Coexpression and Reconstitution of Multiprotein Complexes

in Eukaryotic Nucleotide Excision Repair

BY

YOONJUNG SHIM

B.S., UNIVERSITY OF ILLINOIS AT CHICAGO 2010

DISSERTATION THESIS

Submitted as partial fulfillment of the requirements

for the degree of Doctor of Philosophy in Chemistry

in the Graduate College of the

University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:

Dr. Jung-Hyun Min, Chair and advisor

Dr. Wonhwa Cho

Dr. Leslie Wo-Mei Fung

Dr. Lawrence W. Miller

Dr. Brian Kay, Biological Sciences

This thesis is dedicated to
my husband for love, support, patience and sacrifice,
my parents for unconditional love and prayers,
my in-laws for encouragement and support,
all other family members and friends,
my advisor Prof. Min for guidance during my doctorate research
and God for everything

ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. Jung-Hyun Min for giving me the great opportunity to develop my research interest and guiding me as a better scientist with insightful scientific knowledge and patience. I also would like to thank our group members, Dr. Xuejing Chen for teaching me, Dr. Debamita Paul, Anthony Savushkin and Amirrasoul Tavakoli.

I greatly appreciate my dissertation committee members – Prof. Wonhwa Cho, Prof. Leslie Wo-Mei Fung, Prof. Lawrence W. Miller, and Prof. Brian Kay for their valuable time and support.

Finally, I also thank all my family, friends and church members for their great support.

Contribution of Authors

Chapter 2: Parts of this chapter were previously published as Shim, Y. Duan, MR, Chen, XC, Smerdon MJ, Min JH. (2012) Polycistronic coexpression and nondenaturing purification of histone octamers. Journal of Anal Biochem 427 (2), 190-192. Dr. Mingrui Duan and her advisor Dr. Michael J. Smerdon contributed to validation of histone octamers with NCP reconstitution. Dr. Xuejing Chen and Dr. Jung-Hyun Min contributed to preparation of histone octamers.

Chapter 5: DNA constructs were provided by Dr. Jung-Hyun Min, Dr. Beom Seok Park and Dr. Xuejing Chen. Full length Rad4-Rad23 onstruct was cloned by Dr. Jung-Hyun Min, full length yeast Tfb1 construct was cloned by Dr. Beom Seok Park and truncated Rad4-Rad23 constructs were cloned by Dr. Xuejing Chen.

LIST OF ABBREVIATIONS

ATP	Adenosine Triphosphate
BHD1	β-hairpin domains 1
BHD2	β-hairpin domains 2
BHD3	β-hairpin domains 3
Cy3	Cyanine 3
Cy5	Cyanine 5
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
DTT	1,4-Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium Bromide
FBS	Fetal bovine serum
FRET	Förster resonance energy transfer
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
GG-NER	Globle-genome NER
GST	Glutathione S-transferase tag
LiP	Limited proteolysis
MCS	Multiple cloning site
NCPs	Nucleosome core particles
NER	Nucleotide excision repair
PCR	Polymerase chain reaction
PH domain	Pleckstrin homology domain
RBS	Ribosome-binding sequence
RNAPII	RNA polymerase II
RPA	Replication Protein A
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TC-NER	Transcription-coupled NER
TE	Tris-EDTA
TFIIA	General transcription factor IIA
TFIIB	General transcription factor IIB
TFIID	General transcription factor IID
TFIIE	General transcription factor IIE
TFIIH	General transcription factor IIH
TGD	Transglutaminase-homology domain
Tm	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultra-violet

LIST OF ABBREVIATIONS - Continued

XPA	Xeroderma Pigmentosum, Complementation Group A
XPC	Xeroderma Pigmentosum, Complementation Group C
XPF	Xeroderma Pigmentosum, Complementation Group F
XPG	Xeroderma Pigmentosum, Complementation Group G

SUMMARY

Genetic transactions such as DNA replication, transcription, recombination, and repair are often initiated and controlled by dynamic interactions among intricate multiprotein complexes and chromatin DNA. While it is crucial to study the structural, biochemical and biophysical mechanisms of the interactions involving these molecules, such studies are often challenged by the difficulties in reconstituting large protein and protein-DNA complexes in sufficient purity and quantity.

In this thesis, I focused on developing methods to prepare key such complexes that have seminal roles in several genetic transactions in eukaryotes, including transcription and nucleotide excision repair (NER). First, I describe the co-expression of histone octamer complex and the reconstitution of the nucleosome core particles (NCPs). NCPs are basic organization and packaging units of DNA in all eukaryotic cells, consisting of ~147 bp DNA wrapped around histone octamers. Nucleosomes are folded into compact, highly organized chromatin structures, which form the working platform for various DNA mechanisms including transcription and NER. Second, I discuss the co-expression and purification of the 5-subunit core of the general transcription factor II H complex (TFIIH) complex using MultiBac expression system. TFIIH is essential for trancription initiaton as well as NER DNA damage processing. Lastly, I describe preliminary studies on the interactions between TFIIH and Rad4-Rad23 complex, a key initial damage recognition factor in NER. The TFIIH-Rad4 interaction constitutes a key initiation step in NER that removes various environmentally induced DNA lesions and protect genomic integrity in cells.

Table of Content

1 Introduction 1

1.1 Importance and difficulties on studying multiprotein complexes 1

1.2 Multiprotein complex preparation 1

1.3 Nucleotide excision repair (NER) in chromatin context 2

2 Polycistronic Coexpression and Nondenaturing purification of Histone octamers 6

2.1 Introduction 6

2.1.1 Multiprotein complex reconstitution by coexpression in *E.coli* 6

2.1.2 Histone octamers 8

2.1.3 Nucleosomes 9

2.2 Methods 10

2.2.1 Polycistronic plasmid construction 10

2.2.2 Histone expression 13

2.2.3 Cell lysis and Nickel-affinity chromatography 13

2.2.4 Thrombin digestion 14

2.2.5 Size exclusion chromatography with Superdex 200 10/300 GL 14

2.2.6 NCP reconstitution by stepwise salt dialysis 14

2.2.7 FRET measurements 15

2.3 Results 15

2.4 Discussions 22

3 Improved method for Large scale, High purity NCP preparation 25

3.1 Introduction 25
3.2 Methods 26
3.2.1 Attempts on reconstituting NCP with my histone octamers 26
3.2.2 Control NCP preparation with lyophilized histone powder (conventional method) 27
3.2.2.1 NCP reconstitution by continuous salt dialysis 29
3.2.2.2 NCP purification by Model 491 Prep Cell 30
3.2.3 Improved NCP preparation with coexpressed histone octamers 30
3.3 Results 31
3.4 Discussions 40
4 Reconstitution of General Transcriptional Factor II H complex using Baculovirus expression system 42
4.1 Introduction 42
4.1.1 Multiprotein reconstitution with co-expression in insect cells 42
4.2 General transcription factor II H 44
4.2.1 What is general transcriptional factor II H? 44
4.2.2 Composition of TFIIH 44
4.2.3 TFIIH in transcription 45
4.2.4 TFIIH in Nucleotide Excision Repair 46
4.2.5 TFIIH and disease 46
4.3 Methods 47
4.3.1 Cloning 47

4.3.1.1 Multigene expression cassettes with multiplication module 47
4.3.1.2 Multigene expression cassettes with Cre-loxP recombination 49
4.3.2 Transposition and Transfection..... 52
4.3.3 Proteins expression test..... 53
4.3.4 Large scale protein expression..... 53
4.3.5 Cell lysis and Nickel-affinity chromatography (HisTrap FF 5 mL) 53
4.3.6 Dialysis 54
4.3.7 Cation exchange chromatography (SouceQ 24 mL)..... 54
4.3.8 Size exclusion chromatography (Superdex 200 10/300 GL) 54
4.4 Results..... 55
4.5 Discussions 61
5 Proving Interaction of TFIIH and XPC-Rad23 complex in GG-NER by Biochemical assay 64
5.1 Introduction..... 64
5.2 Methods..... 65
5.2.1 Full length his-Tfb1 with full length his-Rad4-Rad23 complex..... 65
5.2.1.1 Coinfection and colysis..... 65
5.2.1.2 Nickel pull down with HisTrap FF 5 mL colum..... 66
5.2.1.3 Dialysis 66
5.2.1.4 Size exclusion chromatography 66
5.2.2 Full length his-Tfb1 with truncated his-Rad4-Rad23 complex 67
5.2.2.1 Coinfection and colysis..... 67

5.2.2.2 Gravity Nickel Pull down with Ni-NTA agarose 68
5.2.3 Creating co-expressed cassettes of Tfb1-Rad4-Rad23 complex 68
5.2.4 Creating multigene TFIIH complex including Tfb1 69
5.3 Results and discussions 70
6 CONCLUSION 77
APPENDIX 78
1.1. PCR 78
1.2. Restriction Enzyme digestion 79
1.3. Ligation 80
1.4. Transformation and colony identification 81
1.5. Positive clone amplification 82
1.6. Protein production 82
1.6.1. Protein production from bacterial cell culture 82
1.6.1.2. Protein expression 83
1.6.2. Protein production from Insect cell culture 84
1.6.2.1. Transposition and bacmid production 84
1.6.2.2. Transfection and virus production 85
1.6.2.3. Protein expression from Hi5 monolayer cells 87
1.6.2.4. Protein expression from Hi5 suspension cells 88
1.6.3. large scale protein purification (8 L) 89
1.7. Reconstitution of histone octamer 91
1.7.1. Mixing H2A, H2B, H3.1 and H4 91

1.7.2. Dialysis and Concentration 92

1.7.3. HiLoad Superdex 200 16/60 92

1.8. Nucleosome core particle (NCP) reconstitution 93

1.8.1. Small scale set up 93

1.8.2. Large scale set up 93

1.8.3. Purification of NCPs 93

Cited Literature 96

VITA 102

List of Tables

Table 1: PCR primers used in this study..... 12

Table 2: Restriction enzymes and affinity tags used for the histones..... 12

Table 3: DNA oligomers used for α-satellite DNA..... 28

Table 4: Composition of TFIIH subunits in human and yeast..... 45

Table 5: Restriction enzymes used in this study to assemble expression cassettes by using multiplication module..... 49

Table 6: Vector, promoter and restriction enzyme sites used in this study. 50

List of Figures

Figure 1: Lesion recognition by XPC-Rad23B complex on compacted chromatin context..... 4

Figure 2: TFIIH recruitment to the lesion by XPC-Rad23B complex in NER pathway. 5

Figure 3: Schematic map of the histone-coexpressing polycistronic vector used in this study.... 11

Figure 4: Coexpression and copurification of histone octamers. 16

Figure 5: Histone octamers are coexpressed and copurified in simple steps. 17

Figure 6: Characterization of the copurified histones using size-exclusion chromatography. 19

Figure 7: Copurified histones can form nucleosome core particles (NCPs). 21

Figure 8: Characterization of the copurified histones using FRET. 22

Figure 9: Overview of conventional method and new method for histone octamer preparation. 24

Figure 10: Schematic diagram showing how to prepare α-satellite DNA. 28

Figure 11: small scale NCP reconstitution with copurified histone octamers and 601 DNA at different ratio. 33

Figure 12: α-satellite DNA preparation was analyzed on native PAGE. 34

Figure 13: Kurumi histone octamers preparation with HiLoad Superdex 16/60 200 pg. Histone octamers were reconstituted by individually lyophilized histones. 35

Figure 14: Control nucleosomes prepared with α-satellite DNA and individual lyophilized histone powder. 36

Figure 15: Histone octamer preparation with one more purification step (HiLoad Superdex 200). 37

Figure 16: pLMG601-23 digestion by EcoRV-HF. Digestion was analyzed on a 1% agarose gel. 38

Figure 17: Small scale NCP reconstitution with different histone octamers:601 DNA ratio. 39

Figure 18: large scale NCP reconstitution analyzed on a 6 % nave-PAGE..... 40

Figure 19: Generating multigene expression cassettes. 51

Figure 20: Creating Core TFIIH complex multigene expression cassettes. 52

Figure 21: Analysis of Cre loxp recombination product YS63. 56

Figure 22: Coexpression and lysis of core-TFIIH on monolayer Hi5 and suspension Hi5 cells **(A)** Core TFIIH complex expression on monolayer Hi5 cells. 57

Figure 23: Purification of core TFIIH complex (YS63) by HisTrap FF 5mL and SourceQ 24mL. 58

Figure 24: Purification of YS63 (core TFIIH) with 1st Superdex 200 column. 59

Figure 25: Purification of YS63 (core TFIIH) with 2nd Superdex 200 column. 60

Figure 26: Purification of YS63 (core TFIIH) with 3rd Superdex 200 column. 61

Figure 27: Overall procedure to study Tfb1 and Rad4-Rad23 interaction. 67

Figure 28: Truncated Rad4-Rad23 complex used for pull-down assay with Tfb1. 68

Figure 29: Creating Tfb1with truncated Rad4-Rad23 coexpression cassettes. 69

Figure 30: Creating three and four multigene expression cassettes containing Tfb1. 70

Figure 31: Investigating Interaction between full length Tfb1 and full length Rad4-Rad23 complex by pull-down assay and size exclusion chromatography. 71

Figure 32: Coinfection and colysis of his-Tfb1 with truncated Rad4-GST-Rad23. 72

Figure 33: Purification of three subunits (Tfb1, Tfb4 and Ssl1) of TFIIH assembled by baculovirus system. 74

Figure 34: Purification of four subunits (Tfb1, Tfb4, Ssl1 and Tfb5) of TFIIH assembled by baculovirus system. 75

1 Introduction

1.1 Importance and difficulties on studying multiprotein complexes

Cellular processes such as cell duplication, metabolism, and signaling are controlled by dynamic interactions between intricate cellular machineries. Protein-protein interactions (PPI) can be classified based on the half-life of the complexes. Proteins with very stable lasting interactions happen in their complexed form, multiplex, while proteins with transient interactions interact and dissociate transiently *in vivo* [1]. While it is crucial to study the structural, biochemical and biophysical properties of PPI, studies involving multiprotein complexes often pose a significant challenge due to their required preparations, which often requires micro- to milligram quantities of pure, homogeneous protein complex samples. [2]

1.2 Multiprotein complex preparation

Proteins can be isolated from their endogenous host, but in most cases, amount of proteins available from cells are very limited. Often, a large volume of cells is required to isolate a sufficient amount of endogenous proteins from the culture, which it is costly and time consuming. In addition, endogenous sources may not be amenable to extensive protein engineering (*e.g.*, mutational studies) to study the intricate functions. Moreover, proteins may be easily degraded due to proteolytic activity by enzymes even in the presence of protease inhibitors [3]. Recombinant protein expression has thus been a popular tool that can effectively overcome these drawbacks [4].

Recombinant protein expression relies on robust over-expression of proteins in separate host cells [5]. These methods are more beneficial than endogenous protein purification especially for proteins that cannot be produced in a large quantity, easily isolated or manipulated [4]. There

are two major ways to prepare multiprotein complexes. First method, each protein is expressed and purified separately in each respective cell. Then, the individually purified components are mixed together *in vitro* and purified further to reconstitute the complex [6]. Second method, the subunits of the multiprotein complex are co-expressed together inside the same cells and co-purified at once to obtain the complex [7] [8].

1.3 Nucleotide excision repair (NER) in chromatin context

Because DNA contains genetic information of organisms, maintaining its integrity and stability is very critical to life [9]. However, about 10^5 DNA in cells is constantly being damaged by various endogenous and exogenous factors [10] [11].

NER is the major mechanism in eukaryotes that protects the genome from the harms of various environmentally induced DNA lesions, such as UV irradiation, ionizing radiation, chemicals from cigarette smoke, automobile exhaust, and burnt food, [12] evolutionarily conserved from yeast to humans [13].

NER is categorized into two sub pathways: global genome NER (GG-NER) or transcription-coupled NER (TC-NER). GG-NER can arise through the genome, whereas TC-NER detects and repair lesions located on transcribed strand [14] [15]. The location of the lesion detected is the only difference and the remaining steps are the same [16]. Initiation of human GG-NER critically depends on the XPC-Rad23B (yeast ortholog Rad4-Rad23) complex and TC-NER is initiated by blockage of RNA polymerase II at lesions [16].

After a lesion is recognized, general transcription factor IIH (TFIIH) is recruited to the lesion by interacting with XPC-Rad23B to verify the damage and unwind DNA around the lesion. Two ATP-dependent helicase subunits, XPB and XPD, are responsible to unwind DNA around the lesion irreversibly. Then, preincision complex is assembled by recruitment of XPA,

RPA, and XPG proteins. XPA binds to 5' of the DNA bubble and interact with single strand DNA binding protein RPA. An endonuclease XPG is recruited by TFIIH. Upon recruitment of XPF by interaction with XPA, dual incision is mediated by XPF on 5' and by XPG on 3' of the lesion. This results in 24- to 32-mer oligonucleotides excision including the lesion. Then, DNA is synthesized by polymerase δ and ε to fill the gap and ligase I attaches newly synthesized DNA.

It is very crucial how lesions are recognized by XPC-Rad23B to initiate NER for successful repair, but understanding the detailed mechanism of how the lesions are recognized in compacted dynamic chromatin context is unknown (Fig. 1). Moreover, how huge TFIIH multiprotein complex is recruited to the lesion by XPC-Rad23B is also unrevealed (Fig. 2).

In this dissertation, I have focused on preparing multiprotein complexes involved in eukaryotic GG- NER to understand its detailed mechanism and structure. In Chapter 2, histone octamers were prepared by polycistronic coexpression system in *E.coli.* This new simple and fast method enabled me to make histone octamers in a day. In Chapter 3, nucleosome core particle reconstitution protocol has been established in our lab by using coexpressed histone octamers and Widom 601 DNA sequences. In Chapter 4, 5-subunit core TFIIH complex required for NER was prepared by using MultiBac expression system. In chapter 5, I have shown preliminary studies proving interactions between yeast Rad4-Rad23 complex and Tfb1, a subunit of TFIIH.

Altogether, preparation of multiprotein complexes, NCPs and TFIIH, will launch our future studies on NER initiation step.

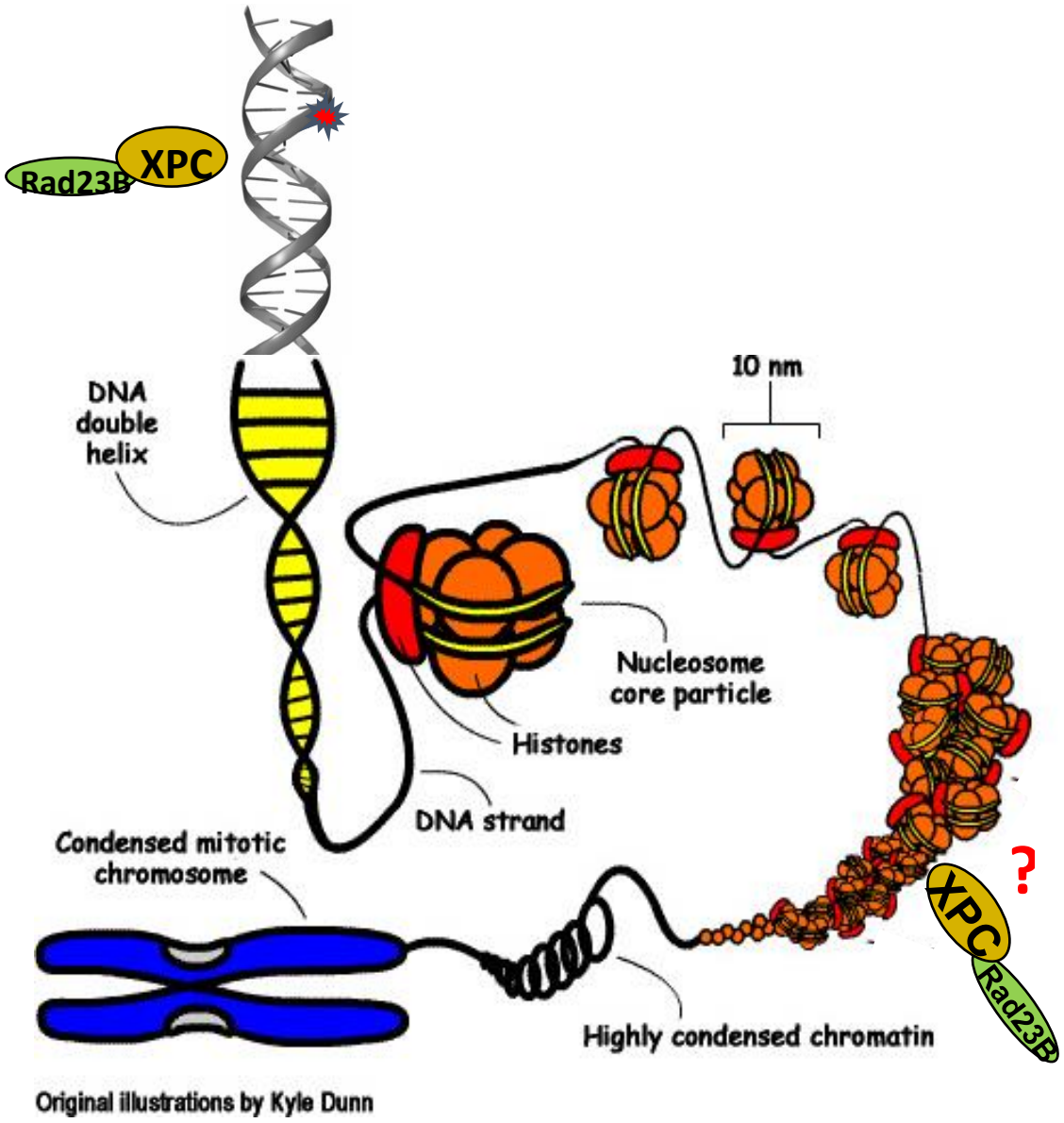


Figure 1: Lesion recognition by XPC-Rad23B complex on compacted chromatin context.

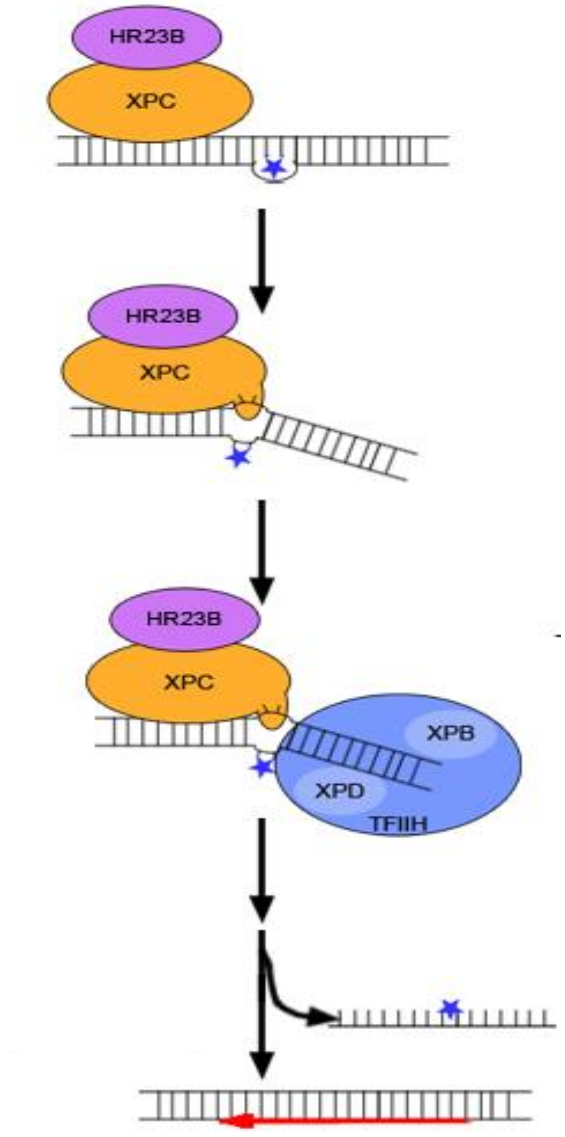


Figure 2: TFIIH recruitment to the lesion by XPC-Rad23B complex in NER pathway.

2 Polycistronic Coexpression and Nondenaturing purification of Histone octamers

Parts of this chapter were previously published as Shim, Y. Duan, MR, Chen, XC, Smerdon MJ, Min JH. (2012) Polycistronic coexpression and nondenaturing purification of histone octamers. Journal of Anal Biochem 427 (2), 190-192

2.1 Introduction

2.1.1 Multiprotein complex reconstitution by coexpression in *E.coli*

E.coli is the most widely used host organism [17]. It is well-established with many modified strains and cultivation approaches [18]. Rapid cell growth and short culturing time with low cost of *E.coli* enable quick and easy large-scale proteins production [19]. In addition, it usually produces high yield due to high protein expression [20]. Expression vector for each protein is constructed and individual proteins are expressed. They are then purified before mixing to form a complex and needs further purification. This method has been widely used [21], but there are some known drawbacks [7].

First, whole process is tiresome because each subunit has to be expressed and purified separately, and then, purified again after combining them together [22]. Moreover, this method can lead to form insoluble inclusion bodies and inactive aggregates from unfolded or misfolded proteins. It requires tedious optimization to make them soluble and fold properly [23]. In addition, protein yields after reconstitution are often minuscule [6]. Therefore, coexpressing in polycistronic system has been developed and will be discussed in the next section.

Multiprotein complex reconstitution by coexpressing together is an alternative method to multiprotein complex reconstitution from combining individually expressed proteins [24]. In the presence of protein folding enzymes and molecular chaperones, individual proteins can fold

properly into multiprotein complex in the cellular context [25]. In addition, time and effort can be saved through this method because complex needs to be purified only once after coexpressing subunits together. Thus, high yield of quality protein complex can be obtained [18].

A typical expression vectors is composed of one translation cassette containing a translational start signal (ribosome binding site, RBS) followed by coding region. One promoter is located before translation cassette and this is where DNA starts to be transcribed. Also, a transcriptional terminator is place after translation cassette. On the other hand, polycistronic expression vectors contains multiple translation cassettes (up to four) flanked by promoter and transcriptional terminator [26].

Polycistronic vectors that are already modified with multiple possible translation cassettes are available. This simplifies the process by only needing to clone the genes into the vector by molecular cloning technique using PCR and particular restriction enzymes [27]. As an alternative, polycistronic vectors can be made from typical monomeric expression vector by conventional method with endonucleases and ligases. First, position of the possible translation cassette should be decided and then cassettes can be created by introducing translation initiation signal (ribosome binding sites) in multiple cloning site (MCS). RBS can be included in forward primer for genes except for the genes positioned in the first cassette for PCR [7]. As a typical bacterial expression system, genes are transformed into competent *E.coli* and cells are amplified in proper media, such as LB and 2xTY. Afterward ,multiprotein complex expression is induced by adding IPTG (Isopropyl β-D-1-thiogalactopyranoside) [28].

There are several parameters that should be considered. The order of expression of genes can be a critical factor to affect the expression level of the complex. Therefore, other possible order of gene expressions should be tested if expression level is low [7]. Restriction enzyme sites

are used to create rooms for translation cassettes and the presence of these sites in genes should be examined. The order of expression of genes can be restricted [7].

2.1.2 Histone octamers

Most of eukaryotic genomes are packaged in the repeating units of nucleosomes, each of which contains approximately 147 base pairs of DNA tightly wrapped around an octameric histone complex [29] [29, 30]. Histone octamers in nucleosomes set the platform for genetic transactions such as replication, transcription and repair. Histones are often the aim of attack of proteins regulating these processes [31].

Studies of detailed genetic and epigenetic mechanisms mandate the preparation of purified homogeneous histone octamers to reconstitute nucleosomes and chromatin structures *in vitro*. Currently, there are two major ways to prepare histones to reconstitute nucleosomes. First, endogenous histones can be isolated from sources such as HeLa cells and calf thymus [32] [33]. However, these endogenous histone complexes are inhomogeneous in their composition and posttranslational modifications. In addition, they are not amenable to molecular cloning and engineering. A second method overcomes this limitation by using recombinant histones. In this method, four histone core subunits (H2A, H2B, H3, and H4) are individually overexpressed and purified from insoluble inclusion bodies of bacteria under denaturing conditions. Octameric complexes are then formed by refolding the denatured histones together [34] [35]. The procedure, however, is labor-intensive and lengthy; it involves chromatography procedures under high concentrations of denaturants (e.g., 6 M guanidinium chloride) and multiple rounds of dialysis and lyophilization that typically take more than a month to complete. The required amount of work can be exacerbated when generating octamers with various histone mutants.

To overcome the limitations of the existing methods, I have developed a simple and fast way to express and purify recombinant histone octamers that reduces the time to 1 day and this work has been published [36].

2.1.3 Nucleosomes

Human genomic DNA is over 3 billion base pairs with over two meters in length and it is organized into a nucleus of a cell which is usually around 6 μm in diameter [37] [9]. Genomic DNA is packaged into a nucleus by wrapping around histone octamers repeatedly every 160 to 240 base pairs throughout the genome [38]. This complex is called nucleosome and nucleosome core is connected to the neighboring nucleosome through a linker DNA [39] . Then, nucleosomes are folded into a highly organized complex of DNA and protein called a chromatin with a diameter of about 30 nm [40].

Chromatin structure in nature is very important for understanding DNA-related cellular processes, such as replication, transcription, recombination and DNA repair [41]. Because chromatin structure is dynamic, DNA is transiently released off the histone octamers by unfolding, disassembly, assembly and refolding. Thus, DNA can access to regulatory factors or regulatory sites [42]. Understanding how DNA in chromatin context gain access to perform these DNA metabolic process must be very crucial [43].

In this chapter, pET-29a vector was modified to a polycistronic vector and H2A, H2B, H3 and H4 genes from Xenopus laevis were incorporated. Then, histone octamers were coexpressed and copurified. Reconstituted histone octamers were purified and confirmed by reconstituting NCPs.

2.2 Methods

2.2.1 Polycistronic plasmid construction

Histone genes from Xenopus laevis were amplified by PCR from the plasmids originally made by the Luger laboratory [34]. To construct a polycistronic plasmid as depicted in Figure 3, RBS was incorporated along with a spacer sequence into the PCR primers for H2B, H3 and H4 (Table 1) [7]. An in-frame hexahistidine (His6) tag was also inserted by PCR in between the pre-existing S-tag and thrombin sites in a pET29a vector so as to precede the N-terminus of H2A (Table 1). The stop codon in the C-terminus of H4 was deleted, and instead, an in-frame thrombin site was added between the C-terminus of H4 and a pre-existing His6-tag in the pET29a vector, using PCR (Table 1). All PCR primers were purified using standard urea-PAGE protocols [44]. Histone genes were assembled the into the engineered pET29a vector using standard cloning techniques using restriction enzymes as indicated in Table 2.

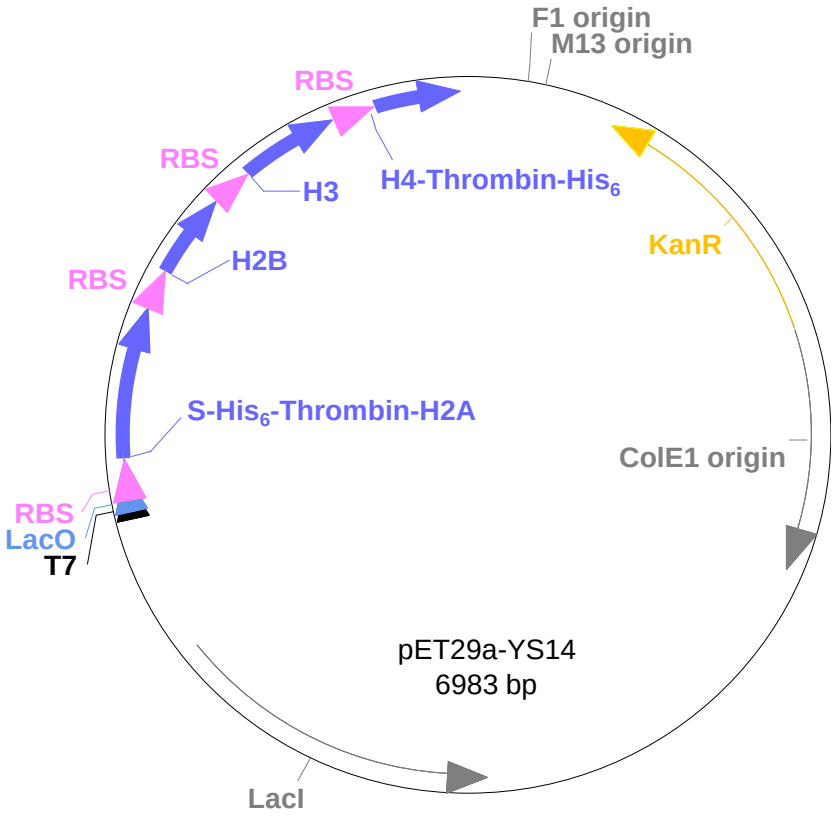


Figure 3: Schematic map of the histone-coexpressing polycistronic vector used in this study. Each histone gene is preceded by a ribosome-binding sequence (RBS). Both H2A and H4 contain His_6-tags which can be removed by thrombin digestion.

Table 1: PCR primers used in this study

Name	**Sequence**
BglII_ His_6_KpnI_F	C AGCCCAGATCTG CACCACCACCACCACCAC GGTACC CTGGTG
BglII_ His_6_KpnI_R	CACCAG GGTACC GTGGTGGTGGTGGTGGTG CAGATCTGGGCT G
NcoI_xeH2A_F	acttga CCATGGgt ATGTCAGGAAGAGGCAAA
EcoRI_xeH2A_R	cacaca GAATTC TCACTTGCTCTTGGCCGA
EcoRI_ RBS_xeH2B_F	acttga GAATTC AATAATTTTGTTTAACTT TAAGAAGGAGATATACAT ATGGCCAAGTCCGCTCCA
SalI_xeH2B_R	cacaca GTCGAC TTACTTGGCGCTGGTGTA
SalI_ RBS_xeH3_F	acttga GTCGAC AATAATTTTGTTTAACTT TAAGAAGGAGATATACAT ATGGCCCGTACCAAGCAG
NotI_xeH3_R	cacaca GCGGCCGC CTAAGCCCTCTCGCCTCG
NotI_ RBS_xeH4_F	Acttga GCGGCCGC ǂ 8AATAATTTTGTTTAACTT TAAGAAGGAGATATACAT ATGTCTGGTCGTGGTAAA
XhoI_Thrombin-xeH4_R	cacaca CTCGAG GCTGCCGCGCGGCACCAGGCCGCTGCT ACCACCGAAACCGTACAG

*Restriction enzyme sites are underlined and RBS's are indicated in red.

Table 2: Restriction enzymes and affinity tags used for the histones.

Histone	**(5') Restriction Enzyme**	**(3') Restriction Enzyme**	**Affinity tags**
H2A	NcoI	EcoRI	An N-terminal S-tag followed by a His_6-tag and a thrombin cleavage site.
H2B	EcoRI	SalI	None
H3	SalI	NotI	None
H4	NotI	XhoI	A C-terminal His_6-tag preceded by a thrombin site.

2.2.2 Histone expression

The resulting plasmid encoding all four core histones was transformed into BL21(DE3) pLysS cells and plated on to an LB agar plate containing kanamycin (50 µg/mL) and chloramphenicol (25 µg/mL). The plate was incubated overnight (~16 hours) at 37 °C. For a 1 L scale preparation, one colony was inoculated into 10 mL 2xTY media containing kanamycin (50 µg/mL) and chloramphenicol (25 µg/mL). This starter culture was shaken at 170 rpm for 4-5 hours at 37 °C until slightly cloudy. The culture was then amplified to 1 L of the same media and was grown for another 6-7 hours at 37 °C. When the *OD600* reached around 0.4, histone coexpression was induced by adding 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The culture was further shaken at 170 rpm at 37 °C for overnight (~20 hours). Cells were harvested by centrifugation at 4,500 x g for 10 min at 4 °C. Cell pellets were processed immediately or stored at -80 °C for future purification.

2.2.3 Cell lysis and Nickel-affinity chromatography

Cell pellets were resuspended in 60 mL of lysis buffer (20 mM Tris-HCl pH 8.0, 2.0 M sodium chloride, 1 mM phenylmethanesulfonylfluoride or Pefabloc SC (Centerchem) and 0.5 mM tris (2-carboxyethyl)phosphine). Resuspended cells were then lysed by EmulsiFlex-C3 high pressure homogenizer and clarified by centrifugation at 38,000 rpm (192,500 x g) for 1 hour at 4 °C in Sorvall Discovery 90SE ultracentrifuge (Thermo Scientific) equipped with a F40L-8x100 rotor. The supernatant was collected and then imidazole stock solution was added to adjust the concentration to 30 mM. The clarified lysate was then loaded onto a 5 mL HisTrap FF column (GE Healthcare) pre-equilibrated in the Ni-buffer A (20 mM Tris-HCl pH 8.0, 2.0 M sodium chloride, 0.5 mM tris(2-carboxyethyl) phosphine) followed by a 10 column volumes (CV) wash with Ni-buffer A containing 30 mM imidazole and lastly, another 10 CV wash with Ni-buffer A.

afterwards, the bound proteins were eluted by increasing the imidazole concentration from 30 mM to 500 mM linearly over 23.5 CV. Each faction was analyzed by 18% SDS-PAGE. Electrophoresis was run at 150 V for 80 min.

2.2.4 Thrombin digestion

Thrombin digestion was carried out by adding purified thrombin (Sigma) in 25:1 mass ratio and incubating the samples at 4 °C for 3 hours. The digestion was confirmed by 18 % SDS-PAGE. Electrophoresis was run at 150 V for 80 min.

2.2.5 Size exclusion chromatography with Superdex 200 10/300 GL

Thrombin digested histones were then concentrated up to 3 mg/mL with ultrafiltration using Amicon YM50 membrane (MWCO 50 kDa) at 4 °C (Millipore). The concentrated sample was then injected onto a Superdex 200 10/300GL column pre-equilibrated in the SD200 buffer (20 mM Tris-HCl pH 8.0, 2.0 M sodium chloride, 10 % glycerol, 1 mM 1,4-Dithiothreitol). Each fraction was analyzed by 18% SDS-PAGE. Electrophoresis was run at 150 V for 80 min. The peak fractions were pooled and concentrated up to 8 mg/mL.

2.2.6 NCP reconstitution by stepwise salt dialysis

Nucleosome reconstitution was done essentially by our collaborator, Dr. Duan from Dr. Smerdon laboratory as described as described in [35] [45]. The histones and DNA were mixed in the molar ratio of histone octamer : DNA at 1.1:1 and dialyzed sequentially against TE buffers (10 mM Tris-HCl pH 8.0 and 1 mM EDTA) containing 1 M, 600 mM and 0 mM sodium chloride, each for at least one hour. The last dialysis was usually done overnight. The results were verified on 5% native polyacrylamide gels run in 0.25x TBE at 130 V for 1 hour at room temperature.

2.2.7 FRET measurements

This work was also done by Dr. Duan. Te DNA substrate was internally labeled at positions 6 and 81 in a 601 sequence with Cy3 and Cy5 respectively (Fig. 8A) [46]. The FRET by the dye pairs was measured as described to validate the formation of NCP by the histones [46].

2.3 Results

The polycistronic vector to coexpress all four histone genes was constructed by incorporating the genes into the multiple cloning site of a pET vector under the control of a single T7 promoter. Each histone-encoding sequence was preceded by a ribosome-binding site to enable the coexpression of the proteins. To facilitate the purification, H2A and H4 were engineered to contain a hexahistidine (His_6) tag and a thrombin site in the N and C termini, respectively. The resulting polycistronic vector indeed showed coexpression of all four histones from bacteria when induced by isopropyl β-d-1-thiogalactopyranoside (Fig. 4A). Overexpression of histones was optimal when the culture was induced at $OD_{600} \cong 0.4$ and grown overnight at 37 °C.

The induced cells were subsequently collected and lysed in a high-salt buffer containing 2 M sodium chloride. Under this condition, histones can dissociate from bulk DNA in cells [29], and it was observed that the histones remained mostly in solution instead of inclusion bodies as they do when expressed by themselves (Fig. 4B).

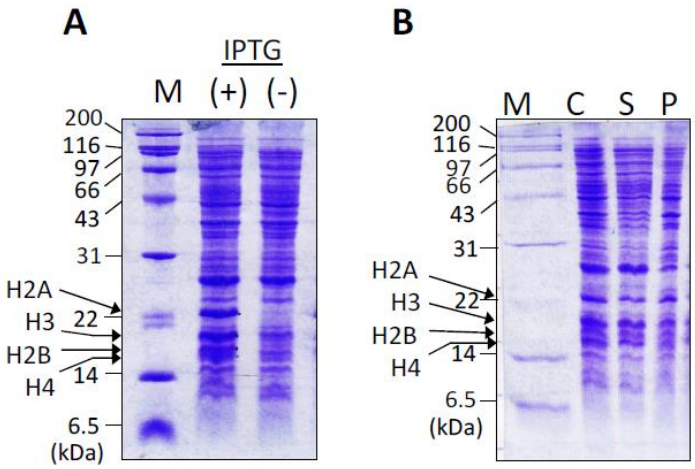


Figure 4: Coexpression and copurification of histone octamers.
(A) All four histones were induced and overexpressed upon adding 0.4 mM IPTG. Induced and uninduced cells are indicated by (+) and (–) IPTG, respectively. M: molecular weight marker. (**B**) Most histones were recovered in the soluble fraction of the cell lysate. C: total cell lysate, S: cell supernatant, P: cell pellet fractions. All lanes represent 50 µl of cell culture.

This allowed us to bypass denaturation and refolding procedures used in conventional methods [35] [45]. Instead, histones could be purified under native conditions using the following two chromatography steps. First, the histone-containing cell extract was loaded onto a nickel affinity chromatography column in the presence of 2 M sodium chloride and 30 mM imidazole. The column was subsequently washed with 20 column volumes of buffer, which effectively removed most other proteins and the majority of cellular DNA from histones. His_6 tags on both H2A and H4 enabled such extensive washing because the interaction between H2A–H2B and H3–H4 can be labile under these conditions. The bound proteins were then eluted over a linear imidazole gradient. Fractions eluted at 110 to 170 mM imidazole contained clean histones of roughly stoichiometric ratio, whereas complexes consisting mainly of H2A–H2B eluted before 110 mM imidazole and those consisting of H3–H4 eluted after 170 mM imidazole (Fig. 5B). The fractions were pooled containing stoichiometric ratios of the four histones and

concentrated them using ultrafiltration. The concentrated histones were subsequently digested with thrombin to remove the affinity tags on H2A and H4. Concentrating histones before digestion significantly accelerates the speed of digestion, and the digestion could be completed in 3 hours (Fig. 5C).

Thrombin-digested samples were subsequently purified over a Superdex 200 size exclusion column. All histones eluted as a single peak at a 12.8-ml elution volume (Fig. 5D).

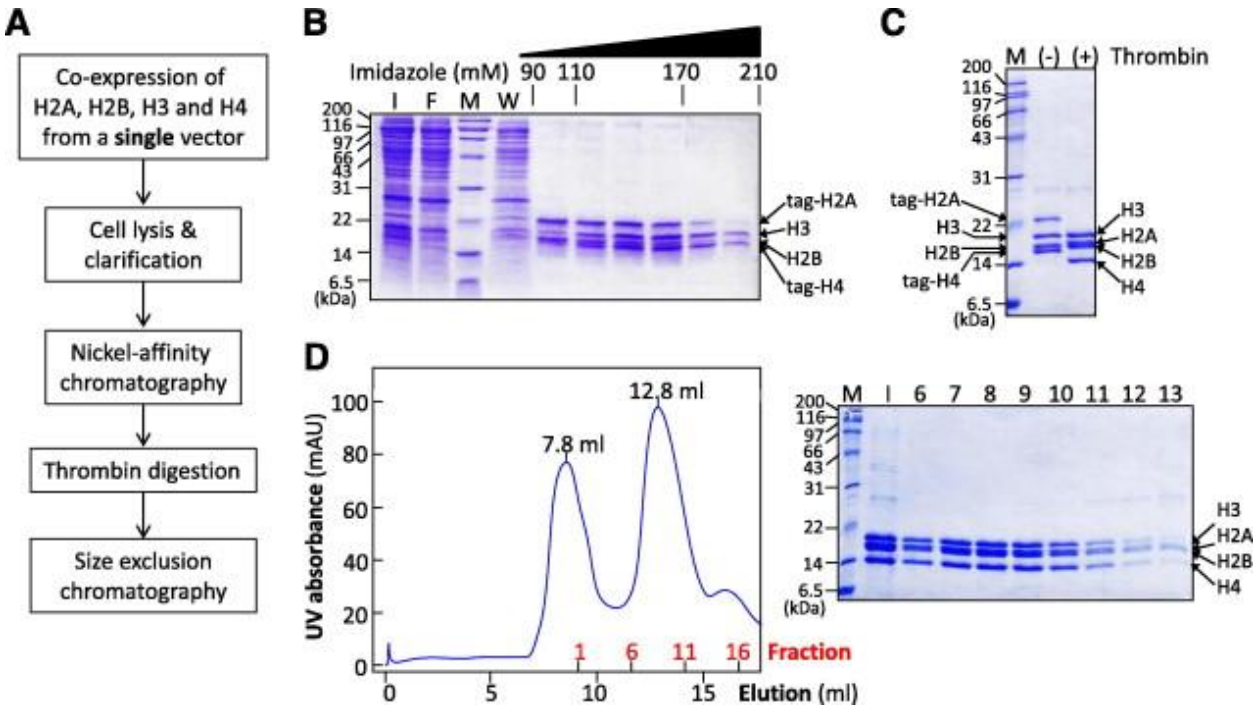


Figure 5: Histone octamers are coexpressed and copurified in simple steps.
(A) Overall procedure of histone octamer preparation presented in this paper. **(B)** Cell supernatant from histone-overexpressing bacteria was first purified over a nickel affinity column. SDS–PAGE shows that histones that contain stoichiometric ratios of histones eluted at 110 to 170 mM imidazole. M, molecular weight marker; I, input; F, flow-through; W, wash. **(C)** The affinity tags on H2A and H4 were removed by thrombin digestion. (−) and (+) thrombin indicate histone samples before and after thrombin digestion, respectively. **(D)** Thrombin-digested histones were finally purified over a Superdex 200 column. The chromatogram (left) and SDS–PAGE (right) show that stoichiometric histone complexes elute at 12.8 mL. UV, ultraviolet; M, molecular weight marker; I, input; lanes 6 to 13 indicate fraction numbers.

This corresponded to 111 kDa, closely matching the calculated molecular weight of the octameric complex (111 kDa) (Fig. 6A). Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis showed stoichiometric ratios of the histones across the histone peak fractions. Together with the molecular weight of the complex, the result indicates that the peak consists mainly of histone octamers. This elution profile is similar to the profile of octamers prepared by a conventional method [45], including the fact that the histone peak was preceded by a void volume peak at 7.8 mL that may contain nonspecific high-molecular-weight aggregates. When the purified histones were subjected to a second round of size exclusion chromatography, the 7.8-mL peak was absent and the histone peak remained at 12.8 mL with a stoichiometric ratio of the subunits (Fig. 6B). This result indicates that one round of size exclusion chromatography can efficiently purify the histone octamer and that the octamer remains intact over these procedures. Finally, SDS–PAGE analysis of the copurified histones and histones prepared by a conventional method (hereafter “control histones”) [46] confirmed their equivalence in the composition and sizes (Fig. 6C).

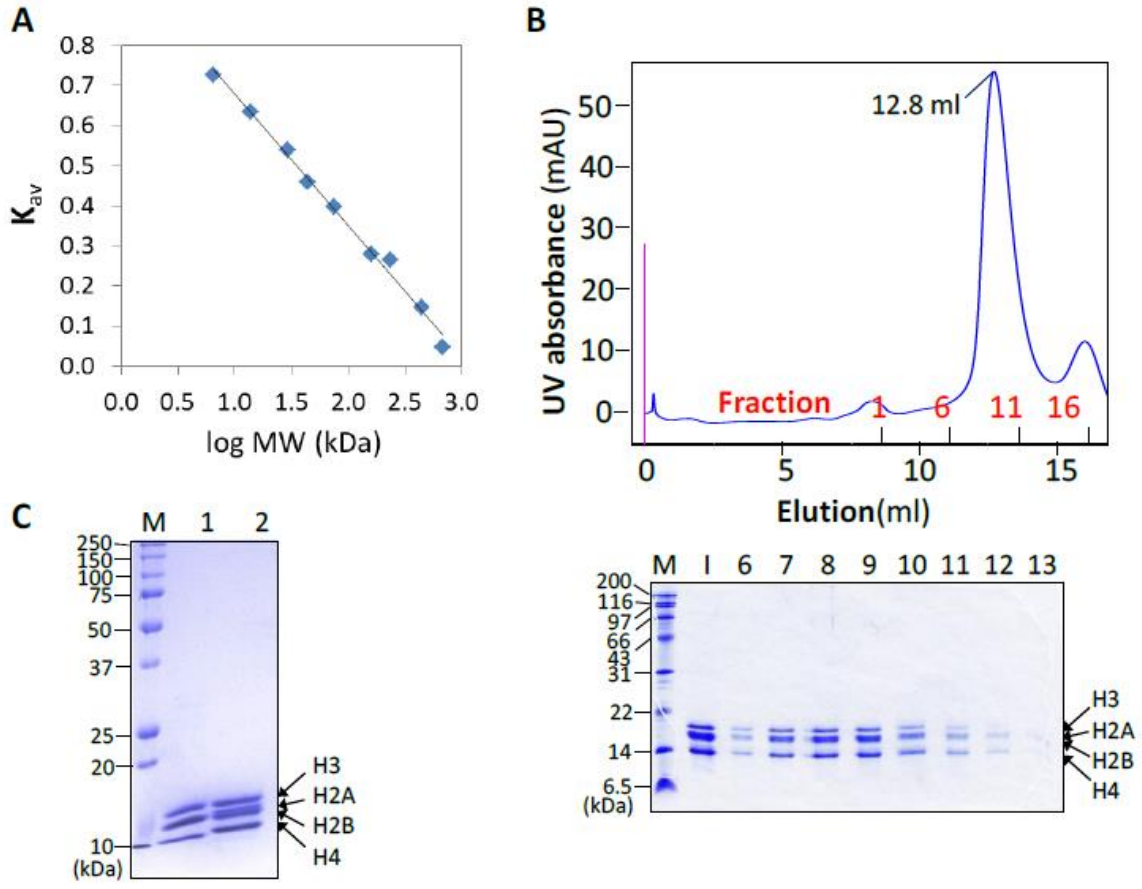


Figure 6: Characterization of the copurified histones using size-exclusion chromatography. **(A)** Calibration curve for the Superdex 200 column used in this study. According to this curve, the molecular weight of the histone complexes eluting at 12.8 ml (Fig. 5D) is estimated to be 111 kDa, closely matching the expected molecular weight of the octamer (111 kDa). Kav and logMW were calculated as described in the product manual for the size-exclusion column molecular weight calibration kit (GE Healthcare). **(B)** Histones purified as shown in Fig. 5D, were subject to a second size-exclusion chromatography. The chromatogram (top) and SDS-PAGE analysis of the fractions (bottom) show that the four histones again elute at 12.8 mL while maintaining the stoichiometric ratio. However, the void volume peak observed in Fig. 5D no longer exists. M, molecular weight marker; I, input, 7-14 indicate fraction numbers. **(C)** SDS-PAGE of histones prepared by a conventional method [6] (lane 1) and the histones copurified in this study (lane 2).

To test whether the copurified histones could form nucleosomes, Dr. Duan compared these histones with the control histones in a nucleosome reconstitution assay [35] [45]. In this assay, histone samples were first mixed with 147-bp double-stranded DNA containing a strong nucleosome-positioning “601” sequence [47] in a 1.1:1 ratio in a buffer containing 2 M Sodium

chloride and dialyzed in a stepwise fashion to gradually remove the salt. The reconstitution products were analyzed on native PAGE. The products of the copurified histone sample migrated similarly to those formed by the control histones (Fig. 7A). This result indicates that the copurified histones can form nucleosome core particles (NCPs) similarly to the histones from conventional methods.

To further validate our results, NCPs were reconstituted using a 601 DNA sequence labeled with a fluorescence donor (Cy3) and an acceptor (Cy5) at specific positions (Fig. 6A) [46]. The 601 DNA sequence has a strong propensity to bind to histone octamers in a specific orientation [48] and shortens the distance between the dye pairs from approximately 26 to 3 nm (Figs. 8A and 8B). Thus, a dramatic increase in the FRET signal is expected on donor excitation when the DNA is specifically assembled into NCPs, but not if the binding occurs in a nonspecific manner. The NCPs reconstituted with copurified histones show a robust FRET signal similar to that of the control histones, further supporting that these copurified histone octamers form canonical NCP structures (Fig. 7B).

FRET of the DNA was also measured the in the reconstituted NCPs as the salt concentration increased. This can also serve as a diagnostic of the nucleosome structure because nucleosomes show increased dissociation with increased salt concentration [49]. After normalization of the FRET signal to the lowest salt concentration (0 mM NaCl (Sodium chloride)), there was no difference between the salt-induced dissociation profiles of the copurified histones and of the control histones (Fig. 7C and 8C).

Based on these three criteria (native gel migration, FRET, and salt dissociation profiles), it was concluded that our copurified histones reconstitute NCPs that are essentially identical to those formed by histones prepared by conventional methods.

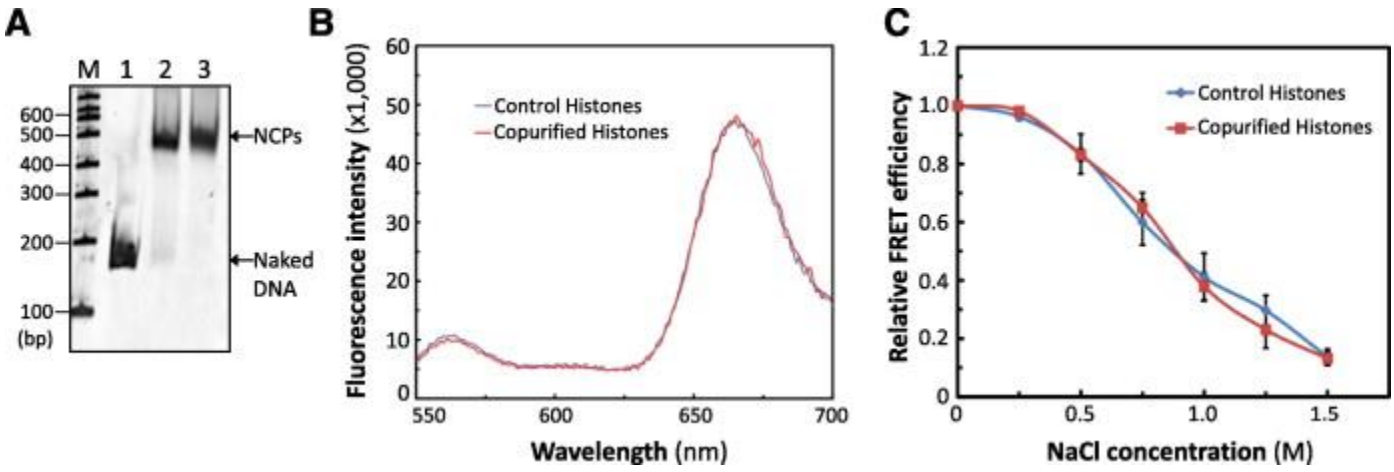


Figure 7: Copurified histones can form nucleosome core particles (NCPs).
(A) NCPs reconstituted using copurified histone octamers run as one band in a native gel analysis at a position equivalent to that of NCPs formed by the control histones. M, molecular weight marker; lane 1, DNA only; lane 2, NCP with control histones; lane 3, NCP with copurified histones. **(B)** The formation of NCPs with the copurified histones was confirmed by FRET. The emission spectra of NCPs (λ_{ex} = 515 nm) with the copurified histones show a robust FRET peak at 670 nm similarly to that of NCPs with control histones. Schematic diagrams of the donor (Cy3) and acceptor (Cy5) fluorophores on a naked DNA and after forming a NCP are shown in Figs. 6A and 6b, respectively. **(C)** Salt-induced NCP dissociation profile of the copurified histones is also similar to that of control histones

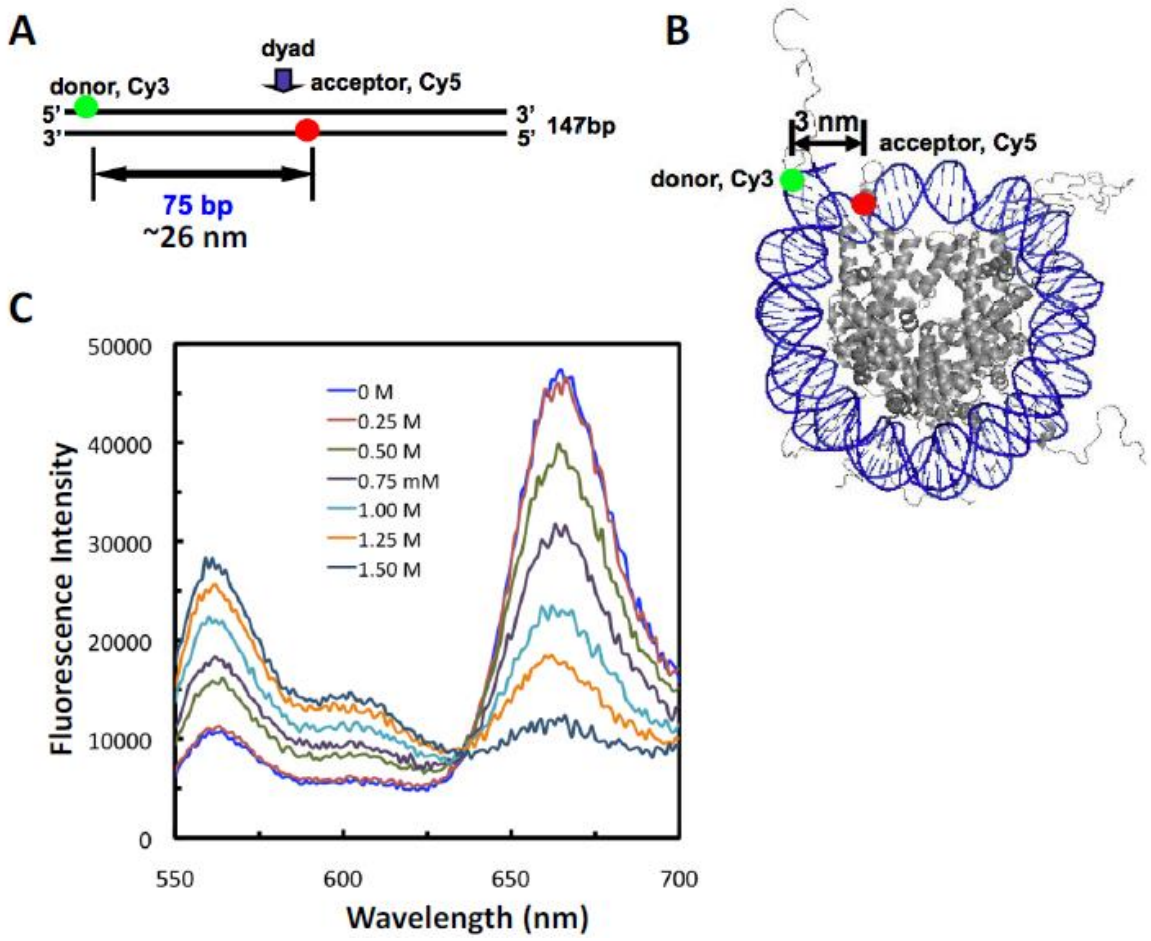


Figure 8: Characterization of the copurified histones using FRET.
(A and B) Schematic diagrams depicting the relative positions of the donor (Cy3) and acceptor (Cy5) fluorophores on a naked "601" DNA **(A)** and after the DNA forms NCP **(B)**. The formation of NCPs with the histones can be confirmed by the FRET peak of the acceptor at 670 nm as shown in Fig. 5B. **(C)** Fluorescence emission spectra of NCPs reconstituted with the copurified histones show that the NCPs dissociate as the salt concentration increases. This salt-induced dissociation profile is similar to that of NCPs with control histones (Fig. 5C)

2.4 Discussions

It has been known for many years that histones can form multimeric complexes containing H2A–H2B dimers and H3–H4 tetramers [50]. Polycistronic vectors have also been developed and used extensively to make protein complexes [7]. Recently, it has also been reported that H2A–H2B dimers and H3–H4 tetramers can each be coexpressed and purified under nondenaturing conditions [51]. Here, one step was taken further with these results and

developed a simple method to prepare histone octamers in the native state by constructing polycistronic histone coexpressing polycistronic vector. Reconstituted histone octamers from this polycistronic vector were confirmed to be identical with the histone octamers prepared from conventional method (*In vitro* reconstitution). Using this protocol, obtaining pure histone octamers to be readily used for NCP reconstitution can be accomplished in a single day, saving weeks of time required using the conventional methods (Fig. 9). It is anticipated that this new method will facilitate endeavors to make various unmodified and modified histone complexes and be used to study biological systems that work on chromatin in vitro.

This polycistronic vector was distributed over 70 laboratories in the world before the vector was deposited at Addgene. After depositing the vector, there have been more than 50 requests so far.

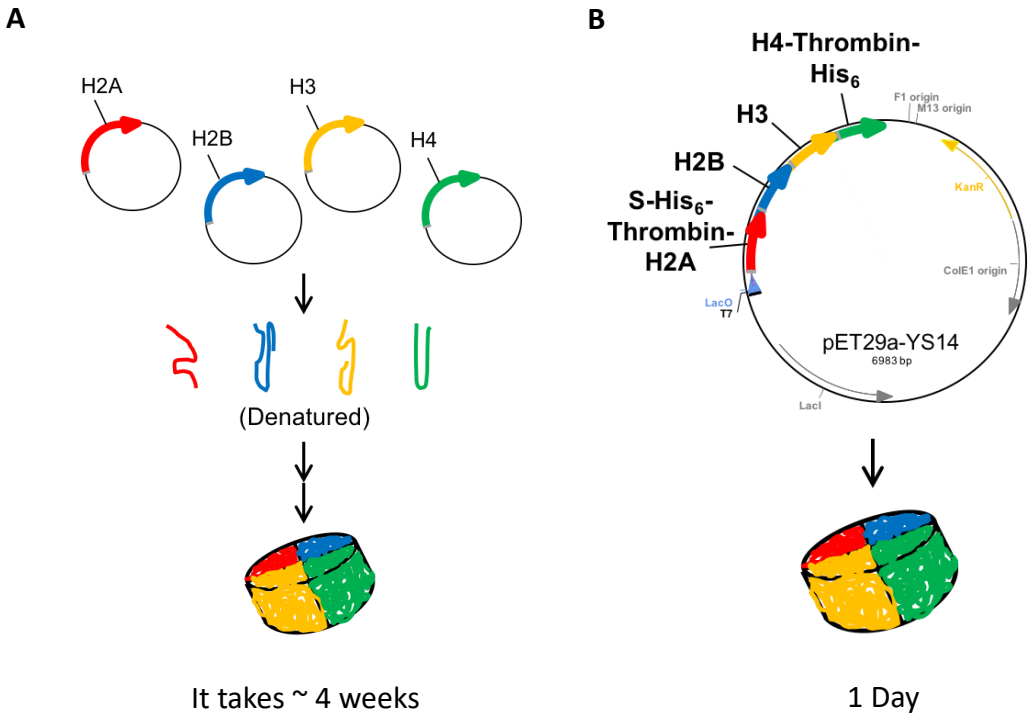


Figure 9: Overview of conventional method and new method for histone octamer preparation. **(A)** Histones are Individually expressed and purified under denaturing conditions. It is labor intensive and lengthy. **(B)** Histones are coexpressed from a single vector and purified under non-denaturing conditions as soluble octamers.

3 Improved method for Large scale, High purity NCP preparation

3.1 Introduction

NCPs were reconstituted by Dr. Duan to test the quality of histones, which were prepared from polycistronic vector as described in Chapter 2. NCPs are required to study NER pathway, which is one of DNA-related repair process, in compacted chromatin context. because our laboratory is focusing on studying NER, it is necessary to reconstitute NCPs with my hands.

To produce large amounts of NCPs, Widom 601 DNA sequences (hereafter "601 DNA") and α-satellite DNA, which are nucleosomal DNA, were prepared first. 147 base pair 601 DNA [47] was synthetically made with high affinity toward histone octamers to form nucleosomes. It has been known to assemble very stable and highly positioned nucleosomes [52] [53]. Dr. Smerdon's laboratory provided the pLMG601-23 plasmid with 23 tandem repeats of 601 DNA sequence [54]. AT rich α-satellite DNA is located at centromeric regions of humans with tandem repeat of about 170 base pairs [55]. From isolated α-satellite DNA containing chromatin [56], sequences of DNA was analyzed and used widely for preparing nucleosomes [57]. Both Widom 601 nucleosome and α-satellite DNA have been used to assemble nucleosomes [58] [30].

Next, Histone octamers were prepared as described in chapter 2. Before reconstituting large scale NCPs, small scale test is required to obtain best ratio of histone octamers to DNA. Otherwise, free excess histone octamers cause low product yield and free extra DNA leads non-native nucleosomal species [35]. As described in [45] [34], optimal molar ratio of histone octamers to nucleosomal DNA is known as around 1:1. However, ratio of histones to DNA forming NCPs with my hands was inconsistent. There may be possibility that purified histone octamers are contaminated by H2A-H2B dimers, $(H3\text{-}H4)_2$ tetramers or hexamers even after purification by Superdex 200 10/300 GL [59]. Elution volumes may not be resolved well with

Superdex 200 10/300 GL, thus, one more purification step of size exclusion chromatography providing higher resolution, HiLoad Superdex 16/60 200 pg (GE Healthcare), was added to puoduce more intact pure histone octamers [60].

3.2 Methods

3.2.1 Attempts on reconstituting NCP with my histone octamers

To prepare 601 DNA sequences, pLMG601-23 plasmid was transformed into XL1-blue cells and plated on to an LB agar plate containing ampicillin (100 µg/mL). The plate was incubated overnight (~16 hours) at 37 °C. The cells from LB agar plates were amplified in 2 L of LB media with ampicillin (100 µg/mL). The culture was shaken at 170 rpm at 37 °C for overnight (at least 12 hours) and purified by NucleoBond Xtra Midi kit (Machery-Nagel). 601 DNA was isolated by EcoRV-HF (20 units/µl, New England Biolabs) digestion of pLMG601-23 plasmid. 5 units of EcoRV-HF was added to 1µg of plasmid and reaction mixture was incubated at 37 °C for overnight. Digestion of plasmid was confirmed by running 1 % agarose gel for 8 min at 180 V. 147 base pair 601 DNA was extracted by Qiaquick Gel Extraction kit (Qiagen) and DNA was eluted with milli-Q H2O.

Histone octamers were prepared as described in Chapter 2 [36]. NCPs were reconstituted according to the protocol as Dr. Duan did as described in [35]. Small scale was set up with different molar ratio of histone octamers to 601 DNA [46] as 1:1 ~ 3.5:1 to determine the best ratio to reconstitute NCPs. DNA concentration was set to 0.7-8 µM and final sample volume was 40 µl. 601 DNA [47] was mixed to with copurified histone octamers in desired molar ratio in TE buffers (10 mM Tris-HCl pH 8.0 and 1 mM EDTA with 1 mM 1,4-Dithiothreitol) containing 2 M sodium chloride. Each sample with different ratio of histone octamers to DNA is placed in a cap of 0.6 ml siliconized microcentrifuge tube (Fisher Scientific) and covered with Spectrum

Spectra/Por 7 Membrane Tubing with 30 MWCO (Spectrum labs). The tube caps containing mixtures were dialyzed sequentially against TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA) containing 1 M and 600 mM sodium chloride, each for three hours. The last dialysis was done against TE buffer without sodium chloride for overnight (at least 12 hours). The results were verified on 5% native PAGE. Electrophoresis was run in 0.25x TBE at 130 V for 1 hour at 4 °C. The gel was incubated in 200 ml of water with 10 µ of ethidium bromide for 5 min at room temperature and visualized under the UV transilluminator.

3.2.2 Control NCP preparation with lyophilized histone powder (conventional method)

DNA oligomers for making α-satellite DNA were provided by Dr. Kurumizaka. The individual histone genes from mouse and lyophilized protein powders of H2A, H2B, H3.1 and H4 were also provided by his laboratory. I've learned how to make control NCPs in his laboratory in Japan.

A 145 bp α-satellite DNA was made with the help of Dr. Kobayashi in Kurumizaka laboratory. Four single stranded fragments were used to make α-satellite DNA (table 3). First, a 71 base pair single stranded DNA a (Top left) was annealed with 74 base pair DNA c (Bottom left), which is complementary to DNA a, and produced double stranded DNA with three-base overhang on the bottom strand. A 74 base pair single stranded DNA b (Top right) was annealed with 71 base pair DNA d (Bottom right), which is complementary to DNA b, and produced double stranded DNA with three-base overhang on top strand. The145 base pair α-satellite DNA was made by ligating two left and right annealed double stranded DNA fragments with overhangs (Fig. 10).

Table 3: DNA oligomers used for making α-satellite DNA

Name	**Sequence (5' → 3')**
a (Top left)	ATCAATATCCACCTGCAGATTCTACCAAAAGTGTATTTGGAAAC TGCTCCATCAAAAGGCATGTTCAGCTG
b (Top right)	TAGTTATAGGTGGACGTCTAAGATGGTTTTCACATAAACCTTTG ACGAGGTAGTTTTCCGTACAAGTCGACTTG-Phosphate
c (Bottom left)	TAGTTATAGGTGGACGTCTAAGATGGTTTTCACATAAACCTTTG ACGAGGTAGTTTTCCGTACAAGTCGACCAA-Phosphate
d (Bottom right)	ATCAATATCCACCTGCAGATTCTACCAAAAGTGTATTTGGAAAC TGCTCCATCAAAAGGCATGTTCAGCTG

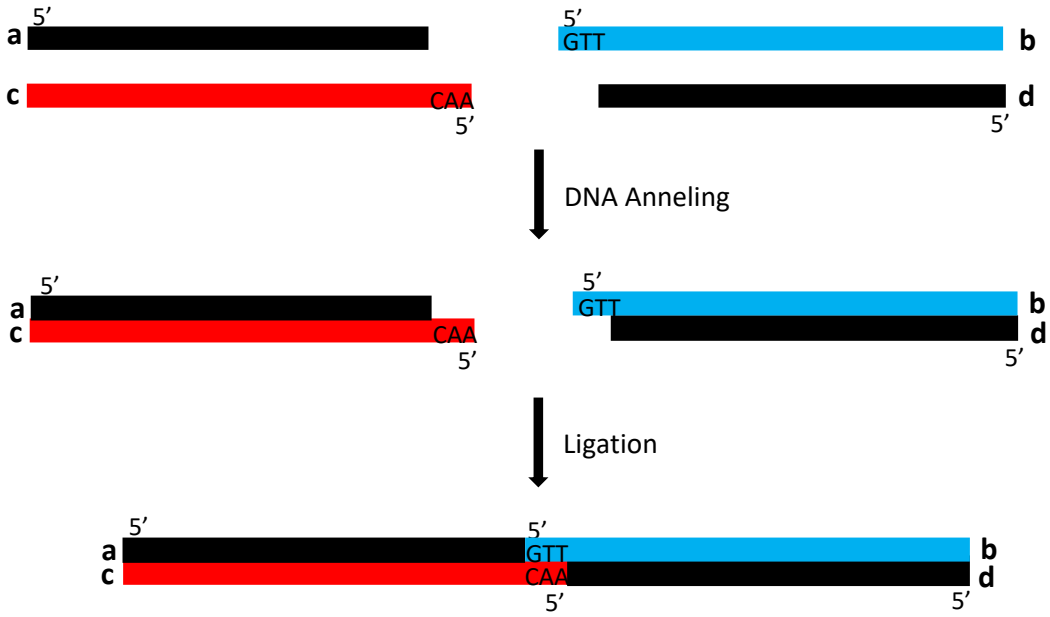


Figure 10: Schematic diagram showing how to prepare α-satellite DNA.
Single stranded DNA a was annealed with complementary DNA c and produced a 5'overhang (same for DNA b and DNA d). Two double stranded DNA were ligated to produce 145 base pair α-satellite DNA.

3.2.2.1 NCP reconstitution by continuous salt dialysis

Equimolar ratio of lyophilized H2A, H2B, H3.1 and H4 were mixed and unfolding buffer (20 mM Tris-HCl pH 7.5, 7.0 M Guanidinium chloride, 2 mM 2-Mercaptoethanol) was added to make 1.5 mg/ml proteins. The proteins were incubated for 1.5 hours while rocking at 4 °C. Histone proteins were dialyzed against 0.5 L of refolding buffer (10 mM Tris-HCl pH 7.5, 2.0 M sodium chloride, 2 mM EDTA, 5 mM 2-Mercaptoethanol) for four times (overnight, 4 hours, 4hours, overnight each) at 4 °C. The dialyzed proteins were injected onto a HiLoad Superdex 16/60 200 pg column pre-equilibrated in the degassed HiLoad 200 buffer (10 mM Tris-HCl pH 7.5, 2.0 M sodium chloride, 1 mM EDTA, 5 mM 2-Mercaptoethanol). Each fraction was analyzed by 18 % SDS-PAGE. Electrophoresis was run at 200 V for 65 min. The peak fractions were pooled and concentrated up to 8 mg/ml and named as Kurumi histone octamers.

Before preparing large scale NCP, small scale with different ratio of histone octamers to 601 DNA or α-satellite DNA was set up. Each mixture with different ratio was dialyzed against 0.4 L of RB high buffer (10 mM Tris–HCl pH 8.0, 2 M potassium chloride, 1 mM EDTA, 1 mM 1,4-Dithiothreitol) using dialysis tubing with 6-8 kD MWCO and 10 mm flat width (Spectrum Labs). The concentration of potassium chloride was gradually decreased from 2 M to 0.25 M by using a peristaltic pump (Cole- Parmer) with 0.8 ml/min flow rate. It took around 33 hours to pump 1.6 L RB low buffer (10 mM Tris–HCl pH 8.0, 0.25 M potassium chloride, 1 mM EDTA, 1 mM 1,4-Dithiothreitol) to 0.4 L of RB high buffer. Then, they were dialyzed against 0.4 L RB low buffer for another 3 hours followed by incubation at 55 °C for 2 hours (heat-shift). They were analyzed on 6 % native PAGE and electrophoresis was run in 0.25 x TBE at 150 V for 1 hour at 4 °C. The gel was incubated in 200 ml of water with 10 µ of ethidium

bromide for 5 min at room temperature and bands were visualized under the gel documentation system. The best ratio was selected to set up large scale NCPs.

3.2.2.2 NCP purification by Model 491 Prep Cell

Large scale NCPs were set up and the mixture was dialyzed followed by heat-shift as small scale NCP reconstitution. The results were verified on 6% native PAGE with 0.25 x TBE at 150 V for 1 hour at 4 °C. The gel was incubated in 200 ml of water with 10 µ of ethidium bromide for 5 min at room temperature and the bands were visualized under the gel documentation system. After presence of NCPs was confirmed, they were purified from free DNA and histone octamers by running Prep Cell (Bio-Rad). The lower chamber contained 3 L of running buffer (0.2x TBE) and upper chamber contained 0.5 L of elution buffer (20 mM Tris–HCl pH 7.5, 1 mM 1,4-Dithiothreitol) in Prep Cell apparatus. Sucrose solution was added to the NCPs to make NCPs with 5 % sucrose and the mixture was loaded on to 6 % native gel inside the Prep Cell apparatus. Electrophoresis was run at 300 V at 4 °C and NCPs were eluted at 1.5 mL/min. Each fraction was verified by 6% native PAGE. Electrophoresis was run in 0.25x TBE at 150 V for 1 hour at 4 °C. The gel was incubated in 200 ml of water with 10 µL of ethidium bromide for 5 min at room temperature and visualized under the gel documentation system. Peak fractions were pooled and concentrated up to 100 µL and stored at 4 °C (hereafter “control NCP”).

3.2.3 Improved NCP preparation with coexpressed histone octamers

Histone octamers were expressed and purified according to the protocol as described in Chapter 2 or in my publication [36]. Histone octamers were further purified by one more size exclusion chromatography (Hiload Superdex 16/60 200 pg column).

Superdex 200 eluates from 4L histone cultures were pooled and concentrated (usually ~ 2 mL with yield of 12-20 mg). Concentrated histones were injected onto a Hiload Superdex 16/60 200 pg column pre-equilibrated in the Hiload 200 degassed buffer (10 mM Tris-HCl pH 7.5, 2.0 M sodium chloride, 1 mM EDTA, 5 mM 2-Mercaptoethanol). Each fraction was analyzed by 18 % SDS-PAGE. Electrophoresis was run at 200 V for 65 min. The peak fractions were pooled and concentrated up to 8 mg/ml.

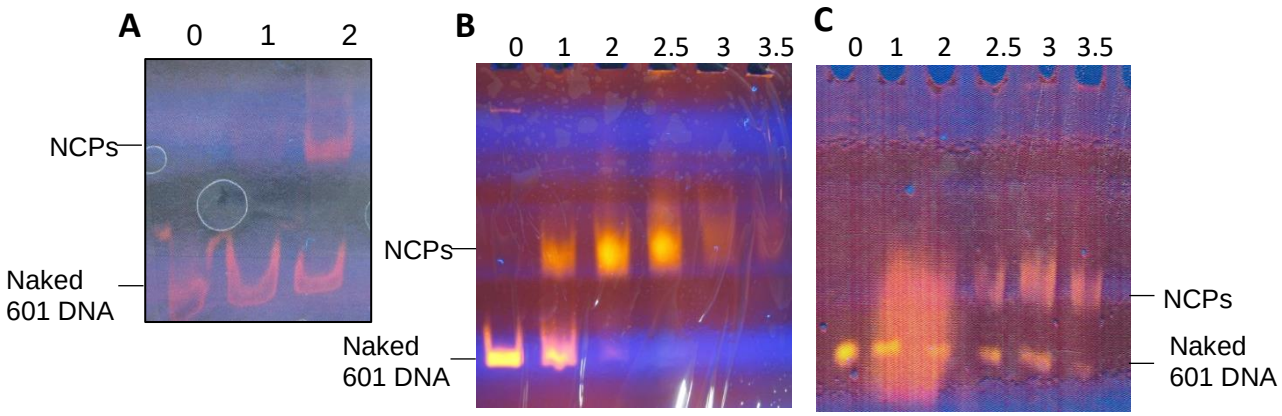
Before preparing large scale NCPs, small scale was set up with different ratio of histone octamers to 601 DNA [47] as 1:1, 2:1 and 3:1. DNA concentration was set to 0.1 µM and final mixture volume was 100 µL with 2M potassium chloride and 1 mM 1,4-Dithiothreitol. The mixtures were dialyzed and heat-shifted with conventional method as described in [38]. They were analyzed by 6 % native PAGE and electrophoresis was run in 0.25 x TBE at 150 V for 1 hour at 4 °C. The bands were visualized under the gel documentation system after treating with ethidium bromide. Large scale NCPs were set up at the best ratio with 500 µL as final volume and maximum concentration of DNA and histone octamers. The mixture was dialyzed, heat-shifted and purified as small-scale NCP reconstitution. The results were analyzed by 6% native PAGE. Electrophoresis was run in 0.25x TBE at 150 V for 1 hour at 4 °C. The gel was incubated in 200 mL of water with 10 µ of ethidium bromide for 5 min at room temperature and visualized under the gel documentation system. Peak fractions were pooled and concentrated up to 100 µl and stored at 4 °C.

3.3 Results

To prepare NCPs in our lab, NCP reconstitution was attempted by mixing 147 base pair 601 DNA containing strong nucleosome-positioning sequence and, coexpressed and copurified histone octamers prepared as described in [36]. Histone octamers were purified by HisTrap FF

5ml and Superdex 200 column. Histone octamers and 601 DNA were mixed with different ratio and NCPs were reconstituted by stepwise dialysis as Dr. Duan did in [46] and they were analyzed by 6 % native PAGE.

It has been known that equimolar ratio of histones to DNA has provided best condition for nucleosome reconstitution [35]. However, ratio of histones to DNA was off and inconsistent even though same DNA and histone octamers were used to reconstitute NCPs (Fig. 11). Even with 2:1 ratio of histone octamers to 601 DNA showed excess DNA band and no NCP reconstitution was observed with 1:1 ratio (Fig. 11A). In addition, the best ratio of histone octamers to DNA was 2.5:1 in Figure 10B, but same ratio (2.5:1) still showed excess DNA band in Figure 10C. It may be due to histone octamers contaminated by H2A-H2B dimers, $(H3\text{-}H4)_2$ tetramers or hexamers even after purification with Superdex 200 10/300GL size exclusion chromatography. I've tried many times reconstituting NCPs with newly prepared DNA and histone octamers to troubleshoot the problem, but it didn't work out.



	Histone octamers : 601 DNA			Histone octamers : 601 DNA
0	0 : 1			
1	1 : 1			
2	2 : 1	2.5	2.5 : 1	
3	3 : 1	3.5	3.5 : 1	

Figure 11: small scale NCP reconstitution with copurified histone octamers and 601 DNA at different ratio.
NCPs are reconstituted by conventional method [35]. NCPs were analyzed on UV transilluminator with EtBr stained 6 % native PAGE **(A)** NCP reconstitution with 1:1 and 2:1 ratio of histone octamers to 601 DNA. Even 2:1 ratio of histone octamers to 601 DNA showed excess free DNA. **(B,C)** 1:1 ~ 3.5:1 ratio of histone octamers to 601 DNA showing inconsistent and different ratio of histone octamers to 601 DNA in NCP formation.

Kurumizaka laboratoty has reconstituted nucleosomes for long time and worked on dynamic regulation and function of chromatin structure. Thus, Dr. Kurumizaka kindly allowed me to learn their NCP reconstitution method in Japan upon my advisor's request.

First, α-satellite DNA was prepared by annealing followed by ligation of four DNA oligomers (a, b, c and d). DNA a was annealed with DNA c and DNA b with DNA d producing a+c and b+d double stranded DNA with 3 base pair overhang. Both annealed DNA a+c and b+d had single bands on 10 % native PAGE proving that DNA was annealed completely (Fig. 12A).

a+c and b+d double stranded DNA were ligated and α-satellite DNA was produced. The final ligated products were analyzed on 6 % native PAGE and two bands were observed on 6 % native PAGE showing that DNA was partially ligated. (Fig. 11B).

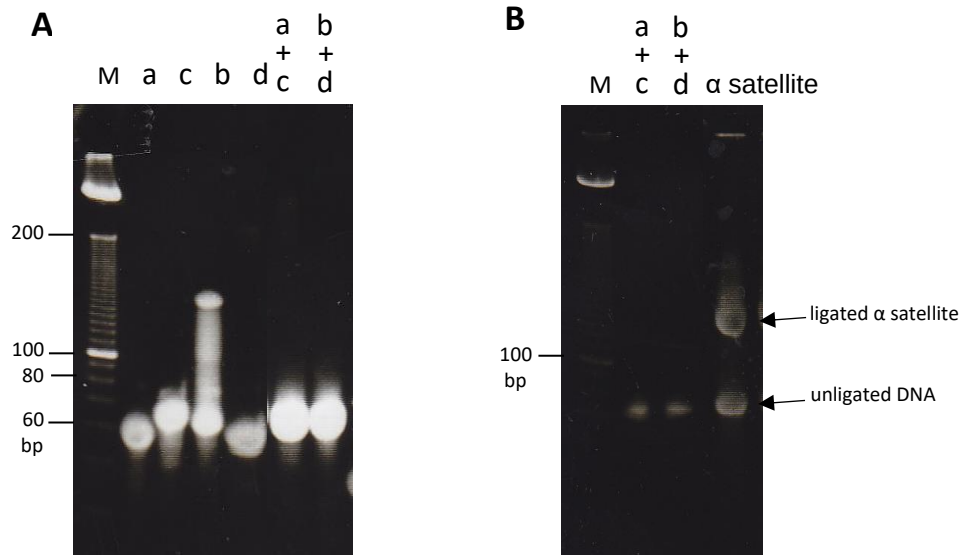


Figure 12: α-satellite DNA preparation was analyzed on native PAGE.
(A) DNA annealing for DNA a with DNA c and DNA b with DNA d. M, 10 bp DNA ladder; a ~ d, DNA a ~ DNA d; a+c, double stranded DNA with 5' overhang after annealing DNA a with DNA b; b+d, double stranded DNA with 5' overhang after annealing DNA b with DNA d. **(B)** Ligation of DNA a+d and DNA b+d. α-satellite DNA, ligation product.

To prepare equimolar H2A, H2B, H3 and H4, 2.52 mg of H2A, 2.48 mg of H2B, 2.70 mg of H3.1 and 2.00 mg of H4 lyophilized powders were mixed with 6.47 mL of unfolding buffer to make 1.5 mg/mL histones. Histones were denatured in unfolding buffer and denatured histones were purified by HiLoad Superdex 200 column. Histones eluted at 67 mL and 82.4 mL elution volume (Fig 13). According to the previously published works, histone elution profile indicates that histone octamers elute at 65 to 68 mL with $(H3\text{-}H4)_2$ tetramers at 72 mL and H2A-H2B dimers at 84 mL [35]. As expected, it was confirmed on a 18 % SDS-PAGE that histone octamers eluted at 67 mL and H2A-H2B dimers eluted at 82.4 mL (Fig. 13). Peak fractions analyzed on a SDS-PAGE showed stoichiometric ratios of the histone octamers across the

histone peak at 67 mL. Histone octamer peak fractions were pooled and concentrated up to 8 mg/mL.

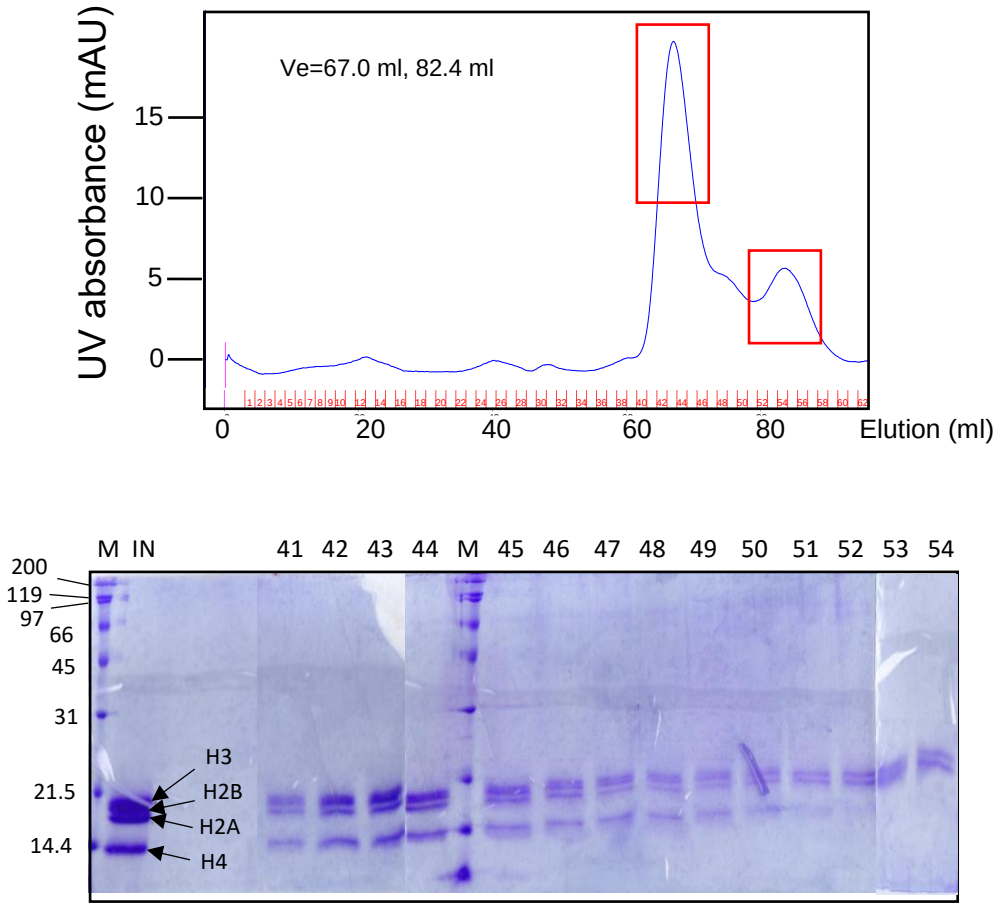


Figure 13: Kurumi histone octamers preparation with HiLoad Superdex 16/60 200 pg. Histone octamers were reconstituted by individually lyophilized histones.
H2A, H2B, H3 and H4 were mixed and denatured followed by reconstitution over a HiLoad Superdex 200 column. The chromatogram (top) and SDS–PAGE (bottom) show that stoichiometric histone complexes elute at 67.0 ml followed by H2A-H2B dimer elution at 82.4 ml. UV, ultraviolet; M, molecular weight marker; I, input; lanes 41 to 52 indicate fraction numbers.

Kurumi histone octamers and α-satellite DNA were mixed with 1:1 to 2:1 ratio and the best condition was observed with 1.5:1 molar ratio of Kurumi histone octamers to α-satellite DNA. Large scale NCPs were reconstituted by continuous dialysis and the reconstituted NCPs

were incubated at 55 °C for 2 hours (heat-shift) so that DNA on nucleosomes move freely to find most favorable positon [42] and they were analyzed by 6 % native PAGE (Fig. 14A). Then, NCPs were purified by Model 491 Prep Cell and control NCP was prepared (Fig. 14B).

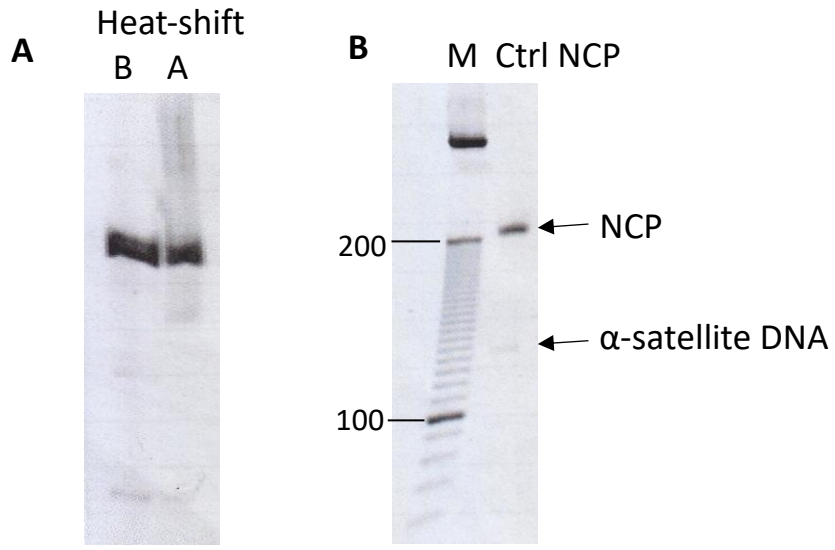


Figure 14: Control nucleosomes prepared with α-satellite DNA and individual lyophilized histone powder.
(A) NCPs were incubated at 55 °C for 2 hours (heat-shift). B, NCPs before heat-shift; A, NCPs after heat-shift. **(B)** NCPs were purified by Model 491 Prep Cell after heat-shift. Ctrl NCP, final concentrated NCPs. Arrows indicate the position of NCP and α-satellite DNA bands.

Many other laboratories including Kurumizaka lab have used HiLoad Superdex 16/60 200 pg column to purify histone octamers because it gives high resolution separation due to steep selectivity of the dextran component and highly cross-linked agarose [35] [58]. My histone octamers prepared from polycistronic vector were subjected to one more purification step with HiLoad Superdex 200. The elution profile was compared with the one of Kurumi histone octamers prepared by individually lyophilized histone powders (Fig. 15). Both histone octamers eluted around 67 ml while maintaining the stoichiometric ratio of each histone shown on a 18 % SDS-PAGE.

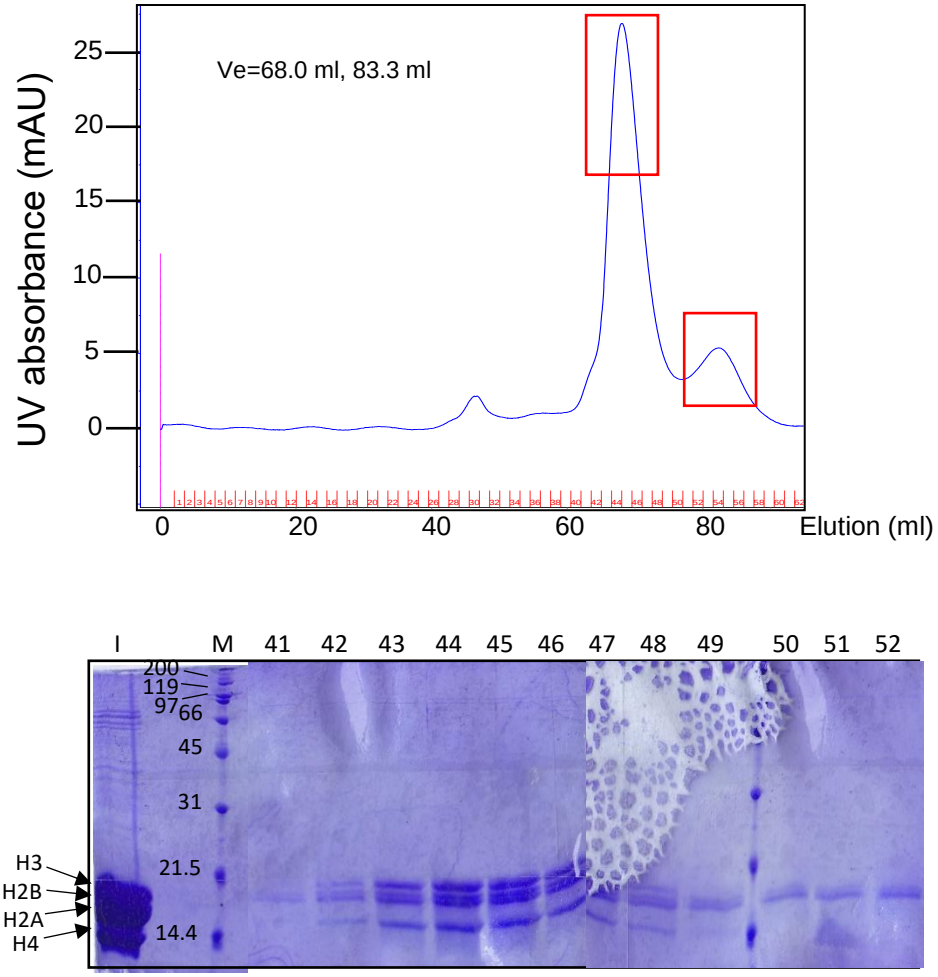


Figure 15: Histone octamer preparation with one more purification step (HiLoad Superdex 200). Histones were coexpressed by polycistronic vector and copurified by HisTrap FF 5 ml and Superdex 200 followed by HiLoad Superdex 200. The chromatogram (top) and SDS–PAGE (bottom) show that stoichiometric histone complexes elute at 68.0 ml followed by H2A-H2B dimer elution at 83.3 ml. UV, ultraviolet; M, molecular weight marker; I, input; lanes 41 to 52 indicate fraction numbers.

pLMG601-23 plasmid from 4 L culture was extracted by midiprep kit and the concentration was adjusted to 1µg/ µL with TE buffer. The plasmid was digestied by EcoRV-HF restriction enzyme to isolate 601 DNA. Before purifying 601 DNA with gel extraction kit, small amount (~ 50-100 µg) was analyzed on a 1 % agarose gel to check complete digestion (Fig.

16A). Because digestion was not complete, reaction mixture was incubated another 6 more hours. Complete digestion was confirmed and 601 DNA was purified. Gel purified 601 DNA products were confirmed on a 1 % agarose gel (Fig. 16B).

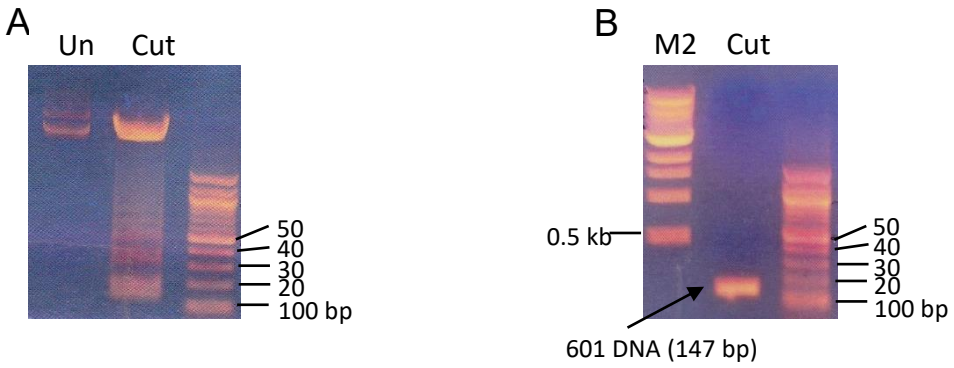


Figure 16: pLMG601-23 digestion by EcoRV-HF. Digestion was analyzed on a 1% agarose gel. **(A)** Many different sizes of DNA fragments were shown which means partial digestion of pLMG601-23 plasmid in 6 hours. **(B)** Purified 601 DNA after overnight digestion of pLMG601-23 plasmid. M1, I00 bp DNA ladder; M2, I kb DNA ladder; Un, uncut circular pLMG601-23 plasmid; Cut, digested pLMG601-23 plasmid with EcoRV-HF restriction enzyme.

Histone octamers prepared from polycistronic vector with one more purification step (HiLoad Superdex 200) and 601 DNA were mixed with 1:1 to 3:1 ratio and the best condition was observed with 1:1 molar ratio of histone octamers to 601 DNA. NCPs were reconstituted by continuous dialysis and the reconstituted products were incubated at 55 °C for 2 hours (heat-shift) so that DNA on nucleosomes move freely to find most favorable position [42] and they were analyzed by 6 % native PAGE. From small scale NCP reconstitution, 1:1 ration of histone octamers to 601 DNA showed the best NCP reconstitution (Fig. 17).

DNA	Histone octamers		601
B0, A0	0	:	1
B1, A1	1	:	1
B2, A2	2	:	1

Figure 17: Small scale NCP reconstitution with different histone octamers:601 DNA ratio. The results were analyzed on a 6 % nave-PAGE. B0~B3, NCP with different ratio before heat-shift; A0-A3, NCP with different ratio after heat-shift at 55°C for 20 min.

Therefore, 1:1 ratio of histone octamers to 601 DNA was selected for large scale NCP reconstitution. NCPs were reconstituted by continuous dialysis and heat-shift followed by purification with Model 491 Prep Cell and final NCPs were analyzed by 6 % native PAGE. (Fig 18).

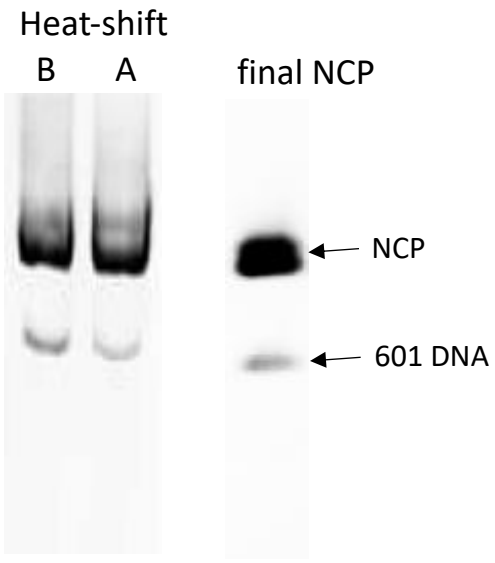


Figure 18: large scale NCP reconstitution analyzed on a 6 % nave-PAGE.
(A) NCPs with 601 DNA and co-purified histone octamers. M, 10 bp DNA ladder; B, before heat-shift; A, after heat-shift. Final NCP, final concentrated product after purification by Prep Cell. **(B)** Control nucleosomes prepared with α-satellite DNA and individual lyophilized histone powder. Ctrl NCP, final concentrated control product after product purification by Prep Cell

3.4 Discussions

In order to understand DNA-related metabolic processes in the context of chromatin, experimental studies should be performed on nucleosome-DNA, not on naked DNA. The first step is to prepare intact pure nucleosomes. Preparation of pure and high-quality histone octamers is very critical step and it can also determine the quality of nucleosomes.

There are several ways to prepare pure histone octameric complex. In this dissertation, I've introduced two methods. First, histones are coexpressed and copurified together as described in Chapter 2 and NCPs were reconstituted by decreasing salt concentration through stepwise dialysis. The ratio of histone octamers to DNA for NCP reconstitution was inconsistent, thus, we couldn't make NCP productively. Second, individual histones are expressed and purified separately and with equimolar ratio of each lyophilized histones are mixed, denatured and reconstituted on a HiLoad Superdex 16/60 200 pg column. Histone octamers are mixed with

nucleosomal DNA in nearly equimolar ratio and NCPs were reconstituted by decreasing salt concentration through continuous dialysis (conventional method).

To separate histone dimers, tetramers or hexamers from octamers, coexpressed histone octamers were further purified HiLoad Superdex 16/60 200 pg, which gives high resolution than superdex 200. Histone octamers shared same elution profile with histone octamers from lyophilized histone powder on size exclusion chromatography and easily separated from H2A-H2B dimers on a column. In addition, continuous dialysis through peristaltic pump was performed instead of stepwise dialysis to dialyze more consistently. Finally, NCPs were reconstituted in nearly equimolar ratio.

Even if both histone octamers prepared by individually lyophilized histones or coexpressed polycistronic vectors, they shared same elution profile on size exclusion chromatography. More validation is required to verify if coexpressed and copurified histones can replace lyophilized histone powder for NCP reconstitution.

When NCPs were purified by Model 491 Prep Cell, excess free DNA still remained in final product. More elaborate protocol is required to prepare pure intact NCPs. Then, NCPs can be used to study how NER initiation factor Rad4-Rad23 can recognize DNA damage in the highly compated

4 Reconstitution of General Transcriptional Factor II H complex using Baculovirus expression system

4.1 Introduction

4.1.1 Multiprotein reconstitution with co-expression in insect cells

Large proteins, membrane proteins and the proteins required post translational modification cannot be produced from *E.coli* [61] [62]. In addition, proteins are often insoluble and/or inactive due to lack of post translational modification [63] or molecular chaperones [64] even if proteins are expressed. Higher eukaryotic system can be used to overcome these limitations [61].

Insect cells as higher eukaryotic system are able to go through complex post translational modifications similar to mammalian cells and enable correct disulfide formation for proper protein folding, unlike the *E.coli* expression system [65]. In addition, ease of large-scale protein production with high level of protein expression allows this system highly versatile for recombinant protein expression [66]. MultiBac, baculorovirus infected insect system, is generating recombinant baculovirus DNA to express eukaryotic protein [67].

Each gene of interest is inserted into baculoviral expression vector and transformed into DH10MultiBac cells containing baculovirus shuttle vector called bacmid, which is used for transfecting insect cells [67]. After monocistronic baculorovirus is generated, viruses are amplified in Sf9 insect cells, a clonal isolate Spodoptera frugiperda Sf21 cells. Multiplicity of infections (MOI)] of each virus, which is the ratio between the number of infectious viruses and the number of viable cells, should be examined to achieve optimal expression. Otherwise, it can vary protein production yield on each proteins resulting in low multiprotein complex yield [68].

The technique on how to estimate virus titer properly is very important. However, virus titration is tedious and tricky. In addition, making bacmid and viruses from each baculoviral

expression vector is time-consuming process. Therefore, co-expressing multiple proteins with MultiBac system has been developed.

Baculovirus transfer vectors are classified as acceptors and donors and they contain two promotors (polh, p10), a terminator, multiple cloning sites (MCS) for each promotor and multiplication module. Multiple genes are assembled into the vector by using multiplication module and/or *In vitro* Cre-loxP fusion. Multigene expression cassettes are integrated into MultiBac bacmid to transfect Sf9 cells. Then, viruses are produced and amplified. Multiprotein complex is produced by infecting cells with MultiBac virus [69].

Two genes can be inserted into a transfer vector because it has two promotors. Then, more genes can be inserted into the same vector by using multiplication module placed in between two translation cassettes with particular restriction enzymes sites (Bstz17I/SpeI, PmeI/AvrII). Or, multiple genes can be assembled by using *In vitro* Cre-loxP fusion with Cre recombinase. Donor transfer vectors with genes can be integrated into a acceptor transfer vector by fusion.

Multigene transfer vectors are integrated into MultiBac baculrovirus shuttle vector (Bacmid) by Cre-loxP fusion or Tn7 transposition for generating recombinant baculoviruses. Monolayer Sf9 insect cells, the lepidopteran cell lines from Spodoptera frugiperda (Sf9), are transfected by Bacmid to produce virus produce and viruses are amplified by Sf9 cells. For large scale protein expression, suspension Hi5 cells, which are derived from Trichoplusia ni, are used. Proteins are produced by infecting Hi5 cells with viruses produced from Sf9 cells [70] [71].

In this dissertation, seven core subunits of general transcription factor IIH (TFIIH) were reconstituted into a multigene transfer vector successfully, but only five proteins were recovered as soluble.

4.2 General transcription factor II H

4.2.1 What is general transcriptional factor II H?

Protein synthesis is very important to our lives in overall because most essential functions of cells are regulated by proteins. Transcription is the first step in protein synthesis, which converts genetic information from DNA to RNA and this process is carried out by an important enzyme called RNA polymerase [9].

General transcription factor II H (hereafter TFIIH), which is a huge multiprotein complex, forms preinitiation complex with other general factors (TFIIA, TFIIB, TFIID, TFIIE, and TFIIF) and recruits RNA Pol II (RNAPII) to the promoters of genes to initiate transcription in eukaryotes [72]. This factor is also known to be involved in RNA polymerase I transcription [73]. In addition, TFIIH has a role in DNA repair suggesting a link between transcription and DNA repair [74] [75].

4.2.2 Composition of TFIIH

TFIIH is composed of 10 subunits with a total mass of about 500 kDa in human [76]. All ten subunit TFIIH is called holo complex and this complex is required for transcription [77]. Holo TFIIH can be resolved as two sub-complexes, a 7 subunit core complex and a 3 subunit CAK (cyclin-dependent kinase (CDK)-activating kinase) complex [78]. The core-TFIIH is composed of seven subunits, XPB/Ssl2, XPD/Rad3, p62/Tfb1, p52/Tfb2, p44/Ssl1, p34/Tfb4, and p8/TTD-A/Tfb5 (names are for the orthologs in human/yeast, respectively), and this complex is required for nucleotide excision repair (NER) [79]. CAK complex contains CDK7/Kin28, cyclin H/Ccl1, and MAT1/Tfb3 [80]. CAK complex is responsible for activating several kinases by phosphorylation [81] [82] and also regulates cell cycle [83]. Sequence analysis shows high sequence conservation of core-TFIIH among various eukaryotes [84].

Table 4: Composition of TFIIH subunits in human and yeast

Sub complex	Human	Predicted protein size in human (kDa)	Yeast	Percent identity between human and yeast (%)
Core	XPB	89.3	Ssl2	49
	XPD	80	Rad3	51
	p62	62	Tfb1	21
	p52	52	Tfb2	34
	p44	44	Ssl1	34
	p34	34	Tfb4	29
	p8	8	Tfb5	28
CAK	MAT1	32	Tfb3	31
	CyclinH	38	Ccl1	23
	CDK7	41	Kin28	41

4.2.3 TFIIH in transcription

There are three steps in transcription cycle: initiation, elongation and termination. In the first step of initiation step, general transcription factors, TFIIA, TFIID, TFIIE, TFIIF and TFIIH [85] form pre-initiation complex (PIC) with RNA polymerase II [86]. Promoter DNA is recognized by TFIID and TFIIB and PIC was load on promoter DNA with aid of TFIIH. Then, TFIIE links TFIIH to RNA polymerase II [87] so that double stranded promoter DNA is unwound by TFIIH with ATP-dependent helicase activity [88].

Holo-TFIIH complex includes two subunit XPB/Ssl2 and XPD/Rad3 possessing ATP-dependent helicase activity [73], but only XPB/Ssl2 ATPase activity is required for DNA unwinding around upstream of the transcription start site in 3’ to 5’ direction and make transcription bubble allowing RNA synthesis to initiate [89] [90]. In addition, TFIIH kinase subunit CDK7 phosphorylates C-terminal domain (CTD) of RNA polymerase II [91]. Formation of transcription bubble and RNA polymerase II phosphorylation by holo-TFIIH enable transcription initiation followed by elongation and termination [92].

4.2.4 TFIIH in Nucleotide Excision Repair

CAK complex dissociates with the assist of XPA from the holo TFIIH complex and only core TFIIH complex is required during nucleotide excision repair [93]. It is known that helicase subunit XPB and XPD function in unwinding the damaged DNA [94]. ATPase activity of XPB contributes to anchoring TFIIH to the DNA around the lesion and separates DNA strands to make XPD configuration favors binding to DNA [95] [96]. From structural study, it was suggested that RED and thumb domain seems to be important with a role in anchoring to the DNA [96]. On the other hand, both helicase and ATPase activity of XPD is required to confirm DNA lesion, move along one DNA strand in the 3' to 5' direction and open 24 to 32 nucleotides around the lesion [95] [97]. In addition, p52 is shown to interact and promote the activity of XPB whereas p44 interact and promote the activity of XPD [98]. The involvement of XPD with the lesion facilitates the association of the preincision complex for successful repair [99]

4.2.5 TFIIH and disease

The human diseases xeroderma pigmentosum (XP), Cockayne syndrome (CS), trichothiodystrophy (TTD) and combined symtoms of XP and Cockayne syndrome (XP/CS) are caused by defects in nucleotide excision repair (NER) pathway [100]. XP have 1000-fold increase in skin and eye cancer and it also leads to progressive neurological degeneration in about 30 % of XP patients in addition to skin defects. CS and TTD patients are sensitive to sunlight, but have normal skin cancer frequencies. In addition, some of CS and TTD patients are extremely short and immature in sexual development. CS patients are often born normally, but postnatal failure of brain development occurs later. TTD patients have sulfur deficient brittle hair. In severe case, they have neurological and somatic defects in development [101]. Disease-related mutations are identified in XPB, XPD and p8 [102].

Helicase activity of XPD is stimulated by the interaction between XPD and p44 [103]. Mutations in XPD residues near interacting surface with p44 causes xeroderma pigmentosum (XP) or trichothiodystrophy (TTD). XPD mutations near DNA- or ATP binding sites of RecA-like domains cause xeroderma pigmentosum combinded with Cockayne syndrome (CS/XP) [104].

ATPase activity of XPB is promoted by the interaction between XPB and p52 [98] and p8 is also known to regulate the activity of XPB by interacting with p52 [105]. Mutations in XPB responsible for this interaction causes xeroderma pigmentosum combinded with Cockayne syndrome (CS/XP) and trichothiodystrophy (TTD) [106] . Moreover, mutations in p8 residue required for this interaction causes trichothiodystrophy (TTD) [107].

4.3 Methods

4.3.1 Cloning

4.3.1.1 Multigene expression cassettes with multiplication module

TFIIH genes from yeast were amplified by PCR from the plasmids former postdoc Dr. Beomseok Park made. Before inserting genes into baculoviral pFL and pUCDM vector, an in frame hexahistidine tag was inserted into pH promoter of pFL vector and named as pFLHT A.

Tfb2 gene was inserted into p10 promoter of pFL between NheI and SphI sites and named as YS18. Ssl1 gene was inserted into pH promoter of YS18 between EcoRI and HindIII sites and named as YS26. Rad3 gene was inserted into pH promoter of pFLHT A vector between NotI and BssHII sites and named as YS54. Tfb1gene was inserted into p10 promoter of YS54 between NheI and SphI sites and named as YS56.

Multi-genes in YS26 (Ssl1 and Tfb2 in pFL) and YS56 (Tfb1 and Rad3 in pFLHT A) were assembled by using multiplication module with Bstz17I, SpeI, PmeI and AvrII restriction

enzymes (Table 5). The entire multigene expression cassette of YS56 was removed by PmeI and AvrII digestion and inserted into multiplication module of YS26 digested with Bstz17I and SpeI. The fragment of expression cassette of YS56 and multiplication module digested YS26 were purified by using Gel extraction kit (Qiagen). Ligation was done because AvrII produced sticky end compatible with SpeI, and PmeI and Bstz17I produced blunt end. This ligation product was named as YS60.

In the same way, Tfb4 gene was inserted into p10 promoter of pFLHT A between NheI and SphI sites and named as YS55. Tfb5 gene with hexahistidine tag was inserted into p10 promoter of another pFLHT A between NheI and SphI RE sites and named as YS48. Ssl2 gene was inserted into pH promoter of YS48 between EcoRI and HindIII sites and named as YS49.

Genes in YS55 (Tfb4 in pFLHT A) and YS49 (his-Tfb5 and Ssl2 in pFLHT A) were assembled by using multiplication module with Bstz17I, SpeI, PmeI and AvrII restriction enzymes. The entire multigene expression cassette of YS49 was removed by PmeI and AvrII digestion and inserted into multiplication module of YS55 digested with Bstz17I and SpeI. The fragment of expression cassette of YS49 and multiplication module digested YS55 were purified by using Gel extraction kit (Qiagen) followed by ligation. This ligation product was named as YS61.

Both YS61 (Tfb4, his-Tfb5 and Ssl2 in pFLHT A) and pUCDMb were digested with AvrII and PmeI. The entire multigene expression cassette of YS61 was inserted into multiplication module of pUCDMb by ligation. DNA fragments containing three genes from YS61 were purified with QIAquick Gel extraction kit (from Qiagen) and the multiplication module digested pUCDMb was purified with same manner. These two fragments were ligated and the plasmid with three genes in pUCDMb was made. It was named as YS62.

Table 5: Restriction enzymes used in this study to assemble expression cassettes by using multiplication module with restriction enzymes.

Restriction enzyme	Sequences (5' → 3')	After digestion	After ligation
Bstz17I	GTA TAC CAT ATG	GTA CAT	GTAAAAC CATTTTG
PmeI	GTTT AAAC CAAA TTTG	AAAC TTTG	
AvrII	C CTAGG GGATC C	C GGATC	CCTAGT GGATCA
SpeI	A CTAGT TGATC A	CTAGT A	

4.3.1.2 Multigene expression cassettes with Cre-loxP recombination

Genes in YS60 (Ssl1, Tfb2, Tfb and Rad3 in pFL) and YS62 (Tfb4, his-Tfb5 and Ssl2 in pUCDMb) were assembled by using *in vitro* fusion of acceptor (pFL) and donor plasmid derivatives (pUCDMb) with Cre recombinase. 3 µg of YS60 and YS62 with 2 U(unit) of Cre recombinase and 2 µL of Cre recombinase reaction buffer were mixed. Milli-Q H2O was added to make the final volume to 20 µL. The reaction mixture was incubated at 37 °C for 30 min followed by 10 min incubation at 70 °C for 10 min. 10 µL of samples were transformed into XL-Blue competent cells and cells were plated onto a LB agar with chloramphenicol (27 µg/mL) and ampicillin (100 µg/mL). Several colonies were selected and 3 mL cultures were grown in LB with chloramphenicol (27 µg/mL) and ampicillin (100 µg/mL) followed by DNA extraction by using miniprep kit. Miniprep DNA was digested with AvrII and PmeI to confirm the existence of YS60 and YS62.

Table 6: Vector, promoter and restriction enzyme sites used in this study.

Core subunit	Vector used	Promoter	RE sites	Construct name
Tfb2	pFL	P10	NheI/SphI	YS18
Ssl1; Tfb2	YS18	PH	EcoRI/HindIII	YS26
Rad3	pFLHT A	PH	NotI/BssHII	YS54
Tfb1; Rad3	YS54	P10	NheI/SphI	YS56
Tfb1, Rad3; Ssl1, Tfb2	YS26	-	Bstz17I/SpeI PmeI/AvrII	**YS60**
p34	pFLHT A	P10	NheI/SphI	YS55
His-p8	pFLHT A	P10	NheI/SphI	YS48
XPB; His-p8	YS48	PH	EcoRI/HindIII	YS49
XPB, His-p8; p34	YS55	-	Bstz17I/SpeI PmeI/AvrII	YS61
XPB, His-p8; p34	pUCDMb	-	PmeI/AvrII	**YS62**
Cre-loxp recombination of **YS60** x **YS62** by *Cre* recombinase → YS63 with all seven subunits				

- XL1-blue competent cells were used for pFLHT A transformation and PirPlus competent cells for pUCDMb transformation.

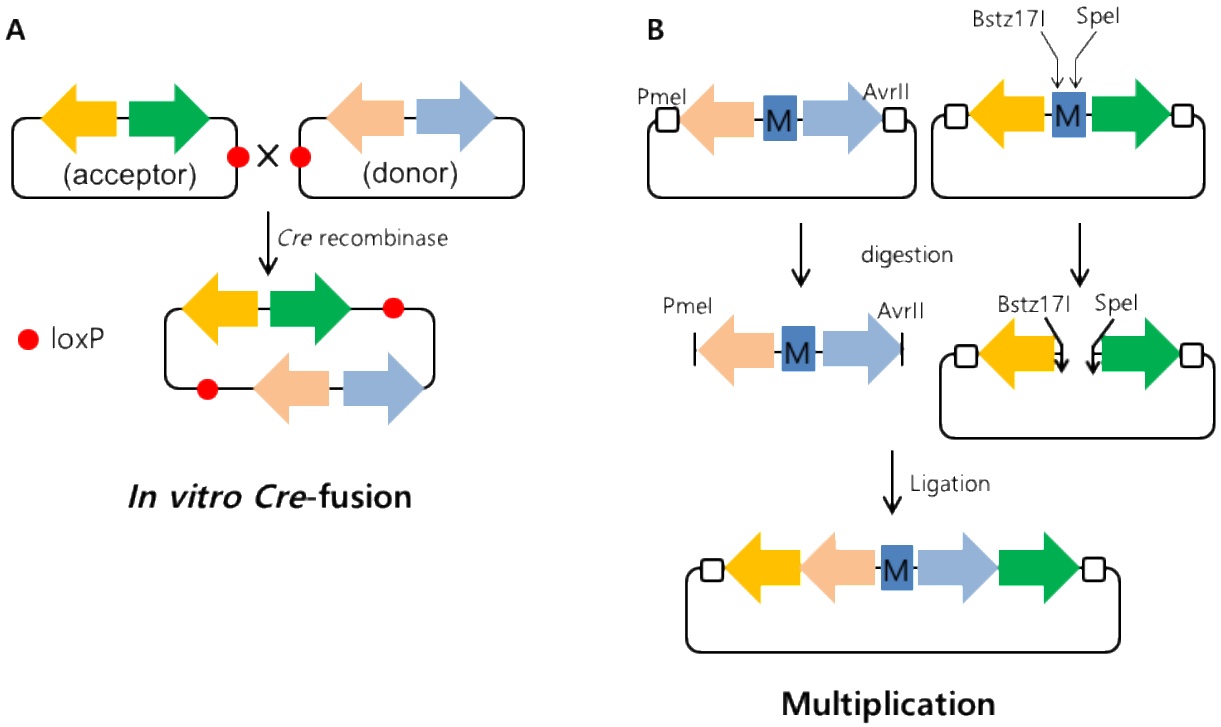


Figure 19: Generating multigene expression cassettes.
(A) Multigene assembly by *in vitro* fusion of donor and acceptor plasmid with Cre recombinase. pFL as acceptor and pUCDM as donor are used in this study. (**B**) Multigene assembly by using the multiplication module.

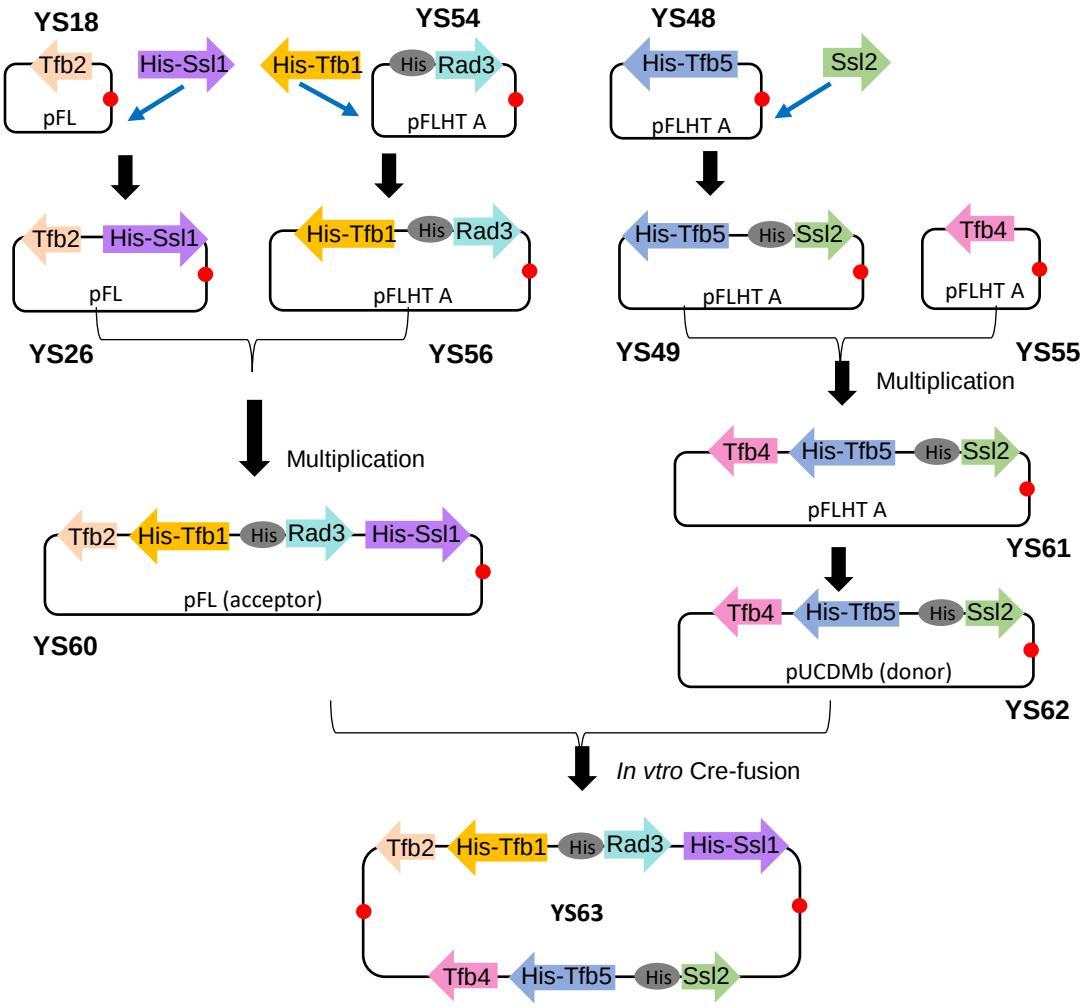


Figure 20: Creating Core TFIIH complex multigene expression cassettes.
Tfb2, his-Ssl1, Tfb1, and his-Rad3 genes were incorporated into pFL transfer vector and his-Tfb5, Ssl2 and Tfb4 were incorporated into pUCDMb transfer vector by using multiplication module with particular restriction enzymes. Two transfer vectors were fused into one vector by using *in vitro* Cre-fusion by using Cre recombinase.

4.3.2 Transposition and Transfection

TFIIH core subunit multigene expression cassettes, YS63 plasmid, was transformed into DH10Bac *E.coli* strain to generate recombinant bacmid DNA. Bacmid DNA was transfected into Sf9 cells with cellfectin II reagent to generate recombinant baculovirus called P0. P0 was amplified to generate P1, P2 and P3 baculoviruses (general method 1.6.2).

4.3.3 Proteins expression test

A Hi5 monolayer plate (150 mm) was prepared, infected with 1 mL of YS63 P1 virus and incubated for 48 hours at 27 °C. Infected cells were harvested and proteins were extracted and separated into cytoplasmic (C), nuclear extract (N) and insoluble pellet (P) in hypotonic buffer condition (general method 1.6.2.3). Proteins in C, N and P were analyzed on a 18 % SDS PAGE at 200 V for 65 min.

200 mL suspension Hi5 cells were also prepared, infected with 5 mL of YS63 P3 virus and incubated for 48 hours at 27 °C with shaking at 95 rpm. Infected cells were harvested and proteins were extracted and separated into soluble supernatant (S) and insoluble pellet (P) by passing through EmulsiFlex-C3 homogenizer twice. Proteins in S and P were analyzed on a 18 % SDS PAGE at 200 V for 65 min.

4.3.4 Large scale protein expression

Suspension Hi5 cells were amplified up to 2L, infected with 40 mL of YS63 P3 virus and infected cells were incubated for 48 hours at 27 °C with shaking at 95 rpm (general method 1.6.3). Cells were harvested by centrifugation at 4,500 x g for 10 min at 4 °C. Cell pellets were processed immediately or stored at -80 °C for future purification.

4.3.5 Cell lysis and Nickel-affinity chromatography (HisTrap FF 5 mL)

Cell pellets were resuspended in 120 mL of lysis buffer (25 mM Tris-HCl pH 7.5, 2.0 M sodium chloride, 1 mM phenylmethanesulfonylfluoride, 0.5 Mm TCEP, 1 µg/mL Leupeptin hemisulfate salt and 1 µg/mL Pepstatin A). Resuspended cells were lysed by EmulsiFlex-C3 high pressure homogenizer and clarified by centrifugation at 38,500 x g at 4 °C for 1 hour. The supernatant was collected and imidazole stock solution was added to adjust the concentration to 30 mM. The clarified lysate was then loaded onto a HisTrap FF 5 mL column pre-equilibrated in

the Ni-buffer A (25 mM Tris-HCl pH 7.5, 2.0 M sodium chloride, 0.5 mM tris(2-carboxyethyl)phosphine followed by a 10 column volumes (CV) wash with Ni-buffer A containing 30 mM imidazole. The bound proteins were eluted by increasing the imidazole concentration from 30 mM to 500 mM linearly over 23.5 CV. Each faction was analyzed by 18% and 15 % SDS-PAGE to detect all expressed protein bands. Electrophoresis was run at 200 V for 65 min.

4.3.6 Dialysis

Pooled fractions from HisTrap FF 5 ml purification were dialyzed against 3.5 L dialysis buffer (25 mM Tris-HCl pH 8.8, 50 mM sodium chloride, 2 mM EDTA, 4 mM 1,4-dithiothreitol) by using Spectrum Spectra/Por 7 Membrane Tubing with 50 MWCO for 3 hours at 4 °C. Dialyzed samples were centrifuged for 10 min at 4,658 x g at 4 °C. Dialyzed proteins were analyzed by 18% SDS-PAGE. Electrophoresis was run at 200 V for 65 min.

4.3.7 Cation exchange chromatography (SouceQ 24 mL)

The dialyzed proteins were loaded onto a SourceQ 24 mL column pre-equilibrated in the SQ A buffer (25 mM Tris-HCl pH 7.5, 2 mM EDTA, 4 mM 1,4- dithiothreitol). The bound proteins were eluted by increasing the sodium chloride concentration from 0 mM to 300 mM linearly over 30 CV. Each faction was analyzed by 18% SDS-PAGE. Electrophoresis was run at 200 V for 65 min.

4.3.8 Size exclusion chromatography (Superdex 200 10/300 GL)

SourceQ 24 mL purified YS63 proteins at 110-135 mM, 136-159 mM and 160-180 mM were pooled separately and concentrated up to 7 mg/ml with ultrafiltration using Amicon YM50 membrane (MWCO 50 kDa) at 4 °C. The concentrated sample was then injected onto a Superdex 200 10/300GL column pre-equilibrated in the SD200 buffer (5 mM Tris-HCl pH 6.8, 800 mM

sodium chloride, 5 mM 1,4-Dithiothreitol). Each faction was analyzed by 18% and 15 % SDS-PAGE. Electrophoresis was run at 200 V for 65 min. The peak fractions were pooled and concentrated up to 8 mg/mL.

4.4 Results

Four subunits of TFIIH (Rad3, Tfb1, Tfb2 and Ssl1) in an acceptor transfer vector, pFL, and three subunits (Ssl2, Tfb4 and Tfb5) in a donor transfer vector, pUCDM were created by using multiplication module. Because many genes and restriction enzymes (NheI, SphI, EcoRI, HindIII, NotI, BssHII, Bstz17I, SpeI, AvrII, PmeI) were used in cloning, designing the order of gene insertion and selection of enzyme sites for each gene, were very tricky and time consuming. In addition, DNA extraction by using gel extraction kit of multigene cassettes digested by PmeI and AvrII was not easy due to similar size of two DNA fragments after digestion.

The transfer vector pUCDM containing three genes (YS62) was fused into a pFL containing four genes (YS60) by *in vitro* Cre-loxP fusion. The reaction mixture was incubated at 37 °C for 30 min and 70 °C for 10 min followed by transformation. Recombined plasmid (YS63) was extracted by miniprep kit and 2 µL of DNA product was digested with AvrII and PmeI to confirm the fusion of two vectors. Because each plasmid contains only one AvrII and PmeI sites, four bands should be generated if two vectors are fused into one. Four bands at correct position were observed on a 1 % agarose gel proving a DNA construct (YS63) with all seven core TFIIH subunit genes was successfully made (Fig. 21).

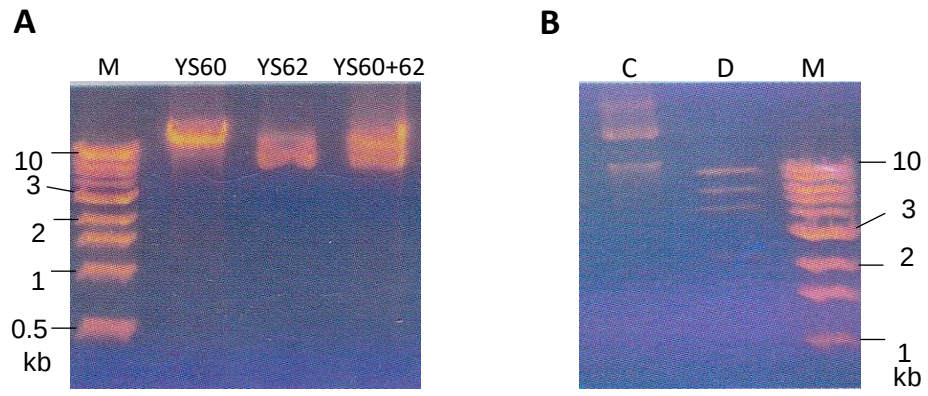


Figure 21: Analysis of Cre loxp recombination product YS63.
YS62 in pUCDM was fused into YS60 in pFL to produce YS63. **(A)** 1□l of miniprep YS60 (~13.7 kb), YS62 (~8.2 kb) and mixture of YS60 and YS62 before recombination were analyzed on a 1 % agarose gel. M, 1 kb DNA ladder; YS60+YS62, mixture of YS60 and YS62 before transformation when recombination was done. (B) recombined product after transformation followed by DNA extraction with total size of ~21.9 kb. 2 uL of recombined product was digested by AvrII and PmeI and generated four bands at ~9.4 kb, 6.0 kb, 4.3 kb and 2.2 kb. M, 1 kb DNA ladder; C, recombined product of YS60 and YS62; D, digested product by AvrII and PmeI.

Different from bacterial expression system, bacmids and viruses were made from YS63 plasmid to infect Hi5 cells for protein production. Recombinant bacmid was made by transposition and it was used to infect Sf9 cells to produce virus. The shape of virus infected Sf9 cells were examined under the microscope (infected cells are bigger in size and more granular) and P0, P1, P2 and P3 viruses were produced sequentially.

Protein expression was tested with both monolayer and suspension Hi5 cells infected by YS63 P1 virus. Proteins of all seven subunits were observed on both monolayer and suspension Hi5 cells, but expression of proteins were better with monolayer His5. Only Five subunits, Tfb1, Ssl1, Tfb2, Tfb4 and Tfb5, were soluble whereas two helicase subunits, Ssl1 and Rad3, were in insoluble pellet fraction.

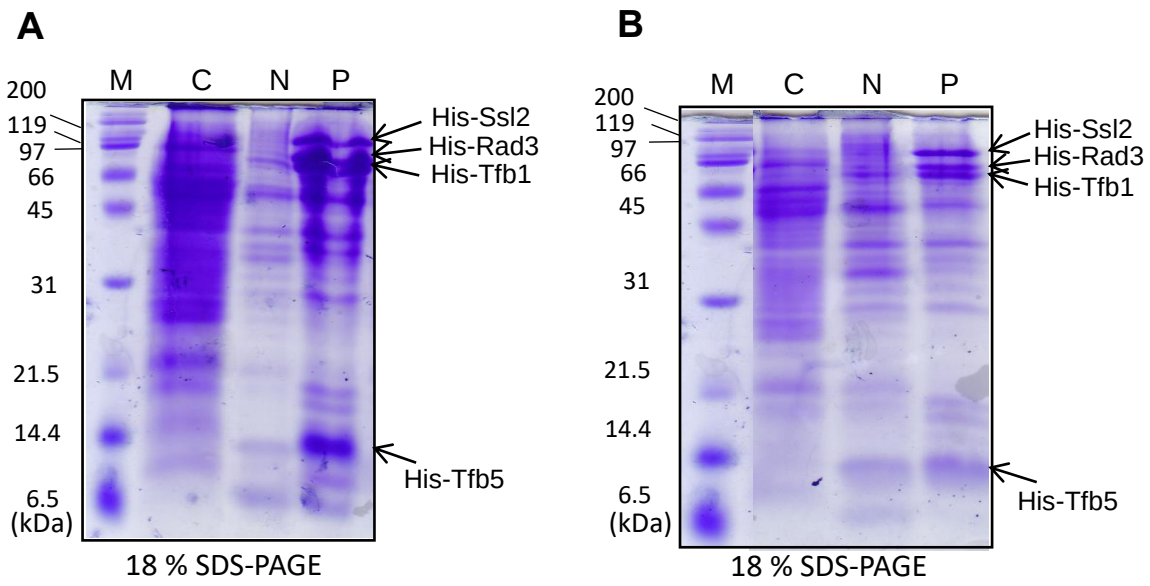


Figure 22: Coexpression and lysis of core-TFIIH on monolayer Hi5 and suspension Hi5 cells **(A)** Core TFIIH complex expression on monolayer Hi5 cells.
Cells were lysed in hypotonic condition. **(B)** Core TFIIH complex expression on suspension Hi5 cells. Cells were lysed in hypotonic condition. Both results were analyzed by18 % SDS-PAGE.

Large scale (2 L) cell culture was infected by 40 mL of YS63 P3 virus and soluble cell supernatant was first purified over HisTrap FF 5 mL column in the presence of 2 M sodium chloride and 30 mM imidazole. The column was subsequently washed with 10 column volumes of buffer, which effectively removed most other proteins. Fractions eluting at 70 to 229 mM imidazole contained clean TFIIH subunit Tfb1, Ssl1, Tfb2, Tfb4 and Tfb5 and they were pooled. (Fig. 23A).

HisTrap FF 5 mL purified samples were subsequently purified by SourceQ 24 mL cation exchange column. Mainly Tfb5 with little Tfb1, Ssl1 and Thb4 eluted at 110-135 mM sodium chloride (Input for 1st Superdex 200) followed by mainly Tfb1, Ssl1 and Thb4 at 136-159 mM sodium chloride (Input for 2nd Superdex 200). Lastly, there was mainly Tfb2 with little Tfb1, Ssl1 and Thb4 at 160 to 180 mM sodium chloride (Input for 3rd Superdex). (Fig. 23B)

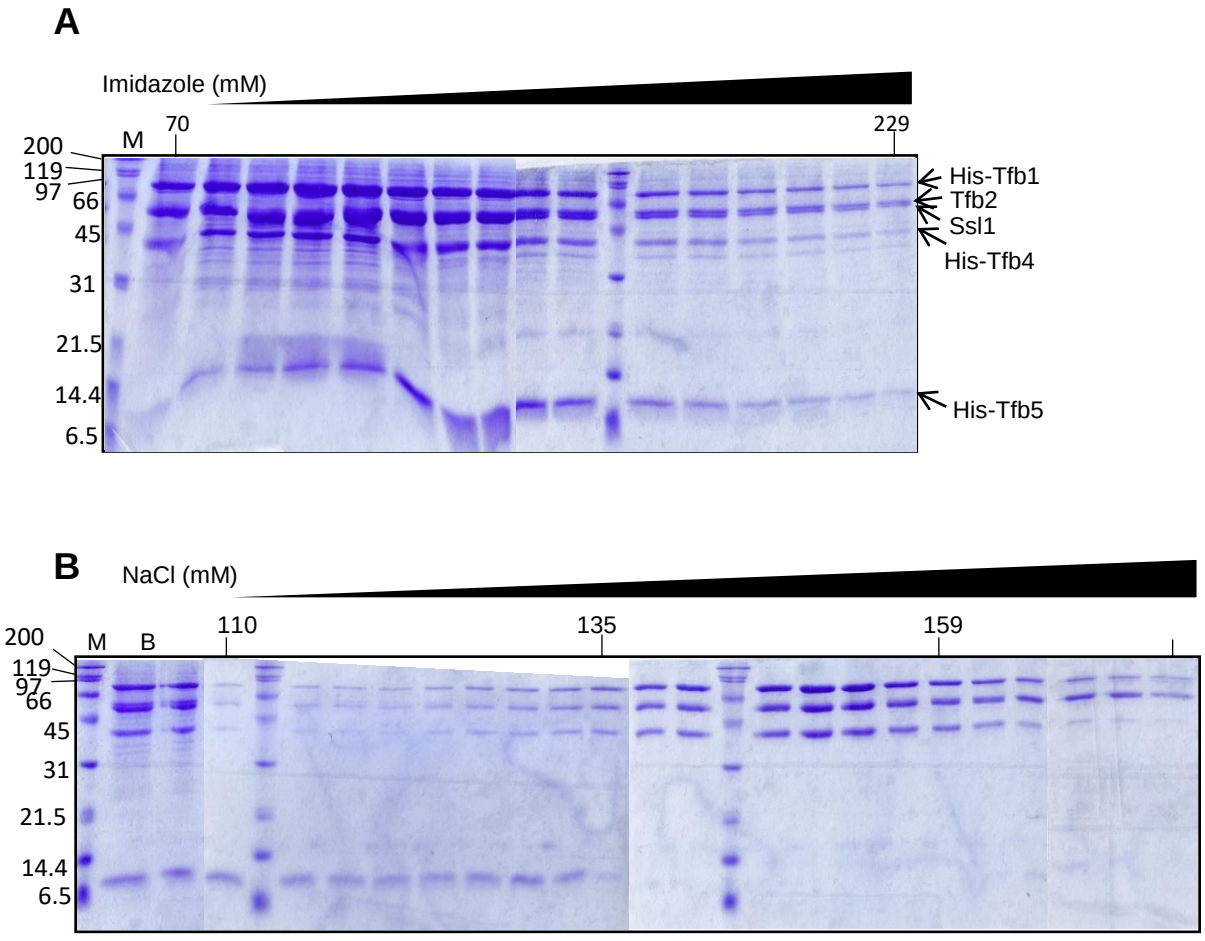


Figure 23: Purification of core TFIIH complex (YS63) by HisTrap FF 5mL and SourceQ 24mL. **(A)** Cell supernatant from core TFIIH (YS63) complex by infecting Hi5 suspension cell was first purified over a nickel affinity column (HisTrap FF 5 ml). SDS-PAGE shows that only five subunits were recovered in soluble cell lysate eluting at 70 mM to 229 mM imidazole. M, molecular weight marker; numbers indicating the concentration of Imidazole. **(B)** HisTrap FF 5 mL purified core TFIIH complex was purified further by SourceQ 24 mL. SDS-PAGE shows that Tfb5 eluted first around 110 to 135 mM sodium chloride followed by Tfb1, Ssl1 and Tfb4 around 136 to 159 mM sodium chloride, but Tfb2 eluted later separately around 160 to 180 mM sodium chloride. M, molecular weight marker; B, before dialysis; A, after dialysis; numbers indicating the concentration of Sodium chloride.

Each pooled fraction was concentrated and subsequently purified over a size exclusion column (Superdex 200 10/300GL column). Proteins eluted as a single peak at 11.7 mL corresponding to 137 kDa with calculated molecular weight of 167 kDa for his-Tfb1, Ssl1 and

his-Tfb4 complex. In SDS-PAGE analysis, Tfb5 was observed in input, but it didn't elute with Tfb1, Ssl1 and Tfb4. Tfb5 may dissociate and ran separately in later fractions, but it wasn't confirmed in this experiment (Fig. 24).

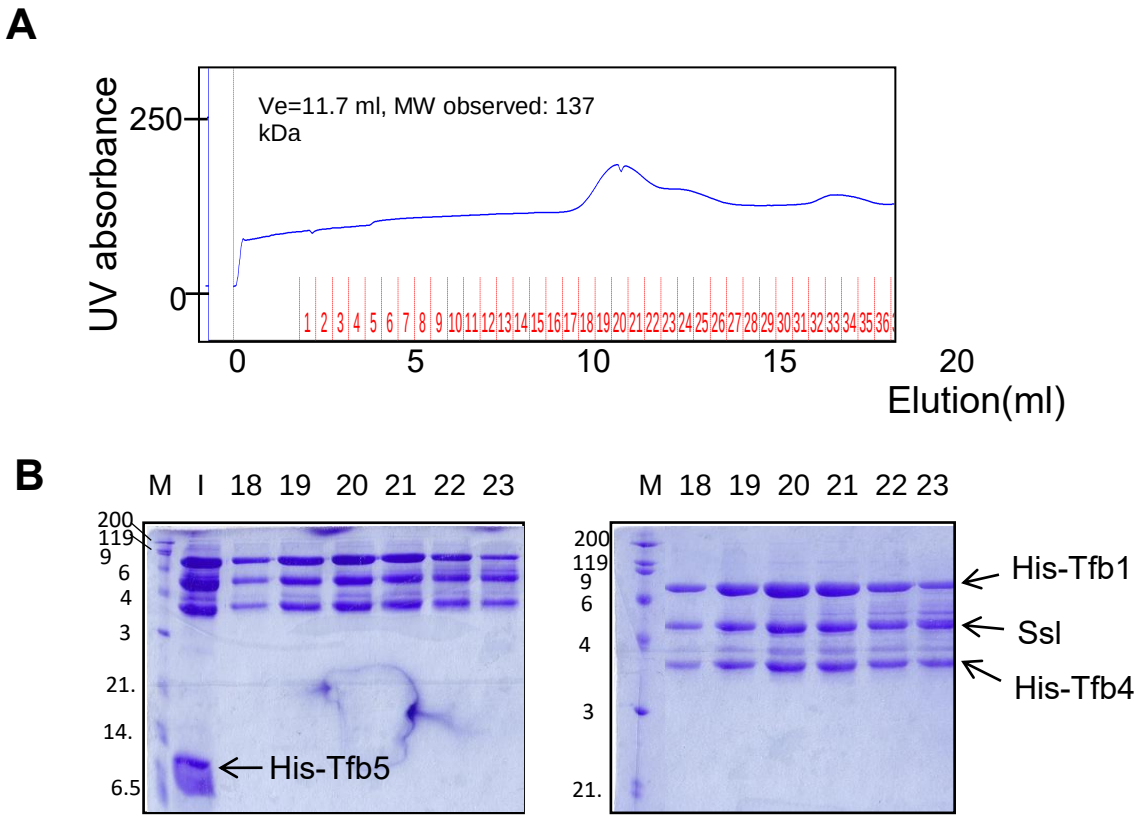


Figure 24: Purification of YS63 (core TFIIH) with 1st Superdex 200 column.
SouceQ 24 mL eluates from 110 to 135 mM sodium chloride was concentrated and finally purified over a Superdex 200 column. The chromatogram **(A)** and SDS-PAGE **(B)** show that Tfb1, Ssl1 and Tfb4 eluted at 11.7 mL, but not with Tfb5. 18 % SDS-PAGE (left) and 15 % SDS-PAGE (right), UV, ultraviolet; M, molecular weight marker; I, input; lanes 18 to 23 indicate fraction numbers.

Similar to the first Superdex 200 run, proteins eluted at 11.4 mL corresponding to 137 kDa with calculated molecular weight of 143.5 kDa. Peak fractions were analyzed on 18 % SDS-PAGE and coelution of Tfb1, Ssl1 and Tfb4 was examined. Second Superdex 200 run presented

a better single elution peak and SDS-PAGE also showed more homogenous Tfb1, Ssl1 and Tfb4 complex. (Fig. 25)

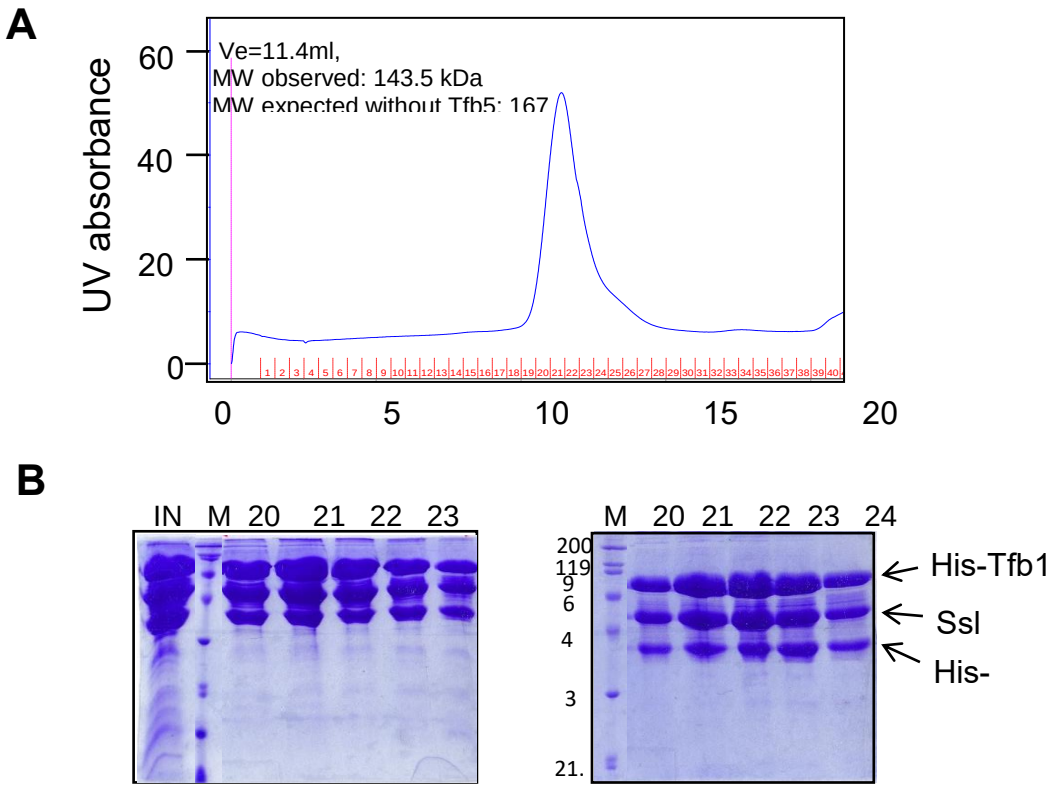


Figure 25: Purification of YS63 (core TFIIH) with 2nd Superdex 200 column.
SouceQ 24 mL eluates from 136 to 159 mM sodium chloride was concentrated and finally purified over a Superdex 200 column. The chromatogram **(A)** and SDS-PAGE **(B)** show that Tfb1, Ssl1 and Tfb4 eluted at 11.4 mL. 18 % SDS-PAGE (left) and 15 % SDS-PAGE (right), UV, ultraviolet; M, molecular weight marker; I, input; lanes 20 to 24 indicate fraction numbers.

The third Superdex run presented two peaks overlapping each other. Two peaks were observed at 11.4 mL and 12.5 mL corresponding 143.5 kDa and 121.9 kDa respectively. From SPS-PAGE, it was observed that first peak fractions contained mainly Tfb1, Ssl1 and Tfb4 followed by Tfb2 at a second peak position. Elution of Tfb1, Ssl1 and Tfb4 complex and Tfb2

was not separated well explaining the reason why two elution peaks were overlapped on a chromatogram (Fig. 26)

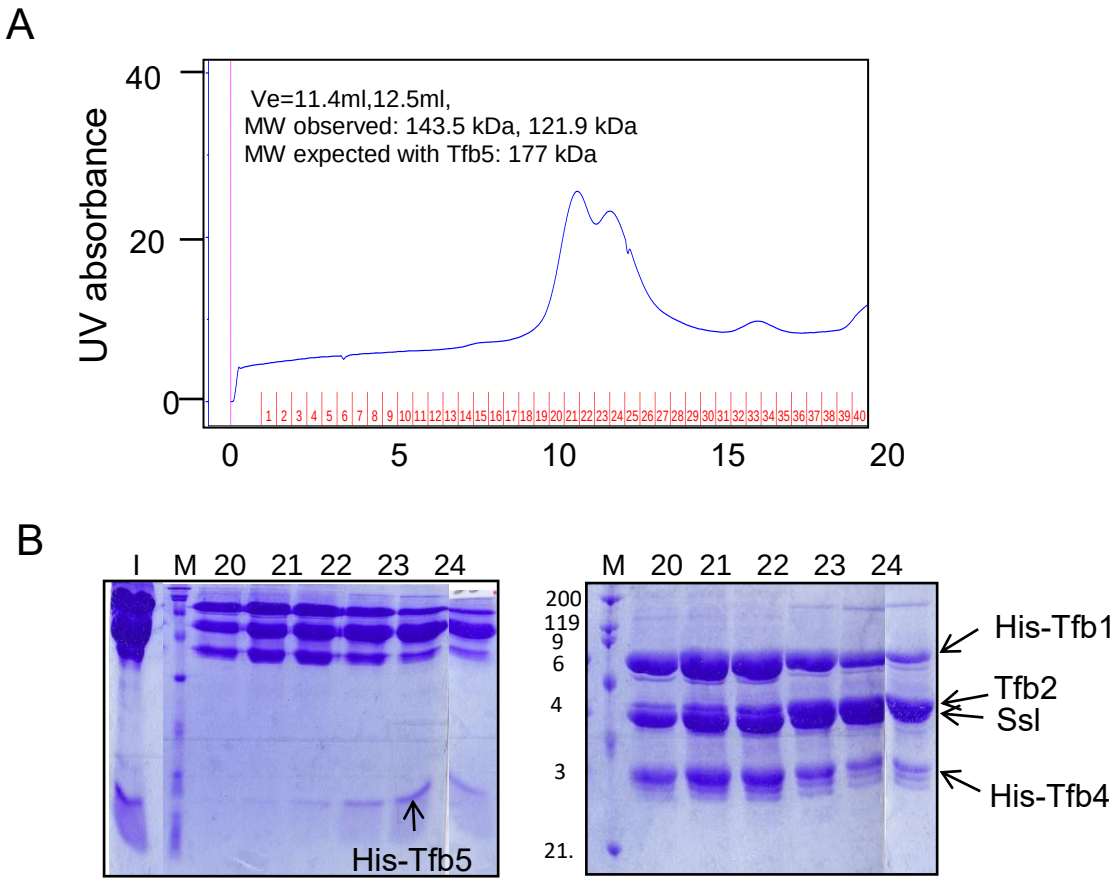


Figure 26: Purification of YS63 (core TFIIH) with 3rd Superdex 200 column.
SouceQ 24 mL eluates from 160 to 180 mM sodium chloride was concentrated and finally purified over a Superdex 200 column. The chromatogram **(A)** and SDS-PAGE **(B)** show that Tfb1, Ssl1, Tfb4 eluted first at 11.4 mL followed by Tfb2 at 12.5 mL. 18 % SDS-PAGE (left) and 15 % SDS-PAGE (right), UV, ultraviolet; M, molecular weight marker; I, input; lanes 20 to 25 indicate fraction numbers.

4.5 Discussions

Multiprotein complex TFIIH has been studied many years to examine the its structure and function in transcription and DNA repair, but it has been challenging to prepare holo or core TFIIH complex has been challenging due to its huge size with multiple subunits.

TFIIH can be prepared by immunopurification from large volume of Hela cells (up to few hundred liters) with monoclonal antibody column. However, growing large volumes of Hela cells and using monoclonal antibody column is costly and time consuming with lots of effort.

In addition, Each subunit of TFIIH can be produced by baculovirus recombinant polypeptides from insect cells [89]. However, obtaining the best virus titers to co-express several proteins equally is not easy. Virus titers should be tested every time when different virus stock is used.

Thus, MultiBac baculovirus expression vector system was introduced here to reconstitute core TFIIH complex. By expressing multiple protein simultaneously with MultiBac baculovirus expression vector system, high cost and tedious virus titers could be overcome. In addition, stability and solubility of proteins could be improved by expressing binding partners [108]. Therefore, seven subunits of core-TFIIH were assembled with MultiBac expression system.

Core-TFIIH seven subunit genes, Ssl2, Rad3, Tfb1, Tfb2, Ssl1, Tfb4 and Tfb5 were successfully assembled in one transfer vector, pFL, by using multiplication module with restriction enzyme and Cre-loxp recombination with Cre recombinase. When multiplication module was used to insert genes in a transfer vector, ligation failed more as plasmid became larger and larger with more genes inserted. Thus, several different ratios of vector to insert was tired for ligation. In addition, isolating DNA fragments by digesting with Bstz17I/SpeI and byAvrII/PmeI was difficult due to their similar sizes after digestion. 1 % agarose was run more about an hour at 150 V in running buffer (1x TAE buffer) at 4 °C in the presence of ethidium bromide for visualization. Fusion of two transfer vectors YS60 and YS62 by using cre recombinase was very simple and successful rate of recombining transfer vectors is quite high.

After cloning, core TFIIH complex (YS63) proteins expression was tested in small scale for both monolayer and suspension Hi5 cells. All seven core TFIIH subunits were confirmed by SDS-PAGE and monolayer showed higher expression of each subunit. However, helicase subunits, Ssl2 and Rad3, were unable to be recovered in soluble fraction after lysis in hypotonic condition for both cells. Because suspension cells are easier to amplify into large volume culture, 2 L of YS63 were expressed in suspension Hi5 cells by infecting with YS63 virus. All seven subunits were expressed and cells were lysed by EmulsiFlex-C3 high pressure homogenizer, but two helicase subunits, Ssl2 and Rad3, were still insoluble.

Five core TFIIH without helicase subunits, Ssl2 and Rad3, didn't elute together on SouceQ 24mL and Superdex 200 column. This might due to the absence of helicase subunits by destabilizing whole complex. Optimization is required to make helicase subunits soluble so that all core-TFIIH subunits are recovered in soluble fractions and used for characterizing the function and structural study of TFIIH in NER.

To solubilize Ssl2 and Rad3, other affinity tags can be tried for Ssl2 and Rad3 because it has been known that fusion tags may coordinate with protein and improve the solubility. In addition, lysis buffer condition can be optimized. Detergent, stabilizer or lower salt concentration (200 mM or 800 mM) have been known to improve solubility of proteins, so they can also help increase in solubility of Ssl2 and Rad3.

5 Proving Interaction of TFIIH and XPC-Rad23 complex in GG-NER by Biochemical assay

5.1 Introduction

In GG-NER pathway, XPC recognizes a lesion and it recruits TFIIH to the DNA damage. To characterize the role of TFIIH in NER in more detail, understanding how TFIIH is recruited to lesions by XPC-Rad23B and how they interact to each other is very important.

The composition, functions and structure of TFIIH has been introduced in Chapter 4. It is difficult to work with human XPC-Rad23B complex because it is prone to degradation due to its long internal loop. Fortunately, the strategy how producing stable and pure intact yeast Rad4-Rad23 complex has been established in our lab [109] , thus, yeast Rad4-Rad23 was used in this study.

The yeast Rad is 754 amino acids long with four domains, transglutaminase-homology domain (hereafter “TGD”, residues 123-432) and three β-hairpin domains (hereafter “BHD1”, residues 433-488; “BHD2”, 489-539; “BHD3”, 540-632). The four domains are sufficient for binding to a lesion and its binding partner Rad23. 100 residues on N-terminal and 122 residues on C-terminal are suggested to lack definite structure or freely folded based on limited proteolysis (LiP). N-terminal region of Rad4 (residues 1-167) is necessary for TFIIH binding [110] and the C-terminal Rad4 also plays important functions in recruiting TFIIH during NER [111].

In addition, XPC/Rad4 has known to directly interact with XPB/Ssl2 and p62/Tfb1 [112] [113]. XPC interacts with N-terminal XPB, which is next to the p52 binding site [114]. Tfb1 is 1926 amino acids long with PH domain followed by two BSD domains and 3-helix bundle motifs [115]. Acidic segment of N-terminal region of Rad4 (residues 76-115) was identified to

interact tightly with Tfb1 PH domain (residues 1-115). This Rad4-TFb1 binding shows that Tfb1 binding motifs is similar as endonuclease Rad2, an NER repair factor recruited by TFIIH [110].

In order to identify minimal component of TFIIH required for the interaction Rad4-Rad23 complex, MultiBac coexpression system was used to make trimeric complex of Tfb1 with Rad4-Rad23 complex. In the future study, the interaction has to be investigated with different truncated Rad4-Rad23 constructs with different TFIIH subunit combination containing Tfb1.

5.2 Methods

5.2.1 Full length his-Tfb1 with full length his-Rad4-Rad23 complex

5.2.1.1 Coinfection and colysis

Tfb1_1-1926 genes from yeast with hexahistidine (His6) tag was cloned into a pFastBac Dual expression vector by former postdoc, Dr. Bumseok Park and named as BP29. Hexahistidine (His6) tagged Rad4_1-754 genes from yeast with Rad23_1-398 was cloned into a pFastBac Dual expression vector by our lab advisor, Dr. Jung-Hyun Min and named as 144. Bacmid and P0-P3 viruses of BP29 and 144 were made. 1 L of Suspension Hi5 cells were infected by BP29 P3 virus (20 mL) and 144 P3 virus (20 mL) together. Coinfected cells were harvested after 48 hours incubation at 27 °C with shaking at 95 rpm.

Cell pellets were resuspended of lysis buffer (25 mM Tris-HCl pH 7.5, 0.8 mM sodium chloride, 1 mM phenylmethanesulfonylfluoride, 0.5 mM TCEP, 1 µg/mL Leupeptin hemisulfate salt and 1 µg/ml Pepstatin A). 60 mL lysis buffer was used per liter of cell cultures. Resuspended cells were lysed by EmulsiFlex-C3 high pressure homogenizer and clarified by centrifugation at 38,500 x g at 4 °C for 1 h. The supernatant was collected.

5.2.1.2 Nickel pull down with HisTrap FF 5 mL colum

Imidazole stock solution was added to the supernatant lysate to adjust 30 mM imidazole in supernatant lysate. Then, lysate was loaded onto a HisTrap FF 5 mL column (GE Healthcare) pre-equilibrated in the Ni-buffer A followed by a 10 column volumes (CV) wash with Ni-buffer A containing 30 mM imidazole. The bound proteins were eluted by increasing the imidazole concentration from 30 mM to 500 mM linearly over 23.5 CV. Each faction was analyzed by 15 % SDS-PAGE. Electrophoresis was run at 200 V for 65 min.

5.2.1.3 Dialysis

Pooled fractions from gravity nickel pull down or HisTrap FF 5 ml purification were dialyzed against 1 L of dialysis buffer (25 mM Tris-HCl pH 8.8, 50 mM sodium chloride, 2 mM EDTA, 4 mM 1,4- dithiothreitol) by using Spectrum Spectra/Por 7 Membrane Tubing with 50 MWCO for 3 hours at 4 °C. Dialyzed samples were centrifuged for 10 min at 4,658 x g at 4 °C. Dialyzed faction was analyzed by 18% SDS-PAGE. Electrophoresis was run at 200 V for 65 min.

5.2.1.4 Size exclusion chromatography

Dialyzed proteins were subjected onto a Superdex 200 10/300GL column pre-equilibrated in the SD200 buffer (5 mM Tris-HCl pH 6.8, 800 mM sodium chloride, 5 mM 1,4-Dithiothreitol). Each faction was analyzed by 18% SDS-PAGE: Electrophoresis was run at 200 V for 65 min.

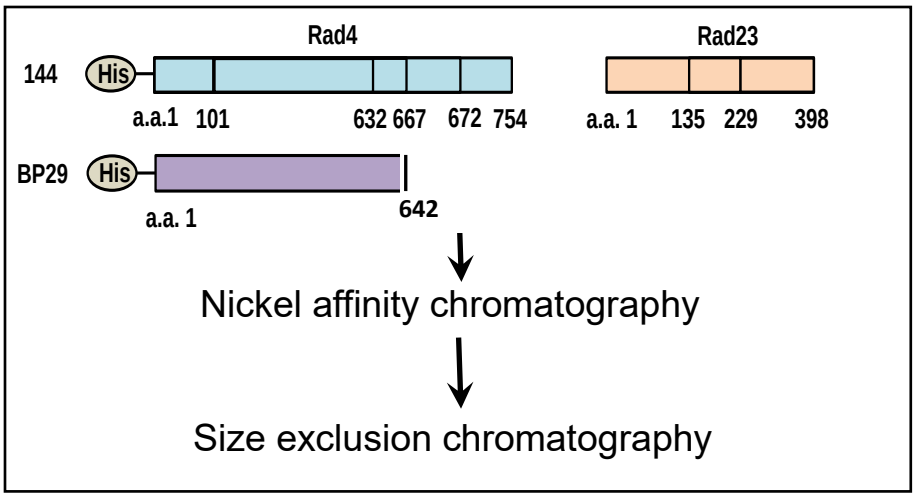


Figure 27: Overall procedure to study Tfb1 and Rad4-Rad23 interaction.
144 and BP29 were coinfected in suspension Hi5 cells and supernatant cell lysate was undertaken nickel affinity pull-down followed by size exclusion chromatography.

5.2.2 Full length his-Tfb1 with truncated his-Rad4-Rad23 complex

5.2.2.1 Coinfection and colysis

Three different versions of truncated Rad4-Rad23 complex, XC42 (Rad4_101-667, GST-Rad23_1-398 with 135-229 deletion), XC43 (Rad4_101-672, GST-Rad23_1-398 with 135-229 deletion) and XC44 (Rad4_101-754, GST-Rad23_1-398 with 135-229 deletion), were made by former students, Dr. Xuejing Chen. Bacmids and P0-P3 viruses were also prepared by her. Hi5 monolayer cells were coinfected by BP29 (Tfb1_1-1926 genes from yeast with hexahistidine (His6) tag) with different truncated Rad4-Rad23 complex. One Hi5 monolayer plate was coinfected by P3 viruses (1 mL each) of BP29 and each Rad4-Rad23 truncated complex. Coinfected cells were harvested after 48 hours incubation at 27 °C and proteins were extracted and separated into cytoplasmic (C), nuclear extract (N) and insoluble pellet (P) in hypotonic buffer condition (general method 1.6.2.3). Proteins in C, N and P were analyzed on a 18 % SDS PAGE at 200 V for 65 min.

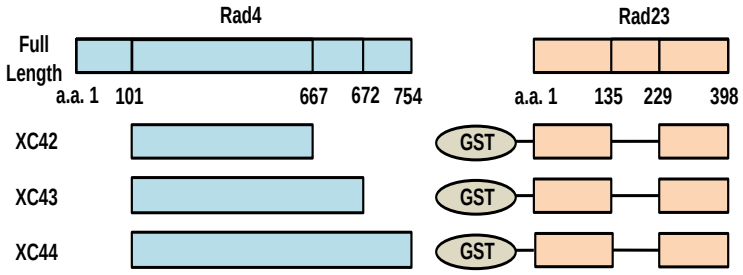


Figure 28: Truncated Rad4-Rad23 complex used for pull-down assay with Tfb1.
Monolayer Hi5 cells were coinfected with each truncated Rad4-Rad23 complex and Tfb1 viruses followed by nickel affinity pull-down.

5.2.2.2 Gravity Nickel Pull down with Ni-NTA agarose

50 uL of Ni-NTA agarose equilibrated in lysis buffer (25 mM Tris pH8.0, 500 mM NaCl , 0.5 mM TCEP, 1 mM PMSF, 1 µg/mL Leupeptin & 1 µg/mL Pepstatin) was added to C and N, and incubated for 1 hour on a nutator at 4 °C. The sample was centrifuged at 1,000 rpm for 1 min and the supernatant was transferred to a 2 ml microcentrifuge tube. The resin was washed with 200 µL of lysis buffer (=W) followed by 100 µL of lysis buffer with 5 mM imidazole (=E5), 100 µL of lysis buffer with 20 mM imidazole (=E20), 100 µL of lysis buffer with 80 mM imidazole (=E80), 100 µL of lysis buffer with 320 mM imidazole (=E320) and 100 µL of lysis buffer with 2 M imidazole (=E2M). Each faction was analyzed by 18% SDS-PAGE. Electrophoresis was run at 200 V for 60 min.

5.2.3 Creating co-expressed cassettes of Tfb1-Rad4-Rad23 complex

Glutathione S-transferase tag (GST) was inserted into p10 promoter between NheI and SphI sites and named as YS78. Rad23 gene was inserted into p10 promoter of pFL between NcoI and NheI and named as YS80. Rad4 gene was inserted into pH promoter of pFL vector between NotI and BssHII sites and named as YS81.

Tfb1, Rad4 and Rad23 genes were assembled by using multiplication module. The entire multigene expression cassette of BP29 was removed by PmeI and AvrII digestion inserted into multiplication module of YS81 digested with Bstz17I and SpeI. The fragment of expression cassette of BP29 and multiplication module digested YS81 were purified by using Gel extraction kit followed by ligation. The ligation product was named as YS82.

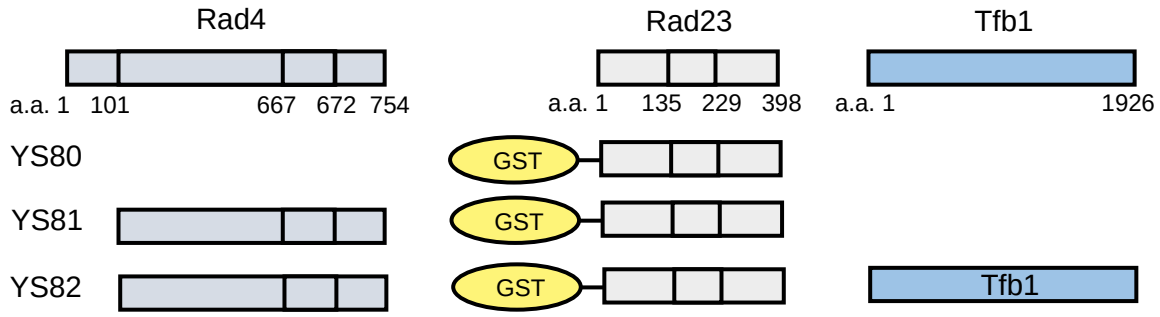


Figure 29: Creating Tfb1with truncated Rad4-Rad23 coexpression cassettes.
YS80, YS81 and YS82 contain Rad4, Rad23 and Tfb1 in one plasmid and Rad23 contains Glutathione S-transferase tag (GST).

5.2.4 Creating multigene TFIIH complex including Tfb1

YS94 containing Tfb1, Tfb4 and Ssl1 and YS95 containing Tfb1, Tfb4, Ssl1 and Tfb5 were made by using multiplication module. After bacmid and P0-P3 viruses were made, each 1 L Hi5 suspension cells were infected with YS94 or YS95 P3 virus. Proteins in soluble supernatant cells were purified by HisTrap FF 5 mL, SourceQ 24 mL followed by Superdex 200 column.

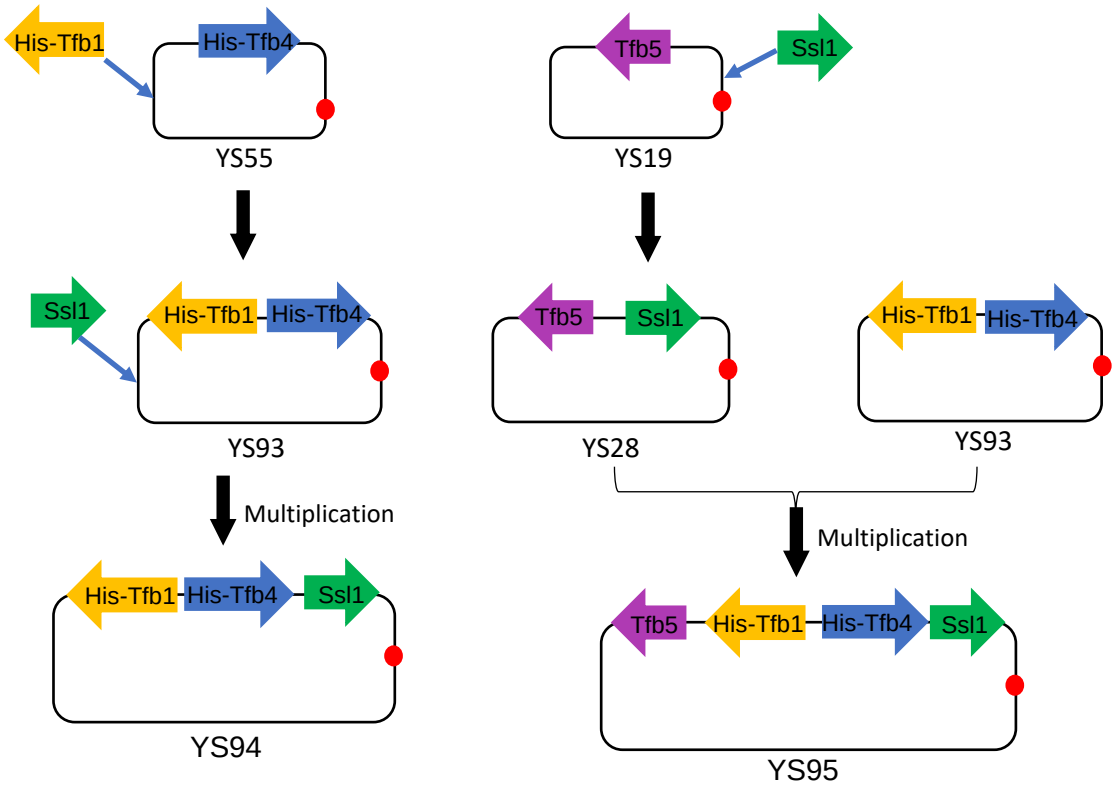


Figure 30: Creating three and four multigene expression cassettes containing Tfb1.

5.3 Results and discussions

Suspension Hi5 cells were coinfected by P3 virus of BP29 (His6 tagged full length Tfb1) and 144 (His6 tagged full length Rad4-Rad23 complex). Supernatant cell lysate was subjected to gravity nickel pull down assay. Nickel resin was washed with buffer containing imidazole and proteins eluted by increasing imidazole concentration to 2 M. Each fraction was analyzed by 15 % SDS-PAGE.

Tfb1 and full length Rad4-Rad23 coeluted on HisTrap FF 5 mL at 104 to 193 mM imidazole (Fig 31A), but this doesn't prove their interaction because both BP29 and 144 proteins contain same His6 tag. Fractions containing both proteins were dialyzed to change the buffer composition and dialyzed proteins were subjected onto a Superdex 200 column. Full length his-Rad4-Rad23 complex eluted first followed by full length his-Tfb1 (Fig.31B). On a Superdex 200

column, proteins eluted together with a single peak, but Rad4-Rad23 eluted little earlier than Tfb1.

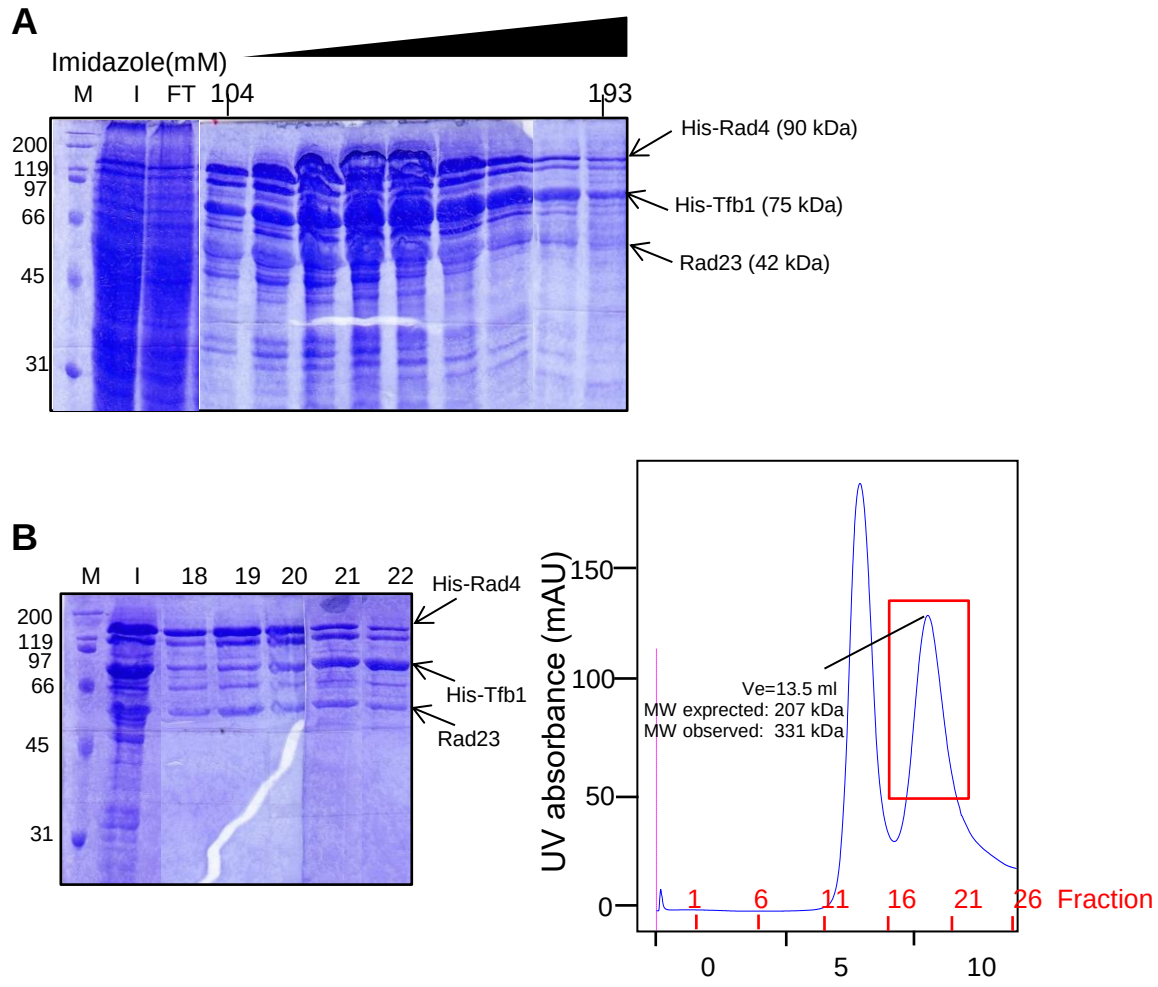


Figure 31: Investigating Interaction between full length Tfb1 and full length Rad4-Rad23 complex by pull-down assay and size exclusion chromatography.
(A) Cell supernatant from co-infected BP29 (his-Tfb1) and 144 (his-Rad4-Rad23 complex) was loaded onto a nickel affinity column (HisTrap FF 5 mL). SDS-PAGE shows Tfb1 and Rad4-Rad23 complex co-eluted at 104-193 mM. M, molecular weight marker; numbers indicating the concentration of Imidazole. **(B)** Co-eluted protein complex from Histrap FF 5 mL was loaded onto a Superdex 200 column. The SDS–PAGE (left) and chromatogram (right) show co-eution of his-Tfb1 and his-Rad4-Rad23 at 13.5 ml, but his-Rad4-Rad23 eluted little earlier than his-Tfb1. UV, ultraviolet; M, molecular weight marker; I, input; lanes 6 to 13 indicate fraction numbers.

As mentioned above, the results were not conclusive to determine whether Rad4-Rad23 complex interact with Tfb1 due to the presence of same affinity tags on Tfb1 and Rad4-Rad23 complex. Thus, XC42, XC43 and XC44 (truncated Rad4- GST-Rad23 complex) made by Dr. Xuejing Chen previously were used with BP29 (his-Tfb1) for GST pull down assay. Only coinfection by BP29 and XC44 (Rad4_101-754, GST-Rad23_1-398 with 135-229 deletion) viruses showed co-elution of Rad4-Rad23 and Tfb1. The expression level and pattern may vary due to inaccurate virus titer. Coinfection is not an ideal method to express multiple proteins as described earlier in introduction.

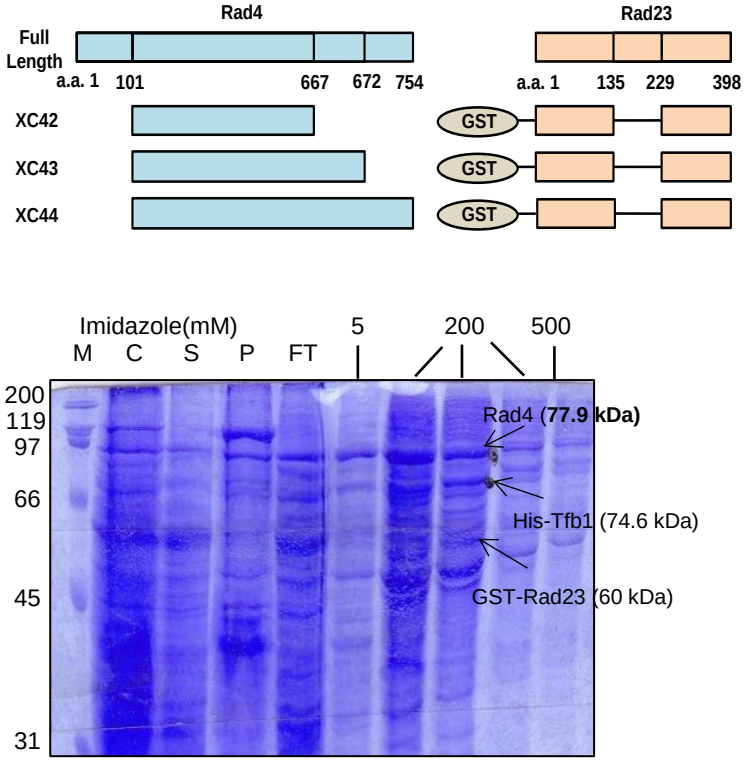


Figure 32: Coinfection and colysis of his-Tfb1 with truncated Rad4-GST-Rad23.
Interaction between Rad4-Rad23 and Tfb1 was examined by Nickel pull-down assay. Three different constructs of Rad4-Rad23 complex used for coinfection with Tfb1 (top). Only XC44 (Rad4_101-754, GST-Rad23_1-398 with 135-229 deletion) and Tfb1 coeluted.

Multigene expression cassettes containing Tfb1 with truncated Rad4-Rad23 complex were made by using MultiBac baculovirus expression system. Because Tfb1 was pull down with XC44 (Rad4_101-754, Rad23_1-398 with 135-229 deletion), same Rad4 and Rad23 residues with GST tagged Rad23 was made (Fig 29.) Bacmids and viruses need to be made and tested and interaction should be confirmed by GST pull down assay in the future.

From core-TFIIH subunit expression (YS63), Ssl2 and Rad3 were insoluble and Tfb2 eluted after Tfb1, Tfb4 and Ssl1 and Tfb5 doesn't elute with other subunits. Thus, only three subunits Tfb1, Tfb4 and Ssl1 seems eluting together stably. Tfb1 is in contact with Tfb4 and Ssl1 and that could be why they are stably forming a complex. Tfb2 is also in contact with Tfb1, but it is also adjacent to Tfb5 and Ssl2. The absence of Ssl2 subunit may promote dissociation of Tfb2 from other complex.

Therefore, YS94 containing Tfb1, Tfb4 and Ssl1 and YS95 containing Tfb1, Tfb4, Ssl1 and Tfb5 were made to use to identify minimal requirement for XPC-TFIIH interactions. YS94 and YS95 were made successfully, expressed and purified by HisTrapFF 5 ml column followed by SourceQ 24 ml and superdex 200 column.

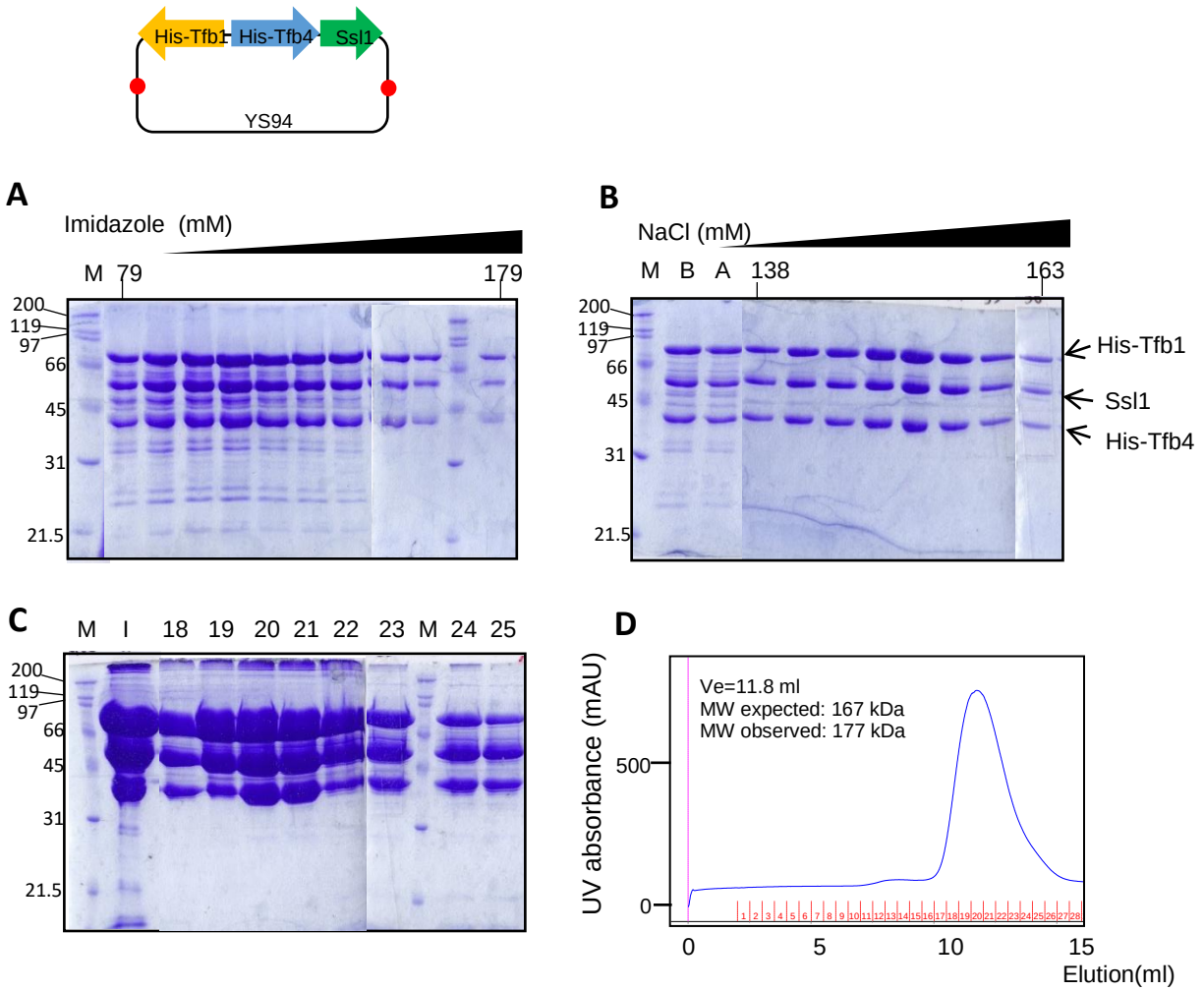


Figure 33: Purification of three subunits (Tfb1, Tfb4 and Ssl1) of TFIIH assembled by baculovirus system.

(A) Cell supernatant from YS94 expression Hi5 suspension cells was first purified over a nickel affinity column (HisTrap FF 5 ml). SDS-PAGE shows that all three subunits were recovered in soluble cell lysate eluting at 70 mM to 179 mM imidazole. M, molecular weight marker; numbers indicating the concentration of Imidazole. **(B)** HisTrap FF 5 ml purified YS94 was purified further by SourceQ 24 ml. SDS-PAGE shows that all three subunits eluted around 138 to 163 mM Sodium chloride. M, molecular weight marker; B, before dialysis; A, after dialysis; numbers indicating the concentration of Sodium chloride. **(C)** SouceQ 24 ml purified YS94 was concentrated and finally purified over a Superdex 200 column. The chromatogram **(C)** and SDS-PAGE **(D)** show that Tfb1, Tfb4 and Ssl1 eluted together at 11.8 ml. UV, ultraviolet; M, molecular weight marker; I, input; lanes 18 to 25 indicate fraction numbers.

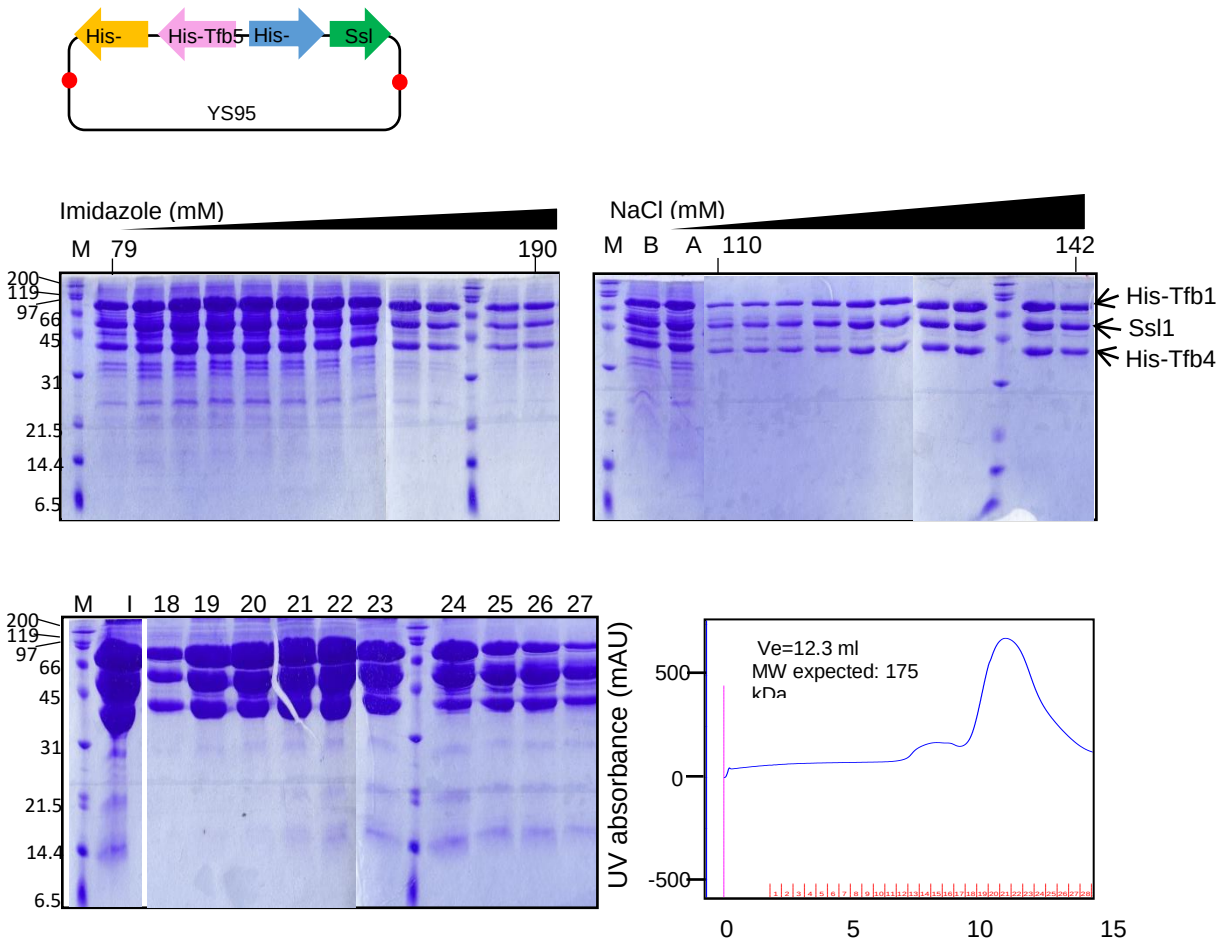


Figure 34: Purification of four subunits (Tfb1, Tfb4, Ssl1 and Tfb5) of TFIIH assembled by baculovirus system.
(A) Cell supernatant from YS95 expression Hi5 suspension cells was first purified over a nickel affinity column (HisTrap FF 5 ml). SDS-PAGE shows that only three subunits without Tfb5 were recovered in soluble cell lysate eluting at 79 mM to 190 mM imidazole. M, molecular weight marker; numbers indicating the concentration of Imidazole. **(B)** HisTrap FF 5 ml purified YS95 was purified further by SourceQ 24 ml. SDS-PAGE shows that three subunits eluted around 110 to 142 mM Sodium chloride. M, molecular weight marker; B, before dialysis; A, after dialysis; numbers indicating the concentration of Sodium chloride. **(C)** SouceQ 24 ml purified YS95 was concentrated and finally purified over a Superdex 200 column. The chromatogram **(C)** and SDS-PAGE **(D)** show that Tfb1, Tfb4 and Ssl1 eluted together at 12.3 ml. UV, ultraviolet; M, molecular weight marker; I, input; lanes 18 to 27 indicate fraction numbers.

To characterize the interaction of TFIIH with NER factor Rad4-Rad23, more biochemical assay (probably pull down) should be performed with more truncated Rad4-Rad 23 constructs in

the future. In addition, more TFIIH sub complexes can be made in the future. If helicase subunits are recovered in soluble fraction, it would be easier to make more stable core-TFIIH sub complexes.

6 CONCLUSION

In this dissertation, I mainly describe the co-expression and reconstitution of NCP and 5-subunit core TFIIH complexes. They are key components of various genetic transactions but their studies have been difficult because of the challenges in preparing the complexes for further mechanistic studies. I envision that the studies described in the dissertation will lay foundations for future studies in the field, in particular on the structure and mechanism of DNA damage recognition and repair by NER in the context of chromatin DNA. With the advantage of cutting-edge structural techniques such as cryo-electron microscopy and biophysical tools to investigate dynamic molecular motions, uncovering the high-resolution structural and dynamical mechanisms of these molecular transactions may be possible in near future.

APPENDIX

1. General Methods

1.1. PCR

- Prepare PCR mixture and the enzyme last.

#	Item (stock conc.)	Volume (µL)
1	autoclaved Milli-Q H_2O	26.0
2	5x Phusion HF Buffer	10.0
3	dNTP (2 mM, final 0.2 mM)	5.0
4	DMSO	2.5
5	Template (~10 ng/ml. Use 10-50 ng)	1.0
6	Primer1 (10 uM, final 0.5 uM)	2.5
7	Primer2 (10 uM, final 0.5 uM)	2.5
8	Phusion HF Polymerase (2 U/ul)	0.5
	Total	50.0

- Run the PCR cycles as follows.

	Temperature (°C)	Time (sec)
Initial denaturation	98 °C	30
Cycles	98 °C 55 °C* 72 °C	5-10 10-30 (15-30) / kb x 30
Final extension	72 °C	5-10 min
Storage	4 °C	infinite

Annealing temperature = Tm+3°C for primers < 20 nucleotides

=Tm for primers > 20 nucleotides

- Run 1 % agarose gel with 1 µL of PCR product for 7-10 min at 180 V.

- If PCR works successfully, load all PCR product on 1 % agarose gel and run for 15-20 min at 150 V.
- Excise the gel-containing DNA of interest and extract DNA with QIAquick Gel Extraction Kit (Qiagen).
- Elute DNA with 45 µl of autoclaved Milli-Q H2O.

1.2. Restriction Enzyme digestion

- <Double digestion of insert with two restriction enzymes>
- Add 10 U(unit) of both restriction enzymes (New England Biolabs) with 5 µL of 10 X reaction buffer, which is compatible for both enzymes, to the purified DNA product from PCR. Final volume is around 50 µL.
- Incubate the reaction mixture at the optimal temperature (usually 37 °C) for both enzymes for 1 hour. If optimal temperature is different, incubate for one hour at each different temperature.
- Purify the insert with QIAquick PCR Purification Kit.
- <Vector digestion>
- Calculate the volume of 5 µg of vector and 10 U(unit) of one of the enzyme with 5 µL of 10 X reaction buffer.
- Incubate the reaction mixture at the optimal temperature for the enzyme for 1 hour.
- Run 1 µL of each incubated mixture and uncut vector (control) on 1% agarose gel to check completeness of digestion. (uncut circular vector runs faster than digested linear one.)
- If digestion is complete, road all on 1 % agarose gel and run at 150 Vfor 15-20 min.

- Excise the gel band containing digested vector and purify it with QIAquick Gel Extraction Kit.
- After eluting vector with 45 µl of autoclaved Milli-Q H_2O, add 10 U(unit) of the other enzyme with 5 µL of 10 X reaction buffer.
- Incubate at the optimal temperature for the enzyme for 1 hour.
- Add 1 µL of Alkaline Phosphatase Calf Intestinal (CIP, New England Biolabs) and incubate at 37 °C for 1 hour.
- Load all the reaction mixture on 1% agarose gel and run at 150 V for 15-20 min.
- Excise the gel band containing DNA and purify it with QIAquick Gel Extraction Kit.
- Elute digested DNA with 30 µL of autoclaved Milli-Q H_2O.

1.3. Ligation

- Run 1 µL of each vector and insert on 1 % gel to estimate the concentration.
- Calculate the volume of about 50 ng of vector and the volume of insert that is two to ten times more than vector in moles. (Two conditions are usually set up for ligation.)

	Amount
Vector	~50 ng
Insert	2 – 10 folds more than vector in moles
10x T4 ligase buffer	2 µL
T4 DNA ligase	1 µL
Autoclaved Milli-Q H_2O	q.s.
Total	20 µL

- Incubate the reaction mixture for 2 hours at room temperature (~25 °C) or at room temperature for 2 hours or for overnight at 16 °C.

1.4. Transformation and colony identification

<Transformation>

- Thaw one tube (100 µL) of XL-Blue competent cells on ice and add 10 uL of ligation reaction mixture to the cells.
- Tap the tube gently several times and incubate them on ice for 30 min.
- Incubate the cells at room temperature for 5 min.
- Put them back to the ice and incubate for 2 min.
- Add 900 µL of SOC medium and incubate for 60 min at 37 °C with vigorous shaking (usually 170-200 rpm).
- Centrifuge for 1 min at 13.3 krpm.
- Discard 700 µL of supernatant and resuspend the cells in the remaining 300 µL media.
- Plate the cells onto LB agar with appropriate antibiotic and spread the cells using glass beads.
- Incubate the plates for overnight (about 16-18 hours) at 37 °C.
- Next day, check the plates and select several colonies for positive clone identification.
- Prepare several (depends on how many colonies are tested) 3 mL of LB medium with antibiotic in 15 mL falcon tube and add each single colony using a pipette tip into the tube.
- Inoculate the media for overnight at 37 °C with vigorous shaking.
- Next day (after 16-18 hours), transfer 2 mL of each culture to 2 mL micro centrifuge tube and centrifuge them for 1 min at 13.3 krpm and aspirate all supernatant.

- Purify DNA with Qiagen Miniprep kit (Qiagen) and elute with 50 µL of Milli-Q H2O.

< Colony identification with double digestion>

	Volume
Miniprep DNA	2 µL
10X reaction buffer	1 µL
Autoclaved Milli-Q H_2O	7 µL
Restriction enzymes used for vector and insert	0.1 µL each
Total volume	10 µL

- Set up the double digestion reaction
- Incubate the reaction mixture for 1 hour at 37 °C.
- Run 5 µL of the product on 1 % agarose gel. Positive clone should have two bands. (one for vector and other one for insert. – check the size of the band, too)
- Send out one or two positive clones for sequencing to confirm.

1.5. Positive clone amplification

- Prepare large amount of plasmid using NucleoBond midiprep kit.

1.6. Protein production

1.6.1. Protein production from bacterial cell culture

1.6.1.1. Transformation

- Thaw 1 tube (100 µL) of appropriate competent cells on ice and add 1 ul of miniprep plasmid into the cells.
- Tap the tube gently several times and incubate them on ice for 15 min.
- Apply heat shock at 42 °C (using water bath) for 45 sec.

- Put them back to the ice and incubate for 2 min.
- Add 900 µl of LB media and incubate for 60 min at 37 °C with vigorous shaking.
- Centrifuge for 1 min at 13.3 krpm.
- Discard 700 µl of supernatant and resuspend the cells in the remaining 300 µl.
- Plate the cells onto LB agar with appropriate antibiotic and spread the cells using glass beads.
- Incubate the plates for overnight (about 16-18 hours) at 37 °C.

1.6.1.2. Protein expression

- Grow 10 mL for expression test (small scale) and 1 L to 8 L for large scale protein production.
- Prepare 10 mL of autoclaved 2xTY medium with appropriate antibiotic in 50 mL falcon tube.
- Pick a single colony with pipette tip from LB agar plate, add to the prepared 10 mL 2xTY medium with antibiotic and incubate at 37 °C with vigorous shaking until culture become slightly cloudy (usually takes 2-3 hours).
- Transfer 10 mL culture to autoclaved 250 mL 2xTY medium in 500 mL flask with antibiotic and incubate for 2-3 hours in same condition.
- Transfer 250 mL culture to autoclaved 750 mL of 2xTY medium in 2 L flasks and grow culture until when the OD_{600} reach ~ 0.4.
- Induce protein expression by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to make final concentration as 0.4 mM.
- Incubate the culture with shaking the for overnight at 37 °C.
- Harvest cells by centrifugation at 4,500 x g for 10 min at 4 °C.

- Store cell pellets at -80 °C for future purification.

1.6.2. Protein production from Insect cell culture

1.6.2.1. Transposition and bacmid production

- Thaw 50 µL of DH10Bac competent cells on ice and transfer all into 15 ml rounded bottom polypropylene tube. (Fisherbrand)
- Add 1 µL of miniprep plasmid to the cells.
- Tap the tube several times gently and incubate it on ice for 30 min.
- Apply hear shock at 42 °C for 45 s in a water bath.
- Incubate it on ice for 2 min.
- Add 950 µL SOC medium (MP Biomedicals) to the tube (total ~ 1 ml).
- Incubate the cells with shaking at 200 rpm at 37 °C for 4 hours.
- Add 10 µL of culture to 90 µL of SOC in 1.5 mL micro centrifuge tube and place all onto a Bac-to-Bac agar plate with following components.

Bac-to-Bac plates	Concentration (µg/mL)
Kanamycin	50
Gentamycin	7
Tetracyclin	10
Bluo-gal	100
IPTG	40

-
- To prevent light to fall on the plate, cover it with aluminum foil and incubate them for 48 hours at 37 °C (White colony: successful transposition).
- Inoculate a single, isolated white colony into 2 ml LB media containing 50 ug/mL Kanamycin, 7 µg/ml Gentamicin and 10 µg/ml Tetracycline. Grow at 37 °C for 24 hours with shaking.

- Transfer 1.5 mL of the culture into a 1.5 mL microcentrifuge tube. Spin down and remove the supernatant.
- Re-suspend the cells in 0.3 mL of cold P1 buffer from Plasmid Maxi Kit (Qiagen).
- Lyse the cells by adding 0.3 mL of room temperature P2 buffer from Plasmid Maxi Kit. Incubate at room temperature for 5 min.
- Neutralize by adding 0.3 mL of P3 buffer from Plasmid Maxi Kit. Incubate on ice for 5-10 min.
- Centrifuge at 13,300 rpm for 10 min.
- Prepare 2.0 mL microcentrifuge tubes with 0.8 mL isopropanol.
- Transfer the supernatant to 2.0 mL microcentrifuge tubes containing 0.8 mL isopropanol.
- Centrifuge at 13,300 rpm for 15 min and discard supernatant. (Pellet might be visible.)
- Add 0.5 mL of 70% cold ethanol to the pellet and invert the tube several times.
- Centrifuge at 13,300 rpm for 5 min.
- Remove as much of the supernatant as possible and air dry the DNA pellet at room temperature for 5 – 10 min until all liquid has evaporated.
- Re-dissolve the bacmid DNA in 40 µL TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

1.6.2.2. Transfection and virus production

- Split the cells from one confluent Sf9 plate (from Falcon, d=150 mm) into 1:12 ratio to 6-well tissue culture plate (from Falcon, d=35 mm) and leave it for one hour until all the cells are attached to the bottom of the plate.
- Wash the cells with 2 mL of unsupplemented Grace’s Insect media (Invitrogen) and remove the media.

- Add 12 µL bacmid to 100 µL of unsupplemented Grace's Insect media in an autoclaved 1.5 mL mircrocentrifuge tube A.
- Add 6 µL of Cellfectin II reagent (Invitrogen) in another autoclaved 1.5 mL mircrocentrifuge tube B.
- Transfer the bacmid mixture from tube A to tube B and incubate the tube at room temperature for 30 min.
- Add 800 µL of unsupplemented Grace's Insect media to the mixture.
- Add all the mixture to the 6-well tissue culture plate containing Sf9 cells and incubate at 27 °C for 5 hours.
- Add 10 % FBS (Atlanta biologicals), 250 fold diluted Pen/Strep/L-Glutamine (Lonza), 1000 fold diluted Fungizone Antimycotic (Gibco) and 10 µg/mL Gentamicin (Gibco) to supplemented Grace's Insect media to make complete Grace's Insect media.
- Remove media which is the transfection mixtures from 5 hour incubated plate and add 2 mL of complete Grace's Insect media.
- Incubate the plate at 27 °C for 72 hours and harvest P0 virus by collecting the supernatant media into a sterile 2 mL cryo tube.
- Add 1 mL of P0 virus to the medium aspirated Sf9 plate.
- Incubate it for 1 hour at room temperature and swirl the plate every 20 min.
- Add 20 mL of complete Grace's Insect media to the plate and incubate for 72 hours at 27 °C.
- Harvest P1 virus by collecting supernatant media into 50 ml centrifuge tube and centrifuge it for 5 min at 3000 rpm.
- Filter the supernatant with 0.22 µm filter into 50 mL centrifuge tube and store at 4 °C.

- Make P2 from P1 and P3 from P2 in the same way.

1.6.2.3. Protein expression from Hi5 monolayer cells

- 1 plate of Hi5 cells can be used for expression test and 10-30 plates for large scale protein production.
- Aspirate the media from a confluent Hi5 plate and add 1 mL of P1 virus.
- Incubate the plate at room temperature for 1 hour and swirl it every 20 min.
- Add 20 mL of complete Grace's Insect media to the plate and incubate it for 48 hours at 27 °C.
- Detach the cells from the plate with pipette aid by using pressure and transfer all the cells to 50 ml centrifuge tube.
- Centrifuge the tube for 10 minutes at 3,000 rpm and discard supernatant media and store cell pellet at -80 °C if they are not used on the same day (The cell pellet volume would be ~200 to 250 µL).
- Resuspend the cells with 200 µL of hypotonic buffer (25 mM Tris pH 7.5, 25 mM sodium chloride, 0.5 mM TCEP) with protease inhibitors.
- Add 1.2 mL of CERI buffer (10 mM Tris pH 8.0, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM TCEP) to the resuspended cells, vortex for 5 sec and incubate on ice for 10 min.
- Add 66 µL of CERII buffer to the tube, vortex for 5 sec and incubate on ice for 1 min.
- Vortex for 5 sec again and centrifuge for 5 min at 13 krpm.
- Transfer the supernatant into a 2.0 mL microcentrifuge tube and place on ice (This is cytoplasmic extract, C).
- Add 600 µL of NER buffer (20 mM Tris pH 8.0, 1.5 mM MgCl2, 0.42 M sodium chloride, 20% glycerol, 0.5 mM TCEP) to the pellet.

- Vortex for 15 seconds to resuspend the pellet and incubate on ice for 10 minutes for 40 minutes (total 40 minutes).
- Centrifuge for 10 minutes at 13 krpm, transfer the supernatant into a 2.0 ml micro centrifuge tube conaining 1260 µL of NER buffer (This is nuclear extract, N).
- Resuspend the pellet in 600 µL of 1x SDS loading buffer (This is pellet, P).
- Prepare samples for gel loading by mixing 10 µL of C and N with 10 µL of 2x SDS loading buffer.
- Load 10 µL of C and N and 5 µL P onto SDS-PAGE and run the gel (usually, 60 min at 200 V with 15-18 % gel)
- The buffer amount used above is for 1 plate expression test and buffer amount should be in proportion to the number of plates.
- If protein of interest has affinity tag (histidine or Glutathione S-transferase), Ni- or GST-pull down can be performed.

1.6.2.4. Protein expression from Hi5 suspension cells

- 20 mL culture is usually used for expression test and 8 L culture for protein purification.
- 200 mL culture is used for expression test if cells need to be passed through EmulsiFlex-C3 homogenizer. (20 ml lysis buffer used)
- Prepare complete Sf-900 media with Sf-900 II SFM (Invitrogen) containing 100-fold diluted Penicillin-Streptomycin (Gibco), 1000-fold diluted Fungizone Antimycotic and 10 µg/ml Gentamicin.
- Count cells on hemocytometer and determine culture density.
- Amplify cells by adding complete Sf-900 media to make culture density about 0.7 x 10^6/mL and grow in suspension culture at 27 °C with shaking at 95 rpm.

- When the culture is 4 L with 4 x 10^6/mL, distribute each 0.5 L culture into 2.8 L flasks and add 20 mL of P3 virus (20 ml virus per 2 x 10^9 cells to the culture).
- Dilute the culture by adding same volume (0.5 L) of complete Sf-900 media and incubate for 48 hours at 27 °C with shaking at 95 rpm.
- Centrifuge for 10 minutes at 4,000 rpm (4,658 x g) at 4 °C in RC3BP+ centrifuge equipped with H6000A.
- Store cell pellet at 80 °C if they are not processed immediately.
- For small scale expression test, separate C, N and P and run them on SDS-PAGE as expression test from Hi5 monolayer cells.

1.6.3. large scale protein purification (8 L)

- Add 60 mL of lysis buffer (depends on proteins) with protease inhibitor to cells per 1 L culture and resuspend cells with 25 ml pipet.
- Lyse resuspended cells by passing through EmulsiFlex-C3 homogenizer (Avestin) twice and centrifuge at 38,000 rpm (192,500 x g) for 1 hour at 4 °C in Sorvall Discovery 90SE ultracentrifuge (Thermo Scientific) equipped with a F40L-8x100 rotor.
- Collect supernatant (S) and add 4 M imidazole to adjust imidazole concentration to 30 mM (pellet is insoluble fraction).
- Load the supernatant onto lysis buffer equilibrated HisTrap FF 5 mL (GE Healthcare) and wash the column with lysis buffer containing 30 mM Imidazole on FPLC.
- Run the column.
- Pool the fractions which have the proteins.

- Dialyze the pooled fractions in Spectrum Spectra/Por 7 Membrane Tubing (Spectrum labs) against 3.5 L dialysis buffer (25 mM Tris-HCl pH 8.8, 50 mM sodium chloride, 2 mM EDTA, 4 mM DTT (1,4-dithiothreitol)) for 3 hours at 4 °C.
- Centrifuge dialyzed sample for 10 min at 4,000 rpm (4,658 x g) at 4 °C in RC3BP+ centrifuge equipped with H6000A
- Load the supernatant onto equilibrated 24 mL Source Q column HR 16/10 (GE Healthcare) and run the column by increasing Sodium chloride concentration from 0 to 1M gradually.

Q Buffer A	25 mM Tris-HCl pH 8.8, 2 mM EDTA, 4 mM DTT
Q Buffer B	25 mM Tris-HCl pH 8.8, 1 M sodium chloride, 2 mM EDTA, 4 mM DTT
Flow rate	8 ml/min
Max pressure	4 MPa

-
- Check the location of protein peak with Bradford (Bio-Rad) by adding 10 µL of fraction to 100 µL Bradford. (brown color changes to blue in the presence of proteins.)
- Pool the fractions and measure the concentration with Bradford at 595 nm.
- Digest proteins with thrombin if affinity tag should be removed in the presence of 10 mM $CaCl_2$ and 10 mM DTT at 4 °C for overnight. (mass ratio between Thrombin and proteins varies)
- Dilute the sample by three to four folds with buffer A.
- Load the diluted sample onto equilibrated Source 15S column HR 10/10 (from GE Healthcare) and run the column by increasing Sodium chloride concentration from 0 to 1M gradually.

S Buffer A	50 mM HEPES pH 7.0, 2 mM EDTA, 4 mM DTT

S Buffer B	50 mM HEPES pH 7.0, 1 M sodium chloride, 2 mM EDTA, 4 mM DTT
Flow rate	4 mL/min
Max pressure	4 MPa

- Check the location of protein peak with Bradford as done with Source Q column and pull the fractions.
- Add 4 M sodium chloride to protein sample to make final sodium chloride concentration 800 mM.
- Concentrate the proteins using Centrifugal Amicon Ultra-15 centrifugal filter devices with proper MWCO (EMD Millipore) or using Amicon ultrafiltration unit with a YM membrane with proper MWCO.
- Inject the sample onto equilibrated Superdex 200 10/300 GL (from GE Healthcare) using 0.5 to 2 mL loop and run the column for 1.5 column volume which is 36 mL.

SD Buffer	5 mM BTP pH 6.8, 800 mM Sodium chloride, 5 mM DTT
Flow rate	0.3 – 0.5 ml/min
Max pressure	1.5 MPa

-
- Check the location of protein peak with Bradford as done with Source Q/S column and pull the fractions.
- Concentrate proteins with Amicon Ultra 0.5 ml centrifugal filters with proper MWCO (EMD Millipore), freeze with liquid Nitrogen and store at - 80 °C.

1.7. Reconstitution of histone octamer

1.7.1. Mixing H2A, H2B, H3.1 and H4

- Add 2.52 mg of H2A, 2.48 mg of H2B, 2.70 mg of H3.1 and 2 mg of H4 in 15 mL falcon tube.
- Add 6.47 mL unfolding buffer to make concentration of 1.5 mg/ml.
- Rock the histone at 4 °C for 1.5 h.

1.7.2. Dialysis and Concentration

- Dialyze histone mixture for 4 times (O/N, 4h, 4h, O/N) at 4 °C against 0.5 L of refolding buffer.
- Concentrate histone octamer with Amicon 30K centrifugal filters (Millipore) at 4000 g for 5 min several times until the volume is ~ 4 mL.

1.7.3. HiLoad Superdex 200 16/60

- V=120 mL.
- Max pressure = 0.5 MPa
- Flow rate 1 ml/min
- Loop volume = 5 mL
- 1.2 CV, 1.5 ml/fraction
- Equilibrate SD200 with 1.2 CV of degassed SD200 (refolding) buffer.
- Inject histones through a 5 mL loop.
- Pool the peak and concentration proteins to ~ 2-300 ul and check the concentration with UV spectrophotometer. (Ve = xx mL, 2nd peak is for H2A-H2B)
- (A280 - A320) / ε = Concentration (M)
- Molarity x MW (Da, g/mol) = Concentration (mg/mL)
- Total proteins (mg) = xx mg/mL x Total volume (mL)
- ~ 200 ul of 100-200 uM obtained

- Aliquot histone octamers with 30 uL.

1.8. Nucleosome core particle (NCP) reconstitution

1.8.1. Small scale set up

- Determine the ratio for DNA : Histone octamers
- Calculate the volume of DNA, # of moles of DNA and the volume of histone octamers for 80 ug of DNA.
- Set the reaction as 0.8 mg/mL for DNA concentration.
- Use 6-8 kDa MWCO spectra/POR1 membrane with 10mm flat widths for dialysis.
- Set the dialysis with perista counter clockwise. (takes ~ 33hrs)
- Dialyze against 400 mL RB Low buffer for 3 hrs and transfer dialyzed sample (~100 uL) to 1.5 mcf tubes at 4 °C.
- Calculate the concentration of DNA in nucleosome with UV spectrophotometer.
- Transfer 30 uL of each sample to PCR tubes and do heatshift (heat at 55 C for 2 hrs).
- Load 10 uL onto a 6 % native gel with 0.2 x TBE running buffer and run a gel at 150 V for 50 min at 4 °C.
- Check the gel and determine the best condition which shows best nucleosome band.

1.8.2. Large scale set up

- Calculate 250 ug of DNA and set the reaction as 0.8 mg/mL for DNA concentration.
- Repeat dialysis with the best condition for small scale.

1.8.3. Purification of NCPs

- Assemble PrepCell and make 6 % gel and degas it before adding Temed for 5-10 min (6.5-7 cm).

- Add isopropanol to make the gel even and add running buffer in the lower chamber (~2 L)
- Put the cell and try to remove all the bubbles in the bottom of the cell.
- Add running buffer in the upper inner chamber and 500 mL of elution buffer to the upper outer cell.
- Remove some elution buffer by using tubing and 20 mL syringe. (3 times)
- Connect PrepCell to the power and Perista (1.5 mL/min) and UV detector.
- Prerun the cell for 1 hour at 10 W (~400 V) at 4 °C .
- Load reconstituted NCP by 1mL syringe with a loop (green) and run for 1.5 hours at 4 °C.

- For 60 min → 6 min/fraction (9 mL/min)
- For last 30 min → 1min/fraction (1.5 mL/min)

- Check the peak position and run 6 % native gel and Pool the peak fractions and concentrate to ~ 100 uL.
- Determine the concentration of NCP at A260 and store at 4 °C.

2. Copyright Permissions

Cited Literature

1. Nooren, I.M. and J.M. Thornton, *Diversity of protein-protein interactions.* EMBO J, 2003. **22**(14): p. 3486-92.
2. Nie, Y., et al., *Getting a grip on complexes.* Curr Genomics, 2009. **10**(8): p. 558-72.
3. Lee, C.H., *A Simple Outline of Methods for Protein Isolation and Purification.* Endocrinol Metab (Seoul), 2017. **32**(1): p. 18-22.
4. Ferrer-Miralles, N., et al., *General introduction: recombinant protein production and purification of insoluble proteins.* Methods Mol Biol, 2015. **1258**: p. 1-24.
5. Sevastsyanovich, Y., S. Alfasi, and J. Cole, *Recombinant protein production: a comparative view on host physiology.* N Biotechnol, 2009. **25**(4): p. 175-80.
6. Basquin, J., M. Taschner, and E. Lorentzen, *Complex Reconstitution from Individual Protein Modules.* Adv Exp Med Biol, 2016. **896**: p. 305-14.
7. Selleck, W. and S. Tan, *Recombinant protein complex expression in E. coli.* Curr Protoc Protein Sci, 2008. **Chapter 5**: p. Unit 5 21.
8. Nie, Y., et al., *Multiprotein complex production in insect cells by using polyproteins.* Methods Mol Biol, 2014. **1091**: p. 131-41.
9. Alberts, B., et al., *Molecular Biology of the Cell 4th Edition: International Student Edition*. 2002: Routledge.
10. Hoeijmakers, J.H., *DNA damage, aging, and cancer.* N Engl J Med, 2009. **361**(15): p. 1475-85.
11. Swenberg, J.A., et al., *Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment.* Toxicol Sci, 2011. **120 Suppl 1**: p. S130-45.
12. Loeb, L.A., *Mutator Phenotype*, in *Encyclopedia of Cancer*, M. Schwab, Editor. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 2416-2419.
13. Wirth, N., et al., *Conservation and Divergence in Nucleotide Excision Repair Lesion Recognition.* J Biol Chem, 2016. **291**(36): p. 18932-46.
14. Gillet, L.C. and O.D. Scharer, *Molecular mechanisms of mammalian global genome nucleotide excision repair.* Chem Rev, 2006. **106**(2): p. 253-76.
15. Hanawalt, P.C. and G. Spivak, *Transcription-coupled DNA repair: two decades of progress and surprises.* Nat Rev Mol Cell Biol, 2008. **9**(12): p. 958-70.
16. Drapkin, R., et al., *Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II.* Nature, 1994. **368**(6473): p. 769-72.
17. Terpe, K., *Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2006. **72**(2): p. 211-22.
18. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Front Microbiol, 2014. **5**: p. 172.
19. Sezonov, G., D. Joseleau-Petit, and R. D'Ari, *Escherichia coli physiology in Luria-Bertani broth.* J Bacteriol, 2007. **189**(23): p. 8746-9.
20. Gopal, G.J. and A. Kumar, *Strategies for the production of recombinant protein in Escherichia coli.* Protein J, 2013. **32**(6): p. 419-25.
21. !!! INVALID CITATION !!! {Busso, 011 #115}.
22. Luger, K., et al., *Characterization of nucleosome core particles containing histone proteins made in bacteria.* J Mol Biol, 1997. **272**(3): p. 301-11.

23. Rudolph, R. and H. Lilie, *In vitro folding of inclusion body proteins.* FASEB J, 1996. **10**(1): p. 49-56.
24. Kerrigan, J.J., et al., *Production of protein complexes via co-expression.* Protein Expr Purif, 2011. **75**(1): p. 1-14.
25. Yasukawa, T., et al., *Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin.* J Biol Chem, 1995. **270**(43): p. 25328-31.
26. Tan, S., *A modular polycistronic expression system for overexpressing protein complexes in Escherichia coli.* Protein Expr Purif, 2001. **21**(1): p. 224-34.
27. Tan, S., R.C. Kern, and W. Selleck, *The pST44 polycistronic expression system for producing protein complexes in Escherichia coli.* Protein Expr Purif, 2005. **40**(2): p. 385-95.
28. Tolia, N.H. and L. Joshua-Tor, *Strategies for protein coexpression in Escherichia coli.* Nat Methods, 2006. **3**(1): p. 55-64.
29. van Holde, K.E., *Chromatin*. 2012: Springer New York.
30. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
31. Campos, E.I. and D. Reinberg, *Histones: annotating chromatin.* Annu Rev Genet, 2009. **43**: p. 559-99.
32. Schnitzler, G.R., *Isolation of histones and nucleosome cores from mammalian cells.* Curr Protoc Mol Biol, 2001. **Chapter 21**: p. Unit 21 5.
33. Shechter, D., et al., *Extraction, purification and analysis of histones.* Nat Protoc, 2007. **2**(6): p. 1445-57.
34. Luger, K., T.J. Rechsteiner, and T.J. Richmond, *Expression and purification of recombinant histones and nucleosome reconstitution.* Methods Mol Biol, 1999. **119**: p. 1-16.
35. Dyer, P.N., et al., *Reconstitution of nucleosome core particles from recombinant histones and DNA.* Methods Enzymol, 2004. **375**: p. 23-44.
36. Shim, Y., et al., *Polycistronic coexpression and nondenaturing purification of histone octamers.* Anal Biochem, 2012. **427**(2): p. 190-2.
37. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
38. McGhee, J.D. and G. Felsenfeld, *Nucleosome structure.* Annu Rev Biochem, 1980. **49**: p. 1115-56.
39. Simpson, R.T., *Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones.* Biochemistry, 1978. **17**(25): p. 5524-31.
40. Fischle, W., Y. Wang, and C.D. Allis, *Histone and chromatin cross-talk.* Curr Opin Cell Biol, 2003. **15**(2): p. 172-83.
41. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.
42. Ehrenhofer-Murray, A.E., *Chromatin dynamics at DNA replication, transcription and repair.* Eur J Biochem, 2004. **271**(12): p. 2335-49.
43. Groth, A., et al., *Chromatin challenges during DNA replication and repair.* Cell, 2007. **128**(4): p. 721-33.
44. Sambrook, J. and D.W. Russell, *Molecular Cloning: A Laboratory Manual*. 2001: Cold Spring Harbor Laboratory Press.

45. Luger, K., T.J. Rechsteiner, and T.J. Richmond, *Preparation of nucleosome core particle from recombinant histones.* Methods Enzymol, 1999. **304**: p. 3-19.
46. Duan, M.R. and M.J. Smerdon, *UV damage in DNA promotes nucleosome unwrapping.* J Biol Chem, 2010. **285**(34): p. 26295-303.
47. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning.* J Mol Biol, 1998. **276**(1): p. 19-42.
48. Makde, R.D., et al., *Structure of RCC1 chromatin factor bound to the nucleosome core particle.* Nature, 2010. **467**(7315): p. 562-6.
49. Li, G. and J. Widom, *Nucleosomes facilitate their own invasion.* Nat Struct Mol Biol, 2004. **11**(8): p. 763-9.
50. Weintraub, H., K. Palter, and F. Van Lente, *Histones H2a, H2b, H3, and H4 form a tetrameric complex in solutions of high salt.* Cell, 1975. **6**(1): p. 85-110.
51. Anderson, M., et al., *Co-expression as a convenient method for the production and purification of core histones in bacteria.* Protein Expr Purif, 2010. **72**(2): p. 194-204.
52. Perales, R., L. Zhang, and D. Bentley, *Histone occupancy in vivo at the 601 nucleosome binding element is determined by transcriptional history.* Mol Cell Biol, 2011. **31**(16): p. 3485-96.
53. Vasudevan, D., E.Y. Chua, and C.A. Davey, *Crystal structures of nucleosome core particles containing the '601' strong positioning sequence.* J Mol Biol, 2010. **403**(1): p. 1-10.
54. Hoch, D.A., J.J. Stratton, and L.M. Gloss, *Protein-protein Forster resonance energy transfer analysis of nucleosome core particles containing H2A and H2A.Z.* J Mol Biol, 2007. **371**(4): p. 971-88.
55. Waye, J.S. and H.F. Willard, *Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome.* Mol Cell Biol, 1986. **6**(9): p. 3156-65.
56. Fittler, F. and H.G. Zachau, *Subunit structure of alpha-satellite DNA containing chromatin from African green monkey cells.* Nucleic Acids Res, 1979. **7**(1): p. 1-13.
57. Waye, J.S. and H.F. Willard, *Chromosome-specific alpha satellite DNA: nucleotide sequence analysis of the 2.0 kilobasepair repeat from the human X chromosome.* Nucleic Acids Res, 1985. **13**(8): p. 2731-43.
58. Kujirai, T., et al., *Structure and function of human histone H3.Y nucleosome.* Nucleic Acids Res, 2016. **44**(13): p. 6127-41.
59. Klinker, H., et al., *Rapid purification of recombinant histones.* PLoS One, 2014. **9**(8): p. e104029.
60. Tachiwana, H., et al., *Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T.* Proc Natl Acad Sci U S A, 2010. **107**(23): p. 10454-9.
61. Demain, A.L. and P. Vaishnav, *Production of recombinant proteins by microbes and higher organisms.* Biotechnol Adv, 2009. **27**(3): p. 297-306.
62. Sahdev, S., S.K. Khattar, and K.S. Saini, *Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies.* Mol Cell Biochem, 2008. **307**(1-2): p. 249-64.

63. Baneyx, F. and M. Mujacic, *Recombinant protein folding and misfolding in Escherichia coli.* Nat Biotechnol, 2004. **22**(11): p. 1399-408.
64. Saibil, H., *Chaperone machines for protein folding, unfolding and disaggregation.* Nat Rev Mol Cell Biol, 2013. **14**(10): p. 630-42.
65. Unger, T. and Y. Peleg, *Recombinant protein expression in the baculovirus-infected insect cell system.* Methods Mol Biol, 2012. **800**: p. 187-99.
66. Ikonomou, L., Y.J. Schneider, and S.N. Agathos, *Insect cell culture for industrial production of recombinant proteins.* Appl Microbiol Biotechnol, 2003. **62**(1): p. 1-20.
67. Berger, I., D.J. Fitzgerald, and T.J. Richmond, *Baculovirus expression system for heterologous multiprotein complexes.* Nat Biotechnol, 2004. **22**(12): p. 1583-7.
68. Aucoin, M.G., M. Perrier, and A.A. Kamen, *Production of adeno-associated viral vectors in insect cells using triple infection: optimization of baculovirus concentration ratios.* Biotechnol Bioeng, 2006. **95**(6): p. 1081-92.
69. Fitzgerald, D.J., et al., *Protein complex expression by using multigene baculoviral vectors.* Nat Methods, 2006. **3**(12): p. 1021-32.
70. Scholz, J. and S. Suppmann, *A new single-step protocol for rapid baculovirus-driven protein production in insect cells.* BMC Biotechnol, 2017. **17**(1): p. 83.
71. Sari, D., et al., *The MultiBac Baculovirus/Insect Cell Expression Vector System for Producing Complex Protein Biologics.* Adv Exp Med Biol, 2016. **896**: p. 199-215.
72. Gerard, M., et al., *Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2.* J Biol Chem, 1991. **266**(31): p. 20940-5.
73. Iben, S., et al., *TFIIH plays an essential role in RNA polymerase I transcription.* Cell, 2002. **109**(3): p. 297-306.
74. Feaver, W.J., et al., *Dual roles of a multiprotein complex from S. cerevisiae in transcription and DNA repair.* Cell, 1993. **75**(7): p. 1379-87.
75. Schaeffer, L., et al., *DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor.* Science, 1993. **260**(5104): p. 58-63.
76. Maldonado, E., R. Drapkin, and D. Reinberg, *Purification of human RNA polymerase II and general transcription factors.* Methods Enzymol, 1996. **274**: p. 72-100.
77. Grunberg, S. and S. Hahn, *Structural insights into transcription initiation by RNA polymerase II.* Trends Biochem Sci, 2013. **38**(12): p. 603-11.
78. Svejstrup, J.Q., et al., *Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome.* Cell, 1995. **80**(1): p. 21-8.
79. Egly, J.M. and F. Coin, *A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor.* DNA Repair (Amst), 2011. **10**(7): p. 714-21.
80. Roy, R., et al., *The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor.* Cell, 1994. **79**(6): p. 1093-101.
81. Solomon, M.J., J.W. Harper, and J. Shuttleworth, *CAK, the p34cdc2 activating kinase, contains a protein identical or closely related to p40MO15.* EMBO J, 1993. **12**(8): p. 3133-42.
82. Aprelikova, O., Y. Xiong, and E.T. Liu, *Both p16 and p21 families of cyclin-dependent kinase (CDK) inhibitors block the phosphorylation of cyclin-dependent kinases by the CDK-activating kinase.* J Biol Chem, 1995. **270**(31): p. 18195-7.
83. Kaldis, P., *The CDK-Activating Kinase (CAK).* 2003: Springer US.
84. Bedez, F., et al., *Functional insights into the core-TFIIH from a comparative survey.* Genomics, 2013. **101**(3): p. 178-86.

85. Buratowski, S., *The basics of basal transcription by RNA polymerase II.* Cell, 1994. **77**(1): p. 1-3.
86. Luse, D.S., *The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled?* Transcription, 2014. **5**(1): p. e27050.
87. Maxon, M.E., J.A. Goodrich, and R. Tjian, *Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance.* Genes Dev, 1994. **8**(5): p. 515-24.
88. Fiedler, U. and H.T. Marc Timmers, *Peeling by binding or twisting by cranking: models for promoter opening and transcription initiation by RNA polymerase II.* Bioessays, 2000. **22**(4): p. 316-26.
89. Tirode, F., et al., *Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7.* Mol Cell, 1999. **3**(1): p. 87-95.
90. Fishburn, J., et al., *Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation.* Proc Natl Acad Sci U S A, 2015. **112**(13): p. 3961-6.
91. Wong, K.H., Y. Jin, and K. Struhl, *TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape.* Mol Cell, 2014. **54**(4): p. 601-12.
92. Cheung, A.C. and P. Cramer, *A movie of RNA polymerase II transcription.* Cell, 2012. **149**(7): p. 1431-7.
93. Coin, F., et al., *Nucleotide excision repair driven by the dissociation of CAK from TFIIH.* Mol Cell, 2008. **31**(1): p. 9-20.
94. Scharer, O.D., *Nucleotide excision repair in eukaryotes.* Cold Spring Harb Perspect Biol, 2013. **5**(10): p. a012609.
95. Coin, F., V. Oksenych, and J.M. Egly, *Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair.* Mol Cell, 2007. **26**(2): p. 245-56.
96. Oksenych, V., et al., *Molecular insights into the recruitment of TFIIH to sites of DNA damage.* EMBO J, 2009. **28**(19): p. 2971-80.
97. Winkler, G.S., et al., *TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair.* J Biol Chem, 2000. **275**(6): p. 4258-66.
98. Fregoso, M., et al., *DNA repair and transcriptional deficiencies caused by mutations in the Drosophila p52 subunit of TFIIH generate developmental defects and chromosome fragility.* Mol Cell Biol, 2007. **27**(10): p. 3640-50.
99. Volker, M., et al., *Sequential assembly of the nucleotide excision repair factors in vivo.* Mol Cell, 2001. **8**(1): p. 213-24.
100. de Boer, J. and J.H. Hoeijmakers, *Nucleotide excision repair and human syndromes.* Carcinogenesis, 2000. **21**(3): p. 453-60.
101. Kraemer, K.H., et al., *Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship.* Neuroscience, 2007. **145**(4): p. 1388-96.
102. Compe, E. and J.M. Egly, *TFIIH: when transcription met DNA repair.* Nat Rev Mol Cell Biol, 2012. **13**(6): p. 343-54.

103. Coin, F., et al., *Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH.* Nat Genet, 1998. **20**(2): p. 184-8.
104. Fan, L., et al., *XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations.* Cell, 2008. **133**(5): p. 789-800.
105. Coin, F., et al., *p8/TTD-A as a repair-specific TFIIH subunit.* Mol Cell, 2006. **21**(2): p. 215-26.
106. Weeda, G., et al., *A mutation in the XPB/ERCC3 DNA repair transcription gene, associated with trichothiodystrophy.* Am J Hum Genet, 1997. **60**(2): p. 320-9.
107. Giglia-Mari, G., et al., *A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A.* Nat Genet, 2004. **36**(7): p. 714-9.
108. Knuesel, M.T., et al., *The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function.* Genes Dev, 2009. **23**(4): p. 439-51.
109. Min, J.H. and N.P. Pavletich, *Recognition of DNA damage by the Rad4 nucleotide excision repair protein.* Nature, 2007. **449**(7162): p. 570-5.
110. Bardwell, A.J., et al., *Yeast nucleotide excision repair proteins Rad2 and Rad4 interact with RNA polymerase II basal transcription factor b (TFIIH).* Mol Cell Biol, 1994. **14**(6): p. 3569-76.
111. Uchida, A., et al., *The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH.* DNA Repair (Amst), 2002. **1**(6): p. 449-61.
112. Yokoi, M., et al., *The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA.* J Biol Chem, 2000. **275**(13): p. 9870-5.
113. Okuda, M. and Y. Nishimura, *Dataset for the NMR structure of the intrinsically disordered acidic region of XPC bound to the PH domain of TFIIH p62.* Data Brief, 2016. **6**: p. 571-7.
114. Bernardes de Jesus, B.M., et al., *Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC.* Mol Cell Biol, 2008. **28**(23): p. 7225-35.
115. Luo, J., et al., *Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH.* Mol Cell, 2015. **59**(5): p. 794-806.

VITA

EUDATION

PhD in Biochemistry
University of Illinois at Chicago, Illinois, USA 2010- 2018
B.S. in Chemistry, University of Illinois at Chicago, Illinois, USA 2008- 2010
Oakton Community College 2004-2008

PUBLICATIONS

1. Base and nucleotide excision repair of oxidatively generated guanine lesions in DNA. Vladimir Shafirovich, Konstantin Kropachev, Thomas Anderson, Zhi Liu, Marina Kolbanovskiy, Brooke D Martin, Kent Sugden, Yoonjung Shim, Xuejing Chen, Jung-Hyun Min, Nicholas E Geacintov
Journal of Biological Chemistry **2016**, 291, 5309-5319

2. Kinetic gating mechanism of DNA damage recognition by Rad4/XPC
Chen, X., Velmurugu, Y., Zheng, G., Park, B., Shim, Y., Kim, Y., Liu, L., Van Houten, B., He, C., Ansari, A., & Min, J.H., Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nature Communications* 6, 5849 (**2015**)

3. Characterization of the histone methyltransferase PRDM9 utilising biochemical, biophysical and chemical biology techniques
Koh-Stenta, X., Joy, J., Poulsen, A., Li, R., Tan, Y., Shim, Y., Min, J.H., Wu, L., Ngo, A., Peng, J., Seetoh, W.G., Cao, J., Wee, J.L., Kwek, P.Z., Hung, A., Lakshmanan, U., Flotow, H., Guccione, E., & Hill, J., *Biochemical Journal* 461 (2), 323-334 (**2014**)

4. Polycistronic coexpression and nondenaturing purification of histone octamers
Shim, Y., Duan, M.R., Chen, X., Smerdon, M.J., & Min, J.H., Polycistronic coexpression and nondenaturing purification of histone octamers. *Analytical Biochemistry* 427 (2), 190-192 (**2012**)

POSTER PRESENTATIONS (CONFERENCE)

1. Polycistronic coexpression and nondenaturing purification of histone octamers, 14th Annual Midwest DNA Repair Symposium, Cincinnati, OH (Poster presentation)

TEACHING EXPERIENCE

1. General Chemistry I (laboratory and discussion), General Chemistry II (laboratory and discussion), Biochemistry laboratory 2010-2018, UIC, USA
2. Supervised undergraduates for research:

- Two undergraduate students won CURA (2012) and LASURI (2012),
- Two undergraduate students went to medical school (2011, 2012)
- All students are Honors students