Structural and mechanistic studies of two enzymes : SARS-associated coronavirus helicase and porcine pancreatic elastase

Liu, Bin

2007

Liu, B. (2007). Structural and mechanistic studies of two enzymes : SARS-associated coronavirus helicase and porcine pancreatic elastase. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/6567>

<https://doi.org/10.32657/10356/6567>

Nanyang Technological University

Downloaded on 26 Apr 2023 17:14:06 SGT

Structural and mechanistic studies of two enzymes: SARS-associated coronavirus helicase and porcine pancreatic elastase

Bin Liu

Ph.D. Thesis

**School of Biological Sciences,
Nanyang Technological University**

2007

To my parents and my wife

CONTENTS

ABSTRACT.....	viii
ACKNOWLEDGEMENTS	ix
ABBREVIATIONS.....	x
CHAPTER 1. INTRODUCTION (I): HELICASES AND SARS-CoV HELICASE.....	1
1.1 HELICASES AND THEIR GENERAL BIOCHEMICAL PROPERTIES	1
1.2 CLASSIFICATION OF HELICASES	1
1.2.1 Substrate directionality.....	1
1.2.2 Sequence level.....	2
1.3 HELICASE MOTIFS	4
1.3.1 SF1 and SF2.....	4
1.3.2 SF3, F4 and F5	6
1.3.3 New conserved motifs	7
1.3.4 Mutagenesis studies on the conserved motifs.....	7
1.4 SARS-CoV HELICASE	9
1.4.1 SARS	9
1.4.2 SARS-associated coronavirus	9
1.4.3 SARS-CoV helicase.....	10
1.5 MECHANISMS OF HELICASE UNWINDING.....	13
1.5.1 Passive mechanism	13
1.5.2 Active mechanism	14
1.5.3 Hexameric helicases	16
1.6 STRUCTURAL STUDIES OF HELICASES.....	18
1.7 HELICASES AS ANTIVIRAL TARGETS.....	20
1.8 PURPOSES OF THIS STUDY	20
CHAPTER 2. CLONING, EXPRESSION, AND PURIFICATION OF SARS-CoV HELICASE	21
2.1 CLONING OF HELICASES	21
2.2 EXPRESSION OF HELICASES	23
2.2.1 6×His tag and no tag.....	23
2.2.2 Glutathione-S-transferase (GST) tag.....	26
2.2.3 Maltose binding protein (MBP) tag.....	28
2.2.4 Improvement of expression	28
2.3 PURIFICATION OF PROTEINS	29
2.3.1 6xHis affinity purification.....	29
2.3.2 GST purification	32
2.3.3 MBP purification	34

CHAPTER 3. FUNCTIONAL STUDIES ON SARS-CoV HELICASE.....	39
3.1 MUTAGENESIS OF SARS-CoV HELICASE	39
3.2 ATPase ACTIVITY	41
3.3 UNWINDING ACTIVITY	48
3.3.1 DNA substrates	48
3.3.2 RNA substrates.....	55
3.4 DISCUSSION	58
3.4.1 Site-directed mutagenesis	58
3.4.2 Effect of polyribonucleotides	59
3.4.3 Role of the zinc finger domain	59
3.4.4 Roles of divalent cations	61
3.4.5 DNA and RNA unwinding assays.....	61
CHAPTER 4. CRYSTALLIZATION AND PRELIMINARY X-RAY ANALYSIS.....	63
4.1 CRYSTALLIZATION OF HELICASES	63
4.1.1 Crystallization of helicases	63
4.1.2 Seeding.....	65
4.1.3 The use of fusion proteins in crystallization	67
4.2 HEAVY ATOM DERIVATIZATION	67
4.2.1 Phase problem	67
4.2.2 Heavy atom soaking.....	68
4.2.3 Selenomethionine-substituted protein	69
4.3 DATA COLLECTION AND PROCESSING	71
4.3.1 Data collected from the in-house X-ray diffractometer.....	71
4.3.2 Synchrotron data.....	74
4.4 PHASE DETERMINATION	76
4.4.1 MIR	76
4.4.2 SAD and MAD	77
4.5 DISCUSSION	78
CHAPTER 5. INTRODUCTION (II): SERINE PROTEASES AND PORCINE PANCREATIC ELASTASE (PPE)	79
5.1 SERINE PROTEASES	79
5.1.1 Chymotrypsin family	79
5.1.2 Structure of the chymotrypsin family of serine proteases	80
5.1.3 The catalytic triad.....	81
5.1.4 The oxyanion hole.....	81
5.1.5 The zymogen activation domain.....	82
5.1.6 The substrate recognition sites	82
5.2 CATALYTIC MECHANISM OF SERINE PROTEASES	86
5.2.1 The basic mechanism and intermediates.....	86
5.2.2 Catalytic triad and charge-relay system.....	87
5.2.3 The role of the oxyanion hole.....	88

5.2.4 Mutagenesis of the catalytic triad and oxyanion hole.....	90
5.2.5 Evidence for tetrahedral intermediates.....	91
5.3 ELASTASE AND PORCINE PANCREATIC ELASTASE (PPE).....	92
5.3.1 Elastase.....	92
5.3.2 The active sites and substrate specificity of elastase	92
5.4 PURPOSE OF THIS STUDY	94
CHAPTER 6. MASS SPECTROMETRIC, KINETIC, AND CRYSTALLOGRAPHIC ANALYSIS OF THE INTERACTION OF SMALL PEPTIDES WITH PPE.....	95
6.1 PEPTIDE SYNTHESIS	95
6.2 HPLC PURIFICATION AND ANALYSIS	96
6.2.1 HPLC purification	96
6.2.2 HPLC analysis	97
6.3 ELECTROSPRAY IONISATION MASS SPECTROMETRY	97
6.4 KINETICS	101
6.4.1 PPE activities at different pH values	101
6.4.2 IC ₅₀ of the peptides.....	101
6.5 CO-CRYSTALLIZATION WITH DIFFERENT PEPTIDES.....	104
CHAPTER 7. MECHANISTIC CRYSTALLOGRAPHY: REACTION OF PEPTIDES WITH THE PPE:BCM7 COMPLEX.....	108
7.1 FIRST TETRAHEDRAL INTERMEDIATE	109
7.2 FORMATION OF AN APPARENT ENZYME:SUBSTRATE COMPLEX	112
7.3 pH JUMP EXPERIMENTS ON THE ASSIGNED TETRAHEDRAL STRUCTURES	115
7.4 BCM7 ANALOGUES	119
7.5 α_1 -ANTITRYPSIN-DERIVED PEPTIDE:MFLE	121
7.6 DISCUSSION	122
7.6.1 Comparison with previously obtained tetrahedral intermediates.....	122
7.6.2 Conformational changes during catalysis	124
7.6.3 Geometry of the scissile amide bond.....	125
7.6.4 Stereoelectronic implications.....	126
7.7 CONCLUSION	129
CHAPTER 8. MATERIALS AND METHODS.....	130
8.1 MATERIALS (I).....	130
8.1.1 General chemicals and reagents.....	130
8.1.2 Molecular cloning and expression.....	130
8.1.3 Protein purification.....	130
8.1.4 Functional and mutagenesis studies	131
8.1.5 Crystallization, data-collection and processing.....	131
8.1.6 Equipment	131
8.2 MOLECULAR CLONING AND PROTEIN EXPRESSION	133
8.2.1 pET24d.....	133
8.2.2 pETM-30.....	133

8.2.3 <i>pET24d + pETM-30</i>	134
8.2.4 <i>pMAL-C₂X</i>	134
8.2.5 <i>Selenomethionine-substituted proteins</i>	134
8.3 PROTEIN PURIFICATION.....	135
8.3.1 <i>Purification of protein with 6×His tag</i>	135
8.3.2 <i>Purification of protein without tag</i>	135
8.3.3 <i>Purification of insoluble protein</i>	136
8.3.4 <i>Purification of protein with GST tag</i>	136
8.3.5 <i>Purification of protein with MBP tag</i>	137
8.4 PROTEIN IDENTIFICATION AND ANALYSIS	137
8.4.1 <i>Western blotting for nsp13 with a 6×His tag</i>	137
8.4.2 <i>Circular dichroism spectrometry</i>	138
8.4.3 <i>Mass spectrometry</i>	138
8.4.4 <i>Dynamic laser light scattering</i>	139
8.4.5 <i>N-terminal protein sequencing</i>	139
8.5 ATPase ASSAY.....	139
8.5.1 <i>Basic ATPase assay</i>	139
8.5.2 <i>Effects of different co-factors on ATPase activity</i>	140
8.6 DNA UNWINDING ASSAY	140
8.6.1 <i>[γ-³²P] labelling</i>	140
8.6.2 <i>Ethidium bromide and silver staining</i>	140
8.7 RNA UNWINDING ASSAY	141
8.8 MUTAGENESIS STUDIES	142
8.9 CRYSTALLIZATION.....	142
8.9.1 <i>Crystallization by vapour diffusion</i>	142
8.9.2 <i>Crystallization screening</i>	143
8.9.3 <i>Seeding to improve crystal quality</i>	143
8.10 DATA COLLECTION	143
8.10.1 <i>In-house data collection</i>	143
8.10.2 <i>Synchrotron data collection</i>	144
8.11 DATA PROCESSING.....	144
8.12 PHASE DETERMINATION	145
8.12.1 <i>Multiple Isomorphous Replacement (MIR)</i>	145
8.12.2 <i>Single-wavelength Anomalous Dispersion (SAD) and Multiple-wavelength Anomalous Dispersion (MAD)</i>	146
8.13 MATERIALS (II)	146
8.13.1 <i>Chemicals</i>	146
8.13.2 <i>Equipment</i>	146
8.14 PEPTIDE SYNTHESIS	147
8.14.1 <i>Peptide with C-terminal amide</i>	147
8.14.2 <i>Peptide with C-terminal carboxylate</i>	147
8.15 HPLC	148
8.15.1 <i>HPLC purification</i>	148

8.15.2 HPLC analysis	148
8.16 ESI-MS IDENTIFICATION AND ANALYSIS	148
8.17 KINETICS	149
8.18 CRYSTALLIZATION.....	149
8.18.1 Crystallization of native PPE.....	149
8.18.2 Co-crystallization of PPE with peptides	150
8.19 SOAKING EXPERIMENTS	150
8.20 pH-JUMP EXPERIMENTS	150
8.21 DATA COLLECTION	151
8.22 DETERMINATION OF CRYSTAL STRUCTURES	151
8.23 COMPARISON OF DIFFERENT STRUCTURES	152
APPENDICES.....	153
REFERENCES	157
PUBLICATIONS.....	169

Abstract

Helicases are enzymes that separate nucleic acid duplexes and play important roles in both DNA and RNA metabolism. Usually helicases have several biochemical properties including ATPase activity and nucleotide duplex unwinding activity. SARS-coronavirus (SARS-CoV) is a positive-sense single-stranded RNA virus belonging to the order Nidovirales. During replication of the virus, the helicase plays a critical role. Therefore inhibitors of the helicase, based on structural and mechanistic studies, are potentially attractive antiviral therapeutic agents.

In this thesis, the SARS-CoV helicase protein was cloned, expressed, and purified. Functional studies including ATPase assays and DNA/RNA substrate unwinding assays were carried out. In addition, different constructs of the SARS-CoV helicase were used for crystallization trials. Crystals of the full-length helicase with a C-terminal 6×His tag were obtained which diffracted to 2.8 Å. MIR (Multiple Isomorphous Replacement) and SAD (Single Wavelength Anomalous Dispersion) data were collected. However, the quality of the phase information was inadequate to solve the structure.

The serine proteases possess one of the best characterised catalytic mechanisms of any enzyme family. Despite several decades of intensive study, the fine detail of the mechanism is still being debated. Structural information on the intermediates in the hydrolysis of natural productive peptide substrates is still lacking. Crystallographic analyses were performed in which an acyl-enzyme complex, formed from elastase and a heptapeptide, was reacted with a series of nucleophilic dipeptides. On certain crystals, a 'pH jump' methodology was employed to trap reaction intermediates. These analyses led to electron density maps consistent with the formation of a tetrahedral species. In certain cases, apparent peptide bond formation at the active site was observed and the electron density maps suggested production of a *cis*-amide, rather than a *trans*-amide.

Acknowledgements

My deepest gratitude must go to Dr. Rupert Charles Wilmouth, my supervisor, for all his support, guidance, endurance and understanding during the development of this work, and his encouragement under difficult circumstances. I also can not forget his precise and serious-minded attitude to science. All these would have a direct impact on my personal and professional life. Thank you for teaching me how to do science and how to deal with the politics of science.

I would also like to acknowledge the chemist, Prof. C.J. Schofield at the University of Oxford, for his kind guidance and suggestions in the serine protease part of my thesis. His keen vision and profound scientific knowledge impresses me deeply.

Next I would like to thank Dr. Liu Chuan Fa for allowing me to carry out my thesis work in his lab and for his kind advice in peptide synthesis. Meanwhile, sincere thanks too to X-ray crystallography staffs in SBS and SLS (Switzerland) whose expertise have ensured that my data collection went as smoothly as possible.

Many thanks also to my friends, Mr. Ji Zhe, Mr. Tan Xiaohong, Ms. Feng Lin, who gave me sincere help and care in my experiments. I would like to thank you for your friendship. And I would not forget your invaluable and honest help during many difficult moments.

I also should like to express my appreciation to all the people who have helped me during the past several years. First those past and present staffs in Dr. Rupert C. Wilmouth's laboratory, Westley, SzeYung, *etc.*, for placing orders and help. Secondly, my classmates in SBS, NTU: Mr. Fan Hui, *etc.*, for help and discussion in many experiments. Then, undergraduate students in the lab: Ms. Sem Xiaohui, *etc.*, for their work. Finally, those in SBS whose help I did not realize.

I am proud to dedicate this dissertation to my parents and my wife for the incredible quantity of their love. Their support was critical to this thesis work. They are indispensable in my life. Their love is the motivation for my progress in science.

Abbreviations

Ac	acetyl
ADP	adenosine diphosphate
ADPNP	5'-adenylyl- β , γ -imidodiphosphate
Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartic acid
ATP	adenosine triphosphate
BCM7	β -casomorphin-7
β -ME	β -mercaptoethanol
BOP	(benzotriazol-1-yloxy)-tris(dimethylamino) phosphonium hexafluorophosphate
BSA	bovine serum albumin
CCD	charge-coupled device
CCP4	Collaborative Computational Project, No. 4
CHES	cyclohexylaminoethanesulphonic acid
CoV	coronavirus
Cys	cysteine
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
DIEA	N,N-diisopropylethylamine
DMAP	4-(dimethylamino)pyridine
DMF	N,N-dimethylformamide
DMSO	dimethylsulphoxide
DTT	dithiothreitol
EB	ethidium bromide
EDTA	ethylenediaminetetraacetic acid
ESI-MS	electrospray ionisation mass spectrometry
Fmoc	9-fluorenylmethoxycarbonyl
FPLC	Fast Performance Liquid Chromatography
Gln	glutamine

Glu	glutamic acid
Gly	glycine
GST	glutathione-S-transferase
HCV	hepatitis C virus
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
His	histidine
HLE	human leukocyte elastase
HNE	human neutrophil elastase
HPLC	High Performance Liquid Chromatography
IC ₅₀	inhibitory concentration 50%
Ile	isoleucine
IPTG	isopropyl-β-D-1-thiogalactopyranoside
kDa	kiloDalton
LB	Luria-Bertani
Leu	leucine
Lys	lysine
MAD	mutiple-wavelength anomalous dispersion
MBP	maltose binding protein
MBHA	4-methylbenzhydrylamine
MES	mesityl-2,4,6-trimethylphenyl
Met	methionine
MIR	multiple isomorphous replacement
MR	molecular replacement
NMR	nuclear magnetic resonance
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
Phe	phenylalanine
PMF	peptide mass fingerprinting
PMSF	phenylmethylsulphonyl fluoride
pNA	<i>para</i> -nitroanilide

PPE	porcine pancreatic elastase
Pro	proline
SAD	single-wavelength anomalous dispersion
SARS	severe acute respiratory syndrome
SDS	sodium dodecyl sulfate
Ser	serine
SPPS	solid phase peptide synthesis
TEV	tobacco etch virus
TEMED	N,N,N',N'-tetramethylethylenediamine
TFA	trifluoroacetic acid
Thr	threonine
TI	tetrahedral intermediate
TIS	trisisopropylsilane
TLS	thin layer chromatography
TRIS	tris(hydroxymethyl)aminomethane
2TY	2 × tryptone yeast
Tyr	tyrosine
Val	valine

Chapter 1 Introduction (I): helicases and SARS-CoV helicase

1.1 Helicases and their general biochemical properties

Helicases are enzymes that unwind nucleic acid duplexes to single strands. The first helicase was discovered in 1976¹. Since that discovery, a large number of enzymes have been identified that possess helicase activity. The principle reactions catalyzed by helicases are the hydrolysis of nucleotide triphosphates and the coupling of the free energy released to the separation of two complementary strands of a duplex nucleic acid. All helicases share several common biochemical properties: NTP binding and hydrolysis, nucleic acid binding, and NTP-dependent unwinding of duplex nucleic acids². Helicases can be found in prokaryotes, eukaryotes and viruses and play important roles in many metabolic processes, including replication, recombination³, DNA repair^{4,5}, transcription⁶, translation, conjugation⁷ and RNA splicing⁸. For human helicases, malfunction can result in serious disease. Six human syndromes have been found to be, at least partly, caused by mutated enzymes that possess helicase activity, including Cockayne's syndrome, Bloom's syndrome, Werner's syndrome, xeroderma pigmentosum, α -thalassemia mental retardation on the X chromosome, and trichothiodystrophy⁹.

RNA helicases have been detected in almost all biological systems where RNA plays a central role, including RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation¹⁰. They are common in viruses, playing critical roles in viral replication. RNA helicases and DNA helicases are normally named from the originally determined function, but they should not be considered to work only on RNA or DNA respectively.

1.2 Classification of helicases

1.2.1 Substrate directionality

Helicases can be classified in various ways, but generally two ways are commonly used. The first is based on activity, and reflects substrate directionality. Helicases either require a 3' flanking strand of single-stranded nucleic acid on the substrate duplex and unwind in the 3'→5' direction or require a 5' tail on the duplex nucleic acid and unwind in the 5'→3' direction. These are known as 3'-5' helicases and 5'-3' helicases respectively¹¹. However, some helicases are also able to unwind from a blunted-ended duplex, for example, RecBCD in *E. coli*¹². Therefore it may be more appropriate to describe the polarity according to the strand

Introduction (I): helicases and SARS-CoV helicase

upon which helicases translocate¹¹. A helicase that displays a polarity in DNA/RNA unwinding is also often assumed to translocate unidirectionally along single-stranded DNA/RNA (ss-DNA/ss-RNA) with the same polarity. In general, helicases are considered to be unidirectional¹³, but bidirectional RNA helicase activity was reported with the eIF-4A/4B complex¹⁴, RH70¹⁵ and p68¹⁶.

1.2.2 Sequence level

Another alternative classification is at the sequence level. Computer-assisted analysis of amino acid sequences has led to the delineation of conserved motifs and patterns in different subsets of the helicases. According to the number of distinct motifs and differences in the consensus sequences for these motifs, helicases have been divided into five main groups¹⁷. Figure 1.1 shows the outline of the structure and arrangement of motifs from four families or superfamilies². Among them, superfamily 1 (SF1), containing Rep, UvrD, RecA, RecB, RecD, TraI, HSV-1, UL5, PcrA, yeast Pif, human coronavirus 229E, SARS-CoV helicase, *etc.*, and superfamily 2 (SF2), containing RecQ, PriA, UvrB, yeast RAD3, eIF-4A, hepatitis C virus NS3, *etc.*, are the two largest and most related groups¹⁸. They contain seven conserved motifs and generally have 3'-5' directionality¹⁹. The similarity in these motifs suggests that SF1 and SF2 might have evolved from a common ancestor²⁰. On the other hand, superfamily 3 (SF3), containing SV40 T antigen, Rep40, RepA, *etc.*, is defined by only three conserved regions, and a smaller family, family 4 (F4) containing *E. coli* DnaB, T7 gene 4 and gene 41, *etc.*, is defined by five conserved motifs and are normally hexameric helicases⁷. Lastly, a very small fifth group contains *E. coli* Rho and RuvB proteins¹⁷.

Introduction (I): helicases and SARS-CoV helicase

Figure 1.1 Schematic diagram showing representative proteins from each of four helicase families with the N-terminii on the left and C-terminii on the right. Open boxes represent the conserved helicase motifs, and the letters inside the boxes are the consensus amino acid sequences of each of motif. Labels above the open boxes are the names assigned to the motifs. The relative positions of motifs and spacing between motifs are arbitrary. In all consensus sequences, a "+" represents a hydrophobic residue, an "o" represents a hydrophilic residue and an "x" represents a residue that is not restricted to being hydrophobic or hydrophilic².

1.3 Helicase motifs

1.3.1 SF1 and SF2

In SF1 and SF2 helicases, seven conserved motifs have been identified. Motif I (Walker A) and II (Walker B) are conserved to the greatest extent across SF1 and SF2. They contain residues that interact with Mg^{2+} and ATP/ADP and are typical ATPase regions²¹ (Figure 1.2). Motif I was classically defined as a GxxxxGKT consensus, and minimally requires the three final residues GKT/GKS; Motif II in SF1 is defined as a DExx consensus, but in SF2, helicases can be divided into at least three subgroups¹¹: DEAD (Asp-Glu-Ala-Asp) box proteins, DEAH box proteins and DEXH box proteins based on differences in motif II¹⁰. The ATPase domain of helicases is a member of the AAA+ ATPase superfamily containing classic Walker A and B motifs. This domain is believed to mediate the oligomerization of the protein, ATP binding and ATP hydrolysis²².

Motifs Ia and IV are involved in oligonucleotide interaction based on the structures of Rep and HCV NS3²³. Motifs III and V show large differences in both length and amino acid sequence among the helicase families. These two motifs participate in several processes: formation of a specific linker between domains, interacting with Mg^{2+} and ATP/ADP, and binding of oligonucleotides¹⁸. In SF1 and SF2, motif VI may play a role in transducing nucleotide-induced conformational changes to the DNA binding site (motif III and IV)²⁴. Figure 1.3 demonstrates the relationship between a SF1 helicase (in this case, PcrA), and an ATP analog (ADPNP) and a DNA duplex. From the structure, it can be observed that motif I, motif II, motif IV, and motif VI are close to ADPNP; motif Ia, motif III, and motif V are situated near the DNA substrate.

Introduction (I): helicases and SARS-CoV helicase

Figure 1.2 Active site of an SF1 helicase (PcrA, PDB code: 1QHG) showing the Walker A and B motifs interacting with Mg^{2+} and ADPNP: Thr38 (Walker A motif) and Asp223 (Walker B motif) are the two residues that interact with Mg^{2+} . K37 and T38 are conserved residues in motif I (Walker A motif); D233 and E224 are conserved residues in motif II (Walker B motif); Q254, R287, K568 and R610 are conserved residues in motifs III, IV, V and VI respectively. The atoms of ADPNP are shown in red, the carbon atoms of the enzyme are shown in green and Mg^{2+} is shown in blue. The $2F_o - F_c$ electron density map (resolution 2.5 Å) for the enzyme is shown in beige and is contoured at 1 σ .

Introduction (I): helicases and SARS-CoV helicase

Figure 1.3 Structure of PcrA complexed with ADPNP and a DNA duplex (PDB code: 3PJR). The DNA substrate is shown in dark red and ADPNP in red. 'N' and 'C' represent the N-terminus and the C-terminus respectively. The seven conserved motifs are highlighted in different colours: motif I in green, motif Ia in blue, motif II in yellow, motif III in pink, motif IV in purple, motif V in cyan, and motif VI in orange. The remaining parts of the enzyme are shown in gray.

1.3.2 SF3, F4 and F5

SF3 helicases contain three conserved motifs (A, B and C)²⁰. Motifs A and C are equivalent to the Walker A and B motifs of SF1/2 helicases respectively. Motif B is similar to motif Ia in SF1/2 helicases. For the F4 and F5 helicases, five conserved motifs (1, 1a, 2, 3, and 4)^{25,26} in the helicase domain have been identified, which may correspond to SF1/2 motifs I, Ia, II, III and IV. Motif 1 and motif 2 contain the Walker A and B sequences. It has been reported that motif 1a plays a role in NTP hydrolysis^{25,27,28}. The role of motif 3 is unclear, but it is assumed to play a role in energy transduction²⁵. Motif 4 may be involved in DNA binding²⁹. It has also been suggested that these two motifs (3 and 4) may also be present in SF3 helicases²⁵. However, F5 helicases do not share significant overall sequence similarity with helicases in the other four families, not even with each other in this family. Therefore, the conserved motifs of F5 helicases have not been shown in Figure 1.1.

Introduction (I): helicases and SARS-CoV helicase

1.3.3 New conserved motifs

More recently, several new conserved helicase motifs have been identified. Two motifs TxGx and QxxR, which were not originally identified as conserved sequence fingerprints, are located between motif Ia and motif II³⁰ and have been found to interact with oligonucleotides¹⁸. The TxGx motif has been observed in helicases from almost all (super)families, but the QxxR motif is specific to the DEAD/DEAH RNA helicases³¹. Two conserved motifs among the various HCV genotypes have been identified. One motif (YRGXDV) forms a loop connecting motif IV and motif V, in which the Arg is the critical residue. Another motif (DFSLDPTF) also forms a loop, which connects motif V and motif VI. The two phenylalanine residues play critical roles. It was observed that both motifs are involved in the unwinding of the nucleic acid duplex³². Another motif, motif Q (GF..PT/S.I/VQ) was identified to be unique to the DEAD box family of RNA helicases, and is located upstream of motif I. It consists of nine amino acid residues, of which the glutamine (Q) is the most essential^{33,34}.

1.3.4 Mutagenesis studies on the conserved motifs

Mutagenesis studies on the conserved motifs are useful to delineate their function. Although most mutagenesis experiments have focused on *in vitro* analysis, including nucleic acid duplex binding and unwinding, ATP binding and hydrolysis, some *in vivo* mutagenesis studies on the conserved motifs have been carried out to identify their roles in DNA replication (motifs I, Ia, II and III) and repair (motifs II, III, IV, V and VI)^{35,36}.

For SF1 helicases, the two critical invariant adjacent residues in motif I, lysine and glycine (Figure 1.1), have been studied. The K35M mutation in UvrD results in a significant decrease in the rate of ATP hydrolysis and unwinding activity, but without altering ATP binding³⁷. Another mutation of this lysine in Upf1, K436A, gave same results²⁴. Mutation of the conserved glycine (G102V) also displayed similar results³⁸. Thus, it appears that motif I in SF1 is critical in ATP hydrolysis, but not ATP binding. Mutation of the threonine residue in Rep68 (T341A) thought to interact with Mg²⁺ demonstrated a 50% reduction in ATPase activity and no change in unwinding activity³⁹. Mutation of the motif Ia phenylalanine residue in PcrA (F64A), was found to affect ATPase activity only slightly, but almost abolish the unwinding activity⁴⁰.

Introduction (I): helicases and SARS-CoV helicase

Motif II (Walker B) in SF1 was also thought likely to be involved in the NTP hydrolysis reaction. Two mutant proteins of UvrD, E221Q and D220N/E221Q, were used to determine the function of motif II⁴¹. It was observed that both mutations significantly reduced rates of ATPase hydrolysis and helicase unwinding. Two different mutant proteins, D572A/E573A from Upf1 and D249A/E250A from HSV UL5, gave identical results. Thus, it is believed that motif I and motif II in SF1 helicases are directly involved in ATP hydrolysis, but are not necessary for ATP binding².

A version of the UL5 helicase (SF1) containing a mutation (G290S) of the conserved glycine residue of motif III (Figure 1.1) was observed to be lacking significant helicase activity³⁸. Mutagenesis studies on the extended motif III sequence of PcrA (246-ICAVGDADQSIYRWR-260) gave a similar result⁴²: the W259A and R260A mutant proteins displayed weak helicase unwinding activity; five Q254 mutations (Q254E, Q254A, Q254N, Q254R, Q254K) reduced ATPase activity but did not affect ssDNA binding. This suggested a possible role for Q254 in coupling the ATP hydrolysis to the unwinding activity.

Motif IV in SF1 is assumed to act as an NTP binding region. The crystal structure of PcrA complexed with a DNA substrate (double-stranded nucleotide with 3' tail) and ADPNP (a non-hydrolysable ATP analog) found that the conserved arginine in motif IV (Figure 1.1) makes direct contact with ADPNP (Figure 1.3)⁴³. A mutation of this conserved arginine (R345K) resulted in a significant decrease in NTP binding and hydrolysis. This gave further support for the role of motif IV in NTP binding and suggested a possible role in catalysing NTP hydrolysis³⁸. Two versions of the UL5 helicase with mutations in motif V (T809I & G815A) were observed to affect nucleic acid binding, which suggested that motif V is a nucleic acid binding region and may play a role in the coupling of free energy to unwinding³⁸. Motif VI is believed to have several roles: NTP binding and hydrolysis, and communicating between the NTP binding site and the DNA binding site by mediating conformational changes^{24,38,44}.

A large number of mutagenesis experiments have been performed to examine the roles of motifs in SF2, the largest of the five helicase families, and they gave similar results as found for SF1². In one experiment, the role of motif Ia was examined by three mutants (S110T, R112A and R113A/F115A) of the UL9 helicase. It was observed that motif Ia is involved in ss-DNA binding and unwinding activity⁴⁵. Fewer mutagenesis studies have been done on SF3,

Introduction (I): helicases and SARS-CoV helicase

F4 and F5 helicases. Mutant proteins from RepA (SF3, motif H4), DnaB (F4, motifs H1, H1a, H2, H3), and RuvB (F5, Walker A motif) have provided functional information on some motifs in these three helicase families⁴⁶⁻⁴⁸. Moreover, the role of the conserved threonine (Figure 1.1) interacting with Mg^{2+} and ATP/ADP was also examined using mutant proteins from F4 and F5 helicases: the mutant DnaB protein (T217A) abolished both the ATPase and unwinding activities⁴⁷. The same result, absence of ATPase activity, was also observed using mutant RuvB proteins (T69S and T69A)⁴⁸.

1.4 SARS-CoV helicase

1.4.1 SARS

Severe acute respiratory syndrome (SARS) is a life-threatening form of pneumonia characterised by high fever, nonprogressive cough, chills, myalgia and lymphopenia⁴⁹. Between November 2002 and June 2003, an epidemic emerged that, facilitated by international air travel, spread within weeks from the first patient in Guangdong Province, China, to many other countries. The SARS coronavirus is thought to have originated via interspecies transmission from the palm civet cat⁵⁰. In order to diagnose this disease, reverse transcription-PCR analysis with random or broadly-specific primers from coronavirus was performed⁵¹. Its rapid transmission, high mortality (~3 to 6%⁵²) and its possible re-emergence make SARS a serious threat for which neither efficient therapy nor vaccine is currently available.

1.4.2 SARS-associated coronavirus

In March, 2003, the putative cause of SARS was identified as a novel coronavirus – the SARS-associated coronavirus⁵³⁻⁵⁶. SARS coronavirus (SARS-CoV) belongs to the Coronaviridae family, which are positive-sense ss-RNA viruses⁵⁷. The new coronavirus helicase belongs to SF1¹³. The genome of SARS-CoV is approximately 29.7 kb and contains 14 open reading frames (ORFs) flanked by 5' and 3'-untranslated regions of 265 and 342 nucleotides, respectively^{54,58}. Following fusion of the viral envelope with the host cell membrane, mediated by the S protein (spike protein), the viral genome RNA is released into the cytoplasm of the infected cell⁵⁹. The translation products contain two polyproteins: a polyprotein with 4382 amino acid residues encoded by ORF1a and another polyprotein with 7073-residue encoded by ORF1b⁶⁰. The two polyproteins, named pp1a and pp1ab, have molecular masses of 486 kDa and 790 kDa respectively⁵⁵. During the viral life cycle, these two polyproteins are processed into 16 putative non-structural proteins (nsp) by

Introduction (I): helicases and SARS-CoV helicase

virally-encoded proteases. These 16 non-structural proteins include key enzymes needed for replicative functions, such as the RNA-dependent RNA polymerase (nsp 12), the helicase as well as other accessory proteins and non-structural proteins whose functions are poorly understood⁶¹.

1.4.3 SARS-CoV helicase

The SARS-CoV helicase (nsp13) is a 601 amino acid cleavage product of pp1ab and belongs to superfamily 1⁶¹. In common with other nidoviruses, the coronavirus helicase consists of two domains separated by a hinge/spacer region: an N-terminal zinc finger domain and a C-terminal catalytic domain.

The N-terminal zinc finger domain is composed of nine conserved Cys residues and three conserved His residues binding four Zn^{2+} ions, resembling the coronavirus model⁶². There are three different types of zinc finger, C_2H_2 (two cysteines and two histidines), C_2C_2 (four cysteine residues) and C_2HC (three cysteines and one histidine), in the zinc finger domain of SARS-CoV nsp13. The term ‘zinc finger’ was first used to describe a 30-residue sequence motif found in a *Xenopus* transcription factor⁶³. It is a small peptide domain, stabilized by a zinc ion bound to cysteine or histidine residues of the finger and results in a DNA-binding region which ‘grips’ the DNA. The helicase domain contains seven conserved motifs, as commonly observed in SF1 and SF2 helicases. Previous work on the SARS helicase has shown that it can unwind both DNA and RNA in the 5'-3' direction^{61,64,65}. It also indicated that the ATPase activity was stimulated by the presence of short fragments of both DNA and RNA. However, no studies have yet been reported on the contribution of the zinc finger domain to the SARS-CoV nsp13 enzymatic activity.

The initial outbreak of SARS led to a cluster of patients in Guangdong province, China (GD-01, *etc.*), a small cluster in Hong Kong (CUHK-W1, *etc.*) and then the SARS virus was subsequently spread from Hong Kong to other countries (Singapore, Taiwan, Vietnam and Canada)^{51,66}. Through alignment of different SARS sequences, it was observed that several variations existed in the nsp13 sequences, including two amino acids in the catalytic domain: one (residue 322) is either a serine or a leucine, another (residue 466) is either a glutamic acid or an aspartic acid (Figure 1.4). These variations may have originated from the molecular evolution of the SARS coronavirus⁶⁶. Neither of these residues lies within the conserved helicase motifs.

Introduction (I): helicases and SARS-CoV helicase

The majority of putative virus-encoded RNA helicases has been identified as belonging to SF1^{13,20}. This superfamily includes proteins from most of positive-stranded plant RNA virus families, several positive-stranded animal RNA viruses: alphavirus, arterivirus [equine arteritis virus (EAV), porcine reproductive and respiratory syndrome virus (PRRSV)], coronavirus [human coronavirus 229E (HCoV-229E), porcine transmissible gastroenteritis coronavirus (TGEV), murine coronavirus (MHV), avian infectious bronchitis virus (IBV), porcine epidemic diarrhoea coronavirus (PEDV), bovine coronavirus (BCoV)], hepatitis E virus (HEV), rubivirus and torovirus [Berne virus (BEV)]; as well as double-stranded RNA viruses, for example, rotavirus^{20,67}. The conserved motifs of SARS-CoV nsp13 are demonstrated through the alignment of several different viral and bacterial SF1 helicase sequences (Figure 1.5).

Figure 1.4 Diagram shows the alignment of 19 SARS-CoV nsp13 amino acid sequences. The first variable amino acid is at residue 322, the second is at residue 466. The names on the left are the abbreviated names of the SARS-CoV sequences from the NCBI database. Only parts of the nsp13 amino acid sequences are displayed.

Introduction (I): helicases and SARS-CoV helicase

Introduction (I): helicases and SARS-CoV helicase

Figure 1.5 Alignment of five SF1 helicase sequences. The diagram at the top shows the two domains of SARS-CoV nsp13: the N-terminal zinc finger domain and the C-terminal catalytic domain. Below is a sequence alignment of the catalytic domains of the RNA helicases from SARS-CoV, IBV (NCBI accession no. NP_740630) and EAV (NCBI accession no. NP_705591) and the PcrA and Rep DNA helicases. The seven conserved motifs are outlined in boxes and the conserved residues within these motifs are highlighted in green. The four residues that were mutated in the functional studies (Chapter 3) (Ser289, Ala308, Leu322 and Glu466) are highlighted in red. The RecA-like subdomains from domains 1 and 2 of PcrA are underlined in blue and orange respectively. The amino acid sequences of subdomains 1B and 2B of PcrA and Rep, which are not present in the SARS-CoV, IBV and EAV helicases, are not included. The C-terminal sections of EAV helicase (residues 447-466), PcrA (residues 675-724) and Rep (residues 667-673) are not shown. The sequence alignment was done manually apart from that of PcrA and Rep which was performed using Vector NTI (InforMax, USA).

1.5 Mechanisms of helicase unwinding

The molecular details of the mechanism of duplex unwinding are not yet fully known for any helicase. It is quite probable that not all helicases unwind nucleic acids by the same mechanism, but it is likely that there are certain common features of the unwinding mechanism. Several helicase unwinding mechanisms have been proposed. However, unwinding mechanisms can generally be classified as either active or passive, depending on whether the helicase actively participates in the unwinding event or merely stabilizes the resultant ss-DNA⁶⁸⁻⁷⁰.

1.5.1 Passive mechanism

In a passive mechanism, a helicase facilitates unwinding indirectly by binding to ss-DNA that becomes available through transient fraying of the duplex caused by thermal fluctuations at the ss-/ds-DNA junction⁷. In this manner, helicase would interact only with ss-DNA using a single site and translocate unidirectionally along ss-DNA to the duplex. Thus, unwinding would be directly coupled to translocation with a step size of one or, at most, a few base pairs⁷⁰. Recently, HCV helicase activity in the absence of ATP was reported, suggesting a possible passive mechanism⁷¹. It was observed that the ATP-independent strand-separating reaction was the result of high affinity of the enzyme for single-stranded DNA and that there was sufficient energy flow during thermal equilibration to separate the base pairs. However, this mechanism still needs direct evidence to support it.

Introduction (I): helicases and SARS-CoV helicase

1.5.2 Active mechanism

An active mechanism generally would require direct binding of the helicase to a region of the duplex and subsequent destabilization of the duplex⁷⁰. The region possesses at least two sites and the helicase can bind both ss- and ds-DNA and both simultaneously during the intermediate (helicase-DNA-NTP) stage. In principle, for non-hexameric helicases, active mechanisms can be of two types.

In the first type, the helicase interacts directly with the ds-DNA at the junction (at least transiently) and actively destabilizes a certain number of base pairs, presumably through conformational changes in the helicase triggered through NTP binding or hydrolysis. This type has two subclasses: the "inchworm" model^{43,72,73}, which is based on a monomeric operational unit, and the "rolling" model⁶⁹, which is based on an active dimeric form of the helicase (Figure 1.6). In the inchworm model, the monomeric helicase is bound to one ss-DNA strand, translocation along the strand to the fork region occurs probably as a result of ATP binding, and then the free energy released from ATP hydrolysis is coupled to base pair separation and one ss-DNA strand is released. For the rolling model, the helicase functions in a dimeric form with both monomers binding to ss-DNA. Upon ATP binding, one subunit releases the ss-DNA and binds to the duplex region. Subsequent destabilization of the duplex and release of one ss-DNA strand are similar to that in the inchworm model. It was reported that UvrD (SF1) unwinds a nucleic acid duplex with an average step size of 4 to 5 base pairs, which is approximately one-half turn of the DNA duplex⁷⁴.

A second type of active mechanism is a "torsional" model in which the helicase does not interact directly with duplex DNA, but binds simultaneously to both single strands at the ss-/ds-DNA junction and unwinds by distorting the adjacent duplex region through an NTP-induced conformational change⁶⁸. It is also based on a dimeric operational unit. This model is also observed in the unwinding mechanism of hexameric helicases (Figure 1.7).

Introduction (I): helicases and SARS-CoV helicase

Figure 1.6 Rolling and inchworm models: two active mechanisms for helicase-mediated unwinding of nucleic acid duplexes. In the rolling model, the two subunits of the dimer are coloured differently in order to distinguish between them⁴³.

Introduction (I): helicases and SARS-CoV helicase

1.5.3 Hexameric helicases

One distinct class of helicases possesses a hexameric functional unit (including T7 gp4 and gp41, RecA, RepA, DnaB, RuvB, Rho, SV40 T antigen, *etc.*). For the hexameric helicases, the basic mechanism that is proposed involves a corkscrew-like motion, in which a single strand of DNA threads through the centre of a doughnut-like helicase, while the complementary single strand passes along the outside of the ring, resulting in the separation of the nucleic acid duplex^{75,76}. This proposal can be divided into three different models: the wedge model, the torsional model, and the helix-destabilization binding model²⁶. Diagrams describing the three models are displayed in Figure 1.7.

In the wedge model, the hexamer helicase interacts with only one ss-DNA strand at the unwinding junction, as shown in Figure 1.7. One of the separated strands binds tightly through the centre, and the excluded strand does not interact tightly with the enzyme. The helicase moves unidirectionally along the strand bound to the centre, and with the release of the free energy from NTP hydrolysis, the helicase destabilizes the base pairs at the junction by a process resembling a wedge.

In the torsional model, the helicase does not interact with the duplex DNA, but interacts with both the ss-DNA strands (Figure 1.7). In the central channel of hexameric helicase, one strand binds tightly, and the other, the excluded strand, interacts tightly with the outer parts of the hexameric ring. The torque generated from these interactions is used to rotate the two strands with respect to each other and thus destabilize the duplex DNA. The enzyme then moves unidirectionally along the ss-DNA in the central channel, and unwinds the DNA by a process resembling the pulling of two threads of a piece of twine.

In the helix-destabilizing model, the helicase interacts with one of the separated strands in the central channel and the duplex DNA in the outer parts of the ring, as depicted in Figure 1.7. The protein-DNA interactions and protein conformational changes induced by NTP hydrolysis leads to the separation of the duplex in the helicase active site. The hexamer then translocates unidirectionally and binds to the newly separated strands.

Introduction (I): helicases and SARS-CoV helicase

Figure 1.7 Proposed models for DNA unwinding by hexameric helicases. In the wedge model, the hexamer interacts with one ss-DNA strand through its central channel. The excluded strand and enzyme do not interact tightly. The force producing unidirectional translocation leads to separation of duplex base pairs in a wedge type of action. In the torsional model, the helicase interacts tightly with both the separated ss-DNA strands. The rotation of one strand with respect to the other results in unwinding of the duplex DNA. In the helix-destabilizing model, the hexamer interacts with one of the separated strands in the central channel, and also interacts with the dsDNA on the outer parts of the hexamer. The duplex region is melted by the helicase, and the helicase translocates unidirectionally along the DNA in the central channel²⁶.

1.6 Structural studies of helicases

The first helicase structure, solved in 1996, was of the PcrA helicase (SF1) from *Bacillus stearothermophilus* co-crystallized with ADP⁷⁷. The crystal structure of *E. coli* Rep helicase (co-crystallized with ADP and ssDNA) was reported the following year²³. Rep and PcrA helicases share a high sequence identity (36.3%) and similar structure. More recently, PcrA co-crystallized with a partial duplex DNA substrate and a non-hydrolysable ATP analogue was solved⁴³. The structures indicated that the helicase may function as a monomer and suggested a possible inchworm unwinding mechanism. The structures of the DNA-binding domain from the replication initiation protein E1 from papillomavirus (SF1 helicase) have also been published^{78,79}. The two SF1 helicases (PcrA and Rep) from bacteria share very low sequence identity (2%) with the SARS-CoV helicase through alignment using Vector NTI. So far, no structures of a human or viral SF1 helicase have been published.

Since the first SF2 helicase from hepatitis C virus was reported^{80,81}, several SF2 helicase structures have been solved, including RecQ⁸², RecG⁸³, UvrB^{84,85}, and eIF4A³¹. For the SF3 helicases, structures of three representative members, SV40 T antigen⁸⁶, Rep 40^{87,88}, and RepA⁸⁹, have been solved. The structures of the N-terminal domain of *E. coli* DnaB, a F4 helicase, were solved by X-ray crystallography and NMR^{90,91}. The structures of another F4 helicase, T7 gene 4, were reported in 1999 and 2003⁹²⁻⁹⁴. The structures of one of the F5 helicases, *E. coli* Rho protein, have been solved by NMR and X-ray crystallography^{95,96}.

Three representative SF1/2 helicase structures are shown in Figure 1.8. It can be found that all of them contain a core of five contiguous parallel β -strands (yellow) and various α -helices. The domain containing the five β -strands is similar to the structure of the RecA protomer core. Moreover, it was found that this RecA-like domain is conserved in all SF1/2 helicases whose structures are known¹⁸. From the three structures in Figure 1.8, it can be observed that PcrA has two extra subdomains, HCV NS3 helicase has one extra domain whilst EIF4A only contains the RecA-like domains. It has been reported that PcrA functions as a monomer, and HCV NS3 functions as a dimer. The translation initiation factor EIF4A, a DEAD-box RNA helicase, belongs to the class of nonprocessive helicases and works in conjunction with EIF4B⁹⁷. The nonprocessive helicases refer to a group of helicases that can function *in vitro*, but have to work together with accessory factors or need modifications *in vivo*⁹⁸. Thus it is possible that the extra subdomains of PcrA and HCV NS3 work as DNA binding partners, which might have similar function as EIF4B.

Introduction (I): helicases and SARS-CoV helicase

Figure 1.8 Topology diagrams (left) and structures (right) of three representative helicases (PcrA, EIF4A and HCV NS3, PDB codes 1PJR, 1FUU and 1HEI)¹⁸. The conserved RecA-related 'core' is shown in yellow, variable elements in the RecA-like domains are shown in red, and the additional structural domains are shown in green or blue.

1.7 Helicases as antiviral targets

Positive-strand RNA viruses belong to a large family of viruses, several members of which cause widespread diseases. The viral genome consists of a single-stranded RNA with positive polarity. To maintain proper viral replication, nascent transcripts must be unwound from their complementary template RNA soon after transcription⁹⁹. The template RNA for transcription is highly structured, especially at its 5' and 3' termini. To synthesize a new strand of RNA, the viral replication machinery must also unwind such structures. Thus, RNA helicases play critical roles in the replication of RNA viruses^{13,100}. If an inhibitor could be found that effectively stops helicase function, viral replication would be blocked, and hence helicases have been treated as potentially attractive targets for antiviral therapy.

Several potential mechanisms might be used to inhibit helicase activity¹⁰¹. These include inhibiting ATPase activity by interfering with ATP binding, and inhibiting ATP hydrolysis or ADP release by blocking conformational shifts in the enzyme structure¹⁰². Also, inhibiting RNA binding, inhibiting unwinding by blocking helicase translocation, and inhibiting the coupling of ATP hydrolysis to unwinding are other possible mechanisms^{103,104}. For SARS-CoV, several different types of inhibitor have been reported: protease inhibitors^{105,106}, viral binding inhibitors of the spike protein^{107,108}, fusion inhibitors¹⁰⁹, and RNA interference¹¹⁰. For the SARS-CoV helicase, an inhibitor which acts on the ATPase activity has been reported¹¹¹. It was found that a chemical named HE602 inhibits the polynucleotide-stimulated ATPase activity of nsp13 *in vitro*.

1.8 Purposes of this study

The main aim of this project is to study the biochemical properties of the SARS-CoV helicase including its ATPase activity and its ability to unwind DNA/RNA substrates. Mutagenesis studies have been performed to understand the roles of residues in the conserved motifs with regards to helicase function. Secondary aims are to obtain the structures of the SARS-CoV helicase both with and without bound cofactors/substrates using X-ray crystallography.

Chapter 2 Cloning, expression, and purification of SARS-CoV helicase

2.1 Cloning of helicases

The orf1ab cDNA from the virus isolate KYK from a Singapore patient infected with SARS (Sin2774) was provided by the Genome Institute of Singapore. Using PCR, two sequences of the SARS genome (16151-17953, representing the full-length helicase and 16928-17953, representing an N-terminally truncated helicase missing the zinc finger domain and hinge/spacer region) were amplified and cloned as NcoI-XhoI or BamHI-PstI fragments into several different plasmids (pET24d, pETM-30, pMAL-C2X) with different affinity tags. Positive clones were screened by digestion with NcoI-XhoI or BamHI-PstI.

The SARS-CoV helicase (nsp13) contains an N-terminal zinc finger domain, hinge/spacer region and a C-terminal catalytic domain. Ten different clones were successfully constructed (Figure 2.1): SARS-CoV helicase without and with a C-terminal 6×His tag (constructs 1 and 2), SARS-CoV helicase with an N-terminal 6×His tag (construct 3), N-terminally truncated SARS-CoV helicase with and without a C-terminal 6×His tag (constructs 4 and 5), SARS-CoV helicase and N-terminally truncated helicase with an N-terminal GST tag (constructs 6 and 7), N-terminally truncated helicase with a special linker (Pro-Ser-Pro) and an N-terminal GST tag (construct 8), SARS-CoV full-length helicase and N-terminally truncated helicase with an N-terminal MBP tag (constructs 9 and 10). The N-terminally truncated helicase (Δ 1-259, deletion of the zinc finger domain and hinge/spacer region) contains only the catalytic domain. All the positive clones were confirmed by DNA sequencing.

Cloning, expression, and purification of SARS-CoV helicase

Figure 2.1 Ten different SARS-CoV helicase (nsp13) constructs. Construct 1 is the SARS full-length helicase [containing the zinc finger domain (residues 1-84), the hinge/spacer region (residues 85-259) and the catalytic domain (residues 260-601)] without any tag. Constructs 2 to 10 contain a combination of various different tags and helicase domains. The black, light-pink, brown, blue, red, green, and pink boxes represent zinc finger domain, hinge/spacer region, catalytic domain, 6xHis tag, maltose binding protein (MBP), glutathione-S-transferase (GST), and Pro-Ser-Pro (PSP) linker respectively.

2.2 Expression of helicases

First, SARS full-length nsp13 and Δ 1-259 nsp13 (construct 1 and 5) were expressed without any tag. However, considering the low amount of soluble expression and the difficulty in purification, expression of fusion proteins with a tag was the next approach. Three different tags: six sequential histidine residues (6 $\times$ His), glutathione-S-transferase (GST) and maltose binding protein (MBP), were utilized to aid expression and purification of SARS-CoV nsp13. Full-length nsp13 and Δ 1-259 nsp13 were always expressed concurrently. The 6 $\times$ His tags, both N-terminal and C-terminal, were the first to be used. The expression of the full-length nsp13 with a C-terminal 6 $\times$ His tag (construct 2) was not high, but could be purified using affinity resins. However, full-length fusion protein with an N-terminal 6 $\times$ His tag (construct 3) and Δ 1-259 nsp13 with a C-terminal 6 $\times$ His tag (construct 4) gave poor soluble expression and could not be purified using affinity resin.

In order to improve expression, other tags were utilized. GST was the second tag used in the study. However, full-length nsp13 with an N-terminal GST tag (construct 6) was always contaminated with the *E. coli* protein GroEL, which was difficult to remove during purification. Δ 1-259 nsp13 with an N-terminal GST tag (construct 7) was found to be degraded during the expression, which might have resulted from the long flexible linker (~15 residues) between the tag and protein. Thus, a different Δ 1-259 nsp13 GST fusion protein with a special linker (-Pro-Ser-Pro-) (construct 8) was expressed. However, as no soluble expression was observed. An MBP tag was tried next to improve expression. Full-length nsp13 with an N-terminal MBP tag (construct 9) gave a certain degree of soluble expression, but it could not be purified using affinity resin. Δ 1-259 nsp13 with an N-terminal MBP tag (construct 10) was expressed and purified using amylose affinity resin.

2.2.1 6 $\times$ His tag and no tag

Full-length nsp13

After transformation into *E. coli* BL21 (DE3), soluble expression of SARS-CoV full-length nsp13 without and with a C-terminal 6 $\times$ His tag was achieved (Figure 2.2). After trials, the best condition for expression of these constructs (1 and 2) was initial growth in 2TY media at 37 °C, addition of 0.4 mM IPTG at OD₆₀₀=0.6, followed by further growth for 14 hours at 30 °C before harvesting.

Cloning, expression, and purification of SARS-CoV helicase

Figure 2.2 Expression of SARS-CoV full-length nsp13 in *E. coli* BL21 (DE3): **A**, construct 1, without a C-terminal 6×His tag; **B**, construct 2, with a C-terminal 6×His tag. In both SDS-PAGE gels, lanes 1 and 2 represent total cell protein of the control (no induction) and after induction by IPTG respectively; lanes 3 and 4 represent soluble protein (in the supernatant) of the control and after induction respectively; lanes 5 and 6 represent insoluble protein (in the pellet) of the control and after induction respectively. Mw: molecular weight standards. The red arrows indicate the helicase without a 6×His tag (left, 67.3 kDa) and helicase with a 6×His tag (right, 68.1 kDa).

For the nsp13 with an N-terminal 6×His tag (construct 3, Figure 2.3), the best expression condition was initial growth in 2TY media at 37 °C, addition of 0.4 mM IPTG at $OD_{600}=0.6$, followed by further growth for 12 hours at 25 °C before harvesting. However, most of this fusion protein was expressed as inclusion bodies and the small amount of soluble protein could not be purified using nickel affinity resin.

Truncated nsp13 ($\Delta 1-259$)

After analyzing the soluble and insoluble fractions, it was found that the expressed truncated SARS-CoV helicases (catalytic domain without zinc finger domain and hinge/spacer region), both with and without C-terminal 6×His tag, were located in inclusion bodies (constructs 4 and 5, Figure 2.4). Several different concentrations of IPTG (0.04 mM - 1 mM), different induction temperatures (18 °C - 37 °C), different induction times (2 hours-overnight) and different induction points ($OD_{600}=0.6$, 1.0, 1.2) were tried in an attempt to obtain soluble expression, but soluble expression could not be achieved. So the best induction condition for expression in inclusion bodies was initial growth in 2TY media at 37 °C, addition of 1 mM IPTG at $OD_{600}=0.6$, followed by further growth for 2 hours at 37 °C before harvesting.

Cloning, expression, and purification of SARS-CoV helicase

Figure 2.3 Expression of SARS-CoV full-length nsp13 with an N-terminal 6×His tag (construct 3) using two different induction conditions: addition of 0.4 mM IPTG at $OD_{600}=0.6$ and then growth for a further 2 hours at 37 °C (gel A) or growth for a further 12 hours at 25 °C (gel B). Lanes 1-3 represent total cell protein (TCP), supernatant (SP), and the pellet (insoluble protein, ISP) of the control (no induction) respectively; lanes 4-6 show TCP, SP and ISP after induction by IPTG respectively. Mw: molecular weight standards. The red arrows indicate SARS-CoV nsp13 with an N-terminal 6×His tag (68.1 kDa).

Figure 2.4 Expression of SARS-CoV $\Delta 1-259$ nsp13 in *E. coli* BL21 (DE3): **A**, the protein with a C-terminal 6×His tag (construct 4); **B**, the protein without a 6×His tag (construct 5). Lanes 1 and 2 in gel A and lanes 3 and 4 in gel B represent soluble protein of the control (no induction) and after induction by IPTG respectively; lanes 3 and 4 in gel A and lanes 5 and 6 in gel B represent insoluble protein of the control and after induction respectively; lanes 5 and 6 in gel A and lanes 1 and 2 in gel B represent total cell protein of the control and after induction respectively. Mw: molecular weight standards. The red arrows indicate $\Delta 1-259$ nsp13 with a 6×His tag (left, 38.4 kDa) and $\Delta 1-259$ nsp13 without a 6×His tag (right, 39.2 kDa).

Cloning, expression, and purification of SARS-CoV helicase

2.2.2 Glutathione-S-transferase (GST) tag

Full-length nsp13

In order to improve the expression of nsp13, the 6xHis tag was replaced with a GST tag. The optimal soluble expression of the SARS-CoV full-length helicase with an N-terminal GST tag (construct 6) was obtained using the following conditions: initial growth in 2TY media at 37 °C, addition of 0.1 mM IPTG at OD₆₀₀=1.2, followed by further growth for 12 hours at 20 °C before harvesting (Figure 2.5A). The expression using these conditions was still relatively low. In all induction conditions, most of expressed protein was located in the insoluble fraction.

Figure 2.5 Expression of SARS-CoV full-length (A) and Δ 1-259 (B) nsp13 with an N-terminal GST tag (constructs 6 and 7 respectively). In gel A, lanes 1-3 represent total cell protein (TCP), supernatant (SP), pellet (insoluble protein, ISP) of the control (no induction) respectively; lanes 4-6 show TCP, SP and ISP after induction by IPTG (0.1 mM) respectively; lanes 7-9 represent TCP, SP and ISP after induction by IPTG (0.4 mM) respectively. In gel B, lanes 1-3 represent TCP, SP and ISP after induction; lanes 4-6 show TCP, SP and ISP of the control. Mw: molecular weight standards. The red arrows indicate the fusion protein (left, 95.9 kDa) and the degraded fragment (right).

Truncated nsp13 (Δ 1-259)

However, for the Δ 1-259 nsp13 fusion protein with an N-terminal GST tag (construct 7), expression could not be achieved despite extensive induction trials, but a large band with similar molecular weight to that of truncated helicase could be seen on the SDS-PAGE gel (Figure 2.5B). In an attempt to overcome this problem (probably a result of *E. coli* proteases), an alternative GST fusion protein with a special linker (-Pro-Ser-Pro-) was constructed. The objective was to obtain complete GST- Δ 1-259 nsp13 fusion protein. As the cyclic structure of

Cloning, expression, and purification of SARS-CoV helicase

proline, which has greater conformational rigidity compared to other amino acids tends to make proteolytic cleavage more difficult, this special linker containing two proline residues was introduced using PCR (construct 8). No degradation was observed showing that the new linker functioned as anticipated, but unfortunately the new fusion protein was initially almost entirely insoluble (Figure 2.6A). However, after altering the induction, a small amount of soluble fusion protein could be obtained using the following expression conditions: initial growth in 2TY media at 37 °C, addition of 0.04 mM IPTG at OD₆₀₀=1.0, followed by further growth for 12 hours at 25 °C before harvesting.

Figure 2.6 Expression of **A:** SARS-CoV Δ1-259 nsp13 with GST tag and PSP linker (construct 8) and **B:** SARS-CoV full-length nsp13 with an N-terminal MBP tag (construct 9). Gel A: lanes 1-3 are TCP (total cell protein) of the control (no induction), after induction (addition of 0.4 mM IPTG at OD₆₀₀=0.6 and then growth for a further 4 hours at 37 °C), and after induction (addition of 0.4 mM IPTG at OD₆₀₀=0.6 and then growth for a further 4 hours at 30 °C) respectively; lanes 4-6 and lanes 7-9 are SP (supernatant protein) and ISP (insoluble protein) of the three conditions respectively. Gel B: lanes 1, 5 and 9 are TCP, SP and ISP of the control (no induction); lanes 2, 6 and 10 are TCP, SP and ISP after induction (addition of 40 μM IPTG at OD₆₀₀=0.6 and then growth for a further 4 hours at 37 °C); lanes 3, 7 and 11 are TCP, SP and ISP after induction (addition of 40 μM IPTG at OD₆₀₀=1.0 and then growth overnight at 30 °C); lanes 4, 8 and 12 are TCP, SP and ISP after induction (addition of 40 μM IPTG at OD₆₀₀=1.0 and then growth overnight at 25 °C). Mw: molecular weight standards. The red arrows indicate GST fusion protein (left, 63.5 kDa) and MBP fusion protein (right, 117.7 kDa).

Cloning, expression, and purification of SARS-CoV helicase

2.2.3 Maltose binding protein (MBP) tag

Considering the poor expression using a GST tag, an MBP tag was then tried. Soluble expression of SARS-CoV Δ 1-259 nsp13 with an N-terminal MBP tag (construct 10) was achieved using the following expression conditions: initial growth in 2TY media at 37 °C, addition of 0.04 mM IPTG at OD₆₀₀=0.6, followed by further growth for 4 hours at 37 °C before harvesting (Figure 2.7). Soluble expression of the MBP and full-length helicase fusion protein could not be obtained despite extensive trials (construct 9, Figure 2.6).

Figure 2.7 Expression of SARS-CoV Δ 1-259 nsp13 with an N-terminal MBP tag (construct 10): **A**: lanes 1-3 are TCP (total cell protein), SP (soluble protein) and ISP (insoluble protein) of the control (no induction); lanes 4-6 are TCP, SP and ISP after induction (addition of 40 μ M IPTG at OD₆₀₀=0.6 and then growth for a further 4 hours at 37 °C); lanes 7-9 represent TCP, SP and ISP after induction (addition of 40 μ M IPTG at OD₆₀₀=1.2 and then growth for a further 4 hours at 37 °C). **B**: lanes 1-3 (OD₆₀₀=0.6) and lanes 4-6 (OD₆₀₀=1.2) are TCP, SP and ISP after induction (addition of 40 μ M IPTG at OD₆₀₀=0.6 or 1.2 and then growth for a further 4 hours at 25 °C); lanes 7-9 (OD₆₀₀=0.6) and lanes 10-12 (OD₆₀₀=1.2) are TCP, SP and ISP after induction (addition of 40 μ M IPTG at OD₆₀₀=0.6 or 1.2 and then growth for a further 4 hours at 30 °C). Mw: molecular weight standards. The red arrow indicates the MBP fusion protein (left, 95.7 kDa).

2.2.4 Improvement of expression

In this study, six sequential histidine residues (6 $\times$ His), glutathione-S-transferase (GST) and maltose binding protein (MBP) are three commonly used tags to aid expression and/or purification of proteins. In the case of the C-terminal 6 $\times$ His tag, the expression of the full-length nsp13 fusion protein was not high, but could be purified using either nickel or cobalt metal affinity resin. A similar expression level of SARS-CoV full-length helicase in pET28a was reported⁶⁴. However, the full-length helicase fusion protein with an N-terminal 6 $\times$ His tag was almost completely insoluble.

Cloning, expression, and purification of SARS-CoV helicase

Often modifications in the expression conditions can increase solubility¹¹². Screening for the best host/expression construct combinations, inducing at lower temperature, lowering the induction agent concentration, choosing different media, and expressing the protein as a fusion construct with a solubilizing partner such as GST or MBP are the principal methods of achieving better solubility. As an example, the GST fusion protein was obtained at low temperature (construct 6, Figure 2.5) and the MBP fusion protein was induced at very low IPTG concentration (construct 10, Figure 2.7). It was found that the point at which IPTG was added could also affect the expression. For example, in the case of the GST full-length nsp13 construct (construct 6), IPTG was added very late (at an OD₆₀₀ of 1.2) to maximise solubility. The overall conclusion appears to be that MBP is better in helping the Δ 1-259 nsp13 fold than the other two tags. Two reported fusion proteins of SARS-CoV nsp13 also used the MBP tag^{61,65}.

2.3 Purification of proteins

2.3.1 6xHis affinity purification

The proteins with a C-terminal 6×His tag were purified using two different affinity columns (Ni²⁺ and Co²⁺ resin), an anion exchange column (MonoQ) followed by a gel filtration column (Superdex 200, 10/300 GL). The final elution solutions were almost pure (> 95% as judged by SDS-PAGE analysis). From the purification trials, it seemed that the Co²⁺ metal affinity resin was better at purification than the Ni²⁺ affinity resin. The protein without 6×His tag (construct 1) was more difficult to purify. In this case, three columns were used: anion exchange (1 mL MonoQ), gel filtration (Superdex 200, 10/300 GL) followed by further anion exchange (1 mL MonoQ). As it was found that the untagged protein had similar functional activity as the protein with a C-terminal 6×His tag, the untagged protein was not used for further functional studies due to the more complicated purification procedure. SARS-CoV nsp13 with a C-terminal 6×His tag (about 96%-99% pure) (construct 2, Figure 2.8) was purified using this procedure, concentrated, desalted and then used in the functional and crystallization studies. This protein was positively detected with a commercial rabbit polyclonal antibody raised against the 6×His tag.

Moreover, this protein (full-length nsp13 with a C-terminal 6×His tag) was also examined using mass spectrometry and N-terminal protein sequencing. The N-terminal protein sequencing result was difficult to interpret because no single significant peak could be

Cloning, expression, and purification of SARS-CoV helicase

observed in each reaction cycle. The Edman degradation is a common method for determining the amino acid sequence of a target protein or peptide¹¹³. However, poor results from N-terminal sequencing are not unusual due to various limitations¹¹⁴. Four amino acids (Ser, Thr, Arg, His) usually have low signals because of the low yield of phenylthiohydantoin amino acids. In addition, for proteins which are unstable or contaminated by other proteins, good sequencing results are difficult to obtain. Furthermore, proteins which have been N-terminally modified (such as acetylation, formylation, *etc.*) cannot undergo Edman degradation. Normally, if protein sequencing works well, then the protein identification is accurate. The poor results in this case meant that an alternative method of protein identification, mass spectrometry, had to be used.

A peak of approximately 68 kDa with low intensity was observed from the mass spectrum using MALDI-TOF (the calculated mass of full-length nsp13 with a C-terminal 6×His tag is 68.1 kDa). The mass spectrometric analysis of the intact helicase did not provide accurate data due to the high molecular weight, but it indicated the molecular weight of the protein was approximately correct. Protein identity was confirmed using peptide mass fingerprinting (PMF). However, it also indicated the presence of *E. coli* glucosamine 6-phosphate synthase (~ 66 kDa) as a minor contaminant. This protein exhibits no ATPase or nucleic acid duplex unwinding activities and therefore is unlikely to interfere with functional assays¹¹⁵.

Meanwhile, the catalytic domain (Δ 1-259 nsp13) with and without a C-terminal 6×His tag (construct 4 and 5) were also purified. Although the two proteins were located largely with inclusion bodies, they gave a significant band on an SDS-PAGE gel. Therefore, a refolding procedure was performed. Only Δ 1-259 nsp13 with a C-terminal 6×His tag was purified, and this was accomplished using an affinity column (Co^{2+} affinity resin), refolding and then an anion exchange column (1 mL MonoQ). It was found that the protein was unstable, and lost activity after storage for one week at 4 °C. However, as the expression was good, this refolded Δ 1-259 nsp13 fusion protein (rapidly purified and stored at -80 °C) was utilized in the functional assays.

Cloning, expression, and purification of SARS-CoV helicase

Figure 2.8 FPLC chromatogram of the third step of the purification of SARS-CoV nsp13 with a C-terminal 6×His tag (construct 2). The protein was purified using a gel filtration column (Superdex 200, 10/300 GL) with 50 mM Tris pH 8.0 and 0.1 M NaCl as the elutant. The inset gel shows peak fractions (A1-A7) run on 10% SDS-PAGE.

Cloning, expression, and purification of SARS-CoV helicase

2.3.2 GST purification

For the full-length nsp13 with an N-terminal GST tag, after being purified by glutathione affinity resin, three bands were obtained in the elutant. Based on the molecular weight, they were assumed to be the intact fusion protein (construct 6), nsp13, and the GST protein. This was confirmed by western blotting and trypsin digestion followed by peptide mass fingerprinting (PMF). Full-length nsp13 was obtained using TEV (tobacco etch virus) protease to cleave the fusion protein (Figure 2.9). The recognition site of TEV protease is Glu-X-X-Tyr-X-Gln-Ser and cleavage occurs between the Gln and Ser residues. It was found that similar results were obtained from on-column cleavage and in-solution cleavage. Final purification was achieved using an anion exchange column (1 mL MonoQ) (Figure 2.10). The GST fusion protein with a PSP (-Pro-Ser-Pro-) linker (construct 8) could not be purified using glutathione affinity resin.

PMF of purified construct 6 indicated the presence of *E. coli* GroEL. Proteins containing a GST tag that are expressed in *E. coli* are often found to be contaminated by GroEL¹¹⁶. GroEL helps proteins fold through interaction with GroES¹¹⁷. Overexpression of GroEL is commonly observed when a protein is difficult to express. 3 M urea was added to the lysis buffer and the wash buffer in an attempt to remove GroEL, but this was not successful.

Figure 2.9 A: cleavage of SARS-CoV full-length nsp13 with an N-terminal GST tag (construct 6) using TEV protease with a molar ratio of 1:20 (protease:sample) in buffer (50 mM Tris pH 8.0, 1 mM EDTA and 1 mM DTT) at 8 °C overnight. Lane 1 and lane 2 represent proteins after cleavage and before cleavage. **B:** purification of Δ 1-259 nsp13 with an N-terminal MBP tag (construct 10) using affinity resin. Lane 1 is the sample (total cell protein) used in the purification; lane 2 shows the elution after purification. Mw: molecular weight standards. The red arrows refer to the GST fusion protein (left, 95.9 kDa) and the MBP fusion protein (right, 95.7 kDa).

Cloning, expression, and purification of SARS-CoV helicase

Figure 2.10 FPLC chromatogram of the second step of the purification of cleaved SARS-CoV full-length nsp13 with an N-terminal GST tag (construct 6). The protein was eluted from an anion exchange column (1 mL MonoQ) at a relatively high NaCl concentration (~ 0.8 M). The inset gel shows peak fractions (A1-A7) run on 10% SDS-PAGE. The orange line represents the NaCl concentration gradient.

Cloning, expression, and purification of SARS-CoV helicase

2.3.3 MBP purification

The fusion protein of full-length nsp13 and an N-terminal MBP tag (construct 9) could not be purified using amylose affinity resin. However, purification of $\Delta 1-259$ nsp13 with an N-terminal MBP tag (construct 10) was successful (Figure 2.9). It was observed that after the fusion protein was cleaved using factor Xa (the substrate specificity of factor Xa is Ile-Glu-Asp-Gly-Arg and cleavage occurs between the Glu and Asp residues, although it will sometimes cleave at other basic residues, depending on the conformation of the protein substrate), the helicase was unstable. Only MBP could be observed on the SDS-PAGE gels (identified using the affinity resin). The intact MBP-catalytic domain fusion protein (~95% purity) could be obtained using MBP affinity resin, MonoQ column and a further gel filtration step (Superdex 200) (Figure 2.11, Figure 2.12).

Figure 2.11 FPLC chromatogram of the second step of the purification of SARS-CoV catalytic domain with an N-terminal MBP tag (construct 10). The protein was eluted using an anion exchange column (1 mL MonoQ) with 50 mM Tris pH 8.0 and 1 M NaCl as elutant. The gradient (orange) of NaCl was up to 95% buffer B in 20 min with a flowrate of 1 mL min⁻¹. The inset gels show corresponding peak fractions (A9-B11) run on 10% SDS-PAGE.

Cloning, expression, and purification of SARS-CoV helicase

Figure 2.12 FPLC chromatogram of the third step of the purification of SARS-CoV catalytic domain with an N-terminal MBP tag (construct 10). The protein was purified using a gel filtration column (Superdex 200, 10/300 GL) eluted with 50 mM Tris pH 8.0 and 0.1 M NaCl. The inset gel shows peak fractions (A1-A4) run on 10% SDS-PAGE.

Cloning, expression, and purification of SARS-CoV helicase

The MBP- Δ 1-259 fusion protein was examined using mass spectrometry. A good PMF (peptide mass fingerprint) was obtained after trypsin digestion. The search result from the Mascot database and further MS/MS analysis confirmed the identification.

Meanwhile, the folding of the MBP- Δ 1-259 fusion protein was also examined using circular dichroism (CD) spectrometry (Figure 2.13). Far-UV CD analyses secondary structures in protein and it was found that there was a trough in the 202-240 nm region at the CD spectrum. The signal at 218 nm is the indicative of β -sheet forms, whilst the signals at 208 nm and 224 nm can be assigned to α -helical structures. These results indicated the existence of secondary structure, suggesting the protein was correctly folded. However, the correct level of folding may not be completely derived from a CD spectrum.

In addition, the protein was analyzed using dynamic laser light scattering (90 plus, Brookhaven Instruments Corporation). Data showed the effective diameter (average diameter) for the particles in the solution was 91.8 nm (Figure 2.14). The lognormal distribution was displayed to visualize the size distribution. The MSD (multimodal size distribution) summary, indicating more than one peak, shows most particles are near to 100 nm, which might be dust particles or aggregated protein particles (Figure 2.15). The diameter of a single protein molecule of 95.7 kDa is approximately 8.34 nm (Zetasizer calculator, www.malvern.co.uk). Therefore, the particle with diameter 20.6 nm observed in the experiment is most likely to be an oligomeric form.

Cloning, expression, and purification of SARS-CoV helicase

Figure 2.13 Far UV circular dichroism spectrum (190 nm - 260 nm) for SARS-CoV Δ 1-259 nsp13 with an N-terminal MBP tag (construct 10) (0.8 mg mL^{-1}) using a 0.1 mm quartz cell.

Cloning, expression, and purification of SARS-CoV helicase

Figure 2.14 Lognormal distribution using the data obtained from SARS-CoV Δ 1-259 nsp13 with an N-terminal MBP tag (0.1 mg mL^{-1}) was displayed at 5% intervals (diameters, in nm). The data were obtained by 660 nm laser light scattering at 90° . The effective diameter of the sample was 91.8 nm. Through three runs, the standard error was below 1%.

Figure 2.15 Bimodal particle size distribution obtained from SARS-CoV Δ 1-259 nsp13 with an N-terminal MBP tag (diameters, in nm). The sample exhibited a bimodal distribution, which indicated a probable oligomer.

Chapter 3 Functional studies on SARS-CoV helicase

In this study, analysis of the ATPase and nucleotide duplex unwinding activities with different ^{32}P -labelled DNA and RNA substrates was performed using full-length and $\Delta 1$ -259 nsp13 with a C-terminal 6 $\times$ His tag (constructs 2 and 4), and eight mutant proteins (construct 2).

3.1 Mutagenesis of SARS-CoV helicase

The original objective was to study the functional differences among the nsp13 proteins from different SARS-CoV genotypes using mutagenesis. Table 3.1 shows the mutations that were completed and their positions in the SARS-CoV nsp13 amino acid sequence. Wild type protein was obtained by rectifying two mutations (S289G and R390K) found in the cDNA obtained from the Genome Institute of Singapore (GIS). The mutation R390K was investigated purely out of scientific curiosity. The mutation S289G is located in motif I, which is known to be involved in ATP binding and hydrolysis. The D466E and L322S/D466E mutations correspond to the BJ01 (from Beijing, China) and GD01 (from Guangdong province, China) SARS-CoV genotypes respectively (Figure 1.4, Chapter 1). The A308V mutation (located in motif Ia) was accidentally introduced (*via* PCR) during correction of the GIS cDNA. As relatively few mutational experiments have been conducted thus far on motif Ia, this accidental mutation was considered a valuable opportunity for further functional study. An SDS-PAGE gel of the nine full-length proteins used in the functional studies is shown in Figure 3.1. Truncated nsp13 ($\Delta 1$ -259) used in the study, consisting of only the catalytic domain with a C-terminal 6 $\times$ His tag (construct 4), yielded insoluble protein which was therefore subjected to a refolding protocol (Chapter 2). As the exact percentage of protein that was correctly refolded was unknown, caveats apply when comparing the specific activity of the full-length and $\Delta 1$ -259 helicases.

Functional studies on SARS-CoV helicase

Proteins Sites	289 (S/G)	308 (A/V)	322 (L/S)	390 (R/K)	466 (D/E)
W.T.	S	A	L	R	D
S289G	G	A	L	R	D
R390K	S	A	L	K	D
S289G/R390K	G	A	L	K	D
D466E	S	A	L	R	E
L322S/D466E	S	A	S	R	E
A308V	S	V	L	R	D
A308V/R390K	S	V	L	K	D
A308V/D466E	S	V	L	R	E

Table 3.1 Nine different SARS-CoV nsp13 constructs: WT refers to wild type protein. Residue 289 is located in motif I, and residue 308 in motif Ia. Residues 322, 390, 466 are located between motif I and II, motif II and III, motif IV and V respectively. The D466E and L322S/D466E proteins are two naturally-occurring variants of nsp13 (Figure 1.4, Chapter 1).

Figure 3.1 Nine different SARS-CoV nsp13 constructs displayed on a 10% SDS-PAGE gel: lane 4 is the wild type protein; lanes 1-3 represent S289G/R390K, S289G, L322S/D466E respectively; lanes 5-9 represent R390K, D466E, A308V, A308V/R390K, and A308V/D466E respectively. Mw: molecular weight standards.

3.2 ATPase activity

Hydrolysis of ATP is an essential part of helicase function. In order to investigate the effects of the mutations on the ATPase activity and to examine whether the proteins were active, nine different SARS-CoV nsp13 constructs with a C-terminal 6×His tag (Figure 3.1) were used in the studies. Additionally, the effects of different polyribonucleotides and divalent cations on the helicase ATPase activity were also investigated.

Using the ^{32}P -ATPase activity assay, the expressed full-length SARS-CoV helicase with a C-terminal 6×His tag was found to hydrolyse ATP effectively, but the activity was perhaps slightly lower than that previously reported¹¹⁸. Initially, 5 mM “cold” (*i.e.* non-radioactive) ATP was added to the reaction solution as suggested in the literature¹¹⁸, but hydrolysis of ATP was not observed. When the assay was repeated without cold ATP and only a small quantity of $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$, release of $^{32}\text{P}_i$ was observed.

Figure 3.2 shows the ATPase activities of nine different proteins (wild type and eight mutants). A308V has the highest activity (380% of the wild-type), whereas, S289G/R390K and A308V/D466E display lower activities. Three different proteins, L322S/D466E, D466E, and wild-type represent helicases from three SARS-CoV strains: SARS-GD01, SARS-CUHKW1, and SARS-Sin2774 respectively (Figure 1.4 from the introduction). Compared to wild-type nsp13, a slight increase in ATPase activity (108% of the wild-type activity) was observed with the D466E mutant and a slight decrease was obtained from the L322S/D466E mutant (78% of the wild-type activity). The activity of the S289G mutant was slightly lower than that of the wild-type enzyme (75% of the original activity). The mutant with the zinc finger and hinge/spacer domains deleted (Δ 1-259, Figure 3.3) retained about 65% of the activity of the wild-type enzyme, which was calculated by comparing the percentage of $^{32}\text{P}_i$ released. Thus, after comparing the ATPase activities of the nine nsp13 constructs, only four different proteins (wild type, Δ 1-259, A308V and S289G/R390K) were utilized in further functional studies.

Functional studies on SARS-CoV helicase

Figure 3.2 ATPase activities of wild type nsp13 and eight different mutants (S289G, L322S/D466E, R390K, D466E, S289G/R390K, A308V, A308V/R390K, A308V/D466E). The activity of wild type nsp13 was set as 100%. The values are the average of at least three measurements and the error bars represent one standard deviation. Scan from the phosphorimager (inset figure) shows the raw data with the release of ^{32}Pi being separated by TLC. The relative ATPase activities were also adjusted according to their protein concentrations used in the study. Lanes 1 and 2 are negative (using heat-denatured helicase) and positive controls (acid-catalyzed hydrolysis) respectively. Lanes 3-11 correspond to the numbers indicated on the graph.

Functional studies on SARS-CoV helicase

Figure 3.3 Graphs showing the effect of the addition of poly(A) (200 μg mL⁻¹), Zn²⁺ (3 mM), with and without EDTA pre-treatment (1 mM EDTA treating protein for 15 min prior to reaction) on refolded wild type nsp13 Δ1-259. The activity of the truncated nsp13 (Δ1-259) without any additive was set as 100%. The values are the average of at least three measurements and the error bars represent one standard deviation. Scan from the phosphoimager (inset figure) shows the raw data with the release of ³²Pi being separated by TLC. Lanes 1 and 2 are negative (using heat-denatured helicase) and positive controls (acid-catalyzed hydrolysis) respectively. Lanes 3-7 correspond to the numbers indicated on the graph.

Functional studies on SARS-CoV helicase

The objective of these further experiments was to study the effects of polyribonucleotides and divalent cations and to clarify the role of the zinc finger domain. For the wild-type helicase, it was found that in the presence of poly(A) ($200\ \mu\text{g mL}^{-1}$) the ATPase activity was reduced to 58% of the original (Figure 3.4). In the case of $\Delta 1-259$, a similar reduction in ATPase activity (57% of the original) was observed upon addition of the same concentration of poly(A) (Figure 3.3). With the A308V mutant protein a similar decrease in activity (47% of the original) was noted (Figure 3.5) suggesting that this mutation in motif Ia did not greatly affect the observed inhibitory effect. Further experiments with different polyribonucleotides [poly(A), poly(U), poly(G), poly(C)] at concentration of 50 and $200\ \mu\text{g mL}^{-1}$ conducted with the A308V mutant protein gave similar reductions in activity (Figure 3.6) suggesting that the observed inhibition was sequence-independent. When a 45 base ssDNA oligomer (see Table 8.1 in Chapter 8) was added to the assay, a decrease in activity was also noted. However, when dsDNA (substrate 4, Table 3.2) or dsRNA (substrate 8, Table 3.2) was added to the assay, a slight increase in activity was observed.

Wild-type and A308V mutant enzyme pretreatment with 1 mM EDTA (magnesium was added in the assay) displayed activity reduced to ~ 50% and 30% of the original respectively (Figure 3.4). The presence of poly(A) further reduced the activity of EDTA-treated enzyme by a similar amount as observed for the untreated enzyme, thus implying that the effect of poly(A) is independent of EDTA pretreatment. However, EDTA pretreatment of $\Delta 1-259$ did not significantly affect the activity (Figure 3.3). Poly(A) reduced the activity of EDTA-treated helicase $\Delta 1-259$ by a similar amount as for untreated enzyme. These results suggest that the presence of the zinc finger domain is required for the observed reduction in activity with EDTA pretreatment and is again suggestive that the effect of poly(A) is independent of that of EDTA. Moreover, when Zn^{2+} (3 mM) was added into the reaction solution with the wild-type, A308V and $\Delta 1-259$ in the presence of poly(A), the ATPase activities were reduced except for S289G/R390K showing a minor increase, which demonstrates the generally negative competitive effect of Zn^{2+} (Figures 3.3, 3.4 and 3.5). Furthermore, a similar decrease was also observed when Zn^{2+} (3 mM) was added into the reaction solution with the A308V mutant protein in the absence of poly(A) (Figure 3.6). In addition, replacing Mg^{2+} with Zn^{2+} (3 mM) plus Mg^{2+} (3 mM) or Mn^{2+} (3 mM) resulted in a decrease in activity (40%) or an increase in activity (20%) respectively (Figure 3.6).

Functional studies on SARS-CoV helicase

Figure 3.4 Graphs showing the effect of the addition of poly(A) (200 µg mL⁻¹), Zn²⁺ (3 mM), with and without EDTA pre-treatment on wild-type, S289G/R390K and A308V SARS-CoV nsp13. The wild-type enzyme without addition of poly(A) and Zn²⁺, and without EDTA pretreatment was given a relative ATPase activity of 100%. The values are the average of at least three measurements and the error bars represent one standard deviation. Scan from the phosphoimager (inset figure) shows the raw data with the release of ³²Pi being separated by TLC. Lanes 1 and 2 are negative (using heat-denatured helicase) and positive controls (acid-catalyzed hydrolysis) respectively. Lanes 3-17 correspond to the numbers indicated on the graph.

Functional studies on SARS-CoV helicase

Figure 3.5 Graphs showing the effect of the addition of poly(A) (200 µg mL⁻¹), Zn²⁺ (3 mM), dsDNA/dsRNA (10 pmol substrate 4/8, Table 3.2) on A308V mutant SARS-CoV nsp13. The activity of the mutant without any additive was set as 100%. The values are the average of at least three measurements and the error bars represent one standard deviation. Scan from the phosphoimager (inset figure) shows the raw data with the release of ³²Pi being separated by TLC. Lanes 1 and 2 are negative (using heat-denatured helicase) and positive controls (acid-catalyzed hydrolysis) respectively. Lanes 3-11 correspond to the numbers indicated on the graph.

Functional studies on SARS-CoV helicase

Figure 3.6 Graphs showing the effect of the addition of polyribonucleotides [poly(A), poly(U), poly(G), poly(C) (50 or 200 µg mL⁻¹)], divalent cations [Zn²⁺ or Mn²⁺ (3 mM)], ssDNA (45 mer, 10 pmol oligo 5, Table 8.1) or dsRNA (10 pmol substrate 4, Table 3.2), with and without EDTA pre-treatment (1 mM EDTA treating protein for 15 min prior to reaction) on A308V SARS-CoV nsp13. The activity of the mutant without any additive was set as 100%. The values are the average of at least three measurements and the error bars represent one standard deviation. Scan from the phosphoimager (inset figure) shows the raw data with the release of ³²Pi being separated by TLC. Lanes 1 and 2 are negative (using heat-denatured helicase) and positive controls (acid-catalyzed hydrolysis) respectively. Lanes 3-19 correspond to the numbers indicated on the graph.

3.3 Unwinding activity

Unwinding activity is another important function of helicases. Experiments demonstrated that separation of the DNA duplexes could not be observed on a 20% PAGE gel using either ethidium bromide (EB) staining or silver staining. Compared to the ^{32}P -based radioassay, about a 20-40 times higher concentration of substrate and enzyme was needed in the assay using EB staining. Therefore, it was considered that EB staining was a poor choice for the DNA and RNA unwinding analysis. It is believed that silver staining is 2-5 times more sensitive than ethidium bromide for DNA and RNA. However, it also has similar difficulty in making the DNA and RNA substrates as that encountered for EB staining. Two single-stranded nucleic acids were annealed to obtain double-stranded substrates. Normally, except for the target substrate, the remaining ss-DNA/RNA also exists in solution, even though a purification step is performed, which affects the unwinding results. This problem can be avoided using ^{32}P to label one ss-DNA/RNA. When using ^{32}P labelling, one single strand is labelled and then another complementary single strand with a much higher concentration (at least 10 times) is added to make a duplex substrate with high efficiency. This is followed by removing the excess complementary single strand with a BioSpin 30 column. A small amount of remaining complementary single strand should not affect the unwinding result on the gel. The trapping strand is then added. Thus, it is more efficient in making the substrates and gives a much clearer result from the unwinding assays. According to our data, ^{32}P labelling is at least 5-10 times more sensitive than silver staining. Therefore, for the unwinding activity studies, ^{32}P labelling appears the best choice. In order to investigate unwinding activity of the proteins, DNA and RNA substrates were utilized in the study.

3.3.1 DNA substrates

Five different DNA substrates were designed in order to determine the polarity of SARS-CoV helicase (Table 3.2, substrates 1-5). The reactions of three substrates containing a 5' flanking strand of single-stranded DNA (substrates 3, 4, and 5) showed separation of the duplex. The other substrates (1 and 2) which contained a 3' tail or no tail gave negative results. Both the negative controls (absence of ATP and the use of an inactive enzyme that had been heated to 95 °C for 10 minutes) gave negative results.

Three full-length helicases with a C-terminal 6×His tag, wild type, mutant A308V and mutant S289G/R390K, were used in the unwinding assay with three DNA substrates (2, 4 and 5, Table 3.2). The three different DNA duplexes: single-stranded DNA (30 mer) plus

Functional studies on SARS-CoV helicase

single-stranded DNA (30 mer), 5'-3' substrate (30 mer ss-DNA annealed with 15 mer ss-DNA to give a 15 base 5'-tail) and 3'-5' substrate (30 mer ss-DNA annealed with 15 mer ss-DNA to give a 15 base 3'-tail) were used (Figures 3.7, 3.8 and 3.9). Positive results were obtained from the former two substrates. Moreover, two different DNA substrates (1 and 3, Table 3.2) were also used to examine the unwinding activity of S289G/R390K in order to confirm the unidirectionality of unwinding (Figure 3.10). Δ 1-259 nsp13 was also used in the unwinding assay. However, only 5'-3' DNA and RNA and 3'-5' DNA substrates were used. Similar positive results from 5'-3' substrates and negative results from 3'-5' substrate were obtained with the wild type and A308V mutant protein (Figure 3.11).

All the nsp13 constructs demonstrated similar unwinding activity except for A308V and the catalytic domain (Δ 1-259). The unwinding activity of the A308V mutant was apparently slightly higher than the wild-type enzyme, but this difference may not be significant. The truncated enzyme (Δ 1-259) was found to retain about half of the unwinding activity.

No.	double-stranded oligonucleotide substrate		trapping strand	positive control	denatured helicase	
1	5'-CTGT CCTGCATGATG 3'-GACAGGACGTACTAC*	3	+	-	-	-
	5'-CTGT CCTGCATGATGTCGGATCGCAGTCAG 3'-GACAGGACGTACTAC*	1	+	-	-	2
	5'-GACTGACGCTAGGCTCTGT CCTGCATGATGTCGGATCGCAGTCAG 3'-GACAGGACGTACTAC*	1	+	+	-	3
	5'-GACTGACGCTAGGCTCATCATGCAGGACAG 3'-GTAGTACGTCCTGTC*	2	+	+	-	4
	5'-CTGT CCTGCATGATGTCG GATCGCAGTCAG 3'-GACAGGACGTACTACTCG GATCGCAGTCAG*	1	+	+	-	5
	5'-GUGUCCUGCAUGAUGUCGGAUCGCAGUCAG 3'-CACAGGACGUACUAC	4	+	-	-	6
	5'-GACUGACGCUAGGCUCAUCAU GCAGGACAC 3'-GUAGUACGUCCUGUG	5	+	+	-	7
	5'-GACUGACGCUAGGCUCAUCAUGCAGGACAC 3'-CUAGUACGUCCUGUG*	-	+	+	-	8

Table 3.2 Substrates used in analysis: substrates 1-5 and 8 are DNA/RNA with the asterisk indicating the position of the 32 P label. Substrates 6 and 7 are RNA with **G** indicating the nucleotide that is potentially 32 P-labelled. The trapping strands used are as follows: **1**, 5' – CTG TCC TGC ATG ATG; **2**, 5' – CAT CAT GCA GGA CAG; **3**, 5' – CTG TCC TGC ATG ATG TCG GAT CGC AGT CAG; **4**, 5' – GUG UCC UGC AUG AUG; **5**, 5' – CAU CAU GCA GGA CAC.

Functional studies on SARS-CoV helicase

Figure 3.7 DNA unwinding assay: SDS-PAGE gels scanned using the phosphorimager demonstrate that positive results were obtained from the ss30+ss30 DNA substrate (No.5, Table 3.2) using **A**: wild type, **B**: A308V mutant and **C**: S289G/R390K mutant nsp13 respectively. On the right is an illustration of the assay with the red asterisks indicating a ^{32}P -labelled strand. Substrate, ss 30mer, 5 min, 20 min and 30min represent the ss30+ss30 DNA substrate, the labelled ss-30 (30 mer single strand) and different reaction times (5, 20 or 30 minutes) respectively. Three different controls were performed: a positive control (+ve) by heating the reaction solution without helicase at 95 °C for 10 minutes, a negative control (-ve heated enzyme) by using heat-denatured helicase and another negative control (-ve no ATP) by using active helicase in the absence of ATP.

Figure 3.8 DNA unwinding assay: SDS-PAGE gels scanned using the phosphorimager demonstrate that positive results were obtained from the 5'-3' DNA substrate (No.4, Table 3.2) using **A**: wild type, **B**: A308V mutant and **C**: S289G/R390K mutant nsp13. On the right is an illustration of the assay with the red asterisks indicating a

Functional studies on SARS-CoV helicase

Figure 3.9 DNA unwinding assay: SDS-PAGE gels scanned using the phosphoimager demonstrate that negative results were obtained from the 3'-5' DNA substrate (No.2, Table 3.2) using **A**: wild type, **B**: A308V mutant and **C**: S289G/R390K mutant nsp13 respectively. On the right is an illustration of the assay with the red asterisks indicating the ^{32}P -labelled strand. Substrate, ss 15mer, 5 min, 20 min and 30 min represent the 3'-5' DNA substrate, the labelled ss-15 (15 mer single strand) and different reaction times (5, 20 or 30 minutes) respectively. Three different controls were performed: a positive control (+ve) by heating the reaction solution without helicase at 95 °C for 10 minutes, a negative control (-ve heated enzyme) by using heat-denatured helicase and another negative control (-ve no ATP) by using active helicase in the absence of ATP.

Functional studies on SARS-CoV helicase

Figure 3.10 DNA unwinding assay: SDS-PAGE gels scanned using the phosphoimager demonstrate that (A) a positive result was obtained from the ss15+ss45 DNA substrate (No.3, Table 3.2), and (B) a negative result was obtained from the double-stranded 15 bp DNA substrate (No.1, Table 3.2) using S289G/R390K mutant nsp13. On the right are illustrations of the assay with the red asterisks indicating the ^{32}P -labelled strand. ss 15mer, 5 min, 20 min and 30 min represent the labelled ss-15 (15 mer single strand) and different reaction times (5, 20 or 30 minutes) respectively. Substrate in gel A and B represent the ss15+ss45 DNA and the double-stranded 15 bp DNA substrates respectively. Three different controls were performed: a positive control (+ve) by heating the reaction solution without helicase at 95 °C for 10 minutes, a negative control (-ve heated enzyme) by using heat-denatured helicase and another negative control (-ve no ATP) by using active helicase in the absence of ATP.

Functional studies on SARS-CoV helicase

Figure 3.11 DNA and RNA unwinding assay: SDS-PAGE gels scanned using the phosphorimager demonstrate that positive results were obtained from (A) the 5'-3' DNA substrate (No.4, Table 3.2) and (B) the 5'-3' RNA substrate (No.8, Table 3.2), and that a negative result was obtained from (C) the 3'-5' DNA substrate (No.2, Table 3.2) using refolded $\Delta 1-259$ nsp13 with a C-terminal 6 $\times$ His tag. On the right are illustrations of the assay with the red asterisks indicating ^{32}P -labelled strand. ss 15mer, 5 min, 20 min and 30min represent the labelled ss-15 (15 mer single strand) and different reaction times (5, 20 or 30 minutes) respectively. Substrate in gel A, B and C represent the 5'-3' DNA, the 5'-3' RNA and 3'-5' DNA substrates respectively. Three different controls were performed: a positive control (+ve) by heating the reaction solution without helicase at 95 °C for 10 minutes, a negative control (-ve heated enzyme) by using heat-denatured helicase and another negative control (-ve no ATP) by using active helicase in the absence of ATP.

Functional studies on SARS-CoV helicase

3.3.2 RNA substrates

In vitro transcription was performed initially to obtain two RNA substrates (substrate 6 and 7, Table 3.2). However, the analysis was complicated by degradation due to RNase contamination or uncompleted transcription (truncated RNA products). Therefore a lower transcription incubation temperature (30 °C) and an extended incubation time of 1 hour were used to reduce uncompleted transcription and obtain better RNA substrates. For the two different RNA substrates, only the one which contained a 5' tail of single-stranded RNA (substrate 7) showed positive separation of the duplex (Figure 3.12). Substrate 6 with a 3' tail gave a negative result (Figure 3.12). Free RNA oligonucleotides were observed on the bottom of the gels. The S289G/R390K mutant nsp13 was used to study the unwinding of the two RNA substrates.

In order to improve the quality of unwinding results, synthesized RNA oligos were then used in the study. However, only one substrate with a 5' tail of single-stranded RNA (substrate 8) was used with this method due to the cost of RNA oligonucleotide synthesis. A positive result was obtained with this substrate using three proteins (wild-type, A308V, and S289G/R390K), and no free RNA oligonucleotides were observed at the bottom of the gel (Figure 3.13). This positive result was clearer than that obtained from RNA substrates synthesized via *in vitro* transcription. Furthermore, it was found that there is a decrease in unwinding activity after adding 3 mM Zn²⁺ to the reaction solution. The catalytic domain helicase (Δ 1-259) also gave positive result on the 5'-3' RNA substrate (Figure 3.11). All the unwinding results concerning the DNA and RNA substrates are displayed in the Table 3.3.

Proteins	DNA substrates				RNA substrate	
	5'-3' (4)	3'-5' (2)	ss15+ss45 (3)	ss30+ss30 (5)	DS15 (1)	5'-3' (8)
W.T.	Positive	Negative	n.d.	Positive	n.d.	Positive
Δ 1-259	Positive	Negative	n.d.	n.d.	n.d.	Positive
A308V	Positive	Negative	n.d.	Positive	n.d.	Positive
S289G/R390K	Positive	Negative	Positive	Positive	Negative	Positive

Table 3.3 Summary of the results from the helicase unwinding assays: Δ 1-259 represents refolded catalytic domain wild type protein; WT indicates the wild type protein. n.d. means not determined. The numbers in parentheses are same as those of substrates in Table 3.2.

Functional studies on SARS-CoV helicase

Figure 3.12 RNA unwinding assay: SDS-PAGE gels scanned using the phosphorimager demonstrate that (A) a positive result was obtained from the 5'-3' RNA substrate (No.7, Table 3.2), and (B) a negative result was obtained from the 3'-5' RNA substrate (No.6, Table 3.2) using S289G/R390K mutant nsp13. The two substrates were obtained through *in vitro* transcription. On the right are illustrations of the assay with the red asterisks indicating the ^{32}P -labelled strand. ss 15mer, 5 min, 20 min and 30 min represent the labelled ss-15 (15 mer single strand) and different reaction times (5, 20 or 30 minutes) respectively. Substrate 1 in gel A is the solution obtained by mixing the labelled ss-15 and an unlabelled 30 mer strand (ss-30), heating at 60 °C for 10 minutes and cooling slowly. Substrate 2 in gel A is the solution obtained by mixing the labelled ss-15 and an unlabelled 30 mer strand (ss-30), heating at 95 °C for 10 minutes and cooling slowly. Substrate in gel B represents the 3'-5' RNA substrate. Three different controls were performed: a positive control (+ve) by heating the reaction solution without helicase at 95 °C for 10 minutes, a negative control (-ve heated enzyme) by using heat-denatured helicase and another negative control (-ve no ATP) by using active helicase in the absence of ATP. The heavy bands at the bottom of the gels are due to release of free RNA oligos.

Functional studies on SARS-CoV helicase

Figure 3.13 RNA unwinding assay: SDS-PAGE gels scanned using the phosphoimager demonstrate that positive results were obtained from the 5'-3' RNA substrate (No.8, Table 3.2) using wild type (**A**), A308V (**B**) and S289G/R390K (**C**) nsp13 respectively. On the right is an illustration of the assay with the red asterisks indicating the ^{32}P -labelled strand. Substrate, 5 min, 20 min and 30 min represent the 5'-3' RNA substrate and different reaction times (5, 20 or 30 minutes) respectively. Zn^{2+} represents 3 mM Zn^{2+} added into the reaction solution (20 min). A drop in unwinding activity was observed in the presence of Zn^{2+} . Three different controls were performed: a positive control (+ve) by heating the reaction solution without helicase at 95 °C for 10 minutes, a negative control (-ve heated enzyme) by using heat-denatured helicase and another negative control (-ve no ATP) by using active helicase in the absence of ATP.

3.4 Discussion

The ATPase activity and nucleotide duplex unwinding activity was weaker than that of other SF1 viral helicases (e.g. HCoV-299E helicase) observed by some researchers¹¹⁸. Three groups have reported analysis of the ATPase activity of SARS-CoV helicase^{61,64,65}. The ability of the helicase purified in this study to hydrolyse ATP is consistent with that reported by Thiel's group⁶¹, where only [γ -³²P]-ATP was added to the reaction solution. Another group demonstrated the enzyme had a stronger ATPase activity, but the malachite green assay was used in the study, not a ³²P-based assay⁶⁴.

3.4.1 Site-directed mutagenesis

From the results of our study, it can be seen that the mutation in motif I (S289G) did not significantly affect the ATPase activity of nsp13. Studies on vaccinia virus nucleoside triphosphate phosphohydrolase I (SF2), RuvB, a superfamily 3 helicase, and two hexameric helicases, DnaB and Rep68 demonstrated that mutation of the threonine in motif I that occupies a similar position to Ser289 to an alanine almost completely abolished ATPase and duplex unwinding activity^{39,47,119,120}. The observation that S289G has similar ATPase activity to the wild-type enzyme suggests that, in the case of the SARS-CoV helicase, the conserved aspartate residue in motif II is sufficient for magnesium binding. This is in agreement with the study on RuvB, where the motif I threonine mutant (T69A) retains the ability to bind ATP and Mg²⁺¹¹⁹. However, explanation of the precise role of Ser-289 must await determination of the three-dimensional structure of the SARS-CoV helicase.

The increase in ATPase activity upon mutation of Ala308 in motif Ia was relatively large considering the relatively conservative change from alanine to valine. If the region of motif Ia is involved in linking nucleotide binding and ATP hydrolysis via conformational changes as previously postulated¹²¹, then this might provide an explanation for the increase in ATP hydrolysis observed for this mutant. Mutation of Phe64 in PcrA (which occupies a similar position in motif Ia to Ala308 in SARS-CoV nsp13) to an alanine demonstrated a large reduction in helicase activity (25-fold), a slight reduction in ssDNA binding (4-fold) and little change in the rate of ATP hydrolysis¹²². This suggested that a more bulky or more hydrophobic residue at this position in motif Ia is favourable for helicase activity and may explain the slight increase in unwinding activity noted for the A308V mutant.

The mutations D466E and L322S/D466E which exist in naturally-occurring variants of nsp13 had little effect on the ATPase activity suggesting that they are phenotypically silent. However, as a caveat, experiments on EAV nsp10 and human coronavirus 229E nsp13 have suggested that there is no simple correlation between the virus phenotype in tissue culture and the enzymatic activities of the helicases measured *in vitro*¹²³.

3.4.2 Effect of polyribonucleotides

The presence of poly(A) (a long polyribonucleotide, M_w : 2×10^5 - 1×10^6) resulted in a ~ 50% reduction in ATPase activity for both wild-type and Δ 1-259 nsp13, suggesting that the inhibitory activity was not due to binding of poly(A) to the zinc finger domain. This was further confirmed by EDTA pre-treatment which did not affect the poly(A)-mediated reduction in activity. Previous reports on coronavirus helicases using either short (24-mer) single-stranded ribonucleotides¹²⁴ or long polyribonucleotides¹²⁵ have suggested a substantial stimulation (between 5-50 fold) of ATPase activity in their presence. An exception was poly(G) which caused only a minor increase in activity. This discrepancy in the effect of poly(A) and poly(U) may be due to differences in the assay conditions or protein constructs.

Studies on HCV NS3 helicase (SF2) have shown either a small (~ 2.5-fold) stimulation [poly(A), poly(U)] or inhibition (~ 4-fold) [poly(G)] of ATPase activity¹²⁶. However, all three [poly(A), poly(G) and poly(U)] substantially inhibited RNA helicase activity. In addition, in the case of helicases from turnip yellow mosaic virus and Semliki forest virus only a slightly increase of ATPase activity was observed by adding them^{127,128}. It was suggested that the polynucleotides modulated ATPase activity by occupying the substrate binding site of the helicase and altering the affinity of the enzyme for ATP. With the A308V mutant protein a similar decrease in activity was noted showing that this mutation in motif Ia did not greatly affect the observed inhibitory effect. Further experiments with different polyribonucleotides conducted with the A308V mutant protein gave similar reductions in activity suggesting that the observed inhibition was sequence-independent.

3.4.3 Role of the zinc finger domain

The Δ 1-259 mutant with the zinc finger and hinge/spacer domains deleted retained half of the activity of the wild-type enzyme. The wild-type and A308V mutant enzyme pretreated with EDTA (magnesium was added in the assay) displayed activity reduced to ~ 50% and 30% of the original respectively. However, EDTA pretreatment of Δ 1-259 did not significantly affect

the activity suggesting that the presence of the zinc finger domain is required for the observed reduction in activity.

Previous experiments have shown that mutations of the conserved cysteine and histidine residues in the N-terminal metal binding domain (MBD) greatly affected the ATPase activity of two SF1 nidovirus helicases, equine arteritis virus (EAV) nsp10 and human coronavirus 229E nsp13^{123,129}. Deletion of the EAV nsp10 MBD abolished enzymatic activity. Similar mutations in the conserved cysteine residues in the C-terminal zinc finger domain of *E. coli* RecQ helicase (SF2) resulted in a reduction in zinc binding and complete loss of DNA-binding ability¹³⁰. However, demetalation of the RecQ helicase by EDTA treatment only slightly reduced the ATPase activity implying that the zinc ion is necessary for proper formation of secondary structure elements during protein folding, but once folded, the structure remains stable even in the absence of zinc. Similar results were found for EAV nsp10; denaturation (using urea) and renaturation in the presence of EDTA abolished ATPase activity, whereas the presence of zinc acetate during renaturation preserved activity¹²³. The relatively small (2-fold) decrease in activity of the EDTA-treated wild-type and A308V SARS-CoV helicase is in agreement with this result.

A comparison of the sequences of SARS-CoV nsp13, IBV (avian infectious bronchitis virus) helicase, EAV nsp10, PcrA and Rep (Figure 1.5, introduction chapter) shows that the catalytic domains of the nidovirus helicases (SARS, IBV and EAV) consist solely of two RecA-like subdomains. PcrA and Rep both have two additional subdomains (1B and 2B) which are thought to contribute towards nucleotide binding. Certain viral RNA helicases, for example hepatitis C, has a third domain in addition to the two core RecA domains (Figure 1.8, introduction chapter). It has been proposed that these accessory domains enhance the strand separation functionality via duplex destabilisation¹³¹, and that the zinc finger domain of the nidovirus helicases possesses a similar function and may play an important role in DNA binding¹³⁰. However, deletion of subdomain 2B of Rep did not appear to affect its duplex unwinding activity leading to the suggestion that this subdomain may be involved in regulation and/or interactions with other proteins, rather than being necessary for helicase activity¹³². A similar proposal has also been made for the accessory domain of hepatitis C helicase¹³³. The essentially preserved unwinding and ATPase activity of Δ 1-259 nsp13 suggests, at least for small DNA and RNA substrates, a non-essential role for the zinc finger

domain in catalysis. It is possible that this domain may aid processivity for longer nucleotide substrates and/or play a role in protein:protein interactions.

3.4.4 Roles of divalent cations

Treating the enzyme with EDTA is likely to remove many of the metal ions from the enzyme solution. The objective is to remove Zn^{2+} from the enzyme solution using EDTA pre-treatment (1 mM) for 15 min. The enzyme was then desalted and buffer-exchanged using a BioSpin-30 column. ATPase activity was measured using similar conditions as that used for the non-pretreated enzyme, with the addition of 3 mM Mg^{2+} to replace that removed by the EDTA.

It is believed that the helicase binds to MgATP/MgADP , and therefore, the NTPase activity is dependent onto the addition of Mg^{2+} or Mn^{2+} , which is a cofactor for ATPase activity¹⁸. The roles of Mg^{2+} or Mn^{2+} in the ATPase activities of helicases have been studied in detail¹⁰¹. Therefore, the removal of Mg^{2+} from the assay solution would be likely to decrease the ATPase activity. However, in our studies, Mg^{2+} is still added to the reaction solution, so ATPase activity can still be observed with EDTA pretreated helicase.

It was observed that in addition to Mg^{2+} and Mn^{2+} , Zn^{2+} , Co^{2+} and Ni^{2+} can also support the ATPase activity of HCV NS3 helicase, but their effect was lower than that of Mg^{2+} and Mn^{2+} ¹³⁴. Furthermore, Cu^{2+} and Hg^{2+} were found to inhibit the ATPase activity. Therefore, it is likely that when Zn^{2+} was added to the reaction solution in our experiments, it would compete with Mg^{2+} , and therefore decrease the ATPase activity. The results from replacing Mg^{2+} by Zn^{2+} (3 mM) plus Mg^{2+} (3 mM), and Mg^{2+} (30 mM) plus Zn^{2+} (3 mM) indicate the importance of Mg^{2+} in ATPase activity, Zn^{2+} having a possible role of inhibiting the helicase ATPase activity by a competitive mechanism.

3.4.5 DNA and RNA unwinding assays

Positive results were obtained with three different DNA substrates with a 5' tail and negative results with two different DNA substrates with a 3' tail or no tail. This demonstrated that SARS-CoV nsp13 has 5'-3' unidirectionality. Unwinding activity was not observed in the absence of ATP, as has been reported previously for HCV NS3 helicase⁷¹. This appears to make a passive mechanism unlikely. The observed 5'-3' polarity of SARS-CoV helicase agrees with a recently published result⁶⁴. The positive results from the RNA 5'-3' substrates

Functional studies on SARS-CoV helicase

and negative result from the RNA 3'-5' substrate proved the 5'-3' unidirectionality of SARS helicase and shows that SARS helicase can unwind both DNA and RNA nucleotides duplexes. From the data obtained on two different substrates, SARS helicase unwinds DNA substrates slightly better than RNA substrates. Furthermore, the data from three different substrates with a 5' flanked strand show that the 5'-3' substrate gave a similar result as that of the substrate formed from two single-stranded DNA (15 mer oligo and 45 mer oligo), the substrate formed from two single-stranded DNA (two 30 mer oligos) gave the least unwinding activity. This data indicates that the presence of a 3' flanked strand might affect the unwinding activity, especially when it is located in the same end as a 5' flanked strand. It was interesting that better unwinding results were obtained from longer duplex substrates (with similar duplex region length but different tail lengths) than from short duplex substrates^{61,64,65}.

For the unwinding activity of SARS-CoV helicase, a similar degree of DNA duplex unwinding was observed compared to that of a previously reported MBP fusion protein⁶¹. The only difference is that substrates with shorter length were used in our studies. However, two other constructs of the SARS-CoV helicase, a different MBP-helicase (using longer length DNA substrates, 200 bp) and helicase with N-terminal 6×His tag (using shorter length DNA substrates, 15 bp)^{64,65}, were observed to have higher activity than that of the nsp13 proteins used in this study.

Chapter 4 Crystallization and preliminary X-ray analysis

4.1 Crystallization of helicases

4.1.1 Crystallization of helicases

Several different expressed proteins were employed for crystallization screening. The C-terminal 6×His-tagged full-length nsp13 (construct 2) was initially used in the crystallization screens. Two days after setting up the initial crystallization trays (sparse-matrix: Hampton CrystalScreen 1 and 2) in the presence and absence of ADPNP and Mg^{2+} , very small crystals were observed with 8% PEG 8000 and 0.1 M Tris pH 8.5 in the absence of ADPNP and Mg^{2+} . Different PEG 8000 concentrations (5%-20%) at different pH values (5.0-9.0) were then grid screened. After two days, formation of small rectangular-shaped crystals with 24% PEG 8000 and 0.1 M HEPES pH 7.5 were observed. After further screening of different PEG solutions (PEG 200-20,000), it was found that 0.1 M sodium citrate pH 6.0, 50 mM KH_2PO_4 , 0.1 M ammonium acetate and 30% PEG 4000, gave several rhombus-like crystals (very weak diffraction). Knife-like crystals with 0.1 M sodium citrate pH 6.0, 50 mM KH_2PO_4 and 20% PEG 8000 were also observed after ten days (Figure 4.1). One crystal from this drop was then mounted on the X-ray diffractometer and low resolution diffraction data were obtained (5-6 Å).

Microcrystals were also observed from Hampton condition 21 (0.2 M magnesium acetate, 0.1 M sodium cacodylate pH 6.5, 30% v/v MPD) in the absence of ADPNP and Mg^{2+} with the full-length nsp13 (construct 6) with its N-terminal GST tag cleaved by TEV protease. After screening around this condition, small crystals, which gave no apparent X-ray diffraction, were observed from 0.2 M magnesium acetate, 0.1 M Bis-Tris pH 6.75, 40% v/v MPD (Figure 4.2). Further screening yielded no improvement in crystal size.

In addition, three other proteins, refolded Δ 1-259 nsp13 with a C-terminal 6×His tag (construct 4), full-length nsp13 with an intact (*i.e.* uncleaved) N-terminal GST tag (construct 6) and Δ 1-259 nsp13 with an intact N-terminal MBP tag (construct 10) were also used for crystallization screening in the presence and absence of ADPNP and Mg^{2+} , but despite extensive trials no protein crystals were obtained. Crystals were observed when screening the MBP-tagged Δ 1-259 protein in a large number of different conditions, but after examination by X-ray diffraction, all these crystals were determined to be salt rather than protein.

Crystallization and preliminary X-ray analysis

Figure 4.1 Knife-like crystals were observed using well solution: 0.1 M sodium citrate pH 6.0, 50 mM KH_2PO_4 and 20% PEG 8000 from SARS-CoV helicase with a C-terminal 6×His tag.

Figure 4.2 Small crystals, which gave no apparent X-ray diffraction, were observed using well solution: 0.2 M magnesium acetate, 0.1 M Bis-Tris pH 6.75, 40% v/v MPD from helicase protein after cleaving tag from SARS-CoV helicase with an N-terminal GST tag.

Crystallization and preliminary X-ray analysis

4.1.2 Seeding

In order to improve the quality of the crystals, various seeding techniques were used. During aggregation of protein molecules, there is equilibrium between the formation of ordered nuclei and the formation of precipitate. With the addition of seeds, the equilibrium might shift towards crystal formation and avoid the random nature of spontaneous nucleation¹³⁴. Microseeding, macroseeding, and streak seeding are three common seeding methods used to improve crystal quality and size. For microseeding, a few small crystals or one large crystal was broken up to powder using a needle. A small amount (usually 1 μ L) of mother liquor containing the powder was added to an aliquot of well solution, making serial dilutions: 1:10, 1:100, 1:1000, *etc.*, and vortexed well. For macroseeding, a single crystal, free from twinning or other crystalline deformations, was selected, washed with well solution several times, and then transferred to a fresh pre-equilibrated drop. A suitable fibre (cat's whisker) was used in streak seeding. The whisker was stroked through a microcrystalline precipitate, and then stroked in a straight line through a fresh drop quickly. Among the three seeding methods, microseeding is the one with the highest success rate and the macroseeding is the most difficult technique to be used successfully¹³⁵.

All three seeding methods were performed on the knife-like crystals, but good results were only obtained using microseeding. Macroseeding, streak seeding and additive screens were not successful in improving crystal quality. The seeds from the knife-like crystals were used to conduct microseeding with the well solution (0.1 M sodium citrate pH 6.0, 50 mM KH_2PO_4 , 0.1 M ammonium acetate and 30% PEG 4000). Crushed crystals diluted 100 or 1000 times were used as the seeding solution. Crystals with better morphology and larger dimensions were obtained with 1000 times dilution seeding (Figure 4.3) than with 100 times dilution seeding (Figure 4.4).

Crystallization and preliminary X-ray analysis

Figure 4.3 Crystals (helicase with a C-terminal 6×His tag) were obtained using 1000× dilution microseeding from the following well solution: 0.1 M sodium citrate pH 6.0, 50 mM KH_2PO_4 , 0.1 M NH_4Ac and 30% PEG 4000.

Figure 4.4 Crystals (helicase with a C-terminal 6×His tag) were obtained using 100× dilution microseeding from the following well solution solution: 0.1 M sodium citrate pH 6.0, 50 mM KH_2PO_4 , 0.1 M NH_4Ac and 30% PEG 4000.

Crystallization and preliminary X-ray analysis

4.1.3 The use of fusion proteins in crystallization

Fusion proteins were also used in an attempt to obtain crystals. The principal objective of using fusion protein constructs (*e.g.* 6×His, GST and MBP tags) is to make expression and/or purification easier. However, the presence of a fusion tag is often considered to be not optimal for crystallization, because of the inherent flexibility of the linker (often containing a protease cleavage sequence) between the tag and the target protein. Therefore, usually the larger tags (*e.g.* GST and MBP) are proteolytically cleaved to provide non-tagged proteins for crystallization trials. Sometimes, however, cleavage by proteases can introduce undesired heterogeneity into the protein sample. Furthermore, efficient cleavage of the linker by proteases is sometimes difficult. One advantage of retaining a large tag on the fusion protein is that the tag could potentially help solve the phase problem (see the next section) through molecular replacement. It must be noted, however, that very few structures have been obtained with an intact GST or MBP tag. This is no doubt due to the flexibility as discussed previously. In our case, both intact and cleaved fusion proteins were used for crystal screening.

4.2 Heavy atom derivatization

4.2.1 Phase problem

The phase problem arises in determining the electron density function of a crystal from X-ray diffraction data, namely that a complete determination requires knowledge of both the magnitudes and phases of the structure factors, but experimental measurements yield only the magnitudes. When X-ray photon waves are diffracted by a crystal, they give rise to diffraction spots. Each diffraction spot corresponds to a point in the reciprocal lattice and represents a wave with amplitude and a relative phase with respect to structure factor. When the diffraction spots on the detector are integrated, only the intensity can be obtained. The phase information is missing, and without this information, the atomic positions cannot be calculated. This is the phase problem, and a large part of crystallography is devoted to solving this problem.

Four different methods can be utilized to solve the phase problem^{136,137}. The first one is the direct method. The programs used in this method include SHELXD, RUNITAN, *etc.* If a crystal is made up of similarly-shaped atoms that all have positive electron density, then there are statistical relationships between sets of structure factors. These statistical relationships can be used to deduce possible values for the phases. Direct methods exploit such relationships,

Crystallization and preliminary X-ray analysis

and can be used to solve small molecule structures easily. Unfortunately, the statistical relationships become weaker as the number of atoms increases, and direct methods are limited to structures with, at most, a few hundred atoms in the unit cell. Although there are developments that push these limits, particularly for crystals that diffract to very high resolution (1.2Å or better), direct methods are not generally applicable to the vast majority of macromolecular crystal structures. However, they do become useful in the context of experimental phasing methods, such as isomorphous replacement and anomalous dispersion, as discussed below.

Molecular replacement is another method used in phase determination. If a good model for a reasonably large fraction of the structure in the crystal is available, this method would be preferred. Generally, the level of resemblance of two protein structures correlates well with the level of sequence identity. As a rule, molecular replacement has a reasonable probability of success if the model is fairly complete and shares at least 40% sequence identity with the unknown structure. It becomes progressively more difficult as the model becomes less complete or shares less sequence identity. Many programs are available, including AMoRe, Molrep, or Phaser. The third and fourth methods of solving phase problem are isomorphous replacement and anomalous scattering.

4.2.2 Heavy atom soaking

In isomorphous replacement, the idea is to make a change to the crystal that will perturb the structure factors and, by the way that they are perturbed, to make some deductions about possible phase values. It is necessary to be able to explain the change to the crystal with only a few parameters, which means that heavy atoms (heavy in the sense that they have a large atomic number, *i.e.* many electrons) need to be used. The introduction of a heavy atom will change the scattered intensity significantly. One reason for this is that "heavy" atoms contribute disproportionately to the overall intensity. On the other hand, all of the electrons in a heavy atom will scatter essentially in phase with one another. Because of this effect, different atoms contribute to the scattered intensity in proportion to the square of the number of electrons they contain.

If two crystals are available, one containing just the protein (native crystal) and one containing in addition bound heavy atoms (derivative crystal), diffraction data from both can be measured. The differences in scattered intensities will largely reflect the scattering

Crystallization and preliminary X-ray analysis

contribution of the heavy atoms, and these differences can be used (for instance) to compute a Patterson map. Once the locations of the heavy atoms in the crystal are determined, then their contribution to the structure factors can be calculated. This allows making some deductions about possible values for the protein phase angles, as follows. If the heavy atom doesn't change the rest of the structure, then the structure factor for the derivative crystal (F_{PH}) is equal to the sum of the protein structure factor (F_P) and the heavy atom structure factor (F_H). However, a special diffraction criterion is required for a promising heavy atom derivative. The derivative crystals must be *isomorphic* with native crystals. At the molecular level, this means that the heavy atom must not disturb crystal packing or the conformation of the protein. Unit cell dimensions are quite sensitive to such disturbances, so heavy atom derivatives whose unit cell dimensions are the same as native crystals are probably isomorphous. The term *isomorphous replacement* comes from this criterion. In the most common technique, crystals of the protein are soaked in solutions of heavy ions. In many cases, such ions bind to one or a few specific sites on the protein without perturbing its conformation or crystal packing. For instance, surface cysteine residues react readily with Hg^{2+} ions and cysteine, histidine, and methionine can form stable Pt adducts when adding Pt compounds. Normally, data from two derivatives are the minimal requirement for providing enough phase information.

Many heavy atom compounds ($KAuCl_4$, $SmCl_3$, $HgCl_2$, K_2PtCl_4 , K_2IrCl_4 , *etc.*) at different concentrations (final concentration: 1 mM, 2 mM, 5mM, *etc.*) were used to soak native crystals. Different soaking times were also tried. The crystals were picked out, placed in a solution formed by mixing the mother liquor and heavy atom compound stock solution and transferred into a cryoprotection solution which contained the heavy atom compound. An alternative method was to increase the concentration of the target heavy atom compound stepwise. Normally, the crystals were kept in the drop, a small amount of compound solution was added, and then same amount of liquid was taken out from the drop in a certain time (usually 10 min). The two steps were repeated until the final concentration was near to the target concentration.

4.2.3 Selenomethionine-substituted protein

X-rays set the electrons in the atoms vibrating, and these vibrations generate radiation of the same frequency, which is propagated in all directions. This is the coherent scattering that gives rise to diffraction effects. Normally, the electrons vibrate in step with the incident beam.

Crystallization and preliminary X-ray analysis

If, however, the incident photons have energy close to a transition energy which can bring the atom to an excited state, the electronic vibration gets out of step. Instead of re-radiating in phase with the incident beam, the radiated energy has a different phase. Also, the intensity of coherent scattering is reduced, because some energy is absorbed to bring about the transition. This effect (shift in amplitude and phase) is called anomalous scattering.

Through this shift, positions of the atoms that contribute to the diffraction pattern with phases could be determined. The anomalous scattering effect depends on the frequency of oscillation being similar to the natural frequency for the atom. So clearly the strength of the anomalous scattering effect depends on the wavelength of the X-rays, which will change both the normal scattering and the out-of-phase scattering of the anomalous scatterers. By collecting data at several wavelengths near the absorption edge of an element in the crystal, phase information analogous to that obtained from MIR can be obtained. This technique is called MAD, for multiple-wavelength anomalous dispersion. Furthermore, another technique, SAD (single-wavelength anomalous dispersion), also utilizes anomalous scattering. However, unlike MAD, SAD only uses measurements at one wavelength. It requires the presence of a heavy atom which often has to be added to the native, but an alternative is to use sulphur which is likely to be present in most proteins. Normally, SAD needs high quality data and good resolution.

One popular way to use MAD is to introduce selenomethionine in the place of methionine residues in a protein. The selenium atoms (which replace the sulphur atoms) have a strong anomalous signal at wavelengths that can be obtained from synchrotron X-ray sources. From the sequence of SARS-CoV nsp13, it can be found the frequency of methionine (9 residues, 1.5% of the total) might have a possibility of providing enough phase from Se anomalous scatterers. In this study, selenomethionine-substituted proteins were expressed using the method of inhibition of the methionine biosynthesis pathway¹³⁸. It does not require an auxotrophic strain and is based on the blocking of methionine biosynthesis by inhibiting aspartokinases in the presence of a high concentration of isoleucine, lysine and threonine. The nonauxotrophic strain is grown in a medium containing a high concentration of amino acids known to inhibit methionine biosynthesis and without methionine (replaced by selenomethionine). Induction by IPTG and harvesting followed the same protocol as for the unsubstituted protein. During purification and crystallization, 1 mM DTT or 10 mM β -ME was added to all the solutions.

Crystallization and preliminary X-ray analysis

In this study, the selenomethionine-substituted proteins were expressed and examined with mass spectrometry. However, the quality of the obtained mass spectra was relatively poor, making quantitation of the selenium substitution difficult. These proteins were then utilized to conduct crystallization screening. However, no crystals from the selenomethionine-substituted proteins could be obtained despite extensive trials. However, soaking heavy atoms into protein crystals (section 4.2.2) can also give rise to anomalous scattering at specific wavelengths (the X-ray absorption wavelength of the metal). Therefore, the derivative crystals were used to collect MAD and SAD data in this study.

4.3 Data collection and processing

Datasets of crystals from the full-length helicase with a C-terminal 6×His tag were collected from both the in-house X-ray diffractometer and a synchrotron and then processed using MOSFLM and other CCP4 suite programs.

4.3.1 Data collected from the in-house X-ray diffractometer

Crystals of the full-length helicase with a C-terminal 6×His tag [from the condition (0.1 M sodium citrate pH 6.0, 50 mM KH₂PO₄, 0.1 M NH₄Ac and 30% PEG 4000) after 1000× dilution microseeding, Figure 4.3] were picked out with a nylon loop and placed into a cryoprotection solution (30% glycerol + 70% well solution). The crystals were then mounted on the X-ray diffractometer at a low temperature (-160 °C or -170 °C) using cold dry nitrogen gas. The crystal to detector distance was normally set between 100 to 200 mm. Initial images were analyzed by MOSFLM to provide information about crystal symmetry and orientation, which are necessary for the design of a data collection strategy. A strategy command in MOSFLM was utilized in order to collect a complete dataset. The software determines the minimal number of degrees that need to be covered to get a full dataset (based on the lattice symmetry or the spacegroup symmetry), and what value of ϕ to start at. The X-ray diffraction data were then processed using MOSFLM and SCALA (CCP4 suite¹³⁹).

Initial autoindexing using MOSFLM suggested that 21 possible trigonal or hexagonal spacegroups in the P3/P6 “family” could fit the data. In order to determine the correct spacegroup, the dataset was scaled in the spacegroups P3 and P6, representing the lowest symmetry trigonal and hexagonal spacegroups. The dataset did not scale well in spacegroup P6, giving an initial R_{merge} before outlier rejection of approximately 41%. The data scaled well

Crystallization and preliminary X-ray analysis

in spacegroup P3 giving an initial R_{merge} before outlier rejection of approximately 21%. Thus, it was obvious that the correct spacegroup should be trigonal. To further determine the spacegroup, systematic absences were checked when processing the datasets. Spacegroups with centred lattices (F, C, and I centred) and translational symmetry elements, either screw axes or glide planes, will have systematic absences - diffraction intensities are zero at certain special conditions. It was found that systematic absences ($00l: l=3n$) existed when datasets were processed as $P3_1$ and $P3_2$ spacegroups, which indicated spacegroups P3, $P321$ and $P312$ could not be possible (there are no systematic absences in the three spacegroups). Therefore, the remaining possible trigonal spacegroups were $P3_1$, $P3_2$, $P3_212$, $P3_221$, $P3_112$ and $P3_121$. Among them, $P3_1$ or $P3_2$ gave the best R_{merge} value.

However, of the 65 chiral spacegroups (no inversion point, mirror planes, or improper axes of rotation as symmetry elements), 11 pairs of spacegroups, including $P3_1$ and $P3_2$, are enantiomorphic spacegroups. An enantiomorphic spacegroup refers to either of a pair of spacegroups that are mirror images of each other but are not identical. The exact spacegroup cannot be determined from the enantiomorphic spacegroup pair ($P3_1$ and $P3_2$) until model building gives additional information. Finally, a possible spacegroup, $P3_1/P3_2$ (trigonal) with the unit cell $a=b=141.37 \text{ \AA}$, $c=185.53 \text{ \AA}$, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$, was suggested by the current 3-5 $\text{\AA}$ resolution data (see Appendices).

After indexing as $P3_1/P3_2$, scaling and merging, the TRUNCATE twinning statistics (moment plots and intensity distribution) indicated that the data may be twinned (Figure 4.5). The twin problem would happen when two or more crystals of the same material inter-grown so that the unit cell of the first is related to the unit cell of the second by a symmetry element. If two structures that are different have grown together, the assembly is not a twin. There are two classes of twin¹⁴⁰. One is the merohedral twins, which means that the lattices of all twin components coincide perfectly in both direct and reciprocal space. In the merohedral twins, the centre of symmetry can always be chosen as the operator. Another is the pseudo-merohedral twins indicating the centre of symmetry is never the twin operator. In this case, the twin belonged to the merohedral twins.

Repeating the processing steps using a lower symmetry spacegroup P3 and other low ranking spacegroups ($P3_112$, and $P3_212$) also gave rise to the similar statistics which suggested a potential twinning problem. This excluded the possibility of incorrect spacegroup assignment

Crystallization and preliminary X-ray analysis

as the cause. Therefore a detwinning procedure was conducted to obtain a detwinned MTZ file. First the MTZ file was examined using the Twinning Server (<http://nihserver.mbi.ucla.edu/Twinning>) to obtain the twin fraction (0.15), and then the structure factors were detwinned using DETWIN (CCP4 suite). All the following processing steps discussed in this chapter were performed using both MTZ files (non-detwinned and detwinned) to compare the differences.

Figure 4.5 Graph of cumulative intensity distribution plotting the percentage of acentric and centric reflections [where Z (height) is less than 0.1, 0.2, up to 1.0] from TRUNCATE output log file using $P3_1$. It can be observed that the acentric observed curve was below the the acentric theoretical curve. The twin fraction is 0.15. The red, black, gray, and green curve represent the theoretical curve from centric reflections, the theoretical curve from acentric reflections, the observed curve from acentric reflections, and the observed curve from centric reflections (there are no centric reflections in this spacegroup) respectively.

In order to explore how many protein molecules were in the asymmetric unit, the Matthews coefficient (V_M) was calculated. Based on the unit cell information, the volume of the unit cell was $3219734 \text{ \AA}^3$. There are three equivalent positions in the spacegroup $P3_1/P3_2$, so Z should be a multiple of 3. In this case, V_M would then equal 16.02 ($V_M = \text{Volume of unit cell} / M_w \times Z$) assuming only a monomer in the asymmetric unit. Thus, there is actually most likely to be a hexamer in the asymmetric unit, which would give $V_M = 2.67$ (V_M is usually between 2.0 to 2.5).

Crystallization and preliminary X-ray analysis

Datasets were also collected from crystals mounted in a capillary at room temperature (Figure 4.6) in order to examine the effect of freezing conditions and cryo-protectant solutions. The room temperature and cryogenic datasets were of similar quality and resolution suggesting that the freezing conditions were optimal.

Figure 4.6 A crystal (full-length nsp13 with a C-terminal 6×His tag) obtained from 0.1 M sodium citrate pH 6.0, 50 mM KH_2PO_4 , 0.1 M NH_4Ac and 30% PEG 4000 was mounted in a quartz capillary (0.2 mm diameter).

Processing of the derivative data collected with the home X-ray diffractometer was performed using the same methodology as that used for native crystal data. After processing, the datasets were used to create difference Patterson maps using FFT. The positions of heavy atoms (Au, Hg, Ir and Sm) bound to amino acid residues (cysteine or histidine) were defined from the maps or the program RSPS. CAD was used to combine the MTZ files, which was followed by SCALEIT to scale the combined files. The combined, scaled structure factors files were used in MLPHARE to refine the heavy atom positions and obtain initial phases. However, only poor quality data (5-7 Å resolution, low isomorphous difference) were obtained.

4.3.2 Synchrotron data

In order to improve resolution, native data at higher resolution (2.8 Å) were collected from a synchrotron (Swiss Light Source, SLS) as only low resolution data could be obtained from the home X-ray machine. The home X-ray diffractometer only produces X-rays at a wavelength

Crystallization and preliminary X-ray analysis

of 1.5418 Å ($\text{CuK}\alpha$), thus SAD and MAD data from derivatives (KAuCl_4 , SmCl_3 , HgCl_2 , K_2PtCl_4 , K_2IrCl_4 , *etc.*) can not be collected using the diffractometer. The crystals soaked with different heavy atoms were mounted on the beamline and scanned using a fluorescence detector. Among them, good fluorescence data were obtained from crystals soaked with K_2IrCl_4 , or SmCl_3 (Figures 4.7 and 4.8). Therefore, data from the two derivatives were collected at peak ($\Delta f''$ maximum) and inflection ($\Delta f'$ minimum). However, no remote wavelength data were collected due to radiation damage. Each derivative crystal was firstly tried to collect MAD data. However, for most of derivatives, due to radiation damage, only one dataset could be collected and therefore SAD had to be used instead of MAD (see Appendices). In addition, many MIR data sets were also collected from the same synchrotron beamline. The processing of synchrotron data was performed in the same way as data collected from the home X-ray diffractometer.

Figure 4.7 Inflection (edge) and peak wavelengths were obtained from crystal (full-length nsp13 with a C-terminal 6×His tag) soaked with SmCl_3 using fluorescence scanning. The x, left y, and right y axes refer to the values of energy of radiation, fluorescence and anomalous scattering factors respectively. The black, blue and red curves represent scan data, $\Delta f'$ minimum value, and $\Delta f''$ maximum value respectively.

Crystallization and preliminary X-ray analysis

Figure 4.8 Inflection (edge) and peak wavelengths were obtained from a crystal (full-length nsp13 with a C-terminal 6×His tag) soaked with K_2IrCl_4 using fluorescence scanning. The x, left y, and right y axes refer to the values of energy of radiation, fluorescence and anomalous scattering factors respectively. The black, blue and red curves represent scan data, $\Delta f'$ minimum value, and $\Delta f''$ maximum value respectively.

4.4 Phase determination

4.4.1 MIR

It was found that the frequencies of cysteine (26 residues, 4.3% of the total) and histidine (12 residues, 2.0% of the total) in the protein sequence (full-length nsp13 with a C-terminal 6×His tag containing 608 residues) would provide many binding sites for heavy atoms. Therefore, MIR, in theory, can be used to solve the phase problem. Data at about 5-7 Å were collected from different heavy atoms [gold compounds: $KAuCl_4$ and $KAuBr_4$, mercury compounds: $Hg(OOCCH_3)_2$, $HgCl_2$, and CH_3HgCl , platinum compounds: K_2PtCl_4 , K_2PtCl_6 , and $PtCl_2(H_2NCH_2CH_2CH_2NH_2)$, samarium compound ($SmCl_3$), iridium compounds (K_2IrCl_4 and $IrCl_3$), iodide compound (KI), rhenium compounds (K_2ReCl_6 , $KReO_4$), silver compound ($AgNO_3$)]. After initial processing using MOSFLM, SCALA and TRUNCATE, the output MTZ files were used to produce isomorphous difference Patterson maps (for example, Figure 4.9). However, after finding the heavy atoms (RSPS) and fixing heavy atom positions (MLPHARE), no convincing locations were found. Refinement and examination of heavy

Crystallization and preliminary X-ray analysis

atom positions was performed using MLPHARE. The PhP_a (phasing_power_acentric) value was too low (e.g. 0.1-0.3, should be over 0.5), CullR_a (Cullis_R_acentric) value was too high (e.g. 0.9-1.5, should be below 1.0), and the overall figure of merit (FOM) was too low (e.g. 0.1-0.25, should be over 0.4, but this is resolution-dependent). Only very few positions were obtained from KAuCl_4 , HgCl_2 , K_2IrCl_4 , and SmCl_3 derivatives that might be correct. However, considering that asymmetric unit being likely to contain six molecules, the possible slight twinning problem, and the low resolution, these putative heavy atom positions could not provide enough phase information. Other methods (SAD or MAD) or different derivatives might need to be utilized.

Figure 4.9 A difference Patterson map at Harker section (0.33) obtained from a crystal (helicase with a C-terminal 6×His tag) soaked with HgCl_2 using FFT is shown. The peaks in the figure give the possible positions of heavy atoms bound to the residues of protein, which can be used to confirm the possible heavy atom position and find the extra positions.

4.4.2 SAD and MAD

SAD and MAD datasets were also collected and processed. All the SAD and MAD data from different heavy atom derivatives were used to create anomalous difference Patterson maps to search possible positions of heavy atoms. Meanwhile, these data were also processed using SOLVE¹⁴¹ to find the convincing positions of heavy atoms. MLPHARE was then used to

Crystallization and preliminary X-ray analysis

refine these positions. Among them, SmCl_3 and K_2IrCl_4 gave better anomalous and isomorphous differences than other derivatives. Four possible heavy atom positions were found containing three from K_2IrCl_4 and one from SmCl_3 . However, after refinement, the overall figure of merit (FOM) was low suggesting either weak phasing or incorrect heavy atom positions. The data including peak and inflection wavelengths were also processed using SOLVE. Similar results were obtained as that from MAD data.

4.5 Discussion

For the crystallization of SARS-CoV nsp13, two constructs (full-length nsp13 with a C-terminal 6×His tag and cleaved nsp13 from the fusion protein with an N-terminal GST tag) gave crystals. However, despite extensive trials, the latter crystals could not be improved.

Based on the peptide mass fingerprinting data from the purified protein (full-length nsp13 with a C-terminal 6×His tag) mentioned in the previous chapter, it was possible that *E. coli* glucosamine-6-phosphate synthase was present as a contaminant. Therefore, the crystals were dissolved in double-distilled water and examined using mass spectrometry, but no convincing mass spectra could be obtained. Therefore, molecular replacement was utilized to exclude the possibility of the presence of the contaminant protein in the crystal. Four different programs were used, AMoRe, MolRep, EPMR¹⁴² and Phaser¹⁴³ using four models (PDB: 1JXA, full-length protein containing both isomerase and glutaminase domains, PDB: 1MOQ/1MOR, isomerase domain, and PDB: 1XFF, glutaminase domain). Top solutions from these models gave high R_{fac} values (over 0.55). Moreover, no α -helical structures could be determined from the electron density maps. The negative results from molecular replacement processing suggested that glucosamine 6-phosphate synthase was a minor contaminant and that the crystals were indeed of SARS-CoV nsp13. However, molecular replacement can fail even for a real case. Thus, this negative result does not rule out the crystals being glucosamine 6-phosphate synthase.

In summary, active SARS-CoV full-length nsp13 has been cloned, expressed, purified and crystallized. SARS-CoV helicase crystals have been shown to diffract to 2.8Å. MIR and SAD data (~5-7 Å) were collected, processed and used to obtain initial phase information. However, because of the poor quality of datasets, additional phasing information is required to solve the structure.

Chapter 5 Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

5.1 Serine proteases

Serine proteases are members of a large group of proteolytic enzymes, named for a nucleophilic serine residue present at the active site. They have been studied intensively over the past several decades, leading to the accumulation of high-quality data in terms of sequences, structures, and kinetic constants for their substrates and inhibitors. The key catalytic residues of a serine protease are Ser-195, His-57 and Asp-102 (the numbering is based on chymotrypsinogen¹⁴⁴). These three residues form a hydrogen bonding system often referred to as the “catalytic triad” or “charge relay” system¹⁴⁴. According to MEROPS¹⁴⁵ nomenclature, there are at least six different clans of serine proteases (peptidases) based on structural similarity, which are typified by chymotrypsin, subtilisin, carboxypeptidase Y, Clp protease, D-Ala-D-Ala peptidase, and LexA repressor¹⁴⁵. Serine proteases are involved in many physiological process, including digestion, hemostasis, apoptosis, signal transduction, reproduction, and immune response¹⁴⁶⁻¹⁵⁰. However, there are also potential hazards. If uncontrolled, they can destroy the protein components of cells and tissues¹⁵¹. It has been reported that serine proteases have a role in fibrinolysis, blood coagulation, complement fixation, and respiratory diseases¹⁵²⁻¹⁵⁵. Thus, understanding the mechanism of serine proteases and designing specific inhibitors for them is important.

5.1.1 Chymotrypsin family

All of the characterised peptidases of the chymotrypsin family are endopeptidases, which can also be called the peptidase family S1¹⁴⁵. There are three main activity types in this family based on their primary substrate specificity: trypsin-like proteases, which cleave substrates with positively-charged amino acid residues (Lys and Arg) at P₁, chymotrypsin-like, and elastase-like proteases. The latter two prefer substrates with aromatic or large hydrophobic amino acids (Phe, Tyr, Trp, Leu, Ile) at P₁ and small uncharged side chain amino acids (Ala, Gly, Val) at P₁ respectively¹⁵⁶. For a long time, this family has been used as a paradigm for investigating the serine protease mechanism. In our study, we chose to focus on porcine pancreatic elastase, secreted from the pancreas into the small intestine and is involved in the digestion of food.

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

Figure 5.1 Structure of porcine pancreatic elastase, showing the catalytic residues (Ser-195, His-57, Asp-102). 'N' and 'C' represent the N-terminus and C-terminus respectively. The structure contains two six-stranded β -barrel domains with the active site located in a cleft between the domains. The figure was created using PyMOL.

5.1.2 Structures of the chymotrypsin family of serine proteases

X-ray crystallographic studies of three serine proteases of the chymotrypsin family: chymotrypsin^{157,158}, trypsin¹⁵⁹ and elastase¹⁶⁰ were some of the earliest enzyme crystal structures obtained and have similar structural characteristics. The structure of pancreatic elastase shows that the 240 residues of the enzyme are arranged in two six-stranded β -barrel domains (Figure 5.1). The active site cleft (containing the catalytic triad) is located between the two β -barrels. However, the structure can also be divided into three functional domains: catalytic, substrate recognition and zymogen activation domain¹⁶¹. This organization is convenient for comparison among different proteases, but these three functional domains should not be considered separate, as they are intricately intertwined. Except the protease domain, auxiliary regulatory domains have also been found in many serine proteases¹⁶². High resolution structures of native porcine pancreatic elastase at 1.1 Å and porcine pancreatic

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

elastase with an inhibitor (N-acetyl-NPI) at 0.95 Å have been solved recently^{163,164}. They provide highly accurate models at atomic resolution to understand and determine the interaction between the enzyme and inhibitor, and to study conformational shifts in the elastase active site after inhibitor binding.

5.1.3 The catalytic triad

Ser-195, His-57, and Asp-102 are conserved in all serine proteases and are close in space (Figure 5.1). The first protease catalytic site was revealed from the structure of α -chymotrypsin¹⁶⁵. Then in 1969, the “charge-relay system” was named for a specific structure: N δ 1 of the imidazole side chain of His-57 is hydrogen-bonded to the carboxyl group of Asp-102, and a different nitrogen atom in the imidazole ring (N ϵ 2) is within hydrogen-bonding distance of the hydroxyl group of Ser-195^{166,167}. Electrons can be conducted through these hydrogen bonds from the carboxylate of Asp-102 through His-57 to Ser-195. This would make the serine hydroxyl group a better nucleophile and may reduce the charge on the carboxylate group of Asp-102. This system was identified in trypsin, elastase and other serine proteases^{160,168}. However, some serine proteases have novel catalytic triads or dyads: Ser-His-Glu, Ser-Lys/His, His-Ser-His, and N-terminal Ser¹⁶⁹. Furthermore, a similar catalytic triad of Asp-32, His-64, and Ser-221 was found in subtilisin^{170,171}.

The functional residues were identified in different ways. Ser-195 was discovered by using diisopropylfluorophosphate (DFP), which inactivates chymotrypsin, and covalently modifies the structure of Ser-195 forming a phosphate ester^{172,173}. The inactivation of chymotrypsin by N-tosyl-L-phenylalanine chloromethyl ketone (TPCK) through a stable covalent bond to His-57 led to the identification of the residue in the catalytic triad¹⁷⁴. But Asp-102 was identified only after a crystal structure had been determined. In 1969, after solving the crystal structure of chymotrypsin, Blow proposed the charge-relay system and identified Asp-102 as the third member of a catalytic triad, which activates Ser-195 for nucleophilic attack by acid-base catalysis¹⁶⁶.

5.1.4 The oxyanion hole

The oxyanion hole is composed of the backbone amide nitrogens of Ser-195 and Gly-193 (Figure 5.1), and was first proposed by Henderson¹⁷⁵. The oxyanion hole of subtilisin is different, it contains one backbone amide nitrogen of Ser-221 and the side chains of Asn-155 and Thr-220¹⁷⁶. These oxyanion holes can provide hydrogen bonds to negatively-charged

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

intermediates; this feature may be required for transition state stabilization. During the serine protease reaction, the oxyanion of the tetrahedral intermediate occupies this region and forms two hydrogen bonds with the two amide nitrogen atoms of Gly-193 and Ser-195. It may also polarise the carbonyl of the ester bond of the acyl-enzyme complex¹⁷⁷.

5.1.5 The zymogen activation domain

Serine proteases are normally synthesized as inactive precursors (zymogens), and then the zymogens are activated by proteolysis. In the chymotrypsin family, chymotrypsinogen and proelastase are activated by trypsin, trypsinogen needs be activated by enteropeptidase^{178,179}. The zymogen contains a 15 residue N-terminal extension. The cleavage of a peptide bond between Arg-15 (Arg/Lys in proelastase) and Ile-16 (Val-16 in proelastase) leads to several conformational changes: a salt bridge forms between Ile-16 and Asp-194, a substrate binding site is revealed which does not exist in the zymogen, and the oxyanion hole is created¹⁸⁰. During this process, several regions undergo conformational change. They are the N-terminus to residue 19, residues 142-152, 184-193 and 216-223 in the zymogens of chymotrypsin and trypsin, which is termed the zymogen activation domain¹⁸¹. Activation of proelastase resembles that of chymotrypsinogen and trypsinogen, which results in the removal of a small peptide (15 residues) from the N-terminus¹⁸².

5.1.6 The substrate recognition sites

The active site of the serine proteases is shaped as a cleft where the polypeptide substrate binds. Generally, the primary substrate binding site (S_1) is responsible for controlling the overall specificity of the enzyme (Figure 5.2). A substrate binding subsite is defined as that portion of the enzyme which interacts with a single residue of the peptide substrate¹⁸³. The subsites are numbered, S_1 , S_2 , *etc.* proceeding toward the N-terminus from the scissile peptide bond (acyl group side) and S'_1 , S'_2 toward the C-terminus from the scissile peptide bond (leaving group side). The corresponding residues of the substrate are designated P_1 , P'_1 , *etc.*^{183,184}. The cleavage is catalyzed between P_1 and P'_1 .

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

Figure 5.2 Schematic diagram of the subsites in a typical serine protease.

Serine proteases can be classified by their primary substrate specificity. P_1 interacts with S_1 , so there should be specific structure in the S_1 position. Actually, the S_1 site is a pocket near Ser-195 (Figure 5.3), formed by residues 189-192, 214-216, 224-228. Residues 189, 216, and 226 are the most important in determining specific binding preference^{185,186}. For example, the S_1 pocket of chymotrypsin is composed of Ser-189, Gly-216, and Gly-226, which form a deep hydrophobic pocket. The feature explains why the S_1 site of chymotrypsin prefers a hydrophobic P_1 residue¹⁸⁷. Asp-189, Gly-216 and Gly-226 of the trypsin S_1 site forms a negatively-charged pocket. This determines trypsin's specificity for positively-charged substrates (Arg or Lys) at P_1 ^{168,188}. The S_1 pocket of elastase, comprised of Thr-216 and Val-226, is smaller than the S_1 sites of trypsin and chymotrypsin, which allows elastase to choose substrates with small uncharged side chains at P_1 ¹⁶⁰. An arginine (Arg-226) interacting with acidic P_1 residues was reported to be responsible for the specificity of granzyme B¹⁸⁹.

Figure 5.3 Schematic diagram showing the specificity (S_1) pockets of chymotrypsin, trypsin, and elastase which locate the side chain of the residue on the C-terminal side of the scissile peptide bond (the P_1 residue). Chymotrypsin prefers aromatic side chains whereas trypsin prefers positively-charged side chains (e.g. Arg or Lys). Elastase has an occluded pocket and can only accept small uncharged side chains (e.g. Ala or Val or Gly). Red dots indicate oxygen atoms and the blue dot represents a nitrogen atom. Figure was adapted from reference¹⁹⁰.

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

Kinetics and X-ray crystallography are two useful methods to study the substrate specificity of an enzyme. Enzyme kinetics makes it possible to establish the catalytic importance of the individual amino acid residues in a polypeptide substrate. Data on P and P' residues can be obtained using this technique. X-ray crystallography provides invaluable information on substrate binding, which can not be obtained from kinetic analysis. A large number of kinetic studies on serine proteases have been conducted over the last three decades. Mostly inhibitors such as boronic acids and trifluoromethyl ketones have been used in X-ray crystallographic studies on serine proteases as substrate analogues. It was found that productive binding of a peptide to the serine protease would result in the formation of an additional anti-parallel β -strand. Between the acyl group side of the substrate and a serine protease, three important hydrogen bonds exist between the amide nitrogen of P₁ and the backbone carbonyl oxygen of S₁ (Ser-214), between the amide nitrogen of P₃ and the carbonyl oxygen of S₃ (Gly/Val-216), and between the carbonyl oxygen of P₃ and the amide nitrogen of S₃ (Gly/Val-216)¹⁹¹. In addition to these three hydrogen bonds, hydrophobic interactions at subsites S₁, S₂ and S₄ can also be observed¹⁹².

Crystallographic studies on serine proteases bound to natural protein inhibitors show that the S'₃ subsite is the limit of the binding cleft in the leaving group direction^{193,194}. For the acyl group side, S₄ appears to be the end of the binding cleft¹⁸³. Furthermore, the location of the substrate-binding sites beyond S₃ and S'₂ was found not to be limited to main-chain interactions. Thus, it was suggested that the substrate residue binding regions beyond S₃ and S'₂ should not be named sites but areas¹⁹⁵. In chymotrypsin, the S₂ site is a shallow hydrophobic groove formed by His-57, Trp-215 and Leu-99. The S₃ site is bounded by Val/Gly-216 and Phe-192. From the structure of α -lytic protease, the S'₁ is located on the enzyme surface which is lined by Ser-41 and His-57¹⁹⁶ and Leu-41 and Thr-142 form a cleft (S'₂) to interact with P₂¹⁹⁷. Furthermore, the kinetic data of α -lytic protease showed that the active site extends over at least six subsites (S₄-S'₂). Much more extended subsites from the structure of bovine trypsinogen and porcine secretory trypsin inhibitor (Kazal) shows the loop of the inhibitor displays a conformation that involves residues P₉, P₇, P₆, P₅, P₃, P₁, P'₁, P'₂ and P'₃. However, it should be noted that the binding site region of trypsinogen may be different from that of active trypsin^{193,198}.

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

Figure 5.4 Catalytic mechanism of the serine proteases: Step 1, enzyme acylation: nucleophilic attack of Ser-195 hydroxyl group onto the substrate peptide bond; first tetrahedral intermediate formation and collapse to form the acyl-enzyme intermediate and the release of the C-terminal part of the substrate polypeptide. Step 2, enzyme deacylation: nucleophilic attack of the hydrolytic water molecule onto the acyl-enzyme intermediate; second tetrahedral intermediate formation, collapse and release of the N-terminal part of the substrate polypeptide.

5.2 Catalytic mechanism of serine proteases

5.2.1 *The basic mechanism and intermediates*

This is one of the best-studied enzyme mechanisms, but new details are still being added. Figure 5.4 shows the generally accepted catalytic mechanism of the chymotrypsin family of serine proteases. The catalysis process can be divided into two main steps: enzyme acylation and enzyme deacylation. The mechanism reaction starts from the binding of the substrate according to specificity of enzyme. This is the formation of the enzyme-substrate complex, which is also called the Michaelis complex¹⁷⁵. The Michaelis complex was proposed as a possible feature of all enzyme reaction in order to explain the phenomenon of saturation kinetics (when the substrate concentration is much greater than the enzyme concentration). Subsequent kinetic studies have proven the reality of the Michaelis complex¹⁹⁹. No crystal structure of a serine protease with a native peptide substrate has yet been reported. After substrate binding to the specificity pocket of the serine protease, Ser-195 is activated by general base catalysis. His-57 polarizes the hydroxyl group of Ser-195, which enhances the nucleophilicity of Ser-195. The negatively-charged Asp-102 promotes the formation of positively-charged His-57.

The alkoxide ion of Ser-195 attacks the peptide carbonyl carbon, forming a tetrahedral intermediate, which is stabilized by the oxyanion hole. The tetrahedral intermediate collapses, the peptide bond of substrate is broken and the amine product (C-terminal part of the substrate) is released. His-57 now acts as a general acid, donating a proton to the leaving group nitrogen of the substrate so that it becomes a neutral amine (NH₂-) group and the C-terminal half of the substrate is free to leave. The enzyme is now covalently attached to the N-terminal part of the substrate as an ester. This is called the acyl-enzyme intermediate. The first structure of an acyl-enzyme intermediate with a native peptide was solved in 1997²⁰⁰.

When the C-terminal part of the peptide substrate leaves the active site, a water molecule enters and hydrogen bonds to His-57. The putative hydrolytic water molecule was observed by X-ray crystallographic experiments in 1993 and 1997^{200,201}. His-57 acts as a general base again to polarise the water molecule, which enhances the nucleophilicity of the water molecule. This is similar to the earlier step with O γ of Ser-195. The nucleophilic oxygen atom of the water molecule then attacks the carbonyl carbon of the substrate forming the second tetrahedral intermediate. A putative second tetrahedral intermediate was observed by an X-ray crystallographic study in 2001²⁰². Just as with the first tetrahedral intermediate, the second is

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

also stabilised by the oxyanion hole. With the help of the negatively-charged Asp-102, a positively-charged His-57 is formed again. The protonated His-57 repeats its role as a general acid, and donates a proton back to Ser-195. At the same time, the covalent bond between the enzyme and substrate is broken, thus releasing the N-terminal part of the peptide substrate.

5.2.2 Catalytic triad and charge-relay system

The basic principle of the charge-relay mechanism is that electrons can be conducted through the catalytic triad via hydrogen bonds from the carboxylate of Asp-102 through His-57 to Ser-195. This kind of hydrogen bond may be a low-barrier hydrogen bond²⁰³. The exposed Nε2 of His-57 carries an excess negative charge, strengthening the hydrogen bond between Ser-195 and His-57, and allowing a hydroxyl proton to be transferred to His-57 from Ser-195²⁰⁴. A proton may also be transferred between Asp-102 and His-57²⁰⁵. In addition, a charge-relay system was also found in the less active precursor chymotrypsinogen, which demonstrated that the system is not enough to make an effective enzyme²⁰⁶.

Two variants of the serine protease mechanism are possible, involving transfer of either one or two protons (Figure 5.5). When the charge-relay system was proposed, the two proton transfer mechanism was considered most likely (to avoid what was thought to be an energetically unfavourable charge separation): protons are transferred from Ser-195 to His-57 and from His-57 to Asp-102^{207,208}. However, later ¹H NMR experiments found the peak representing the proton located between His-57 and Asp-102 did not shift, which indicated that perhaps only one proton was transferred²⁰⁹. Furthermore, other NMR experiments observed that the p*K_a* of His-57 was near to 6.7 in the free enzyme, acyl-enzyme intermediate, and an enzyme-inhibitor complex²¹⁰⁻²¹². They found that when His-57 is neutral or protonated, Nδ1 could act as proton donor to Asp-102; and when His-57 is neutral, Nε2 would act as proton acceptor from Ser-195. This experimental data seem to support a one proton transfer mechanism. However, the hydrogen bonds between Asp-102 and His-57 were suggested to be low barrier hydrogen bonds, which may support a two proton transfer mechanism²¹³⁻²¹⁵. More recently, though, a 0.83 Å structure of α-lytic protease revealed the hydrogen bond between His-57 and Asp-102 was likely to be a conventional, not a low barrier hydrogen bond²¹⁶. The general consensus that has emerged is that a one proton transfer mechanism is the more likely with His-57 becoming protonated at the tetrahedral intermediate steps (and therefore positively-charged) and Asp-102 remaining unprotonated (and therefore negatively-charged).

5.2.3 *The role of the oxyanion hole*

The oxyanion hole stabilises the two tetrahedral intermediates by two strong hydrogen bonds, a kind of electrostatic interaction. Substrate binding to the enzyme can release energy, which can be harnessed to distort bond angles or bond length of certain atoms in the substrate molecule away from ideal situation and hence induce strain²¹⁷. Meanwhile, the active site of the enzyme can also be strained. The distortion of the substrate molecule is frequently towards the transition state. In the serine protease mechanism, this strain results in conformational changes from trigonal to tetrahedral in carbonyl of the scissile amide bond upon nucleophilic addition. The amide bond (sp^2 hybridised) becomes tetrahedral (sp^3 hybridised). This conformational change helps the formation of two strong hydrogen bonds between the oxyanion hole and the negatively-charged oxyanion of the tetrahedral intermediate. X-ray crystal structures of transition state analogues have demonstrated this interaction^{177,188,218-220}. Two NMR results showed that the interaction between the oxyanion hole and the substrate would provide about 6 kcal/mol free energy to stabilize the tetrahedral intermediate^{211,221}. Thus, the hydrogen bonds between the tetrahedral intermediate and the oxyanion hole are critical in transition state stabilization. The oxyanion hole may also polarise the carbonyl group to improve its electrophilicity²²².

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

Figure 5.5 The proposed hydrolysis of the acyl-enzyme complex via two tetrahedral intermediates (**B** and **D**). The mechanism may either involve one proton transfer (**I**) or two proton transfers (**II**).

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

5.2.4 Mutagenesis of the catalytic triad and oxyanion hole

Mutagenesis is an important experimental tool used to examine enzyme mechanisms. In the past decades, many mutagenesis experiments have confirmed the roles of the three catalytic residues. Chemical modification was first used to disrupt the triad. In 1966 and 1970, Weiner and Nakagawa observed that the protease activity was at least 10^4 times lower than that of native protease by removing the hydroxyl of Ser-195 and methylating Nε2 of His-57^{223,224}. Another useful mutagenesis tool is site-directed mutagenesis. The replacement of Asp-102 by Asn 102 decreases trypsin activity 10^4 times^{225,226}. The result indicated the importance of Asp-102 in serine protease catalysis. The k_{cat}/K_m value drops by 10^5 when Ser-195 and His-57 are substituted by alanine²²⁷, thus confirming the indispensable nature of these two residues in the catalytic triad. The double mutant enzyme (H57A/S195A) possessed some catalytic activity indicating that hydrolysis of the amide was still possible, perhaps via an active water molecule²²⁸. This hydroxide activation was observed in both the subtilisin mutant and the trypsin mutant. The exact mechanism for this is still unclear^{225,228}.

In 1987, Carter proposed a novel catalytic mechanism, “substrate-assisted catalysis”, in the study of mutant subtilisin²²⁹. They used a H64A (His-64 is equivalent to His-57) mutant to hydrolyze three different substrates: sFAHF-*p*NA (a His P₂ substrate), sFAAF-*p*NA (an Ala P₂ substrate) and sFAQF-*p*NA (a Gln P₂ substrate) (s = succinyl, *p*NA = *p*-nitroanilide). The results showed that at pH 8.0, the mutant serine protease hydrolyzed the His P₂ substrate 400 times faster than the other two substrates. However, for wild-type native protease, there were similar catalytic efficiencies with the three substrates. Later experiments indicated that the P₁' position also possessed this kind of substrate-assisted catalysis²³⁰. Furthermore, in trypsin, histidine in P₂ or P₁' position is observed to increase the enzyme specificity, which may indicate substrate-assisted catalysis²³¹. According to this mechanism, it was believed that histidine in P₂ or P₁' can replace His-57 in the catalytic triad and interact with Ser-195 and Asp-102.

Experiments in which the oxyanion hole was mutated were only performed in subtilisin. The oxyanion hole of subtilisin is composed of the side chains of Asn-155 and Thr-220 and the backbone NH of Ser-221. The mutagenesis experiments focused on the two former residues. In 1986, Bryan *et al.* replaced Asn-155 with Leu. The mutant protease and wild type protease were used to cleave a substrate, sAAPF- *p*NA. The result showed that the k_{cat}/K_m value of the mutant decreased by 200-300 times, thus demonstrating the importance of Asn-155 in

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

catalysis²³². Later an N155A mutant indicated that O γ of Thr-220 is 4 Å away in the X-ray crystal structure, but it could be the third hydrogen bond donor. They suggested that Thr-220 might move during the catalytic process²³³. In 1991, Braxton *et al.* created four mutants of Thr-220 (T220S, T220C, T200V, and T220A) and found that the hydroxyl of Thr-220 plays an important role in stabilising the transition state through a long-range electrostatic interaction²³⁴.

5.2.5 Evidence for tetrahedral intermediates

Early experiments using kinetic studies on chymotrypsin-catalysed hydrolysis of peptidic *p*-methoxy and *p*-chloroanilide substrates²³⁵ and elastase- and α -lytic protease-catalysed hydrolysis of peptidic *p*-nitroanilide substrates²³⁶ provided evidence for a tetrahedral intermediate between the Michaelis complex and the acyl-enzyme intermediate. Later experiments using ¹³C NMR also supported the existence of tetrahedral intermediates²³⁷. Generally the tetrahedral intermediates exist for a too short a period of time to be observed in X-ray crystallographic studies. However, many crystal structures of serine proteases with inhibitors bound, for example, peptidyl aldehydes¹⁵¹, trifluoromethyl ketones²³⁸, chloromethyl ketones²³⁹, and boronic acids²⁴⁰ have been interpreted as analogues of tetrahedral intermediates. In these complexes His-57 is protonated, and the oxygen atom, deriving from the P₁ carboxyl group, is negatively-charged and forms two strong hydrogen bonds with the oxyanion hole. Although these structures can provide useful information on the stereochemistry of the tetrahedral intermediate, these inhibitors are not productively-bound peptides.

The crystal structure of γ -chymotrypsin showed that a peptide (an autolysis product, assigned as Pro-Gly-Ala-Tyr) had bound to the S subsites and formed an acyl-enzyme intermediate²⁴¹. When this experiment was repeated in hexane, rather than aqueous solution, the density around Ser-195 appeared tetrahedral and was modelled as such²⁴². This possible tetrahedral intermediate, although not repeatable, gave some indication of the structure of the first tetrahedral intermediate.

In 2001, a tetrahedral intermediate model with an unmodified peptide was obtained²⁰². When crystals containing the peptidyl acyl-enzyme intermediate underwent a pH jump and were frozen quickly again, electron density was observed that suggested a tetrahedral intermediate.

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

This structure provided the first indication of the stereochemistry of a genuine second tetrahedral intermediate.

5.3 Elastase and Porcine pancreatic elastase (PPE)

5.3.1 Elastase

Elastases are a group of proteases that have the ability to hydrolyze the peptide bonds of the important connective tissue protein elastin^{243,244}. Elastin possesses the unique property of elastic recoil. Elastic recoil is widely distributed in vertebrate tissue, and is particularly abundant in the skin arteries, ligament and lungs. Elastin is rich in amino acids with small side chains (Ala, Ser, Val) and poor in those with aromatic or basic chains²⁴⁵. Elastase was discovered in 1949²⁴⁶. Human neutrophil elastase and pancreatic elastase are two major serine proteases that cleave elastin. Neutrophil elastase is found in the polymorphonuclear leukocytes, which is also essential for phagocytosis and defence against infection by invading microorganisms²⁴⁷. Pancreatic elastase is stored as an inactive zymogen, proelastase, in the pancreas and is secreted into the intestines where it is activated by trypsin and then participates in digestion.

As already mentioned, elastases can be extremely destructive if not controlled, which results from their property of digesting many connective tissue proteins. The fatal disease pancreatitis, which occurs when pancreatic zymogens are activated and released into the circulation, is concerned with pancreatic elastase. Neutrophil elastase may be linked to several diseases: respiratory disease (cystic fibrosis, chronic obstructive pulmonary and emphysema)¹⁵⁵, pancreatitis²⁴⁸, rheumatoid arthritis²⁴⁹ and other inflammatory disorders²⁵⁰. Thus, research on selectively and effectively inhibiting elastase is of interest to the pharmaceutical industry. Porcine pancreatic elastase (PPE) has a molecular weight of 25,900 Da. The amino acid sequence with 240 residues and four disulfide bridges was reported in 1973²⁵¹. Shotton *et al.* crystallized PPE in 1968 and solved the X-ray crystal structure in 1970^{160,252}. More recently, two atomic level structures of native PPE at 1.1 Å and PPE with an inhibitor (N-acetyl-NPI) at 0.95 Å were solved^{163,164}.

5.3.2 The active sites and substrate specificity of elastase

It has been long-established that the active site of porcine elastase extends to at least seven subsites: five subsites from the acyl group side and at least two subsites from the leaving group side²⁵³. It has been suggested that the two subsites from the leaving group side play

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

indispensable roles in the binding of the substrate and the rate of hydrolysis of the substrate²⁵⁴. Later kinetic studies indicates that the substrate binding site of PPE extends over eight subsites (S_1 to S_5 and S'_1 to S'_3) which implies that the length of the active site is about 29 Å. It was suggested that residue P_4 was of little important for the specificity of porcine pancreatic elastase²⁵⁵. It was also shown that the increased effectiveness of elastase with larger peptides appeared to be mainly associated with the acylation step and to a lesser degree, the deacylation step.

The substrate specificity of human neutrophil elastase (HNE) (also known as human leukocyte elastase, HLE) and porcine pancreatic elastase (PPE) has been well-studied. The S_1 pockets of both elastases are rather hydrophobic and appear as a hemispheric shape. PPE appears to have a preference for an alanine residue, with valine and isoleucine residues only being marginally less favourable. Human neutrophil elastase, which has a larger S_1 subsite pocket, seems to prefer slighter larger residues with valine and methionine being especially favourable²⁵⁶.

The S_2 subsite for the two elatases is bowl-shaped and quite hydrophobic. His-57, Leu-99, and Phe-215 form the S_2 subsite of HLE, which is almost same as that of PPE except for Leu-99 being replaced by Val-99. Hydrophobic side chains of medium size including proline are preferred¹⁸⁴. Equally, the S_3 and S_4 subsites appeared to have a hydrophobic character but their specificity is not particularly marked. It was found that substrates with Lys at P_3 or P_4 are much less reactive than those with Ala²⁵⁷. The S_5 and S_6 subsites of PPE are basic and prefer acidic substrates, but those of HLE have low specificity and can accept hydrophobic, basic and polar-uncharged residues¹⁸⁴.

On the leaving group side of the scissile bond, the S'_1 subsite in HLE is composed of Cys 42, Cys-58 and Phe-41. It is a relatively hydrophobic pocket and can accommodate all amino acid residues at P'_1 subsite, except proline, large hydrophobic residues (e.g. Trp) and N-methylated residues¹⁸⁴. The S'_1 subsite was found to prefer hydrophobic groups in keeping with its strongly hydrophobic character. However good affinity was also noted for lysine residues suggesting the possible presence of an acidic group within the subsite. The S'_2 subsites of both are similarly hydrophobic. PPE has a marked preference for Ala, but HLE has some specificity for Leu at P'_1 ²⁵⁸. Later work suggested that an interaction in the S'_3 subsite might also play a role in elastase catalysis and it was found to have a high aromatic specificity. In

Introduction (II): serine proteases and porcine pancreatic elastase (PPE)

HLE this subsite has no such specificity²⁵⁹. It has also been indicated that the P₄' residue is not of significant importance in the selectivity of elastase²⁵⁵.

5.4 Purpose of this study

Though there have been several reported structures of tetrahedral intermediates from enzyme transition state analogue complexes, no crystal structure of the first tetrahedral intermediate for a native productive substrate has yet been obtained and therefore, an X-ray crystal structure of a true tetrahedral intermediate is required to support the mechanism. Structures of the acyl-enzyme intermediate and the second tetrahedral intermediate with native peptides have recently been obtained. Structures of the enzyme-substrate complex and the enzyme-product complex have not yet been reported. The purpose of this study is to obtain structures of the enzyme-substrate complex (using peptides with different lengths) combining kinetic and ESI-MS data, and the structure of the first tetrahedral intermediate with X-ray crystallography and hence obtain new information on the serine protease catalytic mechanism.

Chapter 6 Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

Proteases are capable of catalysing both the hydrolysis and the synthesis of peptide bonds²⁶⁰. The serine proteases of the chymotrypsin family are the most often used for peptide synthesis, which occurs *via* deacylation of the acyl-enzyme intermediate by water or added nucleophiles, peptides or amino acid derivatives with an unprotected group²⁶¹. When trypsin is used to catalyze the synthesis of peptides, it was found that hydrophobic or bulky amino acid residues were more efficient at the P₁' position, and L-proline and D-amino acid residues were the least²⁶². Furthermore, through studying chymotrypsin and trypsin in peptide synthesis, the ratio of aminolysis to hydrolysis was reported to increase linearly with the concentration of added nucleophile²⁶³. Interestingly, it has been reported that trypsin-catalyzed hydroxyaminolysis of amino acid ester substrates was faster at -23 °C than in a liquid aqueous solution at 1 °C²⁶⁴. Later experiments supported this result and showed that under cold conditions (~ -15 °C) aminolysis is preferred to hydrolysis^{265,266}. It is believed that in protease-catalyzed peptide synthesis, the competitive hydrolysis of the acyl-enzyme and product cleavage can be suppressed at a lower temperature²⁶⁷. The aim of this study, using HPLC and ESI-MS, was to ascertain whether porcine pancreatic elastase can catalyse *de novo* peptide bond synthesis between a dipeptide and human β -casomorphin-7 (YPFVEPI, BCM7) or a BCM7 analogue (AcNPI).

6.1 Peptide synthesis

The solid phase peptide synthesis (SPPS) was first introduced by Merrifield in 1963²⁶⁸. The use of Fmoc (9-fluorenylmethoxycarbonyl) as a protecting group is a common strategy. The procedure involved in Fmoc-based SPPS can be summarized as follows: a special resin providing a solid support and a linkage to the peptide chain, Fmoc deprotection by treatment with 20-50% piperidine in DMF for 20 minutes, coupling steps (adding more residues), monitoring the reaction using ninhydrin to test for the presence of free amino groups, and removal of the side-chain-protecting groups and the peptide from the resin with TFA²⁶⁸.

In this study, fourteen peptides (all with L-amino acids) were synthesized by solid phase peptide synthesis, twelve of these with a C-terminal carboxylate: N-acetyl-Asn-Pro-Ile-OH (AcNPI), NH₂-Ala-Ser-OH, NH₂-Ala-Val-OH, NH₂-Lys-Ala-OH, NH₂-Lys-Gly-OH,

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

NH₂-Lys-Ser-OH, NH₂-Lys-Val-OH, NH₂-Asn-Pro-Ile-OH, NH₂-Asp-Pro-Ile-OH, NH₂-Gln-Pro-Ile-OH, NH₂-Glu-Pro-Ile-OH, NH₂-Tyr-Pro-Phe-Val-Glu-Pro-Ile-OH (BCM7), were synthesized using the Wang resin. The peptides with a C-terminal amide, NH₂-Ala-Ala-NH₂ and NH₂-Lys-Ala-NH₂, were obtained using the Rink amide-MBHA-resin.

6.2 HPLC purification and analysis

In order to detect and purify the possible new peptide product in the solution, High Performance Liquid Chromatography (HPLC) was utilized, which is a popular method of analysis and widely used in biotechnological, biomedical and biochemical research²⁶⁹. The chemical separation is accomplished using HPLC with a particular column and mobile phase. The solid support within the column is the stationary phase in HPLC. Mobile phase refers to the solvent being continuously passed through the column. The migration and separation of components in the sample is determined by the chemical interactions between the mobile phase and the sample through the column. For normal phase HPLC, those samples having stronger interaction with the mobile phase than with the stationary phase will elute from the column faster, and have a shorter retention time. Reversed phase HPLC shows the opposite performance to that of normal phase HPLC. Reversed phase HPLC operates on the basis of hydrophilicity and lipophilicity. The stationary phase consists of a hydrophobic matrix. Therefore, hydrophilic compounds elute more quickly than hydrophobic compounds²⁷⁰. For the analysis and purification of peptides, reversed phase HPLC is the most common method.

6.2.1 HPLC purification

All the crude peptides were purified by reversed phase HPLC with a C18 column. The fractions containing the desired peptide were confirmed by ESI-MS and lyophilized. For the lyophilization of AcNPI, which is a fairly hydrophobic peptide, water was added to dilute the acetonitrile in the peptide solution. The freeze-drying step was repeated several times until an almost white powder was obtained. Although determination of the exact purity of the peptides would require further examination by NMR (structure) and amino acid analysis (peptide content), the peptide purity as determined by HPLC analysis with detection at 214 nm (where the peptide bonds absorb), was considered sufficient for the mass spectrometric and crystallographic experiments.

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

6.2.2 HPLC analysis

The objective was to show the possibility of peptide synthesis by PPE using reversed phase HPLC. PPE and two peptides in a molar ratio of 1:50:50 were dissolved in 25 mM Tris pH 7.5 (50 μ L), incubated at 25 °C or -15 °C for 2, 5, 10, 20, 25 min. At each time point, a 10 μ L aliquot was removed and added to acetone (30 μ L). After centrifugation, the supernatant was added to ddH₂O (360 μ L), and lyophilized. The powder was dissolved in ddH₂O (100 μ L), loaded onto an analytical C18 column, and eluted with a gradient of 0%-60% buffer B (90% acetonitrile + 0.04% TFA) in 40 min. All the peaks were collected and identified with ESI-MS. In the study, either BCM7 (YPFVEPI) or AcNP1 was used as one peptide and either NH₂-Ala-Ala-OH, NH₂-Ala-Gly-OH, or NH₂-Lys-Ser-OH was used as the other peptide. Three additional different molar ratios (1:25:25, 1:100:100 and 1:200:200) and three different pH values (pH 5, pH 6 and pH 8.5) were used in the study. However, in all the HPLC experiments, no peaks corresponding to a newly synthesized peptide could be observed.

Previously, successful peptide synthesis by trypsin and elastase at a low temperature (-15 °C or -23 °C) has been reported^{264,265}. However, sometimes the quantity of peptides was too low, which would result in ESI-MS being an ambiguous method in identification of the peptide. Precise determination of the identity of the enzyme-synthesized peptide would require NMR. The technique requires a substantially larger amount of material than ESI-MS. The exact reason for the lack of expected peptide synthesis by PPE is not clear. It is possible in our case that the ratios of protein to the peptides, the reaction temperature, pH value, or amount of the reaction products were not optimal.

6.3 Electrospray ionisation mass spectrometry

Electrospray ionisation mass spectrometry (ESI-MS) is a 'soft' ionisation process²⁷¹. The ESI-MS technique allows the accurate mass determination of high molecular weight molecules due to its ability to generate multiply-charged ions. These ions produce a series of peaks which lie within the m/z range of standard mass spectrometers. It is claimed that the typical routine accuracy obtainable for proteins under 30 kDa is in the order of 0.01%. The series of peaks can be processed using deconvolution software to give the mass of the target molecules.

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

In this study, ESI-MS was utilized to confirm the identity of fractions that contained the target peptides in the HPLC spectra and to detect the presence of new peptide products or protein:peptide complexes in the reaction solutions. All fourteen peptides were confirmed to have correct mass values. Some peaks are the products of peptides combined with Na^+ .

PPE was originally employed as a model enzyme because it was readily available in a highly purified form and gives a good electrospray mass spectrum²⁰⁰. In the case of peptide synthesis by a serine protease, both the enzyme-substrate complex and the first tetrahedral intermediate would have a mass equal to that of PPE and the two peptides minus a water molecule. Samples for the initial mass spectrometry experiments were prepared by incubating PPE, β -casomorphin-7, and $\text{NH}_2\text{-Lys-Ser-OH}$ or $\text{NH}_2\text{-Ala-Ala-OH}$ (at a molar ratio of 1:5:5) in MilliQ water at 25 °C for 5 or 20 minutes at different pH values, added to 20 μL solution I (98% acetonitrile, 2% formic acid) or 20 μL solution II (5 M ammonium acetate, pH 7.0). The mass spectra results of the proteins were deconvoluted by ProMass (ThermoFinnigan, USA). The aim was to observe the peak of synthesized new peptide directly or the peak of enzyme:substrate complex using deconvolution. However, although four different ratios (1:5:5, 1:10:10, 1:25:25, 1:50:50) of PPE:BCM7:dipeptide, two different dipeptides (KS or AA) and four different pH values (pH 5, 6, 7.5, 8) were used, only in one experiment (PPE:BCM7:KS at a molar ratio of 1:5:5 at 25 °C, pH 7.5 for 20 minutes) was a peak observed that could represent *de novo* peptide synthesis. This peak (1087.7 Da, Figure 6.1) was 8.6 Da higher than the calculated mass of the new nonapeptide (YPFVEPIKS, calculated mass: 1079.1 Da). This inaccuracy in the mass coupled with the relatively low signal means that this result does not provide conclusive evidence of peptide synthesis. Furthermore, only the PPE peak, and not the enzyme:substrate complex, was clearly observed after deconvolution (Figure 6.2).

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

Figure 6.1 Mass spectrum (ESI-MS in negative ion model) of the solution obtained by incubating PPE:BCM7:KS at a molar ratio of 1:5:5 at 25 °C and pH 7.5 for 20 minutes is displayed. The labelled peaks are the mass spectra of PPE (+18 A, for example, indicates the m/z of PPE with 18 negative charges). Calculated masses of BCM7 (YPFVEPI) and KS are 863.9 Da and 233.3 Da respectively. Thus, the calculated mass of the possible product (YPFVEPIKS) should be 1079.1 Da. The peak of BCM7 can be observed from the figure. It can be seen that there is only one significant peak (1087.7 Da) of higher mass than BCM7 which is not part of the PPE multiply-charged series (labelled with a red asterisk).

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

6.4 Kinetics

6.4.1 PPE activities at different pH values

For the crystallographic studies (Chapter 7), it was important to determine the activity of PPE at different pH values. Figure 6.3 shows the data of hydrolysis of substrate (0.2 mM Suc-Ala-Ala-Pro-Ala-pNA) at 25 °C obtained from twelve different buffer solutions with different pH values. It was found that the PPE has the highest activity at pH 8.5 with only a small amount of activity at pH 5.

Figure 6.3 The activity of PPE (0.5 µg) at different pH values with 0.2 mM substrate at 25 °C (measured at 405 nm). The maximum activity was obtained at pH 8.5.

6.4.2 IC_{50} of the peptides

In 1997, BCM7 (YPFVEPI) was found to be a weak competitive inhibitor of PPE ($K_i = 0.76$ mM) and to form a stable acyl-enzyme complex at pH 5.0 as revealed by a crystal structure²⁰⁰. The fact that BCM7 formed a relatively stable acyl-enzyme complex was unusual due to the presence of glutamic acid at the P_3 position. Previous work had indicated the preferential binding of small nonpolar residues in the S_1 - S_4 subsites of PPE¹⁸⁴. However, Meyer *et al.* did suggest the presence of Gln-192 in the active site which may allow the

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

introduction of polar groups into the S₃ pocket of PPE²⁷². The PPE:BCM7 structure indicated the presence of a hydrogen bond (3.2 Å) between the acidic P₃ residue, Glu-5 of BCM7, and the side chain of Gln-192 in PPE. Thus, it was proposed that this hydrogen bond may play a role in the increased stability of the PPE:BCM7 complex compared with those formed by other small peptides. To test this hypothesis, an analogue with alanine at P₃ position was examined using kinetics, and found to be a much weaker inhibitor of PPE ($K_i > 4$ mM)²⁰⁰.

To further investigate the role of the P₃ residue, kinetic studies on BCM7 analogues with a truncated N-terminus were performed²⁷³. In analyzing peptides with a 'reduced' number of residues at the N-terminus, a positively-charged amino terminus is necessarily introduced into the active site of PPE. To nullify a potentially deleterious ionic interaction, peptides were also synthesized with an N-acetyl group to provide a neutral N-terminus. It was found that similar significant inhibition of PPE could be obtained from 'reduced' tri-, tetra-, penta-, and hexapeptides. However the dipeptide, NH₂-Pro-Ile-COOH, demonstrated no inhibition of PPE. Based on these data, it was clear that the P₃ glutamic acid residue is critical in forming the stable acyl-enzyme complex.

In 2002, seven different N-acetylated tripeptides were utilized in a kinetic study to investigate different residues at P₃ position¹⁶⁴. It was observed that AcEPI, AcDPI and AcNPI had similar K_i and IC₅₀ values to BCM7. However, AcQPI and AcOPI (the structure of ornithine is similar to that of lysine but the side chain contains one less methylene group) demonstrated much stronger inhibition to PPE (K_i : 0.095 and 0.018 mM respectively). However, co-crystallization of PPE with these two tripeptides was unsuccessful. Of the seven N-acetylated tripeptides, only AcNPI (K_i : 0.89 mM) could be co-crystallized with PPE and a 0.95 Å structure was obtained¹⁶⁴. No kinetic or crystallographic study of the tripeptides without an N-acetyl group was done. Therefore, it was decided to examine the inhibition of PPE by four unacetylated tripeptides (NPI, EPI, DPI, QPI).

These four tripeptides were synthesized and their inhibitory activity against PPE (IC₅₀) were determined using duplicate to quadruplicate measurements of five different inhibitor concentrations (0.0 mM, 0.1 mM, 0.5mM, 1 mM, 4 mM) at 25 °C, pH 7.5 with 0.2 mM substrate in 2 min. Pre-incubation of inhibitors and PPE (5 min) was also performed. The IC₅₀ values of the four tripeptides, NPI, EPI, DPI, and QPI, with and without pre-incubation are shown in the Table 6.1 (K_i values were not determined because of the large quantity of

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

peptide inhibitor that would be required, given their low inhibitory potency). EPI gave the lowest IC_{50} value (1.97 ± 0.92 mM), which is similar to that obtained in a previous study (1.3 ± 0.2 mM)²⁷³. Compared to the kinetic studies with N-acetylated tripeptides¹⁶⁴, it is clear that NPI, DPI and QPI are much weaker inhibitors of PPE. This suggests that the positively-charged N-terminus of these tripeptides does indeed interact with the S_4 subsite to reduce the binding affinity with PPE. However, the minor decrease in inhibitory potency of EPI compared to AcEPI suggests that the putative hydrogen bond to Gln-192 may, to some extent, overcome the negative effect of the positively-charged N-terminus when EPI is complexed to PPE. The side chains of DPI and NPI may be too short to allow optimal hydrogen-bonding to Gln-192. The reason for the weaker binding of QPI is unclear at present, but may reflect the difference in charge between the carboxylate and amide side chains of EPI and QPI respectively.

All the IC_{50} values were observed to decrease significantly after 5 minutes pre-incubation of enzyme and peptide. It was found that the IC_{50} of NPI without pre-incubation was 25 mM. For the other three tripeptides, the IC_{50} values without pre-incubation were observed at least over 8 mM because it was observed that slight inhibition of these peptides on PPE activity at 8 mM was obtained (exact values were not be calculated due to the large quantity of peptide that would be required).

Peptides	IC_{50} (mM)	
	With pre-incubation	Without pre-incubation
NPI	7.5 ± 0.8	25.0 ± 0.9
EPI	1.97 ± 0.92	> 8
DPI	7.7 ± 0.2	> 8
QPI	5.0 ± 0.2	> 8

Table 6.1 IC_{50} values of NPI, EPI, DPI and QPI were determined at 25 °C, pH 7.5 with 0.2 mM substrate with PPE (0.5 µg). Two different conditions were used: pre-incubation of peptides and PPE (5 min) and without pre-incubation. Data were analyzed using Microsoft Excel.

The differences in the IC_{50} values between with the pre-incubation and without pre-incubation indicate that time- and concentration-dependent approach to steady state. All four tripeptides appear to be slow-binding inhibitors. In principle, there are two different ways in which the

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

slow-binding inhibition can occur: one is the direct binding of inhibitor or the isomerization of two different enzyme-inhibitor complexes²⁷⁴. In the direct-binding model, the binding of the inhibitor to the enzyme active site is either a slow process or both the binding and leaving are slow. For the second model, the inhibitor first binds rapidly to the enzyme, which is the one EI complex, and then this EI complex undergoes a slow conformational change to form the inactive complex EI*.

6.5 Co-crystallization with different peptides

To investigate the potential hydrogen bond between Gln-192 and these tripeptides, crystal structures of their complex with PPE were obtained. Native PPE crystals, for microseeding the complex crystals, were obtained in 24 hours in the solution containing sodium acetate (pH 5.0) and sodium sulphate as described previously²⁰⁰. For the PPE:tripeptide crystal, tripeptides (35 mg mL⁻¹) were dissolved in sodium acetate buffer (pH 5.0, 50 mM), and then PPE (50 mg mL⁻¹) was added. After centrifugation and filtering with a 0.22 µm filter unit, hanging drops were constructed from 2 µL of the protein/peptide solution, 2 µL of sodium sulphate solution (50 mM), and 1 µL of microseeding solution (1000× dilution from crushed native PPE crystals). For all four tripeptides, within 24 hours, crystals (Figure 6.4) were formed that had similar appearance to those of native PPE.

Figure 6.4 Crystals of the PPE:EPI complex grown in sodium acetate (pH 5.0, 20 mM) and sodium sulphate (30 mM).

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

Structures of the four PPE:tripeptide complexes were solved using molecular replacement. It is the method of choice for X-ray crystallography structure determination when structural homologues are available in the Protein Data Bank. Traditional molecular replacement methods (AMoRe, MolRep) are based on the properties of the Patterson function, which is essentially the Fourier transform of the intensities rather than the structure factors. Rotation functions and translation functions in Patterson function constitute the molecular replacement procedures. Another likelihood-based molecular replacement (Phaser) has also been developed recently. In this study, the program MolRep was used to get initial phases with the native PPE model from 1HAZ (Protein Data Bank identification code). First, the top solutions from rotation function between a chosen upper and lower resolution (20.0 and 1.6 or 1.7 or 1.8 Å) were obtained, and then the final solution was created from the translation function. TF_{ent} and R_{fac} values ranged from 35.21 to 40.33 and from 0.32 to 0.39 respectively. Lastly, rigid and restrained refinements (Refmac5) were performed using the output PDB file from MolRep.

In the structures of all four complexes, electron density for the enzyme active site residues is clear. However, due to the low occupancy of the peptides, it was difficult to distinguish electron density corresponding to these tripeptide residues with water molecules binding to the active site. Therefore, no tripeptides could be built in the active site. In order to avoid taking water molecules as disordered binding peptides, the $2F_o - F_c$ map of ligand-free PPE (PDB code: 3EST) is shown in Figure 6.5. Comparing this map with the following two maps (Figures 6.6 and 6.7) the map, it is clear that the electron density fragments in the region of the active site are likely to be disordered peptides. Furthermore, this map (Figure 6.5) combined with the map of PPE:BCM7 (Figure 7.10B) also supports the interpretation of the tetrahedral intermediate and E:S complex in the Chapter 7.

From the maps, it was found that the electron density map of PPE:DPI is similar to that of PPE:NPI. These two electron density maps showed the presence of electron density, which indicated peptide may be located in S subsite. It is interesting that the electron density maps from PPE:QPI and PPE:EPI suggested that the peptides were bound in both the S and S' subsites. The binding direction (forward or backward) of the peptides could not be determined through the electron density maps. However, combining kinetic data with the crystal structures (Figures 6.6 and 6.7), it appears that EPI and QPI inhibit PPE more strongly than DPI and NPI. Crystallographic data are shown in the Appendices.

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

Figure 6.5 Stereo-view of the active site of ligand-free PPE (PDB ID: 3EST). The atoms of the enzyme are shown in green. The $2F_o-F_c$ electron density maps for the enzyme are shown in beige and contoured at 1σ . The red F_o-F_c electron density maps (contoured at 2σ) represent water molecules around the active site residues. The resolution of the maps is $1.65\text{ \AA}$.

Figure 6.6 Stereo-views of the active site of PPE complexed with (A) DPI, and (B) NPI at pH 5 respectively. The atoms of the enzyme are shown in green. The $2F_o-F_c$ electron density maps for the enzyme are shown in beige and contoured at 1σ . The F_o-F_c 'omit' electron density maps for the peptides are shown in blue and contoured at 2σ . The resolution of the maps is 1.7 and $1.8\text{ \AA}$ respectively.

Mass spectrometric, kinetic, and crystallographic analysis of the interaction of small peptides with PPE

Figure 6.7 Stereo-views of the active site of PPE complexed with (A) EPI, and (B) QPI at pH 5 respectively. The atoms of the enzyme are shown in green. The $2F_o - F_c$ electron density maps for the enzyme are shown in beige and contoured at 1. The $F_o - F_c$ 'omit' electron density maps for the peptides are shown in blue and contoured at 2. The resolution of the maps is 1.6 and 1.7 Å respectively.

From Figures 6.6 and 6.7, it seems clear that the four tripeptides did not bind strongly in the active site. In this study, a high concentration ($\sim 20\text{--}50 \text{ mg mL}^{-1}$) of tripeptide was used. However, despite this, the occupancy of the peptides is clearly very low. As mentioned previously, all four tripeptides are poor inhibitors based on the kinetics studies. Therefore, the positively-charged N-terminus at the P_3 position appears to make it difficult for the tripeptides to form stable complex structures with PPE.

Chapter 7 Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

To investigate the structure of the first tetrahedral intermediate and the mechanism of serine protease acylation, a strategy in which dipeptides were soaked into crystals containing a performed acyl-enzyme complex (PPE:BCM7 or PPE:AcNPI) was performed in an attempt to form a peptide bond *via* nucleophilic attack of the N-terminal amine of the dipeptide onto the ester link between BCM7 and Ser-195. A range of L, L-dipeptides was chosen to investigate as nucleophiles for reaction with the PPE:BCM7 complex: NH₂-Ala-Ala-COOH, NH₂-Ala-Gly-COOH, NH₂-Arg-Phe-COOH, NH₂-Asp-Phe-COOH, NH₂-Lys-Ala-CONH₂, NH₂-Lys-Ser-COOH. The dipeptides were soaked into PPE:BCM7 or PPE:AcNPI co-crystals at pH 5 for 30-60 minutes, the crystals were then frozen (-140 °C) and X-ray diffraction data collected (see Appendices). All the structures were solved using molecular replacement (MolRep), as mentioned in Chapter 6.

PPE is crystallised at pH 5 to avoid autolysis which would occur at an active pH (~ 6 – 9.5). Thus, in order to trap the reaction products, a pH jump technique, was utilized to raise the pH within the enzyme crystal rapidly. As PPE has virtually no activity at pH 5 (Figure 6.3) and would become more active with increasing pH value until it reached the most active point at pH 8.5, it is possible to initiate the catalytic reaction by soaking the PPE crystals in a higher pH buffer. In this study, all crystals co-crystallised with a peptide ligand were subjected to an attempted pH jump *via* exposure to pH 9 buffer. The crystals were warmed to room temperature, their pH raised by buffer exchange (pH 5 to pH 9) and then re-frozen. This technique enabled an additional X-ray diffraction dataset to be collected on the *same* crystal that was used for the initial data collection at pH 5. It was found that a slightly higher temperature than usually employed during data collection (-140 °C instead of -160 °C) aided this procedure, perhaps by making the freeze-thaw process less stressful to the crystals, but the small difference in temperature is unlikely to be significant. It was found that the pH jumped crystals frequently cracked or dissolved during the pH jump giving a relatively low success rate. Occasionally, a second immersion in pH 9 buffer and subsequent re-freezing was also possible; this allowed the collection of a third dataset on the same crystal. These experiments were repeated several times for each dipeptide. In total, over 60 X-ray diffraction datasets were collected and refined.

7.1 First tetrahedral intermediate

Initial experiments soaking the dipeptides into PPE:BCM7 co-crystals at pH 5 were unsuccessful as they resulted in weak new density for the dipeptides in the S' subsites, the acyl-enzyme complex was observed as reported²⁰⁰. Removal of sulphate, which is necessary for initial crystallisation of elastase, from the soaking solution greatly improved the apparent occupancy of the dipeptide (a sulphate ion is observed to bind close to the S₁' subsite in native elastase²⁷², but not in the PPE:BCM7 acyl-enzyme structure). A disadvantage of the modified technique was that only short soaking times were possible before the crystals dissolved. In the case of the AlaAla and AlaGly dipeptides, only weak density was observed in the S₁' subsites despite very high concentrations of peptide being used in the soaking solution (70-180 mM). For all the other chosen dipeptides (KS, KANH₂, RF, DF) reasonably good occupancy was observed for the main chain atoms in the S' subsites (the average *B* factors were 36-40 Å², see Appendices), but the side chains were almost invariably disordered, consistent with the selectivity studies^{259,275,276} that imply the absence of defined side-chain binding sites in the S' subsites. The occupancy of BCM7 in the S subsites was concomitantly observed to decrease.

For most dipeptides (see below for exceptions), continuous electron density was observed that linked the amino terminus of the dipeptide, the side-chain oxygen of Ser-195 and the C-terminal carbon of BCM7 (Figures 7.1a, 7.2a-d). In each case, the unrefined electron density surrounding the carbon derived from the carbonyl group of the isoleucine of BCM7 was apparently tetrahedral (with *R* stereochemistry), implying *sp*³ hybridisation. Given the reproducibility of the observation of this species, BCM7 and the dipeptide main chain atoms (but not the side chains) were refined into the structures. The observation of electron density that could be successfully refined for a tetrahedral intermediate both with different crystals and different dipeptides, indicates that the observations of an apparent tetrahedral intermediate are not artefactual. For example, two separate crystals both soaked with ArgPhe, one for 45 min and the other for 50 min (Figures 7.1a/7.2a and 7.2d respectively) demonstrated very similar electron density in the region of Ser-195.

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Figure 7.1 Stereo-views of the active site of PPE complexed with BCM7 and (a) ArgPhe (soaked for 45 min) at pH 5, showing the refined 'tetrahedral structure'. The atoms of the enzyme are shown in green and the atoms of the peptides are shown in gold. All atoms shown in grey were only modelled and not refined against crystallographic data. The $2F_o - F_c$ electron density maps for the enzyme are shown in beige and contoured at 1σ . The $F_o - F_c$ 'omit' electron density maps for the peptides are shown in blue and contoured at 2σ . (b) Close-up stereo-view showing the difference density maps (contoured at $\pm 3 \sigma$) after refinement of the *cis*-E:S complex (shown in purple) from the LysSer structure (Figure 7.3a) into the data from the ArgPhe structure (a). Positive density is shown in blue and negative density is shown in red. The resolution of all maps is $1.7 \text{ \AA}$.

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Figure 7.2 Views of the active site of PPE complexed with BCM7 and different dipeptides at pH 5 showing the refined tetrahedral structures for (a) ArgPhe (soaked for 45 min), (b) AspPhe (soaked for 30 min), (c) LysAla-NH₂ (soaked for 30 min) and (d) a second ArgPhe crystal (soaked for 50 min). The colouring scheme and contouring levels for the atoms and maps is the same as for Figure 7.1a. The resolution of the structures is 1.6 Å [except for (a) which is 1.7 Å].

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

In all cases, the assigned oxyanion was located in the oxyanion hole within hydrogen bonding distances of the amide nitrogens of Gly-193 and Ser-195 (3.2 Å and 2.4 Å respectively). Significantly, a N-H-N hydrogen bond between the α -amino nitrogen atom of the P₁' residue and the amide nitrogen of Gly-193 (2.7 Å) was apparent. This implies that the amide nitrogen of Gly-193 in these structures has a bifurcated hydrogen bond and that the lone pair of the nitrogen atom of the P₁' residue (which would consequently have *S* stereochemistry) is involved in this hydrogen bond whilst the hydrogen atom on the nitrogen projects towards Ser-195. The atoms surrounding the tetrahedral carbon atom (O γ of Ser-195, C α of BCM7 Ile and the oxyanion) and the nitrogen atom of the P₁' residue (C α of the P₁' residue, a hydrogen atom and a lone pair) lie in a fully staggered conformation.

The residues from BCM7 were apparently located in an almost identical position to that observed in the reported PPE:BCM7 acyl-enzyme structure obtained at pH 5, with good occupancy being observed for the P₁ residue and poorer occupancy for the P₂, P₃ and P₄ residues. A similar observation on the relative occupancy of the P₁ residue compared to the P₂-P₄ residues was made in the previous crystallographic analyses on the PPE:BCM7 acyl-enzyme complex²⁰². The residues derived from the nucleophilic dipeptide were observed to be in a similar location as the P' residues observed in the crystal structure of a complex between PPE and a 28 residue non-covalently bound and non-hydrolysed peptide inhibitor that spanned the PPE active site²⁷⁵. The location of the main chain of the dipeptides was clearly identifiable through the position of the carbonyl group of the peptide bond linking the two residues of the dipeptide. There was an apparent hydrogen bond between the amide nitrogen of the P₂' residue and the main chain carbonyl of Thr-41, but insufficient electron density for the side chains of the dipeptide to be refined. The effectiveness of the soaking to remove sulphate from the crystals was demonstrated by examining the average *B* factor of the conserved (amongst all PPE crystal structures) sulphate ion remote from the active site. In the PPE:BCM7 acyl-enzyme structure, the *B* factor was 17 Å². This increased to an average of 62 Å² for the four structures shown in Figure 7.2.

7.2 Formation of an apparent enzyme:substrate complex

For one of the dipeptides (LysSer), significantly different electron density to that described above was observed at a 60 min soaking time (at pH 5). This density did not apparently correspond to a tetrahedral species, neither was there continuous electron density extending from the side chain oxygen of Ser-195 (Figure 7.3a). However, there was an apparent link

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

between the carbonyl derived from the isoleucine of BCM7 and the N-terminal α -amine group derived from the dipeptide lysine. The geometry was as expected for an enzyme:substrate intermediate except that the scissile peptide bond appeared to be in the *cis* configuration. The carbonyl oxygen of the putative new amide bond linking the BCM7 isoleucine and dipeptide lysine was located within the oxyanion hole (2.9 and 2.5 Å from the amide nitrogens of Gly-193 and Ser-195 respectively). The alcohol oxygen atom of the Ser-195 side chain was apparently well-positioned for attack onto the nascent amide with an angle of $\sim 100^\circ$ onto the plane of the peptide bond, in accordance with the optimal geometry proposed by Bürgi and Dunitz for nucleophilic addition to a carbonyl group²⁷⁷. When the geometry of the scissile bond was forcibly refined with the *trans* conformation, the atoms of the new amide bond did not fit well within the ‘omit’ map (Figure 7.3b), and the difference density maps (Figure 7.3c) showed negative density in the region of the *trans* peptide bond after refinement. When the apparent *cis*-E:S complex was refined into the data from the ArgPhe (45 min) tetrahedral intermediate (O γ of Ser-195 was removed to make the map clearer), positive density could be observed in the region of the ‘tetrahedral’ carbon and negative density in the region of the *cis*-amide bond (Figure 7.1b).

The same PPE:BCM7:KS-derived crystal was then removed from the stream of cryo-cooled nitrogen gas, immersed in pH 9 buffer for 30s, re-frozen and an additional X-ray diffraction dataset collected. In this case, the electron density revealed a linkage between Ser-195 and the P₁ isoleucine and P₁' lysine residues, and closely resembled the tetrahedral species observed for the other dipeptide substrates (Figure 7.4a).

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Figure 7.3 Stereo-views of the active site of PPE complexed with BCM7 and (a) LysSer (soaked for 60 min) at pH 5, showing the refined '*cis*-amide enzyme:substrate complex'. The atoms of the enzyme are shown in green and the atoms of the peptides are shown in gold. All atoms shown in grey were only modelled and not refined against crystallographic data. The $2F_o-F_c$ electron density maps for the enzyme are shown in beige and contoured at 1 σ . The F_o-F_c 'omit' electron density maps for the peptides are shown in blue and contoured at 2 σ . (b) Close-up stereo-view showing the same 'omit' map as displayed for (a) and overlaid with the LysSer E:S complex forcibly refined in the *trans*-conformation (shown in purple). This *trans*-conformation is similar to that found in the hybrid squash inhibitor:PPE structure²⁷⁵. (c) Close-up stereo-view showing the F_o-F_c difference density maps (contoured at $\pm 2.8 \sigma$) after refinement of the *trans*-LysSer E:S complex (shown in purple) into the data from the LysSer structure (a). Positive density is shown in blue and negative density is shown in red. The resolution of all maps is 1.7 Å.

7.3 pH Jump experiments on the assigned tetrahedral structures

A pH jump experiment was also carried out on the PPE:BCM7:RF crystal soaked for 50 min. After the first pH jump (pH 9 for 30s), the electron density (Figure 7.4b) was almost identical to that of the pH 5 structure (Figure 7.2d), except for a slight reduction in occupancy for the dipeptide (ArgPhe). After the second exposure to pH 9 buffer (28s), the electron density at the active site had changed and there was no longer continuous electron density for Ser-195 (Figure 7.4c). In this case, the density appeared to resemble the putative enzyme:substrate complex observed in the PPE:BCM7:KS structure (Figure 7.3a). From the F_o-F_c 'omit' map, the occupancy of the dipeptide appeared to have increased relative to the first pH jump, but the quality of the electron density for the peptides was noticeably worse in the $2F_o-F_c$ map and hence they were not refined into the structure. Using error-scaled distance difference matrices²⁷⁸, it was observed that the overall conformational changes were small with an average movement of ~ 0.3 Å and 0.4 Å for all atoms between the pH 5 structures and the first pH jumped structures respectively (Figures 7.5 and 7.6).

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Figure 7.4 Stereo-views of the active site of PPE complexed with BCM7 and different dipeptides showing (a) the results of the pH jump (pH 9, 30s) of the LysSer structure shown in Figure 7.3a, and the results of the pH jumps of the ArgPhe structure shown in Figure 7.2d: (b) one pH jump (pH 9, 30s), and (c) a second pH jump (pH 9, 28s). The colouring scheme and contouring levels for the atoms and maps is the same as for Figure 7.1a. The resolution of the structures is 1.8, 1.7 and 1.9 Å respectively.

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Figure 7.5 Plot of the distance between corresponding atoms in the PDB files from the PPE:BCM7:KS structure (Figure 7.3a) and the structure after one pH jump at 9 for 30 seconds (Figure 7.4a). The x axis represent each residue's number, the y axis is the distance (Å). The distances between corresponding atoms in the two PDB files are displayed after superposition based on the conformationally invariant part (red curve) and the uncertainty of the distance between atoms calculated by error propagation from the coordinate error (blue curve). From the figure, it is clear that the overall conformational changes were small with an average movement of ~0.3 Å and 0.4 Å being observed for all atoms between the pH 5 structure and the first pH jumped structure.

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Figure 7.6 Plot of the distance between corresponding atoms in the PDB files from the PPE:BCM7:RF structure (Figure 7.2d) and the structure after one pH jump at 9 for 30 seconds (Figure 7.4b). The x axis represent each residue's number, the y axis is the distance (Å). The red and blue curve were obtained using the same method as that in Figure 7.5. From the figure, it is clear that the overall conformational changes were also small and the most significant conformational changes are observed in the surface residues (near residue 185 and 245) remote from the active site.

7.4 BCM7 analogues

In order to test whether the observations were limited to BCM7 complexes, two analogues, *N*-acetyl-Asn-Pro-Ile-COOH (AcNPI) and NH₂-Asn-Pro-Ile-COOH (NPI), were then studied. AcNPI is a competitive inhibitor of PPE ($K_i = 0.89$ mM) and an atomic resolution (0.95 Å) structure of AcNPI complexed with PPE confirmed that this peptide formed an almost identical acyl-enzyme intermediate as BCM7, including the presence of an apparently productively positioned water molecule (hydrogen bond to Nε2 of His-57), and demonstrated that the ester bond was not distorted from planarity¹⁶⁴. The N-terminal acetyl group was added to ameliorate any potentially deleterious binding effect due to the introduction of a positive charge at the P₃ position. A PPE:AcNPI complex crystal was soaked with a saturated solution of AlaAla for 30 minutes using the same methodology as for the PPE:BCM7 crystals and an X-ray diffraction dataset collected. The electron density (Figure 7.7a) clearly showed a tetrahedral conformation around Ser-195 analogous to that observed in the PPE:BCM7:diptide crystals (Figures 7.1a, 7.2a-d). This result demonstrates that the formation of the tetrahedral intermediate is not dependent on the presence of a specific sequence for the peptide bound in the S subsites.

As mentioned in the previous chapter, NPI was found to be a very weak inhibitor of PPE based on kinetic data, supporting the proposal that a positive charge at the P₃ position reduces the binding affinity to the active site; the drop in the IC₅₀ value from ~25 mM to 7.5 mM after 5 min pre-incubation of the enzyme was suggestive of a slow-binding process. PPE was co-crystallised with NPI and an X-ray diffraction dataset was collected. The electron density implied that there was no covalent linkage between Ser-195 and the C-terminal carboxylate carbon of the isoleucine of NPI, and the conformation of the Ser-195 side chain was identical to that in native PPE (Figure 7.10e). As with the pH jump studies, electron density for the isoleucine of NPI was clearly visible, but the P₂ proline and P₃ asparagine residues were in lower occupancy. Although this compound only binds poorly to PPE, this dataset may reflect a position adopted by a C-terminal carboxylate product in the PPE active site; such an intermediate was not observed in the previous work on the deacylation of the acyl-enzyme. This difference reveals that there are subtle differences between peptides in their interactions with the PPE active site and that carboxylate product release may in fact be more efficient with (some) longer peptides, perhaps reflecting the fact that many or most of the PPE substrates *in vivo* are likely to be proteins.

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Then the PPE:NPI crystals were soaked with AlaAla at pH 5 and an X-ray diffraction dataset was collected. As before, the C-terminus of NPI was not covalently linked to Ser-195, but the dipeptide was clearly visible in the S₁'-S₂' subsites in reasonably good occupancy. However, it was bound in the reverse direction (C to N rather than N to C) to the other dipeptides examined (Figure 7.7b). The clear location of the carbonyl group of the peptide bond between the two alanines allowed assignment of the dipeptide binding in the reverse direction. 'Reverse binding' of peptides in the PPE active site has been observed previously²⁷⁹, and therefore it seems that peptides containing amino acids with small side chains are flexible with regard to their binding mode in the S' subsites. The fact that no tetrahedral intermediate was observed implies the necessity for formation of a productive acyl-enzyme intermediate before nucleophilic attack by the dipeptide can occur.

Figure 7.7 Stereo-views of the active site of PPE complexed with (a) AcNPI and AlaAla at pH 5 showing a similar tetrahedral intermediate as observed between BCM7 and different dipeptides (Figure 7.1a and Figures 7.2a-d), (b) NPI and AlaAla at pH 5 showing 'reverse' binding (relative to productive binding) of AlaAla in the S' subsites and density that may correspond to NPI in the S subsites. The colouring scheme and contouring levels for the atoms and maps is the same as for Figure 7.1a. The resolution of the three structures is 1.8, and 1.7 Å.

7.5 α_1 -Antitrypsin-derived peptide:MFLE

α_1 -Antitrypsin (α_1 -AT), a serpin (serine protease inhibitor), is a natural inhibitor of elastase that is currently undergoing clinical trials for the treatment of emphysema²⁸⁰. A reactive centre loop (RCL) in α_1 -AT is cleaved by elastase leading to a large conformational change and the formation of stable complex which inactivates the protease. Mass spectrometric and kinetic analyses have shown that the RCL sequence MFLEAIPM, corresponding to residues P₈-P₁, forms a stable complex with PPE and gives rise to apparent uncompetitive inhibition²⁸¹. Systematic truncation of this peptide demonstrated that a shorter tetrapeptide MFLE (residues P₈-P₅ of the RCL) retained uncompetitive inhibitory activity. K_i' (inhibition constant for uncompetitive inhibition) values of MFLEAIPM, MFLEAIP, and MFLE were 0.44 mM, 0.22 mM and 0.37 mM respectively. Although the kinetics of PPE inhibition by MFLE had been previously studied²⁸¹, no crystal structure of the PPE:MFLE complex had yet been obtained.

When this tetrapeptide was co-crystallised with PPE, the electron density in the active site (Figure 7.8) was consistent with the formation of an enzyme:substrate complex with the methionine being located in the S₁ subsite and FLE located in the S₁'-S₃' subsites. The conformation of the amino acids of the inhibitor was similar to that obtained for the BCM7-KS E:S complex that will be discussed in the following chapter, but the methionine residue of MFLE (provided by Prof. C.J.Schofield at the University of Oxford) was located further out of the S₁ subsite, presumably due to the longer length of Met compared to Ile; density corresponding to the methylene carbon atoms of the methionine side chain was visible. Significantly, the conformation of the amino acids of the inhibitor was similar to that obtained for the BCM7-KS E:S complex including the apparent *cis* configuration of the scissile P₁-P₁' amide bond, indicating that such a configuration can be accessed by binding of an intact peptide as well as by reaction of an acyl-enzyme intermediate with a dipeptide (e.g. KS). The PPE:MFLE co-crystals were also subjected to pH jumps (pH 9.0 for 30-60 seconds). Structures after these pH jumps indicated the formation of a first tetrahedral-like intermediate. However, the poor electron density maps made building the peptide into the structure difficult.

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Figure 7.8 Stereo-views of the active site of PPE complexed with MFLE at pH 5, showing the refined ‘*cis*-amide enzyme:substrate complex’. The atoms of the enzyme are shown in green and the atoms of the peptides are shown in gold. All atoms shown in grey were only modelled and not refined against crystallographic data. The $2F_o - F_c$ electron density maps for the enzyme are shown in beige and contoured at 1σ . The $F_o - F_c$ ‘omit’ electron density maps for the peptides are shown in blue and contoured at 2σ . The resolution of the maps is 1.6 Å.

7.6 Discussion

7.6.1 Comparison with previously obtained tetrahedral intermediates

It was previously mentioned in the introduction chapter that several crystallographic studies have examined the conformation of various tetrahedral adducts formed by the reaction of unnatural substrate analogues and inhibitors with serine proteases. Comparison of the conformation of the assigned tetrahedral species described here with these structures of various proposed intermediate analogues reveals certain differences. The structure of PPE complexed with a peptidyl boronic acid²⁸² showed that its peptide residues were located in the S subsites in an almost identical location to the residues of BCM7 in the PPE:BCM7 structures. One of the oxygens of the boronic acid is located in the oxyanion hole as observed for the PPE:BCM7 acyl-enzyme structure. The other oxygen atom of the boronic acid is much closer to the side chain of His-57 than is the P₁' derived nitrogen in the TI structures reported here. The crystal structure of a complex between chymotrypsin and a peptidyl trifluoromethyl ketone inhibitor^{220,283} also revealed a tetrahedral structure adjacent to Ser-195. The location of the hemiketal oxygen atoms in this structure are very close to the location of the boronic acid oxygen atoms in the PPE complex. A similar tetrahedral conformation was observed in the complexes between chymostatin and *Streptomyces griseus* protease A²⁸⁴, a peptidyl boronic acid inhibitor and chymotrypsin²⁸⁵ and leupeptin and trypsin²⁸⁶. The fact that in the TI

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

structures described here the oxygen atoms (or oxygen and nitrogen atoms) are equidistant between the carbonyl oxygen in the acyl-enzyme intermediate gives a new view of the conformation of the tetrahedral intermediate.

In addition, according to the comparison to the reported tetrahedral intermediate structures of γ -chymotrypsin soaked in hexane (see introduction, chapter 5), it was found that the conformation of the assigned chymotrypsin tetrahedral carbon is approximately similar to that in the obtained first tetrahedral intermediate structures reported here, but in the chymotrypsin study, the refined angles around the tetrahedral carbon are distorted with bond angles of 86.6° and 120.6° . The aspartic acid residue that was built into the S_1' subsite of γ -chymotrypsin also appears wrongly positioned for productive binding with the side chain lying in the location of the main chain of the P' peptide. In contrast, it is likely that the reproducibly obtained structures presented here, with defined peptides in both the S and S' subsites, more closely resemble the structure of the first tetrahedral intermediate.

Meanwhile, a comparison of the structures proposed as closely reflecting the first tetrahedral intermediate (TI) described here with the putative structure of the second TI that was previously described²⁰² (Figure 7.10d), indicate an almost identical conformation around the tetrahedral carbon with the oxyanion occupying a similar position in the oxyanion hole. In superimposed structures the two oxygen atoms bound to the central carbon in the tetrahedral species reported here ('first TI') and the oxygen atom and nitrogen atom in the previously reported tetrahedral species ('second TI') are both approximately equidistant from the carbonyl oxygen of the acyl-enzyme intermediate. The evidence for the tetrahedral species presented here is more extensive than that obtained for the TI in the hydrolysis step, where it is likely that more than one species was present, but the close resemblance between the two structures supports the previous assignment.

The distance of the assigned oxyanion to the amide nitrogen of Gly-193 is longer (3.1-3.2 Å in the first TI and 2.9 Å in the second TI) than that to the amide nitrogen of Ser-195 (2.4-2.5 Å in the first TI and 2.6 Å in the second TI). The distance to the amide nitrogen of Ser-195 is longer in the acyl-enzyme complex (2.8 Å) relative to the TIs probably in part because the angle $O\gamma-C=O$ (sp^2 hybridisation) is larger than $O\gamma-C-O^-$ (sp^3 hybridisation). In each case the relatively short distance between the oxyanions in the tetrahedral intermediates and the amide nitrogen of Ser-195 compared to that of Gly-193 is notable²⁸⁷. As the N-H-N

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

hydrogen bond between Gly-193 and the P₁' nitrogen atom has better geometry than the N-H-O hydrogen bond between Gly-193 and the tetrahedral oxyanion, it is possible that the latter has relatively less significance in stabilising the oxyanion. The fact that the oxygen atom of the carbonyl group of the E:S complex (like the acyl-enzyme intermediate) can also form good hydrogen bonds to these two amide nitrogens supports the suggestion that polarisation of the amide or ester carbonyl, and consequent electrophilic catalysis, is likely to be an important driving force for the mechanism. This is in agreement with Warshel's view that electrostatic interactions are of prime importance in the serine protease mechanism²⁸⁸.

Despite the collection of many data sets, there was no clear evidence for a productively-bound PPE:BCM7 acyl-enzyme intermediate with unreacted dipeptide present in the S' subsites. This suggests that under these soaking conditions at pH 5 in the presence of excess dipeptide, the equilibrium of the addition reaction in the PPE crystals lies towards the TI rather than the unreacted dipeptide. The acyl-enzyme intermediates formed from BCM7 or AcNPI appear stable at pH 5 in the crystal except when a peptide substrate is present. In the latter case, the first tetrahedral intermediate appears to be the predominant structure suggesting that nucleophilic addition by the N-terminal peptide amine occurs more readily than that by the hydrolytic water molecule. Although the majority of the structures for the assigned tetrahedral intermediate were obtained at pH 5, the pH-jumped structures of PPE:BCM7:KS and PPE:BCM7:RF (Figures 7.4a and 7.4b) suggest that a similar conformation may occur at higher pH.

7.6.2 Conformational changes during catalysis

It has been suggested that during the course of serine protease catalysis in solution there may be significant movements in some regions of the active site, particularly the side chains of His-57 and Ser-195²⁸⁹, and recent high resolution work on a protease:inhibitor complex suggested slight movements in the position of His-57 between the acyl-enzyme and E:S complexes²⁹⁰. Previous crystallographic work on the deacylation step implied only a small movement for the side chain of His-57 during deacylation. The movement of His-57 was coupled *via* a hydrogen bond network (His-57-Asp-102-Ser-214) to a small movement in the Ser-214-Val-216 β -strand away from the peptide substrate, thus potentially weakening the hydrogen bonding interactions that bind the substrate. This is a possible mechanism for synchronising hydrolysis of the acyl-enzyme intermediate with product release and support for this proposal has come from modelling studies²⁹¹. The refined pH jump technique in the

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

current work enabled use of the *same* crystal for data collection before and after the change in pH, allowing possible conformational changes in the enzyme to be observed more clearly.

Although at the resolution of the reported structures, small differences may not be apparent, the results indicate that the overall conformational changes, particularly in the region of the active site, between the acyl-enzyme intermediate, first TI and E:S complex are small. Unlike in the pH jump experiments with acylating lactam inhibitors²⁹², no significant movement of His-57 was observed. Overall the observations support the concept of a relatively rigid active site architecture into which a more flexible substrate (at least in the region of the scissile bond) can bind.

7.6.3 Geometry of the scissile amide bond

The possibility of a *cis*-peptide conformation being preferred (at least in some cases) for the scissile amide in the enzyme:substrate complex has not been widely discussed. It is likely that the binding energy of a peptide or polypeptide substrate to the active site would be sufficient to enable a *trans/cis* isomerisation ($\sim 10 \text{ kJ mol}^{-1}$ for Xaa-nonPro bonds²⁹³). It also seems probable that an E:S complex in which the scissile bond is *cis* would be closer in structure to the conformation of the first tetrahedral species observed in this work than the corresponding *trans* configuration (Figure 7.3b). Previous structures of peptide *inhibitors* which form stable non-covalent complexes with serine proteases have an intact peptide bond in a *trans* conformation spanning the S and S' subsites of the active site²⁷⁵, including the high-resolution structure of a trypsin-inhibitor complex²⁹⁰. It is possible that these peptides are inhibitors in part because the conformation of their scissile amides is bound in a non-productive *trans* form. It is notable that β -lactams (a four-membered ring) which contain a conformationally-restrained amide bond are inhibitors of a variety of nucleophilic serine enzymes including the D-Ala-D-Ala transpeptidases of bacterial cell wall biosynthesis and serine proteases. In addition, a crystal structure of the GyrA intein (not complexed with a protease) also revealed a *cis* conformation at its scissile peptide bond²⁹⁴.

Presently it is unclear precisely why different peptides may bind as *trans* or *cis* amides at the position of the scissile bond. The available evidence suggests that residues other than those at the P₁ and P₁' positions are involved in determining this variable. The Bowman-Birk inhibitors have a relatively rigid reactive site loop that adopts a fixed 'canonical' conformation with a *trans* peptide bond at the scissile bond position. They also contain a

highly conserved *cis*-proline at the P3' position that is necessary for biological activity²⁹⁵. This rigidity may prevent *trans/cis* isomerisation of the scissile peptide bond and hence allow these proteins to function as inhibitors.

It should be noted that there is no evidence presented here to imply that the *cis* configuration of the scissile amide is necessarily the only productive route; it has been argued that catalysis can proceed *via* more than one stereoelectronically viable arrangement²⁹⁶. However, assuming that *cis-trans* interconversion does not occur during the experiments, the pH jump experiments do imply that the observed conformation for the enzyme:amide complex is productive, at least in the crystalline state used here. This preference for a *cis* amide may also occur in the case of ester substrates where the *cis/trans* isomerisation has a lower energy barrier.

A comparison of the structure of the putative enzyme:substrate complex and acyl-enzyme intermediate shows a significant difference in the location of the amide and ester carbonyl groups respectively. However, the angle of nucleophilic attack by O γ of Ser-195 and the hydrolytic water molecule is almost the same (close to 100°), implying that there is some flexibility in achieving this stereoelectronically preferred arrangement in the active site, perhaps reflecting the fact that *in vivo* elastases probably are required to hydrolyse a range of substrates.

7.6.4 Stereoelectronic implications

Assuming the correct assignment of the hydrogen bond donor-acceptors between the P₁' nitrogen atom of the first TI and the amide nitrogen of Gly-193, the lone pair of the P₁' nitrogen faces away from His-57. This arrangement is predicted by stereoelectronic theory²⁹⁷: following nucleophilic attack of Ser-195, the nitrogen lone pair will be anti-periplanar to the C-O bond between O γ of Ser-195 and the tetrahedral carbon (Figure 7.9). In this conformation it is probably not possible for the leaving group P₁' nitrogen to accept a proton directly from Ne2 of His-57, which in the accepted mechanism has been protonated upon formation of the first TI. One possibility is that a solvent molecule donates a proton to the nitrogen atom²⁹⁸. However, it seems more likely that if this conformation is on the catalytic path, then the bonds about the nitrogen rearrange (by inversion and/or rotation) such that the nitrogen lone pair is directed towards Ne2 of His-57 as shown in Figure 7.9. As well as promoting C-N bond cleavage by protonation, this arrangement would relatively disfavour

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

C-O cleavage compared to C-N cleavage since the nitrogen lone pair is no longer anti-periplanar with respect to the C-O(Ser-195) bond. It should be noted that even after rearrangement the distance between the P₁' nitrogen and Nε2 of His-57 may be too large for hydrogen bond formation; in this case indirect proton transfer, possibly *via* O_γ of Ser-195, may be required.

Although in some non-optimal cases mechanistic variants may occur, for example, the slow deacylation rates observed with certain stable acyl-enzyme complexes^{299,300}, the current structures (Figure 7.10) support an optimal mechanism for serine protease catalysis in which every step occurs in a stereoelectronically favoured manner and in which there is minimal movement of the histidine residue that enables general acid/base catalysis. These mechanistic proposals also have consequences for inhibitor design as peptides which bind in a *trans* configuration may be regarded as high-energy intermediate analogues that do not form a tetrahedral intermediate. It is also notable that the activity of many naturally occurring serine proteases is regulated by peptides/proteins without resorting to the non-peptidic functional groups (e.g. boronic acids, trifluoromethylketones) commonly used in man-made inhibitors of serine proteases; a factor in the mechanisms of these inhibitors may be the configuration of the amides that span the active site.

Figure 7.9 Proposed forms of the first tetrahedral intermediate (A and B) interchangeable *via* rotation about the C-N bond. The amide nitrogen of Ser-195 which forms part of the oxanion hole is not shown for clarity; from the viewpoint shown it lies behind the plane of the picture. Note that steric constraints mean that the conformation in which the nitrogen lone pair is exactly coplanar with the C-O(Ser-195) bond is unlikely.

Mechanistic crystallography: reaction of peptides with the PPE:BCM7 complex

Figure 7.10 Left-hand side: stereo-views of the active site of PPE showing the view from crystal structures that potentially reflect the intermediates. (a) The apparent enzyme:substrate complex, PPE:BCM7:LysSer at pH 5, the same as Figure 7.3a. (b) The apparent tetrahedral structure from PPE:BCM7:LysAla-NH₂ at pH 5, the same as Figure 7.2c. (c) The assigned acyl-enzyme intermediate, PPE:BCM7 at pH 5 (PDB code 1HAX). (d) The apparent tetrahedral structure from previous work, PPE:BCM7 jumped to pH 9 for 1 min (PDB code 1HAZ). (e) PPE:NPI at pH 5 as a possible analogue of an enzyme:product complex. The colouring scheme and contouring levels for the atoms and maps are as for Figure 7.1a, except that the contour level of the $F_o - F_c$ 'omit' electron density maps for (c) and (d) was 2.5σ . Right-hand side: serine protease mechanism showing the hydrogen bonds (dashed lines) to the amide nitrogens of Gly-193 and Ser-195 forming the oxyanion hole, and between Nε2 of His-57 and Ser-195 and the hydrolytic water molecule.

7.7 Conclusion

The structures of the first tetrahedral intermediates in serine protease acylation step with different peptides at P₁' position gave the first indication of its likely conformation. The hydrogen bonds between the enzyme and the first tetrahedral intermediate oxyanion hole are in agreement with the assigned role of the oxyanion in stabilizing the tetrahedral intermediate. Obtaining a structure of the E:S complex from the soaking experiment demonstrated the ability of elastase to synthesize a peptide bond. Thus, the negative results of the HPLC experiments (Chapter 6) are not due to the inability of PPE to catalyse peptide synthesis, but rather are more likely to suggest that the assays conditions used in the HPLC experiments were not optimal for the reverse reaction. The 'pH jump' experiments demonstrated both that PPE is active in the crystalline state, and also that, by the principle of microscopic reversibility, the E:S complex obtained was fully productive. Furthermore, this structural information enabled a detailed new proposal for the pathway of the acylation step to be made. At least in some cases, the hydroxyl oxygen of Ser-195 attacks the *cis*-amide carbonyl in a stereoelectronically favoured manner to give a tetrahedral intermediate which undergoes C-N bond rotation to enable protonation of the leaving group nitrogen.

Chapter 8 Materials and methods

Part I

8.1 Materials (I)

8.1.1 General chemicals and reagents

All the restriction endonucleases used in molecular cloning and factor Xa were from New England Biolabs. TEV protease was a gift from Dr. H.S.Yoon's laboratory. The general buffers were prepared according to the guidelines given in Molecular Cloning (3rd edition, Cold Spring Harbour Laboratory Press, USA): 10× PBS buffer (1.3 M NaCl, 357 mM Na₂HPO₄, 143 mM KH₂PO₄ and 30 mM KCl), 10× TAE buffer (2 M Tris base, 57.1 mL of glacial acetic acid and 0.05 M of EDTA, and adjusted to 1 L solution with MilliQ H₂O), 5× TBE buffer (1.1 M Tris, 900 mM sodium borate, 25 mM EDTA, pH 8.0), 10× SDS-PAGE running buffer (0.25 M Tris base, 2.5 M glycine and 34.68 mM SDS, and adjusted to 1 L solution with MilliQ H₂O), staining solution for SDS-PAGE [0.2% Coomassie Blue R-250 in 45:45:10% (v/v) methanol:water:acetic acid], destaining buffer [25:65:10% (v/v) methanol:water:acetic acid mix]. Primary antibody (DB063 His H20, rabbit polyclonal antibody) and secondary antibody (anti-Rabbit IgG + HRP: Goat-anti rabbit and conjugated with horseradish peroxidase) were from Delta Biolabs, USA. ECL plus western blotting detection reagents were obtained from Amersham Biosciences. Luria-Bertani broth (LB) (10 g L⁻¹ tryptone, 5 g L⁻¹ yeast extract and 10 g L⁻¹ NaCl) and 2TY media (16 g L⁻¹ tryptone, 10 g L⁻¹ yeast extract and 5 g L⁻¹ NaCl) were also prepared according to the guidelines of Molecular Cloning (3rd Edition).

8.1.2 Molecular cloning and expression

pET24d was from Novagen, pETM-30 was a gift from Dr. H.S.Yoon's laboratory, and pMAL-C₂X was a gift from Prof. James P Tam's laboratory. All the competent cell strains [*E. coli* DH5α and BL21 (DE3)] were from Novagen and prepared by CaCl₂ transformation according to Molecular Cloning III. The oligonucleotide primers were from Research Biolabs, 1st Base and Proligo.

8.1.3 Protein purification

GST (glutathione-S-transferase) affinity resin (Glutathione Sepharose 4B) and amylose resin were from Amersham Biosciences. Ni-NTA Superflow resin was from Qiagen, and Co²⁺ metal affinity resin (BD TALON™) was from BD Biosciences. Bradford assay solution and

Materials and methods

BioSpin-30 columns were from BioRad. The fast desalting column (PD-10) was from Amersham Biosciences. Centrifugal protein concentrators were ordered from Millipore.

8.1.4 Functional and mutagenesis studies

The oligonucleotides used in the unwinding and mutagenesis studies were synthesized by Research Biolabs, 1st Base, Qiagen (HPLC grade), or Dharmacon (Colorado, USA). All oligonucleotides (from Prologo, Research Biolabs, 1st Base, or Dharmacon) were solubilized in sterilized water to a final concentration of 100 μ M (the sequences are shown in Table 8.1). ATP, TLC plates, and polyribonucleotides were from Sigma-Aldrich. The Quikchange site-directed mutagenesis kit was ordered from Stratagene.

8.1.5 Crystallization, data-collecting and processing

Crystallization plates and crystallization kits were from either Sigma-Aldrich or Hampton Research. The coverslips were from Molecular Dimensions, UK. Selenomethionine was purchased from Fluka. Heavy atoms were from Sigma-Aldrich and Fluka.

8.1.6 Equipment

The FPLC machine (Fast Performance Liquid Chromatography) was purchased from Amersham Biosciences. Purification columns were from BioRad (Econo Columns) and Amersham Biosciences (XK16/20 column, MonoQ, MonoS, Superdex 75, Superdex 200, Q/S sepharose, HIC column, *etc.*). The phosphoimager (Molecular Imager FX) was from BioRad. Film cassettes were from Kodak BioMax. The TLC chamber was from Sigma-Aldrich. The gel drying system was purchased from BioRad. Image plates were from Kodak Storage Phosphor. The Chirascan CD spectrometer was from Applied Photophysics, UK. Particle size analyzer (90 plus) was from Brookhaven Instruments Corporation, USA. The Rigaku MicroMax-007 X-ray generator and R-axis IV++ image plate detector were from Rigaku, Japan/USA.

Materials and methods

Designation	Sequences (5'-3')	Purpose
Oligo 1	CTG TCC TGC ATG ATG	DNA oligos used to form DNA substrates
Oligo 2	CAT CAT GCA GGA CAG	
Oligo 3	CAT CAT GCA GGA CAG TCG GAT CGC AGT CAG	
Oligo 4	CTG TCC TGC ATG ATG TCG GAT CGC AGT CAG	
Oligo 5	GAC TGA CGC TAG GCT CTG TCC TGC ATG ATG TCG GAT CGC AGT CAG	
Oligo 6	GAC TGA CGC TAG GCT CAT CAT GCA GGA CAG	
Oligo 7	ATG TAA TAC GAC TCA CTA TAG	Oligos used in <i>in vitro</i> transcription
Oligo 8	GTG TCC TGC ATG ATC AGC CTA GCG TCA GTC CTA TAG TGA GTC GTA TTA CAT	
Oligo 9	CTG ACT GCG ATC CGA GAT CAT GCA GGA CAC CTA TAG TGA GTC GTA TTA CAT	
Oligo 10	GAT CAT GCA GGA CAC CTA TAG TGA GTC GTA TTA CAT	
Oligo 11	GTG TCC TGC ATG ATC CTA TAG TGA GTC GTA TTA CAT	Oligos used to form substrates
RNA 1	GAC UGA CGC UAG GCU GAU CAU GCA GGA CAC	
RNA 2	GUG UCC UGC AUG AUC	Oligos used in mutagenesis study
Primer 1	CCT GGT ACT GGT AAG AGT CAT TTT GCC ATC GG	
Primer 2	CCG ATG GCA AAA TGA CTC TTA CCA GTA CCA GG	
Primer 3	GTG TTG TCA ATG CTA GAC TTC GTG CAA AAC	
Primer 4	GTT TTG CAC GAA GTC TAG CAT TGA CAA CAC	
Primer 5	GCA CAC AAG GAG AAG TCA GCT C	
Primer 6	GAG CTG ACT TCT CCT TGT GTG C	
Primer 7	CCC TAT GTG AAA AGG CAT CAA AAT ATT TGC CC	
Primer 8	GGG CAA ATA TTT TGA TGC CTT TTC ACA TAG GG	

Table 8.1 Oligonucleotides used in functional study: oligonucleotides (HPLC grade) used to prepare DNA and RNA substrates, and oligos used in site-directed mutagenesis are shown. DS15 substrate (substrate 1, Table 3.2 in Chapter 3): Oligo 1 + **Oligo 2**, trapping with Primer 4; 3'-5' substrate (substrate 2, Table 3.2): **Oligo 2** + Oligo 4, trapping with Oligo 1; ss15+ss45 substrate (substrate 3, Table 3.2): **Oligo 2** + Oligo 5, trapping with Oligo 1; 5'-3' substrate (substrate 4, Table 3.2): **Oligo 1** + Oligo 6, trapping with Oligo 2; ss30+ss30 substrate (substrate 5, Table 3.2): Oligo 4 + **Oligo 6**, trapping with Oligo 1. Bold words indicate those labelled by T4 polynucleotide kinase and [γ - 32 P]-ATP. Primer 1 and Primer 2 were used in rectifying mutation G289 to S289. Primer 3 and Primer 4 were for rectifying mutation K390 to R390. Primer 5 and Primer 6 were for the mutant D466E. Primer 7 and Primer 8 were for mutant L322S.

8.2 Molecular cloning and protein expression

8.2.1 pET24d

The orflab cDNA from virus isolate KYK from a Singapore patient infected with SARS was kindly provided by the Genome Institute of Singapore. Using PCR, two sequences [16151-17953, representing the full-length helicase (nsp13) and 16928-17953, representing an N-terminally truncated helicase (Δ 1-259) missing the zinc finger domain and hinge/spacer region] of the SARS genome were amplified and cloned as NcoI-XhoI fragments into pET24d. After amplifying in *E. coli* DH5 α cells, they were confirmed by NcoI-XhoI double digestion and DNA sequencing (performed by Research Biolabs, Singapore). Then the positive clone was transformed into *E. coli* BL21 (DE3) cells. In addition, by designing PCR primers containing a 6 $\times$ His sequence, the sequence (16928-17953) with an N-terminal 6 $\times$ His tag was cloned into pET24d with NcoI and XhoI.

A single colony was picked from the plate and cultured in LB or 2TY media containing 30 $\mu\text{g mL}^{-1}$ kanamycin at 37°C or 30°C and shaken at 250 rpm. When the OD₆₀₀ reached a certain value (normally 0.6-0.8), IPTG (isopropyl- β -D-1-thiogalactopyranoside) at a certain final concentration (normally 0.4-1.0 mM) was added to the cells. After certain number of hours, the cells were harvested by centrifuging at 14,000 rpm. The cells were placed in two different lysis buffers, buffer 1 for protein with a 6 $\times$ His tag [25 mM TRIS pH 8, 10% (v/v) glycerol, 20 mM β -ME, 1 mM PMSF (in EtOH), 1 μM pepstatin A, 0.2 M NaCl] and buffer 2 for protein without a 6 $\times$ His tag [25 mM TRIS pH 8, 10% (v/v) glycerol, 20 mM β -ME, 1 mM PMSF (in EtOH), 1 μM pepstatin A, 0.2 M NaCl], and lysed by sonication. Expression was checked using a 10% SDS-PAGE gel. In order to find the optimal induction condition, different induction temperatures, induction times, different IPTG concentrations, and different induction points were tried.

8.2.2 pETM-30

The same two sequences from the SARS-CoV genome (16151-17953 and 16928-17953) were cloned as NcoI-XhoI fragments into a different plasmid: pETM-30. After confirmation of the positive clones, successive transformation into *E. coli* DH5 α and BL21 (DE3) cells were performed by the same methods as that for pET24d. However, after harvesting the cells were placed in a different lysis buffer [1 $\times$ PBS buffer, 20 mM β -ME, 1 mM PMSF (in EtOH), 1 μM pepstatin A].

8.2.3 *pET24d + pETM-30*

The GST gene contained in the vector pETM-30 and the sequence of $\Delta 1-259$ nsp13 from pET24d were cloned by PCR. Two sequences were taken as templates to obtain a new sequence comprising, in order from the N-terminus, a 6×His tag, a GST tag, a Pro-Ser-Pro linker, and $\Delta 1-259$ nsp13. The sequence was cloned into pET24d as an NcoI/XhoI fragment. The protocol for expression was similar to that used for pETM-30.

8.2.4 *pMAL-C₂X*

A construct containing a third tag, MBP, was made by cloning the two sequences as BamHI-PstI fragments into pMAL-C₂X. An alternative method for the screening of positive clones, colony PCR, was used. Single colonies (diameter over 1 mm) were picked into 50 μ L MilliQ H₂O, vortexed and heated at 99 °C (5 min), and centrifuged. The supernatant (10 μ L) was used to perform PCR with the same primers as used in molecular cloning. The positive clones from colony PCR were also checked by double cloning. A different lysis buffer [50 mM TRIS pH 7.4, 10% (v/v) glycerol, 10 mM β -ME, 1 mM PMSF (in EtOH), 1 μ M pepstatin A, 1 mM EDTA, 0.2 M NaCl] was used after harvesting.

8.2.5 *Selenomethionine-substituted proteins*

The selenomethionine-substituted protein used in the crystallization trials for MAD experiments was expressed using different methods comparing to that for native protein expression. M9 medium (42.27 mM Na₂HPO₄, 22 mM KH₂PO₄, 18.7 mM NH₄Cl, 8.56 mM NaCl in 1 L MilliQ H₂O) with several additives [1 M MgSO₄ (1 mL), 40% glucose (10 mL), 0.5% (v/v) thiamine (100 μ L), 0.0276 mM FeSO₄, and 50 mg mL⁻¹ kanamycin (1 mL) in 1 L M9 medium] was used to grow selenomethionine-substituted protein. A small amount (1-5 mL) of overnight culture in LB medium containing the recombinant *E. coli* was centrifuged and resuspended in this M9 medium. The seed liquor was added into a large volume of M9 medium. Seven different amino acids (100 mg L⁻¹ L-lysine, 100 mg L⁻¹ L-phenylalanine, 100 mg L⁻¹ L-threonine, 50 mg L⁻¹ L-isoleucine, 50 mg L⁻¹ L-selenomethionine or 100 mg L⁻¹ DL-selenomethionine, 50 mg L⁻¹ L-leucine, 50 mg L⁻¹ L-valine in 1 L medium) were added into the medium when the OD₆₀₀ reached 0.3. Induction by IPTG and harvesting followed the same protocol as for the unsubstituted protein. During purification and crystallization, 1 mM DTT or 10 mM β -ME was added to all the solutions.

8.3 Protein purification

8.3.1 Purification of protein with 6×His tag

The supernatant of the lysed solution was passed through 0.45 μm and 0.22 μm filters and purified with 6×His tag affinity resins. Two different resins and methods were used. In the first method, the solution was loaded onto Ni-NTA Superflow resin (10 mL) in a XK16/20 column (Amersham Biosciences) mounted on an FPLC (Amersham Biosciences). Elution was performed using a linear gradient of 20-500 mM imidazole in a solution containing 50 mM Tris pH 8.0 and 250 mM NaCl. Fractions containing the helicase were pooled. Another method, Ni-NTA resin (Qiagen) or Co^{2+} metal affinity resin (BD TalonTM) packed in an Econo-Column (BioRad), was also performed. The solution was incubated with the resin in the column overnight at 4 °C, and then the resin was washed with wash buffer (25 mM TRIS pH 8, 20 mM β -ME, 0.25 M NaCl) for at least 5 column volumes. The final solution containing target protein was obtained by washing with elution buffer (25 mM TRIS pH 6, 20 mM β -ME, 0.25 M NaCl).

Fractions or eluted solution containing the helicase were desalted with 1 mM DTT, 50 mM TRIS pH 8 through a fast desalting column (PD-10). The desalted sample solutions were loaded onto a MonoQ column (1 mL) (Amersham Biosciences) mounted on a FPLC. After two column volumes of washing with buffer A (50 mM TRIS pH 8.0 and 1 mM EDTA) at a rate of 1 mL min⁻¹, elution was performed using a linear gradient of 0.0-1.0 M NaCl in a solution containing 50 mM TRIS pH 8.0 and 1 mM EDTA. Fractions containing the helicase at 97-99% purity as judged by SDS-PAGE analysis (10% SDS-PAGE gel) were pooled, and desalted into 10 mM TRIS pH 7.4 using a BioSpin-30 column.

8.3.2 Purification of protein without tag

For the proteins without a 6×His tag, the supernatant was directly loaded onto Q Sepharose FF resin (16 mL, Amersham Biosciences) in a XK16/20 column, and eluted using a linear gradient of 0.0-1.0 M NaCl in a solution containing 50 mM TRIS pH 8.0. The fraction with the desired protein was concentrated to 1-2 mg mL⁻¹ with a protein concentrator (Amicon Ultra-15, 30,000 MWCO, Millipore) and loaded onto a Sephadex 75 column (20 mL, Amersham Biosciences). Fractions containing the target protein were loaded onto a MonoQ column (1 mL), eluted with a linear gradient of 0.0-1.0 M NaCl, and desalted using a BioSpin-30 column.

Materials and methods

8.3.3 Purification of insoluble protein

After harvesting, the cells were sonicated and centrifuged, and the pellet was washed at least 5 times with buffer (0.1 M NaH₂PO₄, 10 mM Tris, pH 8) until the pellet was almost white. Resuspension buffer (0.1 M NaH₂PO₄, 8 M urea, 10 mM Tris, pH 8) was used to dissolve the pellet at 4 °C (overnight). The solution was centrifuged and the supernatant was loaded onto Ni-NTA resin in an Econo-Column followed by washing with buffer (0.1 M NaH₂PO₄, 8 M urea, 10 mM Tris, pH 6). The final solution containing denatured helicase was obtained by elution with buffer (0.1 M NaH₂PO₄, 8 M urea, 0.5 M imidazole, 10 mM Tris, pH 6).

Refolding solution (10 mM Tris pH 8, 0.5 M L-arginine, 1 mM DTT) with a volume 100 times greater than that of sample was used to refold the preliminarily purified denatured protein. The sample was added very slowly (5-10 µL min⁻¹) at 4 °C with slow stirring and then concentrated to a small volume (0.5-2 mL) using a concentrator (Amicon 8200, Millipore). Superdex 75 (20 mL) and a BioSpin-30 column (BioRad) were used to obtain the final solution of almost pure helicase.

8.3.4 Purification of protein with GST tag

Before the supernatant containing the helicase was loaded onto the GST affinity resin (4 mL) in an Econo-Column, at least 3 bed volumes of 1 × PBS were used to equilibrate the gel. The sample was mixed and incubated with resin at room temperature with gentle agitation for 30 min or overnight at 4 °C. The column was washed with a further 5-10 bed volumes of PBS. The final solution was obtained by eluting with 3 mL buffer (10 mM reduced glutathione in 50 mM Tris pH 8).

Ion exchange chromatography (MonoQ, see 5.3.1 for protocol) was used to increase the purity to ~95%. Fractions containing the helicase were cleaved by TEV protease with a molar ratio of 1:20 (protease:sample) in buffer (50 mM Tris pH 8, 1 mM DTT, 1 mM EDTA) at 8 °C for 12 hours. Fractions containing the target protein were loaded onto the GST affinity resin and then passed through the MonoQ column again. On-column cleavage was also attempted. TEV protease (20 µL, 1 mg mL⁻¹) in the reaction buffer (10 mL) was incubated with the sample and GST affinity resin, and then washing and elution were performed. A BioSpin-30 column was used for desalting.

Materials and methods

Purification of the fusion protein with an N-terminal GST tag expressed in pET24d (section 8.2.3) was similar to that of the fusion protein expressed in pETM-30. However, this fusion protein was purified as an intact protein without cleavage.

8.3.5 Purification of protein with MBP tag

The supernatant containing the helicase was loaded onto amylose resin (4 mL) in an Econo-Column, previously equilibrated with at least 5 bed volumes of lysis buffer (50 mM Tris pH 7.4, 0.2 M NaCl, 1 mM EDTA, 10 mM β -ME). The sample solution was incubated with the resin at room temperature with gentle agitation for 1-2 hours or overnight at 4 °C. The resin was then washed with the same 5-10 bed volumes of lysis buffer. The final solution was obtained by eluting with 3 mL lysis buffer supplemented with 10 mM maltose.

A Superdex 200 column (20 mL) (Amersham Biosciences) mounted on a FPLC was equilibrated with at least two column volumes of buffer (50 mM Tris pH 8, 100 mM NaCl). The eluent from the amylose resin was loaded onto the column. Fractions containing helicase (as checked by SDS-PAGE) were cleaved by Factor Xa protease. The optimal reaction condition contained protease and helicase in a 1:50 molar ratio in buffer (1 mM DTT, 50 mM Tris pH 8) at room temperature. The cleaved fractions containing the target protein were repurified with amylose resin, and then loaded onto a MonoQ column to obtain protein of higher purity. On-column cleavage was also attempted. Factor Xa (10 μ L, 1 mg mL⁻¹) dissolved in the reaction buffer (6 mL) was incubated with the sample and amylose resin, and then washed and the helicase eluted as previously described. The final protein sample was desalted using a BioSpin-30 column.

8.4 Protein identification and analysis

8.4.1 Western blotting for *nsp13* with a 6 $\times$ His tag

The PVDF membrane was soaked in methanol for 10 min and transferred into buffer (1 L, pH 8: 0.02 M Tris-base, 0.15 M glycine, 200 mL methanol in MilliQ H₂O). The SDS-PAGE gel containing the target samples was placed on the semi-dry transfer apparatus (BioRad) in the following order from bottom to top: filter paper (two sheets), PVDF membrane, gel, filter paper (two sheets). After the proteins were transferred into the PVDF membrane (20 V, 20 min), the membrane was soaked in 10 mL blocking solution of 5% milk powder in PBST buffer (50 mM sodium phosphate pH 7.5, 155 mM NaCl, 0.05% Tween 20) for 1 hr.

Materials and methods

The membrane was washed with PBST buffer, and 15 μL rabbit polyclonal primary antibody ($200\text{ }\mu\text{g mL}^{-1}$) with a dilution of 1:200 in 3 mL PBST buffer was spread on the surface of the membrane for 1 hour at room temperature or overnight at 4 °C. The membrane was washed with PBST buffer twice with 10 min interval each, and incubated with 1 μL goat-antirabbit secondary antibody (1 mg mL^{-1}) with a dilution of 1:10000 in 10 mL PBST buffer for 1 hour at room temperature. Again the PVDF membrane was washed twice with PBST buffer. Then the ECLplus western blotting detection system (Amersham) was added to the surface of the membrane and incubated at room temperature for 5 min. Finally, the membrane was placed on the top of a film (Kodak BioMax) in a cassette and exposed for 15-30 seconds.

8.4.2 Circular dichroism spectrometry

Far UV-circular dichroism spectra were recorded at 25 °C. Filtered 10 mM Tris pH 7.4 buffer (60 μL) (as a blank) was loaded into a quartz cuvette (Hellma) in a cuvette holder, and placed in the Chirascan CD spectrometer. In the presence of dry nitrogen gas, UV radiation in the range 175 nm to 260 nm with a step of 0.5 nm/s was used to scan the buffer solution in order to obtain the baseline. Then, the target protein, SARS-CoV Δ 1-259 nsp13 MBP fusion protein (construct 10), in the same buffer with a total volume of 60 μL was placed in the cuvette and scanned across the same range. The result was obtained by subtracting the baseline using Chirascan software.

8.4.3 Mass spectrometry

Intact protein (0.1 mg mL^{-1} nsp13 with a 6 $\times$ His tag) dissolved in MilliQ H_2O was analysed by MALDI-TOF-MS (matrix-assisted laser desorption time-of-flight mass spectrometry) (performed by Shimadzu, Singapore). The sample was diluted in 0.1% TFA, purified and mixed with a matrix solution. The results were analyzed on an AXIMA-CFRplus in positive-ion linear mode. The MBP fusion protein (with Δ 1-259 nsp13) was also determined by ESI-MS (Thermo Electron Corporation, LCQ Deca XP Max); the mass value was obtained by deconvolution with ProMass.

GST-tagged full length nsp13 (conducted by Protein & Proteomics Centre, NUS) and MBP-tagged Δ 1-259 nsp13 (conducted by Bruker, Singapore) were identified by in-gel trypsin digestion followed by PMF. The bands of the target protein on the SDS-PAGE gel were excised, and then trypsin digestion was performed (by the company). The final solution was determined by MALDI-TOF-MS. The resulting mass spectrum of peptides, the peptide

Materials and methods

mass fingerprint or PMF, was searched against the sequence database Mascot. Furthermore, some peptide fragments from the MBP fusion protein were analyzed by MS/MS. Possible amino acid sequences for the trypsin-digest peptides were obtained.

8.4.4 Dynamic laser light scattering

Particle size distributions data were obtained at 25 °C using Particle Size Analyzer 90 plus, (Brookhaven Instruments, USA). The sample (0.1 mg mL^{-1}) was dissolved in 10 mM Tris pH 7.4, $0.22 \text{ }\mu\text{m}$ filtered and placed in disposable acrylic square cells (1.5 mL). The laser light (660 nm) scattering was measured at 90°. Three runs (9 min) were performed to obtain a mean value. The results were analysed to give the effective diameter and particle size distributions.

8.4.5 N-terminal protein sequencing

The helicases with 6×His tags were analysed by N-terminal protein sequencing (ABI Procise 494 Protein Sequencer, performed by Protein & Proteomics Centre, NUS, Singapore). The samples (0.1 mg mL^{-1}) were separated by SDS-PAGE, the gels were electroblotted onto PVDF membranes, and then the membranes were stained by Coomassie Blue R-250, and loaded directly on the sequencer. The amino acid sequences were obtained by Edman degradation chemistry.

8.5 ATPase assay

8.5.1 Basic ATPase assay

The ATPase assays were carried out at 37 °C in a total volume of 20 μL using [γ - ^{32}P]-ATP ($3000 \text{ Ci mmol}^{-1}$, $0.2 \text{ }\mu\text{Ci}$) in a solution containing nsp13 (23 pmol), 50 mM HEPES pH 7.4, 2 mM DTT, 3 mM MgCl_2 , 0.1 mg mL^{-1} BSA. The reaction was stopped by the addition of 100 mM EDTA after either 5, 20 or 30 min. A positive control was provided by substituting the enzyme with 0.5 M HCl, and a negative control by using enzyme that had been heated to 95 °C. The mixture ($0.5 \text{ }\mu\text{L}$) was then applied to a 0.1 mm cellulose polyethyleneimine (PEI) thin layer chromatography (TLC) plate (Sigma-Aldrich) and separated with a mixture of 0.15 M methanoic acid and 0.15 M LiCl (pH 3.0). The TLC plate was either exposed onto film at -80 °C or scanned using a phosphoimager (Molecular Imager FX, BioRad).

8.5.2 Effects of different co-factors on ATPase activity

For some of the assays, polyribonucleotide [poly(U), poly(G), poly(A) or poly(C)] at a concentration of either 50 $\mu\text{g mL}^{-1}$ or 200 $\mu\text{g mL}^{-1}$ was added to the reaction solution, but the total volume was kept at 20 μL . When polyribonucleotide was added, a single time point of 20 min was used. In certain assays, different factors: Zn^{2+} (3 mM), Mg^{2+} (30 mM), Mn^{2+} (3 mM), double-stranded DNA (10 pmol), single-stranded DNA (10 pmol), or double-stranded DNA (10 pmol) were added to the reaction solution. Helicase treated with 1 mM EDTA for 15 min and desalted with a BioSpin-30 column was also used.

8.6 DNA unwinding assay

8.6.1 [γ - ^{32}P] labelling

The synthetic DNA substrates used for the assay are shown in Table 1. The appropriate oligonucleotide was labelled with [γ - ^{32}P]-ATP (3000 Ci mmol^{-1} , 10 μCi) using T4 polynucleotide kinase and purified using a BioSpin-30 column. To produce duplex DNA substrates two oligonucleotides were heated to 95 $^{\circ}\text{C}$ for 10 min and slowly cooled to room temperature. The DNA substrates (0.5 pmol) were incubated in a volume of 20 μL with the purified helicase (23 pmol), 25 mM HEPES pH 7.4, 1.5 mM magnesium chloride, 2 mM DTT and 0.1 mg mL^{-1} BSA at 37 $^{\circ}\text{C}$ for either 5, 20 or 30 min. The reaction was initiated by the addition of 2.5 mM ATP and the appropriate trapping strand (10 pmol). The reaction was stopped by the addition of loading buffer [20 mM Tris pH 7.5, 6% (v/v) glycerol, 5 mM EDTA, 0.1% (w/v) SDS, 0.04% (w/v) xylene cyanol, 0.04% (w/v) bromophenol blue]. A positive control was provided by heating the mixture to 95 $^{\circ}\text{C}$ for 10 min and slowly cooling to room temperature. Two negative controls were provided by either using enzyme that had been heated to 95 $^{\circ}\text{C}$ or by omitting ATP. The reaction products were run on a 20% polyacrylamide TBE gel [10 mL containing 6.66 mL 30% acrylamide/bis-acrylamide (29:1) stock, 1.27 mL MilliQ H_2O , 2 mL $5 \times \text{TBE}$, 70 μL 10% APS, and 4 μL TEMED]. The gel was then dried and then exposed either onto X-ray film (Kodak BioMax) at -80 $^{\circ}\text{C}$ and developed or onto an image plate (Kodak Storage Phosphor) and scanned using a phosphoimager (Molecular Imager FX, BioRad). All assays were performed at least in triplicate.

8.6.2 Ethidium bromide and silver staining

The unwinding reaction results were also examined by silver staining and EB (ethidium bromide) staining. The non-radioactive substrate (1 μL) was prepared by mixing two

Materials and methods

single-stranded oligonucleotides (100 mM) and annealed after heating at 95 °C for 10 min. The same unwinding buffer and conditions were used in the unwinding reaction analysis. The gels with the separated bands were soaked in $0.5 \times$ TBE buffer containing 5 μ L EB (10 mg mL⁻¹). After 20 min incubation, results were examined using a UV transilluminator and the image was recorded using a gel documentation system (Gel Doc 2000, BioRad).

For silver staining, 40% methanol was used to pre-fix the polyacrylamide gels (30 min), followed by 10% ethanol (15 min twice). The special further fixing solution (200 mL: 38 mM benzene sulfonic acid, 57.6 mL ethanol in MilliQ H₂O) was used to fix the oligonucleotides (40 min), then the gels were washed three times (3.69 mM benzene sulfonic acid in 200 mL MilliQ H₂O). After fixing, the gels were immersed in silver-staining solution (11.8 mM silver nitrate, 4.4 mM benzene sulfonic acid, 260 μ L formaldehyde in 200mL MilliQ H₂O) for 40 min, washed with MilliQ H₂O (at least 2 min). Then, the gels were placed in development solution (87.4 mM Na₂CO₃, 0.126 mM sodium thiosulfate, 260 μ L formaldehyde in 200mL MilliQ H₂O) for 7 min, soaking in normal SDS-PAGE destaining buffer (30 min). Lastly, the gels were dried under vacuum at 60 °C (at least 1.5 hr). All steps after fixing were performed in glass vessels.

8.7 RNA unwinding assay

The duplex RNA substrates were formed using *in vitro* transcription. The short RNA oligonucleotides were synthesised using two synthetic DNA oligonucleotides (Table 8.1), one of 21 base pairs containing the T7 promoter sequence (ATG TAA TAC GAC TCA CTA TAG) (10 pmol) and the other containing the complementary T7 promoter sequence and the complementary DNA sequence for the RNA oligonucleotides (10 pmol). These were annealed by heating to 95 °C and slow cooling. RNA transcription was carried out in a total volume of 40 μ L using T7 RNA polymerase (20 U) in the presence of 0.5 mM NTP mix, 10 mM DTT, 0.1 mg mL⁻¹ BSA, RNasin (80 U) and either 0.25 mM GTP or [γ -³²P]-GTP (Amersham Biosciences, 400 Ci mmol⁻¹, 10 μ Ci) at 37 °C for 1 hr. The DNA was digested using RNase-free DNase (2U) for 15 min at 37 °C. Two RNA oligonucleotides (one of which was radiolabelled) were annealed to form the duplex RNA substrate by heating to 95 °C for 10 min and slow cooling and purified using a BioSpin-30 column. The unwinding assay was then performed in an identical manner to that of the DNA unwinding assay (see above) with the exception that a radiolabelled trapping strand was used to increase sensitivity. The RNA synthesis and assay were performed in triplicate for both substrates.

Materials and methods

In addition to obtaining RNA substrates by *in vitro* transcription, two 2'-ACE (acetyl) protected RNA oligonucleotides substrates (Dharmacon, USA) were also used in the RNA unwinding analysis. Deprotection buffer (0.1 mM acetic acid adjusted to pH 3.8 with TEMED) was added to the tubes containing RNA (400 μL per tube), vortexed, centrifuged and incubated at 60 °C for 30 min, followed by lyophilization. RNase-free water was added to dissolve the powder, and then the solution was desalted using an RNase-free BioSpin 30 column. The final concentration of RNA for analysis was 10 pmol μL^{-1} . The unwinding assay was then performed in an identical manner to that of the RNA unwinding assay.

8.8 Mutagenesis studies

The mutagenesis experiments were performed according to the instruction of the site-directed mutagenesis kit (Stratagene). Different complimentary HPLC-grade primers (Table 8.1) were used to obtain target DNA sequences by PCR. The final PCR products carrying the appropriate mutations were digested by Dpn I (10 U μL^{-1} , 1 hr), followed by transformation and a plasmid minprep. The target fragments from positive clones, confirmed by double digestion, were ligated into the original vector again after NcoI/XhoI double digestion in order to avoid mutations probably occurring in the vector. All mutants were sequenced in order to verify the mutations and the absence of additional mutations which might have been introduced by the PCR. The mutant proteins were used to conduct the in ATPase and nucleotide unwinding assays with the same methodology as those described above. Polyribonucleotides (50 or 200 $\mu\text{g mL}^{-1}$), single-stranded short DNA (10 pmol), double-stranded short DNA and RNA (10 pmol), and three cations (30 mM Mg^{2+} , 3 mM Mn^{2+} , and 3 mM Zn^{2+}) were also added to observe their effect.

8.9 Crystallization

8.9.1 Crystallization by vapour diffusion

Before crystallization, the protein was concentrated to a final concentration of about 10 mg mL^{-1} , filtered with a 0.22 μm filter unit, and then desalted with a Biospin-30 (BioRad) or a PD-10 desalting column (Amersham Biosciences). The sample concentration was measured by Bradford solution (BioRad) at OD₅₀₅. Initial trials for crystallization were set up using the hanging drop vapour diffusion technique with the Basic Crystal Screen Kit (Sigma-Aldrich). Subsequently, different combinations of precipitant (PEG 4000, PEG 8000, etc.), buffers, and salt solutions were used to screen the optimal conditions for crystallization.

Materials and methods

In the hanging drop method, 500 μL of the crystallization solution was placed in each of the reservoirs of a 4 $\times$ 6 well plate (VDX, Hampton Research). A 1 or 2 μL drop of protein solution with or without 5 mM MgCl_2 and with or without 5 mM ADPNP was placed on plastic coverslips (Molecular Dimensions, USA) and an equal volume of crystallization solution from each reservoir was added to the protein drop. The plate was incubated at 18 $^\circ\text{C}$ and was inspected for crystal formation at periodic intervals. The crystals were checked by X-ray diffraction.

8.9.2 Crystallization screening

If the crystals obtained from preliminary screening were not suitable for high resolution X-ray analysis then different additives (methanol, ethanol, DMSO, detergents, *etc.*) were mixed with the concentrated protein samples prior to setting up crystallisation screens. Two different temperatures (18 $^\circ\text{C}$ and 10 $^\circ\text{C}$) were tried. The influence of vibration on crystallization was checked by setting up identical trays inside or outside the incubator. In certain cases, 10-15 mer single-stranded DNA and RNA oligonucleotides (150-200 μM) were also added to the drop to see if they aided crystallization. Several different formulations screens (SaltRx, PEG/Ion, ammonium sulphate, *etc.*), prepared according to the instruction of Hampton Research, were conducted.

8.9.3 Seeding to improve crystal quality

Microseeding, macroseeding, and streak seeding were used to improve crystal quality and size. For macroseeding, a single crystal, free from twinning or other crystalline deformations, was selected, washed with well solution several times, and then transferred to fresh pre-equilibrated drop. For microseeding, a few small crystals or one large crystal was broken up to powder using a needle. A small amount (1 μL) of mother liquor containing the powder was added to an aliquot of well solution, making serial dilutions: 1:10, 1:100, 1:1000, *etc.*, and vortexed well. A suitable fibre (cat's whisker) was used in streak seeding. The whisker was stroked through a microcrystalline precipitate, and then stroked in a straight line through a fresh drop quickly.

8.10 Data collection

8.10.1 In-house data collection

Normally, crystals of reasonable size were picked out with a nylon loop (Hampton Research) and placed into a cryoprotection solution (30% glycerol + 70% well solution). The crystals

Materials and methods

were then mounted on the X-ray diffractometer (Rigaku MicroMax-007 X-ray generator and R-axis IV++ image plate detector, Rigaku, Japan/USA) at a low temperature (-160 °C or -170 °C) using cold dry nitrogen gas. The crystal to detector distance was normally set between 100 to 200 mm. CrystalClear was used to control the data collection. The oscillation angle was usually 1° and the exposure time for each image was 10-40 min (crystal dependant). Initial images were analyzed by MOSFLM to provide information about orientation (strategy). The total number of images for a full dataset was also calculated by the strategy command in MOSFLM.

Some crystals were mounted on the X-ray diffractometer by capillary at room temperature. The capillary was cut to the right length using special cutting stone, and attached to a pipette with a short piece of tubing. First, the protein crystal was sucked into the capillary. Then most of the mother liquor was removed from around the crystal, leaving only sufficient liquid to allow the crystal to adhere to the wall of the capillary. Lastly, a small amount of mother liquor was added on either side of the crystal (but not touching the crystal) in order to keep the humidity high. Both ends of the capillary were then sealed using wax. The capillary containing the crystal was placed onto the goniometer head and data collection was performed as described previously.

8.10.2 Synchrotron data collection

Data for the native helicase crystals were collected at 100 K at beamline X06SA at the Swiss Light Source, Villingen-PSI, Switzerland at a wavelength of 0.98 Å using a MarResearch CCD detector (MarResearch Charge Coupled Device). A short exposure time was used (1 second per image) and the alignment between the crystal and beamline was performed automatically. For the crystals soaked with different heavy atom solutions, the crystals were initially scanned using a fluorescence detector to determine the exact wavelengths (inflection wavelength, peak wavelength, and remote wavelength) necessary for the MAD experiments. At least 180 images for each dataset were collected.

8.11 Data processing

The reflections on the images were indexed and integrated using the program MOSFLM. Reflections from the diffraction images integrated by MOSFLM were scaled using the program SCALA (from the CCP4 suite of programs). Scaled intensities were sorted using SORTMTZ (CCP4 suite) and converted to a data set of F_{hkl} 's using the CCP4 program:

TRUNCATE (for converting intensities to F). The MNF and 'free R ' flags were incorporated into the reflection files using the CCP4 programs MTZMNF and CAD. The free R flag was assigned to a random 5% of reflections which acted as a test data set to provide an unbiased measure of goodness-of-fit.

8.12 Phase determination

Three different methods were used in an attempt to solve the phase problem necessary for the determination of the crystal structure. They are Multiple Isomorphous Replacement (MIR), Single-wavelength Anomalous Dispersion (SAD), and Multiple-wavelength Anomalous Dispersion (MAD).

8.12.1 Multiple Isomorphous Replacement (MIR)

Different heavy atom compounds [gold compounds: KAuCl_4 and KAuBr_4 , mercury compounds: $\text{Hg}(\text{OOCCH}_3)_2$, HgCl_2 , and CH_3HgCl , platinum compounds: K_2PtCl_4 , K_2PtCl_6 , and $\text{PtCl}_2(\text{H}_2\text{NCH}_2\text{CH}_2\text{CH}_2\text{NH}_2)$, samarium compound (SmCl_3), iridium compounds (K_2IrCl_4 and IrCl_3), iodide compounds (KI), rhenium compounds (K_2ReCl_6 , KReO_4), silver compound (AgNO_3)] at different concentrations (final concentration: 1 mM, 2 mM, 5mM, *etc.*) were used to soak the crystals. Different soaking times were also tried. The crystals were picked out and placed in a solution by mixing the mother liquor and heavy atom compounds stock solution, transferred into the cryoprotection solutions, which contained the heavy atom compound. An alternative method was to increase the concentration of the target heavy atom compound stepwise. Normally, the crystals were kept in the drop, a small amount of compound solution was added, and then same amount of liquid was taken out from the drop in a certain time (10 min). The two steps were repeated until the final concentration was near to the target concentration.

The datasets were used to create difference patterson maps using FFT. The positions of heavy atoms bound to amino acid residues were defined from the maps or the program RSPS. CAD was used to combine the MTZ files, which was followed by SCALEIT to scale the combined files. The combined, scaled structure factors files were used in MLPHARE to refine the heavy atom positions and obtain initial phases. The output MTZ files were used to obtain the map using FFT program.

8.12.2 Single-wavelength Anomalous Dispersion (SAD) and Multiple-wavelength Anomalous Dispersion (MAD)

The datasets collected at the peak wavelength were analyzed using single anomalous dispersion (SAD). After being processed to obtain the unique MTZ files using MTZMNF, several datasets from different heavy atom derivatives were combined together using CAD and then used in one program (SOLVE) running. Here anomalous difference was used in phase determination. The datasets were processed using SOLVE in order to obtain the number and positions of heavy atoms bound to the protein and basic phase information. Meanwhile, the positions were also refined using MLPHARE. For the MAD phase determination, the datasets with the two wavelengths (inflection and peak wavelength) were analyzed by SOLVE using the similar methods to SAD except two datasets for one heavy atom derivative.

Part II

8.13 Materials (II)

8.13.1 Chemicals

PPE and dipeptides (all with L-amino acids) ($\text{NH}_2\text{-Arg-Phe-OH}$, $\text{NH}_2\text{-Ala-Ala-OH}$, $\text{NH}_2\text{-Ala-Gly-OH}$, $\text{NH}_2\text{-Asp-Phe-OH}$) used in the soaking experiments were purchased from Sigma-Aldrich, Singapore. Dichloromethane (DCM), N,N-dimethylformamide (DMF), N,N-diisopropylethylamine (DIEA), trifluoroacetic acid (TFA), tris(isopropyl)silane (TIS), (benzotriazol-1-yloxy)-tris(dimethylamino) phosphonium hexafluorophosphate (BOP) used in peptide synthesis were from Merck, Fluka, or Sigma-Aldrich. Amino acids with 9-fluorenylmethoxycarbonyl (Fmoc) protection group, the Rink-amide-MBHA resin and Wang resin in peptide synthesis were ordered from GL Biochem (Shanghai) Ltd. Plates and coverslips for crystallization were from Hampton Research Corp or Molecular Dimensions, UK. Suc-Ala-Ala-Pro-Ala-pNA for kinetic analysis was from Bachem, Switzerland. All other biochemicals (usually of analytical grade or higher) were purchased from either Sigma-Aldrich, or Fluka Chemicals, or Merck, or Fisher.

8.13.2 Equipment

Microscope (Nikon SMZ1500) used in screening crystals was from Nikon, Japan. Incubators (SANYO MIR-253) for plates were purchased from SANYO, Japan. The UV-Visible spectrophotometer (Cary 300 Bio) was from Varian, USA. High Performance Liquid Chromatography (Agilent 1100 series) was from Agilent, USA. The protein and peptide C18 column (VYDACTM) was from W.R.GRACE & CO. USA. The Luna 5 μ C18 column was

from Phenomenex, USA. The ESI-MS machine (LCQ Deca XP Max) was from Thermo Electron Corporation, USA. Freeze-dryer (Freezemobile 25EL) for lyophilization was from Virtis, USA. Rigaku MicroMax-007 X-ray generator and R-axis IV++ image plate detector were from Rigaku, Japan/USA. The synchrotron data were collected from beamline X06SA in Swiss Light Source, Villigen-PSI, Switzerland. Vacuum concentrator was from Heidolph Instruments, Germany.

8.14 Peptide synthesis

8.14.1 Peptide with C-terminal amide

Synthesis of peptides with a C-terminal amide was achieved by the following solid phase peptide synthesis (SPPS) protocol. Rink-Amide-MBHA-resin (loading 0.3 mmol g^{-1}) was soaked in DCM (15-20 min), dried under vacuum, soaked in DMF (2 min), dried again, and then placed in a reaction vessel. For removing the Fmoc group, the resin was treated twice ($2 \text{ min} \times 1$, dry, and $20 \text{ min} \times 1$) by piperidine (20% v/v in DMF). The wash A procedure (DMF $2 \text{ min} \times 3$, DCM $2 \text{ min} \times 3$, methanol $2 \text{ min} \times 1$, DCM $2 \text{ min} \times 1$) was performed. A small quantity of resin was mixed with the ninhydrin detection solution (1,2,3-indanetrione) and heated to boiling point (110°C , 1-2 min) to check for the presence of free amine groups. Fmoc-amino acid (the C-terminal amino acid), BOP (equal mole), and DIEA (4 times mol) were dissolved in DMF (5 min), transferred to the vessel with the pre-soaked resin (in DMF, 2 min), and shaken for 45-60 min. The reaction was stopped by wash B (DMF $2 \text{ min} \times 1$, DCM $2 \text{ min} \times 1$, MeOH + DCM $2 \text{ min} \times 1$, DCM $2 \text{ min} \times 1$). The ninhydrin test was done to check the reaction efficiency again. Then the coupling process (deprotection, wash A, ninhydrin test, coupling new amino acid, wash B, ninhydrin test) was repeated. After the final coupling was finished, the product was cleaved from the resin with 4 mL solution (95% TFA, 2.5% H_2O , 2.5% TIS, 2-3 hr). Seven volumes of diethyl ether was added to the cleavage solution to precipitate the peptide, and then double-distilled water (adding acetonitrile if necessary) was used to dissolve the precipitate. If unprecipitated, the solution was dried with a vacuum concentrator and then double-distilled water was added. The peptide solution was then lyophilized to produce a powder.

8.14.2 Peptide with C-terminal carboxylate

A similar protocol was followed as that described previously except for two points: the Wang resin was used (loading 1.5 mmol g^{-1}); in the reaction of loading the first amino acid to resin, BOP and DIEA were replaced by dicyclohexylcarbodiimide (DCC) and

Materials and methods

4-(dimethylamino)pyridine (DMAP) and this reaction lasted overnight. The remaining coupling cycles were performed with same procedure. For the acetylation of the N-terminal residue, the resin was incubated with 5 mL solution (10% v/v acetic anhydride, 200 μ L DIEA in DCM) for 1 hr.

8.15 HPLC

8.15.1 HPLC purification

The lyophilized peptides were dissolved in ddH₂O (100 μ L), except that about 10% acetonitrile was added to highly hydrophobic peptides, and loaded onto a C18 column with buffer A (0.045% TFA in ddH₂O) at a flow-rate of 1 mL min⁻¹. Then the peptides were eluted with a gradient of 0%-40% buffer B in 30 min (90% acetonitrile + 0.04% TFA). For highly hydrophobic peptides, a gradient of 10%-80% buffer B was used. The fractions were identified by ESI-MS, and the peaks containing the target peptides were lyophilized and stored at -80 °C.

8.15.2 HPLC analysis

PPE and two peptides in a molar ratio of 1:50:50 were dissolved in 25 mM Tris (50 μ L pH 7.5), incubated at 25 °C or -15 °C for 2, 5, 10, 20, 25 min. At each time point, a 10 μ L aliquot was removed and added to 30 μ L acetone. Either BCM7 (YPFVEPI) or AcNPI was used as one peptide and either NH₂-Ala-Ala-OH, NH₂-Ala-Gly-OH, or NH₂-Lys-Ser-OH was used as the other peptide. Three different molar ratios (1:25:25, 1:100:100 and 1:200:200) and pH values (pH 5, pH 6 and pH 8.5) buffers (25 mM sodium acetate, HEPES and Tris respectively) were also tried in the study. The mixture was vortexed well and centrifuged at maximum speed (13000 rpm, 5 min). The supernatant was added to 360 μ L ddH₂O, and lyophilized. The powder was dissolved in ddH₂O (100 μ L), loaded onto a Luna 5 μ C18 column, and eluted with a gradient of 0%-60% buffer B (90% acetonitrile + 0.04% TFA) in 40 min. All the peaks were collected and identified with ESI-MS.

8.16 ESI-MS identification and analysis

For ESI-MS identification, the peaks from the HPLC were lyophilized and dissolved in 20 μ L ddH₂O. Samples for the initial mass spectrometry experiments were prepared by incubating PPE (40 pmol μ L⁻¹), β -casomorphin-7 (BCM7, 200 pmol μ L⁻¹), and NH₂-Lys-Ser-OH/NH₂-Ala-Ala-OH (200 pmol μ L⁻¹) in MilliQ water (20 μ L) at 25 °C for 5 or 20 min, added to 20 μ L solution I (98% acetonitrile, 2% formic acid) or 20 μ L solution II (5 M ammonium

Materials and methods

acetate, pH 7.0). Meanwhile, three additional different molar ratios (1:10:10, 1:25:25, and 1:50:50) and pH values (pH 5, pH 6 and pH 8.5) buffers (25 mM sodium acetate, HEPES and Tris respectively). All the samples were injected directly into the ESI needle (8 kV) with buffer (0.045% TFA in ddH₂O) at a flowrate of 300 $\mu\text{L min}^{-1}$. The mass spectra were deconvoluted by ProMass.

8.17 Kinetics

Enzyme assays were performed using a Cary 300 Bio UV/Vis spectrophotometer equipped with a thermostatted multi-cell transport system. The hydrolysis of the Suc-Ala-Ala-Pro-Ala-pNA substrates was measured at 405 nm and at a constant temperature of 25 °C in a total volume of 1 mL in 2 min²⁷³. Kinetic measurements for PPE in different pH values were made in seven different 0.1 M buffer solutions [sodium acetate (pH 4.6, pH 5.5), MES (pH 6.08, pH 6.5), HEPES (pH 7.0), Tris (pH 8.0, 8.5), CHES (pH 9.0, 9.6)] with 0.2 mM substrate. The amount of enzyme used for each assay was 0.5 μg of PPE.

For measuring the IC₅₀ of different inhibitors, duplicate to quadruplicate measurements of five different inhibitor concentrations (0.0 mM, 0.1 mM, 0.5mM, 1 mM, 4 mM) were recorded with 0.2 mM substrate. Pre-incubation of inhibitors and PPE (5 min) was also performed. IC₅₀ (half-maximal inhibition) values were calculated using Microsoft Excel.

8.18 Crystallization

8.18.1 Crystallization of native PPE

The hanging drop vapour diffusion method used for this work was adapted from Meyer *et al*²⁷². PPE (50 mg mL⁻¹) was dissolved in sodium acetate buffer (pH 5.0, 50 mM). This PPE solution (2 μL) was mixed with microseed solution (1 μL) and sodium sulphate solution (50 mM, 2 μL) to form drops (5 μL) that were suspended above sodium sulphate solution (50 mM, 400 μL). The microseed solution consisted of mother liquor containing finely ground PPE crystals (1 μL) added to 50 mM sodium sulphate solution (1 mL). The crystallisation trays were kept at a constant temperature of 22 °C. Within about 12 hours, this method gave between 2 and 20 well-formed PPE crystals in the majority of the drops. To freeze the crystals, a cryoprotectant solution was used consisting of sodium sulphate (25 mM), sodium acetate buffer (25 mM, pH 5.0), and glycerol (25% v/v). A soaking time of between 30 and 60 seconds was found to be optimum, minimising both ice formation and mosaicity increase.

8.18.2 Co-crystallization of PPE with peptides

BCM7 (35 mg mL⁻¹) was dissolved in sodium acetate buffer (pH 5.0, 50 mM), and then PPE (50 mg mL⁻¹) was added²⁰⁰. The mixture was centrifuged at maximum speed (13000 rpm) at 4 °C (5 min), filtered with a 0.22 µm filter unit. Hanging drops (5 µL) were constructed from 2 µL of the protein/peptide solution, 2 µL of sodium sulphate solution (50 mM), and 1 µL of microseed solution (as Section 8.18.1). After about 30 days at 22 °C, crystals were formed that had similar appearance to those of native PPE. To freeze the crystals, a cryoprotectant solution of 25 mM sodium sulphate, 25 mM sodium acetate, 25% (v/v) glycerol, and 15 mg mL⁻¹ BCM7, pH 5.0 was used. As with native PPE crystals, the optimum duration of soaking in the cryoprotectant solution was found to be between 30 and 60 seconds. For PPE co-crystallized with AcNPI, QPI, NPI, DPI, EPI, and MFLE, exactly same methodology was followed, except that crystals were formed in 24 hours. In all cases, the pH value of the sodium acetate solution was checked to ensure that it remained at 5.0 after addition of PPE and peptides.

8.19 Soaking experiments

Co-crystallized crystals (PPE:BCM7, PPE:AcNPI, PPE:NPI), and various dipeptides (all with L-amino acids) (NH₂-AA-OH, NH₂-AA-NH₂, NH₂-AG-OH, NH₂-KS-OH, NH₂-KA-NH₂, NH₂-DF-OH, NH₂-RF-OH, NH₂-KV-OH, NH₂-AS-OH) were used in the soaking experiments. The co-crystallized crystals were picked out and washed with 250 mM sodium acetate (pH 5.0) three times. The washed crystals were soaked in a saturated solution of 2 mg dipeptide powder (NH₂-AA-OH, NH₂-AG-OH, NH₂-DF-OH, NH₂-RF-OH) mixed in 0.25 M sodium acetate (pH 5.0, 5 µL) or a stepwise (replacing with 0.5 µL fresh peptide solution per 10 min) increase in dipeptide concentration (NH₂-KS-OH, NH₂-KA-NH₂, NH₂-AA-NH₂, NH₂-KV-OH, NH₂-AS-OH) to reach a final concentration of about 20-50 mg mL⁻¹. The crystals were then soaked in cryoprotectant solution [0.25 mM sodium acetate (pH 5.0), 25% (v/v) glycerol] and rapidly frozen by low temperature dry nitrogen gas (-140 °C).

8.20 pH-jump experiments

The crystals obtained from the soaking experiments were then briefly soaked in cryoprotectant solution [25% (v/v) glycerol, 0.25 M sodium acetate (pH 5.0)] and rapidly frozen in liquid nitrogen. The crystals were then mounted on the X-ray generator and held at -140 °C using a dry nitrogen gas stream (X-Stream, Rigaku). After collecting a full dataset,

Materials and methods

the crystal was removed from the nitrogen gas stream and immediately placed in the pH jump solution [25% (v/v) glycerol, 0.25 M CHES (pH 9.0)] for 30-60 seconds. If the crystal survived then it was rapidly transferred back to the nitrogen gas stream and refrozen at the same temperature (-140 °C). After another set of data were collected, this pH jump procedure was attempted a second time. If successful, a further dataset was collected.

8.21 Data collection

Most of the X-ray diffraction data were collected on an in-house X-ray diffractometer by using a Rigaku MicroMax-007 X-ray generator operating at 40 kV and 20 mA and a Rigaku R-axis IV++ image plate detector, except that data for the AcNPI:AA structure were collected at 100 K at beamline X06SA at the Swiss Light Source, Villingen-PSI, Switzerland at a wavelength of 0.98 Å using a MarResearch CCD detector. Data from the home X-ray machine were collected with an exposure of 2 min or 10 min and 1° oscillations using the CrystalClear program at -140 °C or -170 °C. 90 or 120 images were collected for one dataset. Data with 120 images from SLS were collected with an exposure of 1 second and 1° oscillations at -170 °C. The completeness (over 95%) of all the data was obtained by using the strategy analysis from MOSFLM.

8.22 Determination of crystal structures

All the data were indexed and integrated with the program MOSFLM, which is part of the CCP4 package (Collaborative Computational Project 4¹³⁹), and sorted, scaled, merged, and truncated using programs: SORTMTZ, SCALA, TRUNCATE from CCP4 package. Unique data were obtained by using MTZMNF from CCP4. The program MolRep was used to get initial phases with the native PPE model from 1HAZ (Protein Data Bank identification code). A total of 5% of the reflections in the entire dataset were randomly selected to provide a test set for the R_{free} calculation³⁰¹. First, the top solution from rotation function between a chosen upper and lower resolution (20.0 and 1.6 or 1.8 or 2.0 Å), and then the final solution was obtained from translation function. Using these parameters from final solution contained in the PDB model file by PDBSET program, a rigid refinement (resolution varies for different crystals) was done using Refmac5 program (CCP4 suite v.5). After it was found that the same result could be obtained with or without the MolRep step, the MolRep step was omitted in later structure determination. Further refinement with Refmac (using a maximum likelihood target) was performed to get R_{work} and R_{free} as low as possible. The TLS command was used in the refinement to achieve better R_{work} and R_{free} .

Materials and methods

For the tetrahedral intermediate structures and E:S complexes structures, additional topology parameters were introduced via a LINK entry (in Refmac5). During the latter stages of refinement, water molecules were added to improve the FOM (figure-of-merit) value to about 0.88 by using F_o-F_c difference maps. The PDB models and electronic density maps were displayed with the program O³⁰². F_o-F_c difference maps without peptides built in were also created. The water molecules and the protein model were checked manually using O. The overall b-factors of different chains were obtained using program Moleman2³⁰³. The quality of the final model was assessed by PROCHECK³⁰⁴ and WHATCHECK³⁰⁵. The final coordinates of the structures have been submitted to the Protein Data Bank. Figures showing the detailed structure of active sites were created with the program PyMOL³⁰⁶. To obtain stereopictures, two identical images were rotated -3° (left) and $+3^\circ$ (right) around the Y axis respectively. The labels were added using Adobe Illustrator CS software.

8.23 Comparison of different structures

Different PDB files were aligned using O with special commands: LSQ_EX, LSQ_MOL. The PDB models of various tetrahedral intermediate analogue structures were obtained from the Protein Data Bank (1GG6, 1GGD, 1GMC, 1GMD, 1JRS, 1JRT, 1MCV, 1PO1, 1SGC, 4EST, 5EST, 6EST, 6GCH, 7GCH). The difference between the E:S complex structure and the two pH jump structures was evaluated using the program ESCET²⁷⁸. The difference distance matrix maps of overall structure and active sites were created to identify the possible position shift. The distance between corresponding atoms was plotted based on the conformationally invariant part.

Appendices

Appendices

Table a Data collection statistics (SARS-CoV helicase)

	Native from in-house	Native from SLS	Ir Peak (K ₂ IrCl ₄)	Ir Inflection (K ₂ IrCl ₄)	Sm Peak (SmCl ₃)	Re peak (K ₂ ReCl ₆)
Wavelength (Å)	1.5418	0.9801	1.1052	1.1056	1.8456	1.1770
Resolution (Å)	4.20 (4.43)	3.06 (3.14)	4.25 (4.36)	6.80 (7.17)	4.22 (4.45)	4.22 (4.33)
<i>P</i> ₃₁ / <i>P</i> ₃₂ cell dimensions (Å)						
<i>a</i>	141.37	141.70	141.71	141.57	141.88	141.18
<i>b</i>	141.37	141.70	141.71	141.57	141.88	141.18
<i>c</i>	185.53	184.87	185.28	183.73	185.73	185.85
Total number of observations	58572	256374	143447	24392	162092	170598
Unique reflection set	29649	78444	29323	6801	30082	30507
Completeness (%)	98.8 (98.7)	99.8 (100)	99.9 (100)	96.1 (100)	99.9 (100)	99.9 (100)
$\langle I/\sigma_I \rangle$	7.1 (3.0)	9.3 (2.1)	10.1 (2.9)	7.7 (2.2)	10.3 (3.5)	7.5 (2.9)
* <i>R</i> _{merge} (%)	15.7 (29.7)	14.7 (79.7)	17.0 (46.4)	22.1 (43.3)	18.4 (53.6)	25.0 (50.6)

Values in parentheses are for the highest-resolution shell. * (see below).

Table b Data collection statistics (SARS-CoV helicase)

	Au Peak (KAuCl ₄)	Pt Peak (K ₂ PtCl ₄)	Hg Peak (HgCl ₂)
Wavelength (Å)	1.0367	1.0717	1.0068
Resolution (Å)	4.82 (4.95)	7.05 (7.43)	6.50 (6.85)
<i>P</i> ₃₁ / <i>P</i> ₃₂ cell dimensions (Å)			
<i>a</i>	143.80	143.28	143.88
<i>b</i>	143.80	143.28	143.88
<i>c</i>	186.71	188.08	185.98
Total number of observations	84217	25045	29675
Unique reflection set	20843	6393	8184
Completeness (%)	99.9 (100)	95.7 (100)	96.5 (96.3)
$\langle I/\sigma_I \rangle$	7.2 (2.1)	4.9 (2.0)	13.2 (2.4)
* <i>R</i> _{merge} (%)	27.1 (79.2)	28.2 (42.3)	9.7 (49.8)

Values in parentheses are for the highest-resolution shell. * *R*_{merge} and *R*_{sym} are essentially the average percentage errors of the intensity measurements and often used interchangeably. *R*_{sym} represents the differences in the symmetry related reflections on an image (in MOSFLM) or a single crystal. *R*_{merge} includes these definitions of *R*_{sym} plus any other sources of reflections. *R*_{merge} is usually used to describe the difference between two data sets or between the parts of the data sets from different crystals.

Appendices

Table c Data collection and refinement statistics (elastase)

	PPE:BCM7: RF	PPE:BCM7: DF	PPE:BCM7: KANH ₂	PPE:BCM7: KS	PPE:BCM7: KS 1st pH jump
Experimental condition	pH 5 (45 min)	pH 5 (30 min)	pH 5 (30 min)	pH 5 (60 min)	pH 9 (30 s)
Resolution (Å)	1.70 (1.79-1.70)	1.60 (1.69-1.60)	1.60 (1.69-1.60)	1.70 (1.79-1.70)	1.80 (1.90-1.80)
<i>P</i> 2 ₁ 2 ₁ 2 ₁ cell dimensions (Å)					
<i>a</i>	50.19	49.88	50.41	50.23	50.65
<i>b</i>	57.88	57.58	57.91	57.94	57.59
<i>c</i>	74.60	74.14	74.73	74.77	74.52
Total number of observations	82,936	93,777	125,012	86,085	70,197
Unique reflection set	24,541	28,611	29,238	24,646	20,837
Completeness (%)	99.9 (100)	99.0 (93.9)	99.0 (97.7)	99.9 (100)	99.9 (100)
$\langle I/\sigma_I \rangle$	22.9 (7.7)	30.7 (15.2)	28.7 (11.8)	20.6 (8.6)	20.3 (5.1)
<i>R</i> _{merge} (%)	3.9 (13.7)	2.9 (6.1)	3.6 (11.3)	4.3 (13.0)	4.6 (18.1)
Number of solvent molecules	128	129	129	100	128
Average <i>B</i> -factors					
Protein (Å ²)	20.60	25.07	22.89	19.36	30.06
Peptide S subsite (Å ²)	32.35	36.70	39.35	33.74	41.68
Peptide S' subsite (Å ²)	36.00	40.09	36.88	38.19	54.77
R.m.s. deviations					
Bond lengths (Å)	0.017	0.014	0.015	0.019	0.019
Bond angles (°)	1.593	1.481	1.508	1.717	1.746
<i>R</i> _{work} / <i>R</i> _{free} (%)	16.8 / 20.5	17.3 / 19.7	17.0 / 20.1	18.6 / 22.0	17.4 / 21.2
PDB code	2BD2	2BB4	2BD3	2BD4	2BD5
Figures	7.1 <i>a</i> , 7.2 <i>a</i>	7.2 <i>b</i>	7.2 <i>c</i> , 7.10 <i>b</i>	7.3 <i>a</i> , 7.10 <i>a</i>	7.4 <i>a</i>

Values in parentheses are for the highest-resolution shell.

Appendices

Table d Data collection and refinement statistics (elastase)

	PPE:BCM7: RF	PPE:BCM7: RF 1st pH jump	PPE:BCM7: RF 2nd pH jump	PPE:AcNPI: AA	PPE:NPI: AA	PPE:MFLE
Experimental condition	pH 5 (50 min)	pH 9 (30 s)	pH 9 (28 s)	pH 5 (30 min)	pH 5 (30 min)	pH 5
Resolution (Å)	1.60 (1.69-1.60)	1.70 (1.79-1.70)	1.90 (2.00-1.90)	1.80 (1.90-1.80)	1.70 (1.79-1.70)	1.60 (1.69-1.60)
<i>P</i> 2 ₁ 2 ₁ 2 ₁ cell dimensions (Å)						
<i>a</i>	50.04	50.54	51.38	50.51	50.46	50.08
<i>b</i>	57.70	57.78	57.59	57.67	57.66	57.57
<i>c</i>	74.36	74.76	74.76	73.87	74.23	74.15
Total number of observations	99,200	83,824	60,398	66,057	87,649	100,669
Unique reflection set	29,030	24,742	17,944	20,470	24,480	27,749
Completeness (%)	99.8 (99.5)	99.9 (100)	99.4 (99.0)	99.5 (100)	99.9 (99.9)	96.1 (91.9)
$\langle I/\sigma_I \rangle$	16.8 (4.0)	21.8 (5.3)	16.1 (3.0)	11.1 (2.4)	24.8 (8.2)	21.3 (4.8)
R_{merge} (%)	6.2 (20.4)	4.6 (16.9)	6.3 (32.8)	8.6 (46.0)	3.4 (14.4)	4.2 (20.8)
Number of solvent molecules	128	127	121	120	126	118
Average <i>B</i> -factor						
Protein (Å ²)	25.43	27.84	33.38	17.45	16.46	21.40
Peptide S subsite (Å ²)	38.44	35.97	-	37.55	-	37.71
Peptide S' subsite (Å ²)	40.11	39.78	-	41.76	35.89	38.17
R.m.s. deviations						
Bond lengths (Å)	0.016	0.019	0.012	0.019	0.013	0.012
Bond angles (°)	1.575	1.746	1.274	1.662	1.371	1.413
$R_{\text{work}} / R_{\text{free}}$ (%)	17.7 / 21.1	17.3 / 22.0	17.2 / 22.6	17.1 / 21.2	16.3 / 19.4	18.2 / 20.5
PDB code	2BD7	2BD8	2BD9	2BDA	2BDB	2H1U
Figures	7.2 <i>d</i>	7.4 <i>b</i>	7.4 <i>c</i>	7.7 <i>a</i>	7.7 <i>b</i>	7.8

Values in parentheses are for the highest-resolution shell.

Appendices

Table e Data collection and refinement statistics (elastase)

	PPE:NPI	PPE:DPI	PPE:EPI	PPE:QPI
Experimental condition	pH 5	pH 5	pH 5	pH 5
Resolution (Å)	1.80 (1.90-1.80)	1.70 (1.79-1.70)	1.60 (1.69-1.60)	1.70 (1.79-1.70)
<i>P</i> 2 ₁ 2 ₁ 2 ₁ cell dimensions (Å)				
<i>a</i>	50.15	50.16	50.20	50.26
<i>b</i>	57.87	57.84	57.83	57.79
<i>c</i>	74.53	74.47	74.55	74.48
Total number of observations	88,860	85610	99759	84835
Unique reflection set	19,783	24049	29116	11850
Completeness (%)	95.8 (93.1)	98.4 (96.6)	99.2 (95.3)	95.5 (92.3)
$\langle I/\sigma_I \rangle$	34.6 (23.2)	16.7 (4.5)	18.8 (4.7)	16.7 (2.4)
R_{merge} (%)	3.0 (6.1)	5.9 (22.7)	5.1 (19.5)	6.4 (39.7)
Number of solvent molecules	124	128	128	128
Average <i>B</i> -factors				
Protein (Å ²)	12.91	22.17	22.9	26.98
Peptide S subsite (Å ²)	-	-	-	-
Peptide S' subsite (Å ²)	-	-	-	-
R.m.s. deviations				
Bond lengths (Å)	0.012	0.018	0.014	0.012
Bond angles (°)	1.286	1.598	1.529	1.286
$R_{\text{work}} / R_{\text{free}}$ (%)	17.7 / 22.2	17.4 / 21.1	17.9 / 20.5	17.7 / 22.2
PDB code	2BDC	-	-	-
Figures	6.5 <i>B</i> , 7.10 <i>e</i>	6.5 <i>A</i>	6.6 <i>A</i>	6.6 <i>B</i>

Values in parentheses are for the highest-resolution shell.

References

References

1. Abdel-Monem, M., Durwald, H. & Hoffmann-Berling, H. Enzymic unwinding of DNA. 2. Chain separation by an ATP-dependent DNA unwinding enzyme. *Eur. J. Biochem.* **65**, 441-9 (1976).
2. Hall, M. C. & Matson, S. W. Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* **34**, 867-77 (1999).
3. Ray, B. K., Lawson, T. G., Kramer, J. C., Cladaras, M. H., Grifo, J. A., Abramson, R. D., Merrick, W. C. & Thach, R. E. ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J. Biol. Chem.* **260**, 7651-8 (1985).
4. Sancar, A. Mechanisms of DNA excision repair. *Science* **266**, 1954-6 (1994).
5. Hanawalt, P. C. Transcription-coupled repair and human disease. *Science* **266**, 1957-8 (1994).
6. Brennan, C. A., Dombroski, A. J. & Platt, T. Transcription termination factor Rho is an RNA-DNA helicase. *Cell* **48**, 945-52 (1987).
7. Lohman, T. M. & Bjornson, K. P. Mechanisms of helicase-catalyzed DNA unwinding. *Annu. Rev. Biochem.* **65**, 169-214 (1996).
8. Wassarman, D. A. & Steitz, J. A. RNA splicing. Alive with DEAD proteins. *Nature* **349**, 463-4 (1991).
9. Ellis, N. A. DNA helicases in inherited human disorders. *Curr. Opin. Genet. Dev.* **7**, 354-63 (1997).
10. Luking, A., Stahl, U. & Schmidt, U. The protein family of RNA helicases. *Crit. Rev. Biochem. Mol. Biol.* **33**, 259-96 (1998).
11. Singleton, M. R. & Wigley, D. B. Modularity and specialization in superfamily 1 and 2 helicases. *J. Bacteriol.* **184**, 1819-26 (2002).
12. Roman, L. J. & Kowalczykowski, S. C. Characterization of the helicase activity of the *Escherichia coli* RecBCD enzyme using a novel helicase assay. *Biochemistry* **28**, 2863-73 (1989).
13. Kadare, G. & Haenni, A. L. Virus-encoded RNA helicases. *J. Virol.* **71**, 2583-90 (1997).
14. Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C. & Sonenberg, N. Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol. Cell. Biol.* **10**, 1134-44 (1990).
15. Lee, C. G. RH70, a bidirectional RNA helicase, co-purifies with U1snRNP. *J. Biol. Chem.* **277**, 39679-83 (2002).
16. Huang, Y. & Liu, Z. R. The ATPase, RNA unwinding, and RNA binding activities of recombinant p68 RNA helicase. *J. Biol. Chem.* **277**, 12810-5 (2002).
17. Gorbalenya, A. E. & Koonin, E. V. Helicase: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**, 419-429 (1993).
18. Caruthers, J. M. & McKay, D. B. Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* **12**, 123-33 (2002).
19. Bird, L. E., Subramanya, H. S. & Wigley, D. B. Helicases: a unifying structural theme? *Curr. Opin. Struct. Biol.* **8**, 14-8 (1998).
20. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**, 4713-30 (1989).
21. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J.* **1**, 945-51 (1982).
22. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27-43 (1999).
23. Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**, 635-47 (1997).
24. Weng, Y., Czaplinski, K. & Peltz, S. W. Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol. Cell. Biol.* **16**, 5477-90 (1996).
25. Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C. & Ellenberger, T. Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell* **99**, 167-77 (1999).
26. Patel, S. S. & Picha, K. M. Structure and function of hexameric helicases. *Annu. Rev. Biochem.* **69**, 651-97 (2000).
27. Story, R. M. & Steitz, T. A. Structure of the recA protein-ADP complex. *Nature* **355**, 374-6 (1992).
28. Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621-8 (1994).

References

29. Washington, M. T., Rosenberg, A. H., Griffin, K., Studier, F. W. & Patel, S. S. Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding. *J. Biol. Chem.* **271**, 26825-34 (1996).
30. Korolev, S., Yao, N., Lohman, T. M., Weber, P. C. & Waksman, G. Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases. *Protein Sci.* **7**, 605-10 (1998).
31. Caruthers, J. M., Johnson, E. R. & McKay, D. B. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc. Natl. Acad. Sci. USA* **97**, 13080-5 (2000).
32. Lam, A. M., Keeney, D. & Frick, D. N. Two novel conserved motifs in the hepatitis C virus NS3 protein critical for helicase action. *J. Biol. Chem.* **278**, 44514-24 (2003).
33. Tanner, N. K., Cordin, O., Banroques, J., Doere, M. & Linder, P. The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol. Cell* **11**, 127-38 (2003).
34. Cordin, O., Tanner, N. K., Doere, M., Linder, P. & Banroques, J. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *Embo J.* **23**, 2478-87 (2004).
35. Liu, G. & Carstens, E. B. Site-directed mutagenesis of the AcMNPV p143 gene: effects on baculovirus DNA replication. *Virology* **253**, 125-36 (1999).
36. Zhang, G., Deng, E., Baugh, L. R., Hamilton, C. M., Maples, V. F. & Kushner, S. R. Conserved motifs II to VI of DNA helicase II from *Escherichia coli* are all required for biological activity. *J. Bacteriol.* **179**, 7544-50 (1997).
37. George, J. W., Brosh, R. M., Jr. & Matson, S. W. A dominant negative allele of the *Escherichia coli* UvrD gene encoding DNA helicase II. A biochemical and genetic characterization. *J. Mol. Biol.* **235**, 424-35 (1994).
38. Graves-Woodward, K. L., Gottlieb, J., Challberg, M. D. & Weller, S. K. Biochemical analyses of mutations in the HSV-1 helicase-primase that alter ATP hydrolysis, DNA unwinding, and coupling between hydrolysis and unwinding. *J. Biol. Chem.* **272**, 4623-30 (1997).
39. Walker, S. L., Wonderling, R. S. & Owens, R. A. Mutational analysis of the adeno-associated virus type 2 Rep68 protein helicase motifs. *J. Virol.* **71**, 6996-7004 (1997).
40. Dillingham, M. S., Soultanas, P., Wiley, P., Webb, M. R. & Wigley, D. B. Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc. Natl. Acad. Sci. USA* **98**, 8381-7 (2001).
41. Brosh, R. M., Jr. & Matson, S. W. Mutations in motif II of *Escherichia coli* DNA helicase II render the enzyme nonfunctional in both mismatch repair and excision repair with differential effects on the unwinding reaction. *J. Bacteriol.* **177**, 5612-21 (1995).
42. Dillingham, M. S., Soultanas, P. & Wigley, D. B. Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation. *Nucleic Acids Res.* **27**, 3310-7 (1999).
43. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75-84 (1999).
44. Hall, M. C., Ozsoy, A. Z. & Matson, S. W. Site-directed mutations in motif VI of *Escherichia coli* DNA helicase II result in multiple biochemical defects: evidence for the involvement of motif VI in the coupling of ATPase and DNA binding activities via conformational changes. *J. Mol. Biol.* **277**, 257-71 (1998).
45. Marintcheva, B. & Weller, S. K. Helicase motif Ia is involved in single-strand DNA-binding and helicase activities of the herpes simplex virus type 1 origin-binding protein, UL9. *J. Virol.* **77**, 2477-88 (2003).
46. Ziegelin, G., Niedenzu, T., Lurz, R., Saenger, W. & Lanka, E. Hexameric RSF1010 helicase RepA: the structural and functional importance of single amino acid residues. *Nucleic Acids Res.* **31**, 5917-29 (2003).
47. Soultanas, P. & Wigley, D. B. Site-directed mutagenesis reveals roles for conserved amino acid residues in the hexameric DNA helicase DnaB from *Bacillus stearothermophilus*. *Nucleic Acids Res.* **30**, 4051-60 (2002).
48. Hishida, T., Iwasaki, H., Yagi, T. & Shinagawa, H. Role of Walker motif A of RuvB protein in promoting branch migration of holliday junctions. Walker motif A mutations affect ATP binding, ATP hydrolyzing, and DNA binding activities of RuvB. *J. Biol. Chem.* **274**, 25335-42 (1999).
49. Peiris, J. S. *et al.* Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**, 1767-72 (2003).
50. Normile, D. & Enserink, M. SARS in China. Tracking the roots of a killer. *Science* **301**, 297-9 (2003).
51. Ruan, Y. J. *et al.* Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet* **361**, 1779-85 (2003).

References

52. Bossi, P., Guihot, A., Ansart, S. & Bricaire, F. Severe acute respiratory syndrome: update. *Rev. Med. Intern.* **24**, 672-80 (2003).
53. Peiris, J. S. *et al.* Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319-25 (2003).
54. Marra, M. A. *et al.* The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399-404 (2003).
55. Rota, P. A. *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394-9 (2003).
56. Drosten, C. *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967-76 (2003).
57. Pringle, C. R. Virus taxonomy--1999. The universal system of virus taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses during 1998. *Arch. Virol.* **144**, 421-9 (1999).
58. Snijder, E. J. *et al.* Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**, 991-1004 (2003).
59. Ng, M. L., Tan, S. H., See, E. E., Ooi, E. E. & Ling, A. E. Early events of SARS coronavirus infection in *vero* cells. *J. Med. Virol.* **71**, 323-31 (2003).
60. Ziebuhr, J. Molecular biology of severe acute respiratory syndrome coronavirus. *Curr. Opin. Microbiol.* **7**, 412-9 (2004).
61. Thiel, V. *et al.* Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305-15 (2003).
62. van Dinten, L. C., van Tol, H., Gorbalenya, A. E. & Snijder, E. J. The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion biogenesis. *J. Virol.* **74**, 5213-23 (2000).
63. Pelham, H. R. & Brown, D. D. A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc. Natl. Acad. Sci. USA* **77**, 4170-4 (1980).
64. Tanner, J. A., Watt, R. M., Chai, Y. B., Lu, L. Y., Lin, M. C., Peiris, J. S., Poon, L. L., Kung, H. F. & Huang, J. D. The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5' to 3' viral helicases. *J. Biol. Chem.* **278**, 39578-82 (2003).
65. Ivanov, K. A., Thiel, V., Dobbe, J. C., van der Meer, Y., Snijder, E. J. & Ziebuhr, J. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J. Virol.* **78**, 5619-32 (2004).
66. Guan, Y. *et al.* Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. *Lancet* **363**, 99-104 (2004).
67. Gonzalez, J. M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A. E. & Enjuanes, L. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch. Virol.* **148**, 2207-35 (2003).
68. Lohman, T. M. *Escherichia coli* DNA helicases: mechanisms of DNA unwinding. *Mol. Microbiol.* **6**, 5-14 (1992).
69. Wong, I. & Lohman, T. M. Allosteric effects of nucleotide cofactors on *Escherichia coli* Rep helicase-DNA binding. *Science* **256**, 350-5 (1992).
70. Lohman, T. M. Helicase-catalyzed DNA unwinding. *J. Biol. Chem.* **268**, 2269-72 (1993).
71. Porter, D. J. & Preugschat, F. Strand-separating activity of hepatitis C virus helicase in the absence of ATP. *Biochemistry* **39**, 5166-73 (2000).
72. Hill, T. L. & Tsuchiya, T. Theoretical aspects of translocation on DNA: adenosine triphosphatases and treadmilling binding proteins. *Proc. Natl. Acad. Sci. USA* **78**, 4796-800 (1981).
73. Yarranton, G. T. & Gefter, M. L. Enzyme-catalyzed DNA unwinding: studies on *Escherichia coli* Rep protein. *Proc. Natl. Acad. Sci. USA* **76**, 1658-62 (1979).
74. Ali, J. A. & Lohman, T. M. Kinetic measurement of the step size of DNA unwinding by *Escherichia coli* UvrD helicase. *Science* **275**, 377-80 (1997).
75. Ahnert, P. & Patel, S. S. Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate. *J. Biol. Chem.* **272**, 32267-73 (1997).
76. Jezewska, M. J., Rajendran, S., Bujalowska, D. & Bujalowski, W. Does single-stranded DNA pass through the inner channel of the protein hexamer in the complex with the *Escherichia coli* DnaB Helicase? Fluorescence energy transfer studies. *J. Biol. Chem.* **273**, 10515-29 (1998).
77. Subramanya, H. S., Bird, L. E., Brannigan, J. A. & Wigley, D. B. Crystal structure of a DExx box DNA helicase. *Nature* **384**, 379-83 (1996).
78. Enemark, E. J., Chen, G., Vaughn, D. E., Stenlund, A. & Joshua-Tor, L. Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus. *Mol. Cell* **6**, 149-58 (2000).

References

79. Enemark, E. J., Stenlund, A. & Joshua-Tor, L. Crystal structures of two intermediates in the assembly of the papillomavirus replication initiation complex. *Embo J.* **21**, 1487-96 (2002).
80. Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A. D., Le, H. V. & Weber, P. C. Structure of the hepatitis C virus RNA helicase domain. *Nat. Struct. Biol.* **4**, 463-7 (1997).
81. Cho, H. S., Ha, N. C., Kang, L. W., Chung, K. M., Back, S. H., Jang, S. K. & Oh, B. H. Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J. Biol. Chem.* **273**, 15045-52 (1998).
82. Bernstein, D. A., Zittel, M. C. & Keck, J. L. High-resolution structure of the *E. coli* RecQ helicase catalytic core. *Embo J.* **22**, 4910-21 (2003).
83. Singleton, M. R., Scaife, S., Raven, N. D. & Wigley, D. B. Crystallization and preliminary X-ray analysis of RecG, a replication-fork reversal helicase from *Thermotoga maritima* complexed with a three-way DNA junction. *Acta Crystallogr. D* **57**, 1695-6 (2001).
84. Machius, M., Henry, L., Palnitkar, M. & Deisenhofer, J. Crystal structure of the DNA nucleotide excision repair enzyme UvrB from *Thermus thermophilus*. *Proc. Natl. Acad. Sci. USA* **96**, 11717-22 (1999).
85. Theis, K., Chen, P. J., Skovvaga, M., Van Houten, B. & Kisker, C. Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *Embo J.* **18**, 6899-907 (1999).
86. Li, D. *et al.* Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* **423**, 512-8 (2003).
87. James, J. A., Escalante, C. R., Yoon-Robarts, M., Edwards, T. A., Linden, R. M. & Aggarwal, A. K. Crystal structure of the SF3 helicase from adeno-associated virus type 2. *Structure (Camb)* **11**, 1025-35 (2003).
88. James, J. A., Aggarwal, A. K., Linden, R. M. & Escalante, C. R. Structure of adeno-associated virus type 2 Rep40-ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases. *Proc. Natl. Acad. Sci. USA* **101**, 12455-60 (2004).
89. Niedenzu, T., Roleke, D., Bains, G., Scherzinger, E. & Saenger, W. Crystal structure of the hexameric replicative helicase RepA of plasmid RSF1010. *J. Mol. Biol.* **306**, 479-87 (2001).
90. Fass, D., Bogden, C. E. & Berger, J. M. Crystal structure of the N-terminal domain of the DnaB hexameric helicase. *Struct. Fold. Des.* **7**, 691-8 (1999).
91. Weigelt, J., Brown, S. E., Miles, C. S., Dixon, N. E. & Otting, G. NMR structure of the N-terminal domain of *E. coli* DnaB helicase: implications for structure rearrangements in the helicase hexamer. *Struct. Fold. Des.* **7**, 681-90 (1999).
92. Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**, 589-600 (2000).
93. Toth, E. A., Li, Y., Sawaya, M. R., Cheng, Y. & Ellenberger, T. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Mol. Cell* **12**, 1113-23 (2003).
94. Kato, M., Ito, T., Wagner, G., Richardson, C. C. & Ellenberger, T. Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis. *Mol. Cell* **11**, 1349-60 (2003).
95. Allison, T. J., Wood, T. C., Briercheck, D. M., Rastinejad, F., Richardson, J. P. & Rule, G. S. Crystal structure of the RNA-binding domain from transcription termination factor Rho. *Nat. Struct. Biol.* **5**, 352-6 (1998).
96. Bogden, C. E., Fass, D., Bergman, N., Nichols, M. D. & Berger, J. M. The structural basis for terminator recognition by the Rho transcription termination factor. *Mol. Cell* **3**, 487-93 (1999).
97. Rogers, G. W., Jr., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.* **276**, 30914-22 (2001).
98. Tye, B. K. & Sawyer, S. The hexameric eukaryotic MCM helicase: building symmetry from nonidentical parts. *J. Biol. Chem.* **275**, 34833-6 (2000).
99. Suzich, J. A., Tamura, J. K., Palmer-Hill, F., Warrenner, P., Grakoui, A., Rice, C. M., Feinstone, S. M. & Collett, M. S. Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *J. Virol.* **67**, 6152-8 (1993).
100. Bartenschlager, R. & Lohmann, V. Replication of hepatitis C virus. *J. Gen. Virol.* **81**, 1631-48 (2000).
101. Kwong, A. D., Kim, J. L. & Lin, C. in *Current Topics in Microbiology and Immunology*, Vol.242 The Hepatitis C Viruses (eds. Hagedron, C. H. & Rice, C. M.) 171-196 (Springer-Verlag, Berlin, 2000).
102. Borowski, P., Mueller, O., Niebuhr, A., Kalitzky, M., Hwang, L. H., Schmitz, H., Siwecka, M. A. & Kulikowsk, T. ATP-binding domain of NTPase/helicase as a target for hepatitis C antiviral therapy. *Acta Biochim. Pol.* **47**, 173-80 (2000).
103. Kim, J. L., Morgenstern, K. A., Griffith, J. P., Dwyer, M. D., Thomson, J. A., Murcko, M. A., Lin, C. & Caron, P. R. Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* **6**, 89-100 (1998).

References

104. Yao, N. & Weber, P. C. Helicase, a target for novel inhibitors of hepatitis C virus. *Antivir. Ther.* **3**, 93-7 (1998).
105. Barnard, D. L., Hubbard, V. D., Burton, J., Smee, D. F., Morrey, J. D., Otto, M. J. & Sidwell, R. W. Inhibition of severe acute respiratory syndrome-associated coronavirus (SARSCoV) by calpain inhibitors and β -D-N4-hydroxycytidine. *Antivir. Chem. Chemother.* **15**, 15-22 (2004).
106. Yamamoto, N. *et al.* HIV protease inhibitor nelfinavir inhibits replication of SARS-associated coronavirus. *Biochem. Biophys. Res. Commun.* **318**, 719-25 (2004).
107. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-4 (2003).
108. Sui, J. *et al.* Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc. Natl. Acad. Sci. USA* **101**, 2536-41 (2004).
109. Liu, S. *et al.* Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* **363**, 938-47 (2004).
110. Zhang, Y., Li, T. & Fu, L. Silencing SARS-CoV spike protein expression in cultured cells by RNA interference. *FEBS Lett.* **560**, 141-46 (2004).
111. Kao, R. Y. *et al.* Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetics. *Chem. Biol.* **11**, 1293-9 (2004).
112. Stevens, R. C. Design of high-throughput methods of protein production for structural biology. *Structure* **8**, R177-R185 (2000).
113. Edman, P. Chemistry of amino acids and peptides. *Annu. Rev. Biochem.* **28**, 69-96 (1959).
114. Gevaert, K. & Vandekerckhove, J. Protein identification methods in proteomics. *Electrophoresis* **21**, 1145-54 (2000).
115. Milewski, S. Glucosamine-6-phosphate synthase--the multi-facets enzyme. *Biochim. Biophys. Acta* **1597**, 173-92 (2002).
116. Chen, X. S., Casini, G., Harrison, S. C. & Garcea, R. L. Papillomavirus capsid protein expression in *Escherichia coli*: purification and assembly of HPV11 and HPV16 L1. *J. Mol. Biol.* **307**, 173-82 (2001).
117. Walter, S. Structure and function of the GroE chaperone. *Cell. Mol. Life Sci.* **59**, 1589-97 (2002).
118. Seybert, A., Hegyi, A., Siddell, S. G. & Ziebuhr, J. The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA* **6**, 1056-68 (2000).
119. Martins, A., Gross, C. H. & Shuman, S. Mutational analysis of vaccinia virus nucleoside triphosphate phosphohydrolase I, a DNA-dependent ATPase of the DEXH box family. *J. Virol.* **73**, 1302-8 (1999).
120. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75-84 (1999).
121. Dillingham, M. S., Soultanas, P., Wiley, P., Webb, M. R. & Wigley, D. B. Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc. Natl. Acad. Sci. USA* **98**, 8381-7 (2001).
122. Seybert, A., Posthuma, C. C., van Dinten, L. C., Snijder, E. J., Gorbalenya, A. E. & Ziebuhr, J. A complex zinc finger controls the enzymatic activity of Nidovirus helicases. *J. Virol.* **79**, 696-704 (2005).
123. Tanner, J. A., Watt, R. M., Chai, Y. B., Lu, L. Y., Lin, M. C., Peiris, J. S., Poon, L. L., Kung, H. F. & Huang, J. D. The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5' to 3' viral helicases. *J. Biol. Chem.* **278**, 39578-82 (2003).
124. Seybert, A., van Dinten, L. C., Snijder, E. J. & Ziebuhr, J. Biochemical characterization of the equine arteritis virus helicase suggests a close functional relationship between arterivirus and coronavirus helicases. *J. Virol.* **74**, 9586-93 (2000).
125. Tai, C. L., Chi, W. K., Chen, D. S. & Hwang, L. H. The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J. Virol.* **70**, 8477-84 (1996).
126. Kadare, G., David, C. & Haenni, A. L. ATPase, GTPase, and RNA binding activities associated with the 206-kilodalton protein of turnip yellow mosaic virus. *J. Virol.* **70**, 8169-74 (1996).
127. Rikonen, M., Peranen, J. & Kaariainen, L. ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsp2. *J. Virol.* **68**, 5804-10 (1994).
128. van Dinten, L. C., van Tol, H., Gorbalenya, A. E. & Snijder, E. J. The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion biogenesis. *J. Virol.* **74**, 5213-23 (2000).
129. Liu, J. L., Rigolet, P., Dou, S. X., Wang, P. Y. & Xi, X. G. The zinc finger motif of *Escherichia coli* RecQ is implicated in both DNA binding and protein folding. *J. Biol. Chem.* **279**, 42794-802 (2004).
130. Singleton, M. R. & Wigley, D. B. Modularity and specialization in superfamily 1 and 2 helicases. *J. Bacteriol.* **184**, 1819-26 (2002).

References

131. Cheng, W., Brendza, K. M., Gauss, G. H., Korolev, S., Waksman, G. & Lohman, T. M. The 2B domain of the *Escherichia coli* Rep protein is not required for DNA helicase activity. *Proc. Natl. Acad. Sci. USA* **99**, 16006-11 (2002).
132. Mackintosh, S. G., Lu, J. Z., Jordan, J. B., Harrison, M. K., Sikora, B., Sharma, S. D., Cameron, C. E., Raney, K. D. & Sakon, J. Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus. *J. Biol. Chem.* **281**, 3528-35 (2005).
133. Kyono, K., Miyashiro, M. & Taguchi, I. Characterization of ATPase activity of a hepatitis C virus NS3 helicase domain, and analysis involving mercuric reagents. *J. Biochem. (Tokyo)* **134**, 505-11 (2003).
134. Mcree, D. E. Practical Protein Crystallography (Academic press, San Diego, 1993).
135. Bergfors, T. Seeds to crystals. *J. Struct. Biol.* **142**, 66-76 (2003).
136. Hauptman, H. Phasing methods for protein crystallography. *Curr. Opin. Struct. Biol.* **7**, 672-80 (1997).
137. Abrahams, J. P. & De Graaff, R. A. New developments in phase refinement. *Curr. Opin. Struct. Biol.* **8**, 601-5 (1998).
138. Double, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523-530 (1997).
139. Bailey, S. programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-3 (1994).
140. Yeates, T. O. Detecting and overcoming crystal twinning. *Methods Enzymol.* **276**, 344-58 (1997).
141. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55** (Pt 4), 849-61 (1999).
142. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* **55**, 484-91 (1999).
143. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458-64 (2005).
144. Blow, D. M. The tortuous story of Asp ... His ... Ser: structural analysis of α -chymotrypsin. *Trends Biochem. Sci.* **22**, 405-408 (1997).
145. Rawlings, N. D. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **28**, 323-325 (2000).
146. Neurath, H. Evolution of proteolytic enzymes. *Science* **224**, 350-7 (1984).
147. Johnson, D. E. Noncaspase proteases in apoptosis. *Leukemia* **14**, 1695-703 (2000).
148. Joseph, K., Ghebrehiwet, B. & Kaplan, A. P. Activation of the kinin-forming cascade on the surface of endothelial cells. *Biol. Chem.* **382**, 71-5 (2001).
149. Coughlin, S. R. Thrombin signalling and protease-activated receptors. *Nature* **407**, 258-64 (2000).
150. Barros, C., Crosby, J. A. & Moreno, R. D. Early steps of sperm-egg interactions during mammalian fertilization. *Cell. Biol. Int.* **20**, 33-9 (1996).
151. Powers, J. C. & Harper, J. W. in *Proteinase Inhibitors* (eds. Barrett, A. J. & Salvesen, G.) 55-152 (Elsevier, Amsterdam, 1986).
152. Davidson, D. J., Higgins, D. L. & Castellino, F. J. Plasminogen activator activities of equimolar complexes of streptokinase with variant recombinant plasminogens. *Biochemistry* **29**, 3585-90 (1990).
153. Sim, R. B. & Laich, A. Serine proteases of the complement system. *Biochem. Soc. Trans.* **28**, 545-50 (2000).
154. Collen, D. & Lijnen, H. R. The fibrinolytic system in man. *Crit. Rev. Oncol. Hematol.* **4**, 249-301 (1986).
155. Gardiner, P. J. Neutrophil elastase inhibitors. *Eur. Respir. Rev.* **12**, 373-374 (2002).
156. Kraut, J. Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.* **46**, 331-58 (1977).
157. Blow, D. M. The study of α -chymotrypsin by x-ray diffraction. The Third CIBA Medal Lecture. *Biochem. J.* **112**, 261-8 (1969).
158. Sigler, P. B., Blow, D. M., Matthews, B. W. & Henderson, R. Structure of crystalline α -chymotrypsin. II. A preliminary report including a hypothesis for the activation mechanism. *J. Mol. Biol.* **35**, 143-64 (1968).
159. Stroud, R. M., Kay, L. M. & R.E., D. The structure of bovine trypsin: electron density maps of the inhibited enzyme at 5 Å and at 2-7 Å resolution. *J. Mol. Biol.* **83**, 185-208 (1974).
160. Shotton, D. M. & Watson, H. C. Three-dimensional structure of tosyl-elastase. *Nature* **225**, 811-6 (1970).
161. Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501-24 (2002).
162. Pathy, L. Modular design of proteases of coagulation, fibrinolysis, and complement activation: implications for protein engineering and structure-function studies. *Methods Enzymol.* **222**, 10-21 (1993).
163. Wurtele, M., Hahn, M., Hilpert, K. & Hohne, W. Atomic resolution structure of native porcine pancreatic elastase at 1.1 Å. *Acta Crystallogr. D* **56** (Pt 4), 520-3 (2000).

References

164. Katona, G., Wilmouth, R. C., Wright, P. A., Berglund, G. I., Hajdu, J., Neutze, R. & Schofield, C. J. X-ray structure of a serine protease acyl-enzyme complex at 0.95 Å resolution. *J. Biol. Chem.* **277**, 21962-70 (2002).
165. Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. Three-dimensional structure of tosyl- α -chymotrypsin. *Nature* **214**, 652-6 (1967).
166. Blow, D. M., Birktoft, J. J. & Hartley, B. S. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* **221**, 337-40 (1969).
167. Markley, J. L. & Porubcan, M. A. The charge-relay system of serine proteinases: proton magnetic resonance titration studies of the four histidines of porcine trypsin. *J. Mol. Biol.* **102**, 487-509 (1976).
168. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. *J. Mol. Biol.* **89**, 73-101 (1974).
169. Dodson, G. & Wlodawer, A. Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**, 347-52 (1998).
170. Wright, C. S., Alden, R. A. & Kraut, J. Structure of subtilisin BPN' at 2.5 Å resolution. *Nature* **221**, 235-42 (1969).
171. Drenth, J., Hol, W. G., Jansonius, J. N. & Koekoek, R. Subtilisin Novo. The three-dimensional structure and its comparison with subtilisin BPN'. *Eur. J. Biochem.* **26**, 177-81 (1972).
172. Schaffer, N. K., May, S. C., Jr. & Summerson, W. H. Serine phosphoric acid from diisopropylphosphoryl chymotrypsin. *J. Biol. Chem.* **202**, 67-76 (1953).
173. Oosterbaan, R. A., Kunst, P., Van Rotterdam, J. & Cohen, J. A. The reaction of chymotrypsin and diisopropylphosphorofluoridate. II. The structure of two DP-substituted peptides from chymotrypsin-DP. *Biochim. Biophys. Acta* **27**, 556-63 (1958).
174. Schoellmann, G. & Shaw, E. Direct evidence for the presence of histidine in the active center of chymotrypsin. *Biochemistry* **2**, 252-5 (1963).
175. Henderson, R. Structure of crystalline α -chymotrypsin. IV. The structure of indoleacryloyl- α -chymotrypsin and its relevance to the hydrolytic mechanism of the enzyme. *J. Mol. Biol.* **54**, 341-54 (1970).
176. Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T. & Kraut, J. X-ray crystallographic study of boronic acid adducts with subtilisin BPN' (Novo). A model for the catalytic transition state. *J. Biol. Chem.* **250**, 7120-6 (1975).
177. Bone, R., Sampson, N. S., Bartlett, P. A. & Agard, D. A. Crystal structures of α -lytic protease complexes with irreversibly bound phosphonate esters. *Biochemistry* **30**, 2263-72 (1991).
178. Schick, J., Verspohl, R., Kern, H. & Scheele, G. Two distinct adaptive responses in the synthesis of exocrine pancreatic enzymes to inverse changes in protein and carbohydrate in the diet. *Am. J. Physiol.* **247**, G611-6 (1984).
179. Schwert, G. W. Proteolytic enzymes. *Annu. Rev. Biochem.* **24**, 83-112 (1955).
180. Brayer, G. D., Delbaere, L. T. & James, M. N. Molecular structure of the α -lytic protease from *Myxobacter* 495 at 2.8 Å resolution. *J. Mol. Biol.* **131**, 743-75 (1979).
181. Huber, R. & Bode, W. Structural basis of the activation and action of trypsin. *Acc. Chem. Res.* **11**, 114-122 (1978).
182. Gertler, A. & Birk, Y. Isolation and characterization of porcine proelastase. *Eur. J. Biochem.* **12**, 170-6 (1970).
183. Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-62 (1967).
184. Bode, W., Meyer, E., Jr. & Powers, J. C. Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry* **28**, 1951-63 (1989).
185. Perona, J. J. & Craik, C. S. Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337-60 (1995).
186. Czapinska, H. & Otlewski, J. Structural and energetic determinants of the S1-site specificity in serine proteases. *Eur. J. Biochem.* **260**, 571-95 (1999).
187. Blow, D. M. in *The Enzymes* (ed. Boyer, P. D.) 231-248 (Boca Raton, 1971).
188. Krieger, M., Kay, L. M. & Stroud, R. M. Structure and specific binding of trypsin: comparison of inhibited derivatives and a model for substrate binding. *J. Mol. Biol.* **83**, 209-30 (1974).
189. Waugh, S. M., Harris, J. L., Fletterick, R. & Craik, C. S. The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its specificity. *Nat. Struct. Biol.* **7**, 762-5 (2000).
190. Branden, C. & Tooze, J. *Introduction to protein structure* (Garland Publishing, Inc, New York, 1991).
191. Polgar, L. in *Hydrolytic Enzymes* (eds. Brocklehurst, K. & Neuberger, A.) 159-200 (Elsevier Science, 1987).

References

192. James, M. N., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. & Bauer, C. A. Structures of product and inhibitor complexes of *Streptomyces griseus* protease A at 1.8 Å resolution. A model for serine protease catalysis. *J. Mol. Biol.* **144**, 43-88 (1980).
193. Bolognesi, M., Gatti, G., Menagatti, E., Guarneri, M., Marquart, M., Papamokos, E. & Huber, R. Three-dimensional structure of the complex between pancreatic secretory trypsin inhibitor (Kazal type) and trypsinogen at 1.8 Å resolution. Structure solution, crystallographic refinement and preliminary structural interpretation. *J. Mol. Biol.* **162**, 839-68 (1982).
194. Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry* **22**, 4420-33 (1983).
195. Turk, D., Guncar, G., Podobnik, M. & Turk, B. Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.* **379**, 137-47 (1998).
196. Bauer, C. A., Brayer, G. D., Sielecki, A. R. & James, M. N. Active site of α -lytic protease: enzyme-substrate interactions. *Eur. J. Biochem.* **120**, 289-94 (1981).
197. Schellenberger, V., Turck, C. W. & Rutter, W. J. Role of the S' subsites in serine protease catalysis. Active-site mapping of rat chymotrypsin, rat trypsin, α -lytic protease, and cercarial protease from *Schistosoma mansoni*. *Biochemistry* **33**, 4251-7 (1994).
198. Antonini, E., Ascenzi, P., Bolognesi, M., Gatti, G., Guarneri, M. & Menegatti, E. Interaction between serine (pro)enzymes, and Kazal and Kunitz inhibitors. *J. Mol. Biol.* **165**, 543-58 (1983).
199. Johnson, C. H. & Knowles, J. R. The binding of inhibitors to α -chymotrypsin. *Biochem. J.* **101**, 56-62 (1966).
200. Wilmouth, R. C., Clifton, I. J., Robinson, C. V., Roach, P. L., Aplin, R. T., Westwood, N. J., Hajdu, J. & Schofield, C. J. Structure of a specific acyl-enzyme complex formed between β -casomorphin-7 and porcine pancreatic elastase. *Nat. Struct. Biol.* **4**, 456-62 (1997).
201. Singer, P. T., Smalas, A., Carty, R. P., Mangel, W. F. & Sweet, R. M. The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. *Science* **259**, 669-73 (1993).
202. Wilmouth, R. C., Edman, K., Neutze, R., Wright, P. A., Clifton, I. J., Schneider, T. R., Schofield, C. J. & Hajdu, J. X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. *Nat. Struct. Biol.* **8**, 689-94 (2001).
203. Frey, P. A., Whitt, S. A. & Tobin, J. B. A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* **264**, 1927-30 (1994).
204. Rodgers, P. S., Goaman, L. C. & Blow, D. M. On the correlation between three-dimensional structure and reactivity for a series of locked substrates of chymotrypsin. *J. Am. Chem. Soc.* **98**, 6690-5 (1976).
205. Robillard, G. & Shulman, R. G. High resolution nuclear magnetic resonance study of the histidine--aspartate hydrogen bond in chymotrypsin and chymotrypsinogen. *J. Mol. Biol.* **71**, 507-11 (1972).
206. Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T. & Xuong, N. H. Chymotrypsinogen: 2.5 Å crystal structure, comparison with α -chymotrypsin, and implications for zymogen activation. *Biochemistry* **9**, 1997-2009 (1970).
207. Hunkapiller, M. W., Smallcombe, S. H., Witaker, D. R. & Richards, J. H. Ionization behavior of the histidine residue in the catalytic triad of serine proteases. Mechanistic implications. *J. Biol. Chem.* **248**, 8306-8 (1973).
208. Hunkapiller, M. W., Smallcombe, S. H., Whitaker, D. R. & Richards, J. H. Carbon nuclear magnetic resonance studies of the histidine residue in α -lytic protease. Implications for the catalytic mechanism of serine proteases. *Biochemistry* **12**, 4732-43 (1973).
209. Bachovchin, W. W. Confirmation of the assignment of the low-field proton resonance of serine proteases by using specifically nitrogen-15 labeled enzyme. *Proc. Natl. Acad. Sci. USA* **82**, 7948-51 (1985).
210. Lin, J., Westler, W. M., Cleland, W. W., Markley, J. L. & Frey, P. A. Fractionation factors and activation energies for exchange of the low barrier hydrogen bonding proton in peptidyl trifluoromethyl ketone complexes of chymotrypsin. *Proc. Natl. Acad. Sci. USA* **95**, 14664-8 (1998).
211. Liang, T. C. & Abeles, R. H. Complex of α -chymotrypsin and N-acetyl-L-leucyl-L-phenylalanyl trifluoromethyl ketone: structural studies with NMR spectroscopy. *Biochemistry* **26**, 7603-8 (1987).
212. Bachovchin, W. W., Wong, W. Y., Farr-Jones, S., Shenvi, A. B. & Kettner, C. A. Nitrogen-15 NMR spectroscopy of the catalytic-triad histidine of a serine protease in peptide boronic acid inhibitor complexes. *Biochemistry* **27**, 7689-97 (1988).
213. Cleland, W. W. & Kreevoy, M. M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **264**, 1887-90 (1994).
214. Warshel, A., Papazyan, A. & Kollman, P. A. On low-barrier hydrogen bonds and enzyme catalysis. *Science* **269**, 102-6 (1995).

References

215. Ash, E. L., Sudmeier, J. L., De Fabo, E. C. & Bachovchin, W. W. A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment. *Science* **278**, 1128-32 (1997).
216. Fuhrmann, C. N., Kelch, B. A., Ota, N. & Agard, D. A. The 0.83 Å resolution crystal structure of α -lytic protease reveals the detailed structure of the active site and identifies a source of conformational strain. *J. Mol. Biol.* **338**, 999-1013 (2004).
217. Pauling, L. The Architecture of Molecules. *Proc. Natl. Acad. Sci. USA* **51**, 977-84 (1964).
218. Powers, J. C. Reaction of serine proteases with halomethyl ketones. *Methods Enzymol.* **46**, 197-208 (1977).
219. Bone, R., Frank, D., Kettner, C. A. & Agard, D. A. Structural analysis of specificity: α -lytic protease complexes with analogues of reaction intermediates. *Biochemistry* **28**, 7600-9 (1989).
220. Brady, K., Wei, A. Z., Ringe, D. & Abeles, R. H. Structure of chymotrypsin-trifluoromethyl ketone inhibitor complexes: comparison of slowly and rapidly equilibrating inhibitors. *Biochemistry* **29**, 7600-7 (1990).
221. O'Sullivan, D. B., O'Connell, T. P., Mahon, M. M., Koenig, A., Milne, J. J., Fitzpatrick, T. B. & Malthouse, J. P. ^{13}C NMR study of how the oxyanion pK_a values of subtilisin and chymotrypsin tetrahedral adducts are affected by different amino acid residues binding in enzyme subsites S1-S4. *Biochemistry* **38**, 6187-94 (1999).
222. Warshel, A. Dynamics of enzymatic reactions. *Proc. Natl. Acad. Sci. USA* **81**, 444-8 (1984).
223. Weiner, H., White, W. N., Hoare, D. G. & Koshland, D. E., Jr. The formation of anhydrochymotrypsin by removing the elements of water from the serine at the active site. *J. Am. Chem. Soc.* **88**, 3851-9 (1966).
224. Nakagawa, Y. & Bender, M. L. Methylation of histidine-57 in α -chymotrypsin by methyl *p*-nitrobenzenesulfonate. A new approach to enzyme modification. *Biochemistry* **9**, 259-67 (1970).
225. Craik, C. S., Rocznik, S., Largman, C. & Rutter, W. J. The catalytic role of the active site aspartic acid in serine proteases. *Science* **237**, 909-13 (1987).
226. Sprang, S., Standing, T., Fletterick, R. J., Stroud, R. M., Finer-Moore, J., Xuong, N. H., Hamlin, R., Rutter, W. J. & Craik, C. S. The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. *Science* **237**, 905-9 (1987).
227. Corey, D. R. & Craik, C. S. An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *J. Am. Chem. Soc.* **114**, 1784-1790 (1992).
228. Carter, P. & Wells, J. A. Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564-8 (1988).
229. Carter, P. & Wells, J. A. Engineering enzyme specificity by "substrate-assisted catalysis". *Science* **237**, 394-9 (1987).
230. Carter, P., Abrahmsen, L. & Wells, J. A. Probing the mechanism and improving the rate of substrate-assisted catalysis in subtilisin BPN'. *Biochemistry* **30**, 6142-8 (1991).
231. Corey, D. R., Willett, W. S., Coombs, G. S. & Craik, C. S. Trypsin specificity increased through substrate-assisted catalysis. *Biochemistry* **34**, 11521-7 (1995).
232. Bryan, P., Pantoliano, M. W., Quill, S. G., Hsiao, H. Y. & Poulos, T. Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc. Natl. Acad. Sci. USA* **83**, 3743-5 (1986).
233. Rao, S. N., Singh, U. C., Bash, P. A. & Kollman, P. A. Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *Nature* **328**, 551-4 (1987).
234. Braxton, S. & Wells, J. A. The importance of a distal hydrogen bonding group in stabilizing the transition state in subtilisin BPN'. *J. Biol. Chem.* **266**, 11797-800 (1991).
235. Caplow, M. Chymotrypsin catalysis. Evidence for a new intermediate. *J. Am. Chem. Soc.* **91**, 3639-45 (1969).
236. Hunkapiller, M. W., Forgac, M. D. & Richards, J. H. Mechanism of action of serine proteases: tetrahedral intermediate and concerted proton transfer. *Biochemistry* **15**, 5581-8 (1976).
237. Matheson, N. R., van Halbeek, H. & Travis, J. Evidence for a tetrahedral intermediate complex during serpin-proteinase interactions. *J. Biol. Chem.* **266**, 13489-91 (1991).
238. Imperiali, B. & Abeles, R. H. Inhibition of serine proteases by peptidyl fluoromethyl ketones. *Biochemistry* **25**, 3760-7 (1986).
239. Powers, J. C. in *In Chemistry and Biology of Amino Acids, Peptides and Proteins* (ed. Weinstein, B.) 66-178 (Marcel Dekker, New York, 1977).
240. Kettner, C. A. & Shenvi, A. B. Inhibition of the serine proteases leukocyte elastase, pancreatic elastase, cathepsin G, and chymotrypsin by peptide boronic acids. *J. Biol. Chem.* **259**, 15106-14 (1984).
241. Harel, M., Su, C. T., Frolow, F., Silman, I. & Sussman, J. L. γ -chymotrypsin is a complex of α -chymotrypsin with its own autolysis products. *Biochemistry* **30**, 5217-25 (1991).
242. Yennawar, N. H., Yennawar, H. P. & Farber, G. K. X-ray crystal structure of γ -chymotrypsin in hexane. *Biochemistry* **33**, 7326-36 (1994).

References

243. Werb, Z., Banda, M. J., McKerrow, J. H. & Sandhaus, R. A. Elastases and elastin degradation. *J. Invest. Dermatol.* **79 Suppl 1**, 154s-159s (1982).
244. Bieth, J. G. in Regulation of Matrix Accumulation (ed. Mechan, R. D.) 217-320 (Academic Press, New York, 1986).
245. Sandberg, L. B., Bruenger, E. & Cleary, E. G. Tropoelastin purification: improvements using enzyme inhibitors. *Anal. Biochem.* **64**, 249-54 (1975).
246. Balo, J. & Banga, I. Elastase and elastase-inhibitor. *Nature* **164**, 491 (1949).
247. Janoff, A. & Scherer, J. Mediators of inflammation in leukocyte lysosomes. IX. Elastinolytic activity in granules of human polymorphonuclear leukocytes. *J. Exp. Med.* **128**, 1137-55 (1968).
248. Geokas, M. C., Rinderknecht, H., Swanson, V. & Haverback, B. J. The role of elastase in acute hemorrhagic pancreatitis in man. *Lab. Invest.* **19**, 235-239 (1968).
249. Menninger, H., Putzier, R., Mohr, W., Hering, B. & Mierau, H. D. in Biological Functions of Proteases (eds. Holtzer, H. & Tshesche, H.) 196-206 (Springer, Berlin, 1979).
250. Virca, G. D., Metz, G. & Schnebli, H. P. Similarities between human and rat leukocyte elastase and cathepsin G. *Eur. J. Biochem.* **144**, 1-9 (1984).
251. Shotton, D. M. & Hartley, B. S. Evidence for the amino acid sequence of porcine pancreatic elastase. *Biochem. J.* **131**, 643-75 (1973).
252. Shotton, D. M., Hartley, B. S., Camerman, N., Hofmann, T., Nyburg, S. C. & Rao, L. Crystalline porcine pancreatic elastase. *J. Mol. Biol.* **32**, 155-6 (1968).
253. Thompson, R. C. & Blout, E. R. Evidence for an extended active center in elastase. *Proc. Natl. Acad. Sci. USA* **67**, 1734-40 (1970).
254. Atlas, D., Levit, S., Schechter, I. & Berger, A. On the active site of elastase: Partial mapping by means of specific peptide substrates. *FEBS Lett.* **11**, 281-283 (1970).
255. Atlas, D. The active site of porcine elastase. *J. Mol. Biol.* **93**, 39-53 (1975).
256. McRae, B., Nakajima, K., Travis, J. & Powers, J. C. Studies on reactivity of human leukocyte elastase, cathepsin G, and porcine pancreatic elastase toward peptides including sequences related to the reactive site of α 1-protease inhibitor (α 1-antitrypsin). *Biochemistry* **19**, 3973-8 (1980).
257. Yasutake, A. & Powers, J. C. Reactivity of human leukocyte elastase and porcine pancreatic elastase toward peptide 4-nitroanilides containing model desmosine residues. Evidence that human leukocyte elastase is selective for cross-linked regions of elastin. *Biochemistry* **20**, 3675-9 (1981).
258. Hughes, D. L., Sieker, L. C., Bieth, J. & Dimicoli, J. L. Crystallographic study of the binding of a trifluoroacetyl dipeptide anilide inhibitor with elastase. *J. Mol. Biol.* **162**, 645-58 (1982).
259. Renaud, A., Lestienne, P., Hughes, D. L., Bieth, J. G. & Dimicoli, J. L. Mapping of the S' subsites of porcine pancreatic and human leukocyte elastases. *J. Biol. Chem.* **258**, 8312-6 (1983).
260. Stepanov, V. M. Proteinases as catalysts in peptide synthesis. *Pure & Appl. Chem.* **68**, 1335-1339 (1996).
261. Petkov, D. D. & Stoineva, I. Nucleophile specificity in chymotrypsin peptide synthesis. *Biochem. Biophys. Res. Commun.* **118**, 317-23 (1984).
262. Oka, T. & Morihara, K. Trypsin as a catalyst for peptide synthesis. *J. Biochem. (Tokyo)* **82**, 1055-62 (1977).
263. Riechmann, L. & Kasche, V. Peptide synthesis catalyzed by the serine proteinases chymotrypsin and trypsin. *Biochim. Biophys. Acta* **830**, 164-72 (1985).
264. Grant, N. H. & Alburn, H. E. Acceleration of enzyme reactions in ice. *Nature* **212**, 194 (1966).
265. Schuster, M., Aaviksaar, A. & Jakubke, H. D. Enzyme-catalyzed peptide synthesis in ice. *Tetrahedron* **46**, 8093-8102 (1990).
266. Tougu, V., Meos, H., Haga, M., Aaviksaar, A. & Jakubke, H. D. Peptide synthesis by chymotrypsin in frozen solutions. Free amino acids as nucleophiles. *FEBS Lett.* **329**, 40-2 (1993).
267. Haensler, M., Wehofsky, N., Gerisch, S., Wissmann, J. D. & Jakubke, H. D. Reverse catalysis of elastase from porcine pancreas in frozen aqueous systems. *Biol. Chem.* **379**, 71-4 (1998).
268. Fields, G. B. & Noble, R. L. Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.* **35**, 161-214 (1990).
269. Erickson, B. E. HPLC and the ever popular LC/MS. *Anal. Chem.* **72**, 711A-716A (2000).
270. Tanaka, N., Kimata, K., Hosoya, K., Miyanishi, H. & Araki, T. Stationary phase effects in reversed-phase liquid chromatography. *J. Chromatogr. A* **656**, 265-287 (1993).
271. Edmonds, C. G. & Smith, R. D. Electrospray ionization mass spectrometry. *Methods Enzymol.* **193**, 412-31 (1990).
272. Meyer, E., Cole, G., Radhakrishnan, R. & Epp, O. Structure of native porcine pancreatic elastase at 1.65 Å resolution. *Acta Crystallogr. B* **44 (Pt 1)**, 26-38 (1988).
273. Wright, P. A., Wilmouth, R. C., Clifton, I. J. & Schofield, C. J. Kinetic and crystallographic analysis of complexes formed between elastase and peptides from β -casein. *Eur. J. Biochem.* **268**, 2969-74 (2001).

References

274. Morrison, J. F. & Walsh, C. T. The behavior and significance of slow-binding enzyme inhibitors. *Adv. Enzymol. Relat. Areas Mol. Biol.* **61**, 201-301 (1988).
275. Ay, J., Hilpert, K., Krauss, N., Schneider-Mergener, J. & Hohne, W. Structure of a hybrid squash inhibitor in complex with porcine pancreatic elastase at 1.8 Å resolution. *Acta Crystallogr. D* **59**, 247-54 (2003).
276. Schellenberger, V., Schellenberger, U., Mitin, Y. V. & Jakubke, H. D. Characterization of the S'-subsite specificity of porcine pancreatic elastase. *Eur. J. Biochem.* **179**, 161-3 (1989).
277. Burgi, H. B., Dunitz, J. D. & Shefter, E. Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J. Am. Chem. Soc.* **95**, 5065-5067 (1973).
278. Schneider, T. R. A genetic algorithm for the identification of conformationally invariant regions in protein molecules. *Acta Crystallogr. D* **58**, 195-208 (2002).
279. Clore, G. M., Gronenborn, A. M., Carlson, G. & Meyer, E. F. Stereochemistry of binding of the tetrapeptide acetyl-Pro-Ala-Pro-Tyr-NH₂ to porcine pancreatic elastase. Combined use of two-dimensional transferred nuclear Overhauser enhancement measurements, restrained molecular dynamics, X-ray crystallography and molecular modelling. *J. Mol. Biol.* **190**, 259-67 (1986).
280. Stoller, J. K. & Aboussouan, L. S. α 1-antitrypsin deficiency. *Lancet* **365**, 2225-36 (2005).
281. Wright, P. A., Rostom, A. A., Robinson, C. V. & Schofield, C. J. Mass spectrometry reveals elastase inhibitors from the reactive centre loop of α 1-antitrypsin. *Bioorg. Med. Chem. Lett.* **10**, 1219-21 (2000).
282. Takahashi, L. H., Radhakrishnan, R., Rosenfield, R. E., Jr. & Meyer, E. F., Jr. Crystallographic analysis of the inhibition of porcine pancreatic elastase by a peptidyl boronic acid: structure of a reaction intermediate. *Biochemistry* **28**, 7610-7 (1989).
283. Neidhart, D., Wei, Y., Cassidy, C., Lin, J., Cleland, W. W. & Frey, P. A. Correlation of low-barrier hydrogen bonding and oxyanion binding in transition state analogue complexes of chymotrypsin. *Biochemistry* **40**, 2439-47 (2001).
284. Delbaere, L. T. & Brayer, G. D. The 1.8 Å structure of the complex between chymostatin and *Streptomyces griseus* protease A. A model for serine protease catalytic tetrahedral intermediates. *J. Mol. Biol.* **183**, 89-103 (1985).
285. Bone, R., Shenvi, A. B., Kettner, C. A. & Agard, D. A. Serine protease mechanism: structure of an inhibitory complex of α -lytic protease and a tightly bound peptide boronic acid. *Biochemistry* **26**, 7609-14 (1987).
286. Kurinov, I. V. & Harrison, R. W. Two crystal structures of the leupeptin-trypsin complex. *Protein Sci.* **5**, 752-8 (1996).
287. Gerlt, J. A., Kreevoy, M. M., Cleland, W. & Frey, P. A. Understanding enzymic catalysis: the importance of short, strong hydrogen bonds. *Chem. Biol.* **4**, 259-67 (1997).
288. Warshel, A. & Russell, S. Theoretical correlation of structure and energetics in the catalytic reaction of trypsin. *J. Am. Chem. Soc.* **108**, 6569-6579 (1986).
289. Ash, E. L., Sudmeier, J. L., Day, R. M., Vincent, M., Torchilin, E. V., Haddad, K. C., Bradshaw, E. M., Sanford, D. G. & Bachovchin, W. W. Unusual ¹H NMR chemical shifts support (His) C^ε1...O=C H-bond: proposal for reaction-driven ring flip mechanism in serine protease catalysis. *Proc. Natl. Acad. Sci. USA* **97**, 10371-6 (2000).
290. Fodor, K., Harmat, V., Neutze, R., Szilagyi, L., Graf, L. & Katona, G. Enzyme:substrate hydrogen bond shortening during the acylation phase of serine protease catalysis. *Biochemistry* **45**, 2114-21 (2006).
291. Topf, M. & Richards, W. G. Theoretical studies on the deacylation step of serine protease catalysis in the gas phase, in solution, and in elastase. *J. Am. Chem. Soc.* **126**, 14631-41 (2004).
292. Wright, P. A., Wilmouth, R. C., Clifton, I. J. & Schofield, C. J. 'pH-jump' crystallographic analyses of γ -lactam-porcine pancreatic elastase complexes. *Biochem. J.* **351 Pt 2**, 335-40 (2000).
293. Stewart, D. E., Sarkar, A. & Wampler, J. E. Occurrence and role of *cis* peptide bonds in protein structures. *J. Mol. Biol.* **214**, 253-60 (1990).
294. Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., Jr. & Sacchettini, J. C. Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat. Struct. Biol.* **5**, 31-6 (1998).
295. Brauer, A. B., Domingo, G. J., Cooke, R. M., Matthews, S. J. & Leatherbarrow, R. J. A conserved *cis* peptide bond is necessary for the activity of Bowman-Birk inhibitor protein. *Biochemistry* **41**, 10608-15 (2002).
296. Topf, M., Varnai, P., Schofield, C. J. & Richards, W. G. Molecular dynamics simulations of the acyl-enzyme and the tetrahedral intermediate in the deacylation step of serine proteases. *Proteins* **47**, 357-69 (2002).
297. Deslongchamps, P. Stereoelectronic control in the cleavage of tetrahedral intermediates in the hydrolysis of esters and amides. *Tetrahedron* **31**, 2462-2636 (1975).

References

298. Bizzozero, S. A. & Dutler, H. Stereochemical aspects of peptide hydrolysis catalyzed by serine proteases of the chymotrypsin type. *Bioorg. Chem.* **10**, 46-62 (1981).
299. Knight, W. B., Chabin, R. & Green, B. Inhibition of human serine proteases by substituted 2-azetidinones. *Arch. Biochem. Biophys.* **296**, 704-8 (1992).
300. Wilmouth, R. C., Westwood, N. J., Anderson, K., Brownlee, W., Claridge, T. D., Clifton, I. J., Pritchard, G. J., Aplin, R. T. & Schofield, C. J. Inhibition of elastase by N-sulfonylaryl β -lactams: anatomy of a stable acyl-enzyme complex. *Biochemistry* **37**, 17506-13 (1998).
301. Brunger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475 (1992).
302. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47** (Pt 2), 110-9 (1991).
303. Kleywegt, G. J. & Jones, T. A. Model-building and refinement practice. *Methods Enzymol.* **277**, 208-230 (1997).
304. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291 (1993).
305. Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. Errors in protein structures. *Nature* **381**, 272 (1996).
306. Delano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA, USA, 2002).

Publications

1. **Liu B**, Schofield CJ, Wilmouth RC. Structural analyses on intermediates in serine protease catalysis. *J. Biol. Chem.* **281**, 24024-35 (2006).
2. **Liu B**, Lescar J, Wilmouth RC. The N-terminal zinc finger domain of the SARS coronavirus helicase is not required for ATPase and nucleotide unwinding activity. (Submitted)
3. Shi ML, Sundramurthy K, **Liu B**, Tan SM, Law SK & Lescar J. The crystal structure of the plexin-semaphorin-integrin domain/hybrid domain/I-EGF1 segment from the human integrin $\beta 2$ subunit at 1.8 Å resolution. *J. Biol. Chem.* **280**, 30586-93 (2005).