM, 100 nM and 200 nM TFIIH Core, we established 50 nM as the standard sub-complex concentration for our future unwinding assays.

STUDYING TFIIH INTERACTIONS WITH DNA

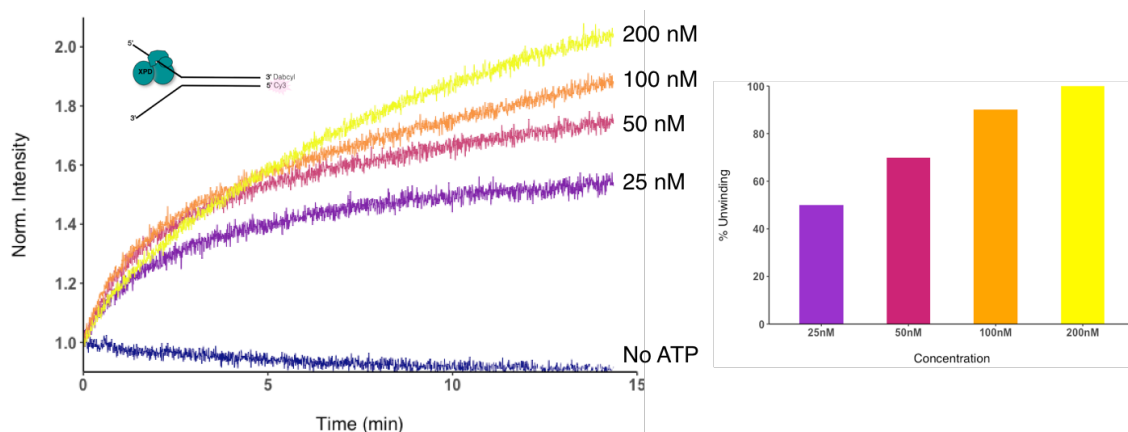


Figure 99: Unwinding of the SD substrate increases with TFIIH Core concentration.

Fluorescence helicase assay showing unwinding of substrate SD (25 nM) by increasing concentrations of the TFIIH Core sub-complex: 25 nM (purple), 50 nM (pink), 100 nM (orange), and 200 nM (yellow). The blue line represents a control reaction in which no ATP was added. The graph shows that unwinding is dependent on TFIIH Core concentration, as duplex opening at a sub-complex concentration of 200 nM is about 50% higher than unwinding at 25 nM.

We then further tested the optimal conditions for our assay by trying different ATP concentrations (figure 100 (A)) and different temperatures (figure 100 (B)). To determine the best ATP concentration the reaction was performed at 25 °C, showing approximately 30% more unwound substrate when the reaction was performed with 1 mM ATP compared to a reaction performed with 0.5 mM ATP. We took 1 mM as the standard ATP concentration for our assays, and verified the best temperature for our experiment by comparing unwinding at 25 °C and 37 °C. As the assay showed that there was only a difference of about 15% between the amount of unwound product obtained at 25 °C and that obtained at 37 °C, we continued to perform our assays at 25 °C, establishing this temperature as the standard condition for further assays.

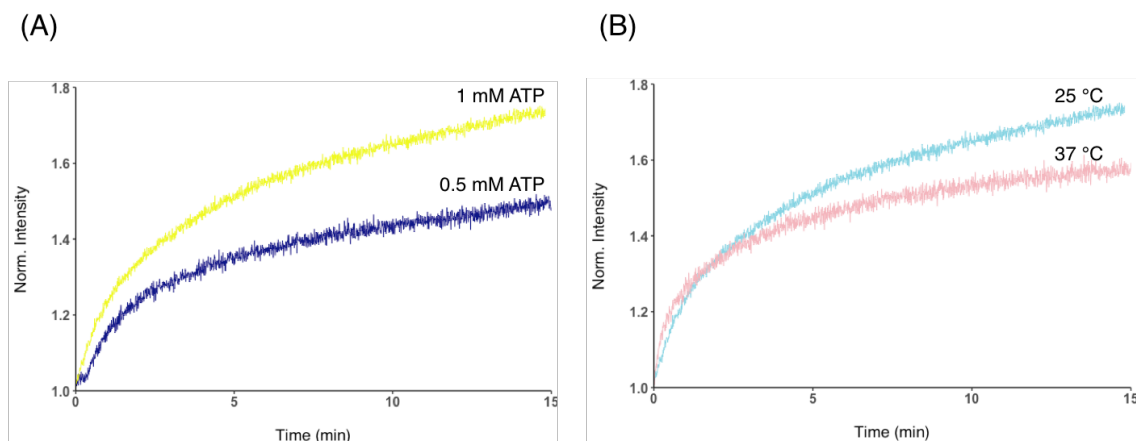


Figure 100: Establishing the ideal ATP concentration and temperature for our assay.

Fluorescence helicase assay showing unwinding of substrate SD (25 nM) by the TFIIH Core sub-complex at different ATP concentrations (A). 1 mM ATP (yellow line) was confirmed as the optimal ATP concentration for the assay after comparing this result with the one obtained in a reaction containing 0.5 mM ATP (dark blue line). Different temperatures were also analyzed (B) by running the assay at 37 °C (pink line) and 25 °C (light blue line), showing no significant changes at the different conditions.

5.2.2.3. Effect of RPA and XPA on TFIIH Core unwinding

The XPA and RPA proteins are the next factors to join the repair bubble after TFIIH is recruited by XPC. Our previous experiments had confirmed that TFIIH can bind to and unwind a dsDNA substrate *in vitro* without the presence of XPC, so we subsequently studied the effect of the addition of RPA, XPA or both on TFIIH Core unwinding of a splayed duplex substrate in an assay performed as described in chapter 2, section 2.5.5.3.

RPA is a heterotrimeric complex essential in multiple cellular processes. As part of the NER machinery, RPA binds to the undamaged, single-stranded portion of the repair bubble and protects it from degradation (Krasikova *et al.*, 2010) until the damaged fragment is removed and the gap left is filled by a polymerase (Riedl *et al.*, 2003). RPA is recruited together with XPA (Krasikova *et al.*, 2010), a modular protein with no enzymatic activity that acts as a fundamental scaffold for interaction with multiple repair factors and other proteins (Sugitani *et al.*, 2016). In NER, XPA is responsible for the

STUDYING TFIIH INTERACTIONS WITH DNA

removal of the CAK sub-complex from TFIIH (Svejstrup *et al.*, 1995) (Coin *et al.*, 2008) and the recruitment of endonucleases XPG and XPF/ERCC1 (Krasikova *et al.*, 2010). The RPA and XPA proteins, expressed in *E. coli*, were produced and purified in our lab by Mrs. Biljana Petrovic-Stojanovska.

Addition of RPA to a reaction already containing the TFIIH Core sub-complex increased the amplitude of the unwinding by approximately 47%. By contrast, addition of XPA alone, or XPA and RPA together to the TFIIH Core-containing reaction did not alter unwinding significantly (figure 101).

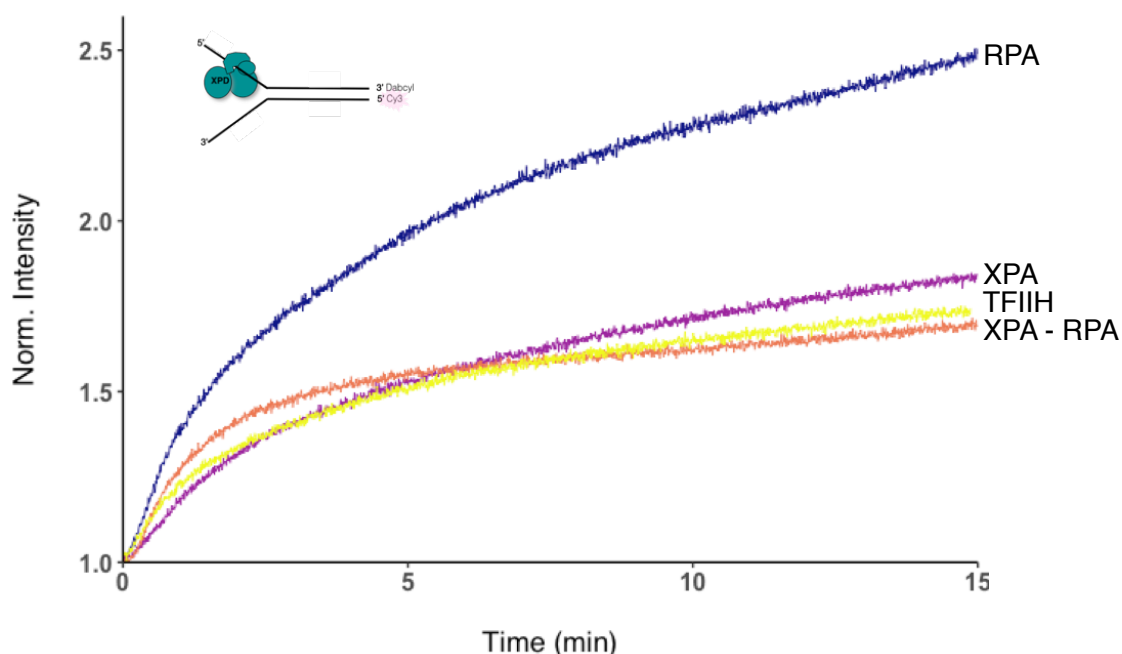


Figure 101: Effect of XPA and RPA on TFIIH Core unwinding of substrate SD.

Fluorescence-based helicase assays showing how the addition of proteins RPA (dark blue line), XPA (purple line) or both (pink line) affect the unwinding of substrate SD (25 nM) by the TFIIH Core sub-complex. The yellow line represents a control in which only the TFIIH Core sub-complex was added to the reaction. Unwinding amplitude was considerably increased by the addition of RPA. Addition of XPA alone or XPA and RPA together didn't alter unwinding by TFIIH Core.

5.2.2.4. Unwinding of damaged DNA substrates

The TFIIH Core sub-complex was also tested for unwinding of two splayed duplex

substrates carrying a fluoro-dT nucleotide mimicking a bulky adduct in the double-stranded portion of the substrate, with substrate SD13 carrying the lesion in the non-translocating strand, while the damage is located in the translocating strand in substrate SD24 (figure 102). The assay showed that the amplitude of unwinding of the damaged substrates was considerably lower than unwinding of the undamaged SD substrate (58% and 73% less unwound product for SD13 and SD24, respectively).

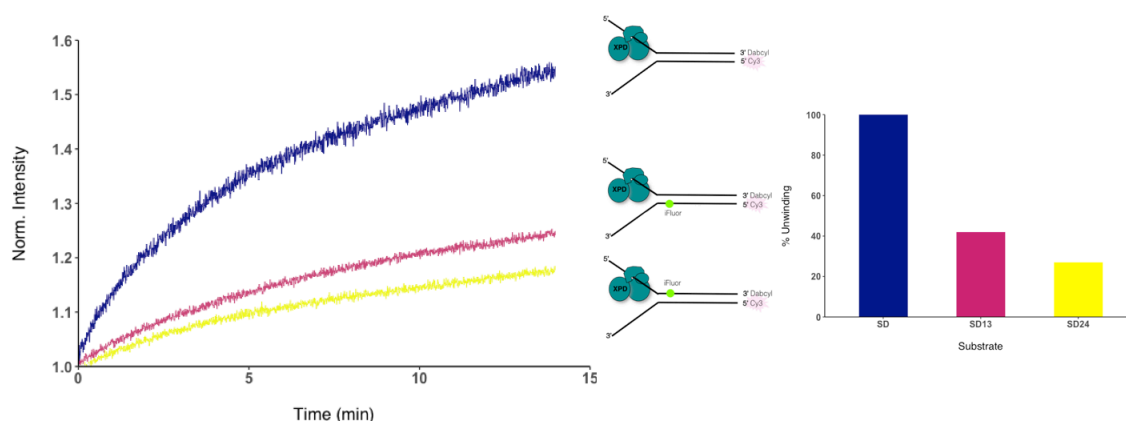


Figure 102: TFIIH Core is stalled by damage in the translocating and the non-translocating strands.

Fluorescence-based helicase assay showing unwinding of an undamaged SD substrate (dark blue line) and two splayed duplexes with a fluorescein molecule mimicking a bulky adduct located in the translocating strand (SD24, yellow line) and the non-translocating strand (SD13, pink line). The presence of a bulky adduct reduces unwinding by 58% approximately when the damage is located in the non-translocating strand, and by 73% when located in the translocating strand. Final concentration for all three substrates was 25 nM.

Analysis of the kinetics for the unwinding of these substrates showed that all of them fitted a bi-exponential model with a burst phase followed by a slow phase, which is in all cases the major component of the reaction (figure 103). Other models including a linear phase representing non-specific binding were also tested, but they did not significantly improve the results compared to the bi-exponential model. The fast component is quite similar in the unwinding of the two damaged substrates (12% for SD13 and 9% for SD24); interestingly, this component is higher for the SD substrate (30%).

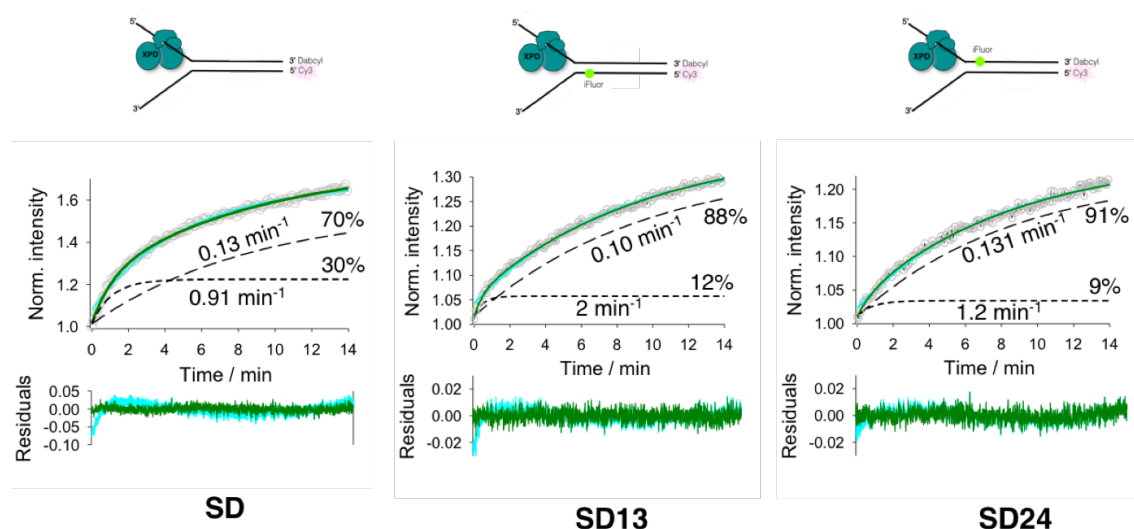


Figure 103: Kinetics of the TFIIH Core unwinding of substrates SD, SD13 and SD24.

TFIIH Core unwinding of undamaged substrate SD and damaged substrates SD13 and SD24 was fitted to a bi-exponential model, with a minor fast component (dotted lines) and a major slow component (dashed lines) in all three cases. The fast component is higher in the unwinding of substrate SD, compared to substrates SD13 and SD24, which showed lower and similar values.

5.3. Discussion

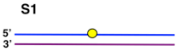

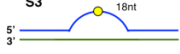
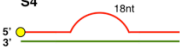
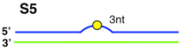
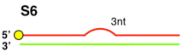
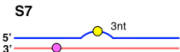
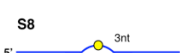
5.3.1. TFIIH binds dsDNA without the presence of XPC

During NER the TFIIH complex is recruited to the damaged site to open and extend the mismatch detected by XPC, in order to allow the subsequent repair factors to access the lesion and repair it. A study by the Yang group reported that the TFIIH Core sub-complex binds a 45-nucleotide ssDNA substrate with a much higher affinity (2.5 nM) than a 45-base pair dsDNA substrate (binding barely detectable in an EMSA assay), and this affinity was not altered by the presence of a lesion in the DNA (Li *et al.*, 2015). Our group studied the binding of the Core sub-complex to a range of 44-mer dsDNA substrates in three different assays: EMSA experiments, fluorescence anisotropy assays and bulk FRET assays, with the K_D values obtained in each experiment summarised in table 6.

STUDYING TFIIH INTERACTIONS WITH DNA

Table 6: Dissociation constants for binding of different TFIIH constructs to a range of substrates in EMSA, fluorescence anisotropy or bulk FRET assays.

The limitations of the EMSA experiment meant that K_D could only be determined as an apparent (approximated) K_D . Data from anisotropy and bulk FRET experiments were fitted to a binding model that assumes 1:1 DNA:protein (see chapter 2, sections 2.5.5.1 and 2.5.5.2).

	K_D (nM)				
	EMSA		Anisotropy		FRET
	TFIIH Core	10xTFIIH	TFIIH Core		TFIIH Core (XPD enriched)
			MES	HEPES	
 S1		16 ± 2	40 ± 3		
 S2			28 ± 7		
 S3			29 ± 5	13 ± 2	
 S4			8 ± 1	42 ± 13	
 S5				15 ± 3	8 ± 1
 S6				29 ± 4	15 ± 3
 S7					16 ± 2
 S8	~ 410				32 ± 3

The EMSA assays we performed with damage-detector XPC didn't offer accurate K_D values for the binding, but they revealed an interesting result: the possible binding of two molecules of the protein to the substrate. We followed the same approach in our study of the binding of the TFIIH Core sub-complex: first we performed an EMSA assay that would offer an overview on the binding, before moving on to more precise techniques, such as fluorescence anisotropy and bulk FRET. Also based on the results obtained in our XPC experiments, we decided to test the binding of the sub-complex to substrate S8 only, a 44-mer dsDNA substrate with a 3-nucleotide bubble with a bulky lesion at the centre of the mismatch. The assay showed that TFIIH Core binds the S8 substrate with an apparent K_D of approximately $0.4 \mu\text{M}$ (table 6) (figure 86). The gel didn't offer any

STUDYING TFIIH INTERACTIONS WITH DNA

other insights regarding the binding mode of the sub-complex, so next we tried to confirm the K_D value obtained here in a fluorescence anisotropy assay.

The 10-subunit TFIIH complex was first construct tested in an anisotropy assay. The full complex includes both the TFIIH Core sub-complex (subunits XPD, XPB, p62, p52, p44, p34 and p8), thoroughly described in previous sections, and the CAK sub-complex (subunits MAT1, cyclin H and cdk7), which inhibits the helicase activity of XPD (Sandrock & Egly, 2001) (Abdulrahman *et al.*, 2013) and is removed from the Core sub-complex after binding to the damaged site in order to complete the repair reaction (Svejstrup *et al.*, 1995) (Coin *et al.*, 2008). The assay showed that the TFIIH complex could bind substrate S1 (a 44-mer dsDNA substrate carrying a bulky adduct at the centre of its sequence) with a K_D of 16 nM (table 6) (figure 87). Although the differences between this assay and the EMSA assay previously performed are too many to allow a real comparison of the two results (different composition of the constructs, substrate S1 carries a lesion but not a mismatch as opposed to substrate S8, higher accuracy of the anisotropy assay compared to EMSA), the great disparity between the two K_D values encouraged us to continue this binding study by means of fluorescence anisotropy assays. This value is also three times higher than the K_D obtained for the binding of XPC to the same substrate, consistent with XPC's role as the mismatch detector (Sugasawa *et al.*, 2001).

The difficulties in obtaining more 10-subunit TFIIH sample led us to focus on the study of the binding to dsDNA of the TFIIH Core sub-complex next. Binding to substrates S1, S2, S3 and S4 (44-bp duplexes carrying a fluorescein molecule at the mismatch, mimicking a bulky adduct (S1, S3), or at its 5' end (S2, S4); S3 and S4 also present an 18-nucleotide mismatch) in a reaction carried out in 20 mM MES pH 6.5 buffer offered K_D s of 40 nM, 28 nM, 29 nM and 8 nM, respectively (table 6) (figure 88). The similarity shown by these four values supports a previous report that stated that TFIIH showed no preference for damaged substrates over undamaged ones (Nocentini *et al.*, 1997).

Substituting our 20 mM MES pH 6.5 reaction buffer for 20 mM HEPES pH 7.0, 50 mM NaCl minimized protein-induced effects on the fluorescent emission of the dye present in our substrates (figure 89), and a repetition of the assay using this new reaction buffer

and substrates S3 and S4 (as a mismatched dsDNA is the natural substrate for the TFIIH complex) offered K_D values of 13 nM for S3 and 42 nM for S4 (table 6) (figure 90). According to this result, the Core sub-complex shows 3-fold higher affinity for a substrate with a lesion accompanying the mismatch than for an undamaged substrate.

TFIIH Core binding experiments to substrates S5 and S6 (44-bp duplexes carrying a 3-nucleotide bubble with a fluorescein molecule at the mismatch, mimicking a lesion (S5), or at its 5' end (S6)) offered a K_D value of 15 nM for S5 and 29 nM for S6 (table 6) (figure 91). The affinity of the sub-complex for the damaged substrate S5 is only 2-fold higher than its affinity for the undamaged substrate S6, similar to our observations for substrates S3 and S4. This reinforces our previous result and allows us to conclude that the binding affinities for both damaged and undamaged substrates are too similar to permit us to unambiguously affirm that TFIIH Core binds preferentially to a damaged substrate (S3, S5) over an undamaged one (S4, S6). Interestingly, the similarity between the K_D values obtained for substrates S3-S4 (carrying an 18-nucleotide bubble) and substrates S5-S6 (carrying a 3-nucleotide bubble) suggests that the size of the mismatch does not play a role in the binding of the sub-complex to the substrate.

A final XPD-enriched, TFIIH Core sub-complex was tested for binding to substrates S5 and S6, showing a K_D of 8 nM and 15 nM, respectively (table 6) (figure 92). This sub-complex shows an affinity for the mismatched substrates 2-times higher than that of the TFIIH Core previously employed in our assays, in which only a residual amount of XPD remained attached to the sub-complex after purification. This result further supports our conclusion that the preference of the Core sub-complex for damaged substrates cannot be asserted. Furthermore, such a small difference in the binding affinity of the two sub-complexes might imply that XPD does not actively participate in the binding of TFIIH Core to the DNA. Further experiments would be required to confirm this last point.

In contrast to the results published by Li and co-workers (Li *et al.*, 2015), our anisotropy fluorescence assays show that TFIIH Core can bind dsDNA with an affinity in the nM range. The K_D s shown in this work are approximately six times lower than the one observed by the Yang group for the binding of the sub-complex to ssDNA when our dsDNA substrate presents a mismatch, and only three times lower when our dsDNA

substrate presents a mismatch and a lesion at the mismatched site, when comparing their results with the ones obtained with our XPD-enriched TFIIH Core construct.

We finally studied the binding of the Core sub-complex to substrates S7 and S8 in bulk FRET assays. Substrates S7 and S8 are 44-mer duplexes carrying a 3-nucleotide bubble with a lesion at its centre and a Cy3 dye at the left (S7) or the right (S8) side of the mismatch. Binding of the Core sub-complex offered a K_D of 16 nM for substrate S7 and 32 nM for S8 (table 6) (figure 93), very similar to the values obtained for substrates S1, S2, S3, S4, S5 and S6, thus further validating the results obtained in our fluorescence anisotropy experiments. Although there is a possibility that the difference in the K_D s obtained for substrates S7 and S8 is due to the location of the Cy3 dye affecting the binding affinity of the Core sub-complex for substrate S8, the 2-fold difference is more likely negligible. A further analysis of the changes on the emitted fluorescence of both the fluorescein and Cy3 dyes present in the S7 and S8 substrates showed no changes in the ratio A, suggesting that the binding of the sub-complex does not significantly alter the conformation of the bubble in our substrates. Further experiments are needed to confirm this point.

5.3.2. A fluorescence-based helicase assay shows TFIIH unwinding of dsDNA

The fact that the TFIIH complex requires two helicases (enzymes known for their high processivity) to open a very small fragment of DNA has been a matter of intense discussion for a long time. The recent redefinition of the 3' to 5' helicase XPB as a translocase (Fishburn *et al.*, 2015) (He *et al.*, 2016) (Schilbach *et al.*, 2017) has shed some light into the possible mechanism of action of the complex, defining different but complementary functions for both XPB and XPD, whose 5' to 3' helicase activity makes it the factor responsible for the actual unwinding of the damaged bubble to allow the subsequent repair factors access to the lesion.

We verified the lack of 3' to 5' helicase activity in an ATP-dependent fluorescence-based unwinding assay using two different substrates: one with a 5' overhang end (SD) and a second one with a 3' overhang end (REV34) available. As expected, the TFIIH Core

sub-complex was only able to open the SD substrate (figure 96). This result was further confirmed in a subsequent assay employing a different set of substrates: two asymmetric duplexes with a 5' overhang (XPD12 and XPD34) and another two with a 3' overhang (REV12 and REV34). XPD12 and REV12 present a majority of purine bases in their translocating strands, while XPD34 and REV34 have a majority of pyrimidine bases composing their translocating strands. The assay confirmed that TFIIH Core cannot open a substrate that doesn't present a 5' overhang end, and interestingly it also showed that the sub-complex opened the XPD12 substrate preferentially over XPD34 (figure 97). This preference had previously been observed by the Penedo group in experiments performed with the archaeal homologue of XPD from *Thermoplasma acidophilum* (unpublished data), and it was further verified by our lab in a second experiment that showed approximately 50% more unwound product for the XPD12 substrate compared to XPD34 (figure 98).

Although our experiments showed that TFIIH Core could open a double-stranded substrate with a 5' overhang efficiently, the amplitude of unwinding of the SD substrate by the sub-complex was 30% higher when compared to unwinding of the XPD12 asymmetric duplex (figure 96). We next analysed the effect of TFIIH Core's concentration on the opening of the SD substrate. The assay showed that unwinding was dependent on the concentration of the sub-complex, with 50% more unwound product at the maximum concentration tested (200 nM) compared to the minimum concentration tested (25 nM) (figure 99). This dependency on concentration has also been reported for the XPD homologue from *T. acidophilum* (Constantinescu-Aruxandei *et al.*, 2016). Different ATP concentrations and temperatures were subsequently tested, showing about 30% more unwound product when the reaction was performed with 1 mM ATP, compared to a reaction performed with 0.5 mM ATP (figure 100 (A)). On the other hand, the assay showed only a 15% more unwound substrate when the reaction was run at 25 °C compared to the same reaction carried out at 37 °C (figure 100 (B)), hinting that within this range temperature doesn't have a significant effect on TFIIH Core unwinding. Taking in consideration all these results, we decided to establish a concentration of 50 nM TFIIH Core and 1 mM ATP, in a reaction performed at 25 °C, as standard conditions to perform any future TFIIH Core unwinding fluorescence assays.

STUDYING TFIIH INTERACTIONS WITH DNA

The arrival of TFIIH to the damaged DNA-XPC complex acts as a signal for the recruitment of repair factors RPA and XPA (Krasikova *et al.*, 2010), which will cause a reorganization of the repair bubble (Tapias *et al.*, 2004). RPA will protect the ssDNA portion of the bubble by binding to it (Krasikova *et al.*, 2010), while XPA triggers the removal of the CAK sub-complex from TFIIH (Svejstrup *et al.*, 1995) (Coin *et al.*, 2008), thus allowing XPD to carry out its unwinding activity (Sandrock & Egly, 2001) (Abdulrahman *et al.*, 2013), as well as promoting the recruitment of endonucleases XPG and XPF/ERCC1 to the repair bubble (Krasikova *et al.*, 2010).

As RPA and XPA are the next proteins to be recruited to ensure the progress of the NER reaction, we aimed to study the effect that the addition of these factors might have on TFIIH Core unwinding of the SD substrate. The addition of RPA on its own to an unwinding reaction already containing TFIIH Core caused an increment of approximately 47% of the obtained unwound product, while addition of XPA or a combination of XPA and RPA to a similar reaction did not affect unwinding by TFIIH Core (figure 101). The increase caused by addition of RPA can probably be attributed to RPA's ability to melt dsDNA in the presence of magnesium, as previously reported by the Seidel group (Kemmerich *et al.*, 2016). This effect is not observable when the protein is added together with XPA, consistent with a previously reported coordinated action of both proteins (Krasikova *et al.*, 2010). The next step in our research will be to investigate the progression of the repair bubble when endonuclease XPF/ERCC1 is added to the reaction.

Having confirmed that TFIIH Core can efficiently unwind an undamaged splayed duplex substrate, we next investigated the effect a lesion placed in the non-translocating strand (substrate SD13) or in the translocating strand (substrate SD24) would have on unwinding by the sub-complex. Compared to the unwinding of the SD substrate, amplitude of the unwinding of SD13 was 58% lower, and unwinding of SD24 was 73% lower (figure 102). Interestingly, unwinding for both the SD13 and SD24 substrates was quite similar, with only a 15% difference between the two of them, suggesting that XPD can be stalled not only by a lesion present in the translocating strand, but also by damage located in the non-translocating strand, further supporting the observations of Mathieu and co-workers (Mathieu *et al.*, 2010) and the Tessmer group (Buechner *et al.*, 2014).

Unwinding of the three substrates fit to a bi-exponential model with a short burst phase (30% of the unwinding reaction for the SD substrate, 12% for SD13 and 9% for SD24) and a long slow phase (70% for SD, 88% for SD13 and 91% for SD24) (figure 103). This extended slow phase was also observed by Constantinescu-Aruxandei and co-workers in their study of the *T. acidophilum* XPD homologue (Constantinescu-Aruxandei *et al.*, 2016), proposing an unwinding mechanism through multiple repeated steps to explain the behaviour of the archaeal XPD. In the case of the TFIIH Core sub-complex, there is a possibility that this slow phase might be caused by the binding of a second TFIIH Core to the substrate, as there is an excess of sub-complex compared to the substrate concentration; this situation is consistent with previous observations made by the Yang group in their study of TFIIH Core (Li *et al.*, 2015) and by the Spies group in their study of XPD's archaeal homologue from *Ferroplasma acidarmanus* (Honda *et al.*, 2009). Another possibility to consider given the difficulties experienced in obtaining active, Core-bound XPD is that our sample might contain a mixed population of Core sub-complexes carrying XPD helicases with different levels of activity. The fact that substrates SD13 and SD24 present similar fast and slow components (similar kinetics) in their unwinding further supports our observation that a lesion in the non-translocating can stall XPD almost as effectively as damage in the translocating strand.

5.4. Conclusions

We have analyzed the binding affinity of the 10-subunit TFIIH complex, plus TFIIH Core with suboptimal amounts of XPD bound to it and an XPD-enriched TFIIH Core construct to a variety of DNA duplexes, both unmodified and carrying an 18-nucleotide or a 3-nucleotide mismatch, by three different methods: EMSA, fluorescence anisotropy, and bulk FRET. We have also analyzed the unwinding activity of the XPD-enriched Core sub-complex in a helicase fluorescence-based assay with a variety of asymmetric duplexes presenting a 5' or a 3' overhang, and a splayed duplex substrate either undamaged or presenting a bulky adduct in the translocating or the non-translocating strand.

Although our EMSA assay seemed to initially support the Yang group's observation that TFIIH Core binds dsDNA with low affinity (Li *et al.*, 2015), further fluorescence anisotropy experiments showed similar binding affinities (in the low nM range) of the TFIIH Core

STUDYING TFIIH INTERACTIONS WITH DNA

sub-complex to substrates that included unmodified duplexes, damaged and undamaged duplexes carrying an 18-nucleotide mismatch and damaged and undamaged duplexes carrying a 3-nucleotide mismatch. Although we consistently observed a 2 to 3-fold difference in the binding affinity of the Core sub-complex to mismatched substrates carrying a bulky adduct at the site of the bubble compared to similar mismatched substrates carrying the lesion at their 5' end, this difference is not significant, and so we conclude that TFIIH Core does not show a binding preference for damaged DNA over undamaged DNA. Interestingly, the similarity of the results obtained for the binding to substrates carrying an 18-nucleotide mismatch and the ones carrying a 3-nucleotide mismatch suggest that the size of the bubble does not affect the binding of the Core sub-complex. Similarly, the analogous K_D values obtained for the TFIIH Core and XPD-enriched TFIIH Core constructs might imply a passive role of XPD in the binding of the sub-complex to DNA. Further experiments are required to confirm this last point.

The results observed in our fluorescence anisotropy assays were further confirmed in bulk FRET assays that analyzed the binding of the Core sub-complex to a damaged substrate carrying a 3-nucleotide mismatch. These assays showed that, although TFIIH Core can effectively bind a dsDNA substrate with high affinity, it doesn't seem to modify the conformation of the substrate, although a more thorough investigation is required to validate this point.

Our study of the unwinding activity of TFIIH Core showed that the sub-complex was only able to open an asymmetric duplex substrate when a 5' overhang end was available, confirming the lack of helicase activity of subunit XPB. Furthermore, the sub-complex showed an evident preference for an asymmetric duplex with a 5' overhang and a majority of purine bases in their translocating strands, compared to unwinding of a similar substrate whose translocating strand was primarily composed of pyrimidine bases. Analysis of the unwinding of an undamaged splayed duplex showed that the opening of the substrate was dependent on the concentration of the sub-complex and concentration of ATP in the reaction, but it was not significantly affected by changes in a temperature range of 25 °C to 37 °C. Similarly, the addition of the RPA and XPA proteins (the repair factors that follow the recruitment of TFIIH in NER) to the reaction did not affect the

unwinding of the splayed duplex by the Core sub-complex. Unsurprisingly, the helicase activity of the sub-complex was considerably reduced by a bulky adduct placed in the translocating strand of the splayed duplex, and, more interestingly, it was reduced almost as effectively by a lesion located in the non-translocating strand.

Future work includes:

- Anisotropy and unwinding experiments with a substrate carrying a cisplatin lesion.
- Study of the effects of our site-directed variants XPD K46R (Coin *et al.*, 1998), XPB K346R (Tirode *et al.*, 1999), XPD Y192A and XPD K196R (Mathieu *et al.*, 2013) on the binding of the sub-complex to DNA and its unwinding activity.
- Study of a complete repair bubble (XPC, TFIIH, RPA, XPA, XPF) by means of single molecule (TIRF and FCS assays).

6. Conclusions and further work

6.1. Summary

Protein XPC initiates the DNA repair reaction by detecting a mismatch in the double helix (Sugasawa *et al.*, 1998), aided by partners HR23 (paralogs A or B) and CETN2 (Sugasawa *et al.*, 1996) (Nishi *et al.*, 2005). This recognition of a distortion in the DNA rather than identification of actual damage (Maillard *et al.*, 2007) is the basis for the extended range of lesions repaired by NER, and the reason why a damage-verification step is needed (Min & Pavletich, 2007) (Shell *et al.*, 2013). The formation of a stable DNA-XPC complex will trigger the arrival of the TFIIH complex, recruited by XPC through interaction with its subunits XPB and p62 (Yokoi *et al.*, 2000) (Uchida *et al.*, 2002) (Bernardes de Jesus *et al.*, 2008). The 10-subunit TFIIH complex will now open and extend the repair bubble through the coordinated actions of helicases XPB and XPD (Schilbach *et al.*, 2017) (Constantinescu-Aruxandei *et al.*, 2016), to allow the subsequent repair factors access to the damaged site. XPD has further been proposed as the factor responsible for the verification of the NER-triggering lesion (Kuper *et al.*, 2012) (Sugasawa *et al.*, 2009) (Li *et al.*, 2015) so that the repair reaction can progress successfully.

Our interest in the NER process led us to clone, express and purify both the damage-detector protein XPC (in its monomeric and heterodimeric forms, as protein CETN2 is reportedly not required in *in vitro* experiments (Okuda *et al.*, 2015)) and the TFIIH complex (sub-complexes CAK and TFIIH Core obtained individually first, and 10-subunit TFIIH complex later obtained by co-infection with a CAK- and TFIIH Core-recombinant virus) as a first step in our investigation to characterize the binding and opening of the repair bubble at a damaged site. All constructs were cloned using the MultiBac™ tool and obtained using *Sf9* cells as an expression system.

Monomeric XPC (reportedly sufficient to bind to the DNA and start a repair reaction (Sugasawa *et al.*, 1996) (Reardon *et al.*, 1996)) was initially purified following a three-step protocol consisting of IMAC, size exclusion and cation exchange chromatography,

CONCLUSIONS AND FURTHER WORK

but the sample consistently presented two unspecific bands of about 110 kDa and 50 kDa even after adding a fourth heparin chromatographic step to our purification process. As MS analysis identified the 110 kDa band as a degraded form of XPC and the 50 kDa band as the *Spodoptera frugiperda* homologue of HR23B, hinting at the importance of their partnership, we decided to clone and express the heterodimeric XPC-HR23B complex for future experiments. Fluorescence anisotropy experiments performed with this construct showed a high affinity for a dsDNA substrate with a simulated lesion inside its sequence (S1), but unexpectedly two duplex substrates presenting an 18-nucleotide bubble (S3 and S4, with S3 also carrying the damage-like fluorescein molecule at the site of the mismatch) presented a K_D similar to the one obtained for an undamaged duplex substrate (S2) rather than a value closer to the K_D obtained in the binding of monomeric XPC to substrate S1.

Expression levels for the heterodimer XPC-HR23B were dramatically reduced after scaling-up our infections, and even after optimising our purification protocol (which was finally reduced to a two-step process, comprising only an IMAC and size exclusion chromatography) the amount of sample obtained for its biochemical characterization was seriously limited. This sample was tested for binding in both EMSA and bulk FRET assays. After our anisotropy assays with monomeric XPC showed that the protein bound to a substrate carrying an 18-nucleotide mismatch with an affinity similar to that for a DNA duplex substrate, we decided to perform our EMSA experiments with two substrates carrying a 3-nucleotide mismatch (S5 and S6, with S5 carrying a lesion-mimicking molecule at the centre of the bubble, similarly to S3) in case the 18-nucleotide bubble had been too big to be identified by XPC as a mismatch. This assay showed that XPC-HR23B bound to both S5 and S6 with an affinity that was higher than that obtained for the duplex S2, and similar to each other, suggesting that the presence of lesion did not increase XPC's binding affinity. Even more interestingly, the assay hinted the possibility that two XPC were bound to substrates S5 and S6, but not to S2, at higher concentrations of the heterodimer. Our bulk FRET experiments with substrates S7 and S8 (equivalent to S5, with S7 carrying a Cy3 dye to the left of the bubble and S8 carrying it to the right) confirmed that binding K_D of XPC-HR23B is in the nM range and similar to the K_D obtained for the binding of monomeric XPC to the substrates carrying an 18-nucleotide bubble. They also confirmed the binding of two XPC molecules to substrates

S7 and S8, as seen for substrates S5 and S6 in our EMSA assay, with binding of the second XPC molecule aided by the binding of the first XPC and the structure of our substrates altered due to the binding of the protein, which seems to be positioned closer to the Cy3 dye in substrate S8 than in S7. Addition of the TFIIH Core sub-complex did not alter XPC's binding.

The TFIIH complex is composed of the 3-subunit CAK sub-complex and the 7-subunit TFIIH Core sub-complex. While all ten subunits are required in transcription, only TFIIH Core participates in NER (Svejstrup *et al.*, 1995), and the binding of repair factor XPA will promote the removal of the CAK sub-complex (Coin *et al.*, 2008) to eliminate its inhibitory action on XPD's helicase activity (Sandrock & Egly, 2001). As our main interest in TFIIH focuses on its role as a repair factor, we initially aimed to clone the Core sub-complex on its own. For this purpose, a short and versatile purification tag (including a 6-residue His tag, a V5 epitope and a TEV cleavage site) was designed and cloned into several TFIIH Core subunits. Different combinations of tagged subunits were assembled using the multiplication module present in the MultiBac™ plasmids, which allowed the cloning of different multi-gene constructs (Bieniossek *et al.*, 2013) that were analysed for expression in *Sf9* insect cells. Infections for the successful constructs were scaled-up and a variety of chromatographic steps, buffers and conditions were tested until an easily reproducible three-step protocol (consisting of an IMAC step, a size exclusion step and a final cation exchange chromatography) was standardized. This protocol yielded up to 1 mg of a pure TFIIH Core sub-complex (with subunits p62 and p52 carrying the 6xHis-V5-TEV tag at their N-terminal end, simply referred to as TFIIH Core in this thesis) for every 2 L culture. Unfortunately, most of the XPD helicase was lost along the purification process, although its presence was confirmed by MS first and later by crosslinking assays that showed a tightly connected structure. Fluorescence anisotropy experiments performed with this construct confirmed that the TFIIH Core sub-complex can bind a dsDNA substrate with an affinity in the low nM range, and the presence of a lesion or a mismatch at the substrate did not alter that affinity significantly. These results were further supported by bulk FRET experiments, which also showed that the structure of the substrate was not altered by the binding of the sub-complex (further experiments are required to confirm this observation).

CONCLUSIONS AND FURTHER WORK

Attempts to increase the yield of XPD in our samples included the modification of our purification tag, increasing the number of His residues and introducing a spacer between the V5 epitope and the TEV cleavage site (8xHis-V5-spacer-TEV), the designing of a new, C-terminal tag exclusively for XPD (TEV-10xHis-V5), and the cloning and expression of the helicase individually to subsequently be combined with the a 6-subunit TFIIH Core construct. We also attempted the cloning and expression of the CAK sub-complex, to later obtain the complete 10-subunit TFIIH complex in a co-infection experiment with a CAK- and a TFIIH Core-recombinant viruses, as a means to stabilize XPD within the Core sub-complex. Unfortunately, none of these approaches succeeded to improve the recovery of XPD; neither did a purification in anaerobic conditions or addition of different detergents to our extraction and purification buffers. The combined use of a Strep-Tactin®XT Superflow® resin with an affinity in the pM range for a matching Twin-Strep-tag®-encoding sequence cloned into XPD finally solved these issues and allowed us to obtain the helicase, and a TFIIH Core sub-complex in which subunits p62 and p52 carrying the N-terminal 6xHis-V5-TEV tag and XPD carried a C-terminal Twin-Strep-tag® (referred to as XPD-enriched TFIIH Core in this thesis) was purified following a protocol that included an ion affinity, a size exclusion and a Streptavidin affinity chromatography steps, with all seven subunits obtained in stoichiometric amounts. This construct was tested for binding to two dsDNA substrates carrying a 3-nucleotide bubble (one of them carrying a lesion-mimicking fluorescein molecule at the centre of the bubble (S5) and the other one carrying the probe at its 5' end) in fluorescence anisotropy assays, showing an affinity only slightly higher than that observed for the binding of TFIIH Core to the same substrates. This suggests a possible passive role of XPD in the binding of the complex to DNA, although a more thorough investigation is required to confirm this point. The study of the unwinding activity of this sub-complex in fluorescence-based helicase assays showed that TFIIH Core can only open a substrate when a 5' overhang end is available, and a preference for a substrate with a majority of purine bases in the translocating strand was further observed. Unwinding was dependent on TFIIH Core and ATP concentration in the reaction, but changes in temperature within a range of 25 °C – 37 °C did not affect the opening of the substrate significantly, nor did the addition of the RPA and XPA repair factors. Lastly, our assays showed that the helicase activity of the TFIIH Core sub-complex was greatly reduced by a lesion-mimicking fluorescein molecule located in either the translocating or the non-translocating strands.

6.2. Conclusions

This thesis described the development and optimization of a protocol for the cloning and expression of the ~ 500 kDa, 10-subunit TFIIH complex, and damage-detector XPC-HR23B, following an efficient and simplified method and aided by powerful molecular biology tools. Both complex and heterodimer were cloned using the multiplication module present in the plasmids that integrate the MultiBac™ system, a BEV conceived for the production of multi-subunit eukaryotic complexes, with every individual subunit expressed in equimolar amounts (Berger *et al.*, 2004).

The design of the synthetic genes for subunits p52 and p8, p44 and p34 and cdk7 and cyclin H as single units, with each pair of genes separated by an autocleavable 2A-like peptide (Luke *et al.*, 2009) (with an observed cleaving efficiency of approximately 90%) further simplified the cloning of the TFIIH complex by reducing the number of steps required to assemble the desired multi-gene construct.

Purification of the complex was assisted by the introduction of the multi-function 6xHis-V5-TEV tag (developed and successfully tested in our lab in a variety of constructs using both bacteria and insect cells as an expression system) into several Core subunits. Moreover, the versatility of the MultiBac™ system allowed the quick assembly of new, different multi-gene constructs by relocating the tag within the sub-complex without requiring the cloning of the whole complex *de novo* by using intermediate constructs obtained in previous cloning processes.

Our most successful purification was achieved with a TFIIH Core sub-complex in which subunits p52 and p62 carried the 6xHis-V5-TEV tag at their N-terminal end, obtaining up to 1 mg of pure sub-complex from a 2 L culture after a 3-step protocol consisting of an IMAC, a size exclusion and a cation exchange chromatography. This yield was higher than any reported result until the publication of the work by the Yang group (Li *et al.*, 2015) two years after the start of this project. The obtaining of the TFIIH Core sub-complex was further optimised to improve the recovery of helicase XPD, finally agreeing on a 3-step IMAC – size exclusion – streptavidin affinity chromatography purification protocol for a TFIIH Core sub-complex in which subunits p52 and p62 still carried the N-

CONCLUSIONS AND FURTHER WORK

terminal 6xHis-V5-TEV tag and subunit XPD carried a C-terminal Twin-Strep-tag®.

Analysis of the binding of monomeric XPC and heterodimeric XPC-HR23B to a variety of dsDNA substrates in EMSA, fluorescence anisotropy and bulk FRET assays led us to conclude that XPC binds preferentially to substrates carrying a mismatch, with an affinity in the low nM range. However, neither the size of that mismatch (18-nucleotide or 3-nucleotide bubbles) nor the additional presence of a lesion at the bubble altered this binding affinity. The presence of partner HR23B in the sample didn't increase the binding affinity of XPC either. Our bulk FRET experiments suggest that two molecules of XPC can bind a duplex substrate carrying a 3-nucleotide mismatch and a lesion-like fluorescein molecule at the centre of the bubble with high cooperativity, and this binding is able to change the conformation of our substrate. Addition of the TFIIH Core sub-complex to this reaction did not alter the binding. Unfortunately, the experiments so far performed with both monomeric XPC and HR23B-XPC were severely limited by the amount of pure protein/heterodimer obtained in our purifications, and a more thorough research will be needed to confirm our preliminary observations.

We also tested the TFIIH Core sub-complex for binding to the same range of substrates tested with XPC and XPC-HR23B in EMSA, fluorescence anisotropy and bulk FRET assays. Our research revealed that the sub-complex binds to duplex DNA with an affinity in the nM range; compared to the binding affinity observed for XPC, TFIIH Core binds to a duplex carrying a lesion-like fluorescein molecule at the centre of its sequence with an affinity at least three times lower. The presence of damage or a mismatch at the substrate did not alter the sub-complex binding affinity, and K_D values obtained for an undamaged duplex and substrates carrying an 18-nucleotide bubble were similar for both the sub-complex and damage detector XPC. Bulk FRET assays further supported these observations, and also suggested that the binding of the sub-complex did not alter the structure of the substrates (further experiments are required to confirm this last observation).

The unwinding activity of TFIIH Core was evaluated in a fluorescence-based helicase assay that confirmed that the sub-complex can only open a substrate when a 5' overhang end (but not a 3' end) is available. The opening of a splayed duplex substrate was

observed to be dependent on both TFIIH Core and ATP concentration, but it was not significantly altered by changes in temperature within a range of 25 °C – 37 °C. Addition of repair factors XPA and RPA did not modify the unwinding activity of the sub-complex either. Lastly, analysis of the opening of two splayed duplexes carrying a fluorescein molecule mimicking a bulky adduct in either the translocating or the non-translocating strand revealed that both lesions can successfully stall helicase XPD, showing a considerably reduced opening of the substrates compared to opening of the undamaged splayed duplex.

6.3. Future work

The results obtained in the course of this PhD offer interesting and challenging perspectives for the future of the project. A pressing issue to tackle is the low expression observed in our XPC-HR23B infections and consequent reduced yield in our purifications. A possible solution to this problem is the change of expression strain, from our current Sf9 cells to High Five™ insect cells, which reportedly offer 5- to 10-fold higher expression in the case of secreted heterologous protein. This change will also require the optimization of our expression and purification protocols, adapting them to the amount of heterodimer obtained from our High Five™ infections. The cloning of a gene encoding XPC's partner HR23B into our TCP-tagged monomeric XPC will also be required before continuing our anisotropy and FRET assays with this construct.

The publication of the cryoEM structure of TFIIH Core as part of the PIC complex (Schilbach *et al.*, 2017) offered new, unprecedented insights about the sub-complex which, together with the structure of XPC's yeast homologue (Min & Pavletich, 2007), allowed our group to develop a model describing the potential unravelling of the repair reaction at its initial stages (figures 104 and 105). A similar model, although offering less detail regarding the role of XPB, has very recently been published by the Schärer group (Mu *et al.*, 2018).

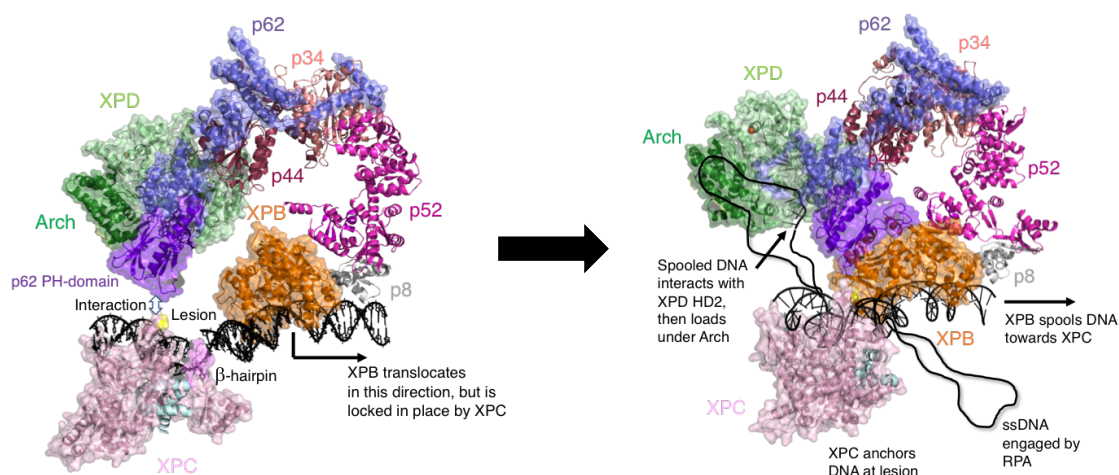


Figure 104: Unravelling the repair bubble.

Structures of TFIIH Core (5OQJ, (Schilbach *et al.*, 2017)) and XPC (2QSG, (Min & Pavletich, 2007)) engaged in a damaged dsDNA substrate are here brought together to model the initiation of NER. XPC's interacting domain with subunit p62 is missing from the structure, but it will potentially extend towards the PH domain of p62. XPC locks XPB in place; as the enzyme will try to translocate in the 3' to 5' direction, the bubble will be extended until ssDNA is engaged by XPD, which will finally unwind the DNA to allow subsequent repair factors access to the lesion.

Our group already possesses many of the tools that would allow the experimental validation of the theoretical model described in figures 104 and 105. We already have not only a TCP-tagged monomeric XPC, but also a TCP-tagged TFIIH Core sub-complex, which have been successfully labelled in our lab. Both TCP-tagged XPC and TFIIH Core could be employed in fluorescence anisotropy and FRET assays with unmodified and mismatched duplex substrates to further characterize the binding of XPC and TFIIH Core. In this sense, it would be very interesting to continue our study of the binding of TFIIH Core to a DNA-bound XPC-HR23B complex, ideally using a cisplatinated substrate, as this lesion is one of the natural targets of NER. Single molecule experiments such as TIRF and FCS would further help us characterize this binding and the changes undergone by the substrate as a consequence of it. Fluorescence-based and radioactive helicase activity assays performed with the cisplatinated substrate would also help us to advance in our characterization of the unwinding activity of the TFIIH Core sub-complex, particularly in the study of the ability of a lesion located in the non-translocating strand to stall helicase XPD almost as

efficiently as a lesion located in the translocating strand. Ideally, these studies would lead to the observation and characterization of the changes taking place in a damaged substrate as a full repair reaction takes place, including mismatch-detection and DNA binding by XPC-HR23B, substrate-opening and potential damage verification by TFIIH Core, reorganization of the bubble caused by addition of XPA and RPA and 5' damaged strand nicking by endonuclease XPF (our lab already has access to all these factors). A very interesting complement to this research would be to investigate the effect on DNA binding and unwinding of TFIIH Core mutant variants affecting the Walker A domain (XPD K46R (Coin *et al.*, 1998) and XPB K346R (Tirode *et al.*, 1999)) and affecting XPD's DNA binding pocket (XPD Y192A and XPD K196R (Mathieu *et al.*, 2013)) (cloning of these mutants is currently under way in our lab).

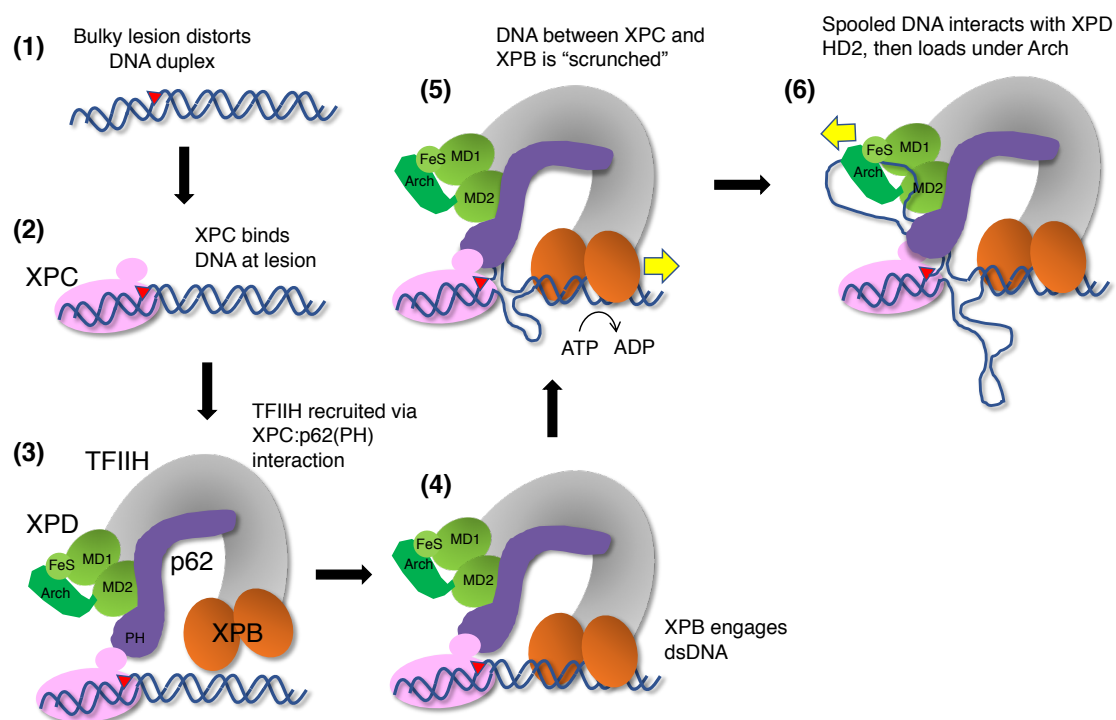


Figure 105: Hypothetical model describing the first stages of a GGR reaction.

The distortion caused in the DNA by the appearance of a lesion (1) will be detected by XPC, which will bind to the damaged site (2) and will recruit the TFIIH complex through interactions with subunits p62 and XPB (3). TFIIH subunit XPB will now engage dsDNA (4) and translocate in the 3' to 5' direction in an ATP-dependent manner, but as the enzyme is locked by XPC this will cause an initial opening of the bubble, with XPB spooling DNA towards XPC (5). The spooled ssDNA will finally reach XPD (6), interacting first with HD2 before the interface between the Arch and

CONCLUSIONS AND FURTHER WORK

4FeS domains opens temporarily to allow ssDNA to pass through the channel formed by Arch, 4FeS and HD1 (Constantinescu-Aruxandei *et al.*, 2016). ssDNA is finally engaged by XPD's HD1, and the helicase will unwind the dsDNA until it becomes stalled by the lesion, potentially acting as a damage-verification factor (Sugasawa *et al.*, 2009) (Li *et al.*, 2015).

BIBLIOGRAPHY

- Abdulrahman, W., Iltis, I., Radu, L., Braun, C., Maglott-Roth, A., Giraudon, C., ... Poterszman, A. (2013). ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8), E633-42.
- Adamczewski, J. P., Rossignol, M., Tassan, J. P., Nigg, E. a, Moncollin, V., & Egly, J. M. (1996). MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH. *The EMBO Journal*, 15(8), 1877–84.
- Adams, S. R., Campbell, R. E., Gross, L. A., Martin, B. R., Walkup, G. K., Yao, Y., ... Tsien, R. Y. (2002). New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *Journal of the American Chemical Society*, 124(21), 6063–6076.
- Ahn, B. (2000). A physical interaction of UvrD with nucleotide excision repair protein UvrB. *Molecules and Cells*, 10(5), 592–7.
- Airenne, K. J., Hu, Y. C., Kost, T. A., Smith, R. H., Kotin, R. M., Ono, C., ... Ylä-Herttuala, S. (2013). Baculovirus: an insect-derived vector for diverse gene transfer applications. *Molecular Therapy*, 21(4), 739–749.
- Akoulitchev, S., Chuikov, S., & Reinberg, D. (2000). TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature*, 407(6800), 102–106.
- Alekseev, S., Ayadi, M., Brino, L., Egly, J. M., Larsen, A. K., & Coin, F. (2014). A small molecule screen identifies an inhibitor of DNA repair inducing the degradation of TFIIH and the chemosensitization of tumor cells to platinum. *Chemistry and Biology*, 21(3), 398–407.
- Araki, M., Masutani, C., Maekawa, T., Watanabe, Y., Yamada, A., Kusumoto, R., ... Hanaoka, F. (2000). Reconstitution of damage DNA excision reaction from SV40 minichromosomes with purified nucleotide excision repair proteins. *Mutation Research - DNA Repair*, 459(2), 147–160.
- Assenberg, R., Wan, P. T., Geisse, S., & Mayr, L. M. (2013). Advances in recombinant protein expression for use in pharmaceutical research. *Current Opinion in Structural Biology*, 23(3), 393–402.
- Bartlett, E. J., Brissett, N. C., & Doherty, A. J. (2013). Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. *Proceedings of the National Academy of Sciences of the United States of America*, 110(22), E1984-91.

BIBLIOGRAPHY

- Bedez, F., Linard, B., Brochet, X., Ripp, R., Thompson, J. D., Moras, D., ... Poch, O. (2013). Functional insights into the core-TFIIH from a comparative survey. *Genomics*, *101*(3), 178–186.
- Berger, I., Fitzgerald, D. J., & Richmond, T. J. (2004). Baculovirus expression system for heterologous multiprotein complexes. *Nature Biotechnology*, *22*(12), 1583–1587.
- Bergink, S., Toussaint, W., Luijsterburg, M. S., Dinant, C., Alekseev, S., Hoeijmakers, J. H. J., ... Vermeulen, W. (2012). Recognition of DNA damage by XPC coincides with disruption of the XPC-RAD23 complex. *Journal of Cell Biology*, *196*(6), 681–688.
- Berico, P., & Coin, F. (2017). Is TFIIH the new Achilles heel of cancer cells? *Transcription*, *1264*(December), 1–5.
- Bernardes de Jesus, B. M., Bjørås, M., Coin, F., & Egly, J. M. (2008). Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC. *Molecular and Cellular Biology*, *28*(23), 7225–7235.
- Bieniossek, C., Garzoni, F., Berger, I. (2013). MultiBac™ - multi-protein expression in insect cells (4.1, Vol. 49). Geneva Biotech.
- Bieniossek, C., Imasaki, T., Takagi, Y., & Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. *Trends in Biochemical Sciences*, *37*(2), 49–57.
- Bieniossek, C., Nie, Y., Frey, D., Olieric, N., Schaffitzel, C., Collinson, I., ... Berger, I. (2009). Automated unrestricted multigene recombineering for multiprotein complex production. *Nature Methods*, *6*(6), 447–450.
- Blau, J., Xiao, H., McCracken, S., O'Hare, P., Greenblatt, J., & Bentley, D. (1996). Three functional classes of transcriptional activation domain. *Molecular and Cellular Biology*, *16*(5), 2044–2055.
- Bootsma, D., & Hoeijmakers, J. H. J. (1993). Engagement with transcription. *Nature Letters*, *363*.
- Boshra, H., Lorenzo, G., Busquets, N., & Brun, A. (2011). Rift Valley fever: recent insights into pathogenesis and prevention. *Journal of Virology*, *85*(13), 6098–6105.
- Botta, E., Nardo, T., Lehmann, A. R., Egly, J., Pedrini, A. M., & Stefanini, M. (2002). Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy. *Human Molecular Genetics*, *11*(23), 2919–2928.
- Buechner, C. N., Heil, K., Michels, G., Carell, T., Kisker, C., & Tessmer, I. (2014). Strand-specific recognition of DNA damages by XPD provides insights into nucleotide excision repair substrate versatility. *Journal of Biological Chemistry*, *289*(6), 3613–3624.

- Bunick, C. G., Miller, M. R., Fuller, B. E., Fanning, E., & Chazin, W. J. (2006). Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. *Biochemistry*, *45*(50), 14965–14979.
- Camenisch, U., Träutlein, D., Clement, F. C., Fei, J., Leitenstorfer, A., Ferrando-May, E., & Naegeli, H. (2009). Two-stage dynamic DNA quality check by xeroderma pigmentosum group C protein. *The EMBO Journal*, *28*(16), 2387–2399.
- Chabot, P. R., Raiola, L., Lussier-Price, M., Morse, T., Arseneault, G., Archambault, J., & Omichinski, J. G. (2014). Structural and functional characterization of a complex between the acidic transactivation domain of EBNA2 and the Tfb1/p62 subunit of TFIIH. *PLoS Pathogens*, *10*(3).
- Chen, J., Larochelle, S., Li, X., & Suter, B. (2003). Xpd/Ercc2 regulates CAK activity and mitotic progression. *Nature*, *424*(6945), 228–232.
- Chung, R., Angwan, V., Patil, S. P., Dudeja, V., Dawra, R. K., Banerjee, S., ... Saluja, A. K. (2012). A preclinical evaluation of Minnelide as a therapeutic agent against pancreatic cancer. *Science Translational Medicine*, *4*(156), 156ra139.
- Coin, F., Auriol, J., Tapias, A., Clivio, P., Vermeulen, W., & Egly, J. M. (2004). Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity. *The EMBO Journal*, *23*(24), 4835–4846.
- Coin, F., De Santis, L. P., Nardo, T., Zlobinskaya, O., Stefanini, M., & Egly, J. M. (2006). p8/TTD-A as a repair-specific TFIIH subunit. *Molecular Cell*, *21*(2), 215–226.
- Coin, F., Marinoni, J. C., Rodolfo, C., Fribourg, S., Pedrini, A. M., & Egly, J. M. (1998). Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nature Genetics*, *20*(2), 184–188.
- Coin, F., Oksenysh, V., & Egly, J. M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Molecular Cell*, *26*(2), 245–256.
- Coin, F., Oksenysh, V., Mocquet, V., Groh, S., Blattner, C., & Egly, J. M. (2008). Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Molecular Cell*, *31*(1), 9–20.
- Compe, E., & Egly, J.-M. (2012). TFIIH: when transcription met DNA repair. *Nature Reviews Molecular Cell Biology*, *13*(7), 476–476.
- Compe, E., & Egly, J.-M. (2016). Nucleotide excision repair and transcriptional regulation: TFIIH and beyond. *Annual Review of Biochemistry*, *85*(1), 265–290.

BIBLIOGRAPHY

- Constantinescu-Aruxandei, D., Petrovic-Stojanovska, B., Penedo, J. C., White, M. F., & Naismith, J. H. (2016). Mechanism of DNA loading by the DNA repair helicase XPD. *Nucleic Acids Research*, *44*(6), 2806–2815.
- Cubeddu, L., & White, M. F. (2005). DNA damage detection by an archaeal single-stranded DNA-binding protein. *Journal of Molecular Biology*, *353*(3), 507–516.
- Dabholkar, M., Thornton, K., Vionnet, J., Bostick-Bruton, F., Yu, J. J., & Reed, E. (2000). Increased mRNA levels of xeroderma pigmentosum complementation group B (XPB) and Cockayne's syndrome complementation group B (CSB) without increased mRNA levels of multidrug-resistance gene (MDR1) or metallothionein-II (MT-II) in platinum-resistant human . *Biochemical Pharmacology*, *60*(11), 1611–1619.
- de Laat, W., Jaspers, N., & Hoeijmakers, J. (1999). Molecular mechanism of nucleotide excision repair. *Genes & Development*, *13*(7), 768–785.
- Desai, A., Yan, Y., & Gerson, S. L. (2018). Advances in therapeutic targeting of the DNA damage response in cancer. *DNA Repair*, *66–67*(April), 24–29.
- Diderich, K., Alanazi, M., & Hoeijmakers, J. H. J. (2011). Premature aging and cancer in nucleotide excision repair-disorders. *DNA Repair*, *10*(7), 772–780.
- Dip, R., Camenisch, U., & Naegeli, H. (2004). Mechanisms of DNA damage recognition and strand discrimination in human nucleotide excision repair. *DNA Repair*, *3*(11), 1409–1423.
- Dubaele, S., De Santis, L. P., Bienstock, R. J., Keriél, A., Stefanini, M., Van Houten, B., & Egly, J. M. (2003). Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Molecular Cell*, *11*(6), 1635–1646.
- Eisen, J. A., & Hanawalt, P. C. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutation Research* (Vol. 435).
- Eme, L., Spang, A., Lombard, J., Stairs, C. W., & Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. *Nature Reviews Microbiology*, *15*(12), 711–723.
- Epshtein, V., Kamarthapu, V., McGary, K., Svetlov, V., Ueberheide, B., Proshkin, S., Mironov, A., & Nudler, E. (2014). UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature*, *505*(7483), 372-377.
- Evans, E., Moggs, J. G., Hwang, J. R., Egly, J. M., & Wood, R. D. (1997). Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *The EMBO Journal*, *16*(21), 6559–6573.

- Fairman-Williams, M. E., Guenther, U. P., & Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology*, 20(3), 313–324.
- Fan, L., Arvai, A. S., Cooper, P. K., Iwai, S., Hanaoka, F., & Tainer, J. A. (2006). Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Molecular Cell*, 22(1), 27–37.
- Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., ... Tainer, J. A. (2008). XPD Helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell*, 133(5), 789–800.
- Feaver, W. J., Svejstrup, J. Q., Bardwell, L., Bardwell, A. J., Buratowski, S., Gulyas, K. D., ... Kornberg, R. D. (1993). Dual roles of a multiprotein complex from *S. cerevisiae* in transcription and DNA repair. *Cell*, 75(7), 1379–1387.
- Fishburn, J., Tomko, E., Galburt, E., & Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(13), 3961–6.
- Flores, O., Lu, H., & Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 87(23), 9158–9162.
- Fortini, P., Parlanti, E., Sidorkina, O. M., Laval, J., & Dogliotti, E. (1999). The type of DNA glycosylase determines the base excision repair pathway in mammalian cells. *Journal of Biological Chemistry*, 274(21), 15230–15236.
- Fuertes, M. A., Alonso, C., & Perez, J. M. (2003). Biochemical modulation of cisplatin mechanisms of action: enhancement of antitumor activity and circumvention of drug resistance. *Chemical Reviews*, 103(3), 645–662.
- Gaillard, P. H., Martini, E. M., Kaufman, P. D., Stillman, B., Moustacchi, E., & Almouzni, G. (1996). Chromatin assembly coupled to DNA repair: a new role for chromatin assembly factor I. *Cell*, 86(6), 887–96.
- Gary, R., Ludwig, D. L., Cornelius, H. L., MacInnes, M. A., & Park, M. S. (1997). 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. *Journal of Biological Chemistry*, 272(39), 24522–24529.
- Gervais, V., Lamour, V., Jawhari, A., Frindel, F., Wasielewski, E., Dubaele, S., ... Poterszman, A. (2004). TFIIH contains a PH domain involved in DNA nucleotide excision repair. *Nature Structural & Molecular Biology*, 11(7), 616–622.

BIBLIOGRAPHY

- Giglia-Mari, G., Coin, F., Ranish, J. A., Hoogstraten, D., Theil, A., Wijgers, N., ... Vermeulen, W. (2004). A new, tenth subunit TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nature Genetics*, *36*(7), 714–719.
- Giglia-Mari, G., Miquel, C., Theil, A. F., Mari, P. O., Hoogstraten, D., Ng, J. M. Y., ... Vermeulen, W. (2006). Dynamic interaction of TTDA with TFIIH is stabilized by nucleotide excision repair in living cells. *PLoS Biology*, *4*(6), 0952–0963.
- Gillet, L. C. J., & Schärer, O. D. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chemical Reviews*, *106*(2), 253–276.
- Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., & Lieber, M. R. (1998). DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Molecular Cell*, *2*(4), 477–484.
- Greber, B. J., Nguyen, T. H. D., Fang, J., Afonine, P. V., Adams, P. D., & Nogales, E. (2017). The cryo-electron microscopy structure of human transcription factor IIH. *Nature*, *549*(7672):414-417
- Griffin, B. A., Adams, S. R., & Tsien, R. Y. (1998). Specific covalent labeling of recombinant protein molecules inside live cells. *Science*, *281*(1998), 269.
- Hanawalt, P. C. (1987). Preferential DNA repair in expressed genes. *Environmental Health Perspectives*, *76*(6), 9–14.
- Hanawalt, P. C., & Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Reviews. Molecular Cell Biology*, *9*(12), 958–970.
- Hawtin, R. E., Zarkowska, T., Arnold, K., Thomas, C. J., Gooday, G. W., King, L. A., ... Possee, R. D. (1997). Liquefaction of *Autographa californica* nucleopolyhedrovirus-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. *Virology*, *238*(2), 243–253.
- He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., & Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. *Nature*, *533*(7603), 359–65.
- Helleday, T., Petermann, E., Lundin, C., Hodgson, B., & Sharma, R. A. (2008). DNA repair pathways as targets for cancer therapy. *Nature Reviews Cancer*, *14*(3), 1291–1295.
- Hill A.V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *The Journal of Physiology*.
- Hilton, B., Gopal, S., Xu, L., Mazumder, S., Musich, P. R., Cho, B. P., & Zou, Y. Z. (2016). Dissociation dynamics of XPC-RAD23B from damaged DNA is a determining factor of NER efficiency. *PLoS ONE*, *11*(6), 1–21.

- Hirsh, A. E., & Fraser, H. B. (2001). Protein dispensability and rate of evolution. *Nature*, *411*(6841), 1046–1049.
- Holstege, F. C., van der Vliet, P. C., & Timmers, H. T. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *The EMBO Journal*, *15*(7), 1666–1677.
- Honda, M., Park, J., Pugh, R. A., Ha, T., & Spies, M. (2009). Single-molecule analysis reveals differential effect of ssDNA-binding proteins on DNA translocation by XPD helicase. *Molecular Cell*, *35*(5), 694–703.
- Hoogstraten, D., Bergink, S., Ng, J. M. Y., Verbiest, V. H. M., Luijsterburg, M. S., Geverts, B., ... Houtsmuller, A. B. (2008). Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. *Journal of Cell Science*, *121*(23), 3991–3991.
- Ishino, S., Nishi, Y., Oda, S., Uemori, T., Sagara, T., Takatsu, N., ... Ishino, Y. (2016). Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. *Nucleic Acids Research*, *44*(7), 2977–2986.
- Ito, S., Tan, L. J., Andoh, D., Narita, T., Seki, M., Hirano, Y., ... Tanaka, K. (2010). MMXD, a TFIIH-independent XPD-MMS19 protein complex involved in chromosome segregation. *Molecular Cell*, *39*(4), 632–640.
- Janićijević, A., Sugasawa, K., Shimizu, Y., Hanaoka, F., Wijgers, N., Djurica, M., ... Wyman, C. (2003). DNA bending by the human damage recognition complex XPC-HR23B. *DNA Repair*, *2*(3), 325–336.
- Jasin, M., & Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harbor Perspectives in Biology*, *5*(11), 1–18.
- Kainov, D. E., Vitorino, M., Cavarelli, J., Poterszman, A., & Egly, J.-M. (2008). Structural basis for group A trichothiodystrophy. *Nature Structural & Molecular Biology*, *15*(9), 980–984.
- Kainulainen, M., Habjan, M., Hubel, P., Busch, L., Lau, S., Colinge, J., ... Weber, F. (2014). Virulence factor NSs of Rift Valley fever virus recruits the F-box protein FBXO3 to degrade subunit p62 of general transcription factor TFIIH. *Journal of Virology*, *88*(6), 3464–3473.
- Kalveram, B., Lihoradova, O., & Ikegami, T. (2011). NSs Protein of Rift Valley fever virus promotes posttranslational downregulation of the TFIIH subunit p62. *Journal of Virology*, *85*(13), 6234–6243.
- Kass, E. M., & Jasin, M. (2010). Collaboration and competition between DNA double-strand break repair pathways. *FEBS Letters*, *584*(17), 3703–3708.

BIBLIOGRAPHY

- Kellenberger, E., Dominguez, C., Fribourg, S., Wasielewski, E., Moras, D., Poterszman, A., ... Kieffer, B. (2005). Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions. *Journal of Biological Chemistry*, *280*(21), 20785–20792.
- Kelman, Z., & White, M. F. (2005). Archaeal DNA replication and repair. *Current Opinion in Microbiology*, *8*(6), 669–676.
- Kemmerich, F. E., Daldrop, P., Pinto, C., Levikova, M., Cejka, P., & Seidel, R. (2016). Force regulated dynamics of RPA on a DNA fork. *Nucleic Acids Research*, *44*(12), 5837–5848.
- Kim, J. S., Saint-André, C., Lim, H. S., Hwang, C. S., Egly, J. M., & Cho, Y. (2015). Crystal structure of the Rad3/XPD regulatory domain of Ssl1/p44. *Journal of Biological Chemistry*, *290*(13), 8321–8330.
- Kisker, C., Kuper, J., & Van Houten, B. (2013). Prokaryotic nucleotide excision repair. *Cold Spring Harbor Perspectives in Biology*, *5*(3), 1–18.
- Krasikova, Y. S., Rechkunova, N. I., Maltseva, E. A., Petrusheva, I. O., & Lavrik, O. I. (2010). Localization of xeroderma pigmentosum group A protein and replication protein A on damaged DNA in nucleotide excision repair. *Nucleic Acids Research*, *38*(22), 8083–8094.
- Krokan, H. E., Standal, R., & Slupphaug, G. (1997). DNA glycosylases in the base excision repair of DNA. *Biochemical Journal*, *325*(1), 1–16.
- Kuper, J., Braun, C., Elias, A., Michels, G., Sauer, F., Schmitt, D. R., ... Kisker, C. (2014). In TFIIH, XPD helicase is exclusively devoted to DNA repair. *PLoS Biology*, *12*(9).
- Kuper, J., Wolski, S. C., Michels, G., & Kisker, C. (2012). Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *The EMBO Journal*, *31*(2), 494–502.
- Kusumoto, R., Masutani, C., Sugasawa, K., Iwai, S., Araki, M., Uchida, A., ... Hanaoka, F. (2001). Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. *Mutation Research - DNA Repair*, *485*(3), 219–227.
- Kwiatkowski, N., Zhang, T., Rahl, P. B., Abraham, B. J., Reddy, J., Ficarro, S. B., ... Gray, N. S. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature*, *511*(7511), 616–620.
- Labahn, J., Schärer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L., & Ellenberger, T. E. (1996). Structural basis for the excision repair of alkylation-damaged DNA. *Cell*, *86*(2), 321–329.

- Langlois, C., Mas, C., Di Lello, P., Miller Jenkins, L. M., Legault, P., & Omichinski, J. G. (2008). NMR structure of the complex between the Tfb1 subunit of TFIIH and the activation domain of VP16: structural similarities between VP16 and p53. *Journal of the American Chemical Society*, *130*(32), 10596–10604.
- Le May, N., Mota-Fernandes, D., Vélez-Cruz, R., Iltis, I., Biard, D., & Egly, J. M. (2010). NER Factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Molecular Cell*, *38*(1), 54–66.
- Lehmann, A. R. (2003). DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie*, *85*(11), 1101–1111.
- Li, C. L., Golebiowski, F. M., Onishi, Y., Samara, N. L., Sugasawa, K., & Yang, W. (2015). Tripartite DNA lesion recognition and verification by XPC, TFIIH, and XPA in nucleotide excision repair. *Molecular Cell*, *59*(6), 1025–1034.
- Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Research*, *18*(1), 85–98.
- Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *Journal of Biological Chemistry*, *283*(1), 1–5.
- Liu, H., Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., ... White, M. F. (2008). Structure of the DNA repair helicase XPD. *Cell*, *133*(5), 801–812.
- Lorenz, M., & Diekmann, S. (2006). Distance determination in protein-DNA complexes using fluorescence resonance energy transfer. *Methods in Molecular Biology*, *335*, 243–255.
- Lu, H., Zawel, L., Fisher, L., Egly, J. M., & Reinberg, D. (1992). Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature*, *358*(6388), 641–645.
- Luke, G. A., Escuin, H., Felipe, P. De, & Ryan, M. D. (2009). 2A to the fore – research, technology and applications. *Biotechnology and Genetic Engineering Reviews*, *26*(1), 223–260.
- Luo, J., Cimermancic, P., Viswanath, S., Ebmeier, C. C., Kim, B., Dehecq, M., ... Ranish, J. (2015a). Architecture of the human and yeast general transcription and DNA repair factor TFIIH. *Molecular Cell*, *59*(5), 794–806.
- Lyles, D. S. (2000). Cytopathogenesis and inhibition of host gene expression by RNA viruses. *Microbiology and Molecular Biology Reviews*, *64*(4), 709–724.
- Maillard, O., Solyom, S., & Naegeli, H. (2007). An aromatic sensor with aversion to damaged strands confers versatility to DNA repair. *PLoS Biology*, *5*(4), 717–728.

BIBLIOGRAPHY

- Marteijn, J. a, Lans, H., Vermeulen, W., & Hoeijmakers, J. H. J. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews. Molecular Cell Biology*, *15*(7), 465–81.
- Mathieu, N., Kaczmarek, N., & Naegeli, H. (2010). Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17545–17550.
- Mathieu, N., Kaczmarek, N., Rüthemann, P., Luch, A., & Naegeli, H. (2013). DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. *Current Biology*, *23*(3), 204–212.
- Matthews, L. A., & Simmons, L. A. (2014). Bacterial nonhomologous end joining requires teamwork. *Journal of Bacteriology*, *196*(19), 3363–3365.
- Mckenzie, J., & El-guindy, A. (2015). Epstein Barr Virus Volume 2, *391*, 237–261.
- Mellon, I., Spivak, G., & Hanawalt, P. C. (1987). Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell*, *51*(2), 241–249.
- Mimitou, E. P., & Symington, L. S. (2009). Nucleases and helicases take center stage in homologous recombination. *Trends in Biochemical Sciences*, *34*(5), 264–272.
- Min, J.-H., & Pavletich, N. P. (2007). Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, *449*(7162), 570–5.
- Mishina, Y., Duguid, E. M., & He, C. (2006). Direct reversal of DNA alkylation damage. *Chemical Reviews*, *106*(2), 215–232.
- Mocquet, V., Lainé, J. P., Riedl, T., Yajin, Z., Lee, M. Y., & Egly, J. M. (2008). Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *The EMBO Journal*, *27*(1), 155–167.
- Morgan, D. O. (1995). Principles of CDK regulation. *Nature*, *374*(6518):131-4.
- Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L. H. F., & Fousteri, M. I. (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III α in a cell-cycle-specific manner. *Molecular Cell*, *27*(2), 311–323.
- Moser, J., Volker, M., Kool, H., Alekseev, S., Vrieling, H., Yasui, A., ... Mullenders, L. H. F. (2005). The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA Repair*, *4*(5), 571–582.

- Mu, H., Geacintov, N. E., Broyde, S., Yeo, J.-E., & Schärer, O. D. (2018). Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair. *DNA Repair*, 71, 33–42.
- Naegeli, H., & Sugasawa, K. (2011). The xeroderma pigmentosum pathway: decision tree analysis of DNA quality. *DNA Repair*, 10(7), 673–683.
- Nazarenko, I., Pires, R., Lowe, B., & Rashtchian, A. (2002). Effect of primary and secondary structure of oligodeoxyribonucleotides on the fluorescent properties of conjugated dyes. *Nucleic Acids Research*, 30(9), 2089–2095.
- Ng, J. M. Y., Vermeulen, W., van der Horst, G. T. J., Bergink, S., Sugasawa, K., Vrieling, H., & Hoeijmakers, J. H. J. (2003). A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. *Genes & Development*, 17(13), 1630.
- Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., ... Hanaoka, F. (2005). Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Molecular and Cellular Biology*, 25(13), 5664–5674.
- Nishi, R., Sakai, W., Tone, D., Hanaoka, F., & Sugasawa, K. (2013). Structure-function analysis of the EF-hand protein centrin-2 for its intracellular localization and nucleotide excision repair. *Nucleic Acids Research*, 41(14), 6917–6929.
- Nocentini, S., Coin, F., Saijo, M., Tanaka, K., & Egly, J. M. (1997). DNA damage recognition by XPA protein promotes efficient recruitment of transcription factor II H. *Journal of Biological Chemistry*, 272(37), 22991–22994.
- Ogi, T., Limsirichaikul, S., Overmeer, R. M., Volker, M., Takenaka, K., Cloney, R., ... Lehmann, A. R. (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Molecular Cell*, 37(5), 714–727.
- Öğrünç, M., Becker, D.F., Ragsdale, S.W., Sancar, A. (1998). Nucleotide excision repair in the third kingdom. *Journal of Bacteriology*, 180(21), 5796–5798.
- Ohkuma, Y., & Roeder, R. G. (1994). Regulation of TFIIH ATPase and kinase activities by TFIIIE during active initiation complex formation. *Nature*, 368(6467), 160–163.
- Oksenysh, V., De Jesus, B. B., Zhovmer, A., Egly, J. M., & Coin, F. (2009). Molecular insights into the recruitment of TFIIH to sites of DNA damage. *The EMBO Journal*, 28(19), 2971–2980.
- Okuda, M., Kinoshita, M., Kakumu, E., Sugasawa, K., & Nishimura, Y. (2015). Structural insight into the mechanism of TFIIH recognition by the acidic string of the nucleotide excision repair factor XPC. *Structure*, 23(10), 1827–1837.

BIBLIOGRAPHY

- Okuda, M., Nakazawa, Y., Guo, C., Ogi, T., & Nishimura, Y. (2017). Common TFIIH recruitment mechanism in global genome and transcription-coupled repair subpathways. *Nucleic Acids Research*, *45*(22), 13043–13055.
- Okuda, M., Tanaka, A., Satoh, M., Mizuta, S., Takazawa, M., Ohkuma, Y., & Nishimura, Y. (2008). Structural insight into the TFIIE-TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. *The EMBO Journal*, *27*(7), 1161–1171.
- Okuda, Y., Nishi, R., Ng, J. M. Y., Vermeulen, W., Van Der Horst, G. T. J., Mori, T., ... Sugawara, K. (2004). Relative levels of the two mammalian Rad23 homologs determine composition and stability of the xeroderma pigmentosum group C protein complex. *DNA Repair*, *3*(10), 1285–1295.
- Orphanides, G., Lagrange, T., & Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes & Development*, *10*(21), 2657–83.
- Ossipow, V., Tassan, J. P., Nigg, E. A., & Schibler, U. (1995). A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell*, *83*(1), 137–146.
- Park, H., Hanson, G. T., Duff, S. R., & Selvin, P. R. (2004). Nanometre localization of single ReAsH molecules. *Journal of Microscopy*, *216*(3), 199–205.
- Peng, Y., Wang, H., Santana-santos, L., & Kisker, C. (2011). *Chemical Carcinogenesis*. (T. M. Penning, Ed.). Springer New York.
- Polo, S. E., Roche, D., & Almouzni, G. (2006). New histone incorporation marks sites of UV repair in human cells. *Cell*, *127*(3), 481–493.
- Pullara, F., Guerrero-Santoro, J., Calero, M., Zhang, Q., Peng, Y., Spåhr, H., ... Levine, A. S. (2013). A general path for large-scale solubilization of cellular proteins: from membrane receptors to multiprotein complexes. *Protein Expression and Purification*, *87*(2), 111–119.
- Rabik, C. A., & Dolan, M. E. (2007). Molecular mechanisms of resistance and toxicity associated with platinating agents. *Cancer Treatment Reviews*, *33*(1), 9–23.
- Reardon, J. T., Mu, D., & Sancar, A. (1996). Nucleic acids , protein synthesis , and molecular genetics: overproduction , purification , and characterization of the XPC subunit of the human DNA repair excision nuclease. *Journal of Biological Chemistry*, *271*(32), 19451–19456.
- Reid, S. L., Parry, D., Liu, H. H., & Connolly, B. A. (2001). Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: a study using fluorescent oligonucleotides and fluorescence polarization. *Biochemistry*, *40*(8), 2484–2494.

- Riedl, T., Hanaoka, F., & Egly, J. M. (2003). The comings and goings of nucleotide excision repair factors on damaged DNA. *The EMBO Journal*, *22*(19), 5293–5303.
- Robertson, A. B., Klungland, A., Rognes, T., & Leiros, I. (2009). Base excision repair: the long and short of it. *Cellular and Molecular Life Sciences*, *66*(6), 981–993.
- Robinson, C. V., Sali, A., & Baumeister, W. (2007). The molecular sociology of the cell. *Nature*, *450*(7172), 973–982.
- Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences*, *21*(9), 327–335.
- Rosenberg, B., van Camp, L., Trosko, J. E., & Mansour, V. H. (1969). Platinum compounds: a new class of potent antitumour agents. *Nature*, *222*(5191), 385–386.
- Rouillon, C., & White, M. F. (2010). The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: Implications for eukaryal and archaeal nucleotide excision repair. *Journal of Biological Chemistry*, *285*(14), 11013–11022.
- Rouillon, C., & White, M. F. (2011). The evolution and mechanisms of nucleotide excision repair proteins. *Research in Microbiology*, *162*(1), 19–26.
- Roy, R., Adamczewski, J. P., Seroz, T., Vermeulen, W., Tassan, J. P., Schaeffer, L., ... Egly, J. M. (1994). The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell*, *79*, 1093–1101.
- Rudolf, J., Makrantonis, V., Ingledew, W. J., Stark, M. J. R., & White, M. F. (2006). The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Molecular Cell*, *23*(6), 801–808.
- Sancar, A., & Rupp, W. D. (1983). A novel repair enzyme: UVRABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell*, *33*(1), 249–260.
- Sandrock, B., & Egly, J. M. (2001). A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH. *Journal of Biological Chemistry*, *276*(38), 35328–35333.
- Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H. J., ... Egly, J. (1993). Repair helicase : a component of BTF2 (TFIIH) basic transcription factor. *Science*, *260* (April).
- Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., & Cramer, P. (2017). Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature*, *551*(7679), 204–209.

BIBLIOGRAPHY

- Schmitt, D. R., Kuper, J., Elias, A., & Kisker, C. (2014). The structure of the TFIIH p34 subunit reveals a Von Willebrand Factor a like fold. *PLoS ONE*, *9*(7).
- Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., & Egly, J. M. (2000). Molecular structure of human TFIIH. *Cell*, *102*(5), 599–607.
- Scrima, A., Koníčková, R., Czyzewski, B. K., Kawasaki, Y., Jeffrey, P. D., Groisman, R., ... Thomä, N. H. (2008). Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell*, *135*(7), 1213–1223.
- Sedgwick, B. (2004). Repairing DNA-methylation damage. *Nature Reviews Molecular Cell Biology*, *5*(2), 148–157.
- Sedgwick, B., Robins, P., Totty, N., & Lindahl, T. (1988). Functional domains and methyl acceptor sites of the *Escherichia coli* Ada protein. *Journal of Biological Chemistry*, *263*(9), 4430–4433.
- Selby, C. P., & Sancar, A. (1993). Molecular mechanism of transcription-repair coupling. *Science*, *260*(5104), 53–58.
- Selby, C. P., & Sancar, A. (1997). Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II. *Journal of Biological Chemistry*, *272*(3), 1885–1890.
- Seroz, T., Winkler, G. S., Auriol, J., Verhage, R. A., Vermeulen, W., Smit, B., ... Hoeijmakers, J. H. (2000). Cloning of a human homolog of the yeast nucleotide excision repair gene MMS19 and interaction with transcription repair factor TFIIH via the XPB and XPD helicases. *Nucleic Acids Research*, *28*(22), 4506–4513.
- Shell, S. M., Hawkins, E. K., Tsai, M. S., Hlaing, A. S., Rizzo, C. J., & Chazin, W. J. (2013). Xeroderma pigmentosum complementation group C protein (XPC) serves as a general sensor of damaged DNA. *DNA Repair*, *12*(11), 947–953.
- Silva, I. A. L., Cox, C. J., Leite, R. B., Cancela, M. L., & Conceição, N. (2014). Evolutionary conservation of TFIIH subunits: implications for the use of zebrafish as a model to study TFIIH function and regulation. *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology*, *172–173*(1), 9–20.
- Smith, G. E., Summers, M. D., & Fraser, M. J. (1983). Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Molecular and Cellular Biology*, *3*(12), 2156–65.
- Staresincic, L., Fagbemi, A. F., Enzlin, J. H., Gourdin, A. M., Wijgers, N., Dunand-Sauthier, I., ... Schäfer, O. D. (2009). Coordination of dual incision and repair synthesis in human nucleotide excision repair. *The EMBO Journal*, *28*(8), 1111–1120.

- Stracy, M., Jaciuk, M., Uphoff, S., Kapanidis, A. N., Nowotny, M., Sherratt, D. J., & Zawadzki, P. (2016). Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living *Escherichia coli*. *Nature Communications*, 7, 1–9.
- Sugasawa, K. (2016). Molecular mechanisms of DNA damage recognition for mammalian nucleotide excision repair. *DNA Repair*, 44, 110–117.
- Sugasawa, K., Akagi, J. ichi, Nishi, R., Iwai, S., & Hanaoka, F. (2009). Two-Step Recognition of DNA damage for mammalian nucleotide excision repair: directional binding of the XPC complex and DNA strand scanning. *Molecular Cell*, 36(4), 642–653.
- Sugasawa, K., Masutani, C., Uchida, a, Maekawa, T., van der Spek, P. J., Bootsma, D., ... Hanaoka, F. (1996). HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair *in vitro*. *Molecular and Cellular Biology*, 16(9), 4852–61.
- Sugasawa, K., Ng, J. M., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P., ... Hoeijmakers, J. H. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Molecular Cell*, 2(2), 223–232.
- Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., & Hanaoka, F. (2001). A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes and Development*, 15(5), 507–521.
- Sugitani, N., Sivley, R. M., Perry, K. E., Capra, J. A., & Chazin, W. J. (2016). XPA: a key scaffold for human nucleotide excision repair. *DNA Repair*, 44, 123–135.
- Svejstrup, J. Q., Wang, Z., Feave, W. J., Wu, X., Bushnell, D. A., Donahue, T. F., ... Kornberg, R. D. (1995). Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. *Cell*, 80(1), 21–28.
- Szalat, R., Samur, M. K., Fulciniti, M., Lopez, M., Nanjappa, P., Cleynen, A., ... Munshi, N. C. (2017). Nucleotide excision repair is a potential therapeutic target in multiple myeloma. *Leukemia*, 32(1), 111–119.
- Takagi, Y., Masuda, C. A., Chang, W. H., Komori, H., Wang, D., Hunter, T., ... Kornberg, R. D. (2005). Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage. *Molecular Cell*, 18(2), 237–243.
- Tapias, A., Auriol, J., Forget, D., Enzlin, J. H., Schärer, O. D., Coin, F., ... Egly, J. M. (2004). Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *Journal of Biological Chemistry*, 279(18), 19074–19083.
- Theis, K., Skorvaga, M., MacHius, M., Nakagawa, N., Van Houten, B., & Kisker, C. (2000). The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutation Research - DNA Repair*, 460(3–4), 277–300.

BIBLIOGRAPHY

- Tirode, F., Busso, D., Coin, F., & Egly, J.-M. (1999). Reconstitution of the transcription factor TFIIH. *Molecular Cell*, *3*, 87–95.
- Titov, D. V., Gilman, B., He, Q. L., Bhat, S., Low, W. K., Dang, Y., ... Liu, J. O. (2011). XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nature Chemical Biology*, *7*(3), 182–188.
- Tomkinson, A. E., Chen, L., Dong, Z., Leppard, J. B., Levin, D. S., Mackey, Z. B., & Motycka, T. A. (2001). Completion of base excision repair by mammalian DNA ligases. *Progress in Nucleic Acid Research and Molecular Biology*, *68*, 151–164.
- Tomkinson, A. E., Vijayakumar, S., Pascal, J. M., & Ellenberger, T. (2006). DNA ligases: structure, reaction mechanism, and function. *Chemical Reviews*, *106*(2), 687–699.
- Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., & Sedgwick, B. (2002). Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature*, *419*(6903), 174–178.
- Uchida, A., Sugasawa, K., Masutani, C., Dohmae, N., Araki, M., Yokoi, M., ... Hanaoka, F. (2002). The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH. *DNA Repair*, *1*(6), 449–461.
- Van Gool, A. J., Citterio, E., Rademakers, S., Van Os, R., Vermeulen, W., Constantinou, A., ... Hoeijmakers, J. H. J. (1997). The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. *The EMBO Journal*, *16*(19), 5955–5965.
- Van Houten, B., Croteau, D. L., DellaVecchia, M. J., Wang, H., & Kisker, C. (2005). “Close-fitting sleeves”: DNA damage recognition by the UvrABC nuclease system. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *577*(1–2 SPEC. ISS.), 92–117.
- Van Houten, B., Kuper, J., & Kisker, C. (2016). Role of XPD in cellular functions: to TFIIH and beyond. *DNA Repair*, *44*, 136–142.
- Van Oers, M. M., Pijlman, G. P., & Vlak, J. M. (2015). Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *Journal of General Virology*, *96*(1), 6–23.
- Vashisht, A. A., Yu, C. C., Sharma, T., Ro, K., & Wohlschlegel, J. A. (2015). The association of the xeroderma pigmentosum group D DNA helicase (XPD) with transcription factor IIH is regulated by the cytosolic iron-sulfur cluster assembly pathway. *Journal of Biological Chemistry*, *290*(22), 14218–14225.
- Verhoeven, E. E. A., Van Kesteren, M., Moolenaar, G. F., Visse, R., & Goosen, N. (2000). Catalytic sites for 3' and 5' incision of *Escherichia coli* nucleotide excision repair are both located in UvrC. *Journal of Biological Chemistry*, *275*(7), 5120–5123.

- Verhoeven, E. E. A., Wyman, C., Moolenaar, G. F., & Goosen, N. (2002). The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands. *The EMBO Journal*, *21*(15), 4196–4205.
- Walker, A. S., Rablen, P. R., & Schepartz, A. (2016). Rotamer-restricted fluorogenicity of the bis-arsenical ReAsH. *Journal of the American Chemical Society*, *138*(22), 7143–7150.
- Weber, S. (2005). Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Biochimica et Biophysica Acta - Bioenergetics*, *1707*(1 SPEC. ISS.), 1–23.
- Weeda, G., Eveno, E., Donker, I., Vermeulen, W., Chevallier-Lagente, O., Taïeb, a, ... Sarasin, a. (1997). A mutation in the XPB/ERCC3 DNA repair transcription gene, associated with trichothiodystrophy. *American Journal of Human Genetics*, *60*(2), 320–329.
- Welsh, C., Day, R., McGurk, C., Masters, J. R. W., Wood, R. D., & Köberle, B. (2004). Reduced levels of XPA, ERCC1 and XPF DNA repair proteins in testis tumor cell lines. *International Journal of Cancer*, *110*(3), 352–361.
- White, M.F., Allers, T. (2018). DNA Repair in the Archaea - an emerging picture. *FEMS Microbiology Reviews*, *42*(4):514-526.
- White, M. F. (2003). Archaeal DNA repair: paradigms and puzzles. *Biochemical Society Transactions*, *31*(Pt 3), 690–693.
- White, M. F., & Dillingham, M. S. (2012). Iron-sulphur clusters in nucleic acid processing enzymes. *Current Opinion in Structural Biology*, *22*(1), 94–100.
- Whitley, R. J., & Roizman, B. (2001). Herpes simplex virus infections. *The Lancet*, *357*(9267), 1513–1518.
- Wilson, D. M., & Barsky, D. (2001). The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutation Research - DNA Repair*, *485*(4), 283–307.
- Wolski, S. C., Kuper, J., Hänzelmann, P., Truglio, J. J., Croteau, D. L., Van Houten, B., & Kisker, C. (2008). Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biology*, *6*(6), 1332–1342.
- Wong, K. H., Jin, Y., & Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Molecular Cell*, *54*(4), 601–612.
- Wysocka, J., & Herr, W. (2003). The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends in Biochemical Sciences*, *28*(6), 294–304.

BIBLIOGRAPHY

- Xiao, H., Pearson, a, Coulombe, B., Truant, R., Zhang, S., Regier, J. L., ... Ingles, C. J. (1994). Binding of basal transcription factor TFIID to the acidic activation domains of VP16 and p53. *Molecular and Cellular Biology*, 14(10), 7013–7024.
- Xie, Z., Liu, S., Zhang, Y., & Wang, Z. (2004). Roles of Rad23 protein in yeast nucleotide excision repair. *Nucleic Acids Research*, 32(20), 5981–5990.
- Yokoi, M., Masutani, C., Maekawa, T., Sugasawa, K., Ohkuma, Y., & Hanaoka, F. (2000). The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIF to damaged DNA. *Journal of Biological Chemistry*, 275(13), 9870–9875.
- Yuzhakov, A., Kelman, Z., Hurwitz, J., & O'Donnell, M. (1999). Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme. *The EMBO Journal*, 18(21), 6189–6199.
- Zhang, E. T., He, Y., Grob, P., Fong, Y. W., Nogales, E., & Tjian, R. (2015). Architecture of the human XPC DNA repair and stem cell coactivator complex. *Proceedings of the National Academy of Sciences*, 112(48), 14817–14822.
- Zurita, M., & Cruz-Becerra, G. (2016). TFIID: new discoveries regarding its mechanisms and impact on cancer treatment. *Journal of Cancer*, 7(15), 2258–2265.

APPENDIX A: Synthetic gene sequences

TFIIH Core synthetic gene sequences (5' → 3')

XPB

GGATCCCATGGGCAAAAGAGACCGAGCGGACCGCGACAAGAAGAAATCCAGGA
AGCGGCACTATGAGGATGAAGAGGATGATGAAGAGGACGCCCCGGGGAACGAC
CCTCAGGAAGCGGTTCCCTCGGCGGCGGGGAAGCAGGTGGATGAGTCAGGCAC
CAAAGTGGATGAATATGGAGCCAAGGACTACAGGCTGCAAATGCCGCTGAAGGA
CGACCACACCTCCAGGCCCTCTGGGTGGCTCCCGATGGCCATATCTTCTTGGAA
GCCTTCTCTCCAGTTTACAAATATGCCCAAGACTTCTTGGTGGCTATTGCAGAGCC
AGTGTGCCGACCAACCCATGTGCATGAGTACAACTAACTGCCTACTCCTTGTAT
GCAGCTGTCAGCGTTGGGCTGCAAACCAGTGACATCACCGAGTACCTCAGGAAG
CTCAGCAAGACTGGAGTCCCTGATGGAATTATGCAGTTTATTAAGTTGTGTA
CAGCTATGGAAAAGTCAAGCTGGTCTTGAAGCACAAACAGATACTTCGTTGAAAGTT
GCCACCCTGATGTAATCCAGCATCTTCTCCAGGACCCCGTGATCCGAGAATGCCG
CTTAAGAACTCTGAAGGGGAGGCCACTGAGCTCATCACAGAGACTTTCACAAGC
AAATCTGCCATTTCTAAGACTGCTGAAAGCAGTGGTGGGCCCTCCACTTCCCGAG
TGACAGATCCACAGGGTAAATCTGACATCCCAATGGACCTGTTTGACTTCTATGAG
CAAATGGACAAGGATGAAGAAGAAGAAGAAGAGACACAGACAGTGTCTTTTGAAG
TCAAGCAGGAAATGATTGAGGAACTCCAGAAACGTTGCATCCACCTGGAGTACCC
TCTGTTGGCAGAATATGACTTCCGGAATGATTCTGTCAACCCTGATATCAACATTG
ACCTAAAGCCCACAGCTGTCTCAGACCCTATCAGGAGAAGAGCTTGCGAAAGAT
GTTTGGAAACGGGCGTGACGTTCCGGGGTCAATTGTTCTTCCCTGCGGTGCTGG
AAAGTCCCTGGTTGGTGTGACTGCTGCATGCACTGTCAGAAAACGCTGTCTGGTG
CTGGGCAACTCAGCTGTTTCTGTGGAGCAGTGGAAAGCCCAGTTCAAGATGTGGT
CCACCATTGACGACAGCCAGATCTGCCGGTTCACCTCCGATGCCAAGGACAAGC
CCATCGGCTGCTCCGTTGCCATTAGCACCTACTCCATGCTGGGCCACACCACCAA
AAGTCCCTGGGAGGCCGAGCGAGTCATGGAGTGGCTCAAGACCCAGGAGTGGG
GCCTCATGATCCTGGATGAAGTGCACACCATAACCAGCCAAGATGTTCCGAAGGGT
GCTCACCATCGTGCAGGCCCACTGTAAGCTGGGTTTACTGCGACCCTCGTCCG
CGAAGATGACAAAATTGTGGATTTAAATTTTCTGATTGGGCCTAAGCTCTACGAAG

APPENDIX A

CTAACTGGATGGAGCTGCAGAATAATGGCTACATCGCCAAAGTCCAGTGTGCTGA
GGTCTGGTGCCCTATGTCTCCTGAATTTTACCGGGAATATGTGGCAATCAAAACCA
AGAAACGAATCTTGCTGTACACCATGAACCCCAACAAATTTAGAGCTTGCCAGTTT
CTGATCAAGTTTCATGAAAGGAGGAATGACAAGATTATTGTCTTTGCTGACAATGT
GTTTGCCCTAAAGGAATATGCCATTCGACTGAACAAACCCTATATCTACGGACCTA
CGTCTCAGGGGGAAAGGATGCAAATTCTCCAGAATTTCAAGCACAACCCCAAAT
TAACACCATCTTCATATCCAAGGTAGGTGACACTTCGTTTGATCTGCCGGAAGCAA
ATGTCCTCATTAGATCTCATCCCACGGTGGCTCCAGGCGTCAGGAAGCCCAAAG
GCTAGGGCGGGTGCTTCGAGCTAAAAAAGGGATGGTTGCAGAAGAGTACAATGC
CTTTTTCTACTCACTGGTATCCCAGGACACACAGGAAATGGCTTACTCAACCAAGC
GGCAGAGATTCTTGGTAGATCAAGGTTATAGCTTCAAGGTGATCACGAACTCGC
TGGCATGGAGGAGGAAGACTTGGCGTTTTTCGACAAAAGAAGAGCAACAGCAGCT
CTTACAGAAAGTCCTGGCAGCAACTGACCTGGATGCCGAGGAGGAGGTGGTGGC
TGGGGAATTTGGCTCCAGATCCAGCCAGGCATCTCGGCGCTTTGGCACCATGAG
TTCTATGTCTGGGGCCGACGACACTGTGTACATGGAGTACCACTCATCGCGGAGC
AAGGCGCCAGCAAACATGTACACCCGCTCTTCAAGCGCTTTAGGAAATGAGGTA
CCGTCGAC

XPD

GGATCCCATGGAGCTCAACGTGGACGGGCTCCTGGTCTACTTCCCGTACGACT
ACATCTACCCCGAGCAGTTCTCCTACATGCGGGAGCTCAAACGCACGCTGGACG
CCAAGGGTCATGGAGTCCTGGAGATGCCCTCAGGCACCGGGAAGACAGTATCCC
TGTTGGCCCTGATCATGGCATAACAGAGAGCATATCCGCTGGAGGTGACCAA
ACTCATCTACTGCTCAAGAACTGTGCCAGAGATTGAGAAGGTGATTGAAGAGCTTCGA
AAGTTGCTCAACTTCTATGAGAAGCAGGAGGGCGAGAAGCTGCCGTTTCTGGGAC
TGGCTCTGAGCTCCCGCAAAA
ACTTGTGTATTACCCCTGAGGTGACACCCCTGCG
CTTTGGGAAGGACGTCGATGGGAAATGCCACAGCCTCACAGCCTCCTATGTGCG
GGCGCAGTACCAGCATGACACCAGCCTGCCCCACTGCCGATTCTATGAGGAATTT
GATGCCACGGGCGTGAGGTGCCCTCCCCGCTGGCATCTACAACCTGGATGAC
CTGAAGGCCCTGGGGCGGCGCCAGGGCTGGTGCCATACTTCTTGCTCGATAC
TCAATCCTGCATGCTAATGTGGTGGTTTATAGCTACCACTACCTCCTGGACCCCAA
GATTGCAGACCTGGTGTCTAAGGAACTGGCCCGCAAGGCCGTCGTGGTCTTCGA
CGAGGCCCAACATTGACAACGTCTGCATCGACTCCATGAGCGTCAACCTCACC

CGCCGGACCCTTGACCGGTGCCAGGGCAACCTGGAGACCCTGCAGAAGACGGT
GCTCAGGATCAAAGAGACAGACGAGCAGCGCCTGCGGGACGAGTACCGGCGTCT
GGTGGAGGGGCTGCGGGAGGCCAGCGCCGCCGGGAGACGGACGCCACCTG
GCCAACCCCGTGCTGCCCGACGAAGTGTGTCAGGAGGCAGTGCCTGGCTCCATC
CGCACGGCCGAGCATTTCCTGGGCTTCCTGAGGCGGCTGCTGGAGTACGTGAAG
TGGCGGCTGCGTGTGCAGCATGTGGTGCAGGAGAGCCCGCCCGCCTTCCTGAG
CGGCCTGGCCAGCGCGTGTGCATCCAGCGCAAGCCCCTCAGATTCTGTGCTGA
ACGCCTCCGGTCCCTGCTGCATACTCTGGAGATCACCGACCTTGCTGACTTCTCC
CCGCTCACCTCCTTGCTAACTTTGCCACCCTTGTGAGCACCTACGCCAAAGGCT
TCACCATCATCATCGAGCCCTTTGACGACAGAACCCCGACCATTGCCAACCCCAT
CCTGCACTTCAGCTGCATGGACGCCTCGCTGGCCATCAAACCCGTATTTGAGCGT
TTCCAGTCTGTCATCATCACATCTGGGACACTGTCCCCGCTGGACATCTACCCTAA
GATCCTGGACTTCCACCCCGTCACAATGGCAACCTTCACAATGACGCTGGCACGG
GTCTGCCTCTGCCCTATGATCATCGGCCGTGGCAATGACCAGGTGGCCATCAGCT
CCAAATTTGAGACCCGGGAGGATATTGCTGTGATCCGGAACCTATGGGAACCTCCT
GCTGGAGATGTCCGCTGTGGTCCCTGATGGCATCGTGGCCTTCTTCACCAGCTAC
CAGTACATGGAGAGCACCGTGGCCTCCTGGTATGAGCAGGCGATCCTTGAGAAC
ATCCAGAGGAACAAGCTGCTCTTTATTGAGACCCAGGATGGTGCCGAAACCAGTG
TCGCCCTGGAGAAGTACCAGGAGGCCTGCGAGAATGGCCGCGGGGCCATCCTG
CTGTCAGTGGCCCGGGGCAAAGTGTCCGAGGGAATCGACTTTGTGCACCACTAC
GGGCGGGCCGTCATCATGTTTGGCGTCCCCTACGTCTACACACAGAGCCGCATT
CTCAAGGCGCGGCTGGAATACCTGCGGGACCAGTTCCAGATTTCGTGAGAATGAC
TTTCTTACCTTCGATGCCATGCGCCACGCGGCCAGTGTGTGGGTTCGGGCCATC
AGGGGCAAGACGGACTACGGCCTCATGGTCTTTGCCGACAAGCGGTTTGCCCGT
GGGACAAGCGGGGGAAGCTGCCCGCTGGATTCAGGAGCACCTCACAGATGC
CAACCTCAACCTGACCGTGGACGAGGGTGTCCAGGTGGCCAAGTACTTCCTGCG
GCAGATGGCACAGCCCTTCACCCGGGAGGATCAGCTGGGCCTGTCCCTGCTCAG
CCTGGAGCAGCTAGAATCAGAGGAGACGCTGAAGAGGATAGAGCAGATTGCTCA
GCAGCTCTGAGGTACCGTCGAC

p62

GGATCCCATGGCAACCTCATCTGAAGAAGTTTTGCTGATTGTAAAGAAAGTGCG
TCAAAGAAGCAGGATGGAGCTCTGTACCTCATGGCAGAAAGAATTGCTTGGGCA

APPENDIX A

CCTGAAGGCAAAGATAGATTTACAATCAGCCATATGTATGCAGATATTAATGCCA
GAAAATTAGTCCAGAAGGAAAAGCTAAAATTCAGCTTCAGCTGGTCCTACATGCAG
GGGACACAACCTAECTTCCATTTTTCCAATGAAAGCACAGCAGTGAAAGAGCGAGA
TGCAGTAAAAGACCTTCTTCAGCAGCTGCTGCCCAAATTC AAGAGGAAAGCAAAT
AAAGA ACTGGAAGAGAAGAACAGAATGCTGCAAGAAGATCCTGTTTTGTTTCAGCT
TTATAAAGACCTTGTTGTGAGTCAAGTGATCAGTGCTGAGGAATTCTGGGCCAATC
GTTTAAATGTGAATGCAACAGATAGTTCTTCCACATCCAATCATAAGCAGGATGTT
GGCATTCTGCTGCATTTCTGGCTGATGTCCGGCCTCAA ACTGATGGCTGTAACG
GTCTAAGATATAATTTAACTTCTGATATCATTGAGTCCATATTTAGGACCTATCCAG
CAGTAAAATGAAATATGCAGAAAATGTTCCCCACAACATGACAGAGAAGGAATTC
TGGACACGTTTTTTCCAGTCCCATTATTTTCACAGGGATCGGCTGAATACAGGGTC
AAAGGATCTCTTGCAGAATGTGCCAAAATAGATGAAAAAGGCCTAAAAACAATGG
TTTCATTAGGAGTGAAAAACCCACTACTAGATTTAACAGCTTTGGAAGATAAACCA
TTAGATGAGGGCTATGGCATTTCCTCTGTGCCATCTGCTTCCAATTCTAAATCCAT
AAAAGAGAATAGTAATGCTGCCATCATCAAGAGATTTAACCATCACAGTGCTATGG
TCCTGGCAGCTGGACTCAGAAAACAAGAAGCACAAAATGAACAACTAGTGAGCC
TAGCAACATGGATGGAAATCCGGAGATGCAGACTGCTTTCAGCCAGCAGTCAAA
AGGGCGAAATTACAAGAGTCCATTGAATATGAAGACTTGGGGAAAATAATTCTGT
AAAAACGATTGCACTAAACCTCAAGAAGTCAGATAGGTATTATCATGGTCCA ACTC
CAATCCAGTCACTACAGTATGCAACAAGTCAGGACATTATTAATTCTTTTCAAAGTA
TTAGACAAGAAATGGAAGCTTATACACCCAAGTTAACTCAGGTTCTCTCAAGTAGT
GCTGCCAGTAGTACCATCACAGCACTGTCACCTGGAGGGGCACTTATGCAGGGA
GGAACACAGCAAGCCATAAACCAGATGGTGCCAAATGATATTCAATCTGAATTGAA
ACACTTATATGTAGCTGTTGGAGA ACTTCTACGACATTTCTGGTCCTGCTTTCTG
TTAATACGCCATTCCTAGAAGAAAAGGTAGTGAAAATGAAAAGTAATTTGGAACGA
TTCCAAGTTACGAAGCTCTGTCCATTCCAAGAAAAGATTCGGAGACAGTATTTAAG
CACAAATTTGGTAAGTCACATAGAAGAGATGCTCCAGACAGCCTACAACAAGCTC
CACACATGGCAGTCACGGCGTCTGATGAAGAAAACGTGAGGTACCGTCGAC

p52-V5-F2A-p8

GGATCCCATGGAGAGCACCCCTTCAAGGGGACTGAACCGAGTACACCTACAAT
GCAGGAATCTGCAGGAATTCTTAGGGGGCCTGAGCCCTGGGGTATTGGACCGAT
TGTATGGGCACCCTGCTACATGTCTGGCTGTCTTCAGGGAGCTCCCATCCTTGGC

TAAGAACTGGGTGATGCGGATGCTCTTTCTGGAGCAGCCTTTGCCACAGGCTGCT
GTAGCTCTGTGGGTAAAGAAGGAATTCAGCAAGGCTCAGGAGGAAAGTACAGGG
CTGCTGAGCGGCCTCCGGATCTGGCACACACAGCTGCTCCCAGGCGGGCTCCAG
GGCCTCATCCTCAACCCCATTTTCCGCCAGAACCTCCGCATTGCCCTTCTGGGTG
GGGGGAAGGCCTGGTCTGATGACACAAGTCAGCTGGGACCAGACAAGCATGCC
GGGACGTTCCCTCCCTTGACAAGTACGCCGAGGAGCGATGGGAGGTGGTCTTGC
ACTTCATGGTGGGCTCCCCAGTGCAGCTGTCAGCCAGGACTTGGCTCAGCTCC
TCAGCCAGGCTGGGCTCATGAAGAGTACTGAACCTGGAGAGCCGCCCTGCATTA
CTTCCGCTGGCTTCCAGTTCCTGTTGCTGGACACCCCGGCTCAGCTCTGGTACTT
TATGTTGCAGTATTTGCAGACAGCCCAGAGCCGGGGCATGGACCTGGTAGAGATT
CTCTCCTTCTCTTCCAGCTCAGCTTCTCTACTCTGGGCAAGGATTAATCTGTGGA
AGGTATGAGTGATTCTCTGTTGAACTTCCTGCAACATCTGCGTGAGTTTGGGCTTG
TTTTCCAGAGGAAGAGGAAATCTCGGCGTACTACCCTACACGCCTGGCCATCAA
TCTCTCATCAGGTGTCTCTGGAGCTGGGGGCACTGTGCATCAGCCAGGTTTCATT
GTCGTGGAAACCAATTACCGACTGTATGCCTACACGGAGTCGGAGCTGCAGATTG
CCCTCATTGCCCTCTTCTCTGAGATGCTCTATCGGTTCCCTAACATGGTGGTGGC
GCAGGTGACCCGGGAGAGTGTGCAGCAGGCAATCGCCAGTGGCATCACAGCCC
AGCAGATAATCCATTTCTAAGGACAAGAGCCCACCCAGTGATGCTCAAACAGAC
ACCTGTGCTGCCCCCACCATCACCGACCAGATCCGGCTCTGGGAGCTGGAAAG
GGACAGACTCCGGTTCCTGAGGGTGTCTGTATAACCAGTTCCTGTGCAAGTG
GACTTTGAGCTGCTGCTGGCCCACGCGGGGAGCTGGGCGTGCTCGTGTTGAG
AACTCGGCCAAGCGGCTCATGGTGGTGACCCCGGCCGGGCACAGCGACGTCAA
GCGCTTTTGGAAGCGGCAGAAACATAGCTCCGGCAAGCCCATCCCCAACCCCT
GCTGGGCCTGGACAGCACCCAGACCCTGAACTTCGACCTGCTGAAGCTGGCCGG
CGACGTGGAGAGCAACCCCGGCCCTATGGTCAACGTCTTGAAAGGAGTGCTTATA
GAATGTGATCCTGCCATGAAGCAGTTTCTGCTGTACTTGGATGAGTCAAATGCCCT
GGGGAAGAAGTTCATCATTCAAGACATTGATGACACTCACGTCTTTGTAATAGCAG
AATTGGTTAATGTCCTCCAGGAGCGAGTGGGTGAATTAATGGACCAAATGCTTTT
TCCCTTACCCAGAAATGAGGTACCGTCGAC

p44-T2A-p34

GGATCCCATGGATGAAGAACCTGAAAGAACTAAGCGATGGGAAGGAGGCTATG
AAAGAACATGGGAGATTCTTAAAGAAGATGAATCTGGATCACTTAAAGCTACAATA

APPENDIX A

GAAGACATTCTATTCAAGGCAAAGAGAAAAAGAGTATTTGAGCATCATGGACAAGT
TCGACTTGGAATGATGCGCCACCTTTATGTGGTAGTAGATGGATCAAGAACAATG
GAAGACCAAGATTTAAAGCCTAATAGACTGACGTGTACTTTAAAGTTGTTGGAATA
CTTTGTAGAGGAATATTTTGATCAAATCCTATTAGTCAGATTGGAATAATTGTAAC
TAAGAGTAAAAGAGCTGAAAAATTGACTGAACTTTCAGGAAACCCAAGAAAACATA
TAACGTCTTTGAAGAAAGCTGTGGATATGACCTGCCACGGAGAGCCATCTCTTTAT
AATCCCTAAGCATAGCTATGCAGACTCTAAAACACATGCCTGGACATACAAGTCG
AGAAGTACTAATCATCTTTAGCAGCCTTACAACCTTGCATCCATCTAATATTTATGA
TCTAATCAAGACCCTAAAGGCAGCTAAAATTAGAGTATCTGTTATTGGATTGTCTG
CAGAAGTTCGCGTTTGCCTGTACTTGCTCGTGAAACTGGTGGCAGTACCATGT
TATTTTAGATGAAAGCCATTACAAAGAGTTGCTCACACATCATGTTAGTCCTCCTC
CTGCTAGCTCAAGTTCTGAATGCTCACTTATTCGTATGGGATTTCTCAGCACACC
ATTGCTTCTTTATCTGACCAGGATGCAAACCCTCTTTCAGCATGGCGCATTGGA
TGGCAATACTGAGCCAGGGCTTACATTAGGAGGCTATTTCTGCCACAGTGTCGG
GCAAAGTACTGTGAGCTACCTGTTGAATGTAATCTGTGGTCTTACTTTGGTGTC
TGCTCCCCACTTGGCACGGTCTTACCATCATTTGTTTCCTTTGGATGCTTTTCAAG
AAATCCCCTAGAAGAATATAATGGAGAAAGATTTTGTATGGATGTCAGGGGGAA
TTGAAAGACCAACATGTTTATGTTTGTGCTGTGTGCCAAAATGTTTTCTGTGTGGA
CTGTGATGTTTTTGTTCATGATTCTCTACTGTTGCCCTGGCTGTATTCATAAGAT
TCCAGCTCCTTCAGGTGTTGAGGGCCGGGGCAGCCTGCTGACCTGCGGGCAGCGT
GGAGGAGAACCCCGGCCCTATGGTTTCAGACGAAGATGAATTGAATCTTCTGGTT
ATTGTAGTTGATGCCAACCTATTTGGTGGGGAAAGCAAGCATTAAAGGAATCTCA
GTTCACTTTATCCAAATGCATAGATGCCGTGATGGTGTGGGAAATTCGCATTTAT
TCATGAATCGTTCCAACAACTTGTGTGATAGCAAGTCACATTCAAGAAAGCCGA
TTCTTATATCCTGGAAAGAATGGCAGACTTGGAGACTTCTTCGGAGACCCTGGCA
ACCCTCCTGAATTTAATCCCTCTGGGAGTAAAGATGGAAAATACGAACTTTTAACC
TCAGCAAATGAAGTTATTGTTGAAGAGATTAAGATCTAATGACCAAAGTGACAT
AAAGGGTCAACATACAGAACTTTGCTGGCAGGCTCCCTGGCCAAAGCCCTTTGC
TACATTCATAGAATGAACAAGGAAGTTAAAGACAATCAGGAAATGAAATCAAGGAT
ATTGGTGATTAAGGCTGCAGAAGACAGTGCCTTGCAGTATATGAACTTCATGAATG
TCATCTTTGCAGCACAGAAACAGAATATTTGATTGATGCCTGTGTTTTAGACTCC
GACTCAGGGCTCCTCCAACAGGCTTGTGACATCACGGGAGGACTGTACCTGAAG
GTGCCTCAGATGCCTTCTCTTCTGCAGTATTTGCTGTGGGTGTTTCTCCCGATCA
AGATCAGAGATCTCAGTTAATCCTCCCACCCCGGTTTCATGTTGACTACAGGGCT

GCTTGCTTCTGTCATCGAAATCTCATTGAAATTGGTTATGTCTGTTCTGTGTGTTTG
 TCAATATTCTGCAATTTTCAGCCCCATTTGTACTACGTGCGAGACAGCCTTTAAAATT
 TCTCTGCCTCCAGTGCTGAAAGCCAAGAAAAAGAACTGAAAGTGTCTGCCTGAG
GTACCGTCGAC

CAK synthetic gene sequences

MAT1

GGATCCCATGGACGATCAGGGTTGCCCTCGGTGTAAGACCACCAAATATCGGAA
 CCCCTCCTTGAAGCTGATGGTGAATGTGTGCGGACACACTCTCTGTGAAAGTTGT
 GTAGATTTACTGTTTGTGAGAGGAGCTGGAACTGCCCTGAGTGTGGTACTCCAC
 TCAGAAAGAGCAACTTCAGGGTACAACCTTTGAAGATCCCCTGTTGACAAGGA
 GGTTGAGATCAGGAAAAAAGTGCTAAAGATATACAATAAAAGGGAAGAAGATTTTC
 CTAGTCTAAGAGAATACAATGATTTCTTGGAAGAAGTGGAAGAAATTGTTTTCAAC
 TTGACCCACAATGTGGATTTGGACAACACCAAAAAGAAAATGGAGATATACCAAAA
 GGAAAACAAAGATGTTATTCAGAAAAATAAATTAAGCTGACTCGAGAACAGGAAG
 AACTGGAAGAAGCTTTAGAAGTGGAACGACAGGAAAATGAACAAAGAAGATTATTT
 ATACAAAAGAAGAACAACACTGCAGCAGATTCTAAAAAGGAAGAATAAGCAGGCTTT
 TTTAGATGAGCTGGAGAGTTCTGATCTCCCTGTTGCTCTGCTTTTGGCTCAGCATA
 AAGATAGATCTACCCAATTAGAAATGCAACTTGAGAAACCCAAACCTGTAAAACCA
 GTGACGTTTTCCACAGGCATCAAAATGGGTCAACATATTTCACTGGCACCTATTCA
 CAAGCTTGAAGAAGCTCTGTATGAATACCAGCCACTGCAGATAGAGACATATGGA
 CCACATGTTCCCTGAGCTTGAGATGCTAGGAAGACTTGGGTATTTAAACCATGTCAG
 AGCTGCCTCACCACAGGACCTTGCTGGAGGCTATACTTCTTCTTCTTGTGTCACA
 GAGCACTACAGGATGCATTCAGTGGGCTTTTCTGGCAGCCCAGTTAA**GGTACCGT**
CGAC

Cdk7-T2A-Cyclin H

GGATCCCATGGGCTCTAGACGTGAAGTCTCGGGCAAAGCGTTATGAGAAGCTGG
 ACTTCCTTGGGGAGGGACAGTTTGCCACCGTTTACAAGGCCAGAGATAAGAATAC
 CAACCAAATTGTCGCCATTAAGAAAATCAAACCTTGACATAGATCAGAAGCTAAAG
 ATGGTATAAATAGAACCGCCTTAAGAGAGATAAAATTATTACAGGAGCTAAGTCAT
 CCAAATATAATTGGTCTCCTTGATGCTTTTGGACATAAATCTAATATTAGCCTTGTC

APPENDIX A

TTTGATTTTATGGAACTGATCTAGAGGTTATAATAAAGGATAATAGTCTTGTGCTG
ACACCATCACACATCAAAGCCTACATGTTGATGACTCTTCAAGGATTAGAATATTTA
CATCAACATTGGCTCCTACATAGGGATCTGAAACCAAACAACCTTGTTGCTAGATGA
AAATGGAGTTCTAAAACCTGGCAGATTTTGGCCTGGCCAAATCTTTTGGGAGCCCC
AATAGAGCTTATACACATCAGGTTGTAACCAGGTGGTATCGGGCCCCCGAGTTAC
TATTTGGAGCTAGGATGTATGGTGTAGGTGTGGACATGTGGGCTGTTGGCTGTAT
ATTAGCAGAGTTACTTCTAAGGGTTCCTTTTTTGGCAGGAGATTCAGACCTTGATC
AGCTAACAAGAATATTTGAAAACCTTTGGGCACACCAACTGAGGAACAGTGGCCGGA
CATGTGTAGTCTTCCAGATTATGTGACATTTAAGAGTTTCCCTGGAATACCTTTGC
ATCACATCTTCAGTGCAGCAGGAGACGACTTACTAGATCTCATACAAGGCTTATTC
TTATTTAATCCATGTGCTCGAATTACGGCCACACAGGCACTGAAAATGAAGTATTT
CAGTAATCGGCCAGGGCCAACACCTGGATGTCAGCTGCCAAGACCAAACCTGTCC
AGTGGAACCTTAAAGGAGCAATCAAATCCAGCTTTGGCAATAAAAAGGAAAAGAA
CAGAGGCCTTAGAACAAGGAGGATTGCCCAAGAACTAATTTTTGAAGGTTCGCGG
GAGTCTCCTCACGTGTGGGGATGTCTGAAGAAAACCCCGGCCCTATGTACCACAAC
AGTAGTCAGAAGCGGCACTGGACCTTCTCCAGCGAGGAGCAGCTGGCAAGACTG
CGGGCTGACGCCAACCGCAAATTCAGATGCAAAGCCGTGGCCAACGGGAAGGTT
CTTCCGAATGATCCAGTCTTTCTTGAGCCTCATGAAGAAATGACACTCTGCAAATA
CTATGAGAAAAGGTTATTGGAATTCTGTTCCGGTGTTTAAGCCAGCAATGCCAAGAT
CTGTTGTGGGTACGGCTTGTATGTATTTCAAACGTTTTTATCTTAATAACTCAGTAA
TGGAATATCACCCCAGGATAATAATGCTCACTTGTGCATTTTTTGGCCTGCAAAGTA
GATGAATTCAATGTATCTAGTCCTCAGTTTGTGGAAACCTCCGGGAGAGTCCTCT
TGGACAGGAGAAGGCACTTGAACAGATACTGGAATATGAACTACTTCTTATACAGC
AACTTAATTTCCACCTTATTGTCCACAATCCTTACAGACCATTTGAGGGCTTCCTCA
TCGACTTAAAGACCCGCTATCCCATATTGGAGAATCCAGAGATTTTGAGGAAAACA
GCTGATGACTTTCTTAATAGAATTGCATTGACGGATGCTTACCTTTTATACACACCT
TCCCAAATTGCCCTGACTGCCATTTTATCTAGTGCCTCCAGGGCTGGAATTAAT
GGAAAGTTATTTATCAGAGAGTCTGATGCTGAAAGAGAACAGAACTTGCCTGTCAC
AGTTACTAGATATAATGAAAAGCATGAGAACTTAGTAAAGAAGTATGAACCACCC
AGATCTGAAGAAGTTGCTGTTCTGAAACAGAAGTTGGAGCGATGTCATTCTGCTG
AGCTTGCCTTAACGTAATCACGAAGAAGAGGAAAGGCTATGAAGATGATGATTA
CGTCTCAAAGAAATCCAACATGAGGAGGAAGAATGGACTGATGACGACCTGGTA
GAATCTCTCTAAGGTACCGTCGAC

XPC Synthetic gene sequence

GGATCCCATGGCTCGGAAACGCGCGGCCGGCGGGGAGCCGCGGGGACGCGA
ACTGCGCAGCCAGAAATCCAAGGCCAAGAGCAAGGCCCGGCGTGAGGAGGAGG
AGGAGGATGCCTTTGAAGATGAGAAACCCCAAAGAAGAGCCTTCTCTCCAAAGT
TTCACAAGGAAAGAGGAAAAGAGGCTGCAGTCATCCTGGGGGTTTCAGCAGATGG
TCCAGCAAAAAGAAAGTGGCCAAGGTGACTGTTAAATCTGAAAACCTCAAGGTTA
TAAAGGATGAAGCCCTCAGCGATGGGGATGACCTCAGGGACTTTCCAAGTGACCT
CAAGAAGGCACACCATCTGAAGAGAGGGGCTACCATGAATGAAGACAGCAATGAA
GAAGAGGAAGAAAGTGAAAATGATTGGGAAGAGGTTGAAGAACTTAGTGAGCCTG
TGCTGGGTGACGTGAGAGAAAGTACAGCCTTCTCTCGATCTCTTCTGCCTGTGAA
GCCAGTGGAGATAGAGATTGAAACGCCAGAGCAGGCGAAGACAAGAGAAAGAAG
TGAAAAGATAAACTGGAGTTTGAGACATATCTTCGGAGGGCGATGAAACGTTTCA
ATAAAGGGGTCCATGAGGACACACACAAGGTTACCTTCTCTGCCTGCTAGCAA
TGGCTTCTATCGAAATAACATCTGCAGCCAGCCAGATCTGCATGCTATTGGCCTGT
CCATCATCCCAGCCCGCTTTACCAGAGTGCTGCCTCGAGATGTGGACACCTACTA
CCTCTCAAACCTGGTGAAGTGGTTCATTGGAACATTTACAGTTAATGCAGAACTTT
CAGCCAGTGAACAAGATAACCTGCAGACTACATTGGAAGGAGATTTGCTATTTAC
TCTGCTCGAGATGATGAGGAATTGGTCCATATATTCTTACTGATTCTCCGGGCTCT
GCAGCTCTTGACCCGGCTGGTATTGTCTCTACAGCCAATTCCTCTGAAGTCAGCA
ACAGCAAAGGGAAAGAAACCTTCCAAGGAAAGATTGACTGCGGACCCAGGAGGC
TCCTCAGAACTTCCAGCCAAGTTCTAGAAAACCACACCAAACCAAAGACCAGCA
AAGGAACCAAACAAGAGGAAACCTTTGCTAAGGGCACCTGCAGGCCAAGTGCCA
AAGGGAAGAGGAACAAGGGAGGCAGAAAGAAACGGAGCAAGCCCTCCTCCAGC
GAGGAAGATGAGGGCCCAGGAGACAAGCAGGAGAAGGCAACCCAGCGACGTCC
GCATGGCCGGGAGCGGCGGGTGGCCTCCAGGGTGTCTTATAAAGAGGAGAGTG
GGAGTGATGAGGCTGGCAGCGGCTCTGATTTTGAGCTCTCCAGTGGAGAAGCCT
CTGATCCCTCTGATGAGGATTCCGAACCTGGCCCTCAAAGCAGAGGAAAGCCC
CCGCTCCTCAGAGGACAAAGGCTGGGTCCAAGAGTGCCTCCAGGACCCATCGTG
GGAGCCATCGTAAGGACCCAAGCTTGCCAGCGGCATCCTCAAGCTCTTCAAGCA
GTAAAAGAGGCAAGAAAATGTGCAGCGATGGTGAGAAGGCAGAAAAAGAAGCAT
AGCTGGTATAGACCAGTGGCTAGAGGTGTTCTGTGAGCAGGAGGAAAAGTGGGT
ATGTGTAGACTGTGTGCACGGTGTGGTGGGCCAGCCTCTGACCTGTTACAAGTAC
GCCACCAAGCCCATGACCTATGTGGTGGGCATTGACAGTGACGGCTGGGTCCGA

APPENDIX A

GATGTCACACAGAGGTACGACCCAGTCTGGATGACAGTGACCCGCAAGTGCCGG
GTTGATGCTGAGTGGTGGGCCGAGACCTTGAGACCATAACCAGAGCCCATTTATGG
ACAGGGAGAAGAAAGAAGACTTGGAGTTTCAGGCTAAACACATGGACCAGCCTTT
GCCACTGCCATTGGCTTATATAAGAACCACCCTCTGTATGCCCTGAAGCGGCAT
CTCCTGAAATATGAGGCCATCTATCCCGAGACAGCTGCCATCCTTGGGTATTGTC
GTGGAGAAGCGGTCTACTCCAGGGATTGTGTGCACACTCTGCATTCCAGGGACA
CGTGGCTGAAGAAAGCAAGAGTGGTGAGGCTTGGAGAAGTACCCTACAAGATGG
TGAAAGGCTTTTCTAACCGTGCTCGGAAAGCCCGACTTGCTGAGCCCCAGCTGCG
GGAAGAAAATGACCTGGGCCTGTTTGGCTACTGGCAGACAGAGGAGTATCAGCC
CCCAGTGGCCGTGGACGGGAAGGTGCCCCGGAACGAGTTTGGGAATGTGTACCT
CTTCCTGCCAGCATGATGCCTATTGGCTGTGTCCAGCTGAACCTGCCAATCTA
CACCGCGTGGCCCGCAAGCTGGACATCGACTGTGTCCAGGCCATCACTGGCTTT
GATTTCCACGGCGGCTACTCCCATCCCGTGACTGATGGATACATCGTCTGCGAGG
AATTCAAAGACGTGCTCCTGACTGCCTGGGAAAATGAGCAGGCAGTCATTGAAAG
GAAGGAGAAGGAGAAAAAGGAGAAGCGGGCTCTAGGGAAGTGGAAAGTTGCTGGC
CAAAGGTCTGCTCATCAGGGAGAGGCTGAAGCGTACGGGCCCAAGAGTGA
GGCAGCAGCTCCCCACACAGATGCAGGAGGTGGACTCTCTTCTGATGAAGAGGA
GGGGACCAGCTCTCAAGCAGAAGCGGCCAGGATACTGGCTGCCTCCTGGCCTCA
AAACCGAGAAGATGAAGAAAAGCAGAAGCTGAAGGGTGGGCCCAAGAAGACCAA
AAGGGAAAAGAAAGCAGCAGCTTCCACCTGTTCCCATTTGAGCAGCTGTGAGGT
ACCGTCGAC

***HR23B* Synthetic gene sequence**

GGATCCCATGGCAGTCACCCTGAAAACATTGCAGCAGCAGACCTTTAAGATCGA
CATTGATCCCGAAGAACTGTGAAGGCCTTGAAGGAGAAAATAGAGAGCGAAAAA
GGCAAGGACGCATTTCTGTGGCCGGTCAGAAGCTCATCTATGCCGGGAAAATCT
TGAATGATGACACCGCTCTGAAAGAGTACAAGATTGATGAAAAAACTTCGTGGTC
GCCATGGTGACCAAACCCAAAGCAGTGTCCACACCAGCACCAGCTACAACCTCAGC
AGTCAGCTCCTGCCAGCACTACAGCAGTTACTTCCCTCCACCACGACAACCTGTGGC
TCAGGCTCCAACCCCTGTCCCTGCCTTGGCCCCACTTCCACACCTGCATCCATC
ACTCCAGCATCAGCGACAGCATCTTCTGAACCTGCACCTGCTAGTGCAGCTAAAC
AAGAGAAGCCTGCAGAAAAGCCAGCAGAGACACCAGTGGCTACTAGCCCAACAG
CAACTGACAGTACATCGGGTGATTCTTCTCGGTCAAACCTTTTTGAAGATGCAACG

AGTGCACTTGTGACGGGTCAGTCTTACGAGAATATGGTAACTGAGATCATGTCAAT
GGGCTATGAACGAGAGCAAGTAATTGCAGCCCTGAGAGCCAGTTTCAACAACCCT
GACAGAGCAGTGGAGTATCTTTTAATGGGAATCCCTGGAGATAGAGAAAGTCAGG
CTGTGGTTGACCCCCCTCAAGCAGCTAGTACTGGGGCTCCTCAGTCTTCAGCAGT
GGCTGCAGCTGCAGCAACTACGACAGCAACAACACTACAACAACAAGTTCTGGAGGA
CATCCCCTTGAATTTTTACGGAATCAGCCTCAGTTTCAACAGATGAGACAAATTATT
CAGCAGAATCCTTCCTTGCTTCCAGCGTTACTACAGCAGATAGGTCGAGAGAATC
CTCAATTACTTCAGCAAATTAGCCAACACCAGGAGCATTTTATTCAGATGTTAAATG
AACCAGTTCAAGAAGCTGGTGGTCAAGGAGGAGGAGGTGGAGGTGGCAGTGGA
GGAATTGCAGAAGCTGGAAGTGGTCATATGAACTACATTCAAGTAACACCTCAGG
AAAAAGAAGCTATAGAAAGGTTAAAGGCATTAGGATTTCTGAAGGACTTGTGATA
CAAGCGTATTTTGCTTGTGAGAAGAATGAGAATTTGGCTGCCAATTTCTTCTACA
GCAGAACTTTGATGAAGATTGAGGTACCGTCGAC

APPENDIX B: Plasmids and bacmids

Table B1: Plasmids employed in the cloning of the different TFIID constructs described in this thesis. Purification tags (highlighted in red) are indicated in brackets before the subunit they were attached to.

Sub-complex TFIID Core (subunits p52, XPD tagged with N-terminal 6xHis-V5-TEV)
pACEBac2-(6xHis)p52-F2A-p8
pACEBac2-p44-T2A-p34
pACEBac2-p62
pACEBac2-XPB
pACEBac2-(6xHis)XPD
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34-p62
pACEBac2-XPB-(6xHis)XPD
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34-p62-XPB-(6xHis)XPD
Sub-complex TFIID Core (subunits p52, p62 tagged with N-terminal 6xHis-V5-TEV)
pACEBac2-(6xHis)p52-F2A-p8
pACEBac2-p44-T2A-p34
pACEBac2-(6xHis)p62
pACEBac2-XPB
pACEBac1-XPD
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34-(6xHis)p62
pACEBac2-XPB-XPD
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34-(6xHis)p62-XPB-XPD
6-Subunit TFIID Core (subunit p52 tagged with N-terminal 8xHis-V5-spacer-TEV)
pACEBac2-(8xHis)p52-F2A-p8
pACEBac2-p44-T2A-p34
pACEBac2-p62
pACEBac2-XPB
pACEBac2-(8xHis)p52-F2A-p8-p44-T2A-p34

APPENDIX B

pACEBac2-(8xHis)p52-F2A-p8-p44-T2A-p34-p62
pACEBac2-(8xHis)p52-F2A-p8-p44-T2A-p34-p62-XPB
Subunit XPD (tagged with N-terminal 8xHis-V5-spacer-TEV)
pACEBac2-(8xHis)XPD
Sub-complex TFIIH Core (subunits p52, XPD tagged with N-terminal 8xHis-V5-spacer-TEV)
pACEBac2-(8xHis)p52-F2A-p8
pACEBac2-p44-T2A-p34
pACEBac2-p62
pACEBac2-XPB
pACEBac2-(8xHis)XPD
pACEBac2-(8xHis)p52-F2A-p8-p44-T2A-p34
pACEBac2-(8xHis)p52-F2A-p8-p44-T2A-p34-p62
pACEBac2-XPB-(8xHis)XPD
pACEBac2-(8xHis)p52-F2A-p8-p44-T2A-p34-p62-XPB-(8xHis)XPD
Subunit XPD (tagged with C-terminal TEV-10xHis-V5)
pACEBac2-(10xHis)XPD
Sub-complex CAK (subunit MAT1 tagged with N-terminal 8xHis-V5-spacer-TEV)
pACEBac2-(8xHis)MAT1
pACEBac2-cdk7-T2A-cyclin H
pACEBac2-(8xHis)MAT1-cdk7-T2A-cyclin H
Sub-complex CAK (subunit cdk7 tagged with N-terminal 8xHis-V5-spacer-TEV)
pACEBac2-MAT1
pACEBac2-(8xHis)cdk7-T2A-cyclin H
pACEBac2-MAT1-(8xHis)cdk7-T2A-cyclin H
Sub-complex CAK (subunits MAT1, cdk7 tagged with N-terminal 8xHis-V5-spacer-TEV)
pACEBac2-(8xHis)MAT1
pACEBac2-(8xHis)cdk7-T2A-cyclin H
pACEBac2-(8xHis)MAT1-(8xHis)cdk7-T2A-cyclin H
Sub-complex TFIIH Core (subunits p52, XPD tagged with N-terminal 6xHis-V5-TEV and subunit XPD tagged with a C-terminal Twin-Strep-tag®)
pACEBac2-(6xHis)p52-F2A-p8
pACEBac2-p44-T2A-p34

pACEBac2-p62
pACEBac2-XPB
pACEBac2-(Strep)XPD
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34-p62
pACEBac2-XPB-(Strep)XPD
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34-p62-XPB-(Step)XPD
Monomeric XPC (tagged with N-terminal 8xHis-V5-spacer-TEV)
pACEBac2-(8xHis)XPC
Heterodimer XPC-HR23B (both tagged with N-terminal 8xHis-V5-spacer-TEV)
pACEBac2-(8xHis)XPC
pACEBac2-(8xHis)HR23B
pACEBac2-(8xHis)XPC-(8xHis)HR23B
Monomeric XPC (tagged with N-terminal 8xHis-V5-spacer-TEV, TCP)
pACEBac2-(8xHis)(TCP)XPC
Sub-complex TFIIH Core (subunits p52, p62 tagged with N-terminal 6xHis-V5-TEV, XPB tagged with N-terminal TCP)
pACEBac2-(6xHis)p52-F2A-p8
pACEBac2-p44-T2A-p34
pACEBac2-(6xHis)p62
pACEBac2-(TCP)XPB
pACEBac1-XPD
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34-(6xHis)p62
pACEBac2-(TCP)XPB-XPD
pACEBac2-(6xHis)p52-F2A-p8-p44-T2A-p34-(6xHis)p62-(TCP)XPB-XPD

APPENDIX B

Table B2: Recombinant bacmids employed to transfect *Sf9* insect cells for the expression of the different proteins, complexes and sub-complexes described in this thesis. Purification tags (highlighted in red) are indicated in brackets before the subunit they were attached to.

Subunit / Sub-complex	Bacmid	Genes transposed
TFIIH Core	AcMNPV	(6xHis) <i>p52-F2A-p8-p44-T2A-p34-p62-XPB-(6xHis)XPD</i>
TFIIH Core	AcMNPV	(6xHis) <i>p52-F2A-p8-p44-T2A-p34-(6xHis)p62-XPB-XPD</i>
TFIIH Core	AcMNPV	(8xHis) <i>p52-F2A-p8-p44-T2A-p34-p62-XPB</i>
TFIIH Core	AcMNPV	(8xHis) <i>p52-F2A-p8-p44-T2A-p34-p62-XPB-(8xHis)XPD</i>
TFIIH Core	AcMNPV	(6xHis) <i>p52-F2A-p8-p44-T2A-p34-p62-XPB-(Step)XPD</i>
TFIIH Core	AcMNPV	(6xHis) <i>p52-F2A-p8-p44-T2A-p34-(6xHis)p62-(TCP)XPB-XPD</i>
XPD	AcMNPV	(8xHis) <i>XPD</i>
XPD	AcMNPV	(10xHis) <i>XPD</i>
CAK	AcMNPV	(8xHis) <i>MAT1-cdk7-T2A-cyclin H</i>
CAK	AcMNPV	<i>MAT1-(8xHis)cdk7-T2A-cyclin H</i>
CAK	AcMNPV	(8xHis) <i>MAT1-(8xHis)cdk7-T2A-cyclin H</i>
XPC	AcMNPV	(8xHis) <i>XPD</i>
XPC	AcMNPV	(8xHis)(TCP) <i>XPC</i>
XPC-HR23B	AcMNPV	(8xHis) <i>XPC-(8xHis)HR23B</i>

APPENDIX C: PCR oligonucleotides

Table C1: TFIIH Core sub-complex PCR primers

Oligonucleotide	Sequence 5' to 3'
XPD Fwd	TACATGCGGGAGCTCAAACGC
XPD Rev	TCTGCTCTATCCTCTTCAGC
XPB Fwd	GGCCATATCTTCTTGGAAGC
XPB Rev	GCCAGGACTTTCTGTAAGAGC
p62 Fwd	CCTCATCTGAAGTTTTGC
p62 Rev	CGTTTTCTTCATCAGACGCCG
p52-p8 Fwd	GGATCTGGCACACACAGCTGC
p52-p8 Rev	CTTGAATGATGAACTTCTTCC
p44-p34 Fwd	TTAGAGTATCTGTTATTGG
p44-p34 Rev	GACCCTTTATGTCACTTTTGG

APPENDIX D: DNA substrates

Table D1: Oligonucleotides used in this thesis

Strand	Sequence (5' to 3')
A	CGCCCTCGGTGCTAAGTTCATGC/iFluorT/GGTACTCGGAGTATCCCGGG
B	/56-FAM/CGCCCTCGGTGCTAAGTTCATGCTGGTACTCGGAGTATCCCGGG
C	CCCGGGATACTCCGAGTACCAGCATGAACTTAGCACCGAGGGCG
D	CCCGGGATACTCCGTTTTTTTTTTTTTTTTTTTAGCACCGAGGGCG
E	CCCGGGATACTCCGAGTACTTTGAACTTAGCACCGAGGGCG
F	CCCGGGATACT/By3/CCGAGTACTTTCATGAACTTAGCACCGAGGGCG
G	CCCGGGATACTCCGAGTACTTTCATGAACTT/By3/AGCACCGAGGGCG
SD-Dab	AGCTACCATGCCTGCACGAATTAAGCAATTCGTAATCATGGTCATAGCT/Dab cyl/
SD-Cy3	/Cy3/AGCTATGACCATGATTACGAATTGCTTGAATCCTGACGAACTGTAG
SD1	AGCTACCATGCCTGCACGAATTAAGCAATTCGTAATCATGGTCATAGCT/Dab cyl/
SD2	AGCTACCATGCCTGCACGAATTAAGCAA/iFluorT/TCGTAATCATGGTCATAGC T/Dabcyl/
SD3	/Cy3/AGCTATGACCATGATTACGAA/iFluorT/TGCTTGAATCCTGACGAACTG TAG
SD4	/Cy3/AGCTATGACCATGATTACGAATTGCTTGAATCCTGACGAACTGTAG
XPD1	AGCTACCATGCCTGCACGAATGAGGAAGGGAGGAAAGGAAGA/Dabcyl/
XPD2	/Cy3/TCTTCCTTTCCTCCCTTCTC
XPD3	AGCTACCATGCCTGCACGAATCTCCTTCCTCCTTTCCTTCT/Dabcyl/
XPD4	/Cy3/AGAAGGAAAGGAGGGAAGGAG
REV1	/Cy3/AGAAGGAAAGGAGGGAAGGAGTAAGCACGTCCGTACCATCGA
REV2	CTCCTTCCTCCTTTCCTTCT/Dabcyl/
REV3	/Cy3/TCTTCCTTTCCTCCCTTCTAAGCACGTCCGTACCATCGA
REV4	GAGGAAGGGAGGAAAGGAAGA/Dabcyl/

APPENDIX D

Table D2: DNA substrates used in this thesis

Substrate	Constituent oligonucleotides
S1	A, C
S2	B, C
S3	A, D
S4	B, D
S5	A, E
S6	B, E
S7	A, F
S8	A, G
SD	SD-Dab, SD-Cy3
SD13	SD1, SD3
SD24	SD2, SD4
XPD12	XPD1, XPD2
XPD34	XPD3, XPD4
REV12	REV1, REV2
REV34	REV3, REV4

