

**Crystallographic and Functional Studies on the
central domain of *Drosophila* Dribble**

CHENG, Tat Cheung

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in
Molecular Biotechnology

The Chinese University of Hong Kong
March 2011

UMI Number: 3492043

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent on the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3492043

Copyright 2011 by ProQuest LLC.

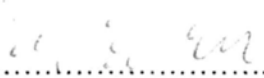
All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

DECLARATION

I declare that this dissertation is entirely my own work except where otherwise stated, either in the form of citation of published work, or acknowledgement of the source of any unpublished material.

Signed.....

Cheng, Tat Cheung

Thesis Committee

Professor WONG, Kam Bo (Chair)

Professor CHAN, Ho Yin Edwin (Thesis Supervisor)

Professor NGAI, Sai Ming (Thesis Supervisor)

Professor KWAN, Kin Ming (Committee Member)

Dr. SANDERSON, Mark (External Examiner)

Abstract

Eukaryotic ribosome biogenesis is mediated via the pre-ribosomal ribonucleoprotein (pre-rRNP) complexes. The *trans*-factors, consisting of ~200 proteins and small nucleolar ribonucleoproteins (snoRNPs), are important structural components of the pre-rRNP complexes, such that depletion of one *trans*-factor is enough to halt ribosome biogenesis and is therefore lethal to the organism. Dribble (DBE) family is a class of essential and evolutionarily conserved protein *trans*-factor, and it carries an uncharacterized central domain. The *Drosophila* DBE and the yeast member (Krr1p) of the family were previously found to be directly and genetically interacting with ribosomal protein S9 (RpS9) respectively, suggesting that DBE may function in recruiting ribosomal proteins to ribosome. This study focuses on structural determination of the uncharacterized central domain of DBE by X-ray crystallography, and function characterization of the central domain of DBE in terms of RpS9 binding. The DBE protein carries a central domain that is flanked by extensive structural disorders. Deletion mutants of DBE were designed to facilitate protein crystallization, and the DBE₁₆₋₁₉₇ protein can be expressed, purified and crystallized. However, the DBE₁₆₋₁₉₇ crystal did not lead to structural determination due to certain technical difficulties, and the yeast and human correspondents of DBE₁₆₋₁₉₇, i.e. Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ could also be crystallized, but they did not show diffraction. In short, the potential DBE domain as represented by DBE₁₆₋₁₉₇, Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ could be crystallized but they did not lead to structural determination. Strategies for additional search for crystal form are proposed. On the other hand, the RpS9 binding region was located within the residue 34 to 229 of DBE, as mapped by glutathione S-transferase (GST)-RpS9 pull down assay, suggesting RpS9 affinity might be partly contributed by the DBE domain.

Abstract (Chinese version)

真核細胞透過前核糖體核糖核蛋白 (pre-ribosomal ribonucleoprotein) 進行核糖體合成 (ribosome biogenesis)。這些前核糖體核糖核蛋白則由為數大約200個的跨因子 (*trans*-factors)組成，這些跨因子或者是蛋白，或者是細小核仁核糖核蛋白 (small nucleolar ribonucleoprotein)。這些跨因子是前核糖體核糖核蛋白的重要結構元件，除去一個跨因子足以癱瘓核糖體合成並引致生物死亡。Dribble (DBE) 是個必不可少並且進化上保守的蛋白跨因子家族，並且這家族有個結構未知的結構域(domain)。果蠅內的DBE蛋白發現與核糖蛋白S9有直接連接，並且酵母菌內的DBE家族成員Krr1p與核糖蛋白S9 有遺傳作用，這意味著DBE蛋白家族可能幫助核糖體收錄核糖蛋白。是項研究集中於使用X射線結晶學的方法解析DBE結構，並致力於了解DBE如何與核糖蛋白S9連接。DBE蛋白中央部份有一個結構域，而結構域兩邊帶有大量結構散亂的位置。為了促進蛋白結晶，是項研究設計了數個DBE移除突變體 (deletion mutant)，其中DBE₁₆₋₁₉₇ 蛋白能被表達，純化並結晶。基於某些技術困難，DBE₁₆₋₁₉₇晶體結構不能被解出。另外，酵母菌與人類的DBE₁₆₋₁₉₇對應物Krr1p₁₃₋₁₉₆蛋白與HRB2₂₈₋₂₁₅蛋白皆可結晶，不過它們都沒有產生衍射圖案。簡單的說，DBE₁₆₋₁₉₇，Krr1p₁₃₋₁₉₆與HRB2₂₈₋₂₁₅蛋白皆可結晶，但卻不能解出結構。因此，是項研究提出數項搜尋另外結晶狀態(crystal form)的策略建議。另一方面，穀胱甘肽S-轉移酶沉聚 (glutathione S-transferase pull down) 實驗顯示DBE通過第34粒至229粒胺基酸殘基中間的斷落連接核蛋白S9，顯示DBE結構域可能部份參與RpS9的連接。

ACKNOWLEDGEMENTS

My deepest gratitude goes to my supervisor Prof. Edwin H. Y. Chan for his generous support and the freedom he gave me while undertaking the project.

I am sincerely grateful to Prof. K.B. Wong for his help in designing DBE mutants; Dr. Yu Wai Chen for his help during my visit to Diamond Light Source; Prof. Jacky Ngo and Prof. Terence Lau for their advice in crystallography; Dr. Thomas Terwilliger for his advice in tackling technical issues in crystallography; Dr. Siu Hong Chan for his help in N-terminal sequencing.

I would like to thank all members of the Laboratory of *Drosophila* Research, CUHK, especially Dr. Priscilla Chan, Mr. Ho Choi, Mr. Chi Chung Wu, Mr. Yat Pang Cheng, Mr. Hon Wing Luk, Ms. Teresa Fung, Ms. Aurora Chan and Ms. Azaria Wong for their help and suggestions.

I would also like to thank all the people in Prof. K.B. Wong's lab including Dr. Ka Ming Lee, Mr. Eric Chan, Mr. Kwok Ho Chan, Mr. Ivan Fong, Mr. Chi Pang Chuk, Ms. Connie Yu and Ms. Tsz Ha Yu for their help in protein crystallography.

LIST OF ABBREVIATIONS

Abbreviation used Term

APS	Ammonium persulfate
BSA	Bovine serum albumin
CD	Circular dichroism
cDNA	Complementary deoxyribonucleic acid
DAS	Diffraction Anisotropic Server
DBE	Dribble
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxynucleoside triphosphates
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediamine tetraacetic acid
EM	Electron microscopy
ETS	External transcribed spacer
GST	Glutathione-S-transferase
HRB2	HIV-1 rev-binding protein
HS	Poly histidine and SUMO
IPTG	Isopropyl β -D-thiogalactopyranoside
ITS	Internal transcribed spacer
kDa	KiloDalton
KH domain	K homology domain
LB	Luria-Bertani medium
LBA	LB with ampicillin
LBAC	LB with ampicillin and chloramphenicol
LBAK	LB with ampicillin and kanamycin
LBC	LB with chloramphenicol
LBK	LB with kanamycin
MAD	Multi-wavelength anomalous dispersion
MIR	Multiple isomorphous replacement
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PMSF	Phenylmethanesulfonyl fluoride

Pre-rRNA	Ribosomal ribonucleic acid precursor
Pre-rRNP	Ribosomal ribonucleoprotein precursor
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	revolutions per minute
rprotein	Ribosomal protein
RpS9	Ribosomal protein S9
SAD	Single anomalous dispersion
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEN1c	SUMO-specific protease 1 mutant C603S
SERp	Surface Entropy Reduction prediction
SIRAS	Single isomorphous replacement with anomalous scattering
smFRET	single molecular fluorescence resonance energy transfer
snoRNA	Small nucleolar RNA
snoRNP	Small nucleolar RNA-containing ribonucleoprotein particle
SSRF	Shanghai synchrotron radiation facility
SSU	Small subunit
SUMO	Small ubiquitin-related modifier
TBE	Tris-borate-EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	N',N',N',N'-Di-(dimethylamino)ethane
Tris-HCl	Tris-hydrochloric acid

LIST OF TABLES

- Table 2.1** Primers used in the PCR generation of DBE deletion mutants
- Table 2.2** Expression vector and their corresponding expression conditions
- Table 2.3** Heavy atom compounds used in the JBScreen
- Table 4.1** Data collection and diffraction statistics of native DBE₁₆₋₁₉₇ crystal
- Table 4.2** Data collection and diffraction statistics of native DBE₁₆₋₁₉₇ crystal for sulfur SAD
- Table 4.3** Fifteen heavy metal atoms used for deriving heavy atom substituted DBE₁₆₋₁₉₇ crystals.
- Table 4.4** Data collection and diffraction statistics of a Pt K₂Cl₄ substituted DBE₁₆₋₁₉₇ crystal
- Table 4.5** Data collection and diffraction statistics of an iodide substituted DBE₁₆₋₁₉₇ crystal
- Table 4.6** Data collection and diffraction statistics of a DBE₁₆₋₁₉₇ crystal grown in the presence of I3C.
- Table 4.7** Data collection and diffraction statistics of a selenomethionyl DBE₁₆₋₁₉₇ crystal
- Table 5.1** Initial crystallization conditions of Krr1p₁₃₋₁₉₆
- Table 5.2** Initial crystallization conditions of HRB2₂₈₋₂₁₅

LIST OF FIGURES

- Figure 1.1** Schematic representation of ribosome assembly
- Figure 1.2 (A)** Schematic representation of pre-rRNA cleavage in yeast **(B)** Schematic representation of pre-rRNA cleavage in *Drosophila*
- Figure 1.3** A model for 90S pre-ribosome functioning in yeast
- Figure 1.4 (A)** Far-UV CD spectrum of DBE at 20, 55, and 110 °C **(B)** Thermal denaturation profile of DBE as monitored by CD spectroscopy at 222 nm
- Figure 3.1** Protein sequence alignment of DBE family
- Figure 3.2** Graphic representation of disorder prediction of DBE by *Disprot VL3H*
- Figure 3.3** Schematic illustration of the design of 9 DBE deletion mutants
- Figure 3.4 (A)** Absence of obvious protein expression of tag-less DBE₂₆₋₁₉₇, DBE₂₆₋₂₁₃ and DBE₂₆₋₂₂₉ **(B)** The inability of column binding of the DBE₃₄₋₁₉₇ protein
- Figure 3.5** Comparison of the expression between the DBE₁₆₋₁₉₇, DBE₁₆₋₂₁₃ and DBE₁₆₋₂₂₉ proteins.
- Figure 3.6** Purification profile of tag-less DBE₁₆₋₁₉₇
- Figure 3.7 (A)** Absence of obvious expression of tag-less DBE₂₄₋₁₉₇ **(B)** Low expression level of His-FXa-DBE₂₄₋₁₉₇ **(C)** Inability to separate the HS tag from DBE₂₄₋₁₉₇
- Figure 3.8** Purification profile of DBE₁₆₋₁₉₇ with the aid of HS tag
- Figure 3.9 (A)** Far-UV CD spectrum of DBE₁₆₋₁₉₇ **(B)** Thermal denaturation profile of DBE₁₆₋₁₉₇
- Figure 3.10 (A)** Elution profile of DBE₁₆₋₁₉₇ and molar mass estimation by static light scattering **(B)** Dynamic light scattering analysis of DBE₁₆₋₁₉₇
- Figure 4.1** Images of DBE₁₆₋₁₉₇ crystals
- Figure 4.2** Optimization of the expression of selenomethionyl HS-DBE₁₆₋₁₉₇
- Figure 4.3** Diffraction anisotropy of the DBE₁₆₋₁₉₇ crystal
- Figure 5.1** Expression and purification profile of HS-DBE₁₆₋₁₉₇
- Figure 5.2 (A)** Expression optimization of HS-DBE₁₆₋₂₂₉ **(B)** Purification profile of DBE₁₆₋₂₂₉ with the aid of HS tag
- Figure 5.3** Design of the DBE₁₆₋₁₉₇ mutants
- Figure 5.4** Purification profiles of DBE₁₆₋₁₉₇ mutants R87K, K97R, K138R, QE29A, EE38A and KKE135A.
- Figure 5.5 (A)** Protein sequence alignment of DBE, yeast homolog Krr1p and human homolog HRB2 **(B)** Expression and purification scheme of Krr1p₁₃₋₁₉₆ and **(C)** HRB2₂₈₋₂₁₅ with the aid of HS tag
- Figure 5.6** Images of Krr1p₁₃₋₁₉₆ crystals

Figure 5.7 Images of HRB2₂₈₋₂₁₅ crystals

Figure 5.8 Schematic representation of the experimental setup for crystal cross-linking

Figure 5.9 Possible positions for DBE₁₆₋₁₉₇ point mutants

Figure 6.1 (A) Protein inputs for GST-RpS9 pull down experiments **(B)** GST-RpS9 of DBE deletion mutants including DBE, DBE₁₋₂₂₉, DBE₃₄₋₂₂₉, DBE₁₆₋₁₉₇ and DBE₃₄₋₁₉₇

Figure 6.2 Schematic representation of the GST-RpS9 pull down result

Figure 6.3 Expression of GST-RpS9

Figure 6.4 (A) Expression of tag-less RpS9 **(B)** Expression of HIS tagged RpS9

Figure 6.5 Co-expression of DBE₁₆₋₂₂₉ with RpS9/HS-RpS9

TABLE OF CONTENTS

	Page
ABSTRACT	I
ABSTRACT (Chinese Version)	II
ACKNOWLEDGEMENTS	III
LIST OF ABBREVIATIONS	IV
LIST OF TABLES	VI
LIST OF FIGURES	VII
1. INTRODUCTION	1
1.1 Introduction of ribosome biogenesis	1
1.1.1 Textbook knowledge and frontiers in ribosome research	1
1.1.2 Ribosome biogenesis and diseases	3
1.1.3 Three stakeholders in ribosome biogenesis	4
1.1.4 Two central processes in ribosome biogenesis	7
1.1.4.1 Ribosome assembly	7
1.1.4.2 Ribosomal RNA processing	10
1.1.5 Ribosome biogenesis proceeds from nucleolus to cytoplasm	13
1.1.6 The 90S pre-ribosome is an early pre-rRNP	13
1.2 The Dribble (DBE) family	16
1.2.1 Dribble family is essential for rRNA processing and ribosome assembly	16
1.2.2 Dribble family's function can be understood via pre-rRNP	17
1.2.3 A structured region is possibly hidden in DBE	18
1.3 Aim of research	21

2. MATERIALS AND METHODS	24
2.1 General techniques	24
2.1.1 Preparation of <i>E. coli</i> competent cells	24
2.1.2 Transformation of <i>E. coli</i> competent cells	25
2.1.3 Agarose gel electrophoresis of DNA	25
2.1.4 Polymerase chain reaction (PCR)	26
2.1.4.1 Condition of PCR	26
2.1.5 Restriction digestion of DNA	27
2.1.6 Ligation of DNA fragments	28
2.1.7 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)	28
2.2 Construction of expression vectors	31
2.2.1 Construction of <i>pET-DBE</i> deletion mutants	31
2.2.2 Construction of <i>pET-Fx-DBE₁₆₋₁₉₇</i> and <i>pET-Fx-DBE₂₄₋₁₉₇</i>	33
2.2.3 Construction of <i>pET-HS-DBE₁₆₋₁₉₇</i> , <i>pET-HS-DBE₂₄₋₁₉₇</i> and <i>pET-HS-DBE₁₆₋₂₂₉</i>	34
2.2.4 Construction of <i>pET-HS-DBE₁₆₋₁₉₇</i> point mutants	36
2.2.5 Construction of <i>pET-HS-Krr1p₁₃₋₁₉₆</i>	38
2.2.6 Construction of <i>pET-HS-HRB₂₈₋₂₁₅</i>	40
2.2.7 Construction of <i>pET-RpS9</i> and <i>pET-HS-RpS9</i>	41
2.2.8 Construction of <i>pCola-DBE₁₆₋₂₂₉</i>	42
2.3 Protein expression	42
2.3.1 Native protein expression	42
2.3.2 Expression of selenomethionyl DBE ₁₆₋₁₉₇	44
2.3.3 GST and GST-RpS9	45
2.4 Protein purification	46
2.4.1 Purification of SENP1c	46

2.4.2 Purification of tag-less DBE and DBE deletion mutants	47
2.4.3 Purification of HS tagged proteins	49
2.5 Biophysical characterization of DBE₁₆₋₁₉₇	51
2.5.1 Circular Dichroism (CD) spectroscopy	51
2.5.2 Dynamic light scattering	52
2.5.3 Static light scattering	52
2.6 Protein crystallization condition screening	53
2.6.1 Screening for crystallization condition of DBE ₁₆₋₁₉₇	53
2.6.2 Screening for crystallization condition of DBE ₁₆₋₂₂₉ , HS-DBE ₁₆₋₁₉₇ , DBE ₁₆₋₁₉₇	54
2.6.3 Screening for crystallization condition of Krr1p ₁₃₋₁₉₆ and HRB2 ₂₈₋₂₁₅	55
2.7 X-ray diffraction and phasing	57
2.7.1 DBE ₁₆₋₁₉₇ crystals	57
2.7.1.1 Native DBE ₁₆₋₁₉₇ crystals	57
2.7.1.2 Heavy atom soaking	58
2.7.1.3 5-amino-2,4,6-triiodoisophthalic acid (I3C) soaking	59
2.7.1.4 Iodide/ bromide quick soak	60
2.7.1.5 Selenomethionyl DBE ₁₆₋₁₉₇ crystals	61
2.7.2 Krr1p ₁₃₋₁₉₆ and HRB2 ₂₈₋₂₁₅ crystals	61
2.8 GST-RpS9 pull down	62
2.9 Buffers and reagents	64
2.9.1 Bacterial culture and competent cell preparation	64
2.9.2 Agarose gel electrophoresis buffer	65
2.9.3 SDS-PAGE buffer	66
2.9.4 Buffers for protein purification	66

2.9.5 Buffers for protein crystallization	67
2.9.6 Buffers for GST pull down assay	68
2.10 Primer sequences	68
2.10.1 Primers for cloning <i>dbe</i> deletion mutants and point mutants	68
2.10.2 Primers for cloning <i>krr1</i> ₁₃₋₁₉₆ and <i>hrb2</i> ₂₈₋₂₁₅	71
2.10.3 Primers for cloning <i>rps9</i>	71
3. Identification of DBE₁₆₋₁₉₇ as the working construct and biophysical characterization of DBE₁₆₋₁₉₇	72
3.1 Introduction	72
3.2 Bioinformatics analysis of DBE	73
3.2.1 Homology alignment	73
3.2.2 Disorder prediction	75
3.3 Construction of DBE deletion mutants	77
3.4 Identification of DBE₁₆₋₁₉₇ as the working construct	79
3.4.1 The DBE ₁₆₋₁₉₇ construct was initially chosen	79
3.4.2 Selection between DBE ₁₆₋₁₉₇ and DBE ₂₄₋₁₉₇	84
3.4.3 Expression and purification of HS-DBE ₁₆₋₁₉₇	87
3.5 Biophysical characterization of DBE₁₆₋₁₉₇	89
3.5.1 Far-UV circular dichroism (CD) analysis	89
3.5.2 Light scattering	92
3.6 Discussion	95
3.6.1 A conserved yet disordered segment is identified	95
3.6.2 The DBE ₁₆₋₁₉₇ protein was the most feasible construct to work with	96
3.6.3 Implications from biophysical studies	98

4. X-ray crystallography studies on DBE₁₆₋₁₉₇	101
4.1 Introduction	101
4.2 Crystallization of DBE₁₆₋₁₉₇	103
4.2.1 Screening of crystallization condition of DBE ₁₆₋₁₉₇	103
4.3 X-ray diffraction analysis of DBE₁₆₋₁₉₇ crystals	106
4.4 Phasing of DBE₁₆₋₁₉₇ crystals	108
4.4.1 Sulfur SAD	109
4.4.2 Heavy atom MIR/SIRAS	111
4.4.2.1 Heavy metal atoms	111
4.4.2.2 Halides	115
4.4.3 Selenium based MAD	118
4.5 Discussion	122
4.5.1 Twinning and diffraction anisotropy analyses of the DBE ₁₆₋₁₉₇ crystal	122
4.5.2 Why molecular replacement failed?	125
4.5.3 Why experimental phasing failed?	126
4.5.3.1 Sulfur SAD	126
4.5.3.2 Heavy atom MIR/SIRAS	126
4.5.3.3 Selenium based MAD	128
5. Searching for alternative crystal form of DBE₁₆₋₁₉₇	131
5.1 Introduction	131
5.2 Seeding and additives screen	132
5.3 Crystallization condition screening of HS-DBE₁₆₋₁₉₇ and DBE₁₆₋₂₂₉	134
5.4 Crystallization condition screening of DBE₁₆₋₁₉₇ point mutants	138

5.4.1 Design of 3 charged-swapping DBE ₁₆₋₁₉₇ mutants	138
5.4.2 Design of 3 clustered residues to alanine DBE ₁₆₋₁₉₇ mutants	139
5.4.3 Expression, purification and crystallization screening of the 6 DBE ₁₆₋₁₉₇ mutants	141
5.5 Crystallization of 2 DBE₁₆₋₁₉₇ orthologs, Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅	144
5.5.1 Construct of Krr1p ₁₃₋₁₉₆ and HRB2 ₂₈₋₂₁₅	144
5.5.2 Expression, purification and crystallization screening of Krr1p ₁₃₋₁₉₆ and HRB2 ₂₈₋₂₁₅	144
5.5.3 Attempts to improve Krr1p ₁₃₋₁₉₆ crystal diffraction	152
5.6 Discussion	156
6. Identification of Rps9 binding region in DBE	160
6.1 Introduction	160
6.2 The GST-Rps9 pull down assay on DBE deletion mutants	161
6.2.1 Purification of DBE deletion mutants	161
6.2.2 The result of GST-Rps9 pull down assay	163
6.3 Further attempts to understand interaction between Rps9 and DBE₁₆₋₂₂₉	166
6.4 Discussion	170
7. General discussion	173
8. Concluding remark	179
9. References	181

1. Introduction

1.1 Introduction of ribosome biogenesis

1.1.1 Textbook knowledge and frontiers in ribosome research

Ribosome has been discovered for decades, it is now textbook knowledge that ribosome is composed of about 65% ribosomal RNA (rRNA) and 35% ribosomal proteins; the eukaryotic ribosome, 80S in size, is composed of a 60S large subunit and a 40S small subunit; the ribosome is the cellular protein synthesizing machine that makes proteins from amino acids carried by transfer RNA (tRNA) according to the instructions from messenger RNA (mRNA) (Nelson and Cox, 2005; Alberts et al., 2008). Yet ribosome research is still gaining momentum.

Research interests are from different perspectives. For instance from a nanotechnology viewpoint, the ribosome can be regarded as a nano-scale (~20 nm in diameter) physical system that excels at making proteins (Whitesides, 2005). Scientists with an engineering bent are now developing orthogonal ribosomes with new genetic codes in hope of expanding the range of amino acids that can be incorporated into proteins (Rackham and Chin, 2005; Wang et al., 2007; Neumann et al., 2010). Moreover, creation of “artificial ribosome” was announced in 2009 winter, which was hailed by some as a milestone towards “artificial life”, although the finding

is yet to be published.

(http://www.harvardscience.harvard.edu/foundations_articles_taking-a-stride-toward-synthetic-life).

Although some interesting discoveries in ribosome engineering are noted, it is the structural investigation of ribosome that is the triumph of the past 10 years. A few high resolution structures of ribosomes, including archaean *Thermus thermophilus* (Ban et al., 2000; Schluenzen et al., 2000; Wimberly et al., 2000), prokaryotes *Escherichia coli* (Mitra et al., 2005; Schuwirth et al., 2005) are reported, whereas the ribosome structure of eukaryotes *Saccharomyces cerevisiae* are of medium resolution (Connell et al., 2007; Taylor et al., 2009). Structural studies have discovered that the active site of ribosome is made of rRNA only, thus confirming the ribosome as essentially a ribozyme (Nissen et al., 2000). While the structural understanding of ribosome architecture is growing, researchers are now trying to look at the structural dynamics of the ribosome during the process of translation using a variety of biophysical techniques. For instance, crystal and cryo-electron microscopy (EM) structures provide valuable snapshots of the translation process (Selmer et al., 2006; Connell et al., 2007; Gilbert et al., 2007; Passmore et al., 2007); some researchers have even suggested that single molecular fluorescence resonance energy transfer

(smFRET) can be used for real time monitoring of the tRNA transit during translation (Blanchard et al., 2004; Munro et al., 2007; Uemura et al., 2010). In short, scientists thrive on working out the molecular details of, not just the ribosome, but the whole translation process, although the picture is as yet far from complete.

However, compared with the knowledge about mature ribosome, much less is known about how a ribosome comes into being, i.e. ribosome biogenesis, which is an even more complicated process (Dinman, 2009). Findings in ribosome biogenesis may have important roles in the future development of ribosome engineering.

1.1.2 Ribosome biogenesis and diseases

Before introducing ribosome biogenesis, it should be noted that its clinical relevance is gradually being established. Ribosome biogenesis exemplifies differences between prokaryotic and eukaryotic ribosomes. Scientists take advantage of the difference between the two processes in order to develop new antibacterial drugs to target proteins that are specific for bacterial ribosome biogenesis, leaving the eukaryotic one undisturbed (reviewed by (Comartin and Brown, 2006)).

For humans, ribosomopathies, a class of genetic diseases recently discovered, is

caused by impaired ribosome biogenesis due to genetic abnormalities (reviewed by (Narla and Ebert, 2010)). There are reports showing a predisposition of cancer to ribosomopathies patients (Taskinen et al., 2008; Alter et al., 2009). The linkage of ribosome biogenesis to cancer gains support from studies in zebrafish models (Amsterdam et al., 2004; MacInnes et al., 2008), mouse model (Grisendi et al., 2005) and the study of human breast cancer cells *in vitro* (Belin et al., 2009). It has been suggested that mutations in the ribosomal proteins can cause the bypass of the normal surveillance mechanism of p53-mediated apoptosis, which may help explain why ribosomopathies patients are prone to cancer development (reviewed by (Deisenroth and Zhang, 2010)). Ribosome biogenesis is also under the regulation of the Myc protein family, a class of transcription factors that is related to tumor genesis (reviewed by (van Riggelen et al., 2010)). The possibility of taking ribosome biogenesis as a chemotherapeutic target is being evaluated (Burger et al., 2010), therefore, findings in ribosome biogenesis may have important roles in the future development of medications.

1.1.3 Three stakeholders in ribosome biogenesis

Eukaryotic ribosome biogenesis is a complicated and dynamic process that involves the making of 4 rRNAs, 70 ribosomal proteins and about 200 *trans*-factors (reviewed

by (Fromont-Racine et al., 2003; Henras et al., 2008)). As most research in ribosome biogenesis are conducted in yeast, findings in yeast are for the moment considered to reflect the blueprint in eukaryotic systems, as the basic features and *trans*-factors involved in the ribosome biogenesis machineries are largely conserved across eukaryotes (Raska et al., 2006).

Ribosomal RNAs

Ribosomal RNA makes up the catalytic unit of the ribosome and hence it is central to the function of ribosome (Nissen et al., 2000). There are four rRNAs in the mature yeast ribosome, the 25S, 5.8S and 5S rRNA are found in the 60S large subunit while the 18S rRNA is found in the 40S small subunit. Although found in different ribosomal subunits, the 25S, 18S and 5.8S rRNA share the same origin from the long 35S pre-ribosomal RNA (pre-rRNA), while the 5S rRNA is transcribed separately. The rRNA receives extensive processing for maturation, which is further discussed in the section on rRNA processing (Section 1.1.4.2).

Ribosomal proteins

Over 70 ribosomal proteins are involved in ribosome biogenesis. They are thought to play no direct role in the ribozymatic peptidyl-transferase activity of the ribosome

(Nissen et al., 2000). Some ribosomal proteins, however, are suggested to have roles that may facilitate the production of the ribosome, for example aiding the nuclear export of precursor rRNAs (Ferreira-Cerca et al., 2005); depletion of ribosomal proteins resulting in compromised ribosome production (Robledo et al., 2008).

trans-factors

The *trans*-factors are either protein or small nucleolar ribonucleoprotein (snoRNP) in nature, snoRNP consists of small nucleolar RNA (snoRNA) and proteins, and they combine to function as a unit (reviewed by (Kressler et al., 1999)). Eukaryotes have about 200 *trans*-factors, which are more elaborate than the prokaryotic counterpart (reviewed by (Hage and Tollervey, 2004; Henras et al., 2008)). Most eukaryotic *trans*-factors are strictly essential such that depletion of them individually can halt ribosome biogenesis and are therefore lethal to the organism, but they act only transiently and are not found in mature ribosomes (reviewed by (Hage and Tollervey, 2004)).

Several types of function have been suggested for *trans*-factors (reviewed by (Kressler et al., 1999; Fromont-Racine et al., 2003; Hage and Tollervey, 2004)). The first type acts as a RNA helicase to assist rRNA folding and rearrange protein-RNA

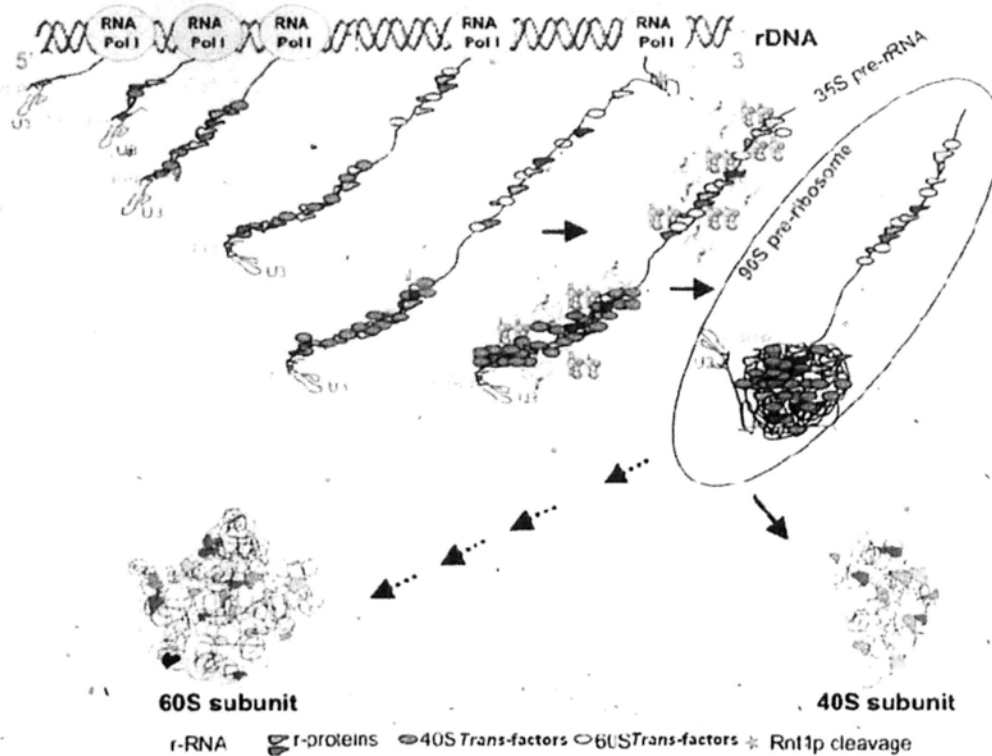
interactions (reviewed by (Kressler et al., 1999)). The second type possesses enzymatic activities for rRNA processing (reviewed by (Kressler et al., 1999)). The third type guides the precise interaction of rRNAs with ribosomal proteins (reviewed by (Kressler et al., 1999)). The fourth type helps large scale structural reorganization of pre-ribosomal subunits (Hage and Tollervey, 2004). The fifth type helps translocation of the pre-ribosomal subunits from nucleus to cytoplasm (Fromont-Racine et al., 2003).

1.1.4 Two central processes in ribosome biogenesis

Ribosome biogenesis is in fact the interplay of two processes, i.e. ribosome assembly and rRNA processing. It is generally held that the two processes are coupled and co-regulated, which means rRNA processing occurs concomitantly inside the complexes formed during ribosome assembly, and the complexes are largely composed of and shaped by the *trans*-factors (reviewed by (Fromont-Racine et al., 2003; Tschochner and Hurt, 2003; Henras et al., 2008)).

1.1.4.1 Ribosome assembly

Ribosome assembly is the process of associating ribosomal proteins to maturing rRNAs in a sequential order (reviewed by (Fromont-Racine et al., 2003)).



1.1 Schematic representation of ribosome assembly. Pre-rRNA is transcribed by RNA polymerase I. Ribosome assembly begins with the transcription of the nascent pre-rRNA. The *trans*-factors for 40S subunit synthesis associate first with the pre-rRNA in the 5'-end, and *trans*-factors for 60S subunit assemble later near the 3'-end. The full pre-rRNA-transcript is released by cleavage mediated by a RNase named Rnt1p. The released pre-rRNA is contained in a particle called 90S pre-ribosome, in which rproteins and numerous *trans*-factors are already assembled. The 90S pre-ribosome is then split in 2, and further mature into 60S and 40S subunit. It should be noted that *trans*-factors constituents of 90S pre-ribosome outnumber the rproteins. This figure is adapted from (Hage and Tollervey, 2004).

Ribosome assembly begins with the transcription of the nascent pre-rRNA and the continual assembly of ribosomal proteins onto the rRNA gives rise to a series of ribosomal intermediates known as ribosomal ribonucleoprotein precursor (pre-rRNP) complexes (reviewed by (Hage and Tollervey, 2004)). The pre-rRNPs are largely constituted by *trans*-factors, for instance, 35 *trans*-factors, in contrast to only 5 ribosomal proteins, are found to be the components in an early pre-rRNP named 90S pre-ribosome (Grandi et al., 2002; Bernstein et al., 2004). The plethora of *trans*-factors are an important structural component of the pre-rRNP and hence depletion of one individual *trans*-factor is enough to halt ribosome biogenesis (reviewed by (Hage and Tollervey, 2004)). Although there may be more pre-rRNPs, most researchers agreed to group the various pre-rRNPs into three main classes, the 90S pre-ribosomes, the pre-40S particles and the pre-60S particles (reviewed by (Henras et al., 2008)). The 90S pre-ribosome is the largest and earliest detectable pre-rRNP, and it is then split and gives rise to the pre-40S and pre-60S particles, which are further processed to generate the 2 final ribosome subunits as illustrated in Figure 1.1.

1.1.4.2 Ribosomal RNA processing

Ribosomal RNA processing involves 2 tasks, pre-rRNA cleavage and base modification. Ribosomal RNA processing proceeds within the 90S pre-ribosome that is composed of numerous *trans*-factors (reviewed by (Hage and Tollervey, 2004)).

Pre-rRNA cleavage

Immediately after transcription, the pre-rRNA undergoes a series of cleavages (reviewed by (Granneman and Baserga, 2004)). The cleavage reaction starts from the 35S pre-rRNA, generating a number of defined intermediates, and finally ends with the 3 mature 25S, 18S and 5.8S rRNAs (reviewed by (Granneman and Baserga, 2004; Decatur et al., 2007)). The yeast pre-rRNA cleavage reaction is illustrated in Figure 1.2 A. Pre-ribosomal RNA cleavage is largely conserved among eukaryotes, as observed by the common features between the pre-rRNA cleavage pattern of yeast and that of *Drosophila* (Figure 1.2 B), for example, the organization of the nascent pre-rRNA transcript is similar in that 18S rRNA precedes 5.8S rRNA and then 25S (yeast)/28S (*Drosophila*) rRNA. The cleavage is precise and well-ordered; and the 18S rRNA is generated early via a specific cleavage at A₂ and site 3 of yeast and *Drosophila*. Of note, the cleavage position that separates the 18S rRNA is critical, because it is where separation of the 40S and 60S subunit synthesis begins. *trans*-factors play important

roles in rRNA processing, for instance, the U3 snoRNP, which is ubiquitously present in eukaryotes (Marz and Stadler, 2009), functions in the early pre-rRNP and mediates endonucleolytic pre-rRNA cleavage by base pairing of the U3 snoRNA with the pre-rRNA (Hughes and Ares, 1991; Beltrame and Tollervey, 1995).

Ribosomal RNA modifications

Three kinds of modifications, pseudouridylations, ribose- and base- methylations, are required for rRNA processing (reviewed by (Decatur and Fournier, 2002)). It is estimated that about 100 modifications occur in yeast, and even 200 in humans (Piekna-Przybylska et al., 2008). Modifications takes place during early ribosome biogenesis within the 90S pre-ribosome, such that specific sites at rRNAs are targeted by base pairing of the snoRNA to snoRNP *trans*-factors and the modifications are concentrated in the conserved regions of mature rRNAs among eukaryotes (reviewed by (Decatur and Fournier, 2002; Decatur and Fournier, 2003; Hage and Tollervey, 2004). The rRNA modification findings have been recently archived into a database (Piekna-Przybylska et al., 2008). The rRNA modifications are important for many functional aspects of the ribosome, such as translation initiation, translocation, maintenance of reading frame, peptide bond formation, to name a few (reviewed by (Decatur and Fournier, 2002)).

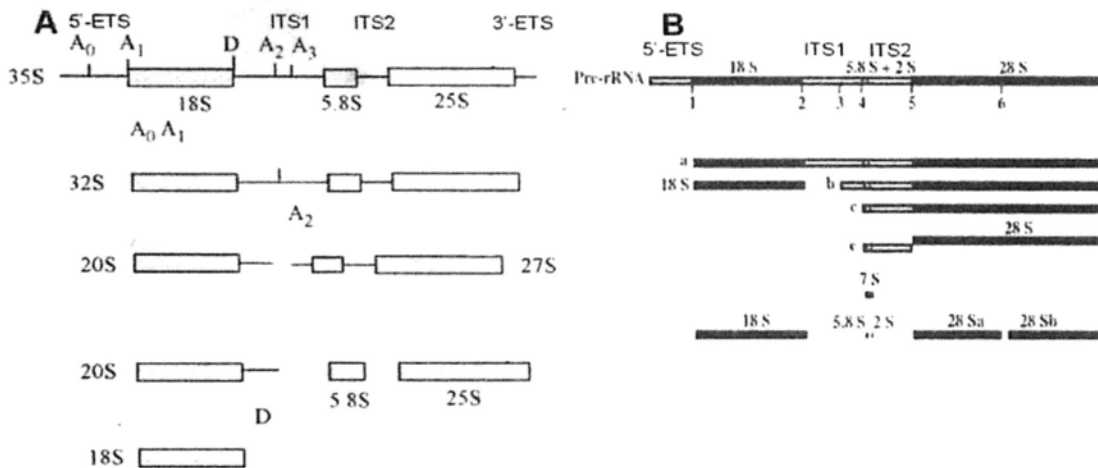


Figure 1.2 (A) Schematic representation of pre-rRNA cleavage in yeast. The mature rRNAs (18S, 5.8S and 25S) are transcribed as a long 35S pre-rRNA containing external and internal transcribed spacers (ETS and ITS). The pre-rRNA is first cleaved at sites A₀, A₁ and then A₂, yielding the 20S pre-rRNA and 27S pre-rRNA. The 20S pre-rRNA matures into 18S rRNA and the 27S pre-rRNA is further processed to give the mature 5.8S and 25S rRNA. This figure is adapted from (Liang et al., 2010). **(B) Schematic representation of pre-rRNA cleavage in *Drosophila*.** Similarly to yeast, the mature rRNAs (18S, 5.8S and 28S) are transcribed as a long pre-rRNA containing ETS and ITS. Cleavage at site 3 generates the 18S, and a precursor for 5.8S and 28S rRNA that further matures into 5.8S and 28S rRNA. It is interesting to note that in both yeast and *Drosophila* the organization of the nascent pre-rRNA is similar in that 18S rRNA precedes 5.8S and 25S/28S rRNA; the cleavage reaction is precise and well-ordered; and 18S rRNA is separated early by cleavage at A₂ and site 3 of the yeast and *Drosophila* system. This figure is adapted from (Chan et al., 2001).

1.1.5 Ribosome biogenesis proceeds from nucleolus to cytoplasm

Ribosome biogenesis is not localized in one location, the pre-rRNPs are actively transported from the nucleolus to the nucleus and then to cytoplasm (reviewed by (Fromont-Racine et al., 2003)). It is exactly these 3 different places that characterize the 3 stages of ribosome biogenesis (reviewed by (Fromont-Racine et al., 2003)). The early stage happens in the nucleolus where the 90S pre-ribosome is assembled and then split; the middle stage happens in the nucleoplasm where further rRNA processing in pre-60S particles occurs; and the final stage sees the export of the pre-40S and pre-60S particles through the nuclear pore complexes to the cytoplasm, where they mature to give the final ribosomal subunits (reviewed by (Fromont-Racine et al., 2003)).

1.1.6 The 90S pre-ribosome is an early pre-rRNP

The subcellular localizations of *trans*-factors can help reveal their role in ribosome biogenesis, for instance, all the *trans*-factors in the 90S pre-ribosome localize in the nucleolus (Dragon et al., 2002; Bernstein et al., 2004). The 90S pre-ribosome, alternatively known as the small subunit (SSU) processome, is a large pre-rRNP complex where the 35S pre-rRNA cleavage and rRNA modifications begins, it leads to the release of the precursor of 18S rRNA and hence it acts as an assembly

intermediate of the 40S subunit (Dragon et al., 2002; Grandi et al., 2002; Bernstein et al., 2004). The importance of *trans*-factors in 90S pre-ribosome is highlighted by the fact that it consists of 35 *trans*-factors but only 5 ribosomal proteins (Dragon et al., 2002; Bernstein et al., 2004); and some specific *trans*-factor components are found to coordinate rRNA transcription and rRNA processing (Gallagher et al., 2004). U3 snoRNP, is one of the *trans*-factors in the 90S pre-ribosome (Grandi et al., 2002; Bernstein et al., 2004) and is ubiquitously present in eukaryotes (Marz and Stadler, 2009). This exemplifies the importance of *trans*-factors in that they mediate endonucleolytic pre-rRNA cleavage by base pairing of the U3 snoRNA with the pre-rRNA (Hughes and Ares, 1991; Beltrame and Tollervey, 1995).

One interesting finding of the 90S pre-ribosome is that very few or even no large subunit ribosomal protein is identified (Dragon et al., 2002; Grandi et al., 2002), which can be explained by a model established by electron microscopy (EM) images that have captured the rRNA transcription process (Osheim et al., 2004). The model is illustrated in Figure 1.3. Consistent with other biochemical findings, the model suggests that 90S pre-ribosome is the first component of the machinery to initiate processing of the 35S pre-rRNA, and trigger the assembly of the pre-40S particle (Dragon et al., 2002; Grandi et al., 2002; Osheim et al., 2004).

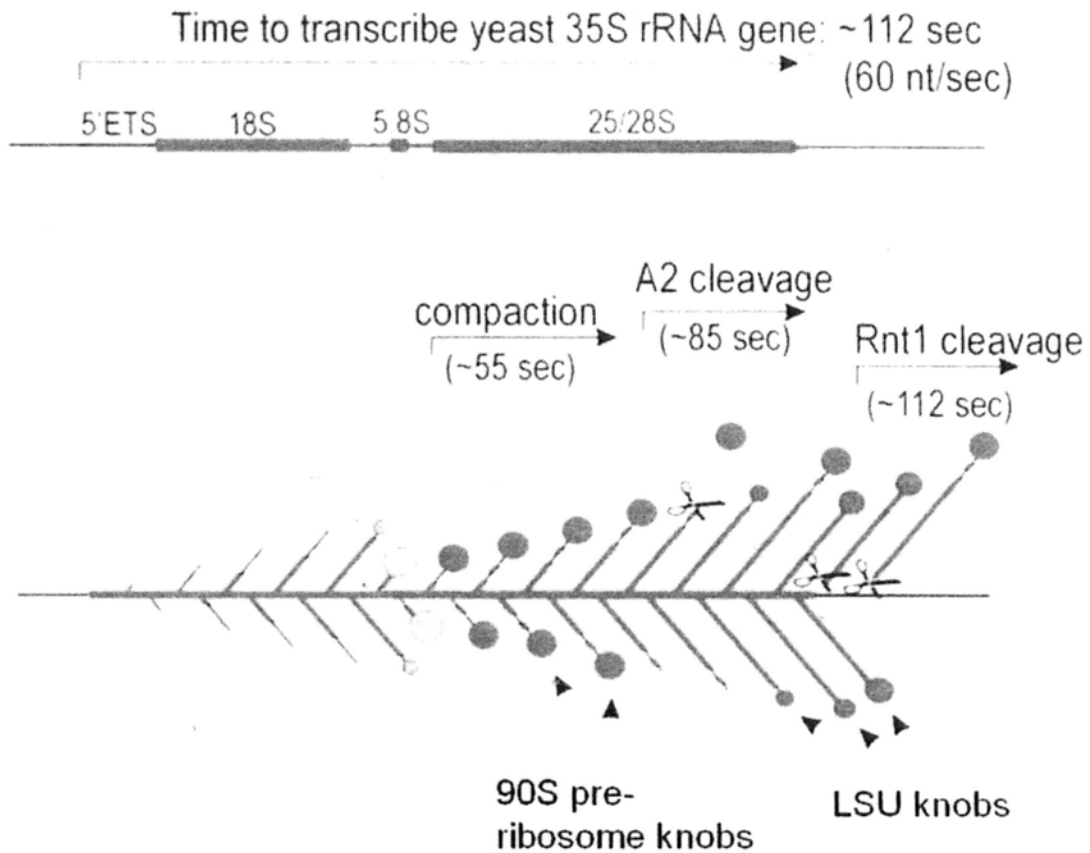


Figure 1.3 A model for 90S pre-ribosome functioning in yeast. Pre-rRNA transcripts start with the 18S sequence, followed by the 5.8S and 25S sequence. The transcription process lasts for about 112 seconds. Just after transcription of the 18S sequence, the unfinished transcripts compact into the 90S pre-ribosome which is seen as knobs from the EM images (Osheim et al., 2004). About 30 seconds later, the premature transcript is cleaved at the A₂ site, causing the release of the 90S pre-ribosome together with the 18S sequence. The large subunit (LSU) processing machineries continue assembling on the 5'- end of the remaining transcript until the whole transcript is cleaved by the RNase, Rnt1p. Note that cleavage at the A₂ site is random, leaving some pre-rRNA full transcript with 90S pre-ribosome attached. This figure is adapted from (Osheim et al., 2004).

Whilst EM is a useful tool for kinetics study of 90S pre-ribosome functioning, it has yet to give a clue as to how the 90S pre-ribosome is organized and how it is assembled on the rRNA. To have a deeper understanding on 90S pre-ribosome, investigating its structural components, i.e. the *trans*-factors, is necessary.

1.2 The Dribble (DBE) family

1.2.1 Dribble family is essential for rRNA processing and ribosome assembly

Dribble (DBE) was originally identified as a *Drosophila* protein, but it also represents a family of well conserved essential nucleolar proteins in eukaryotes (Sasaki et al., 2000; Chan et al., 2001; Xin et al., 2005). Dribble is a *trans*-factor that involves in both rRNA processing and ribosome assembly. It is essential for rRNA processing such that early aberrant rRNA cleavage in *dbe* null mutant fly larvae is detected, and the mutant larvae dies in early stage (the first instar larvae) of development when maternal gene products are presumably used up (Chan et al., 2001). DBE helps ribosomal assembly such that it associates with a high molecular weight complex and is found to have many ribosomal proteins as interacting partners, for example, RpS9. (Choi, 2007)

The yeast member of DBE family, Krr1p, is also involved in rRNA processing and

ribosome assembly, for instance, yeast *krr1* mutant leads to reduced 18S rRNA production and 40S small subunit synthesis (Sasaki et al., 2000), and it interacts physically and genetically with many ribosomal proteins and *trans*-factors, (Grandi et al., 2002; Gromadka et al., 2004). Interestingly RpS9 is also found to be one of the genetically interacting ribosomal proteins of Krr1p in yeast (Gromadka et al., 2004), in addition to being directly interacting with *Drosophila* DBE.

1.2.2 Dribble family's function can be understood via pre-rRNP

Of note, Krr1p is one of the *trans*-factor components of the yeast 90S pre-ribosome, suggesting that the DBE family is no exception to the fact that *trans*-factors organize themselves into the pre-rRNP in order to function (Dragon et al., 2002; Grandi et al., 2002; Bernstein et al., 2004). Being nucleolus localized (Sasaki et al., 2000) and found in 90S pre-ribosome, but not in the 40S particle and 60S particle (Bassler et al., 2001; Harnpicharnchai et al., 2001; Schafer et al., 2003), Krr1p possibly plays its role in the early ribosome biogenesis only via the context of the early pre-rRNP, 90S pre-ribosome.

It is interesting to note that yeast RpS9 is also present in the 90S pre-ribosome, together with Krr1p (Bernstein et al., 2004). It might be suggested that the early

pre-rRNP is the platform where Krr1p, and possibly other DBE family members, function in ribosome biogenesis. Even more interestingly, the suggestion that the early pre-rRNP acts as a platform for the DBE family to function gives clues on its role on rRNA processing. For instance, carrying no enzymatic property itself, DBE is unlikely to mediate rRNA cleavage by itself, however, DBE can be an important structural component of the early pre-rRNP where rRNA processing takes place, such that the early aberrant rRNA cleavage observed in *dbe* null *Drosophila* mutant (Chan et al., 2001) might be due to a structural imbalance of the pre-rRNP in the absence of DBE so that rRNA aberrant cleavage occurs by mis-pairing of rRNA and snoRNA.

1.2.3 A structured region is possibly hidden in DBE

Before any further findings on DBE's role in the pre-rRNP, a few biophysical findings on DBE may give information about its functions. Dribble demonstrated RNA affinity, and its extensive structural flexibility was stabilized upon binding to heparin, which is a nucleic acid homolog (Yiu et al., 2006). Far-UV circular dichroism (CD) analysis suggests that some structurally stable region(s) are possibly hidden, as DBE only began to denature at high temperature (~ 90°C) (Figure 1.4). From the deconvolution of the CD spectrum, DBE is estimated to contain approximately 55% secondary structure, which corresponds to about 190 amino acid residues of the 345

amino acid residues of the full length protein (Figure 1.4) (Yiu et al., 2006). However, database searches in both *Pfam* (Finn et al., 2008) and *Superfamily* (Wilson et al., 2009) reveal that DBE contains a type I KH-like domain spanning between residue 100 and 200, and segment from residue 20 to 119 is also predicted to be structured but the structural identity is currently unknown. Both type I and type II KH domains are well studied nucleic acid binding domains (Grishin, 2001; Valverde et al., 2008). The presence of a KH-like domain may help explain the RNA binding ability of DBE, but it is the structural obscurity of the fragment between residue 20 and 119 and its unreported organization with KH domain that is intriguing.

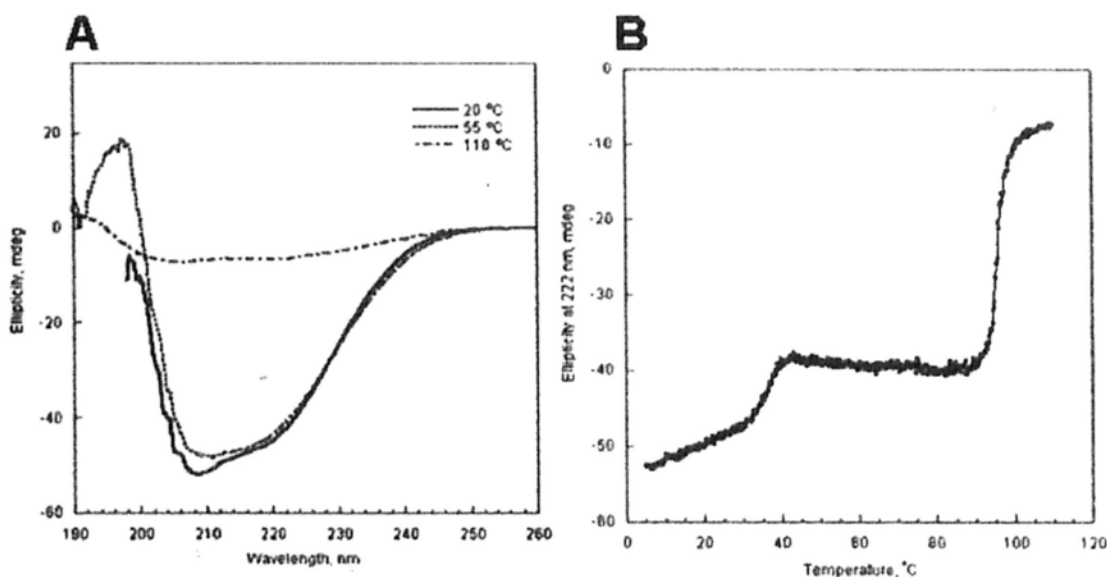


Figure 1.4 (A) Far-UV CD spectrum of DBE at 20, 55, and 110 °C. A large absorption with a peak around 210 nm is observed from both the CD spectra of DBE at 20 and 55 °C, suggesting that the protein is relatively stable up to 55 °C, whereas CD absorption vanishes when the measurement temperature is raised to 110 °C indicating that the protein is denatured. From the deconvolution of CD spectrum of DBE at 20 °C, DBE is estimated to have approximately 55% secondary structure, corresponding to about 190 amino acid residues of the 345 amino acid residues of the full length protein. The figure is adapted from (Yiu et al., 2006). **(B) Thermal denaturation profile of DBE as monitored by CD spectroscopy at 222 nm.** The protein shows a two step denaturation, the first has a mid-point around 35 °C, and the second begins at high temperature 90 °C. In between the two transitions from 40 °C to 90 °C, the protein remains relatively stable, suggesting that DBE possibly consists of structurally compact region(s) that may confer to its thermostability. This figure is adapted from (Yiu et al., 2006).

1.3 Aim of research

Three main challenges in understanding ribosome biogenesis, i.e. the order of ribosome assembly, nuclear export and structural biology have been suggested (Dinman, 2009). These challenges reside in the fact that ribosome biogenesis is a dynamic process that involves the pre-rRNP complexes, which are in turn composed of numerous transiently acting *trans*-factors. It has been suggested that pre-rRNPs are the platforms where *trans*-factors function, but how exactly the pre-rRNP machineries organize, assemble and mediate rRNA cleavage is unknown. Dribble represents a family of essential and evolutionarily conserved *trans*-factors, which may provide a window in studying the pre-rRNP machineries. It has been suggested that DBE, as an essential structural component, is indispensable in that it maintains the intricate structural balance of the early pre-rRNP in which it may be involved. As proper pre-rRNA cleavage is mediated by the snoRNA in the pre-rRNP, the possible involvement of DBE into the pre-rRNP may help explain how DBE affects rRNA processing, for example, by maintaining the proper pairing of rRNA with snoRNA. The genetic and direct interaction of RpS9/ DBE interaction as observed in yeast and *Drosophila* can also be explained by their possible involvement into the early pre-rRNP, where DBE may help in assembling ribosomal proteins like RpS9.

It should be noted that correct rRNA processing is the key to the production of a functional ribosome. With continuous processing, modifications, attaching and detaching of the *trans*-factors, it is no wonder rRNA undergoes extensive structural rearrangements throughout ribosome biogenesis. Structural characterization is thus essential in understanding ribosome biogenesis, yet it is not reported whether of 90S pre-ribosome or other pre-rRNP complexes. While the difficulty to structurally characterize the whole unit of pre-rRNP is noted, possibly due to their dynamic and transient nature, it is feasible to determine the structure of individual constituents, from which a glimpse of the whole unit can be reconstituted. The first attempt using this approach in pre-40S particle is recently reported, in which previously determined structures of 6 *trans*-factors are mapped to the corresponding rRNA positions on the pre-40S particle, and the pre-40S particle is then reconstituted by computational modeling (Granneman et al., 2010).

As DBE probably contains a structured region that is partly undefined and partly KH-like, structural determination of the region may not only reveal its structural identity, but also act as the first step in revealing the structural organization of the early pre-rRNP as in the case of pre-40S particle. Moreover, as it has been demonstrated that DBE directly interacts with RpS9 (Choi, 2007), mapping the RpS9

binding region in DBE may characterize in terms of RpS9 binding the functional value in ribosome assembly of the unreported structured region of DBE.

Hence, the objectives of this project are structural characterization of the structured region of DBE by X-ray crystallography and functional characterization of the structured region of DBE in terms of RpS9 binding.

2. Materials and methods

2.1 General techniques

2.1.1 Preparation of *E. coli* competent cells

Escherichia coli strain *DH5 α* , *BL21 (DE3) pLysS*, *C41* or *B834S* was streaked from a frozen stock onto the surface of an LB agar plate (Section 2.9.1) for *DH5 α* and *C41* or a LBC agar plate (Section 2.9.1) for *BL21 (DE3) pLysS*. The agar plate was incubated at 37°C for 14 to 16 hours. A single colony from the plate was inoculated into 1 ml of LB medium with appropriate antibiotics (Section 2.9.1) and shaken at 37°C for 14 to 16 hours. The culture was transferred to 100 ml of LB medium with appropriate antibiotics and continued shaking was continued at 37°C until OD₆₀₀ readings reached 0.4 to 0.5. The culture was then chilled on ice for 5 minutes and then centrifuged at 3,900g (Beckman rotor, JLA 16.250) for 10 minutes at 4°C. The bacterial pellet was resuspended in 40 ml RF1 buffer (Section 2.9.1) and then kept on ice for 5 minutes. The bacterial suspension was then centrifuged again at 3,900g for 10 minutes at 4°C. The bacterial pellet was then resuspended in 4 ml of RF2 buffer (Section 2.9.1) and kept on ice for 15 minutes. Aliquots of 100 μ l cells were made and stored at -80°C until used.

2.1.2 Transformation of *E. coli* competent cells

An aliquot of 100 μ l of competent cells was quickly thawed by hand. For plasmid DNA, 0.5 μ g of DNA was added to the cells and the competent cells were spread immediately on LB agar plate containing appropriate antibiotics (Section 2.9.1) and then incubated at 37°C for 16 hours. For ligation products, all the ligation products were added to the cells. The cells were left on ice for 30 minutes, followed by 42°C for 2 minutes. The tube was then cooled on ice for 10 minutes. 400 μ l of LB medium (Section 2.9.1) was added to the cells and incubated at 37°C with shaking at 250 rpm for 45 minutes. The cells were then pelleted by centrifugation at 10,000g (model 5415R, Eppendorf) at room temperature for 1 minute. The cells were resuspended in 100 μ l of LB medium and then were spread to a LB agar plate containing appropriate antibiotics and incubated at 37°C for 16 hours.

2.1.3 Agarose gel electrophoresis of DNA

Agarose gels were prepared with 1% (w/v) agarose (Bio-Rad) dissolved in 1X TBE electrophoresis buffer (Section 2.9.2) containing 0.5 μ g/ml ethidium bromide. Samples were mixed with 6X DNA loading buffer (Section 2.9.2) to a final concentration of 1X. Gel electrophoresis was performed at a constant voltage of 120 V in a gel tank containing an appropriate volume of 1X TBE buffer.

2.1.4 Polymerase Chain Reaction (PCR)

All PCR reactions were performed under the following conditions unless otherwise specified.

2.1.4.1 Condition of PCR

Polymerase Chain Reaction was performed in a 50 μ l scale in iCycler or C1000 Thermocycler (Bio-Rad). Either *Pfu* DNA polymerase (Promega) or Phusion® High Fidelity Polymerase (Finnzymes) was used. The reaction mix was prepared as follows:

	<i>Pfu</i>	Phusion®
Sterile distilled H ₂ O	42 μ l	36.25 μ l
Reaction buffer	5 μ l	10 μ l
10 mM dNTPs (2.5 mM each of dATP, dTTP, dGTP and dCTP)	1 μ l	1 μ l
Forward primer (0.1 nmole/ μ l)	0.5 μ l	1 μ l
Reverse primer (0.1 nmole/ μ l)	0.5 μ l	1 μ l
Template DNA (~ 0.1 μ g)	0.5 μ l	0.5 μ l
DNA polymerase	0.5 μ l	0.25 μ l
Total	50 μl	50 μl

Reactions were carried out using the following thermocycling profile:

Step	Temperature (°C) (<i>Pfu</i> / Phusion®)	Duration (minutes) (<i>Pfu</i> / Phusion®)
1	94 / 98	5 / 3
2	94 / 98	0.5 / 0.3
3	50 / 50	0.5 / 0.25
4	72 / 72	1 / 0.5
5	72 / 72	7 / 5
6	16 / 16	Until retrieval

Step 2 to 4 was repeated for 30 times.

After reaction, the mixture was analyzed by agarose gel electrophoresis. The PCR products were purified by DNA/RNA extraction kit (Viogene) if need be.

2.1.5 Restriction digestion of DNA

Restriction digestion of DNA was carried out routinely in a volume of 20 μ l in an appropriate buffer as suggested by the supplier (NEB). 1 μ g of plasmid DNA or DNA from all PCR products after PCR purification was taken for restriction digestion. The reaction was performed with 10 U of each enzyme at 37°C for 3 hours. The result was analyzed by agarose gel electrophoresis. The DNA fragments were retrieved from agarose gel by Gel Extraction Kit (Qiagen) if need be.

2.1.6 Ligation of DNA fragments

Ligation reaction was performed in a volume of 20 μl at 16°C for 20 hours. 200 U of T4 DNA ligase (NEB) was used with reaction buffer provided by manufacturer.

Reaction mixture was prepared in the order specified below:

T4 DNA ligase	0.5 μl
10X ligase buffer	2 μl
Linearized vector	X μl
DNA insert	Y μl
Sterile distilled H ₂ O	(17.5 – X- Y) μl
<hr/>	
Total	20 μl

After ligation, the reaction product was added to freshly thawed *DH5 α* competent cells (Section 2.1.1) for transformation. Single colonies were picked and grown in 5 ml of LB medium overnight with appropriate antibiotics (Section 2.9.1), and plasmid DNA was extracted using a DNA-Spin™ Kit (iNtRON).

2.1.7 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

All SDS-polyacrylamide gels were run using a Mini-PROTEAN III electrophoresis cell (Bio-Rad). The glass plates were assembled into the casting frame according to

manufacturer's instructions (Bio-Rad). Separating gel solution of 12.5% acrylamide was prepared as follows:

Separating gel	12.5%
dH ₂ O	0.83 ml
Separating gel buffer (1 M Tris, pH 8.8)	1.40 ml
30% acrylamide (19 % acrylamide, 1% bis-acrylamide)	1.54 ml
10% SDS	70 µl
10% ammonium persulfate	40 µl
TEMED	5 µl

The separating gel solution was loaded into the casting frame, followed by adding 200 µl of isopropanol on top. After 30 minutes, the separating gel solution polymerized completely, the isopropanol was purged and the gel surface was rinsed with distilled water. A stacking gel solution of 4.5% acrylamide was prepared as follows:

Stacking gel	4.5%
dH ₂ O	1 ml
Stacking gel buffer (1 M Tris, pH 6.8)	0.54 ml
30% acrylamide (19% acrylamide, 1% bis-acrylamide)	0.27 ml
10% SDS	20 µl
10% ammonium persulfate	10 µl
TEMED	2 µl

The stacking gel solution was loaded on top of the separating gel and a comb 0.75 mm thick with either a 10 or 15 well former inserted.

Appropriate volume of 6X SDS sample buffer (Section 2.9.3) was added to the sample. The samples were denatured by heating at 99°C for 10 minutes. Electrophoresis was performed with 1X SDS electrophoresis buffer (Section 2.9.3) at 60 mA per gel until the bromophenol blue dye reached the bottom of the running gel. The gel was then taken out for staining.

The gel was immersed in about 30 ml of Coomassie staining solution (Section 2.9.3) for at least 1 hour at room temperature with continuous shaking. The gel was then destained with Coomassie destaining solution (Section 2.9.3) until a protein pattern could be clearly observed.

2.2 Construction of expression vectors

2.2.1 Construction of *pET-DBE* deletion mutants

The DNA fragments encoding 10 *DBE* deletion mutants, *dbe*₁₆₋₁₉₇, *dbe*₂₄₋₁₉₇, *dbe*₂₆₋₁₉₇, *dbe*₃₄₋₁₉₇, *dbe*₁₆₋₂₁₃, *dbe*₂₆₋₂₁₃, *dbe*₃₄₋₂₂₉, *dbe*₁₆₋₂₂₉, *dbe*₂₆₋₂₂₉ and *dbe*₃₄₋₂₂₉ were generated by PCR using combinations of 4 forward primers, DBE_V16f_*Nde*I, DBE_I24f_*Nde*I, DBE_A26f_*Nde*I and DBE_H34f_*Nde*I (Section 2.10.1), and 3 reverse primers, DBE_P197r_*Xho*I, DBE_D213r_*Xho*I and DBE_K229r_*Xho*I (Section 2.10.1). DNA fragments encoding deletion mutants *dbe*₁₋₁₉₇ and *dbe*₁₋₂₂₉ were generated by *T7* primer and DBE_P197r_*Xho*I/DBE_K229r_*Xho*I (Section 2.10.1). The combination of primers and corresponding *DBE* deletion mutants are listed in Table 2.1.

Table 2.1 Primers used in the PCR generation of various DBE deletion mutants.

Fragment	Primers combinations	Protein size (kDa)
<i>dbe</i> ₁₆₋₁₉₇	DBE_V16f_NdeI x DBE_P197r_XhoI	20
<i>dbe</i> ₂₄₋₁₉₇	DBE_I24f_NdeI x DBE_P197r_XhoI	19
<i>dbe</i> ₂₆₋₁₉₇	DBE_A26f_NdeI x DBE_P197r_XhoI	19
<i>dbe</i> ₃₄₋₁₉₇	DBE_H34f_NdeI x DBE_P197r_XhoI	18
<i>dbe</i> ₁₆₋₂₁₃	DBE_V16f_NdeI x DBE_D213r_XhoI	22
<i>dbe</i> ₂₆₋₂₁₃	DBE_A26f_NdeI x DBE_D213r_XhoI	21
<i>dbe</i> ₃₄₋₂₁₃	DBE_H34f_NdeI x DBE_D213r_XhoI	20
<i>dbe</i> ₁₆₋₂₂₉	DBE_V16f_NdeI x DBE_K229r_XhoI	23
<i>dbe</i> ₂₆₋₂₂₉	DBE_A26f_NdeI x DBE_K229r_XhoI	22
<i>dbe</i> ₃₄₋₂₂₉	DBE_H34f_NdeI x DBE_K229r_XhoI	21
<i>dbe</i> ₁₋₁₉₇	T7 x DBE_P197r_XhoI	22
<i>dbe</i> ₁₋₂₂₉	T7 x DBE_K229r_XhoI	25

The *pET-DBE* vector was previously cloned and used as a PCR template. The DNA polymerase *Pfu* (Promega) was used for PCR and the condition and thermocycling profile was listed in Section 2.1.4. For the 10 PCR products encoding the *DBE* deletion mutants, *dbe*₁₆₋₁₉₇, *dbe*₂₄₋₁₉₇, *dbe*₂₆₋₁₉₇, *dbe*₃₄₋₁₉₇, *dbe*₁₆₋₂₁₃, *dbe*₂₆₋₂₁₃, *dbe*₃₄₋₂₂₉, *dbe*₁₆₋₂₂₉, *dbe*₂₆₋₂₂₉ and *dbe*₃₄₋₂₂₉, the PCR products were digested with *NdeI* (NEB) and *XhoI* (NEB) at 37 °C for 3 hours. The digested PCR products were ligated into a *NdeI/XhoI* digested *pET507* vector, a modified *pET3d* vector, using T4 DNA ligase

(NEB) at 16°C overnight. For *dbe*₁₋₁₉₇ and *dbe*₁₋₂₂₉, the PCR products were digested with *Nco*I (NEB) and *Xho*I at 37°C for 3 hours. The digested PCR product was ligated into a *Nco*I/*Xho*I digested *pET507* vector by T4 DNA ligase at 16 °C overnight. The ligation products were transformed into *E. coli DH5α* competent cells. Positive clones were selected by PCR screening and *Nde*I/*Xho*I or *Nco*I/*Xho*I digestion and the insert identities were confirmed by DNA sequencing (TechDragon Ltd). The resulting plasmids encoded the corresponding DBE deletion mutants of sizes ranged from 18 to 25 kDa without any tag (Table 2.1). The vectors were named *pET-DBE*₁₆₋₁₉₇, *pET-DBE*₂₄₋₁₉₇, *pET-DBE*₂₆₋₁₉₇, *pET-DBE*₃₄₋₁₉₇, *pET-DBE*₁₆₋₂₁₃, *pET-DBE*₂₆₋₂₁₃, *pET-DBE*₃₄₋₂₁₃, *pET-DBE*₁₆₋₂₂₉, *pET-DBE*₂₆₋₂₂₉, *pET-DBE*₃₄₋₂₂₉, *pET-DBE*₁₋₁₉₇ and *pET-DBE*₁₋₂₂₉.

2.2.2 Construction of *pET-Fx-DBE*₁₆₋₁₉₇ and *pET-Fx-DBE*₂₄₋₁₉₇

DNA fragments encoding *dbe*₁₆₋₁₉₇ and *dbe*₂₄₋₁₉₇ were generated by PCR using a forward primer DBE_V16f_Fx/ DBE_I24f_Fx and DBE_P197r_*Xho*I (Section 2.10.1).

The vector *pET-DBE* was used as a PCR template. The DNA polymerase *Pfu* (Promega) was used for PCR and the condition and thermocycling profile was listed

in Section 2.1.4. The PCR products were digested with *NdeI* (NEB) and *XhoI* (NEB) at 37 °C for 3 hours. The digested PCR products were ligated to *NdeI/XhoI* digested *pET507a* vector by T4 DNA ligase (NEB) at 16°C overnight. The ligation products were transformed into *E. coli* strain *DH5α* competent cells (Section 2.1.1). Positive clones were selected by PCR screening and *NdeI/XhoI* digestion and the insert identities were confirmed by DNA sequencing (TechDragon Ltd). The resulting plasmids encoded a 23 kDa/ 22 kDa protein with poly-histidine (6X) tag, a Factor Xa recognition sequence, and DBE₁₆₋₁₉₇/ DBE₂₄₋₁₉₇. The vectors were named *pET-Fx-DBE₁₆₋₁₉₇* and *pET-Fx-DBE₂₄₋₁₉₇*.

2.2.3 Construction of *pET-HS-DBE₁₆₋₁₉₇*, *pET-HS-DBE₂₄₋₁₉₇* and *pET-HS-DBE₁₆₋₂₂₉*

The DNA fragments encoding *dbe₁₆₋₁₉₇* and *dbe₂₄₋₁₉₇* were generated by PCR using forward primer DBE_V16f_AgeI/ DBE_I24f_AgeI and reverse primer DBE_P197r_EcoRI (Section 2.10.1).

The vector *pET-DBE* was used as a PCR template. The DNA polymerase *Pfu* (Promega) was used for PCR and the condition and thermocycling profile was listed in Section 2.1.4. The PCR products were digested with *AgeI* (NEB) and *EcoRI* (NEB)

at 37°C for 3 hours. The digested PCR products were ligated into an *AgeI/EcoRI* digested *pRSETa-SUMO* vector by T4 DNA ligase (NEB) at 16°C overnight. The ligation products were transformed into *E. coli DH5α* competent cells (Section 2.1.1). Positive clones were selected by PCR screening and *AgeI/EcoRI* digestion and insert identities were checked by DNA sequencing (TechDragon Ltd). The resulting plasmids encoded a protein containing DBE₁₆₋₁₉₇/ DBE₂₄₋₁₉₇ with an upstream poly-histidine (6X) and a SUMO tag (HS tag).

The fragments encoding *hs-dbe*₁₆₋₁₉₇ or *hs-dbe*₂₄₋₁₉₇ were then cloned into *pET507a* vector using forward primer HSf and reverse primer DBE_P197r_ *XhoI* (Section 2.10.1). The DNA polymerase *Pfu* (Promega) was used for PCR and the condition and thermocycling profile was listed in Section 2.1.4. The PCR products were digested with *NcoI* (NEB) and *XhoI* (NEB) at 37°C for 3 hours. The digested PCR products were ligated into a *NcoI/XhoI* digested *pET507a* vector by T4 DNA ligase (NEB) at 16°C overnight. The ligation products were transformed into *E. coli* strain *DH5α* competent cells. Positive clones were selected by PCR screening and *NcoI/XhoI* digestion and the insert identities were confirmed by DNA sequencing (TechDragon Ltd). The resulting plasmids encoded a 32 kDa/ 31 kDa protein containing DBE₁₆₋₁₉₇/ DBE₂₄₋₁₉₇ with an upstream HS tag. The vectors were named as *pET-HS-DBE*₁₆₋₁₉₇

and *pET-HS-DBE*₂₄₋₁₉₇.

A DNA fragment encoding *dbe*₁₆₋₂₂₉ was generated by PCR using primer DBE_V16f_AgeI and DBE_K229r_XhoI (Sectoin 2.10.1). The vector *pET-DBE* was used as a PCR template and *Pfu* was used for PCR and the condition and the thermocycling profile was listed in Section 2.1.4. The PCR products were digested with *AgeI* and *XhoI* at 37°C for 3 hours. The digested PCR products were ligated to *AgeI/XhoI* digested *pET-HS-DBE*₁₆₋₁₉₇ vector by T4 DNA ligase at 16°C overnight. The ligation products were transformed into *E. coli* strain *DH5α* competent cells. Positive clones were selected by PCR screening and *AgeI/XhoI* digestion and the insert identity was checked by DNA sequencing (TechDragon Ltd). The vector was named as *pET-HS-DBE*₁₆₋₂₂₉.

2.2.4 Construction of *pET-HS-DBE*₁₆₋₁₉₇ point mutants

Three point mutants of *dbe*₁₆₋₁₉₇, i.e. *R87K*, *K97R* and *K138R* were generated by quick change mutagenesis. The vector *pET-HS-DBE*₁₆₋₁₉₇ was used as the PCR template. Phusion® High Fidelity Polymerase (Finnzymes) with GC buffer was used for PCR reaction with 18 thermocycles of 98°C for 30 seconds, 50°C for 30 seconds and 72°C for 3.5 minutes. Three pairs of specific mutagenesis primers were used,

namely R87Kf, R87Kr, K97Rf, K97Rr, K138Rf and K138Rr (Section 2.10.1). The mutant *R87K* was generated by primer pair R87Kf/ R87Kr, *K97R* by K97Rf/ K97Rr and *K138R* by K138f/ K138r. The PCR products were digested with *DpnI* (NEB) at 37°C for 3 hours. The digestion products were transformed into *E. coli* strain *DH5α* competent cells (Section 2.1.1). Positive clones were selected by PCR and *AgeI/XhoI* digestion and the mutagenesis were confirmed by DNA sequencing (BGI). The vectors were named as *pET-HS-DBE₁₆₋₁₉₇_R87K*, *pET-HS-DBE₁₆₋₁₉₇_K97R* and *pET-HS-DBE₁₆₋₁₉₇_K138R*.

Three point mutants of *dbe₁₆₋₁₉₇*, i.e. *QE29A*, *EE38* and *KKE135A* were generated by overlapping PCR. Three pairs of specific mutagenesis primer were used, namely QE29Af, QE29Ar, EE38Af, EE38Ar, KKE135Af and KKE135Ar (Section 2.10.1). Three rounds of PCR reactions were performed for each mutagenesis and *pET-HS-DBE₁₆₋₁₉₇* was used as PCR template. The first round of PCR used primers DBE_V16f_AgeI (Section 2.10.1) and QE29Ar/ EE38Ar/ KKE135Ar while the second used primers DBE_P197r_XhoI (Section 2.10.1) and QE29Af/ EE38Af/ KKE135Af. Phusion® High Fidelity Polymerase (Finnzymes) was used for PCR reaction with 23 thermocycles of 98°C for 20 seconds, 45°C for 15 seconds and 72°C for 30 seconds. The third round of PCR used DBE_V16f_AgeI and DBE_P197r_XhoI

as primers and 0.2 μ l each of PCR products from the first and second round PCR as template with 25 thermocycles of 98°C for 20 seconds, 45°C for 15 seconds and 72°C for 30 seconds. The PCR products were digested with *Age*I and *Xho*I at 37°C for 3 hours. The digested PCR products were ligated into an *Age*I/*Xho*I digested *pET-HS-DBE*₁₆₋₁₉₇ vector by T4 DNA ligase (NEB) at 16°C overnight. The ligation products were transformed into *E. coli* strain *DH5 α* competent cells. Positive clones were selected by PCR screening and *Age*I/*Xho*I digestion and the insert identities were checked by DNA sequencing (BGI). The vectors were named as *pET-HS-DBE*_{16-197_QE29A}, *pET-HS-DBE*_{16-197_EE38A} and *pET-HS-DBE*_{16-197_KKE135A}.

2.2.5 Construction of *pET-HS-Krr1p*₁₃₋₁₉₆

Preparation of Krr1p specific yeast cDNA

100 μ l of overnight *Saccharomyces cerevisiae* culture was mixed with 1 ml Trizol reagent (Invitrogen). The mixture stood at room temperature for 5 minutes, followed by addition of 200 μ l of chloroform and mixed by vortexing. The mixture was then allowed to stand at room temperature for 5 minutes. Phase separation was achieved by centrifugation at 12,000g (model 5415R, Eppendorf), 4°C for 10 minutes. The upper aqueous phase was transferred to a new tube, followed by addition of 500 μ l of isopropanol and mixed by vortexing. The mixture stood at -20°C for 1 hour and

centrifuged at 12,000g, 4°C for 10 minutes. Supernatant was discarded and the pellet was washed by 70% ethanol. The pellet was then allowed to air dry and then dissolved in 20 µl of DEPC-treated water.

The dissolved RNA was then mixed with 1 µl of 50 mM *krr1*-specific reverse primer Krr1p_P196r_*XhoI* (Section 2.10.2) and the mixture was denatured at 70°C for 5 minutes. The mixture was then cooled rapidly on ice. 1 µl of ImProm-II™ Reverse Transcriptase (Promega), 2.4 µl of 25 mM magnesium chloride, 0.5 µl of 25 mM dNTP, 0.5 µl of RNase OUT was added to the mixture. The mixture was then incubated at 25°C for 5 minutes, then 42°C for 1 hour, followed by denaturation at 72°C for 15 minutes. The cDNA was then used for PCR to generate *krr1*₁₃₋₁₉₆ DNA fragment.

The *krr1*₁₃₋₁₉₆ fragment was generated by PCR using primer Krr1p_T13f_*AgeI* and Krr1p_P196r_*XhoI* (Section 2.10.2). The cDNA described above was used as a PCR template and Phusion® High Fidelity Polymerase (Finnzymes) was used for the PCR reaction and the condition and thermocycling profile was listed in Section 2.1.4. The PCR products were digested with *AgeI* (NEB) and *XhoI* (NEB) at 37°C for 3 hours. The digested PCR products were ligated into an *AgeI/XhoI* digested

*pET-HS-DBE*₁₆₋₁₉₇ vector by T4 DNA ligase (NEB) at 16°C overnight. The ligation products were transformed into *E. coli* strain *DH5α* competent cells (Section 2.1.1). Positive clones were selected by PCR screening and *AgeI/XhoI* digestion and the insert identity was checked by DNA sequencing (BGI). The resulting plasmid encoded a 32 kDa protein of a HS tag and *Krr1p*₁₃₋₁₉₆ and was named as *pET-HS-Krr1p*₁₃₋₁₉₆.

2.2.6 Construction of *pET-HS-HRB2*₂₈₋₂₁₅

DNA fragments encoding *hrb2*₂₈₋₂₁₅ were generated by PCR using primers HRB2_Q28f_ *AgeI* and HRB2_P215r_ *XhoI* (Section 2.10.2).

The vector *pET-HRB2* was previously cloned and was used as a PCR template. Phusion® High Fidelity Polymerase (Finnzymes) was used for PCR reaction and the condition and thermocycling profile was listed in Section 2.1.4. The PCR products were digested with *AgeI* (NEB) and *XhoI* (NEB) at 37°C for 3 hours. The digested PCR products were ligated into an *AgeI/XhoI* digested *pET-HS-DBE*₁₆₋₁₉₇ by T4 DNA ligase (NEB) at 16°C overnight. The ligation products were transformed into *E. coli* strain *DH5α* competent cells (Section 2.1.1). Positive clones were selected by PCR screening and *AgeI/XhoI* digestion and the insert identity was confirmed by DNA

sequencing (BGI). The resulting plasmid encoded a 32 kDa protein with a HS tag and HRB2₂₈₋₂₁₅ and was named *pET-HS-HRB2₂₈₋₂₁₅*.

2.2.7 Construction of *pET-RpS9* and *pET-HS-RpS9*

The DNA fragments encoding *rps9* were generated by PCR using primer pairs RpS9f_*NdeI* and RpS9r_*XhoI* for cloning *rps9* with no tag, or RpS9f_*AgeI* and RpS9r_*XhoI* for cloning *rps9* with a HS tag (Section 2.10.3). The *rps9* sequence was previously cloned in *pGEX-RpS9*, which was used as a PCR template, and Phusion® High Fidelity Polymerase (Finnzymes) was used for PCR reaction. The PCR condition and thermocycling profile was listed in Section 2.1.4. The PCR products were digested with *NdeI* (NEB) and *XhoI* (NEB)/ *AgeI* (NEB) and *XhoI* at 37°C for 3 hours. The digested PCR products were ligated into a *NdeI/XhoI* digested *pET507*, or *AgeI/XhoI* digested *pET-HS-DBE₁₆₋₁₉₇* by T4 DNA ligase (NEB) at 16°C overnight. The ligation products were transformed into *E. coli* strain *DH5α* competent cells (Section 2.1.1). Positive clones were selected by PCR screening and *AgeI/XhoI* digestion and the insert identities were confirmed by DNA sequencing (BGI). The resulting plasmids that encoded a 23 kDa protein RpS9 with no tag and a 32 kDa protein with a HS tag and RpS9 and was named *pET-RpS9* and *pET-HS-RpS9*.

2.2.8 Construction of *pCola-DBE₁₆₋₂₂₉*

The DNA fragment encoding *dbe₁₆₋₂₂₉* was generated by PCR using primers DBE_V16f_ *Nde*I and DBE_P197r_ *Xho*I (Section 2.10.1). Phusion® High Fidelity Polymerase (Finnzymes) was used for PCR reaction and the condition and thermocycling profile was listed in Section 2.1.4. The PCR products encoding *dbe₁₆₋₂₂₉* were digested with *Nde*I (NEB) and *Xho*I (NEB) at 37°C for 3 hours. The digested PCR products were ligated into a *Nde*I/*Xho*I digested *pCola* by T4 DNA ligase at 16°C overnight. The ligation products were transformed into *E. coli* strain *DH5α* competent cells. Positive clones were selected by PCR screening and *Nde*I/*Xho*I digestion and the insert identity was confirmed by DNA sequencing (BGI). The resulting plasmid encoded for RpS9 and DBE₁₆₋₂₂₉ and named as *pCola-DBE₁₆₋₂₂₉*.

2.3 Protein expression

2.3.1 Native protein expression

E. coli strain *BL21 (DE3) pLysS* or *C41* was transformed with the protein expression vectors. Expression vector *pET-SENPIc* was previously cloned to express a SUMO-specific protease that was needed in the course of purification of HS tagged protein. All other expression vectors were described in Section 2.2. The expression

conditions for various vectors are listed in Table 2.2. Transformed *BL21* or *C41* cells were selected by agar plate containing appropriate antibiotics (Section 2.9.1). A single colony was inoculated into 10 ml LB medium with appropriate antibiotics (Section 2.9.1) and the culture medium was shaken for about 5 hours. The 10 ml starter culture was then inoculated into 1 L of LB medium with appropriate antibiotics. The cells were grown at 37°C with shaking at 250 rpm. The cells were allowed to grow overnight at 20°C with shaking at 250 rpm. Cell harvest was performed by centrifugation at 3,700 x *g* (Beckman rotor, JLA 16.25) for 5 minutes at 4°C. The cell pellet was stored at - 80°C

Table 2.2 Expression vectors and their corresponding expression conditions

Vector	<i>E. coli</i> strain	OD ₆₀₀ readings	IPTG concentration	Antibiotics used
<i>pET-DBE</i> and deletion mutants (Section 2.2.1 and 2.2.2)	<i>BL21</i>	0.4	0.1 mM	Ampicillin and chloramphenicol
<i>pET-HS-DBE</i> ₁₆₋₁₉₇ (Section 2.2.3) and point mutants (Section 2.2.4)	<i>BL21</i>	0.4	0.1 mM	Ampicillin and chloramphenicol
<i>pET-HS-DBE</i> ₁₆₋₂₂₉ and <i>pET-HS-DBE</i> ₂₄₋₁₉₇ (Section 2.2.3)	<i>C41</i>	0.1	0.05 mM	Ampicillin
<i>pET-HS-Krr1p</i> ₁₃₋₁₉₆ and <i>pET-HS-HRB2</i> ₂₈₋₂₁₅ (Section 2.2.5 and 2.2.6)	<i>C41</i>	0.3	0.1 mM	Ampicillin
<i>pET-RpS9/ pET-HS-RpS9</i> (Section 2.2.7)	<i>C41</i>	0.6	1 mM	Ampicillin
<i>pCola-DBE</i> ₁₆₋₂₂₉ and <i>pET-RpS9/ pET-HS-RpS9</i> co-expression (Section 2.2.7 and 2.2.8)	<i>C41</i>	0.6	1 mM	Kanamycin, ampicillin
<i>pET-SEN1c</i>	<i>BL21</i>	0.6	0.1 mM	Ampicillin and chloramphenicol

2.3.2 Expression of selenomethionyl DBE₁₆₋₁₉₇

The *E. coli* strain *B834S* was constructed. It was derived from *E. coli* strain *B834* which was transformed with a *pLysS* plasmid. The *B834S* competent cells were prepared as described in Section 2.1.1.

The strain *B834S* was transformed with *pET-HS-DBE₁₆₋₁₉₇*. Transformed cells were selected by LBAC agar plates (Section 2.9.1). A single colony was inoculated into 120 ml LBAC medium (Section 2.9.1) as starter culture. The starter culture was shaken at 37°C and at 250 rpm until OD₆₀₀ reading reached 2.5. The cell pellet was obtained by centrifugation of the culture medium at 3,700g (Beckman rotor, JLA 16.25) for 10 minutes at room temperature. Supernatant was discarded and the cell pellet was washed with 2 ml of M9 medium (Molecular Dimensions). The pellet was then resuspended in 2 ml of M9 minimal medium and transferred to 1 L of M9 medium such that OD₆₀₀ readings reached 0.2. 4 ml of SelenoMethionine solution (Molecular Dimensions) and 50 ml of nutrient supplements (Molecular Dimensions) were added. The culture was then shaken at 37°C and at 250 rpm for 20 minutes. Protein expression was then induced by adding 0.05 mM IPTG (Section 2.9.1). The cells were allowed to grow overnight at 20°C at 250 rpm. Cells were harvested by centrifugation at 3,700g (Beckman rotor, JLA 16.25) for 5 minutes at 4°C. The cell pellet was stored at -80°C.

2.3.3 GST-RpS9 and GST

The *E. coli* strain *C41* was transformed with *pGEX-RpS9* to express GST-RpS9 or *pGEX-6p-1* to express GST alone. Transformed C41 cells were selected by LBA agar

plate (Section 2.9.1). A single colony was inoculated into 10 ml LBA medium (Section 2.9.1) and the culture was shaken for about 5 hours. The 10 ml starter culture was then inoculated into 1 L of LBA medium supplemented with 2% glucose. The cells were grown at 37°C with shaking at 250 rpm until the OD₆₀₀ reading reached 0.4. Protein expression was induced by adding 0.2 mM IPTG (Section 2.9.1) to the medium. For GST-RpS9 expression, the cells were allowed to grow overnight at 22°C with shaking at 250 rpm, whereas for GST expression, the cells were allowed to grow at 30°C for 4 hours. Cell harvest was performed by centrifugation at 3,700g (Beckman rotor, JLA 16.25) for 5 minutes at 4°C. The cell pellet was stored at - 80°C.

2.4 Protein Purification

2.4.1 Purification of SENP1c

The cell pellet was disrupted by sonication in 50 ml of GST affinity column binding buffer GST_A (Section 2.9.4). Cells were sonicated for 10 minutes at 1 minute pulse at 80 % output, pulse for 4 seconds, on ice using normal probe (Sonic & Materials, Vibra Cell). After sonication, lysed cells were centrifuged at 19,000g (Beckman rotor, JA 20) for 30 minutes at 4°C. The supernatant was collected for column purification. A GSTrap[™] FF column (GE Healthcare) was connected to ÄKTA prime machine (GE Healthcare) and equilibrated with buffer GST_A. The

sample was introduced into the column at 0.8 ml/minute. The column was cleaned of unbounded cellular proteins by washing the column with buffer GST_A until the OD₂₈₀ reading became steady. SENP1c protein was eluted with elution buffer GST_B (Section 2.9.4). Aliquots of 250 µl of protein solution were made and frozen in liquid nitrogen and stored in -80°C until used.

2.4.2 Purification of tag-less DBE and DBE deletion mutants

Tag-less DBE and DBE deletion mutants expressed from *pET* expression vectors (Section 2.2.1 and 2.3.1) were purified by the following procedures.

The cell pellet was disrupted by sonication in 50 ml of cation exchanger binding buffer SP_A (Section 2.9.4). Cells were sonicated for 30 minutes using a 1 minute pulse at 80% output on ice using normal probe (Sonic & Materials, Vibra Cell). The lysed cells were centrifuged at 19,000g (Beckman rotor, JA 20) for 30 minutes at 4°C.

Supernatant was collected for column purification. A HiTrap[™] SP HP column (GE Healthcare) was connected to ÄKTA prime machine (GE Healthcare) and equilibrated with buffer SP_A. The sample was introduced into the column at 5 ml/minute. The column was cleaned of unbounded cellular proteins by washing the column with

buffer SP_A until the OD₂₈₀ reading became steady. A 100 ml linear gradient of 0 to 1 M NaCl with elution buffer SP_B (Section 2.9.4) was applied at a flow rate of 5 ml/minute and fractions of 5 ml were collected. The protein was eluted at about 0.5 M NaCl.

Fractions containing target proteins were pooled and dialyzed against 2 L of hydrophobic interaction column binding buffer Phenyl_A (Section 2.9.4) at 4°C overnight. The protein sample was then ready for column purification. A HiTrap^{1m} Phenyl HP column (GE Healthcare) was connected to ÄKTA prime machine (GE Healthcare) and equilibrated with buffer Phenyl_A. The sample was introduced into the column at 5 ml/minute. Unbound proteins were removed by washing the column with buffer Phenyl_A until OD₂₈₀ reading became steady. A 100 ml linear gradient of 2 to 0 M NaCl with elution buffer Phenyl_B (Section 2.9.4) was applied at a flow rate of 5 ml/minute and fractions of 5 ml were collected. The protein was eluted at about 0.1 M NaCl.

Fractions containing target proteins were pooled and concentrated to about 4 ml before loading to a HiLoad 26/60 Superdex 75 column (GE Healthcare) connected to ÄKTA prime machine (GE Healthcare) pre-equilibrated with running buffer GF_1

(Section 2.9.4). The flow rate was kept at 3 ml/minute and the elution volume was about 180 ml.

2.4.3 Purification of HS tagged proteins

The HS tagged protein including DBE₁₆₋₁₉₇, DBE₁₆₋₁₉₇ point mutants, DBE₂₄₋₁₉₇, DBE₁₆₋₂₂₉, Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ were expressed to aid purification of the target proteins (Sections 2.2.3, 2.2.4, 2.2.5, 2.2.6 and 2.3.1). The HS tag was removed during the course of the purification; no extra residues were left on the target proteins. Purification of DBE₁₆₋₁₉₇, DBE₁₆₋₁₉₇ point mutants, DBE₂₄₋₁₉₇ and DBE₁₆₋₂₂₉ involved the use of buffers Ni_A, Ni_B and GF_1 (Section 2.9.4), while that of Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ involved the use of buffers Ni_C, Ni_D and GF_2 (Section 2.9.4).

The cell pellet was disrupted by sonication in 50 ml of Ni²⁺ affinity column binding buffer Ni_A or Ni_C. Cells were sonicated for 30 minutes at 1 minute pulse at 80% output on ice using normal probe (Sonic & Materials, Vibra Cell). The lysed cells were centrifuged at 19,000g (Beckman rotor, JA 20) for 30 minutes at 4°C. Supernatant was collected for Ni²⁺ affinity column purification. A HiTrap[™] IMAC HP column (GE Healthcare) was connected to ÄKTA prime machine (GE Healthcare) and equilibrated with Ni²⁺ ion solution (Section 2.9.4). The sample was introduced

into the column at 5 ml/minute. The column was cleaned of unbound cellular proteins by washing the column with buffer Ni_A or Ni_C until OD₂₈₀ reading became steady. A 100 ml linear gradient of 0 to 300 mM imidazole in elution buffer Ni_B or Ni_D was applied at a flow rate of 5 ml/minute and fractions of 5 ml were collected. The protein was eluted at about 200 mM imidazole.

Fractions containing the target proteins were pooled. 250 µl of SENP1c (Section 2.4.1) was added to the pooled fractions to cleave the HS tag from the target protein. The cleavage reaction was performed at room temperature for 25 minutes. The cleavage product was diluted 1 fold with binding buffer Ni_A or Ni_C before loading onto column to separate the HS tag from target proteins. A HiTrap[™] IMAC HP column (GE Healthcare) was connected to ÄKTA prime machine (GE Healthcare) and equilibrated with Ni²⁺ ion solution. The sample was introduced into the column at 5 ml/minute. Target proteins were collected as flowthrough, and the HS tag was eluted in buffer Ni_B or Ni_D.

The target protein was concentrated to about 4 ml before loading to a HiLoad 26/60 Superdex 75 column (GE Healthcare) connected to ÄKTA prime machine (GE Healthcare) pre-equilibrated with running buffer GF_1 or GF_2. The flow rate was

kept at 3 ml/minute and the elution volume was about 180 ml.

2.5 Biophysical characterization of DBE₁₆₋₁₉₇

2.5.1 Circular dichroism (CD) spectroscopy

Far-UV CD spectrum

Protein samples (20 μ M) were equilibrated in 10 mM NaH₂PO₄, pH 7.4. Molar ellipticity from far-UV of wavelength 190 nm to 260 nm was measured at 25°C by 1 mm path length cuvette using JASCO J810 spectropolarimeter equipped with a Peltier-type temperature control unit. Three measurements for each sample were taken and averaged. Measurements were analyzed by the program Spectra Manager. The secondary content was estimated by deconvolution of the CD spectrum using the program CDNN.

Thermal denaturation profile

Protein samples (20 μ M) were equilibrated in 10 mM NaH₂PO₄, pH 7.4 and thoroughly degassed. The samples were heated in a 1 mm path length cuvette from 25°C to 110°C at a heating rate of 1 °C/minute using JASCO J810 spectropolarimeter equipped with a Peltier-type temperature control unit. The denaturation profile was monitored at wavelength 222 nm. The cuvette was securely stoppered to ensure there

was no loss in volume due to evaporation. Three measurements for each sample were taken and averaged. Measurements were analyzed by the program Spectra Manager.

2.5.2 Dynamic Light Scattering

Dynamic light scattering measurement was powered by Dynapro Titan (Wyatt). Dynapro Titan was connected to software Dynamics 6.7.6. All buffers and samples were filtered through a 0.2 μm filter and all containers and instruments were dusted to minimize measurement interference. 20 μl of protein sample in GF_1 (Section 2.9.4) was loaded to a cuvette for measurement. Ten acquisitions, with 10 seconds per acquisition, were taken for each measurement. Measurements were analyzed by Dynamics 6.7.6.

2.5.3 Static Light Scattering

An analytical gel filtration column Superdex 75 (GE Healthcare) was connected to miniDawn Tristar light scattering detector and Optilab DSP refractometer (Wyatt). All buffers used in static light scattering were filtered through 0.1 μm filter. A flow rate of 0.5 ml/minute was maintained throughout light scattering experiment. The analytical gel filtration column was equilibrated with running buffer GF_1 (Section 2.9.4) for 16 hours. Then the column was connected to the miniDawn Tristar and Optilab DSP

refractometer (Wyatt) and equilibrated for another 3 hours until there was no difference in refractive index between the sample cell and reference cell of the refractometer. 100 μ l of 5 mg/ml protein sample was injected and the light scattering intensity and the refractive index of the protein were monitored by the software Astra V. Two-detectors, a light scattering detector and a refractive index detector, approximation was used to determine the molecular weight.

2.6 Protein crystallization condition screening

2.6.1 Screening for crystallization condition of DBE₁₆₋₁₉₇

The purified DBE₁₆₋₁₉₇ fragment (Section 2.4.3) was concentrated to 10 mg/ml under the gel filtration running buffer GF_1 (Section 2.9.4) for protein crystallization screening. Screening for crystallization condition was carried out using the Crystal Screen 1 and 2, Index, SaltRx, PEG/ion screen kits (Hampton Research) and JCSG+, JCSG core I, II, III, IV (QIAGEN), at 16°C using sitting drop on Greiner CrystalQuick plates with 3 protein to well solution ratios, 0.1 μ l: 0.2 μ l, 0.15 μ l: 0.15 μ l and 0.2 μ l: 0.1 μ l using PhoenixTM RE system (Rigaku). Only one initial hit was observed, the hit was 0.4 M ammonium phosphate (monobasic) from Crystal Screen 1. Extensive gridding followed and the optimal condition for crystal growth was found to be using a 24-well sitting drop plate, with 500 μ l of well solution of DBE_Crys

(Section 2.9.5), i.e. 0.6 M ammonium phosphate, pH 4.8, protein to well solution ratio of 2 μ l: 4 μ l at 16°C. Crystals of maximum size could be observed after 4 days of incubation.

2.6.2 Screening for crystallization condition of DBE₁₆₋₂₂₉, HS-DBE₁₆₋₁₉₇ and DBE₁₆₋₁₉₇ point mutants

The DBE₁₆₋₂₂₉, HS-DBE₁₆₋₁₉₇ and the six DBE₁₆₋₁₉₇ point mutants R87K, K97R, K138R, QE29A, EE38A and KKE135A (Section 2.4.3) were concentrated under the gel filtration running buffering GF_1 (Section 2.9.4). Purified DBE₁₆₋₂₂₉ was concentrated to 6 mg/ml and a slight precipitation was observed. The HS-DBE₁₆₋₁₉₇ protein was concentrated to 8 mg/ml with a moderate precipitation. The DBE₁₆₋₁₉₇ mutants QE29A, EE38A and KKE135A were concentrated to 6 mg/ml with moderate precipitation, while mutants R87K, K97R, K138R were concentrated to 10 mg/ml without precipitation. Screening for crystallization condition was carried out using Crystal Screen 1 and 2, Index, SaltRx, PEG/ion, Natrrix (Hampton Research) and JCSG+, JCSG core I, II, III, IV (QIAGEN), at 16 °C using sitting drop on Greiner CrystalQuick plates with 3 protein to well solution ratios, 0.1 μ l: 0.2 μ l, 0.15 μ l: 0.15 μ l and 0.2 μ l: 0.1 μ l using PhoenixTM RE system (Rigaku). No crystals were observed.

2.6.3 Crystallization condition screening of Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅

Purified Krr1p₁₃₋₁₉₆ or HRB2₂₈₋₂₁₅ fragment was concentrated to 12 mg/ml using the gel filtration running buffer GF_2 (Section 2.9.4) for protein crystallization screening. Screening for crystallization condition was carried out using the Crystal Screen 1 and 2, Index, SaltRx, PEG/ion, Natrix (Hampton Research) and JCSG+, JCSG core I, II, III, IV (QIAGEN), at 16°C using sitting drop on Greiner CrystalQuick plates with 3 protein to well solution ratios, 0.1 µl: 0.2 µl, 0.15 µl: 0.15 µl and 0.2 µl: 0.1 µl using PhoenixTM RE system (Rigaku). Several hits were observed for both proteins.

For Krr1p₁₃₋₁₉₆, 6 initial hits were observed and they are listed in the following table.

Initial crystallization conditions of Krr1p₁₃₋₁₉₆

Crystallization screen	Components	Reproducible in 24-well plate?
Natrix	100 mM KCl, 10 mM MgCl ₂ , 50 mM Tris-Cl, pH 8.5, 30 % (v/v) PEG400	Yes
JCSG +	200 mM Mg formate, 20 % (w/v) PEG3350	No
JCSG I	50 mM Li ₂ SO ₄ , Na ₂ SO ₄ , 50 mM Tris-Cl, pH 8.5, 30 % (v/v) PEG400	Yes
JCSG II	200 mM NaCl, 100 mM Na/K H ₂ PO ₄ pH 6.2, 20 % (w/v) PEG1000	No
Index	50 mM CaCl ₂ , 100 mM BisTris, pH 6.5, 30 % (v/v) PEG500 monomethyl ether	No
Index	50 mM MgCl ₂ , 100 mM HEPES, pH 7.5, 30 % (v/v) PEG500 monomethyl ether	No

Only 2 of the initial hits can be reproduced in 24-well sitting drop plates. Extensive gridding followed and the best condition for crystal growth was found by using sitting drop, with 500 µl of well solution of Kr_Crys (Section 2.9.5) i.e. 36% (v/v) PEG400, 50 mM Tris-Cl pH 9.0, 50 mM Li₂SO₄, 50 mM Na₂SO₄ and 100 mM MgCl₂, protein

to precipitant ratio of 4 μ l: 2 μ l at 16°C. Crystals of maximum size could be observed 7 days post incubation.

For HRB2₂₈₋₂₁₅, 3 initial hits were observed and they are listed in the following table.

Initial crystallization conditions of HRB2₂₈₋₂₁₅

Crystallization screen	Components
JCSG +	200 mM NaCl, 100 mM Na/K H ₂ PO ₄ , pH 6.2, 50 % (v/v) PEG200
JCSG IV	100 mM HEPES, pH 7.5, 5 % (w/v) PEG3000, 40 % (v/v) Ethylene glycol
Index	50 mM CaCl ₂ , 100 mM BisTris, pH 6.5, 30 % (v/v) PEG500 monomethyl ether

2.7 X-ray diffraction analysis and phasing

2.7.1 DBE₁₆₋₁₉₇ crystals

2.7.1.1 Native DBE₁₆₋₁₉₇ crystals

A single plate-like crystal was transferred from mother liquor and immersed in cryoprotectant solution Cryo_1 (Section 2.9.5) for 1 minute before X-ray diffraction.

X-ray diffraction data were collected on beamline I04 at the Diamond Light Source

(UK) with a wavelength of 0.9704 Å at 100 K using an ADSC Q315 detector. Data processing was performed using *MOSFLM* (Leslie, 2006) and *SCALA* (Evans, 2006) via *CCP4i* suite (Potterton *et al.*, 2003).

2.7.1.2 Heavy atom soaking

Twelve heavy atom compounds from JBScreen Heavy (Jena Bioscience) heavy atom screening kits (listed in Table 2.3), samarium (III) acetate, lead (II) chloride and cesium chloride were used. Sitting drop vapor diffusion setup was used. The single plates-like DBE₁₆₋₁₉₇ crystals were carefully transferred from their crystallizations drops to heavy atom solution drops. Heavy atom solutions were prepared by adding heavy atom into the DBE₁₆₋₁₉₇ crystallization buffer DBE_Crys (Section 2.9.5). Concentrations of the heavy atom solution ranged from 50 µM to 50 mM and soaking times ranged from 10 minutes to over 2 weeks. Crystals that remained largely intact after soaking were cryoprotected in Cryo_1 (Section 2.9.5) and then subject to X-ray diffraction. X-ray diffraction data were collected with a copper anode generator MicroMaxTM-007 HF (Rigaku) at a wavelength of 1.54 Å, at 100 K using R-AXIS IV⁺⁺ image plates (Rigaku) detector.

Table 2.3 Heavy atom compounds used in the JBScreen

Chemical name	Molecular formulae	Heavy atom
Potassium tetrachloroplatinate (II)	Pt K ₂ Cl ₄	Pt
Potassium dicyanoaurate (I)	KAu(CN) ₂	Au
Dipotassium tetraiodomercurate (II)	K ₂ HgI ₄	Hg
Mercury (II) chloride	HgCl ₂	Hg
Sodium 4-(Chloromercuri)benzenesulfonic acid	C ₆ H ₄ ClHgO ₃ SNa	Hg
Potassium tetranitroplatinate (II)	K ₂ Pt(NO ₂) ₄	Pt
Potassium tetracyanoplatinate (II) trihydrate	K ₂ Pt(CN) ₄	Pt
Mercury (II) acetate	Hg(CH ₃ COO) ₂	Hg
<i>cis</i> -Dichlorodiammine platinum (II)	<i>cis</i> -Pt(NH ₃) ₂ Cl ₂	Pt
Ethyl mercurithiosalicylate (Thimerosal)	C ₉ H ₉ HgNaO ₂ S	Hg
Potassium tetrachloroaurate (III) hydrate	KAuCl ₄	Au
Potassium hexachloroiridate (IV)	K ₂ IrCl ₆	Ir

2.7.1.3 5-amino-2,4,6-triiodoisophthalic acid (I3C)

I3C (Sigma) stock solution was prepared in 1 M LiOH. I3C substitution was tried in 2 different ways, by quick-soaking or by co-crystallization. X-ray diffraction images were collected using a copper anode generator MicroMaxTM-007 HF (Rigaku) at a wavelength of 1.54 Å, at 100 K using R-AXIS IV⁺⁺ image plates (Rigaku) detector.

Quick-soaking

Cryoprotectant solutions Cryo_1 (Section 2.9.5) containing 0.75 M, 0.5 M and 0.35 M of I3C were prepared. Single plate-like DBE₁₆₋₁₉₇ crystals were immersed in the I3C containing Cryo_1 for 1 minute to 10 minutes before X-ray diffraction analysis.

Co-crystallization

Three concentrations of I3C, 0.01 M, 0.05 M and 0.1 M were added to the DBE₁₆₋₁₉₇, crystallization buffer DBE_Crys (Section 2.9.5) for co-crystallization of DBE₁₆₋₁₉₇. The rest of the crystallization conditions were exactly the same as Section 2.6.1. Small crystals were observed in wells containing 0.01 M and 0.05 M of I3C. The crystals were cryoprotected in Cryo_1 for 1 minute before X-ray diffraction analysis.

2.7.1.4 Iodide/ bromide quick soak

Cryoprotectant solutions Cryo_1 (Section 2.9.5) containing 0.75 M, 0.5 M and 0.35 M of potassium iodide or sodium bromide were prepared. Single plate-like DBE₁₆₋₁₉₇ crystals were immersed in the halide containing Cryo_1 from 20 seconds

to 1 minute before X-ray diffraction analysis.

2.7.1.5 Selenomethionyl DBE₁₆₋₁₉₇ crystals

Selenomethionine labeled protein was expressed by *E. coli* strain *B834S* mentioned in Section 2.3.2. Except for adding 1 mM TCEP in buffers Ni_A, Ni_B and GF_1 (Section 2.9.4), purification of the labeled protein was the same as that of the native DBE₁₆₋₁₉₇ mentioned in Section 2.4.3. The crystallization condition was exactly the same as that of the native DBE₁₆₋₁₉₇ protein mentioned in Section 2.6.1. Anomalous diffraction data were collected on beamline BL17U at the Shanghai Synchrotron Radiation Facility (China) at 100K using MAR 225 CCD detector.

2.7.2 Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ crystals

For both crystals, X-ray diffraction images were taken using a copper anode generator MicroMaxTM-007 HF (Rigaku) at a wavelength of 1.54 Å, at 100 K using R-AXIS IV⁺⁺ image plates (Rigaku) detector. For Krr1p₁₃₋₁₉₆, the crystals were mounted for X-ray diffraction with or without cryoprotection. For X-ray diffraction with cryoprotection, several cryoprotectant solutions were prepared by including 30% (v/v) ethylene glycol, 30% (v/v) glycerol and 30% (w/v) xylitol into the crystallization solution Kr_Crys (Section 2.9.5).

Without cryoprotection, X-ray diffraction was carried out at room temperature. The crystals were covered with MicroRT™ capillary (MiTeGen) before they were subjected to X-ray irradiation.

For HRB2₂₈₋₂₁₅, the crystals were covered with MicroRT™ capillary for X-ray diffraction at room temperature.

2.8 GST-RpS9 pull down

The GST and GST-RpS9 proteins were expressed as mentioned in Section 2.3.3. Full length DBE and its deletion mutants were expressed and purified as mentioned in Section 2.3.1 and 2.4.2. A 10 mg/ml lysozyme solution was prepared by dissolving lysozyme powder (Sigma) into distilled water.

A cell pellet containing GST or GST-RpS9 was disrupted by sonication in 15 ml of 1X PBS (Section 2.9.6). Cells were sonicated for 10 minutes at 1 minute pulse at 80% output, pulse for 4 seconds, on ice using normal probe (Sonic & Materials, Vibra Cell). After sonication, lysed cells were centrifuged at 19,000g (Beckman rotor, JA 20) for 30 minutes at 4 °C. Supernatant was collected. 1 µl of supernatant was mixed with

50 μ l of GST beads (Glutathionine Sepharose™ 4 Fast Flow, GE Healthcare). GST or GST-RpS9 was captured by GST beads by rotating the mixture at 4°C for 1 hour. The GST beads were separated from the supernatant by centrifugation at 2,500g (Eppendorf Centrifuge 5415R), 4°C for 30 seconds. Supernatant was removed and the GST beads were washed using 1 ml of Wash buffer (Section 2.9.6) for 5 times. After washing, 1 ml of 1X PBS was added to the GST beads, followed by 20 μ g of protein, including DBE, DBE₁₋₂₂₉, DBE₁₆₋₁₉₇, DBE₃₄₋₁₉₇, DBE₃₄₋₂₂₉ (Section 2.4.2) and lysozyme. The mixture was then rotated at 4°C for 1 hour, followed by washing with 1 ml of 1X PBS for 5 times. Proteins were released from the GST beads by adding 40 μ l of 6X SDS sample buffer (Section 2.9.3) and heated at 99°C for 10 minutes.

2.9 Buffers and reagents

2.9.1 Bacterial culture and competent cells preparation

Ampicillin and Kanamycin stock solution

Ampicillin stock solution was prepared in 100 mg/ml ampicillin (USB) in double distilled water, while kanamycin stock solution was prepared in 50 mg/ml kanamycin (USB) in double distilled water. The solutions were sterilized by filtration through a 0.2 µm filter. The stock solution was kept at -20°C.

Chloramphenicol stock solution

A stock solution of 100 mg/ml chloramphenicol (USB) was prepared by dissolving in absolute ethanol. The stock solution was kept at -20°C.

Luria-Bertani (LB) medium

Luria-Bertani medium was prepared by dissolving 20 g of LB broth powder (USB) in 1 L of double distilled water. The medium was sterilized by autoclaving and stored at room temperature

LB/ LBA/ LBK/ LBC/ LBAC agar plate

Luria-Bertani agar solution was prepared by dissolving 20 g of LB broth powder (USB) and 15 g of bacto-agar (USB) in 1 L of double distilled water. It was sterilized by autoclaving. LB agar plate was made by pouring about 25 ml of LB agar solution into a 90-mm Petri dish and let the solution gel at room temperature. For LBA, LBK, LBAC agar plates, the corresponding antibiotics were added before the LB agar solution gelled. LBA agar plates contained 100 µg/ml ampicillin, LBK agar plates contained 50 µg/ml kanamycin, LBC agar plates contained 100 µg/ml chloramphenicol and LBAC agar plates contained 100 µg/ml each of ampicillin and

chloramphenicol.

Isopropyl β -D-thiogalactopyranoside (IPTG)

A stock solution of 1 M IPTG (Shanghai Biocolors) was prepared in double distilled water. The stock solutions were stored at -20°C .

Buffers for preparation of competent cells

RF1: 30 mM KAc, 100 mM RbCl₂, 10 mM CaCl₂, 50 mM MnCl₂, 15% glycerol.

Acidity was adjusted to pH 5.8 by acetic acid. The buffer was sterilized by filtration through 0.2 μm filter.

RF2: 10 mM MOPS, 75 mM CaCl₂, 10 mM RbCl₂, 15% glycerol. Acidity was adjusted to pH 6.5 with KOH. The buffer was sterilized by filtration through 0.2 μm filter.

2.9.2 Agarose gel electrophoresis buffer

1X Tris-borate-EDTA (TBE) electrophoresis buffer

Stock TBE buffer (5X) contained 0.45 M of Tris-Cl, pH 8.0, 0.45 M of boric acid and 10 mM of EDTA. 1 X TBE electrophoresis buffer was prepared by diluting the stock buffer 5 fold in double distilled water. Both stock and working solutions were kept at room temperature.

6X DNA loading buffer

The loading buffer contained 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, and 40% (v/v) glycerol in autoclaved double distilled water. Stock was kept as aliquots of 1 ml at -20°C .

2.9.3 SDS-PAGE Buffer

6X SDS sample buffer

SDS sample buffer (6X) contained 0.02 % (w/v) of bromophenol blue, 0.1 M of Tris-Cl, pH 6.8, 2 % (w/v) of SDS, 20 % (v/v) of glycerol, and 5 % (v/v) of β -mercaptoethanol (Sigma) in double distilled water. Aliquots of 1.5 ml were stored at -20°C.

1X SDS electrophoresis buffer

25 mM Tris-Cl, pH 8.3, 0.19 M glycine and 1 % (w/v) SDS in double distilled water.

Coomassie staining solution

The solution contained 0.25% Coomassie Brilliant Blue R250 (USB), 45 % (v/v) methanol and 10 % (v/v) acetic acid dissolved in double distilled water.

Coomassie destaining solution

The solution contained 45 % (v/v) methanol and 10 % (v/v) acetic acid dissolved in double distilled water.

2.9.4 Buffers for protein purification

GST_A: Binding buffer for GST affinity column purification containing 20 mM Tris-Cl, pH 7.8, 150 mM NaCl, 5 mM EDTA, 5 mM DTT and 1 mM PMSF.

GST_B: Elution buffer for GST affinity column purification containing 20 mM Tris-Cl, pH 7.8, 150 mM NaCl, 5 mM EDTA, 5 mM DTT, 20 mM reduced

glutathione and 1 mM PMSF.

SP_A: Binding buffer for cation exchanger purification containing 50 mM NaH₂PO₄, pH 7.4 and 1 mM PMSF.

SP_B: Elution buffer for cation exchanger purification containing 50 mM NaH₂PO₄, pH 7.4, 1 M NaCl and 1 mM PMSF.

Phenyl_A: Binding buffer for hydrophobic interaction column purification containing 50 mM NaH₂PO₄, pH 7.4, 2 M NaCl and 1 mM PMSF.

Phenyl_B: Elution buffer for hydrophobic interaction column purification containing 50 mM NaH₂PO₄, pH 7.4, and 1 mM PMSF.

GF_1: Running buffer for gel filtration purification containing 50 mM NaH₂PO₄, pH 7.4 and 0.4 M NaCl.

Ni_A: Binding buffer for Ni²⁺ affinity column purification containing 50 mM NaH₂PO₄, pH 7.4, 0.4 M NaCl, 50 M imidazole and 1 mM PMSF.

Ni_B: Elution buffer for Ni²⁺ affinity column purification containing 50 mM NaH₂PO₄, pH 7.4, 0.4 M NaCl, 0.3 M imidazole and 1 mM PMSF.

Ni²⁺ ion solution: A solution for charging Ni²⁺ to the HiTrap™ IMAC column containing 0.1 M NiCl₂.

Ni_C: Binding buffer for Ni²⁺ affinity column purification containing 20 mM Tris-Cl, pH 7.4, 0.3 M NaCl, 0.05 mM imidazole and 1 mM PMSF.

Ni_D: Elution buffer for Ni²⁺ affinity column purification containing 20 mM Tris-Cl, pH 7.4, 0.3 M NaCl, 0.3 M imidazole and 1 mM PMSF.

GF_2: Running buffer for gel filtration purification containing 20 mM Tris-Cl, pH 7.4 and 0.3 M NaCl, 5 mM DTT.

2.9.5 Buffers for protein crystallization

DBE_Crys: Optimal crystallization condition of DBE₁₆₋₁₉₇ containing 0.6 M ammonium phosphate, pH 4.8.

Cryo_1: Cryoprotectant solution for DBE₁₆₋₁₉₇ crystals containing 0.6 M ammonium phosphate, pH 4.8 and 30 % (v/v) ethylene glycol.

Kr_Crys: Optimal crystallization condition of Krr1p₁₃₋₁₉₆ containing 50 mM Tris-Cl pH 9.0, 50 mM Li₂SO₄, 50 mM Na₂SO₄, 100 mM MgCl₂ and 36 % (v/v) PEG400.

2.9.6 Buffers for GST pull down assay

1X PBS: Binding buffer for GST pull down containing 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄ and 2 mM KH₂PO₄, pH7.4.

Wash buffer: wash buffer to minimize non-specific binding containing 1X PBS and 2 mM reduced glutathionine.

2.10 Primer sequences

2.10.1 Primers for cloning *dbe* deletion constructs and point mutants

DBE_V16f_NdeI: A forward primer encoding DBE fragment that starts with the residue Val16 of DBE. The *NdeI* restriction site is underlined.

5'-AAC TCG CATATG GTG GAC AAT GCG TGG TCC-3'

DBE_I24f_NdeI: A forward primer encoding DBE fragment that starts with the residue Ile24 of DBE. The *NdeI* restriction site is underlined.

5'-AAC TCG CATATG ATC CCT GCC TTC AGG CAG-3'

DBE_A26f_NdeI: A forward primer encoding DBE fragment that starts with the residue Ala26 of DBE. The *NdeI* restriction site is underlined.

5'-AAC TCG CATATG GCC TTC AGG CAG GAG GAC-3'

DBE_H34f_NdeI: A forward primer encoding DBE fragment that starts with the residue His34 of DBE. The *NdeI* restriction site is underlined.

5'-AAC TCG CATATG CAT GGG ATG GTG GAG GAG-3'

DBE_P197r_XhoI: A reverse primer encoding DBE fragment that ends with the residue Pro197 of DBE. The *XhoI* restriction site is underlined.

5'-AAC TCG CTCGAG TTA TCA GGG GTG CAC ATT GTT CAT-3'

DBE_D213r_XhoI: A reverse primer encoding DBE fragment that ends with the residue Asp213 of DBE. The *XhoI* restriction site is underlined.

5'-AAC TCG CTCGAG TTA TCA ATC CTT CAT CAG CTC CCG-3'

DBE_K229r_XhoI: A reverse primer encoding DBE fragment that ends with the residue Lys229 of DBE. The *XhoI* restriction site is underlined.

5'-AAC TCG CTCGAG TTA TCA CTT GAA CTT GGG CAG GAA-3'

T7: A T7 promoter sequence for cloning the DBE fragment that start with the first residue of DBE.

5'-TAA TAC GAC TCA CTA TAG GG-3'

DBE_V16f_Fx: A forward primer encoding DBE fragment that starts with the residue V16 of DBE and Factor Xa recognition sequence. The *NdeI* restriction site is underlined

5'- AAC TCG CATATG ATC GAA GGC CGA GTG GAC AAT GCG TGG TCC-3'

DBE_I24f_Fx: A forward primer encoding DBE fragment that starts with the residue Ile24 of DBE and Factor Xa recognition sequence. The *NdeI* restriction site is underlined

5'- AAC TCG CATATG ATC GAA GGC CGA ATC CCT GCC TTC AGG CAG-3'

DBE_V16f_AgeI: A forward primer encoding DBE fragment that starts with the residue Val16 of DBE. The *AgeI* restriction site is underlined.

5'-AAC TCG ACCGGT GGA GTG GAC AAT GCG TGG TCC

DBE_I24f_AgeI: A forward primer encoding DBE fragment that starts with the residue Ile24 of DBE. The *AgeI* restriction site is underlined.

5'-AAC TCG ACCGGT GGA ATC CCT GCC TTC AGG CAG-3'

DBE_P197r_EcoRI: A reverse primer encoding DBE fragment that ends with the residue Pro197 of DBE. The *EcoRI* restriction site is underlined.

5'-AAC TCG GAATTC TTA TCA GGG GTG CAC ATT GTT CAT-3'

HSf: A forward primer encoding SUMO fragment that starts with the first residue of SUMO tag. The *NcoI* restriction site is underlined.

5'-AAC TCG CCATGG GAA TGC GGG GTT CTC ATC AT-3'

The following primer pair covered the mutagenesis from the residue Arg87 of DBE to lysine.

R87Kf: 5'-GTG GTG AAG ACC AGT AAG AAG ACC TGG GAT CCC-3'

R87Kr: 5'-GGG ATC CCA GGT CTT CTT ACT GGT CTT CAC CAC-3'

The following primer pair covered the mutagenesis from the residue Lys97 of DBE to arginine.

K97Rf: 5'-CCC TAC ATC ATC ATC CGG GCG CGG GAT ATG ATC

K97Rr: 5'-GAT CAT ATC CCG CGC CCG GAT GAT GAT GTA GGG

The following primer pair covered the mutagenesis from the residue Lys138 of DBE to arginine

K138Rf: 5'-GTG CAC AAG AAG GAG CGG TTC GTG AAG CGG CGA-3'

K138Rr: 5'-TCG CCG CTT CAC GAA CCG CTC CTT CTT GTG CAC-3'

The following primer pair covered the mutagenesis from the residues Glu29 and Glu30 of DBE to alanine.

QE29Af: 5'-AAG ATC CCT GCC TTC AGG GCA GCA GAC AAC CCG CAT GGG ATG-3'

QE29Ar: 5'-CAT CCC ATG CGG GTT GTC TGC TGC CCT GAA GGC AGG GAT CTT-3'

The following primer pair covered the mutagenesis from the residues Glu38 and Glu39 of DBE to alanine.

EE38Af: 5'-AAC CCG CAT GGG ATG GTG GCA GCA AGC TCC TTC GCC ACG CTG-3'

EE38Ar: 5'-CAG CGT GGC GAA GGA GCT TGC TGC CAC CAT CCC ATG CGG GTT-3'

The following primer pair covered the mutagenesis from the residues Lys135 and Lys136 and the Glu137 of DBE to alanine.

KKE135Af: 5'- ATC GGA AAC CTT GTG CAC GCA GCA GCA AAG TTC GTG AAG CGG CGA-3'

KKE135Ar: 5'-TCG CCG CTT CAC GAA CTT TGC TGC TGC GTG CAC AAG GTT TCC GAT-3'

DBE_V16f_EcoRI: A forward primer encoding DBE fragment that starts with the residue Val16 of DBE. The *EcoRI* restriction site is underlined.

5'-AAC TCG GAATTC ATG GTG GAC AAT GCG TGG TCC

DBE_V16f_sh_EcoRI: A forward primer encoding DBE fragment that starts with the residue Val16 of DBE. Two nucleotides were added to correct for the frameshifting in

the vector. The *EcoRI* restriction site is underlined.

5'-AAC TCG GAATTC TGG TGG ACA ATG CGT GGT CC-3'

DBE_P197r_HindIII: A reverse primer encoding DBE fragment that ends with the residue Pro197 of DBE. The *HindIII* restriction site is underlined.

5'-AAC TCG AAGCTT TTA TCA GGG GTG CAC ATT GTT CAT-3'

2.10.2 Primers for cloning *krr1*₁₃₋₁₉₆ and *hrb2*₂₈₋₂₁₅

Krr1p_T13f_AgeI: A forward primer encoding Krr1p fragment that starts with the residue Thr16 of Krr1p. The *AgeI* restriction site is underlined.

5'-AAC TCG ACCGGT GGA ACG GAT GAT ATT GAT AAA-3'

Krr1p_P196r_XhoI: A reverse primer encoding Krr1p fragment that ends with the residue Pro196 of Krr1p. The *XhoI* restriction site is underlined.

5'-AAC TCG CTCGAG TTA TCA AGG GTG AAT ATT TTT CAT-3'

HRB2_Q28f_AgeI: A forward primer encoding HRB2 fragment that starts with the residue Glu28 of HRB2. The *AgeI* restriction site is underlined.

5'-AAC TCG ACCGGT GGA CAA GAT GAA TCA GAA CTC-3'

HRB2_P215r_XhoI: A reverse primer encoding HRB2 fragment that ends with the residue Pro215 of HRB2. The *XhoI* restriction site is underlined.

5'-AAC TCG CTCGAG TTA TCA TGG ATG AAT ATT CTT CAT-3'

2.10.3 Primers for cloning *rps9*

RpS9f_AgeI: A forward primer encoding RpS9. The *AgeI* restriction site is underlined.

5'-AAC TCG ACCGGT GGA GTG AAC GGC CGC ATA CCC-3'

RpS9f_NdeI: A forward primer encoding RpS9. The *NdeI* restriction site is underlined.

5'-AAC TCG CATATG GTG AAC GGC CGC ATA CCC-3'

RpS9r_XhoI: A reverse primer encoding RpS9. The *XhoI* restriction site is underlined.

5'-AAC TCG CTCGAG TCA TTA GTC CTC CTC CTC TTC AGC-3'

3. Identification of DBE₁₆₋₁₉₇ as the working construct and biophysical characterization of DBE₁₆₋₁₉₇

3.1 Introduction

It has been estimated by a bioinformatics analysis that 33% of eukaryotic proteins in general, and 36.6% of *Drosophila* proteins in particular, carry long (over 30 amino acid residues) disordered segments (Ward et al.,2004). Ribosome biogenesis, nucleic acid binding and nucleolus localization, as some of the functional keywords, are found to be correlated with the occurrence of long disordered segments (Ward et al.,2004). The finding could be interpreted as that a large number of protein factors contributing to early ribosome biogenesis contain long stretch of disordered amino acid residues. Dribble is one of those proteins, and yet DBE carries ~155 disordered amino acid residues (Yiu et al.,2006), much longer than the 30 amino acid residues mentioned by Ward et al. 2004.

Dribble protein was estimated to contain about 45% random coils and 55% secondary structures by deconvolution of CD spectrum (Yiu et al.,2006). The estimated secondary structural content corresponds to about 190 amino acid residues of the 345 amino acids residues of the full length protein. The secondary structural content agrees with protein database searches in *Pfam* (Finn et al.,2008) and

Superfamily (Wilson et al.,2009), such that both searches reveal a structured region of 190 amino acid residues, in which a type I KH-like domain exists in the segment between residue 120 to 210, and a segment of unknown structure around residue 20 to 119. The structured region that is possibly carried in DBE is half undefined and half KH-like, as this structural organization has not been reported, it is thus tentatively referred as DBE domain. The technique X-ray crystallography is employed to structurally characterize the DBE domain, therefore the extensive structural disorders in DBE that may hinder crystallization of the DBE domain needed to be removed.

This chapter reports the work to eliminate the structural disorders of DBE and locate the DBE domain, express it and characterize it by biophysical means.

3.2 Bioinformatics analysis of DBE

3.2.1 Homology alignment

Protein sequence of 8 members of DBE family, including fruit fly, budding yeast, human, rat, mouse, frog, mosquito and roundworm, were aligned by CustalW (Figure 3.1). The most conserved region concentrates in the middle, around residue 20 to 270 of DBE, suggesting that functional importance is carried through the middle region

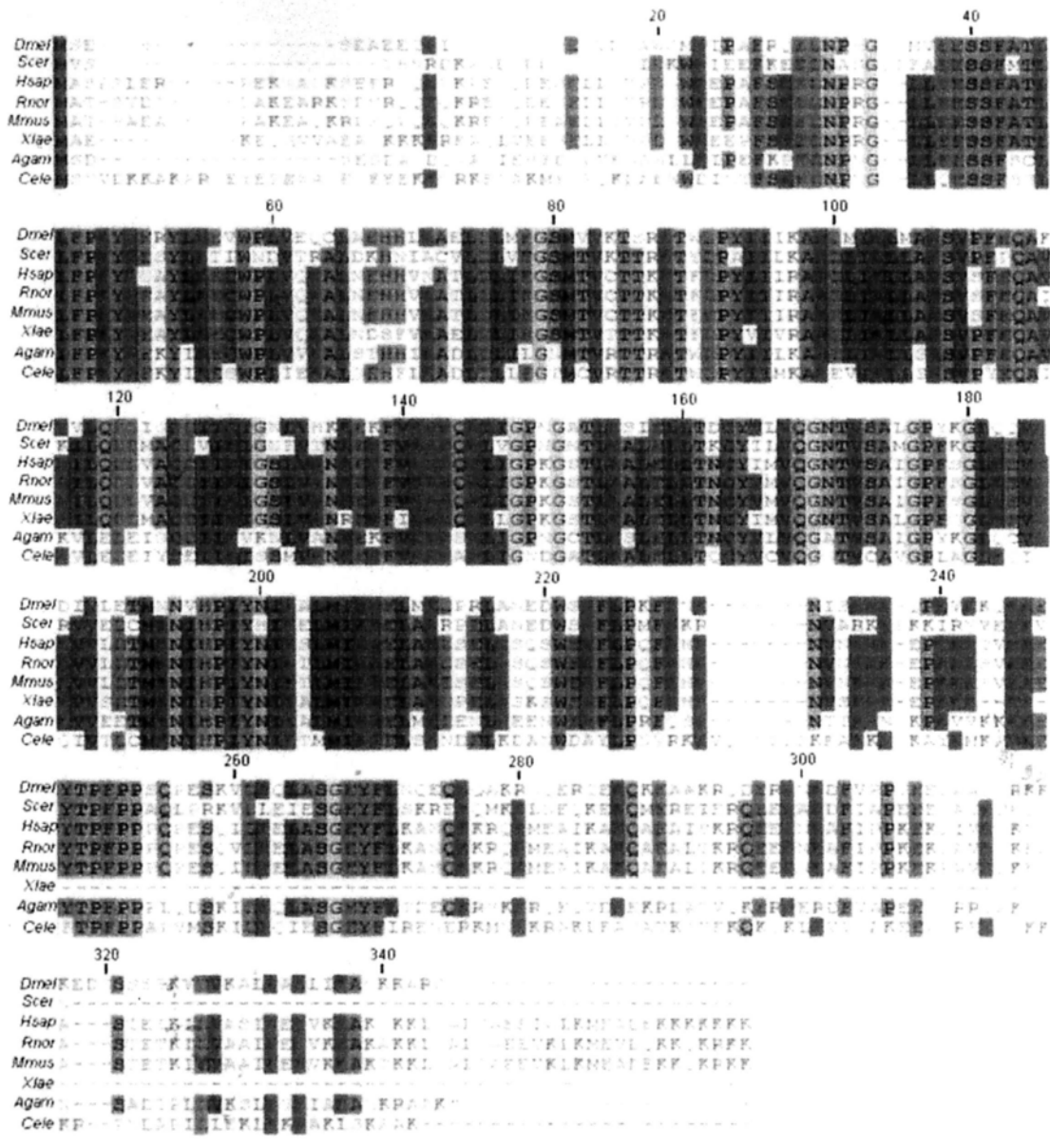


Figure 3.1 Protein sequence alignment of DBE family. Eight homologous proteins of DBE family are aligned by CustalW, including species: fruit fly, *Dmel*, *Drosophila melanogaster*; budding yeast, *Scer*, *Saccharomyces cerevisiae*; human, *Hsap*, *Homo sapiens*; rat, *Rnor*, *Rattus norvegicus*; mouse, *Mmus*, *Mus musculus*; frog, *Xlae*, *Xenopus laevis*; mosquito, *Agam*, *Anopheles gambiae*; roundworm, *Cele*, *Caenorhabditis elegans*. Amino acid residues that are largely conserved are highlighted in boxes. The most conserved region, i.e. the most highlighted region, concentrates in the middle, from around the residue 20 to 270 of DBE.

3.2.2 Disorder prediction

Protein sequence of DBE was subjected to disorder analysis by *Disprot VL3H* (Peng et al.,2005). A disorder score was generated for every amino acid residue of DBE such that amino acid residues with disorder scores over 0.5 were considered to be disordered while those below or at 0.5 were considered to be stable. The disorder prediction is illustrated in Figure 3.2.

About 20 amino acid residues in the N- terminus and about 130 amino acid residues in the C- terminus are considered to be disordered. The stable region lies between residues 20 and 210, with a short disordered stretch around the residue 130. Interestingly, the disorder prediction result agreed with *Pfam* and *Superfamily* database search to the extent that they both agreed on a stable region of about 190 amino acid residues where two structural identities could be found.

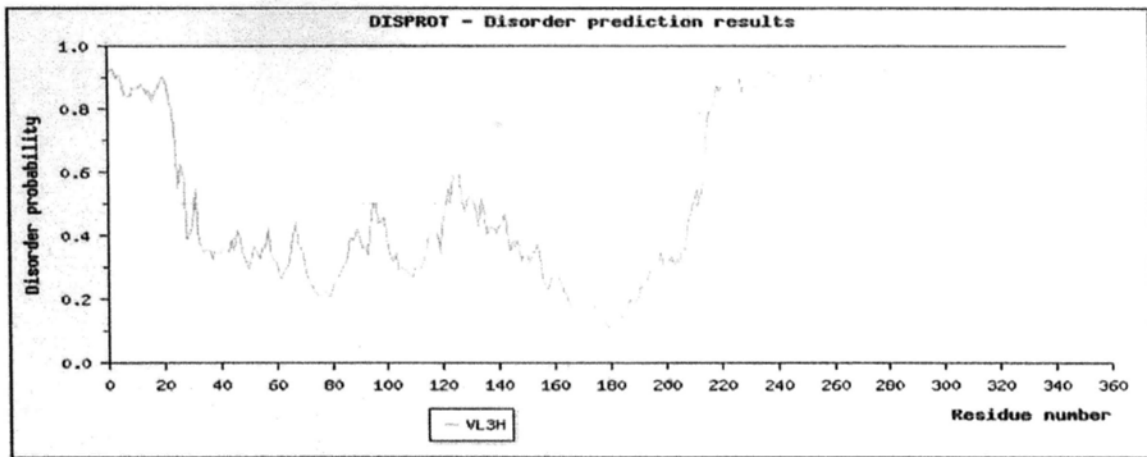


Figure 3.2 Graphic representation of disorder prediction of DBE by *Disprot VL3H*. The DBE protein sequence (345 amino acid residues) was subjected to disorder prediction by *Disprot VL3H*. Disorder probability is scored to every amino acid residues of DBE such that if the score is over 0.5, the residues are considered to be disordered; if the score is below or equal to 0.5, the residues are considered to be stable. The most stable region lies in the middle, between residue 20 and 210. The regions around the N- and C- terminal are disordered.

3.3 Construction of DBE deletion mutants

To facilitate structural studies on the DBE domain, deletion mutants were designed to eliminate the N- and C- terminal disorders and to locate the DBE domain. The design is illustrated in Figure 3.3. The design aimed to locate the DBE domain around residue 20 to 210 of DBE as suggested by disorder prediction analysis (Figure 3.2).

Proline residues, which often act as disruptor of secondary structure due to its conformational rigidity, were chosen as markers to avoid disrupting in the middle of helices or sheets. Three positions at the N- terminus, the Val16, Ala26 and His34, were chosen as they are located immediately after the proline residues. Three positions at the C- terminus, the Pro197, Asp213 and Lys229 were chosen which are proximate to the nearby proline residues (Figure 3.3).

Nine deletion mutants were designed by permutation of the 3 N- and 3 C-terminal positions, i.e. DBE₁₆₋₁₉₇, DBE₁₆₋₂₁₃, DBE₁₆₋₂₂₉, DBE₂₆₋₁₉₇, DBE₂₆₋₂₁₃, DBE₂₆₋₂₂₉, DBE₃₄₋₁₉₇, DBE₃₄₋₂₁₃ and DBE₃₄₋₂₂₉. The 9 deletion mutants were cloned into modified *pET3d* vector with no tag for bacterial expression.

1 - MV16SEEAETKISTEPVDNAWSMKIPAFRQEDNPHGMVEESSFATLFPKYRERYLKEVWPL -60
 61 - VEQCLAA26EHHLKAELDLMEGSMVVKTSRKTWDPYIIHKARDMIKLMARSVPFEQAKRVLQD - 120
 121 - DIGCDH34HIKIGNLVHKKEKFKRRQRLLIGPNGATLKSIELLTDCYVIVQGNTVSAIGPYKG - 180
 181 - LP197QVRDIVLETMNNVHPIYNIKALMIKRELMKDPRLANEDWSRFIPKFKNKNISKRRKQPK - 240
 241 - VD213KKQKKEYTPFPSQPESKVDKQLASGEYFLNQEQKQAKRNRQERTEKQKEAAKRDERRN - 300
 300 - KDFVPPTTEESAASSRKKEDGSSSSKVDVKALKAKLIKANKKARSS - 345

Figure 3.3 Schematic illustration of the design of 9 DBE deletion mutants. The protein sequence of DBE is shown as above. Three positions at the N- terminus, the Val16, Ala26 and His34 residues, and 3 positions at the C- terminus, the Pro197, Asp213 and Lys229 residues were chosen to eliminate structural disorders of DBE. By permutation of the 3 N- terminus positions and 3 C- terminus positions, 9 DBE deletion mutants were constructed, i.e. DBE₁₆₋₁₉₇, DBE₁₆₋₂₁₃, DBE₁₆₋₂₂₉, DBE₂₆₋₁₉₇, DBE₂₆₋₂₁₃, DBE₂₆₋₂₂₉, DBE₃₄₋₁₉₇, DBE₃₄₋₂₁₃ and DBE₃₄₋₂₂₉.

3.4 Identification of DBE₁₆₋₁₉₇ as the working construct

3.4.1 The DBE₁₆₋₁₉₇ construct was initially chosen

The deletion mutants were expressed without any tag. It took lots of trial and error to select the optimal construct to work with.

The constructs DBE₂₆₋₁₉₇, DBE₂₆₋₂₁₃ and DBE₂₆₋₂₂₉ did not show obvious expression (Figure 3.4 A), indicating constructs starting with the residue Ala26 were not feasible to work with. The constructs DBE₃₄₋₁₉₇ and DBE₃₄₋₂₂₉ expressed normally but had difficulty in column binding: less than a half of the protein bound to column even after repeated loading, indicating that constructs starting with the residue His34 possibly harbored a problem of soluble aggregates (Figure 3.4 B). The constructs DBE₁₆₋₂₁₃ and DBE₁₆₋₂₂₉ did not express very well, a significant portion in insoluble fraction was observed after centrifugation of the cell lysate (Figure 3.5).

Contrasting to other constructs, the construct DBE₁₆₋₁₉₇ was optimal in that it expressed well and had no soluble aggregate concern. The protein was expressed in *E. coli* strain *BL21 (DE3) pLysS*. The protein was then subjected to a 3-step purification. The construct DBE₁₆₋₁₉₇ has a predicted pI of 8.9, so the first step was cation exchanger HiTrapTM SP HP purification. The eluted protein was then dialyzed

in a buffer containing 2 M NaCl and subjected to hydrophobic interaction purification by HiTrap™ Phenyl HP. The eluted protein was then concentrated to around 5 ml before gel filtration purification by HiLoad 26/60 Superdex G75. The DBE₁₆₋₁₉₇ protein was eluted at around 180 ml. The purified protein was of high purity in spite of an obvious degradation product (Figure 3.6). A yield of ~6 mg of protein was obtained per liter bacterial culture.

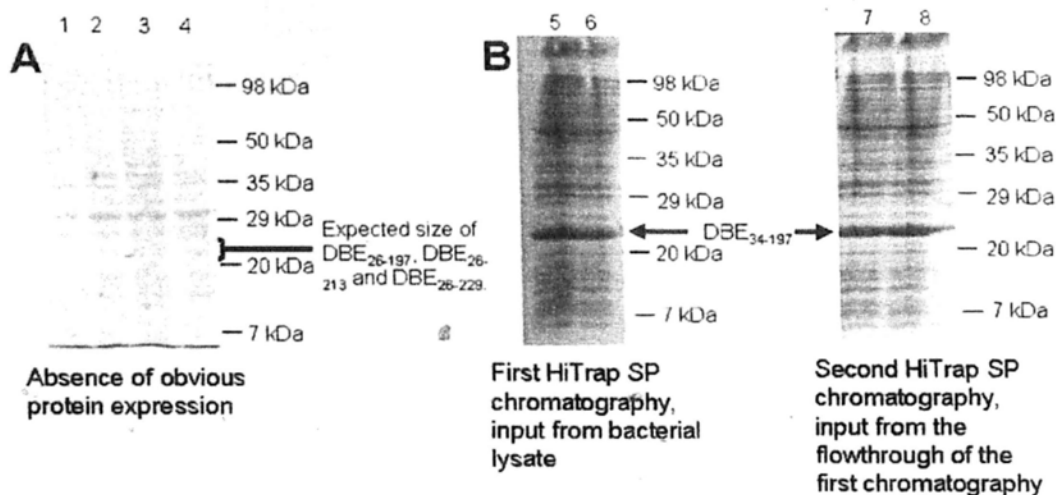


Figure 3.4 (A) Absence of obvious protein expression of tag-less DBE₂₆₋₁₉₇, DBE₂₆₋₂₁₃ and DBE₂₆₋₂₂₉. Expression of DBE₂₆₋₁₉₇, DBE₂₆₋₂₁₃ and DBE₂₆₋₂₂₉ were tested. Samples taken during the course of expression were analyzed by SDS-PAGE, lane 1, *E. coli* strain *BL21 (DE3) pLysS* before induction of protein expression; lane 2, bacterial sample for expression of DBE₂₆₋₁₉₇ (20 kDa) collected at cell harvest; lane 3, bacterial sample for expression of DBE₂₆₋₂₁₃ (22 kDa) collected at cell harvest; lane 4, bacterial sample for expression of DBE₂₆₋₂₂₉ (24 kDa) collected at cell harvest. The expected sizes of the proteins are marked. No obvious protein expression was observed, even protein expression was induced by elevated concentration of IPTG (1 mM) and bacterial culture was grown at 37°C overnight. **(B) The inability of column binding of the DBE₃₄₋₁₉₇ protein.** The DBE₃₄₋₁₉₇ has a predicted pI of 8.9 and so it was first subjected to cation exchange chromatography. Lane 5, bacterial lysate input for the first HiTrap[™] SP column purification; lane 6, flowthrough of the first column purification; lane 7, flowthrough of the first column purification was used as the input of the second HiTrap[™] SP purification; lane 8, flowthrough of the second column purification. It was observed that a large portion of the DBE₃₄₋₁₉₇ protein was not captured by the column and remained in the flowthrough as shown in the left panel as lane 5 and 6. The flowthrough obtained from the first purification was then used as the input for another round of cation exchanger purification. Similar to the first purification, a large portion of the DBE₃₄₋₁₉₇ protein was not captured by the column and remained in the flowthrough as shown in lane 7 and 8 in the right panel. As the HiTrap[™] SP column was perfectly functional, the inability of column binding after repeated loading suggested DBE₃₄₋₁₉₇ possibly harbored a problem of soluble aggregate.

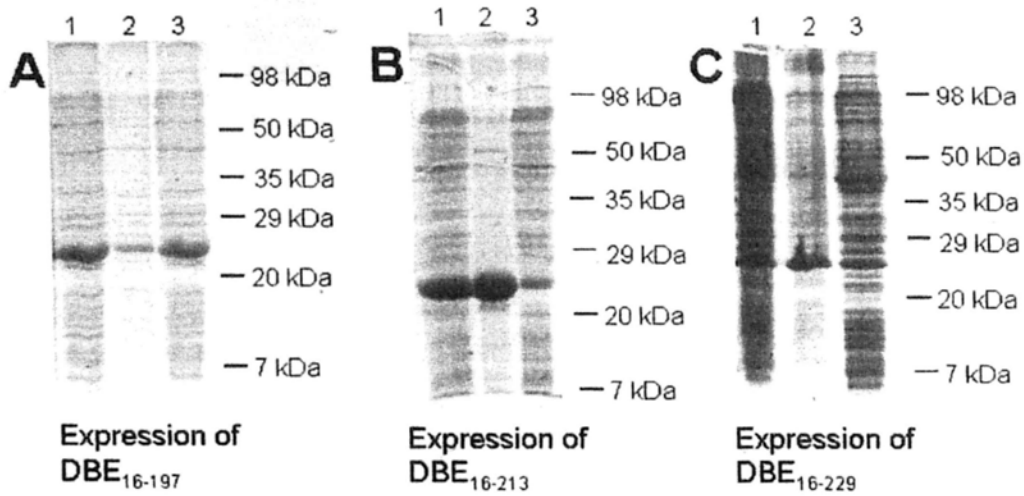


Figure 3.5 Comparison of the expression between the DBE₁₆₋₁₉₇, DBE₁₆₋₂₁₃ and DBE₁₆₋₂₂₉ proteins. The expression of DBE₁₆₋₁₉₇, DBE₁₆₋₂₁₃ and DBE₁₆₋₂₂₉ was compared. Samples taken during the course of expression were analyzed by SDS-PAGE, lane 1, lysed *E. coli* strain *BL21 (DE3) pLysS* sample; lane 2, pelleted sample after centrifugation of bacterial lysate; lane 3, supernatant sample after centrifugation of bacterial lysate. Expression of DBE₁₆₋₁₉₇, DBE₁₆₋₂₁₃ and DBE₁₆₋₂₂₉ are shown in panel (A), (B) and (C) respectively. It can be observed that in the expression of DBE₁₆₋₂₁₃, a large fraction of DBE₁₆₋₂₁₃ was pelleted after centrifugation of bacterial lysate as shown in (B) lane 2. And in the expression of DBE₁₆₋₂₂₉, only half of the DBE₁₆₋₂₂₉ protein was found in the soluble fraction as shown in (C) lane 3. Contrasting to DBE₁₆₋₂₁₃ and DBE₁₆₋₂₂₉, most of the expression of DBE₁₆₋₁₉₇ was found in the soluble fraction, leaving a small portion pelleted as shown in (A) lane 2 and 3.

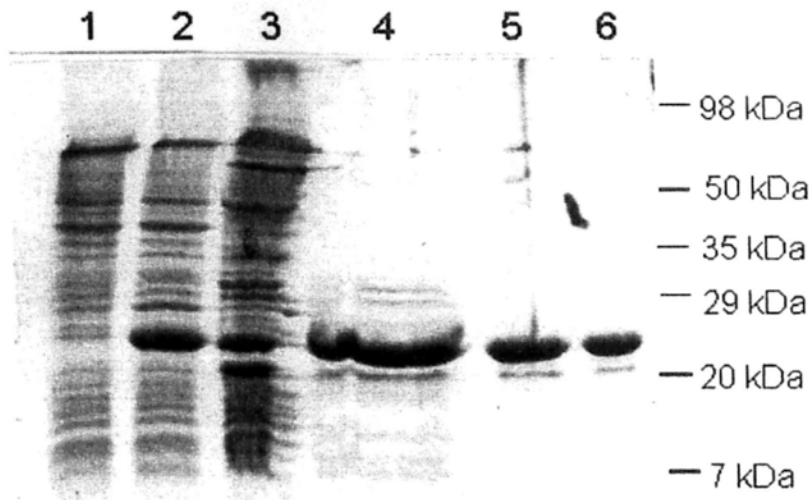


Figure 3.6 Purification profile of tag-less DBE₁₆₋₁₉₇. Samples taken during the course of expression and purification of tag-less DBE₁₆₋₁₉₇ were analyzed by SDS-PAGE: lane 1, *E. coli* strain *BL21 (DE3) pLysS* before induction of protein expression; lane 2, bacterial cell sample at cell harvest; lane 3, sample input for cation exchanger HiTrapTM SP; lane 4, sample input for hydrophobic interaction column HiTrapTM Phenyl; lane 5, sample input for gel filtration Superdex G75 purification; lane 6, sample after all purification steps. The DBE₁₆₋₁₉₇ protein can be expressed, and purified to about high purity despite observable contamination. The impurity band just below the DBE₁₆₋₁₉₇ major band was later identified to be a degradation product in the N-terminus of DBE₁₆₋₁₉₇ that began at the Ile24 residue of DBE.

3.4.2 Selection between DBE₁₆₋₁₉₇ and DBE₂₄₋₁₉₇

As an effort to further optimize the design of DBE deletion construct, the degradation product was extracted and subjected to N-terminal sequencing (performed by Dr. Chan, Siu Hong). The result indicated that the degradation product started with the residue Ile24 of DBE. Thus, *dbe*₂₄₋₁₉₇ was cloned into modified *pET3d* expression vector with no tag like other DBE deletion mutants as in Section 3.3. Similar to DBE₂₆₋₁₉₇, tag-less DBE₂₄₋₁₉₇ did not show obvious protein expression (Figure 3.7 A).

To improve expression and purification of DBE₂₄₋₁₉₇, a poly histidine tag and a Factor Xa recognition sequence were added upstream of the protein, so that the poly histidine tag could be separated from the target protein by Factor Xa protease cleavage. However, an observable portion of expressed protein was found in the insoluble fraction and the expression level of soluble protein was not high (Figure 3.7 B).

Alternatively, a poly histidine tag and a SUMO tag (HS) were added upstream of the protein. SUMO stands for small ubiquitin-related modifier, which is a protein of size 11 kDa that can be recognized by SUMO-specific protease 1 mutant C603S

(SEN1c), such that no extra amino acid residue is left on the C- terminus after cleavage (Xu et al.,2006). By exploiting the specific recognition and cleavage of SUMO by SEN1c, many DBE deletion mutants including DBE₂₄₋₁₉₇, DBE₁₆₋₁₉₇ (Section 3.4.3), DBE₁₆₋₁₉₇ point mutants (Section 5.4.3) and DBE₁₆₋₁₉₇ homologs (Section 5.4.4) were tagged with a HS tag to aid expression and purification. Expression of HS tagged DBE₂₄₋₁₉₇ was observed in *E. coli* strain *C41* but not in *BL21 (DE3) pLysS*. It was purified by Ni²⁺ affinity column. However, the HS tag could not be removed by SEN1c (Figure 3.7 C).

Expression of DBE₁₆₋₁₉₇ was being optimized concurrent to that of DBE₂₄₋₁₉₇. Addition of a poly histidine tag and a Factor Xa recognition sequence to DBE₁₆₋₁₉₇ resulted in protein expression in inclusion bodies like the case of DBE₂₄₋₁₉₇. Unlike DBE₂₄₋₁₉₇, expression and purification of DBE₁₆₋₁₉₇ was made easier with the help of a HS tag as described in the following section.

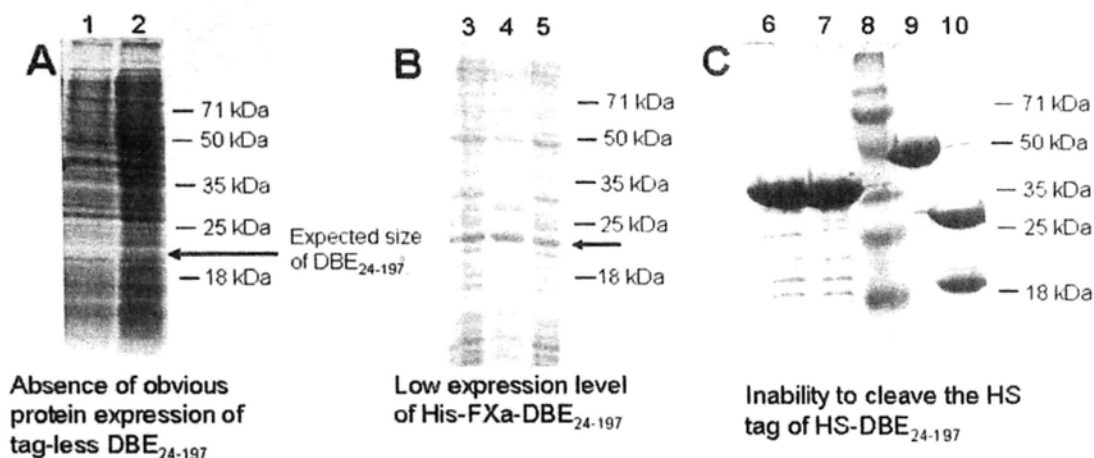


Figure 3.7 (A) Absence of obvious expression of tag-less DBE₂₄₋₁₉₇. The protein DBE₂₄₋₁₉₇ was identified as a degradation product of DBE₁₆₋₁₉₇. Expression of tag-less DBE₂₄₋₁₉₇ was tested. Samples were analyzed in SDS-PAGE: lane 1, *E. coli* strain *BL21 (DE3) pLysS* before induction of protein expression; lane 2, bacterial cell sample at cell harvest. The expected size of tag-less DBE₂₄₋₁₉₇ (20 kDa) is marked. No obvious protein expression of DBE₂₄₋₁₉₇ was observed, even protein expression was induced by elevated concentration of IPTG (1 mM) and bacterial culture was grown at 37°C overnight. **(B) Low expression level of His-FXa-DBE₂₄₋₁₉₇.** The DBE₂₄₋₁₉₇ protein was tagged with a poly histidine and Factor Xa recognition sequence (His-FXa) in hope of aiding expression. Lane 3, lysed *E. coli* strain *BL21 (DE3) pLysS* sample; lane 4, pelleted sample after centrifugation of bacterial lysate; lane 5, supernatant sample after centrifugation of bacterial lysate. The band corresponded to His-FXa-DBE₂₄₋₁₉₇ is marked. It was observed that the expression level of soluble protein was not high, and a fraction of protein was found pelleted after centrifugation of bacterial lysate. **(C) Inability to cleave the HS tag from HS-DBE₂₄₋₁₉₇.** The DBE₂₄₋₁₉₇ protein was HS-tagged to aid expression and purification. The protein was expressed and purified by a Ni²⁺ affinity chromatography. Lane 6, HS-DBE₂₄₋₁₉₇ with 50 μM SENP1c; lane 7, HS-DBE₂₄₋₁₉₇ with 500 μM SENP1c; lane 8, protein standards; lane 9, negative control, HS-GFP with 10 μM SENP1c and NEM as SENP1c inhibitor; lane 10, positive control, HS-GFP with 10 μM SENP1c. The cleavage reaction was carried out at room temperature for 3 hours. The SDS-PAGE analysis demonstrated that SENP1c mediated specific cleavage of HS tag from HS-GFP, leaving 2 major bands as observed in lane 10, while HS-GFP could not be cleaved by inhibited SENP1c as seen in lane 9. However, the HS tag of DBE₂₄₋₁₉₇ cannot be cleaved even at elevated concentrations of SENP1c (lane 6 and 7).

3.4.3 Expression and purification of HS-DBE₁₆₋₁₉₇

Expression of HS-DBE₁₆₋₁₉₇ in *E. coli* strain *BL21 (DE3) pLysS* was observed. The tagged protein was first purified by a Ni²⁺ charged HiTrapTM IMAC column. The protein was eluted at around 200 mM imidazole. The HS tag was recognized and cleaved by SENP1c. The HS tag was then separated from DBE₁₆₋₁₉₇ by passing through a Ni²⁺ charged HiTrapTM IMAC column for a second time, as the HS tag was captured to the column, but DBE₁₆₋₁₉₇ carried no intrinsic Ni²⁺ affinity and was collected in the flowthrough. The DBE₁₆₋₁₉₇ protein was then concentrated to around 5 ml before gel filtration purification by HiLoad 26/60 Superdex G75. The DBE₁₆₋₁₉₇ protein was eluted at around 180 ml. The degradation product observed in the previous tag-less DBE₁₆₋₁₉₇ purification was still present (Figure 3.8). A doubling of protein yield (~13 mg per liter bacterial culture) was obtained in the HS tagged DBE₁₆₋₁₉₇ expression and purification scheme comparing with the tag-less purification scheme (~6 mg per liter bacterial culture). The purified DBE₁₆₋₁₉₇ was then subjected to biophysical characterizations described as below.

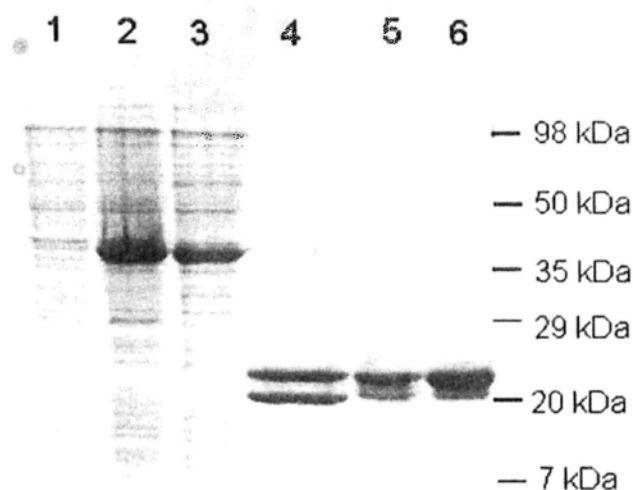


Figure 3.8 Purification profile of DBE₁₆₋₁₉₇ with the aid of HS tag. A HS tag was added to DBE₁₆₋₁₉₇ to aid expression and purification. Samples taken on the course of expression and purification of DBE₁₆₋₁₉₇ were analyzed by SDS-PAGE: lane 1, *E. coli* strain *BL21 (DE3) pLysS* cell sample before induction of protein expression; lane 2, bacterial cell sample at cell harvest; lane 3, sample input for Ni²⁺ affinity column purification; lane 4, sample after SENP1c cleavage for another Ni²⁺ affinity column purification, the upper major band was DBE₁₆₋₁₉₇, the lower major band was the HS tag and the minor band around 35 kDa corresponded to the uncleaved HS-DBE₁₆₋₁₉₇; lane 5, sample input for gel filtration Superdex G75 purification; lane 6, sample after all purification steps. The HS-DBE₁₆₋₁₉₇ protein can be expressed, and purified to about high purity, although the degradation band can still be observed just below the DBE₁₆₋₁₉₇ band. The HS tag expression and purification scheme led to a doubling of protein yield (~13 mg per liter bacterial culture) when compared with the tag-less purification scheme (~6 mg per liter bacterial culture).

3.5 Biophysical characterization of DBE₁₆₋₁₉₇

3.5.1 Far-UV circular dichroism (CD) analysis

Far-UV CD spectroscopy was used to estimate the secondary structure content and the thermostability of DBE₁₆₋₁₉₇. Twenty micromolar of purified DBE₁₆₋₁₉₇ protein was equilibrated in 10 mM NaH₂PO₄, pH 7.4 before measurements by a JASCO J810 spectropolarimeter.

For secondary structure analysis, ellipticity from far-UV of wavelength 190 nm to 260 nm was measured at 25°C. Large absorption (below -20 mdeg) around 200 nm to 230 nm, with an absorption peak at 208 nm and large absorption around 222 nm was observed (Figure 3.9 A). By deconvolution of the CD spectrum using the program CDNN, it was estimated that DBE₁₆₋₁₉₇ carried 53% helices, 12% strands, 21% turns and 14% random coils. The result suggested that DBE₁₆₋₁₉₇ was largely folded.

The thermal denaturation profile DBE₁₆₋₁₉₇ was also measured by CD. Protein sample of DBE₁₆₋₁₉₇ was heated from 25°C to 110°C at a heating rate of 1 °C/minute, and stability was monitored by CD spectroscopy at 222 nm. In contrast to the thermostability of DBE, thermal denaturation of DBE₁₆₋₁₉₇ began at around 45°C and the midpoint of denaturation around 50°C was observed (Figure 3.9 B). The result

indicated that DBE₁₆₋₁₉₇ had a different thermostability than DBE, suggesting there are thermostability contributions by other regions in DBE. The reason of the thermostability disparity between DBE and DBE₁₆₋₁₉₇ is discussed in Section 3.6.3.

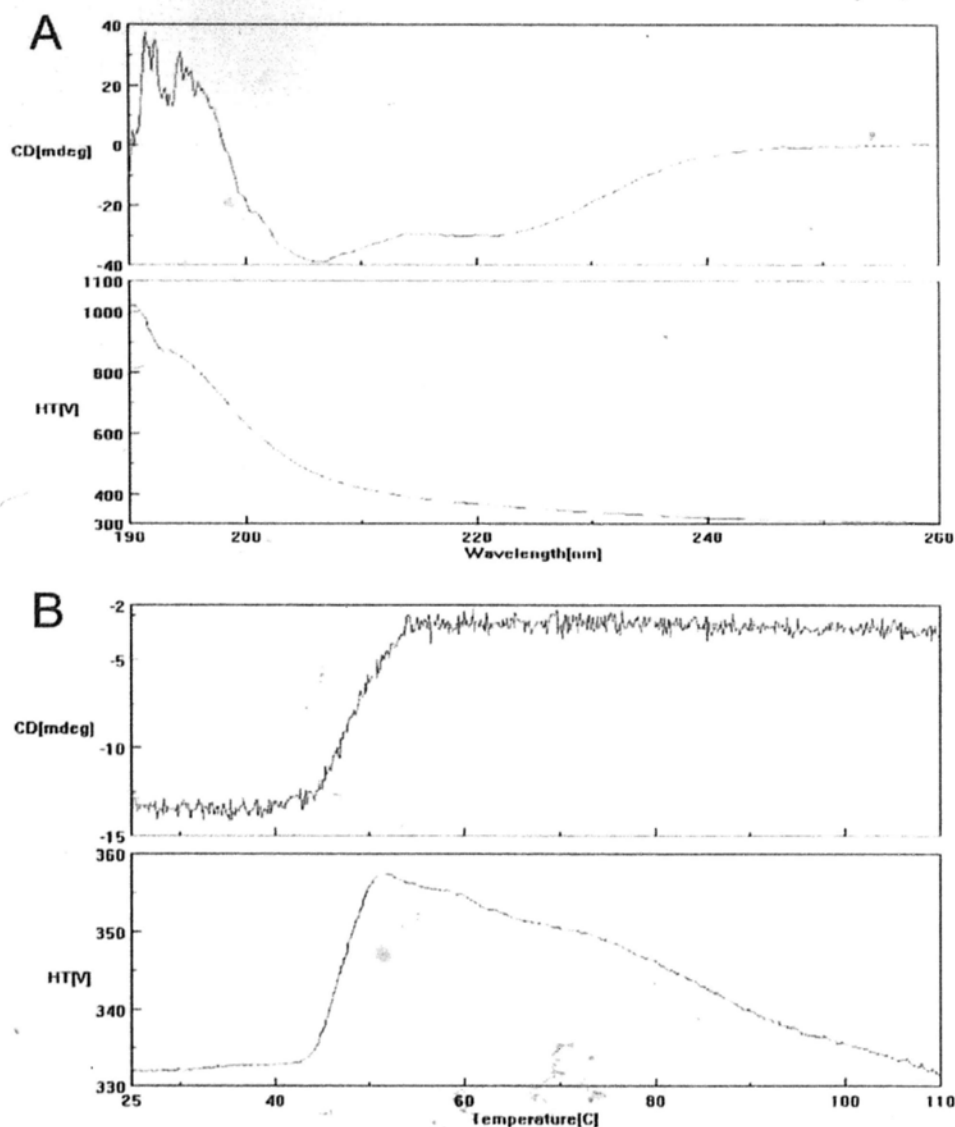


Figure 3.9 (A) Far-UV CD spectrum of DBE₁₆₋₁₉₇. The far-UV CD spectrum of DBE₁₆₋₁₉₇ protein was monitored so as to estimate the secondary structural content of the protein, the ellipticity (mdeg) and dynode value (V) were plotted against wavelength from 190 nm to 260 nm. A large absorption (below -20 mdeg) around 200 nm to 230 nm, with an absorption peak at 208 nm and a large absorption around 222 nm was observed. By deconvolution of the CD spectrum using the program CDNN, it was estimated that DBE₁₆₋₁₉₇ carried 53% helices, 12% strands, 21% turns and 14% random coils, suggesting that DBE₁₆₋₁₉₇ was largely folded. **(B) Thermal denaturation profile of DBE₁₆₋₁₉₇.** The thermal denaturation of DBE₁₆₋₁₉₇ protein was monitored by CD spectroscopy at wavelength 222 nm, so that ellipticity (mdeg) and dynode value (V) were plotted against temperature from 25°C to 110°C. A single step of thermal denaturation was observed. It was observed that denaturation began at around 45°C with a midpoint of denaturation around 50°C.

3.5.2 Light scattering

Both static and dynamic light scattering were performed. Static light scattering was coupled with size exclusion chromatography (SEC) to monitor the absolute molecular weight of purified DBE₁₆₋₁₉₇ protein in solution, and thereby the multimeric state of the protein can be estimated. Dynamic light scattering was used to estimate the hydrodynamic radius and polydispersity content, which reflects the compactness and homogeneity of the purified DBE₁₆₋₁₉₇ protein respectively.

For static light scattering, an analytical gel filtration column Superdex G75 was connected to miniDawn Tristar light scattering detector and Optilab DSP refractometer, and measurements were analyzed by program Astra V. Five hundred micrograms of DBE₁₆₋₁₉₇ was injected for analysis. A major peak with 2 small peaks was observed (Figure 3.10 A). The 2 small peaks might correspond to minor populations of aggregates, multimeric forms or impurities, while the major peak showed a calculated molar mass of 21,000 g/mol, corresponding to a molecular weight of 21 kDa. The experimentally determined molecular weight matched with the molecular weight of DBE₁₆₋₁₉₇ deduced from protein sequence, i.e. 20, 947 Da, which indicated that the major population of purified DBE₁₆₋₁₉₇ is monomeric.

For dynamic light scattering, the measurements were powered by Dynapro Titan and analyzed by Dynamics 6.7.6. Twenty microliters of DBE₁₆₋₁₉₇ of concentration 5 mg/ml was used for analysis. Ten acquisitions, with 10 seconds each acquisition, were taken for each measurement. A major peak with 2 small peaks was observed (Figure 3.10 B). The 2 small peaks might correspond to minor populations of aggregates, multimeric forms of DBE₁₆₋₁₉₇ or impurities, while the major peak that represented DBE₁₆₋₁₉₇ had a calculated hydrodynamic radius of 2.7 nm and 27.7% polydispersity. The implication of the result is discussed in Section 3.6.3.

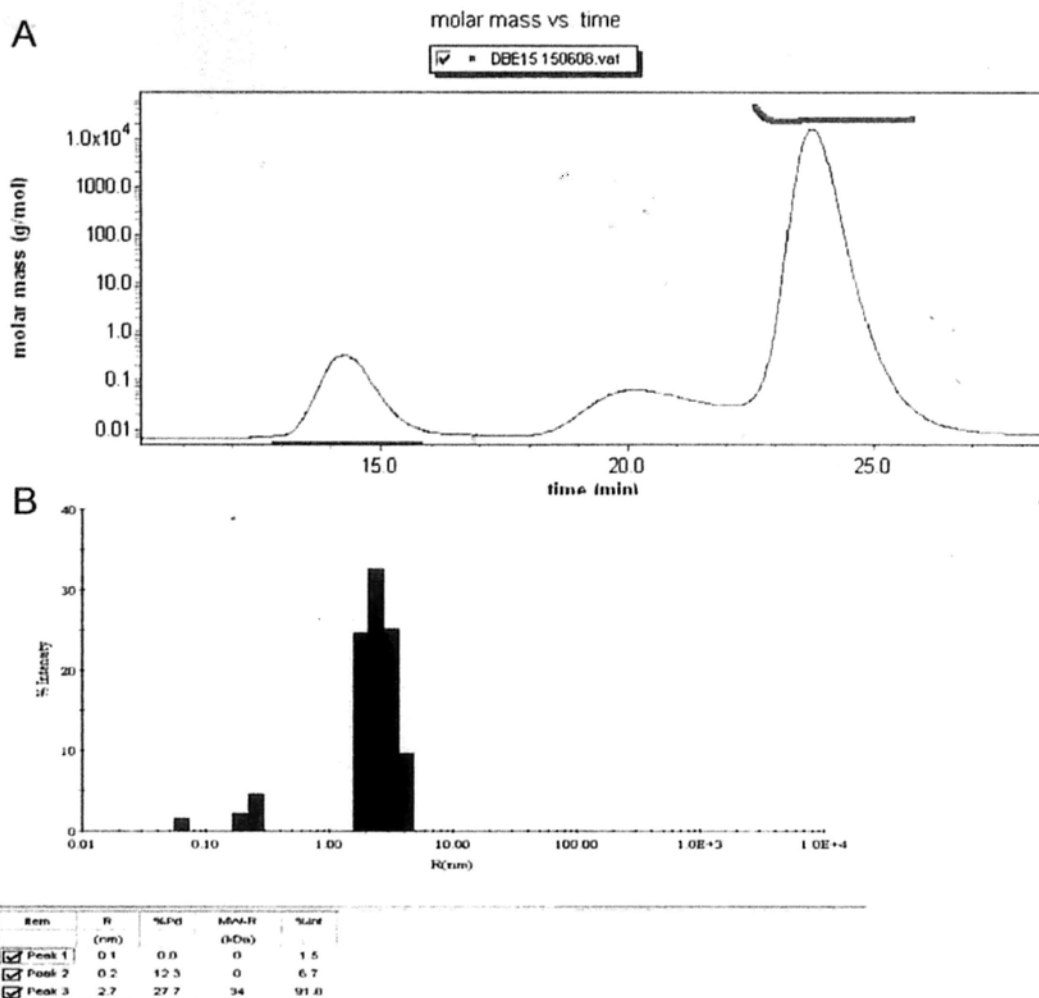


Figure 3.10 (A) Elution profile of DBE₁₆₋₁₉₇ and molar mass estimation by static light scattering. Static light scattering was coupled with size exclusion chromatography (SEC) for determining the absolute size of the purified DBE₁₆₋₁₉₇ protein in solution. The elution profile of DBE₁₆₋₁₉₇ was plotted against time. A major peak and 2 small peaks were observed, and the major peak was estimated to be of a molar mass of 21,000 g/ml, corresponding to a molecular weight of 21 kDa, which indicated that purified DBE₁₆₋₁₉₇ protein (20,947 Da) was mainly a monomer in solution. The 2 small peaks might correspond to minor populations of aggregates, multimeric forms or impurities **(B) Dynamic light scattering analysis of DBE₁₆₋₁₉₇.** The hydrodynamic radius and polydispersity of DBE₁₆₋₁₉₇ was estimated by dynamic light scattering analysis. A major peak and 2 small peaks were observed, the major peak was DBE₁₆₋₁₉₇ (peak 3) that had an estimated hydrodynamic radius of 2.7 nm and 27.7% polydispersity. The 2 small peaks might correspond to minor populations of aggregates, multimeric forms or impurities.

3.6 Discussion

3.6.1 A conserved yet disordered segment is identified

Bioinformatics analyses give insight on the distribution of the disordered regions. For instance, homology alignment (Figure 3.1) suggests that the segment between residue 20 and 270 of DBE are most conserved, while disorder prediction (Figure 3.2) suggests that residue 20 to 210 of DBE are likely to be folded. Taken together, the two pieces of information suggest that a structured region of ~190 amino acid residues is found within the evolutionarily conserved region (~250 amino acid residues); and a region of ~60 amino acid residues spanning from residues 210 to 270 is evolutionarily conserved and yet being structurally disordered. The conserved and disordered segment contains 16 charged residues (13 lysine and 3 arginine) and 7 hydrophobic residues (4 phenylalanine, 1 tryptophan and 2 tyrosine), such that they may possibly contribute to RNA affinity via electrostatic interaction and base stacking respectively. Upon binding to RNA, the disordered region may turn structured and thereby the interaction with other functional components may be facilitated by exposing the otherwise buried residues.

3.6.2 The DBE₁₆₋₁₉₇ protein was the most feasible construct to work with

It is understood that the exact domain boundary is not easy to predict, even though bioinformatics analyses can give a rough idea on the location. The extensive structural disorder of DBE that may deter structural determination needs to be eliminated. Therefore, the aim of creating the 9 DBE deletion mutants was to remove structural disorders and to locate the DBE domain (Figure 3.3). The design tried to avoid interfering in the middle of helices or sheets by choosing proline residues as markers, since proline often acts as disruptor of secondary structures. It is in fact more complicated in the case of DBE, since the number of deletion mutants increases as a product of the number of designated positions at the N- and C- termini, than proteins that carry structural disorder at only one side in which the number of deletion mutants equals the number of designated positions. So, it was thought that 9 deletion mutants were a reasonable number to start with. It was hoped that at least 1 of the 9 constructs could eventually lead to successful crystallization and structural determination.

Among the 9 DBE deletion mutants, DBE₁₆₋₁₉₇ was the most feasible construct to work with in terms of expression level and solubility. Limited proteolysis is often employed to identify a stable domain from structural disorders (Fontana et al.,2004).

It was, however, unexpected that DBE₂₄₋₁₉₇, discovered from a natural degradation

product from DBE₁₆₋₁₉₇ (Figure 3.6), did not give a better construct to work with, since DBE₂₄₋₁₉₇ did not express well in tag-less form and was inseparable from the HS tag (Figure 3.7). The possible reason of inability to cleave the HS tag is that the expression of HS-DBE₂₄₋₁₉₇ rendered the specific cleavage site on the SUMO tag inaccessible to the specific protease SENP1c.

Therefore, it was DBE₁₆₋₁₉₇ that was finally taken as the construct to work with. A HS tag was added to DBE₁₆₋₁₉₇ to help purification and increase protein yield, so that a yield of 13 mg of protein per liter bacterial culture could be obtained, as opposed to a yield of 6 mg of protein per liter bacterial culture in the tag-less expression and purification scheme.

It should be noted that the selection of DBE₁₆₋₁₉₇ as the working construct to eliminate structural disorders and to locate the DBE domain was largely out of practical reasons; whether it could be crystallized or not can only be known from hindsight.

3.6.3 Implications from biophysical studies

The biophysical properties of DBE and DBE₁₆₋₁₉₇ are compared. Parities and disparities are noted.

The first disparity is noted that DBE contains extensive structural disorders (Yiu et al.,2006), while DBE₁₆₋₁₉₇ is largely folded as estimated by CD spectroscopy (Figure 3.9 A). This disparity is expected since DBE₁₆₋₁₉₇ was designed exactly to eliminate structural disorder of DBE.

Another disparity is that a difference in the thermal denaturation profiles of DBE and DBE₁₆₋₁₉₇ is noted. Previous CD spectrum analysis of DBE reveals a 2 step thermal denaturation, with the first transition of mid-point at ~35°C and a second transition at 90°C, which indicates that DBE is relatively stable in between the transitions, i.e. from 40°C to 90°C (Yiu et al.,2006). However, DBE₁₆₋₁₉₇ began to denature at 45°C (Figure 3.9 B), suggesting that the regions outside DBE₁₆₋₁₉₇ are conducive to DBE's thermostability. The difference in thermostability can be explained in 2 ways as follows. The first explanation assumes that the first transition (~35°C) in DBE corresponds to the thermostability of the DBE domain, and hence the observed thermostability of DBE₁₆₋₁₉₇ indeed matches that of the DBE domain. The

second explanation hinges on the few eliminated conserved and disordered hydrophobic residues in DBE₁₆₋₁₉₇: 3 tyrosine residues, 4 phenylalanine residues and 1 tryptophan residue are found in the segment outside of DBE₁₆₋₁₉₇ from residue 198 to 345 of DBE. It has been noted that the hydrophobic residues of DBE are more exposed than that of fully folded proteins (Yiu et al.,2006), and the few disordered hydrophobic residues may enhance hydrophobic interaction at elevated temperatures, thus conferring extra thermostability to DBE.

Two common features between DBE and DBE₁₆₋₁₉₇ are noted. First, they are both monomeric in solution. Second, DBE is estimated to be in a state of molten globule (Yiu et al.,2006), while DBE₁₆₋₁₉₇ had an estimated hydrodynamic radius of 2.7 nm by dynamic light scattering (Figure 3.10 B), which falls within the range of the theoretical hydrodynamic radius (2.4 nm to 2.7 nm) of a molten globule as estimated by the equations below (Uversky,2002).

$$\text{Log}(V_h) = (2.46 \pm 0.13) + (1.020 \pm 0.053) \cdot \log(N)$$

$$V_h = 4/3 \cdot \pi \cdot R_h^3$$

Where V_h is the hydrodynamic volume, N is the number of amino acid residues and R_h is the hydrodynamic radius.

It should be noted that the calculation above assumes that the protein is spherical, while for the moment it is not known how spherical DBE₁₆₋₁₉₇ is before structural determination. Therefore, although both DBE and DBE₁₆₋₁₉₇ are estimated to be in a state of molten globule by the same set of equations, (Uversky.2002), it is not certain how correct the calculation can be before knowing how spherical both proteins are. If the assumption that DBE₁₆₋₁₉₇ is spherical is true, then DBE₁₆₋₁₉₇ protein as in the state of molten globule suggested that it is collapsed and it consists of native secondary structures and also a dynamic tertiary structure, which agrees with the far-UV CD spectroscopy analysis that DBE₁₆₋₁₉₇ is largely folded. Therefore, although DBE₁₆₋₁₉₇ was estimated to have a 27.7% polydispersity, which is not ideal (< 15%) for crystallization, DBE₁₆₋₁₉₇ was thought to be ready for screening of crystallization condition.

4. X-ray crystallography studies on DBE₁₆₋₁₉₇

4.1 Introduction

Difficulty in structural studies on the whole pre-rRNP unit that formed during ribosome biogenesis is noted, which is possibly due to the dynamic nature of the numerous transiently acting *trans*-factors that constitute the pre-rRNP (Henras et al., 2008; Dinman, 2009). The difficulty can be circumvented by taking a bottom-up approach, i.e. determine the structure of individual components first, from which the structure of the whole unit can be reconstituted by computational modeling. The first attempt taking this approach in reconstituting pre-40S particle is recently reported (Granneman et al., 2010) and the DBE family is possibly a *trans*-factor constituent of early pre-rRNP as discussed in Section 1.2.2. Therefore, similar to the pre-40S particle, structural determination of DBE can act as a first step toward a structural understanding of the early pre-rRNP complex. Also, structural determination of DBE can reveal the unreported identity of the DBE domain (Section 3.1). With an aim to locate the DBE domain, DBE₁₆₋₁₉₇ was constructed to facilitate structural determination by X-ray crystallography (Chapter 3).

X-ray crystallography is currently the most used technique for high resolution macromolecular structural determination as archived in the Protein Data Bank (PDB)

database. Structural determination by X-ray crystallography mandates the crystallization of the macromolecule, and then the crystal is subjected to X-ray diffraction (Blow, 2002; Rhodes, 2006). The intensities of the diffracted beams are recorded and combined with the phase information of the diffracted beams, a electron density map can then be calculated from which a model for the mean positions of the atoms in the crystal can be derived (Blow, 2002; Rhodes, 2006). It should be noted that the phase information is more important than the intensity information as required by the mathematics of X-ray crystallography (Blow, 2002; Rhodes, 2006). Phase information can either be obtained by molecular replacement (MR) from previously determined homologous structure, or by experimental techniques like Multiple Isomorphic Replacement (MIR), Single Isomorphous Replacement with Anomalous Scattering (SIRAS), Single Anomalous Dispersion (SAD) and Multi-wavelength Anomalous Dispersion (MAD) (Blow, 2002; Rhodes, 2006). Although it is generally held that successful crystallization is the key to structural determination by X-ray crystallography, it is not uncommon occurrence that complications can deter structural determination, and particularly difficult cases are known as pathological crystallography (Dauter et al., 2005).

This chapter reports the attempt to determine DBE₁₆₋₁₉₇ structure by X-ray

crystallography; it describes the crystallization of DBE₁₆₋₁₉₇, diffraction analysis and phasing attempts of the crystal structure.

4.2 Crystallization of DBE₁₆₋₁₉₇

4.2.1 Screening of crystallization condition of DBE₁₆₋₁₉₇

The construct DBE₁₆₋₁₉₇ was previously selected out of the 9 DBE deletion mutants to locate the DBE domain, and it was purified with the aid of a HS tag (Chapter 3). Screening for crystallization conditions was performed. The purified DBE₁₆₋₁₉₇ was concentrated to 10 mg/ml under the same buffering condition as Superdex G75 column, i.e. 50 mM NaH₂PO₄, pH 7.4 and 0.4 M NaCl, for protein crystallization. Ten crystallization screening kits were tested, and they were Crystal Screen 1 and 2, Index, SaltRx, PEG/ion screen kits (Hampton Research) and JCSG+, JCSG core I, II, III and IV (QIAGEN). Screening was performed at 16°C using sitting drop on Greiner CrystalQuick 96 well plates with 3 protein to well solution ratios, 0.1 µl: 0.2 µl, 0.15 µl: 0.15 µl and 0.2 µl: 0.1 µl. Only one initial hit was observed, and condition for the hit was 0.4 M ammonium phosphate (monobasic) from Crystal Screen 1 (Figure 4.1). The optimal crystallization condition was achieved after extensive screening from various pH (3.0 to 7.0) and concentrations of ammonium phosphate (0.1 to 1 M), incubation temperature (4°C, 16°C and 25°C), various ratios of protein to well

solution (1:1, 1:2, 2:1, 3:1), and hanging drop/ sitting drop setting. The optimal condition for crystal growth was: in a 24 well sitting drop plate, with 500 μ l of well solution of 0.6 M ammonium phosphate, pH 4.8, protein to precipitant ratio of 2 μ l: 4 μ l at 16°C. Crystals of maximum size could be observed after 4 days of incubation. Crystal often grew as stacked plates with occasional single plates of $\sim 0.3 \times 0.4 \times 0.02$ mm after 4 days of incubation (Figure 4.1).

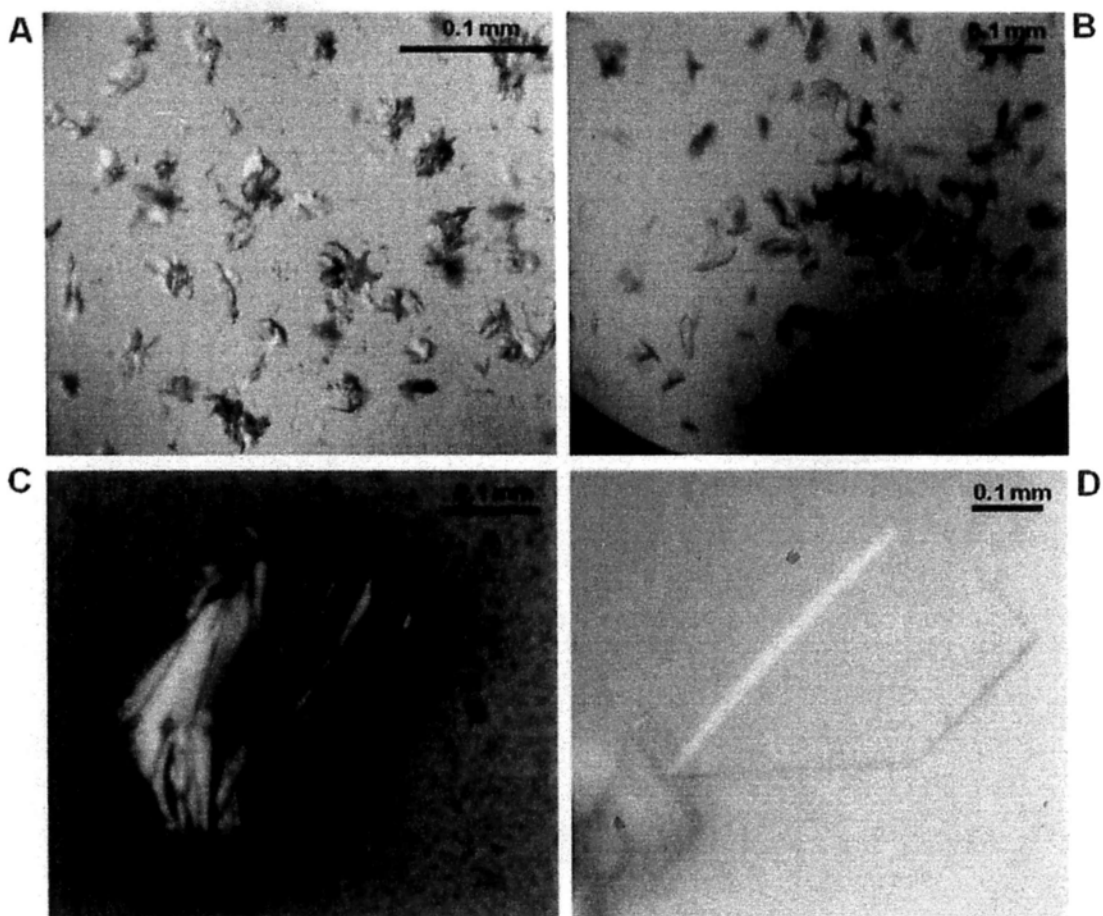


Figure 4.1 Images of DBE₁₆₋₁₉₇ crystals. (A) Crystal images captured from initial screening well. The crystals were clustered and too irregular and too small to use. Optimization of crystallization conditions was needed. (B) After optimization, most crystals grew as stacked plates with occasional small single plates. (C) At 0.6 M ammonium phosphate, pH 4.8, protein to precipitant ratio of 2 μ l: 4 μ l at 16°C in 24 well sitting drop plate, 1-2 crystals could be obtained per well. Most crystals were stacked plates as illustrated. (D) Only a few plate-like single crystals, around 4 per 24 well sitting drop plates could be obtained. The single crystal could often be grown to a size of $\sim 0.3 \times 0.4 \times 0.02$ mm as illustrated.

4.3 X-ray diffraction analysis of DBE₁₆₋₁₉₇ crystals

Cryoprotectant condition screening was performed with an in-house copper anode X-ray source MicroMaxTM-007 HF (Rigaku). It was found that the optimal cryoprotectant condition of DBE₁₆₋₁₉₇ crystals was in Cryo_1 solution (Section 2.9.5), i.e. 30% ethylene glycol with 0.6 M ammonium phosphate, pH 4.8. X-ray diffraction data on single DBE₁₆₋₁₉₇ crystal were collected on beamline I04 at the Diamond Light Source (UK) with a wavelength of 0.9704 Å at 100 K using an ADSC Q315 detector. Data processing was performed using *MOSFLM* (Leslie, 2006) and *SCALA* (Evans, 2006) via *CCP4i* suite (Potterton *et al.*, 2003).

	Native crystal
Beamline	IO4, Diamond (UK)
Wavelength (Å)	0.9704
Detector	ADSC Q315
Total rotation range (°)	280
Resolution range (Å)	52.13-1.98 (2.09-1.98)
Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	$a = 68.99, b = 78.76, c = 79.57,$ 90, 90, 90
Mosaicity (°)	0.73
Total no. of measured intensities	322,992
Unique reflections	30,870
Multiplicity	10.5 (10.8)
Mean $I/\sigma(I)$	11.0 (3.9)
Completeness (%)	99.9 (100)
$^\dagger R_{\text{merge}}$ (%)	9.6 (39.2)
Overall B factor from Wilson plot (Å ²)	25.5

$^\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ are the observed intensity and the mean intensity of related reflections, respectively.

Table 4.1 Data collection and processing statistics of native DBE₁₆₋₁₉₇ crystal. The numbers in parentheses correspond to the values in the highest resolution shell. Diffraction data was integrated by *MOSFLM* and scaled by *SCALA*.

Although a problem of diffraction anisotropy was noted, diffraction data were integrated and scaled to a maximum resolution of 1.98 Å. The space group belonged to $P2_12_12_1$. The unit-cell parameters were $a = 68.99$, $b = 78.76$, $c = 79.57$ Å. Matthews coefficient (V_M) analysis (Matthews, 1968) suggests that each asymmetric unit contains two protein molecules, giving a V_M value of $2.58 \text{ \AA}^3 \text{ Da}^{-1}$ and a solvent content of 52.4%. Statistics of data processing is listed in Table 4.1.

4.4 Phasing of DBE₁₆₋₁₉₇ crystals

After data collection of DBE₁₆₋₁₉₇ crystals, molecular replacement was attempted by *PHENIX AutoMR* (Adams et al., 2010) and *molrep*, *phaser* via *CCP4i* suite (Potterton et al., 2003). Search models of closely related protein Dim2p (PDB code: 1tua and 2e3u) were used, but no solution was identified.

Experimental phasing was then attempted to solve the structure. An in-house copper anode X-ray generator MicroMaxTM-007 HF (Rigaku) that generated X-ray at 1.54 Å was used for experimental phasing except for the selenium MAD experiment mentioned in Section 4.4.3. Data processing and analysis was performed using *MOSFLM* (Leslie, 2006) and *SCALA* (Evans, 2006) via *CCP4i* suite (Potterton et al., 2003), and *PHENIX* suite (Adams et al., 2010).

4.4.1 Sulfur SAD

The technique using anomalous signal given by sulfur atom is known as sulfur SAD. Sulfur atom can give detectable anomalous signal at X-ray of wavelength 1.54 Å, which is accessible by in-house copper anode X-ray source. The DBE₁₆₋₁₉₇ protein carries 10 sulfur atoms, 3 from cysteine residues and 7 from methionine residues. Phasing DBE₁₆₋₁₉₇ structure by the anomalous signal of sulfur at wavelength 1.54 Å was attempted. A dataset was collected, however, no anomalous signal was detected by *phenix.xtriage* analysis. Statistics of data processing is summarized in Table 4.2.

	Native crystal
X-ray source	MicroMax TM -007 HF
Wavelength (Å)	1.54
Detector	R-AXIS IV
Total rotation range (°)	130
Resolution range (Å)	39.78-2.60 (2.74-2.60)
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit-cell parameters (Å, °)	<i>a</i> = 69.69, <i>b</i> = 78.58, <i>c</i> = 79.58, 90, 90, 90
Mosaicity (°)	0.91
Total no. of measured intensities	69,970
Unique reflections	13,789
Multiplicity/ Anomalous Multiplicity	5.1 (5.1) / 2.7 (2.7)
Completeness/ Anomalous Multiplicity (%)	98.9 (98.3)/ 98.7 (97.3)
Mean <i>I</i> /σ(<i>I</i>)	11.5 (3.7)
[†] <i>R</i> _{merge} (%)	7.4 (32.2)
Overall <i>B</i> factor from Wilson plot (Å ²)	47.2

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ are the observed intensity and the mean intensity of related reflections, respectively.

Table 4.2 Data collection and processing statistics of native DBE₁₆₋₁₉₇ crystal for sulfur SAD. The numbers in parentheses correspond to the values in the highest resolution shell. Diffraction data was integrated by *MOSFLM* and scaled by *SCALA*.

4.4.2 Heavy atom MIR/ SIRAS

Introducing heavy atoms into DBE₁₆₋₁₉₇ crystals were attempted, so that the heavy atom derivative of DBE₁₆₋₁₉₇ crystal can help phase the crystal structure by techniques like MIR or SIRAS.

4.4.2.1 Heavy metal atoms

Fifteen heavy metal atoms in total, twelve heavy metal atoms from JBScreen Heavy (Jena Bioscience), and samarium (III) acetate, lead (II) chloride and cesium chloride were used (Table 4.3). Single plates of DBE₁₆₋₁₉₇ crystals were carefully transferred to the heavy metal atom solution drops. The heavy atom solution drops were prepared by adding heavy atom into the optimal growing condition of DBE₁₆₋₁₉₇ crystals, i.e. 0.6 M ammonium phosphate, pH 4.8. Concentrations of the heavy atom solution ranged from 50 mM to 50 μ M and soaking times ranged from 10 minutes to over 2 weeks at 16°C.

Soaking of DBE₁₆₋₁₉₇ crystals in 50 mM of potassium tetrachloroplatinate (II) (PtK₂Cl₄) from JBScreen Heavy for 10 minutes resulted in a change of space group from *P*2₁2₁2₁ to *C*222. All other conditions resulted in crystal cracking, lack of diffraction pattern in spite of crystal looking intact, or no heavy atom substitution of

crystal (Table 4.3). Anomalous signal given by platinum atom was detected, but only extended to 4.8 Å as analyzed by *phenix.xtriage*. Statistics of data processing is listed in Table 4.4. With one single heavy atom substitution and a change of space group, and also weak anomalous signal, it was not likely to phase DBE₁₆₋₁₉₇ crystal structure.

Chemical name	Molecular formula	Heavy atom
Potassium tetrachloroplatinate (II)	Pt K ₂ Cl ₄	Pt
Potassium dicyanoaurate (I)	KAu(CN) ₂	Au
Dipotassium tetraiodomercurate (II)	K ₂ HgI ₄	Hg
Mercury (II) chloride	HgCl ₂	Hg
Sodium 4-(Chloromercuri)benzenesulfonic acid	C ₆ H ₄ ClHgO ₃ Na	Hg
Potassium tetranitroplatinate (II)	K ₂ Pt(NO ₂) ₄	Pt
Potassium tetracyanoplatinate (II) trihydrate	K ₂ Pt(CN) ₄	Pt
Mercury (II) acetate	Hg(CH ₃ COO) ₂	Hg
<i>cis</i> -Dichlorodiammine platinum (II)	<i>cis</i> -Pt(NH ₃) ₂ Cl ₂	Pt
Ethyl mercurithiosalicylate (Thimerosal)	C ₉ H ₉ HgNaO ₂ S	Hg
Potassium tetrachloroaurate (III) hydrate	KAuCl ₄	Au
Potassium hexachloroiridate (IV)	K ₂ IrCl ₆	Ir
Samarium (III) acetate	Sm(CH ₃ COO) ₃	Sm
Lead (II) chloride	PbCl ₂	Pb
Cesium chloride	CsCl	Cs

Table 4.3 Fifteen heavy metal atoms used for deriving heavy metal substituted DBE₁₆₋₁₉₇ crystals. The heavy atom solution drops were prepared by adding heavy atom into the optimal growing condition of DBE₁₆₋₁₉₇ crystals, i.e. 0.6 M ammonium phosphate, pH 4.8. Concentrations of the heavy atom solution ranged from 50 mM to 50 μ M and soaking times ranged from 10 minutes to over 2 weeks at 16°C. Only the condition that soaking DBE₁₆₋₁₉₇ crystal in 50 mM of Pt K₂Cl₄ for 10 minutes led to heavy atom substitution of the crystal. Other soaking conditions led to crystal cracking, or lack of diffraction, or no heavy atom substitution of crystal

Pt K ₂ Cl ₄ soaked crystal	
X-ray source	MicroMax TM -007 HF
Wavelength (Å)	1.54
Detector	R-AXIS IV
Total rotation range (°)	95
Resolution range (Å)	43.79-3.00 (3.16-3.00)
Space group	C222
Unit-cell parameters (Å, °)	$a = 71.30, b = 79.22, c = 77.78,$ $90, 90, 90$
Mosaicity (°)	0.90
Total no. of measured intensities	16,530
Unique reflections	4,615
Multiplicity/ Anomalous Multiplicity	3.6 (3.7) / 1.9 (2.0)
Completeness (%)	99.8 (100)/ 98.4 (97.8)
Mean $I/\sigma(I)$	10.8 (3.1)
[†] R_{merge} (%)	5.5 (34.8)
Overall B factor from Wilson plot (Å ²)	75.5

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ are the observed intensity and the mean intensity of related reflections, respectively.

Table 4.4 Data collection and processing statistics of PtK₂Cl₄ substituted DBE₁₆₋₁₉₇ crystal. The numbers in parentheses correspond to the values in the highest resolution shell. Diffraction data was integrated by *MOSFLM* and scaled by *SCALA*.

4.4.2.2 Halides

Iodide or bromide can be introduced into crystals by quick soaking (Dauter et al., 2000). The DBE₁₆₋₁₉₇ crystals were soaked in 0.75 M, 0.5 M and 0.35 M iodide or bromide containing cryoprotectant from 1 minute to 20 seconds. Soaking in bromide solution resulted in cracking of crystals. Soaking in 0.35 M iodide solution for 20 seconds resulted in an iodide substituted crystal, prolonged soaking or increased iodide concentration deteriorated diffraction quality. Data was collected and statistics of data processing is listed in Table 4.5. Anomalous signal was detected but it only extended to at most 5.4 Å as analyzed by *phenix.xtriage*, which was not strong enough to phase the structure.

5-amino-2,4,6-triiodoisophthalic acid (I3C) carries 3 iodide atoms and is recently taken to phase crystal structure (Beck et al., 2008; Beck et al., 2009). Two approaches were attempted. First, DBE₁₆₋₁₉₇ crystals were soaked in 0.75 M, 0.5 M and 0.35 M of I3C containing cryoprotectant from 10 to 1 minute, no substitution of I3C was observed. Second, 0.01 M, 0.05 M and 0.1 M of I3C were added to DBE₁₆₋₁₉₇ protein drops for co-crystallization. The crystallization conditions were exactly the same as that of native DBE₁₆₋₁₉₇ crystals. Small crystals were observed in well containing 0.01 M and 0.05 M of I3C, but not in 0.1 M of I3C. Diffraction data was collected

(Table 4.6), but no anomalous signal was detected by *phenix.xtriage*, which indicated no I3C substitution of DBE₁₆₋₁₉₇ crystal occurred, and hence no phase information could be derived.

	Iodide soaked crystal
X-ray source	MicroMax TM -007 HF
Wavelength (Å)	1.54
Detector	R-AXIS IV
Total rotation range (°)	100
Resolution range (Å)	43.38-2.70 (2.85-2.70)
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit-cell parameters (Å, °)	<i>a</i> = 68.93, <i>b</i> = 78.86, <i>c</i> = 79.05, 90, 90, 90
Mosaicity (°)	1.28
Total no. of measured intensities	46,539
Unique reflections	11,752
Multiplicity/ Anomalous Multiplicity	4.0 (4.0) / 2.1 (2.1)
Completeness/ Anomalous Multiplicity (%)	95.9 (94.9)/ 95.6 (93.6)
Mean <i>I</i> /σ(<i>I</i>)	8.4 (2.6)
[†] <i>R</i> _{merge} (%)	8.4 (39.3)
Overall <i>B</i> factor from Wilson plot (Å ²)	56.4

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ are the observed intensity and the mean intensity of related reflections, respectively.

Table 4.5 Data collection and processing statistics of iodide substituted DBE₁₆₋₁₉₇ crystal. The numbers in parentheses correspond to the values in the highest resolution shell. Diffraction data was integrated by *MOSFLM* and scaled by *SCALA*.

	Crystal grown in I3C
X-ray source	MicroMax TM -007 HF
Wavelength (Å)	1.54
Detector	R-AXIS IV
Total rotation range (°)	150
Resolution range (Å)	43.55-3.10 (3.27-3.10)
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit-cell parameters (Å, °)	<i>a</i> = 69.52, <i>b</i> = 78.54, <i>c</i> = 79.48, 90, 90, 90
Mosaicity (°)	1.01
Total no. of measured intensities	47,677
Unique reflections	8,157
Multiplicity/ Anomalous Multiplicity	5.8 (6.0) / 3.2 (3.2)
Completeness/ Anomalous Multiplicity (%)	98.4 (97.4)/ 98.3 (97.1)
Mean <i>I</i> /σ(<i>I</i>)	9.8 (4.4)
[†] <i>R</i> _{merge} (%)	11.8 (30.8)
Overall <i>B</i> factor from Wilson plot (Å ²)	57.4

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ are the observed intensity and the mean intensity of related reflections, respectively.

Table 4.6 Data collection and processing statistics of DBE₁₆₋₁₉₇ crystal grown in the presence of I3C. The numbers in parentheses correspond to the values in the highest resolution shell. Diffraction data was integrated by *MOSFLM* and scaled by *SCALA*.

4.4.3 Selenium based MAD

Selenium based MAD is nowadays a commonly used technique for phasing crystal structure. Selenium was introduced into DBE₁₆₋₁₉₇ by substituting the 7 natural methionine residues by selenomethionine. An *E. coli* methionine auxotroph *B834* was used to ensure selenomethionine incorporation, as it depended on external supply of selenomethionine to survive. Expression of selenomethionyl DBE₁₆₋₁₉₇ was conducted in M9 minimal medium with the addition of selenomethionine. An improved strain *B834S*, which was made by introducing a *pLysS* plasmid into the original strain *B834*, was used for selenomethionyl DBE₁₆₋₁₉₇ expression; however, the protein was initially expressed in inclusion bodies (Figure 4.2 A). Optimization of expression condition was performed to express selenomethionyl DBE₁₆₋₁₉₇ in soluble form as described in Section 2.3.2 (Figure 4.2 B).

The purification scheme of selenomethionyl DBE₁₆₋₁₉₇ was the same as that of native protein, except 1 mM TCEP was necessarily required in all purification buffers to prevent selenium oxidation. The crystallization condition was exactly the same as that of native DBE₁₆₋₁₉₇ crystals. The crystals were frozen and transported to the synchrotron facility.

Diffraction data were collected at beamline BL17U at Shanghai Synchrotron Radiation Facility (SSRF) (China) at 100K using MAR 225 CCD detector. Diffraction data at 3 wavelengths, corresponding to peak, inflection and remote, were collected and the statistics are listed in Table 4.7. Anomalous signal was detected, but the selenium atom positions could not be identified by on site *SHELXD* and later *PHENIX hys*s analysis, which meant no phase information of DBE₁₆₋₁₉₇ crystal was learnt from this technique probably because of 2 problems, i.e. bad radiation damage and translational pseudo-symmetry. Bad radiation damage was indicated by a gradual loss of diffraction pattern and a sharp drop in B factors as analyzed by *SCALA*, while translational pseudo-symmetry was detected in the selenomethionyl DBE₁₆₋₁₉₇ crystal by native patterson analysis by *phenix.xtriage* such that a significant (42.9%) off origin peak at (-0.500, 0.500, 0.039) was identified.

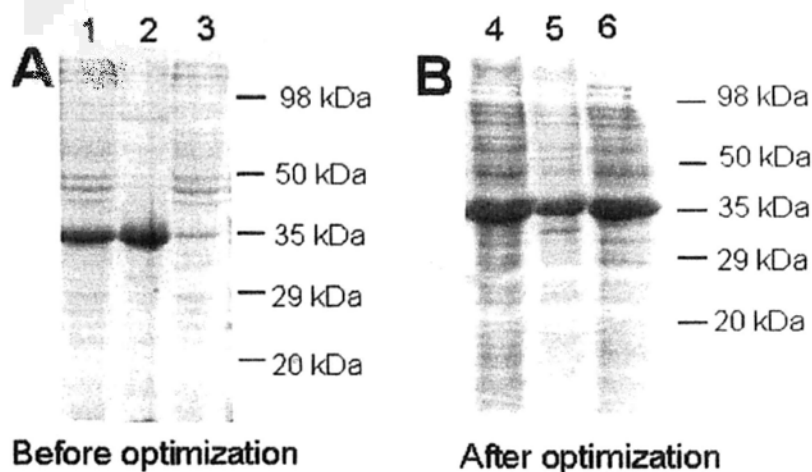


Figure 4.2 Optimization of the expression of selenomethionyl HS-DBE₁₆₋₁₉₇. Selenomethionyl HS-DBE₁₆₋₁₉₇ was expressed to enable selenium based MAD phasing experiment. The samples were analyzed in SDS-PAGE. Before optimization (**A**), the protein was expressed mainly in insoluble fraction (lane 2) as shown in the left panel (lane 1-3): lane 1, lysed *E. coli* strain *B834S* sample; lane 2, pelleted sample after centrifugation of bacterial lysate; lane 3, supernatant sample after centrifugation of bacterial lysate. After optimization (**B**), a significant portion of soluble protein expression was observed (lane 6) as shown in the right panel (lane 4-6); lane 4, lysed *E. coli* strain *C41* sample; lane 5, pelleted sample after centrifugation of bacterial lysate; lane 6, supernatant sample after centrifugation of bacterial lysate. The optimal expression condition consists in first growing the *E. coli* strain *B834S* in LB medium, then changed to selenomethionine containing minimal M9 medium for protein expression. Details of the optimal expression condition were described in Section 2.3.2.

	Selenium labeled crystal
Beamline	BL17U, SSRF (China)
Detector	MAR 225 CCD
Space group	$P2_12_12_1$
Unit-cell parameters (\AA , $^\circ$)	$a = 69.84$, $b = 78.51$, $c = 79.36$, 90, 90, 90

Datasets	Peak	Inflection	Remote
Wavelength (\AA)	0.9794	0.9796	0.9719
Total rotation range ($^\circ$)	150	180	180
Resolution range (\AA)	39.68-2.49 (2.62-2.49)	38.95-2.75 (2.90-2.75)	39.6-2.80 (2.95-2.80)
Total no. of measured intensities	94,745	79,136	73,047
Unique reflections	15,830	11,284	11,168
Multiplicity/ Anomalous	6.0 (6.1) /	6.5 (6.6) /	7.0 (7.1) / 3.8
Multiplicity	3.2 (3.2)	3.5 (3.5)	(3.8)
Completeness/ Anomalous	99.9 (100) /	98.1 (97.6) /	98.9 (98.5) /
Multiplicity (%)	99.9 (100)	98.0 (97.3)	98.8 (98.4)
Mean $I/\sigma(I)$	11.6 (4.7)	10.8 (3.3)	10.0 (4.3)
$^\dagger R_{\text{merge}}$ (%)	7.5 (22.4)	8.7 (36.5)	10.2 (30.9)
Overall B factor from Wilson plot (\AA^2)	39.8	54.1	52.7

$^\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ are the observed intensity and the mean intensity of related reflections, respectively.

Table 4.7 Data collection and processing statistics of selenomethionyl DBE₁₆₋₁₉₇ crystal. The numbers in parentheses correspond to the values in the highest resolution shell. Diffraction data was integrated by *MOSFLM* and scaled by *SCALA*.

4.5 Discussion

4.5.1 Twinning and diffraction anisotropy analyses of the DBE₁₆₋₁₉₇ crystal

Although a twin law $(-h, l, k)$ operator of the DBE₁₆₋₁₉₇ crystal can be identified, no twinning was indicated by the various twinning analyses (L-test, H-test, Britton analysis and RvsR analysis) in *phenix.xtriage*,

The DBE₁₆₋₁₉₇ crystals often grew as stacked plates with occasional single plates, about 4 single crystal plates per 24 well sitting drop plate. A single plate-like crystal was taken for data collection. Although the data can be integrated and scaled to 1.98 Å, a problem of diffraction anisotropy was noted (Figure 4.3). Diffraction anisotropy results from directional dependence in diffraction quality, and it can stall structure refinement at high R-factors and electron density maps can appear featureless, thus inhibiting model building efforts (Strong et al., 2006). In DBE₁₆₋₁₉₇ crystals, it can be observed that the diffraction pattern is more anisotropic in some angles than others (Figure 4.3 A and B). Also, a systematic anisotropy analysis by *Scala* indicated that the mean amplitude of an axis falls off much earlier than the other 2 axes (Figure 4.3 C). However, it was impossible to avoid integrating the “bad” images without compromising data completeness because of the space group $P2_12_12_1$ limitation. The occurrence of diffraction anisotropy may indicate DBE₁₆₋₁₉₇ crystal

packing interactions being more uniform in one direction than another.

Detection of diffraction anisotropy prompted the search for other DBE₁₆₋₁₉₇ crystal forms (Chapter 5), but none could be found at the moment. The problem, however, could be alleviated by the recently established Diffraction Anisotropic Server (DAS) (Strong et al., 2006). This server could help improve the problem by providing ellipsoidal truncation and anisotropic scaling (Strong et al., 2006). Hence, even though severe anisotropy was noted in DBE₁₆₋₁₉₇ crystals, it was thought that the problem could be circumvented by the help of DAS. Hence, attempts to determine DBE₁₆₋₁₉₇ structure with the current crystal form proceeded.

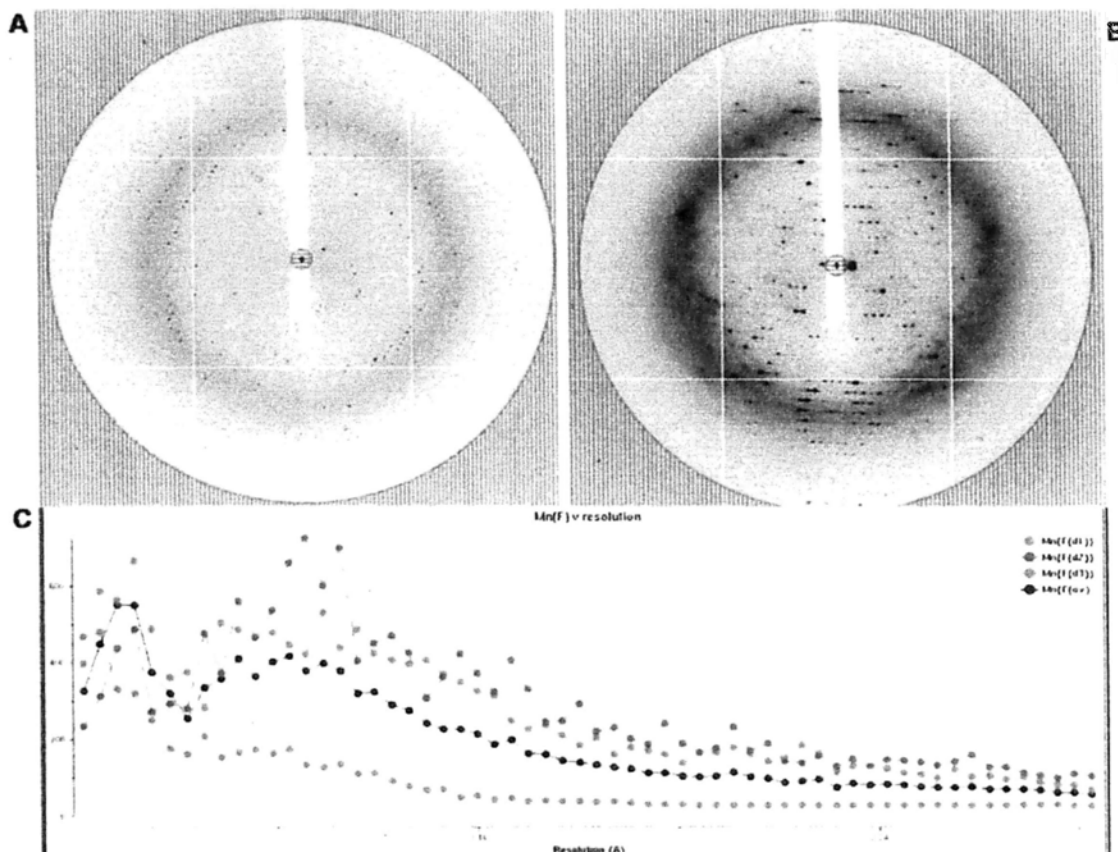


Figure 4.3 Diffraction anisotropy of the DBE₁₆₋₁₉₇ crystal. (A) A diffraction image of a DBE₁₆₋₁₉₇ crystal. (B) A diffraction image of the same DBE₁₆₋₁₉₇ crystal after turning the crystal 90°. It should be noted that the diffraction pattern is more anisotropic in (B) than that in (A); the spots in (A) are circular, while the spots in (B) are more streaked. (C) A systematic analysis on the extent of diffraction anisotropy of DBE₁₆₋₁₉₇ crystal was performed by *Scala* via *CCP4i* after merging. The overall mean amplitude (black) and the mean amplitudes of 3 different axes (red, blue and green) are plotted against resolution. It can be seen that the mean amplitude of one axis (red) falls off at around 4 Å which is much earlier than that of the overall mean (black) and the other two axes (blue and green).

4.5.2 Why molecular replacement failed?

It has been suggested that DBE family is closely related to Dim2p family, a class of conserved KH domain protein that is found in archaeas and eukaryotes, whereas DBE family is found only in eukaryotes (Vanrobays et al., 2004). The Dim2p protein is one of the *trans*-factors for 40S subunit synthesis (Vanrobays et al., 2004; Vanrobays et al., 2008; Jia et al., 2010). The archaea *Aeropyrum pernix* and *Pyrococcus horikoshii* Dim2p crystal structures (PDB code: 1tua and 2e3u) had been determined such that 2 KH domains are found in one single Dim2p protein molecule (Jia et al., 2007).

Dim2p crystal structures were originally thought to be helpful in phasing DBE₁₆₋₁₉₇ crystal structure, because the two families are reported to be evolutionarily closely related and Dim2p is a KH domain protein whereas DBE possibly carries a KH-like domain. However, molecular replacement using the two Dim2p crystal structures as search models did not work. The failure to phase the crystal structure may be explained by the fact that DBE shares only 14.6% and 20% sequence identity with 1tua and 2e3u Dim2p crystal structures while over 30% sequence identity is a rule of thumb for molecular replacement.

4.5.3 Why experimental phasing failed?

Phase information is critical in determination of crystal structure. Experimental phasing was needed to determine DBE₁₆₋₁₉₇ crystal structure when suitable search models for molecular replacement were absent. Various experimental phasing techniques including sulfur SAD, heavy atom MIR/SIRAS and selenium MAD were tried.

4.5.3.1 Sulfur SAD

Sulfur atom can give detectable anomalous signal at X-ray of wavelength 1.54 Å, but it is often reported that the signal is too weak (Dauter et al., 2002; Dodson, 2003). Nevertheless, it was worth trying since DBE₁₆₋₁₉₇ carries 10 sulfur atoms, so a dataset of native DBE₁₆₋₁₉₇ crystal at 1.54 Å was collected (Table 4.2). However, no anomalous signal was detected as analyzed by *phenix.xtriage*.

4.5.3.2 Heavy atom MIR/SIRAS

Phasing can also be done by introducing heavy atoms into the crystals. The technique MIR often requires substitution of at least two heavy atoms without significantly altering the unit cell parameters of the native crystal. Sometimes, the

anomalous signal generated by a single heavy atom can be used to help phasing, the technique is known as SIRAS, for instance, iodine ($f''= 6.93$) and platinum ($f''= 7.24$) atoms give significant anomalous signal at 1.54 Å. The heavy atoms are often introduced by soaking the native crystals in heavy atom containing solution, or by quick soaking in halide/ I3C containing cryoprotectant solution (Dauter et al., 2000; Beck et al., 2009). It is impossible to predict the soaking condition, so it took a lot of trial and error in varying the heavy atom species, concentration and soaking time, to reach a successful soaking condition. But it takes even more effort in the case of DBE₁₆₋₁₉₇ crystals, as only 4 useful DBE₁₆₋₁₉₇ single crystals per 24-well sitting drop plate on average could be obtained, which meant that the burden to produce quality DBE₁₆₋₁₉₇ crystals for soaking was exacerbated.

Only one condition, namely soaking in 50 mM of PtK₂Cl₄ for 10 minutes, led to heavy metal atom substitution of DBE₁₆₋₁₉₇ crystals (Table 4.4). The substitution resulted in a change of space group from $P2_12_12_1$ to $C222$ and anomalous signal that extended to 4.8 Å was detected. With only one heavy atom substitution and change of space group, no phase information can be derived from the non-isomorphous substituted crystal.

Soaking in 0.35 M iodide solution for 20 seconds resulted in an iodide substituted crystal (Table 4.5). The anomalous signal was not strong enough to phase the structure as it could only be extended to 5.4 Å. The reason of weak signal might be due to low occupancy or incorporation of iodide into less ordered regions.

Other conditions might be too drastic to DBE₁₆₋₁₉₇ crystals that resulted in either cracking the crystals or loss of diffraction pattern despite crystals looked intact. Or the heavy atom species simply could not be accommodated in the DBE₁₆₋₁₉₇ crystals, for instance, the size of the I3C triangle (with a side of 6 Å) might hinder its substitution.

4.5.3.3 Selenium based MAD

Selenium based MAD is a commonly used technique for phasing crystal structure but it requires the use of synchrotron facility. Phasing with selenium MAD was attempted once synchrotron radiation at the recently built Shanghai Synchrotron Radiation Facility (SSRF) was available. The 7 natural methionine residues in DBE₁₆₋₁₉₇ were substituted by selenomethionine, and the 7 selenium atoms should provide enough phasing power for DBE₁₆₋₁₉₇, a 21 kDa protein.

Datasets of diffraction at 3 wavelengths, corresponding to peak, inflection, and

remote, were collected, although the data looked decent in terms of diffraction statistics (Table 4.7) and anomalous signals extended to 2.8 Å in the peak dataset, no selenium atom positions could be identified by on site *SHELXD* and later *PHENIX* *hyss* analysis because of two problems, i.e. severe radiation damage and translational pseudo-symmetry, as detected by *SCALA* and *phenix.xtriage* analysis.

Severe radiation damage was detected in selenium incorporated DBE₁₆₋₁₉₇ crystal as indicated by a gradual loss of diffraction power of the crystal and a sharp drop in B factors as analyzed by *SCALA*. Synchrotron radiation is powerful and often leads to crystal damage (Murray and Garman, 2002). Selenomethionine is particularly prone to radiation damage that leads to the loss of anomalous signal and diffractive power of crystals (Holton, 2007).

Pseudo-symmetry can arise when more than a single molecule is present in the asymmetric unit where noncrystallographic symmetry (NCS) operators are close to true crystallographic symmetry (Zwart et al., 2008). Translational pseudo-symmetry occurs when two molecules are related by translational symmetry but have orientations that differ by a root-mean-square displacement (r.m.s.d.) that is below a threshold of 3 Å, and it can lead to problems in deducing the correct unit cell and

space group (Zwart et al., 2008). The DBE₁₆₋₁₉₇ crystals have two protein molecules per asymmetric unit, and translational pseudo-symmetry was detected in both the native and selenomethionyl DBE₁₆₋₁₉₇ crystals by native Patterson analysis in *phenix.xtriage* such that significant (43.8% and 42.9%) off origin peaks at fractional coordinates (-0.500, 0.500, 0.38) and (-0.500, 0.500, 0.039) were identified respectively. Therefore, the problem of translational pseudo-symmetry is an intrinsic problem of the current crystal form of DBE₁₆₋₁₉₇.

In addition to the problems of severe radiation damage and translational pseudo-symmetry, severe diffraction anisotropy was also detected in selenomethionyl DBE₁₆₋₁₉₇ crystal. The 3 problems together made structural determination of DBE₁₆₋₁₉₇ crystal complicated. Hence, it was thought that efforts should be spent on the search for alternative crystal form.

5. Searching for alternative crystal form of DBE₁₆₋₁₉₇

5.1 Introduction

It is generally held that crystallization is the most critical step for structural determination by X-ray crystallography. However, obtaining a crystal does not necessarily guarantee getting a structure. Some complications can arise, e.g. crystals do not diffract; resolution limit too low to be useful; the problem of twinning, pseudo-symmetry and diffraction anisotropy; and some particularly difficult cases are known as pathological crystallography (Dauter et al., 2005). The complications often relate to specific packing problem of the current crystal form. While some problems, e.g. diffraction anisotropy and twinning (Strong et al., 2006), might be circumvented on the course of data processing, some problems can remain intractable, e.g. absence of diffraction or low resolution limit (Heras and Martin, 2005). A search for a new crystal form may thus help bypass the difficulties encountered in the original crystal form, as it is not uncommon for macromolecules to be crystallized in more than one condition.

The DBE₁₆₋₁₉₇ protein can be crystallized and its diffraction data can be integrated and scaled to 1.98 Å (Chapter 4). The current crystal form of DBE₁₆₋₁₉₇ harbored 2 problems, severe diffraction anisotropy and translational pseudo-symmetry, and

selenomethionyl DBE₁₆₋₁₉₇ crystal suffered from bad radiation damage, therefore, the 3 problems together hindered phasing attempt. Hence, alternative crystal form of DBE₁₆₋₁₉₇ was needed.

The current chapter presents the work to look for alternative crystal form of DBE₁₆₋₁₉₇ by various approaches, including the modification of the original chemical environment of the crystallization condition, and manipulation of the protein constructs for crystallization screening

5.2 Seeding and additives screen

The original optimal DBE₁₆₋₁₉₇ crystallization condition was in 500 μ l well solution of DBE_Crys (i.e. 0.6 M ammonium phosphate, pH 4.8), 2 μ l: 4 μ l of protein to well solution ratio in a 24-well sitting drop setting at 16°C for 4 days. In the initial screening for optimal crystallization condition, conditions such as pH and concentration of well solution, growing temperature, ratios of protein to well solution, and hanging drop/ sitting drop had already been tried (Section 4.2.1). New crystal form might be found by modifying the chemical environment of the original crystallization condition, so seeding and additives screening were attempted.

Seeding

Two ways of seeding, i.e. macro-seeding and micro-seeding were attempted. In macro-seeding, a small crystal was carefully rinsed in the stabilizing solution, i.e. 0.6 M ammonium phosphate, pH 4.8. The small crystal was then transferred to a pre-equilibrated DBE₁₆₋₁₉₇ protein solution and allowed for crystal growth. No extra growth of the crystal was observed.

In micro-seeding, a seed stock was made by crushing a number of small crystals, and it was serially diluted by a factor of 5 to produce a dilution series. The dilution series was then streaked to a pre-equilibrated DBE₁₆₋₁₉₇ protein solution and allowed for crystal growth. However, no crystal growth was observed.

Additives screen

It was noted that small molecules can affect the solubility and crystallization of macromolecules. These small molecules can perturb and manipulate sample-sample and sample-solvent interactions, and they can stabilize or engender conformity by specific interaction with the macromolecules. Additive screen HT (Hampton Research) that had 96 chemical species was used. Additive screening was performed in a 96-well plate sitting drop plate, 10 μ l of each additive was added to 90 μ l of DBE_Crys

solution in the reservoir, and 1 μ l of DBE₁₆₋₁₉₇ was mixed with 1 μ l of the additive containing well solution, and it was incubated at 16°C. No crystallization was observed.

5.3 Crystallization condition screening of HS-DBE₁₆₋₁₉₇ and DBE₁₆₋₂₂₉

A HS tag was linked upstream of DBE₁₆₋₁₉₇ to aid expression and purification (Section 3.4.3). As the crystal structure of the SUMO tag was determined (PDB code: 2uyz) (Knipscheer et al., 2007), it was thought that the SUMO tag may help phase the structure, so HS-DBE₁₆₋₁₉₇ was subjected to screening for crystallization condition. The construct was expressed in *E. coli* strain *BL21 (DE3) pLysS* and purified by a Ni²⁺ affinity HiTrap[™] IMAC HP column, followed by a gel filtration HiLoad 26/60 Supdex G75 purification (Figure 5.1). The HS-DBE₁₆₋₁₉₇ protein was concentrated to 8 mg/ml under gel filtration column buffer, i.e. 50 mM NaH₂PO₄, pH 7.4 and 0.4 M NaCl, with slight precipitation. The protein yield was ~15 mg per liter bacterial culture.

The DBE₁₆₋₂₂₉ protein was then found to be an interacting partner of RpS9 (Chapter 6), so it was taken for crystallization screen. The HS-DBE₁₆₋₂₂₉ protein was originally expressed in insoluble form, but it was later optimized to express in soluble form

(Figure 5.2 A). It was first purified by a Ni^{2+} affinity HiTrap[™] IMAC HP column, followed by SENP1c cleavage to cleave the HS tag, then another Ni^{2+} affinity HiTrap[™] IMAC HP column to separate the HS tag from DBE₁₆₋₂₂₉ as the HS tag was captured by the column and DBE₁₆₋₂₂₉ carries no intrinsic Ni^{2+} affinity and was collected as flowthrough. The target protein DBE₁₆₋₂₂₉ was then purified by a gel filtration HiLoad 26/60 Superdex G75 (Figure 5.2 B). Purified DBE₁₆₋₂₂₉ was concentrated to 6 mg/ml under gel filtration buffer, i.e. 50 mM NaH_2PO_4 , pH 7.4 and 0.4 M NaCl, slight precipitation was observed. The yield was ~7 mg per liter bacterial culture.

Concentrated HS-DBE₁₆₋₁₉₇ and DBE₁₆₋₂₂₉ sample were centrifuged to remove precipitation and the samples were then subjected to screening for crystallization condition. Ten crystallization screening kits were tried, i.e. Crystal Screen 1 and 2, Index, SaltRx, PEG/ion screen kits (Hampton Research) and JCSG+, JCSG core I, II, III and IV (QIAGEN). Screening for crystallization condition was performed at 16°C using sitting drop on Greiner CrystalQuick 96 well plates with 3 protein to well solution ratios, 0.1 μl : 0.2 μl , 0.15 μl : 0.15 μl and 0.2 μl : 0.1 μl . No crystal was found.

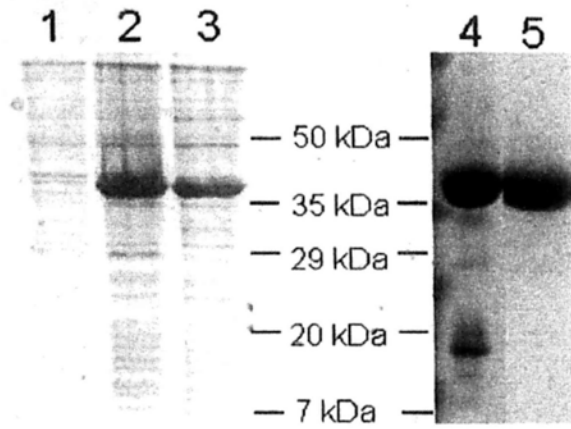


Figure 5.1 Expression and purification profile of HS-DBE₁₆₋₁₉₇. The whole HS-DBE₁₆₋₁₉₇ construct was expressed, purified and subjected to crystallization screening without SENP1c cleavage. Samples taken during the course of expression and purification were analyzed by SDS-PAGE: lane 1, *E. coli* strain *BL21 (DE3)* *pLysS* sample before induction of protein expression; lane 2, bacterial cell sample at cell harvest; lane 3, sample input for Ni²⁺ affinity column purification; lane 4, sample input for gel filtration Superdex G75 purification; lane 5, sample after all purification steps. The HS-DBE₁₆₋₁₉₇ protein can be expressed, and purified to high purity without obvious contamination or degradation.

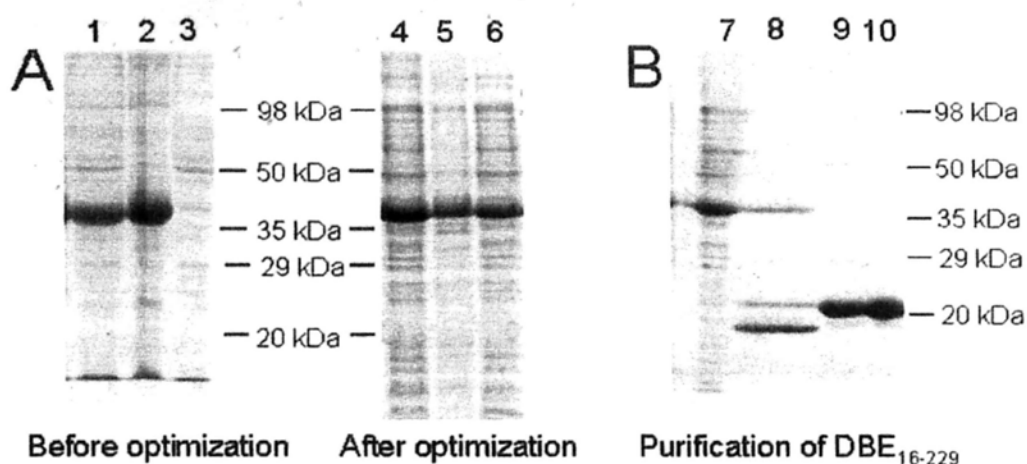


Figure 5.2 (A) Expression optimization of HS-DBE₁₆₋₂₂₉. Optimization of DBE₁₆₋₂₂₉ expression in the presence of HS tag was needed. Samples taken during the course of expression and purification of DBE₁₆₋₂₂₉ were analyzed by SDS-PAGE. Before optimization, the protein was expressed in insoluble fraction (lane 2) as shown in the left panel (lane 1-3): lane 1, lysed *E. coli* strain *C41* sample; lane 2, pelleted sample after centrifugation of bacterial lysate; lane 3, supernatant sample after centrifugation of bacterial lysate. A significant portion of soluble protein expression (lane 6) was observed after optimization as shown in the right panel (lane 4-6); lane 4, lysed *E. coli* strain *C41* sample; lane 5, pelleted sample after centrifugation of bacterial lysate; lane 6, supernatant sample after centrifugation of bacterial lysate. The optimal expression condition was: lowered IPTG concentration (0.05 mM) induction at low OD₆₀₀ reading (0.1) and bacterial culture was grown at 20°C. **(B) Purification profile of DBE₁₆₋₂₂₉ with the aid of HS tag.** The DBE₁₆₋₂₂₉ purification scheme was similar to that of DBE₁₆₋₁₉₇. Lane 7, sample input for Ni²⁺ affinity column purification; lane 8, sample input for another Ni²⁺ affinity column purification after SENP1c cleavage, the target DBE₁₆₋₂₂₉ protein corresponded to the middle band, and the lower band was the cleaved HS tag, and the uppermost band was the uncleaved HS-DBE₁₆₋₂₂₉; lane 9, sample input for gel filtration Superdex G75 purification; lane 10, sample after all purification steps. The DBE₁₆₋₂₂₉ protein was purified to high purity despite minor contaminations.

5.4 Crystallization condition screening of DBE₁₆₋₁₉₇ point mutants

Besides manipulating the protein constructs as mentioned in Section 5.3, the DBE₁₆₋₁₉₇ protein can be modified to replace some long and charged amino acid residues by those that might favor crystallization. Two groups of mutants were constructed so that the first group was designed by swapping of lysine/arginine amino acid residues, and the second group was designed by mutating clustered flexible amino acid residues to alanine.

5.4.1 Design of 3 charged-swapping DBE₁₆₋₁₉₇ mutants

Lysine is considered to be a side chain of high conformational entropy, hence it is suggested that mutating lysine to arginine helped crystallization of protein (Czepas et al., 2004). The DBE₁₆₋₁₉₇ protein contains 16 lysine residues, mutating all those to arginine would generate too many constructs to work with. A selection scheme based on protein sequence alignment of DBE family was designed, so that three positions, R87, K97 and K138, where lysine/ arginine residues interchanged across DBE family were identified (Figure 5.3). The evolutionarily allowed swapping of lysine/ arginine suggested that replacing the original 3 positions with lysine/arginine would not affect protein function significantly. Therefore, 3 point mutants R87K, K97R, K138R were designed. The 3 mutant constructs were generated by quick change mutagenic PCR.

Expression, purification and crystallization of the protein are described in Section 5.4.3.

5.4.2 Design of 3 clustered residues to alanine DBE₁₆₋₁₉₇ mutants

Besides lysine, glutamine, glutamate and aspartate are also considered to be carrying high entropy that may hinder protein crystallization, thus mutation of these residues to alanine may help crystallization (Goldschmidt et al., 2007). A server named Surface Entropy Reduction prediction (SERp) was recently established to predict amino acid residues of the highest surface entropy (Goldschmidt et al., 2007). With the help of SERp server, 3 DBE₁₆₋₁₉₇ mutants were designed by having clustered disordered amino acid residues substituted by alanine, i.e. QE29A, EE38A, KKE135A (Figure 5.3). The 3 mutant constructs were generated by overlapping PCR. Expression, purification and crystallization of the protein are described as below.

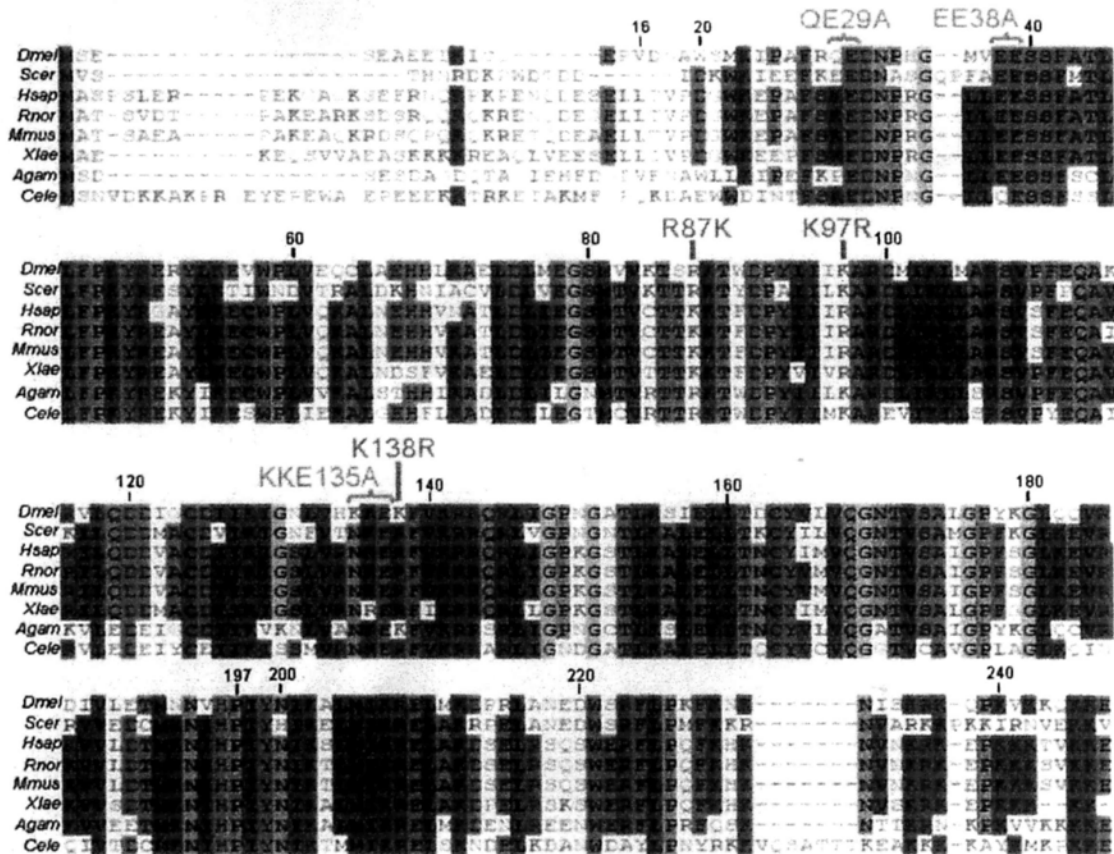


Figure 5.3 Design of the DBE₁₆₋₁₉₇ mutants. The protein sequences of 8 members of DBE family are aligned by CustalW and region from the first to around 240th amino acid residue of DBE is shown. The 8 species are: fruit fly, *Dmel*, *Drosophila melanogaster*; budding yeast, *Scer*, *Saccharomyces cerevisiae*; human, *Hsap*, *Homo sapien*; rat, *Rnor*, *Rattus norvegicus*; mouse, *Mmus*, *Mus musculus*; frog, *Xlae*, *Xenopus laevis*; mosquito, *Agam*, *Anopheles gambiae*; roundworm, *Cele*, *Caenorhabditis elegans*. Amino acid residues that are largely conserved are highlighted in boxes. The positions of the residue 16 and 197 of DBE are marked. Three positions that lysine and arginine residues interchange among the 8 proteins are marked, corresponding to the 3 charged-swapping mutants R87K, K97R and K138R. Three positions that carry flexible clustered amino acid residues as suggested by SERp server are marked, corresponding to clustered residues to alanine mutants QE29A, EE38A and KKE135A. The 6 DBE₁₆₋₁₉₇ mutants were constructed in hope of obtaining alternative DBE₁₆₋₁₉₇ crystal form.

5.4.3 Expression, purification and crystallization screening of the 6 DBE₁₆₋₁₉₇ mutants

All the six DBE₁₆₋₁₉₇ mutants shared the same expression and purification scheme as that of HS tagged DBE₁₆₋₁₉₇. They were expressed in *E. coli* strain *BL21 (DE3)* *pLysS*. The tagged protein was first purified by a Ni²⁺ charged HiTrapTM IMAC column. The eluted protein was then subjected to SENP1c cleavage to cleave the HS tag from DBE₁₆₋₁₉₇. The HS tag was then separated from DBE₁₆₋₁₉₇ by passing through a Ni²⁺ charged HiTrapTM IMAC column again, as the HS tag was captured to the column, but DBE₁₆₋₁₉₇ mutants carried no intrinsic Ni²⁺ affinity and were collected in the flowthrough. The proteins were then concentrated to around 5 ml before gel filtration purification by HiLoad 26/60 Superdex G75 (Figure 5.4). The DBE₁₆₋₁₉₇ mutant proteins were eluted at around 180 ml. The proteins were concentrated under gel filtration buffer, i.e. 50 mM NaH₂PO₄, pH 7.4 and 0.4 M NaCl.

The 3 DBE₁₆₋₁₉₇ mutants R87K, K97R and K138R were concentrated to 10 mg/ml without precipitation, giving a yield of ~13 mg per liter bacterial culture; while DBE₁₆₋₁₉₇ mutants QE29A, EE38A and KKE135A were concentrated to 6 mg/ml with slight precipitation, giving a yield of ~5 mg per liter bacterial culture.

Concentrated protein samples were centrifuged to remove precipitation and then subjected to crystallization screening. Ten crystallization screening kits were tested and they were Crystal Screen 1 and 2, Index, SaltRx, PEG/ion screen kits (Hampton Research) and JCSG+, JCSG core I, II, III and IV (QIAGEN). Screening for crystallization condition was performed at 16°C using sitting drop on Greiner CrystalQuick 96 well plates with 3 protein to well solution ratios, 0.1 µl: 0.2 µl, 0.15 µl: 0.15 µl and 0.2 µl: 0.1 µl. No crystal was found.

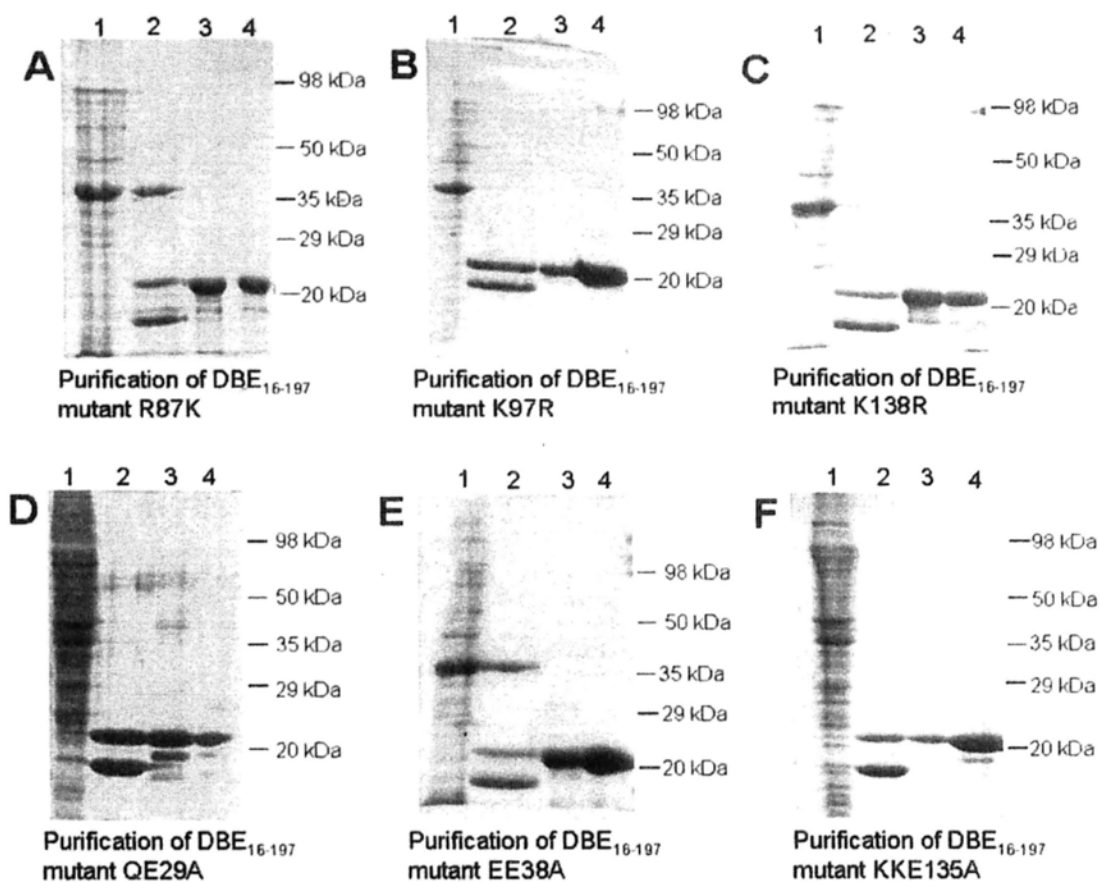


Figure 5.4 Purification profiles of DBE₁₆₋₁₉₇ mutants R87K, K97R, K138R, QE29A, EE38A and KKE135A. Six DBE₁₆₋₁₉₇ mutants were constructed in hope of obtaining alternative DBE₁₆₋₁₉₇ crystal form. The mutant proteins were HS tagged to aid purification, such that the purification scheme of the mutants (A) R87K, (B) K97R, (C) K138R, (D) QE29A, (E) EE38A and (F) KKE135A were similar to that of the DBE₁₆₋₁₉₇ construct. Samples taken during the course of purification were analyzed by SDS-PAGE: lane 1, sample input for Ni²⁺ affinity column purification; lane 2, sample after SENP1c cleavage ready for another Ni²⁺ affinity column purification, the band just above 20 kDa indicated the target mutant proteins, and the lower band was the HS tag; lane 3, sample input for gel filtration Superdex G75 purification; lane 4, sample after all purification steps. All mutant proteins can be expressed, and purified to high purity despite minor contaminations (lane 4).

5.5 Crystallization of 2 DBE₁₆₋₁₉₇ orthologs, Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅

5.5.1 Construction of Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅

Besides manipulating DBE₁₆₋₁₉₇ constructs, crystallizing the corresponding regions of the yeast and human members of DBE family, i.e. Krr1p and HRB2, were attempted. Protein sequence of DBE, Krr1p and HRB2 were aligned, the DBE₁₆₋₁₉₇ corresponding regions are Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ (Figure 5.5 A). Full length Krr1p and HRB2 share 52% and 53% sequence identity to DBE, while Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ share 63% and 66% sequence identity to DBE₁₆₋₁₉₇. The *krr1p*₁₃₋₁₉₆ and *hrb2*₂₈₋₂₁₅ sequence were cloned, similar to *dbe*₁₆₋₁₉₇, with an upstream HS tag into modified *pET3d* expression vector.

5.5.2 Expression, purification and crystallization screening of Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅

The Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ proteins were expressed in *E. coli* strain *C41*. Their purification schemes were the same as that of HS tagged DBE₁₆₋₁₉₇. The HS tagged Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ proteins were first purified by a Ni²⁺ charged HiTrapTM IMAC column. The eluted proteins were then mixed with SENP1c to cleave the HS tag from the target proteins. The HS tag was then separated from the target proteins by passing through a Ni²⁺ charged HiTrapTM IMAC column again, as the HS tag was

captured to the column, but the target proteins carried no intrinsic Ni²⁺ affinity and were collected in the flowthrough. The target proteins were then concentrated to around 5 ml before gel filtration purification by HiLoad 26/60 Superdex G75 (Figure 5.5 B and C). The Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ proteins were eluted at around 180 ml. The proteins were concentrated under gel filtration buffer, i.e. 20 mM Tris, pH 7.4 and 0.3 M NaCl. Both proteins were concentrated to 13 mg/ml without precipitation. The yields of both proteins were ~ 10 mg per liter bacterial culture.

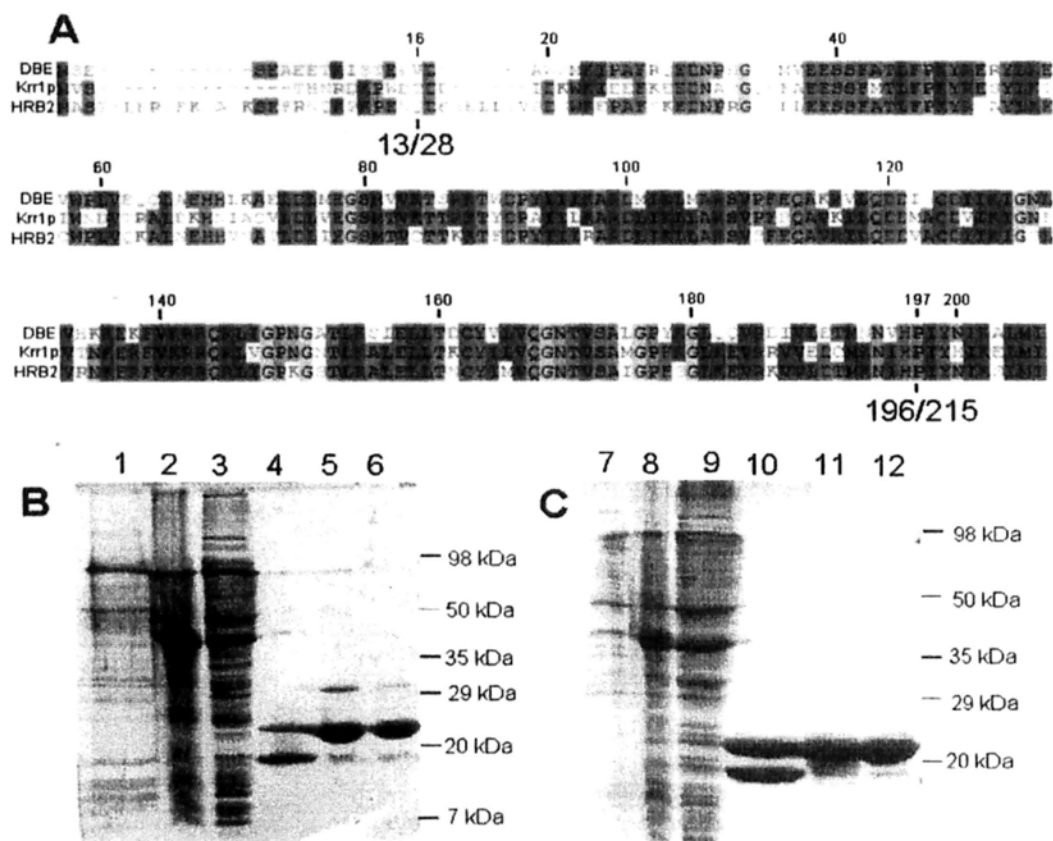


Figure 5.5 (A) Protein sequence alignment of DBE, yeast homolog Krr1p and human homolog HRB2. The protein sequences of DBE, yeast homolog Krr1p and human homolog HRB2 were aligned by CustalW. The alignment from the first residue to residue 206 of DBE is shown. Amino acid residues that are largely conserved are highlighted by boxes. The residues 13 of Krr1p and 28 of HRB2 correspond to the residue 16 of DBE; while the residues 196 of Krr1p and 215 of HRB2 correspond to the residue 197 of DBE as marked in the alignment. Therefore, Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ are the corresponding constructs of DBE₁₆₋₁₉₇. **(B) Expression and purification scheme of Krr1p₁₃₋₁₉₆ and (C) HRB2₂₈₋₂₁₅ with the aid of HS tag.** Both the Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ protein were HS tagged to aid expression and purification. Samples taken on the course of expression and purification were analyzed by SDS-PAGE: lanes 1/7, *E. coli* strain C41 sample before induction of protein expression; lanes 2/8, bacterial cell sample at cell harvest; lanes 3/9, sample input for Ni²⁺ affinity column purification; lanes 4/10, sample after SENP1c cleavage ready for another Ni²⁺ affinity column purification, the band just above 20 kDa was the target protein and the lower band was the HS tag; lanes 5/11, sample input for gel filtration G75 purification; lanes 6/12, sample after all purification steps. Both proteins can be expressed, and purified to high purity despite minor contaminations (lane 6/12).

Concentrated protein samples were subjected to screening for crystallization condition. Ten crystallization screening kits were tested, i.e. Crystal Screen 1 and 2, Index, SaltRx, PEG/ion screen kits (Hampton Research) and JCSG+, JCSG core I, II, III and IV (QIAGEN). Screening for crystallization condition was performed at 16°C using sitting drop on Greiner CrystalQuick 96 well plates with 3 protein to well solution ratios, 0.1 µl: 0.2 µl, 0.15 µl: 0.15 µl and 0.2 µl: 0.1 µl.

Six initial hits were observed in Krr1p₁₃₋₁₉₆, but only 2 of them can be reproduced in 24-well plates (Table 5.1). The optimal condition for crystal growth was identified: in a 24 well sitting drop plate, with 500 µl of Kr_Crys solution: 36% (v/v) PEG400, 50 mM Tris-Cl pH 9.0, 50 mM Li₂SO₄, 50 mM Na₂SO₄ and 100 mM MgCl₂, protein to precipitant ratio of 4 µl: 2 µl at 16°C. The optimal crystal growing condition was achieved after extensive screening from various pH (7.0 to 10.0) and concentrations of Li₂SO₄, (0 to 500 mM), Na₂SO₄ (0 to 500 mM), MgCl₂ (0 to 1 M) and PEG400 (15% to 40%), incubation temperature (4°C, 16°C and 25°C), various ratios of protein to well solution (1:1, 1:2, 2:1, 3:1), and hanging drop/ sitting drop setting. Crystals often grew in the shape of needle with a size of ~ 0.2 x 0.03 x 0.03 mm after 7 days of incubation (Figure 5.6). However, Krr1p₁₃₋₁₉₆ crystals did not show diffraction either with cryo-cooling or at room temperature.

Crystallization screen kits	Components	Reproducible in 24-well plate?
Natrix	100 mM KCl, 10 mM MgCl ₂ , 50 mM Tris-Cl, pH 8.5, 30 % (v/v) PEG400	Yes
JCSG +	200 mM Mg formate, 20 % (w/v) PEG3350	No
JCSG I	50 mM Li ₂ SO ₄ , Na ₂ SO ₄ , 50 mM Tris-Cl, pH 8.5, 30 % (v/v) PEG400	Yes
JCSG II	200 mM NaCl, 100 mM Na/K H ₂ PO ₄ , pH 6.2, 20 % (w/v) PEG1000	No
Index	50 mM CaCl ₂ , 100 mM BisTris, pH 6.5, 30 % (v/v) PEG500 monomethyl ether	No
Index	50 mM MgCl ₂ , 100 mM HEPES, pH 7.5, 30 % (v/v) PEG500 monomethyl ether	No

Table 5.1 Initial crystallization conditions of Krr1p₁₃₋₁₉₆. Only 2 conditions can be reproduced out of the 6 initial hits.

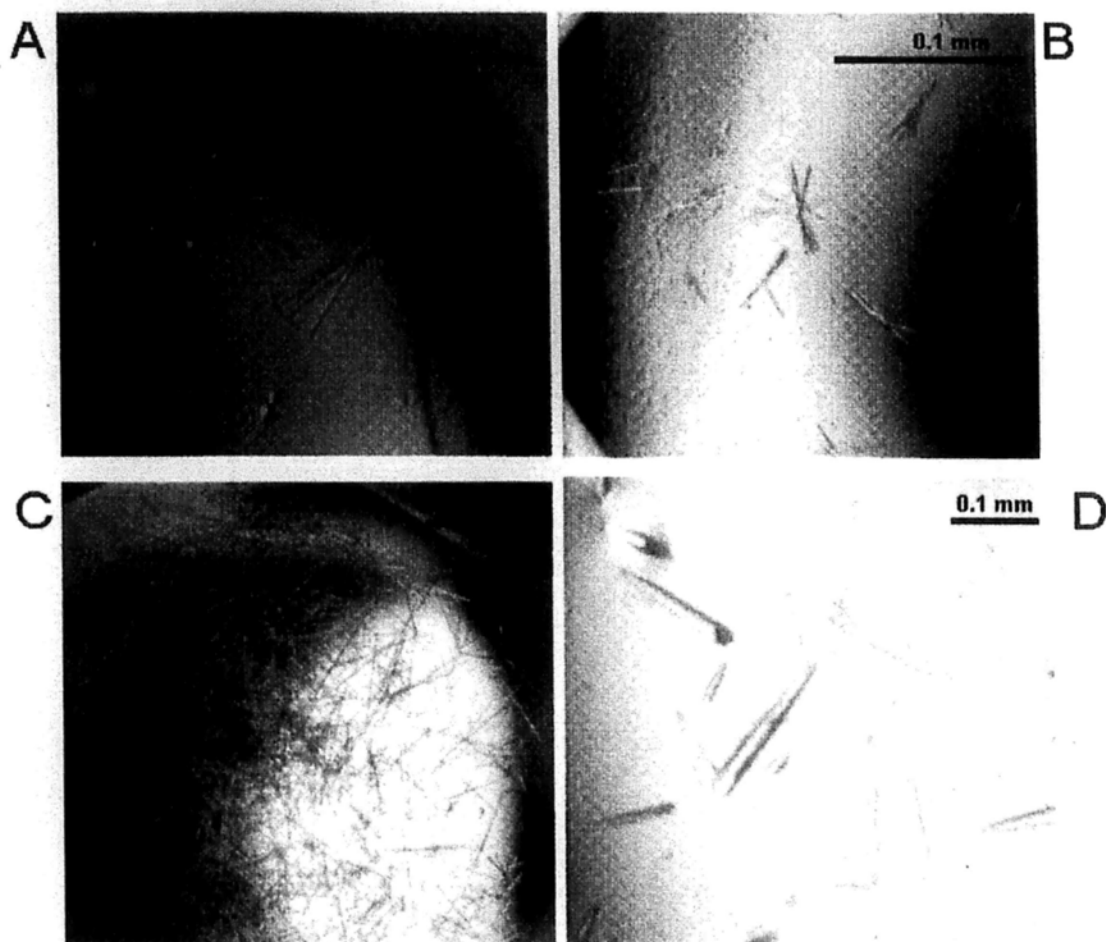


Figure 5.6 Images of Krr1p₁₃₋₁₉₆ crystals. (A) The initial clustered needle shaped crystals were obtained in one condition of Natrix screening kit, i.e. 100 mM KCl, 10 mM MgCl₂, 50 mM Tris-Cl, pH 8.5, 30 % (v/v) PEG400. (B) The initial clustered needle shaped crystals were obtained in one condition of JCSG I screening kit, i.e. 50 mM Li₂SO₄, Na₂SO₄, 50 mM Tris-Cl, pH 8.5, 30 % (v/v) PEG400. (C) During the optimization process, most crystals grew in long (~0.3 mm) but very thin and clustered needle shaped crystals. (D) Single Krr1p₁₃₋₁₉₆ crystals (~ 0.2 x 0.03 x 0.03) were obtained after optimization. The optimal crystallization condition was, in a 24 well sitting drop plate, with 500 μ l of Kr_Crys solution: 36% (v/v) PEG400, 50 mM Tris-Cl pH 9.0, 50 mM Li₂SO₄, 50 mM Na₂SO₄ and 100 mM MgCl₂, protein to precipitant ratio of 4 μ l: 2 μ l at 16°C for 7 days.

Three initial hits were observed in HRB2₂₈₋₂₁₅ as listed in Table 5.2 and Figure 5.7.

The outlook of HRB2₂₈₋₂₁₅ crystals very much resembled that of Krr1p₁₃₋₁₉₆ crystals, they were all clustered and needle shaped. The HRB2₂₈₋₂₁₅ crystal from the JCSG+ screening kit, i.e. 200 mM NaCl, 100 mM Na/K H₂PO₄, pH 6.2, 50 % (v/v) PEG200 was large enough for X-ray diffraction, but no diffraction was observed at room temperature. Crystals from the other 2 conditions were too small to diffract.

Crystallization screen	Components
JCSG +	200 mM NaCl, 100 mM Na/K H ₂ PO ₄ , pH 6.2, 50 % (v/v) PEG200
JCSG IV	100 mM HEPES, pH 7.5, 5 % (w/v) PEG3000, 40 % (v/v) Ethylene glycol
Index	50 mM CaCl ₂ , 100 mM BisTris, pH 6.5, 30 % (v/v) PEG500 monomethyl ether

Table 5.2 Initial crystallization conditions of HRB2₂₈₋₂₁₅. Three initial hits of HRB2₂₈₋₂₁₅ were observed.

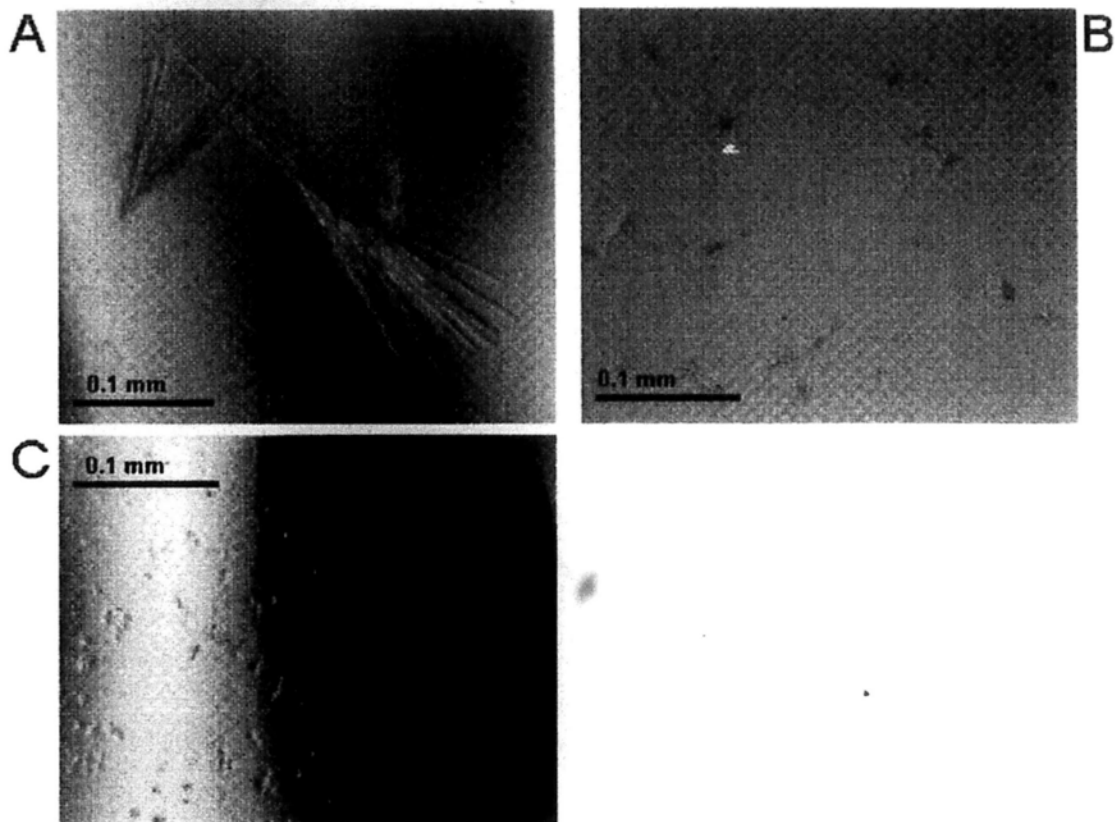


Figure 5.7 Images of HRB₂₈₋₂₁₅ crystals. (A) The initial clustered needle shaped crystals were obtained in one condition of the JCSG+ screening kit, i.e. 200 mM NaCl, 100 mM Na/K H₂PO₄, pH 6.2, 50 % (v/v) PEG200. (B) Very small and clustered needle shaped crystals were obtained in one condition of the JCSG IV screening kit, i.e. 100 mM HEPES, pH 7.5, 5 % (w/v) PEG3000, 40 % (v/v) Ethylene glycol. (C) Very small and clustered needle shaped crystals were obtained in one condition of the Index screening kit, 50 mM CaCl₂, 100 mM BisTris, pH 6.5, 30 % (v/v) PEG500 monomethyl ether.

5.5.3 Attempts to improve Krr1p₁₃₋₁₉₆ crystal diffraction

All Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ crystals looked alike as clustered needles and they did not show diffraction pattern at room temperature, suggesting that the lack of diffraction was not due to problems in cryo-cooling. As crystallization condition had been optimized in Krr1p₁₃₋₁₉₆ crystals, it was hoped that further treatments on Krr1p₁₃₋₁₉₆ crystals might improve diffraction. Several techniques, including seeding, additive screen, crystal desiccation and crystal cross-linking, were attempted to help improve Krr1p₁₃₋₁₉₆ crystals.

Seeding

Both macro-seeding and micro-seeding were attempted. In macro-seeding, a small crystal was carefully fished out and rinsed in the Kr_Crys solution i.e. 36% (v/v) PEG400, 50 mM Tris-Cl pH 9.0, 50 mM Li₂SO₄, 50 mM Na₂SO₄ and 100 mM MgCl₂. The small crystal was then transferred to a pre-equilibrated Krr1p₁₃₋₁₉₆ protein solution and allowed for crystal growth. No extra growth of crystal was observed.

In micro-seeding, a seed stock was made by crushing a number of small crystals, and it was serially diluted by a factor of 5 to produce a dilution series. The dilution series was then streaked to a pre-equilibrated Krr1p₁₃₋₁₉₆ protein solution and allowed

for crystal growth. However, no crystal growth was observed.

Additives screen

Additive screen HT (Hampton Research) that contained 96 chemical species was used. Additive screening was performed in a 96-well sitting drop plate, 10 μ l of each additive was added to 90 μ l of Kr_Crys solution in the reservoir, and 1 μ l of Krr1p₁₃₋₁₉₆ was mixed with 1 μ l of the additive containing Kr_Crys solution, and it was incubated at 16°C. No crystal growth was observed.

Crystal desiccation

It was noted that fast desiccation of protein crystal could help improve diffraction (Abergel, 2004; Heras and Martin, 2005). Hence, the non-diffracting frozen Krr1p₁₃₋₁₉₆ crystals were unmounted and quickly plunged into a well containing cryoprotectant (25% glycerol/ 25% ethylene glycol) added Kr_Crys solution and soaked for 15 minutes to 1 hour. The crystals were then remounted for X-ray diffraction. However, no diffraction pattern was observed.

Crystal cross-linking

It was reported that crosslinking protein crystal by glutaraldehyde can improve

diffraction quality (Heras and Martin, 2005; Salem et al., 2010). Crystal cross-linking was induced by vapor-diffusion of glutaraldehyde. Sitting drop plate was used, and crystal cross-linking was induced by vapor-diffusion of glutaraldehyde, so that the bridge contained 5 μ l of 25% glutaraldehyde, and Krr1p₁₃₋₁₉₆ crystals were hanged in Kr_Crys solution under the cover slip (Figure 5.8). The reaction was allowed to stand for 60 minutes. The crystals were then taken to X-ray diffraction, however, no diffraction was observed.

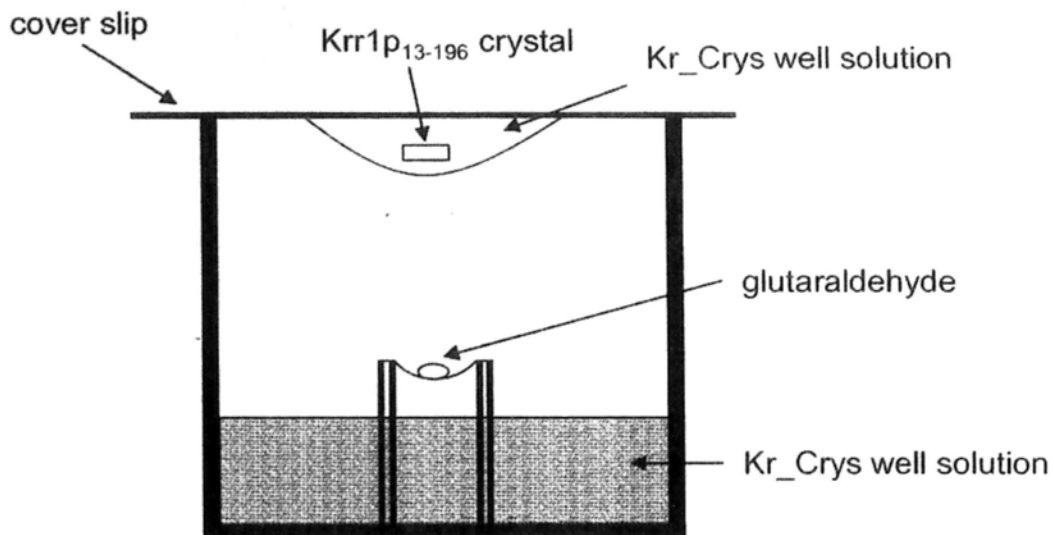


Figure 5.8 Schematic representation of the experimental setup for crystal cross-linking. The Krr1p₁₃₋₁₉₆ crystals were subjected to crystal cross-linking in the hope of improving diffraction. In a 24-well sitting drop setting that contained 500 μl of Kr_Crys well solution (Section 5.5.3), the Krr1p₁₃₋₁₉₆ crystal was hanged in the Kr_Crys well solution under a cover slip, and 5 μl of 25% glutaraldehyde was placed on the bridge. Crystal was cross-linked by vapor diffusion of glutaraldehyde.

5.6 Discussion

It has been estimated that only about 15% of protein targets that were purified would then crystallize in a form that allowed structural determination (Dong et al., 2007). A number of techniques have been devised to help improve crystals, like modifying the chemical environment of crystallization conditions, modifying protein constructs, or working on other homologous proteins (Derewenda, 2010). But still, obtaining a crystal that allows structural determination remains largely trial and error, which means there is virtually no end in screening and tinkering with crystallization conditions. Although the various attempts mentioned in previous sections failed to obtain quality crystals that allowed structural determination, still more can be done to improve crystallization. Three directions are proposed here.

In situ proteolysis

The technique *in situ* proteolysis, i.e. adding trace amounts of protease into protein samples for crystallization screen, has been reported to help crystallization (Dong et al., 2007; Wernimont and Edwards, 2009). Thirty eight structures were determined from 270 proteins that had failed in the past to produce crystal, a success rate of 12.6% was noted by this approach in the test cases (Wernimont and Edwards, 2009). Specific proteases such as trypsin, chymotrypsin, thermolysin, papain and V8

protease were used that resulted in crystals that were suitable for structure determination (Wernimont and Edwards, 2009). Therefore, it remains possible a new crystal form may arise by using the above mentioned specific proteases to treat DBE₁₆₋₁₉₇, Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅, and then take the treated samples for screening for crystallization condition.

More DBE₁₆₋₁₉₇ mutant constructs

The 3 DBE₁₆₋₁₉₇ point mutants, R87K, K97R and K138R did not yield crystals. Apart from lysine to arginine mutations, changing lysine to alanine and glutamate to alanine mutations were also reported to help crystallization (Garrard et al., 2001; Longenecker et al., 2001). The DBE₁₆₋₁₉₇ protein contains 16 lysine and 13 glutamate residues. From sequence alignment of DBE family, 3 lysine residues, K115, K135, K179, and 2 glutamate residues, E62, E73 are found to be not strictly conserved (Figure 5.9). The 5 positions can be chosen to create new DBE₁₆₋₁₉₇ mutants, for example, K115A, K115R, K135A, K135R, K179A, K179R, E62A and E73A. The new mutants can be expressed and purified in scheme similar to that of the original DBE₁₆₋₁₉₇ protein and then subjected to screening for crystallization condition.

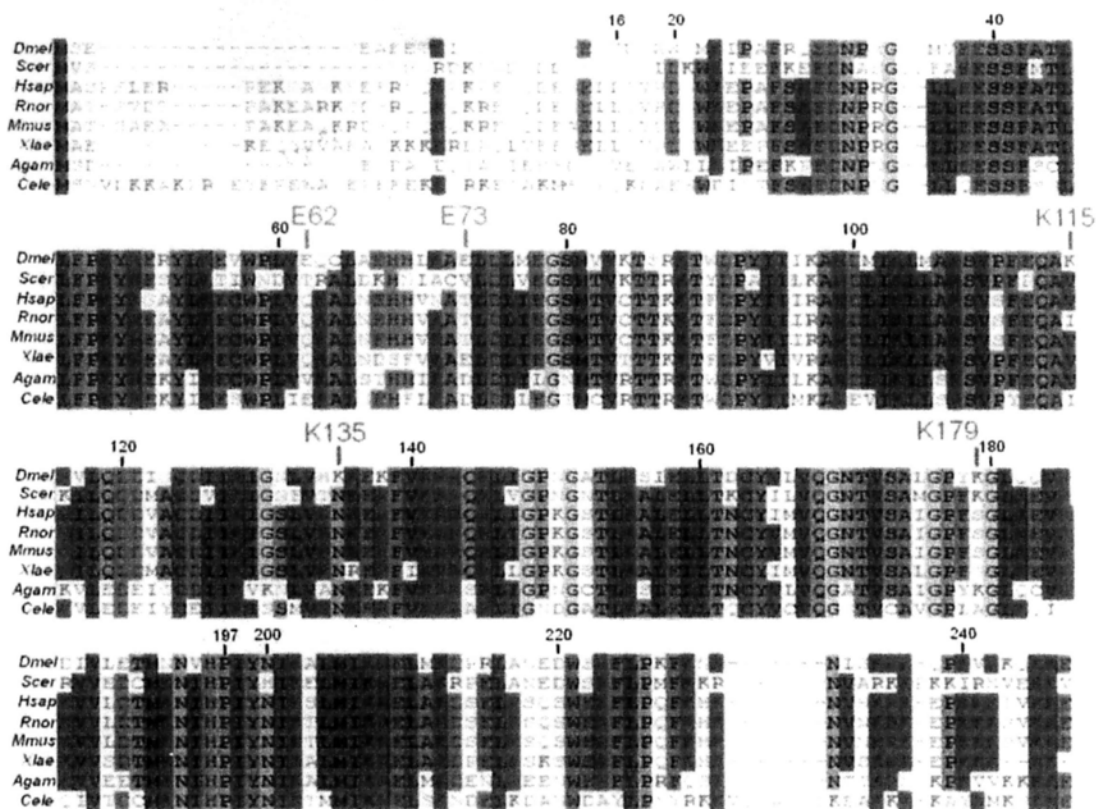


Figure 5.9 Possible positions for DBE₁₆₋₁₉₇ point mutants. The protein sequences of 8 members of DBE family are aligned by CustalW and region from the first residue to residue 247 of DBE is shown. The 8 species are: fruit fly, *Dmel*, *Drosophila melanogaster*; budding yeast, *Scer*, *Saccharomyces cerevisiae*; human, *Hsap*, *Homo sapiens*; rat, *Rnor*, *Rattus norvegicus*; mouse, *Mmus*, *Mus musculus*; frog, *Xlae*, *Xenopus laevis*; mosquito, *Agam*, *Anopheles gambiae*; roundworm, *Cele*, *Caenorhabditis elegans*. Amino acid residues that are largely conserved are highlighted in boxes. The positions of the residues 16 and 197 of DBE are marked. Two glutamate residues E62 and E73, and 3 lysine residues K115, K135 and K179 that are not strictly conserved across the 8 proteins are marked, such that they can be mutated to generate lysine to arginine/alanine or glutamate to alanine mutants to facilitate crystallization.

Redesigning of deletion construct of DBE₁₆₋₁₉₇

The construct DBE₁₆₋₁₉₇, as well as the corresponding Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅, did not yield crystal that is suitable for structural determination. Therefore, it remains possible that manipulating the N- and C- deletion positions of DBE may help crystallization. From previous observations, DBE₁₆₋₁₉₇ expression was good, while DBE₂₄₋₁₉₇ did not express and DBE₁₆₋₂₁₃ did not express in soluble form (Chapter 3). It implies that soluble protein expression level decreased from residues 16 to 24 in the N- terminus, and from residues 197 to 213 residues in the C- terminus. Therefore, if every 4 residues are to be deleted per step, 2 more positions in between the residue 16 to 24, and 4 more positions in between the residue 197 to 213 can be designed that yield 8 more protein constructs in total. While there is no golden rule for the number of residues per cut, 4 residues per cut can be considered to be reasonable (Graslund et al., 2008). The newly designed constructs can be cloned, expressed and purified similarly to that of HS-DBE₁₆₋₁₉₇ and then subjected to screening for crystallization condition.

6. Identification of RpS9 binding region in DBE

6.1 Introduction

Ribosome biogenesis proceeds via large pre-rRNP complexes in which pre-rRNA processing and ribosome assembly occur simultaneously (reviewed by (Fromont-Racine et al., 2003; Henras et al., 2008)). Therefore mutating a single ribosomal protein, e.g. RpS9 (Ferreira-Cerca et al., 2005) or *trans*-factor, e.g. DBE (Chan et al., 2001) is enough to affect pre-rRNA processing and halt ribosome synthesis. The earliest detectable pre-rRNP complex for yeast ribosome biogenesis is named 90S pre-ribosome, in which RpS9 and Krr1p, the yeast homologue of DBE family, are found (Dragon et al., 2002; Grandi et al., 2002; Bernstein et al., 2004). Besides being found in the pre-rRNP complex for early ribosome biogenesis, RpS9/DBE family interaction are noted in the following findings. First, RpS9 and HRB2, the human homolog of DBE, were observed to share similar kinetics of nucleolar localization after the inhibition of rDNA transcription, suggesting that they may participate in ribosome biogenesis at similar stage (Andersen et al., 2005). Second, RpS9 genetically interacted with Krr1p, such that the growth defect phenotype of *krr1* mutant was rescued by overexpression of RpS9 (Gromadka et al., 2004). Third, RpS9 showed a direct physical interaction to DBE (Choi, 2007). Taken all 3 findings together, it might be implied that the RpS9/DBE pairing is conserved and important

for early ribosome biogenesis.

It is, however, not known how DBE mediates RpS9 binding. In this chapter, the efforts to map the RpS9 binding region in DBE by GST-pull down assay and attempts to further characterize the DBE/ RpS9 interaction are reported.

6.2 The GST-RpS9 pull down assay of DBE deletion mutants

6.2.1 Purification of DBE deletion mutants

Full length DBE and 4 DBE deletion mutants DBE₁₆₋₁₉₇, DBE₃₄₋₁₉₇, DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉ were used to map the RpS9 binding region on DBE protein. The constructs DBE₁₆₋₁₉₇, DBE₃₄₋₁₉₇ and DBE₃₄₋₂₂₉ were previously designed to locate the potential DBE domain (Chapter 3). The *dbe*₁₋₂₂₉ sequence was cloned into modified *pET3d* expression vector. Expression and purification of DBE₁₆₋₁₉₇ was described in Section 3.4.1. Full length DBE, DBE₃₄₋₁₉₇, DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉ proteins were expressed in *E. coli* strain *BL21 (DE3) pLysS*. The proteins were first purified by a cation exchanger HiTrapTM SP HP and then followed by a gel filtration chromatography HiLoad 26/60 Superdex G75 (Figure 6.1 A).

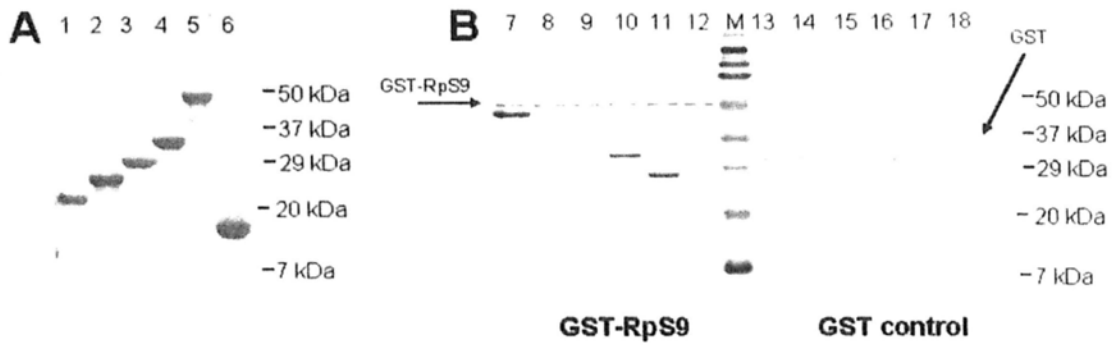


Figure 6.1 (A) Protein inputs for GST-RpS9 pull down experiments. The DBE full length protein and deletion mutants DBE₁₆₋₁₉₇, DBE₃₄₋₁₉₇, DBE₃₄₋₂₂₉ and DBE₁₋₂₂₉ were purified for GST-RpS9 pull down assay. The proteins were of high purity without major contamination as shown in lane 1, DBE₁₆₋₁₉₇; lane 2, DBE₃₄₋₁₉₇; lane 3, DBE₃₄₋₂₂₉; lane 4 DBE₁₋₂₂₉; lane 5, full length DBE; lane 6, lysozyme as negative control. **(B) GST-RpS9 of DBE deletion mutants including DBE, DBE₁₋₂₂₉, DBE₃₄₋₂₂₉, DBE₁₆₋₁₉₇ and DBE₃₄₋₁₉₇.** The RpS9 binding region in DBE was mapped by GST-RpS9 pull down assay. The result of GST-RpS9 pull down was shown in the left panel while that of GST control was shown in the right panel. The GST-RpS9 and GST bands were marked. Lanes 7/13, full length DBE; lanes 8/14, DBE₃₄₋₁₉₇; lanes 9/15, DBE₁₆₋₁₉₇; lanes 10/16, DBE₁₋₂₂₉; lanes 11/17, DBE₃₄₋₂₂₉; lanes 12/18, lysozyme as negative control; lane M, protein standard. It can be observed that only DBE, DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉ showed comparable GST-RpS9 affinity as shown in lanes 7/10/11, and they showed lowered affinity to GST control as shown in lanes 13/16/17, suggesting the GST-RpS9 affinity of these proteins were specific. The DBE₃₄₋₁₉₇ (lanes 8/14) and DBE₁₆₋₁₉₇ (lanes 9/15) proteins showed reduced affinity to both GST-RpS9 and GST. And lysozyme did not show affinity to both GST-RpS and GST as shown in lanes 12/18.

6.2.2 The result of GST-RpS9 pull down assay

The GST-RpS9 and the control GST proteins were expressed by *E. coli* strain *C41*. Bacterial lysates with GST-RpS9 or GST overexpressed were loaded onto Glutathione Sepharose 4 Fast Flow beads before binding with various purified DBE proteins. The beads loaded with GST-RpS9 or GST were washed to remove unbound proteins and then subsequently incubated with 20 µg of various purified DBE proteins and then washed extensively to remove unbound proteins.

The GST-RpS9 pull down experiment used full length DBE and 4 DBE deletion mutants DBE₁₆₋₁₉₇, DBE₃₄₋₁₉₇, DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉ and lysozyme as a negative control (Figure 6.1 A). The result showed that the amounts of DBE, DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉ being pulled down by GST-RpS9 were significantly more than that of the GST control, suggesting that their interactions with GST-RpS9 were specific. On the contrary, DBE₁₆₋₁₉₇ and DBE₃₄₋₁₉₇ showed reduced GST-RpS9 affinity than that of DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉; and DBE₁₆₋₁₉₇ and DBE₃₄₋₁₉₇ showed similar affinity to both GST-RpS9 and GST control, suggesting DBE₁₆₋₁₉₇ and DBE₃₄₋₁₉₇ did not carry specific RpS9 affinity (Figure 6.1 B). Lysozyme as a negative control did not show affinity to either GST-RpS9 or GST (Figure 6.1 B). The results indicated that protein constructs ending at the residue 229, i.e. DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉, but not those ending

at the residue 197, retained similar RpS9 affinity as DBE full length protein, (Figure 6.2). In other words the crystallizable DBE₁₆₋₁₉₇ protein did not carry RpS9 affinity.

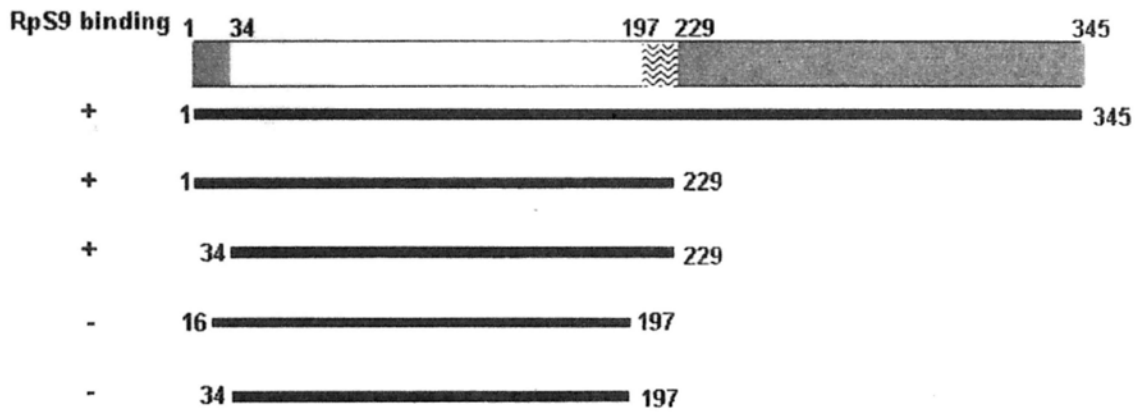


Figure 6.2 Schematic representation of the GST-RpS9 pull down result. As shown by the GST-RpS9 pull down result, the deletion constructs that showed RpS9 binding are indicated as “+”, while constructs that did not show RpS9 binding are indicated as “-”. By comparing the RpS9 affinity of the deletion constructs, it was found that deletion constructs ending with the residue 229, i.e. DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉ retained similar specific RpS9 affinity as full length DBE, and constructs ending with the residue 197, i.e. DBE₁₆₋₁₉₇, DBE₃₄₋₁₉₇, did not show specific RpS9 affinity. It indicated that the RpS9 binding region lies within the region from between residue 34 to 229 of DBE, and the region from the residue 197 to 229 of DBE is necessary for RpS9 binding. In other words, it means the crystallizable DBE₁₆₋₁₉₇ protein did not show RpS9 affinity.

6.3 Further attempts to understand interaction between RpS9 and DBE₁₆₋₂₂₉

In order to further characterize the interaction between RpS9 and DBE, the construct DBE₁₆₋₂₂₉ that carried RpS9 affinity was taken for further experiments. It was originally planned to express the 2 proteins, purify them and then possibly crystallize the RpS9/ DBE₁₆₋₂₂₉ complex. The RpS9 protein could be expressed with the aid of a GST-tag, but the expression level was too low (Figure 6.3). Alternative expression scheme of RpS9 was needed.

Two expression schemes were first attempted to improve RpS9 expression level so that the *rpS9* sequence was cloned into modified bacterial expression vector *pET3d* either with no tag or with a poly histidine and SUMO (HS) tag, but no obvious expression was observed (Figure 6.4 A, B). Then the *dbe₁₆₋₂₂₉* sequence was cloned into *pCola* expression vector that carries the kanamycin resistance gene. The *pCola* vector encoding DBE₁₆₋₂₂₉, together with *pET* vectors (ampicillin resistance) encoding RpS9 or HS-RpS9, was used for co-transformation for bacterial co-expression of the 2 proteins. However, no obvious expression of either RpS9 or HS-RpS9 was observed (Figure 6.5).

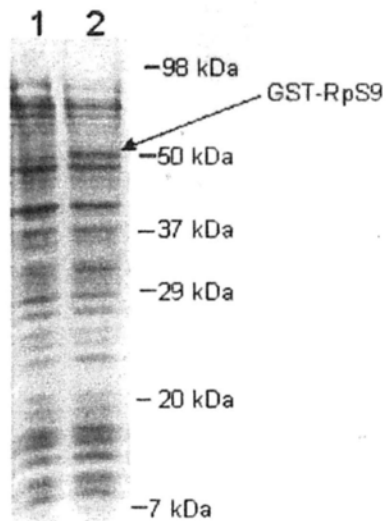


Figure 6.3 Expression of GST-RpS9. Samples taken during the course of expression were analyzed by SDS-PAGE. Lane 1: *E. coli* strain *C41* cell sampled before induction of protein expression. Lane 2: Bacterial cell sample at cell harvest. The GST-RpS9 protein is 49 kDa in size, expression of GST-RpS9 can be observed in lane 2 but the expression level was not high. The band that corresponded to GST-RpS9 protein is marked.

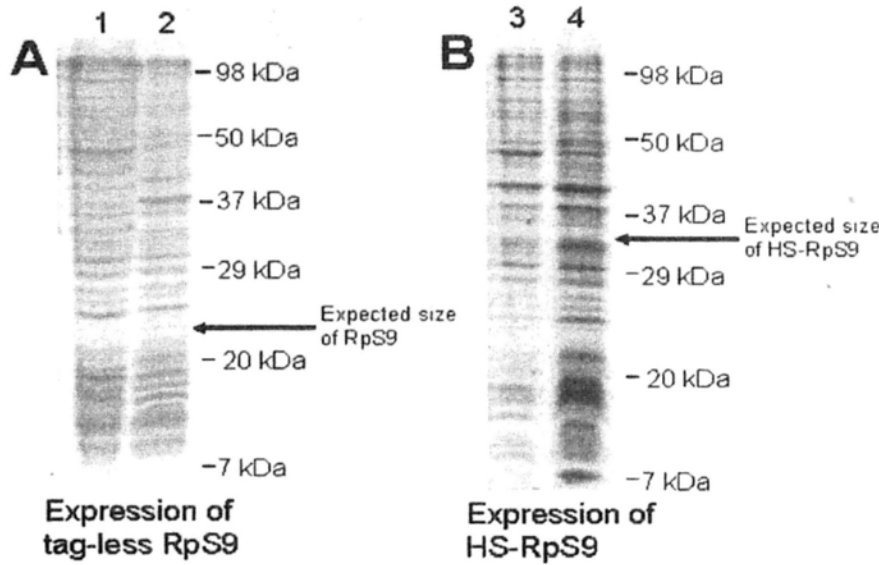


Figure 6.4 (A) Expression of tag-less RpS9. Expression of tag-less RpS9 was attempted with an aim to produce RpS9 in large quantity. Samples taken on the course of expression were analyzed by SDS-PAGE. Lane 1: *E. coli* strain C41 cell sample before induction of protein expression. Lane 2: Bacterial cell sample at cell harvest. The expected size of tag-less RpS9 is 23 kDa in size and marked. No obvious protein expression was observed, even protein expression was induced by elevated concentration of IPTG (1 mM) and the bacterial culture was grown at 37°C overnight

(B) Expression of HS tagged RpS9. The RpS9 protein was HS tagged in hope of aiding expression and purification. Lane 3: *E. coli* strain C41 cell sample before induction of protein expression. Lane 4: Bacterial cell sample at cell harvest. The expected size of HS-RpS9 is 35 kDa in size and marked. No obvious protein expression was observed, even protein expression was induced by elevated concentration of IPTG (1 mM) and the bacterial culture was grown at 37°C overnight

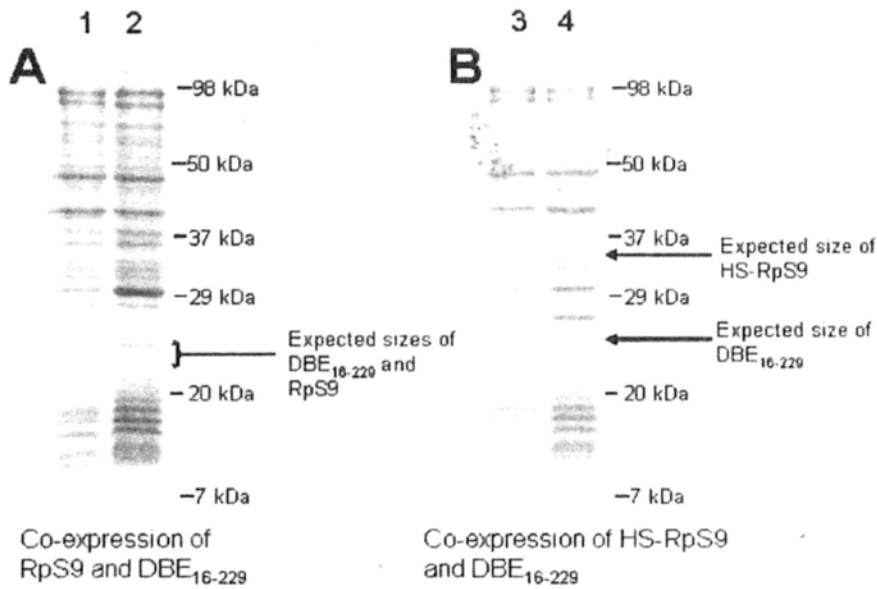


Figure 6.5 Co-expression of DBE₁₆₋₂₂₉ with RpS9/HS-RpS9. It was hoped that expression level of RpS9 or HS-RpS9 would be improved by co-expressing with its interacting partner DBE₁₆₋₂₂₉. Samples taken during the course of expression were analyzed by SDS-PAGE. The result of co-expression with DBE₁₆₋₂₂₉ and tag-less RpS9 was shown in the left panel (**A**), while that of DBE₁₆₋₂₂₉ and HS-RpS9 was shown in the right panel (**B**). Lanes 1 and 3, *E. coli* strain *C41* sample before induction of protein expression; lanes 2 and 4, bacterial cell sample at cell harvest. The expected sizes of DBE₁₆₋₂₂₉ (25 kDa), RpS9 (23 kDa) and HS-RpS9 (35 kDa) were marked. No obvious protein expression was observed in both panels, even protein expression was induced by elevated concentration of IPTG (1 mM) and the bacterial culture was grown at 37°C overnight.

While RpS9 expression seemed difficult to deal with, it was thought that structural insight of DBE₁₆₋₂₂₉ alone could hint on RpS9/ DBE interaction, and DBE₁₆₋₂₂₉ could also provide an alternative construct to search for alternative crystal form of DBE domain. Therefore, DBE₁₆₋₂₂₉ was expressed, purified and then subjected to crystallization screen as described in Section 5.3. However, no crystal was obtained.

6.4 Discussion

The GST-RpS9 pull down experiments further defined the RpS9 binding region on DBE, so that DBE deletion proteins ending at the residue 229, i.e. DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉, retained similar RpS9 affinity as DBE full length protein, but not those ending at the residue 197, i.e. DBE₁₆₋₁₉₇ and DBE₃₄₋₁₉₇ (Figure 6.2). It means that RpS9 binding is mediated through the middle region between residues 34 to 229 of DBE, and the region from residues 197 to 229 is necessary for RpS9 binding. The segment from residue 197 to 229 contains 32 residues; however, no motif could be identified in residues 197 to 229 in several databases like *PROSITE* (Sigrist et al., 2010), *ProDom* (Bru et al.,2005), *PRINTS* (Gaudet et al.,2011) and *Pfam* (Finn et al.,2008)

It is interesting to note that as both DBE and RpS9 carry a high predicted pI (9.7

and 10.6), and many positively charged amino acid residues of DBE lies in the disordered region near the C-terminus and less so around the DBE central domain region, so it was thought that DBE mediates RpS9 binding mainly via the DBE domain region due to less electrostatic repulsion. The experimental result turned out partly as expected so that RpS9 binding was largely mediated via the central region (residues 34 to 229), and the long disordered region outside of the residue 229 was not necessary for RpS9 binding, which suggests a possible contribution to RpS9 affinity by the DBE domain; however, it also turned out partly unexpected as the crystallizable DBE₁₆₋₁₉₇ protein was not sufficient for RpS9 binding. Of note, the segment from the residue 210 to 270 of DBE is conserved and disordered (Section 3.6.1) and the segment from the residue 197 to 229 is necessary for RpS9 binding, suggesting that RpS9 binding might be contributed in part by the structurally disordered region of DBE, and in part by the DBE domain, such that the crystallizable DBE₁₆₋₁₉₇ protein may not carry the complete RpS9 binding region. However, the attempt to further investigate how RpS9 and DBE₁₆₋₂₂₉ interacts and the value of DBE domain in RpS9 binding was thwarted by the fact that RpS9 did not show satisfactory expression level in various expression schemes, and DBE₁₆₋₂₂₉ did not yield protein crystal.

To further validate the DBE/RpS9 interaction in cells, co-immunoprecipitation between DBE and DBE mutants with RpS9 in fly cells can be performed. Moreover, the GST-RpS9 pull down experiment can be a starting point to investigate how DBE might function in ribosome assembly. As DBE interacts with other small subunit ribosomal proteins besides RpS9, including RpS3a, RpS8, RpS15, RpS16 and RpS23 (Choi, 2007), the technique GST pull down can be extended to these ribosomal proteins to investigate how DBE mediates the binding. By using DBE deletion mutants, the ribosomal proteins binding region in DBE can be mapped, such that the molecular functions of DBE in ribosome assembly might be revealed, for example, the functional value of the long conserved disordered region and that of the DBE domain in recruiting ribosomal proteins can be evaluated and compared.

7. General discussion

The present project adopts a structural/functional approach in studying the DBE domain.

It was planned to determine the structure of the DBE domain by X-ray crystallography, therefore DBE deletion mutants were rationally designed so as to remove the extensive structural disorders in the N- and C- termini of DBE to facilitate protein crystallization (Section 3.3). The construct DBE₁₆₋₁₉₇ was selected out of the 9 DBE deletion mutants (Section 3.4). It could be purified to high homogeneity despite an N- terminus degradation product was noted (Section 3.4). It was then found to be largely folded and compacted by biophysical characterizations like far-UV CD spectroscopy and light scattering (Section 3.5). The DBE₁₆₋₁₉₇ protein was crystallized (Section 4.2) but the crystals did not lead to structural determination. The reason of the failure to determine the DBE₁₆₋₁₉₇ crystal structure lied in the intrinsic problems in the current DBE₁₆₋₁₉₇ crystal form such that diffraction anisotropy and translational pseudo-symmetry were noted (Section 4.5.3). In addition to the 2 problems, the selenomethionyl DBE₁₆₋₁₉₇ crystal suffered from severe radiation damage that also hinder phasing attempt using selenium based MAD (Section 4.5.3). The complications in this DBE₁₆₋₁₉₇ crystal prompted the search for alternative crystal

form. The DBE₁₆₋₁₉₇ corresponding yeast homolog Krr1p₁₃₋₁₉₆ and human homolog HRB2₂₈₋₂₁₅ led to crystallization, while all other attempts failed (Section 5.2, 5.3 and 5.4). The Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅ crystals were clustered and needle shaped, however these crystals did not diffract, even after several attempts were made to improve diffraction (Section 5.5).

It was planned to understand the function of the DBE domain in terms of RpS9 affinity. Therefore, how DBE mediates RpS9 binding was investigated, it was found that protein constructs ending with the residue 229, i.e. DBE₁₋₂₂₉ and DBE₃₄₋₂₂₉, retained similar specific RpS9 affinity as the DBE full length protein, but not those ending with the residue 197, i.e. DBE₁₆₋₁₉₇, and DBE₃₄₋₁₉₇ (Section 6.2). In other words, the crystallizable DBE₁₆₋₁₉₇ protein did not show RpS9 affinity and the region between the residue 197 and 229 residues of DBE was necessary for RpS9 binding. The experimental results suggested that RpS9 affinity of DBE might be mediated partly via the DBE domain, and partly via the disordered regions. Therefore, the fact that crystallizable DBE₁₆₋₁₉₇ protein did not show RpS9 binding may be due to the lack of complete RpS9 binding region. However, further attempt to investigate the interactions between DBE and RpS9 was thwarted by the fact that RpS9 could not be expressed in large quantity (Section 6.3) and the DBE₁₆₋₂₂₉ protein did not lead to

crystallization (Section 5.3).

Although the project did not proceed as originally planned, structural understanding of the DBE domain may pave the way for future structural investigation of the early pre-rRNP. As *trans*-factors are integral structural components of the pre-rRNPs (Henras et al., 2008), such that depletion of one individual *trans*-factor, like DBE family, is enough to halt ribosome biogenesis and lethal to the organism (Sasaki et al., 2000; Chan et al., 2001; Gromadka et al., 2004). While the difficulty in structural characterization of the whole pre-rRNPs is noted (reviewed by (Dinman, 2009)), structural determination of individual *trans*-factors component, like DBE, is feasible. Combining the structures of *trans*-factors, the structural organization of the pre-rRNP machineries can be revealed by computational modeling, for example, the recently reported reconstitution of the pre-40S particle (Granneman et al., 2010). The pre-40S particle is reconstituted by computational modeling by first having structures of 6 *trans*-factors determined and their cognate rRNA positions mapped (Granneman et al., 2010). Therefore, it was planned that after structural determination of the DBE domain, the cognate rRNA binding site of DBE was to be mapped, so that the structural disorders of the full length DBE could be stabilized upon RNA binding (Yiu et al., 2006), and the DBE/ RNA complex could be

crystallized hopefully. Therefore, similarly to the case of the pre-40S particle, DBE structure and its cognate rRNA position can enable future reconstitution of the early pre-rRNP in which DBE may be involved (Section 1.2.2).

Without suitable binding partner to stabilize the structural disorders in DBE, structural determination of the DBE domain was first planned instead of the full length DBE protein. Structural understanding of the DBE domain is interesting in that it is predicted to be half-unknown and half KH-like (Section 1.2.3). The KH domain is a well studied RNA binding domain (Grishin, 2001; Valverde et al., 2008), and it is often found associating with other domains like TUDOR domain and S1 domain as archived in the *Pfam* database (Finn et al., 2008), but seldom is it found to be involved in ribosome biogenesis (Vanrobays et al., 2004). Therefore, structural determination of the DBE domain may possibly reveal a novel structural organization in which a KH-like domain is found.

Also, the crystal structure of the DBE domain can be compared with that of Dim2p (Jia et al., 2007). Although Dim2p is a *trans*-factor reported to be evolutionarily related to DBE (Vanrobays et al., 2004), their disparities are noted: the Dim2p protein carries 2 KH domains as revealed by 2 crystal structures (PDB code: 1tua and 2e3u),

but DBE is predicted to carry a single KH-like domain only; and the sequence identity of DBE is not high to the Dim2p protein of crystal structures (14.6% and 20%). Therefore, how DBE differs from Dim2p or how the 2 proteins are related can be understood by comparing the structures of DBE and Dim2p. Besides Dim2p, the DBE domain structure can be compared with other structures that are deposited in the Protein Data Bank (PDB) database by using the Dali server analysis, so as to get clues on the DBE domain function (Holm and Rosenstrom, 2010).

It was thought that DBE mediates RpS9 binding via the DBE domain region because both DBE and RpS9 carry a high predicted pI (9.7 and 10.6), and many positively charged amino acid residues of DBE lies in the disordered region near the C-terminus and less so around the central DBE domain region. The experimental result, however, turned out unexpectedly as the crystallizable DBE₁₆₋₁₉₇ protein was not sufficient for RpS9 binding, suggesting that DBE₁₆₋₁₉₇ may not carry the complete RpS9 binding region. Although the DBE domain as represented by the crystallizable DBE₁₆₋₁₉₇ protein did not show sufficient RpS9 affinity, it remains possible that the DBE domain may function to recruit other DBE interacting ribosomal proteins, like RpS3a, RpS8, RpS15, RpS16 and RpS23, as determined previously (Choi, 2007).

Besides interacting with ribosomal proteins, the DBE domain may carry RNA binding property via its KH-like domain. And identification of cognate rRNA binding site may also help crystallization of the full length DBE. Therefore, it is suggested that mapping the rRNA binding site of DBE is to be attempted, so that, firstly, the cognate rRNA site where DBE binds and functions can be identified; secondly, the RNA affinity of DBE domain and full length DBE can be assessed and compared; thirdly, as the extensive disorders of DBE can be stabilized upon nucleic acid analog binding (Yiu et al., 2006), the RNA may stabilize the structural disorders of DBE that may lead to co-crystallization of the DBE/ RNA complex, which may facilitate future structural understanding of the pre-rRNP.

8. Concluding Remarks

The project aimed to structurally characterize the structured region of DBE by X-ray crystallography, and to characterize the molecular function of the structured region of DBE in terms of RpS9 binding.

For the structural determination part, 9 DBE deletion mutants were constructed by rational design to eliminate structural disorders in both N- and C- termini that may hinder crystallization. Out of the 9 constructs, DBE₁₆₋₁₉₇ could be expressed and purified to high homogeneity despite an N- terminus degradation product was identified. The DBE₁₆₋₁₉₇ protein could be crystallized. The diffraction data of the DBE₁₆₋₁₉₇ crystal could be integrated and scaled to 1.98 Å. Molecular replacement, however, failed to phase the structure. Various experimental phasing techniques were attempted. Soaking DBE₁₆₋₁₉₇ crystal in a platinum derivative potassium tetrachloroplatinate (II), and quick soak in iodide containing cryo-protectant could result in heavy atom substituted crystals but they did not give strong enough signal for phasing. Selenomethionyl MAD could not lead to phasing of the structure, as 3 problems were detected in the datasets of, i.e. severe diffraction anisotropy, bad radiation damage and translational pseudo-symmetry. To bypass the current crystal form of DBE₁₆₋₁₉₇, alternative crystal form was actively sought after by modifying the

chemical environment in which DBE₁₆₋₁₉₇ crystallized; and modifying the DBE₁₆₋₁₉₇ constructs for crystallization screen, but the efforts did not lead to protein crystallization. On the contrary, the DBE₁₆₋₁₉₇ corresponding yeast and human homologs, Krr1p₁₃₋₁₉₆ and HRB2₂₈₋₂₁₅, could yield clustered and needle shaped crystals. However, these crystals did not show diffraction even after efforts to improve crystal quality like seeding, additives screening, crystal desiccation and crystal cross-linking. In short, the DBE domain as represented by the *Drosophila* DBE₁₆₋₁₉₇, yeast Krr1p₁₃₋₁₉₆ and human HRB2₂₈₋₂₁₅ could all be crystallized, but they did not lead to structure determination.

For the functional characterization part, it could be identified that the region spanning from the residue 34 to 229 of DBE, where the potential DBE domain lies, was sufficient to mediate RpS9 binding, while the residues 197 to 229 was necessary for RpS9 binding. In other words, the crystallizable construct DBE₁₆₋₁₉₇ did not show RpS9 affinity. It is hence suggested that the RpS9 affinity of DBE may be mediated partly by the DBE domain, and partly by the disordered region of DBE.

9. References

- Abergel, C. (2004) Spectacular improvement of X-ray diffraction through fast desiccation of protein crystals. **Acta Crystallogr D Biol Crystallogr** 60(Pt 8): 1413-6.
- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. **Acta Crystallogr D Biol Crystallogr** 66(Pt 2): 213-21.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2008). Molecular biology of the cell, fifth edition. New York: Garland **978-0815341055**
- Alter, B.P., Giri, N., Savage, S.A. and Rosenberg, P.S. (2009) Cancer in dyskeratosis congenita. **Blood** 113(26): 6549-57.
- Amsterdam, A., Sadler, K.C., Lai, K., Farrington, S., Bronson, R.T., Lees, J.A. and Hopkins, N. (2004) Many ribosomal protein genes are cancer genes in zebrafish. **PLoS Biol** 2(5): E139.
- Andersen, J.S., Lam, Y.W., Leung, A.K., Ong, S.E., Lyon, C.E., Lamond, A.I. and Mann, M. (2005) Nucleolar proteome dynamics. **Nature** 433(7021): 77-83.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. **Science** 289(5481): 905-20.
- Bassler, J., Grandi, P., Gadal, O., Lessmann, T., Petfalski, E., Tollervey, D., Lechner, J. and Hurt, E. (2001) Identification of a 60S preribosomal particle that is closely linked to nuclear export. **Mol Cell** 8(3): 517-29.
- Beck, T., da Cunha, C.E. and Sheldrick, G.M. (2009) How to get the magic triangle and the MAD triangle into your protein crystal. **Acta Crystallogr Sect F Struct Biol Cryst Commun** 65(Pt 10): 1068-70.
- Beck, T., Krasauskas, A., Gruene, T. and Sheldrick, G.M. (2008) A magic triangle for experimental phasing of macromolecules. **Acta Crystallogr D Biol Crystallogr** 64(Pt 11): 1179-82.
- Belin, S., Beghin, A., Solano-Gonzalez, E., Bezin, L., Brunet-Manquat, S., et al. (2009) Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. **PLoS One** 4(9): e7147.
- Beltrame, M. and Tollervey, D. (1995) Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. **EMBO J** 14(17): 4350-6.
- Bernstein, K.A., Gallagher, J.E., Mitchell, B.M., Granneman, S. and Baserga, S.J.

- (2004) The small-subunit processome is a ribosome assembly intermediate. **Eukaryot Cell** 3(6): 1619-26.
- Blanchard, S.C., Gonzalez, R.L., Kim, H.D., Chu, S. and Puglisi, J.D. (2004) tRNA selection and kinetic proofreading in translation. **Nat Struct Mol Biol** 11(10): 1008-14.
- Blow, D. (2002). Outline of Crystallography for Biologists. Oxford; New York: Oxford University Press **978-0198510512**
- Burger, K., Muhl, B., Harasim, T., Rohmoser, M., Malamoussi, A., et al. (2010) Chemotherapeutic drugs inhibit ribosome biogenesis at various levels. **J Biol Chem** 285(16): 12416-25.
- Chan, H.Y., Brogna, S. and O'Kane, C.J. (2001) Dribble, the *Drosophila* KRR1p homologue, is involved in rRNA processing. **Mol Biol Cell** 12(5): 1409-19.
- Choi, C.G. (2007) Identification of Protein Interactors of Dribble — a Single KH-Domain Nucleolar Protein in *Drosophila*. Unpublished M. Phil Thesis, The Chinese University of Hong Kong
- Comartin, D.J. and Brown, E.D. (2006) Non-ribosomal factors in ribosome subunit assembly are emerging targets for new antibacterial drugs. **Curr Opin Pharmacol** 6(5): 453-8.
- Connell, S.R., Takemoto, C., Wilson, D.N., Wang, H., Murayama, K., et al. (2007) Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. **Mol Cell** 25(5): 751-64.
- Czepas, J., Devedjiev, Y., Krowarsch, D., Derewenda, U., Otlewski, J. and Derewenda, Z.S. (2004) The impact of Lys-->Arg surface mutations on the crystallization of the globular domain of RhoGDI. **Acta Crystallogr D Biol Crystallogr** 60(Pt 2): 275-80.
- Dauter, Z., Botos, I., LaRonde-LeBlanc, N. and Wlodawer, A. (2005) Pathological crystallography: case studies of several unusual macromolecular crystals. **Acta Crystallogr D Biol Crystallogr** 61(Pt 7): 967-75.
- Dauter, Z., Dauter, M. and Dodson, E. (2002) Jolly SAD. **Acta Crystallogr D Biol Crystallogr** 58(Pt 3): 494-506.
- Dauter, Z., Dauter, M. and Rajashankar, K.R. (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. **Acta Crystallogr D Biol Crystallogr** 56(Pt 2): 232-7.
- Decatur, W.A. and Fournier, M.J. (2002) rRNA modifications and ribosome function. **Trends Biochem Sci** 27(7): 344-51.
- Decatur, W.A. and Fournier, M.J. (2003) RNA-guided nucleotide modification of ribosomal and other RNAs. **J Biol Chem** 278(2): 695-8.
- Decatur, W.A., Liang, X.H., Piekna-Przybylska, D. and Fournier, M.J. (2007)

- Identifying effects of snoRNA-guided modifications on the synthesis and function of the yeast ribosome. **Methods Enzymol** 425: 283-316.
- Deisenroth, C. and Zhang, Y. (2010) Ribosome biogenesis surveillance: probing the ribosomal protein-Mdm2-p53 pathway. **Oncogene**.
- Derewenda, Z.S. (2010) Application of protein engineering to enhance crystallizability and improve crystal properties. **Acta Crystallogr D Biol Crystallogr** 66(Pt 5): 604-15.
- Dinman, J.D. (2009) The eukaryotic ribosome: current status and challenges. **J Biol Chem** 284(18): 11761-5.
- Dodson, E. (2003) Is it jolly SAD? **Acta Crystallogr D Biol Crystallogr** 59(Pt 11): 1958-65.
- Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., et al. (2007) In situ proteolysis for protein crystallization and structure determination. **Nat Methods** 4(12): 1019-21.
- Dragon, F., Gallagher, J.E., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., et al. (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. **Nature** 417(6892): 967-70.
- Evans, P. (2006) Scaling and assessment of data quality. **Acta Crystallogr D Biol Crystallogr** 62(Pt 1): 72-82.
- Ferreira-Cerca, S., Poll, G., Gleizes, P.E., Tschochner, H. and Milkereit, P. (2005) Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. **Mol Cell** 20(2): 263-75.
- Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., et al. (2008) The Pfam protein families database. **Nucleic Acids Res** 36(Database issue): D281-8.
- Fontana, A., de Laureto, P.P., Spolaore, B., Frare, E., Picotti, P. and Zamboni, M. (2004) Probing protein structure by limited proteolysis. **Acta Biochim Pol** 51(2): 299-321.
- Fromont-Racine, M., Senger, B., Saveanu, C. and Fasiolo, F. (2003) Ribosome assembly in eukaryotes. **Gene** 313: 17-42.
- Gallagher, J.E., Dunbar, D.A., Granneman, S., Mitchell, B.M., Osheim, Y., Beyer, A.L. and Baserga, S.J. (2004) RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. **Genes Dev** 18(20): 2506-17.
- Garrard, S.M., Longenecker, K.L., Lewis, M.E., Sheffield, P.J. and Derewenda, Z.S. (2001) Expression, purification, and crystallization of the RGS-like domain from the Rho nucleotide exchange factor, PDZ-RhoGEF, using the surface entropy reduction approach. **Protein Expr Purif** 21(3): 412-6.
- Gilbert, R.J., Gordiyenko, Y., von der Haar, T., Sonnen, A.F., Hofmann, G., Nardelli,

- M., Stuart, D.I. and McCarthy, J.E. (2007) Reconfiguration of yeast 40S ribosomal subunit domains by the translation initiation multifactor complex. **Proc Natl Acad Sci U S A** 104(14): 5788-93.
- Goldschmidt, L., Cooper, D.R., Derewenda, Z.S. and Eisenberg, D. (2007) Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. **Protein Sci** 16(8): 1569-76.
- Grandi, P., Rybin, V., Bassler, J., Petfalski, E., Strauss, D., et al. (2002) 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. **Mol Cell** 10(1): 105-15.
- Granneman, S. and Baserga, S.J. (2004) Ribosome biogenesis: of knobs and RNA processing. **Exp Cell Res** 296(1): 43-50.
- Granneman, S., Petfalski, E., Swiatkowska, A. and Tollervey, D. (2010) Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. **EMBO J**.
- Graslund, S., Sagemark, J., Berglund, H., Dahlgren, L.G., Flores, A., et al. (2008) The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. **Protein Expr Purif** 58(2): 210-21.
- Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K. and Pandolfi, P.P. (2005) Role of nucleophosmin in embryonic development and tumorigenesis. **Nature** 437(7055): 147-53.
- Grishin, N.V. (2001) KH domain: one motif, two folds. **Nucleic Acids Res** 29(3): 638-43.
- Gromadka, R., Karkusiewicz, I., Rempola, B. and Rytka, J. (2004) Functional and physical interactions of Krr1p, a *Saccharomyces cerevisiae* nucleolar protein. **Acta Biochim Pol** 51(1): 173-87.
- Hage, A.E. and Tollervey, D. (2004) A surfeit of factors: why is ribosome assembly so much more complicated in eukaryotes than bacteria? **RNA Biol** 1(1): 10-5.
- Harnpicharnchai, P., Jakovljevic, J., Horsey, E., Miles, T., Roman, J., et al. (2001) Composition and functional characterization of yeast 66S ribosome assembly intermediates. **Mol Cell** 8(3): 505-15.
- Henras, A.K., Soudet, J., Gerus, M., Lebaron, S., Caizergues-Ferrer, M., Mouglin, A. and Henry, Y. (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. **Cell Mol Life Sci** 65(15): 2334-59.
- Heras, B. and Martin, J.L. (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. **Acta Crystallogr D Biol Crystallogr** 61(Pt 9): 1173-80.
- Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D.

Nucleic Acids Res 38 Suppl: W545-9.

- Holton, J.M. (2007) XANES measurements of the rate of radiation damage to selenomethionine side chains. **J Synchrotron Radiat** 14(Pt 1): 51-72.
- Hughes, J.M. and Ares, M., Jr. (1991) Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. **EMBO J** 10(13): 4231-9.
- Jia, M.Z., Horita, S., Nagata, K. and Tanokura, M. (2010) An archaeal Dim2-like protein, aDim2p, forms a ternary complex with a/eIF2 alpha and the 3' end fragment of 16S rRNA. **J Mol Biol** 398(5): 774-85.
- Jia, M.Z., Ohtsuka, J., Lee, W.C., Nagata, K. and Tanokura, M. (2007) Crystal structure of Dim2p: a preribosomal RNA processing factor, from *Pyrococcus horikoshii* OT3 at 2.30 Å. **Proteins** 69(2): 428-32.
- Knipscheer, P., van Dijk, W.J., Olsen, J.V., Mann, M. and Sixma, T.K. (2007) Noncovalent interaction between Ubc9 and SUMO promotes SUMO chain formation. **EMBO J** 26(11): 2797-807.
- Kressler, D., Linder, P. and de La Cruz, J. (1999) Protein trans-acting factors involved in ribosome biogenesis in *Saccharomyces cerevisiae*. **Mol Cell Biol** 19(12): 7897-912.
- Leslie, A.G. (2006) The integration of macromolecular diffraction data. **Acta Crystallogr D Biol Crystallogr** 62(Pt 1): 48-57.
- Longenecker, K.L., Garrard, S.M., Sheffield, P.J. and Derewenda, Z.S. (2001) Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. **Acta Crystallogr D Biol Crystallogr** 57(Pt 5): 679-88.
- MacInnes, A.W., Amsterdam, A., Whittaker, C.A., Hopkins, N. and Lees, J.A. (2008) Loss of p53 synthesis in zebrafish tumors with ribosomal protein gene mutations. **Proc Natl Acad Sci U S A** 105(30): 10408-13.
- Marz, M. and Stadler, P.F. (2009) Comparative analysis of eukaryotic U3 snoRNA. **RNA Biol** 6(5): 503-7.
- Matthews, B.W. (1968) Solvent content of protein crystals. **J Mol Biol** 33(2): 491-7.
- Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C.L., 3rd, Ban, N. and Frank, J. (2005) Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. **Nature** 438(7066): 318-24.
- Munro, J.B., Altman, R.B., O'Connor, N. and Blanchard, S.C. (2007) Identification of two distinct hybrid state intermediates on the ribosome. **Mol Cell** 25(4): 505-17.
- Murray, J. and Garman, E. (2002) Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography. **J**

Synchrotron Radiat 9(Pt 6): 347-54.

- Narla, A. and Ebert, B.L. (2010) Ribosomopathies: human disorders of ribosome dysfunction. **Blood** 115(16): 3196-205.
- Nelson, D.L. and Cox, M.M. (2005). *Lehninger Principles of Biochemistry*, Fourth Edition. New York: Freeman, W. H. **978-0716743392**
- Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. and Chin, J.W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. **Nature** 464(7287): 441-4.
- Nissen, P., Hansen, J., Ban, N., Moore, P.B. and Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. **Science** 289(5481): 920-30.
- Osheim, Y.N., French, S.L., Keck, K.M., Champion, E.A., Spasov, K., Dragon, F., Baserga, S.J. and Beyer, A.L. (2004) Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. **Mol Cell** 16(6): 943-54.
- Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R. and Ramakrishnan, V. (2007) The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. **Mol Cell** 26(1): 41-50.
- Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K. and Obradovic, Z. (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. **J Bioinform Comput Biol** 3(1): 35-60.
- Piekna-Przybylska, D., Decatur, W.A. and Fournier, M.J. (2008) The 3D rRNA modification maps database: with interactive tools for ribosome analysis. **Nucleic Acids Res** 36(Database issue): D178-83.
- Potterton, E., Briggs, P., Turkenburg, M. and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. **Acta Crystallogr D Biol Crystallogr** 59(Pt 7): 1131-7.
- Rackham, O. and Chin, J.W. (2005) A network of orthogonal ribosome x mRNA pairs. **Nat Chem Biol** 1(3): 159-66.
- Raska, I., Shaw, P.J. and Cmarko, D. (2006) Structure and function of the nucleolus in the spotlight. **Curr Opin Cell Biol** 18(3): 325-34.
- Rhodes, G. (2006). *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*, third edition. Amsterdam; Boston: Academic Press **978-0125870733**
- Robledo, S., Idol, R.A., Crimmins, D.L., Ladenson, J.H., Mason, P.J. and Bessler, M. (2008) The role of human ribosomal proteins in the maturation of rRNA and ribosome production. **RNA** 14(9): 1918-29.

- Salem, M., Mauguen, Y. and Prange, T. (2010) Revisiting glutaraldehyde cross-linking: the case of the Arg-Lys intermolecular doublet. **Acta Crystallogr Sect F Struct Biol Cryst Commun** 66(Pt 3): 225-8.
- Sasaki, T., Toh, E.A. and Kikuchi, Y. (2000) Yeast Krr1p physically and functionally interacts with a novel essential Kri1p, and both proteins are required for 40S ribosome biogenesis in the nucleolus. **Mol Cell Biol** 20(21): 7971-9.
- Schafer, T., Strauss, D., Petfalski, E., Tollervey, D. and Hurt, E. (2003) The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. **EMBO J** 22(6): 1370-80.
- Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., et al. (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. **Cell** 102(5): 615-23.
- Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M. and Cate, J.H. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. **Science** 310(5749): 827-34.
- Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. **Science** 313(5795): 1935-42.
- Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D. and Eisenberg, D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. **Proc Natl Acad Sci U S A** 103(21): 8060-5.
- Taskinen, M., Ranki, A., Pukkala, E., Jeskanen, L., Kaitila, I. and Makitie, O. (2008) Extended follow-up of the Finnish cartilage-hair hypoplasia cohort confirms high incidence of non-Hodgkin lymphoma and basal cell carcinoma. **Am J Med Genet A** 146A(18): 2370-5.
- Taylor, D.J., Devkota, B., Huang, A.D., Topf, M., Narayanan, E., Sali, A., Harvey, S.C. and Frank, J. (2009) Comprehensive molecular structure of the eukaryotic ribosome. **Structure** 17(12): 1591-604.
- Tschochner, H. and Hurt, E. (2003) Pre-ribosomes on the road from the nucleolus to the cytoplasm. **Trends Cell Biol** 13(5): 255-63.
- Uemura, S., Aitken, C.E., Korlach, J., Flusberg, B.A., Turner, S.W. and Puglisi, J.D. (2010) Real-time tRNA transit on single translating ribosomes at codon resolution. **Nature** 464(7291): 1012-7.
- Uversky, V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. **Protein Sci** 11(4): 739-56.
- Valverde, R., Edwards, L. and Regan, L. (2008) Structure and function of KH domains. **FEBS J** 275(11): 2712-26.
- van Riggelen, J., Yetil, A. and Felsher, D.W. (2010) MYC as a regulator of ribosome

- biogenesis and protein synthesis. **Nat Rev Cancer** 10(4): 301-9.
- Vanrobays, E., Gelugne, J.P., Caizergues-Ferrer, M. and Lafontaine, D.L. (2004) Dim2p, a KH-domain protein required for small ribosomal subunit synthesis. **RNA** 10(4): 645-56.
- Vanrobays, E., Leplus, A., Osheim, Y.N., Beyrer, A.L., Wacheul, L. and Lafontaine, D.L. (2008) TOR regulates the subcellular distribution of DIM2, a KH domain protein required for cotranscriptional ribosome assembly and pre-40S ribosome export. **RNA** 14(10): 2061-73.
- Wang, K., Neumann, H., Peak-Chew, S.Y. and Chin, J.W. (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. **Nat Biotechnol** 25(7): 770-7.
- Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. **J Mol Biol** 337(3): 635-45.
- Wernimont, A. and Edwards, A. (2009) In situ proteolysis to generate crystals for structure determination: an update. **PLoS One** 4(4): e5094.
- Whitesides, G.M. (2005) Nanoscience, nanotechnology, and chemistry. **Small** 1(2): 172-9.
- Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C. and Gough, J. (2009) SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. **Nucleic Acids Res** 37(Database issue): D380-6.
- Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vornrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. **Nature** 407(6802): 327-39.
- Xin, D.D., Wen, J.F., He, D. and Lu, S.Q. (2005) Identification of a Giardia krr1 homolog gene and the secondarily anucleolate condition of Giardia lamblia. **Mol Biol Evol** 22(3): 391-4.
- Xu, Z., Chau, S.F., Lam, K.H., Chan, H.Y., Ng, T.B. and Au, S.W. (2006) Crystal structure of the SENP1 mutant C603S-SUMO complex reveals the hydrolytic mechanism of SUMO-specific protease. **Biochem J** 398(3): 345-52.
- Yiu, C.P., Beavil, R.L. and Chan, H.Y. (2006) Biophysical characterisation reveals structural disorder in the nucleolar protein, Dribble. **Biochem Biophys Res Commun** 343(1): 311-8.
- Zwart, P.H., Grosse-Kunstleve, R.W., Lebedev, A.A., Murshudov, G.N. and Adams, P.D. (2008) Surprises and pitfalls arising from (pseudo)symmetry. **Acta Crystallogr D Biol Crystallogr** 64(Pt 1): 99-107.