Structural Analysis of Bacterial and Viral Infections and a Cysteine Proteinase Legumain

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades

einer Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von   Wei Li
aus   Shandong, China

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikation

Hofmann, J., Heider, C., Li, W., Krausze, J., Roessle, M., Wilharm, G. (2013) Recombinant production of *Yersinia enterocolitica* pyruvate kinase isoenzymes PykA and PykF. Protein Expr. Purif. **88**, 243-247

Tagungsbeiträge

Li, W., Pietschmann, T., van den Heuvel, J., Heinz, D. W. : Structural analysis of hepatitis C virus (HCV) - host receptor interactions. (Poster) 1st Grad School Public Retreat, Quedlinburg, Germany (2010). 1st Poster prize.

Li, W., Krausze, J., Heider, C., Whiharm, G., Heinz, D. W. : Structural analysis of *Yersinia* Glycolytic enzymes - crystallization and X-ray structure analysis of PykF. (Talk) 14th Heart of Europe bio-Crystallography Meeting, Zagan, Poland (2011).

Li, W., Pietschmann, T., van den Heuvel, J., Heinz, D. W. Structural analysis of hepatitis C virus (HCV) - host receptor interactions. (Talk) 2nd HZI Grad School Public retreat, Goslar Hahnenklee, Germany (2011).

Li, W., Pietschmann, T., van den Heuvel, J., Krausze, J., Heider, C., Whiharm, G., Heinz, D. W. Structural analysis of hepatitis C virus (HCV) and *Yersinia* glycolytic enzymes. (Talk) 3rd HZI Grad School Annual Retreat, Bad Bevensen, Germany (2012).

Li, W., Krausze, J., Heider, C., Whiharm, G., Heinz, D. W. : Structural analysis of bacterial infection mechanisms. (Talk) 4th Annual Retreat, Goslar Hahnenklee, Germany (2013).

Li, W., Krausze, J., Heider, C., Whiharm, G., Heinz, D. W. : Structural analysis of *Yersinia* pyruvate kinase PykF. (Poster) 21st Annual Meeting of German Crystallographic Society, Freiberg, Germany (2013).

Li, W., Pietschmann, T., van den Heuvel, J., Heinz, D. W. : Structural analysis of hepatitis C virus (HCV) - host receptor interactions. (Poster) HZI Grad School Evaluation, Braunschweig, Germany (2013).

Li, W., Krausze, J., Heider, C., Whiharm, G., Heinz, D. W. : Structural analysis of *Yersinia* Glycolytic enzymes PykF and PykA. (Poster) International School on Biological Crystallization, Granada, Spain (2013).

Contents

Contents

Contents

IV

Abbreviations

AMP	Adenosine monophosphate
Å	Angstrom (1 Å = 0.1 nm)
ADP	Adenosine diphosphate
AEP	Asparagine endopeptidase
AP	Activation peptide
ATP	Adenosine triphosphate
CD-spectroscopy	Circular dichroism spectroscopy
DESY	Deutsches Elektronen-Synchrotron
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
F16BP	Fructose 1, 6-diphosphate
F26BP	Fructose 2, 6-diphosphate
GFP	Green fluorescent protein
HCl	Hydrochloric acid
HCV	Hepatitis C virus
Hepes	2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid
His$_6$	6 consecutive histidine residues
IPTG	Isopropyl-β-D-thiogalactopyranoside
kDa	Kilodalton
LB	Luria-Bertani broth
LSAM	legumain stabilization and activity modulation
M cells	Microfold cells
MR	Molecular replacement
MW	Molecular weight
NaCl	Sodium chloride
NAG	N-acetylglucosamine
NAM	N-acetylmuramic acid
NCS	Non-crystallographic symmetry
Ni-NTA	Nickel nitrilotriacetic acid
OD	Optical density
PBS	Phosphate buffered saline
PBST	PBS containing 0.1 % (v/v) TWEEN-20
PDB	Protein data bank
PEG	Polyethylene glycol

pI	Isoelectric point
PK	Pyruvate kinase
PKL	Pyruvate kinase liver isozyme
PKM	Pyruvate kinase muscle isozyme
PKR	Pyruvate kinase red cell isozyme
PVDF	Polyvinylidene difluoride
PykA	Pyruvate kinase A
PykF	Pyruvate kinase F
r.m.s.d.	Root mean square deviation
RNA	Ribonucleic acid
rpm	Rotations per minute
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
T3SS	Type-3 secretion system
TLS	Translation Libration Screw motion
Tris	Tris - (hydroxymethyl) aminomethane
XDS	X-ray Detector Software
Yop	*Yersinia* outer protein
YscM	Yop proteins translocation protein M

Zusammenfassung

Strukturanalyse von Yersinia-Pyruvat-Kinasen offenbart die allosterische Regulierung

Lebensmittelinfektionen werden in letzter Zeit immer häufiger. *Yersinia enterocolitica* ist eine der pathogenen Bakterien, die die durch Lebensmittel übertragene Krankheit Yersiniose verursachen. Die Untersuchung des Stoffwechselsystems von *Yersinia enterocolitica* wäre eine gute Strategie zur Kontrolle der Infektion. Die Pyruvat-Kinase, eines der glykolytischen Enzyme, ist hoch reguliert und spielt eine wichtige Rolle bei der Steuerung des Stoffwechselflusses. *Yersinia* hat zwei Arten von Pyruvat-Kinasen, YePykF und YePykA, die jeweils allosterisch durch Fructose 1, 6-Bisphosphat (F16BP) bzw. AMP reguliert werden. Ungetaggte *Yersinia* Pyruvat-Kinasen (PK) YePykF und YePykA wurden aus dem *E. coli*-Expressionssystem exprimiert und aufgereinigt. Beide PKs wurden allein kristallisiert und auch mit ihren allosterischen Effektoren co-kristallisiert. Die Röntgenkristallstrukturen von *Yersinia*-PKs zeigten einen ähnlichen Aufbau wie andere bereits bekannte Pyruvat-Kinasen. Der Vergleich zwischen den Strukturen von YePykF- und des YePykF+FBP-Komplexes offenbarte Konformationsänderungen sowohl am allosterischen wie am aktiven Zentrum. Bei F16BP-Bindung ändert eine Loop-Region des allosterischen Zentrums ihre Konformation und die B-Domäne dreht im Durchschnitt um $14,6°$ in Richtung A-Domäne, was zu einer partiellen Schließung des aktiven Zentrums führt. Die YePykA-Struktur wurde nur im Komplex mit AMP aufgeklärt. Aufgrund der geringen Qualität des Datensatzes ist die AMP-Bindungsstelle noch nicht klar.

Strukturanalyse von Hepatitis C-Virus (HCV) und Wirtsrezeptor-Wechselwirkungen

Das Hepatitis C-Virus (HCV) ist der Haupterreger von akuten und chronischen Lebererkrankungen, Leberzirrhose, hepatozellulärem Karzinom und Leberversagen. Rund 160 Millionen Menschen sind weltweit infiziert, mit geschätzten 130 Millionen chronisch Infizierten. Es ist kein Impfstoff verfügbar und die aktuellen Therapien sind nur gegen einige

Genotypen wirksam, aber nicht gegen alle. Das Eindringen von HCV in die Wirtszelle erfordert das Zusammenwirken zahlreicher Faktoren in zeitlich und räumlich geordneter Weise. Die Interaktion des HCV-Hüllproteins E2 und einem Scavenger-Rezeptor Klasse B Typ I (SRB1) wird als notwendig für den Prozess des Viruseintritts erachtet. Die Strukturuntersuchung von E2 und SRB1 würde wertvolle Informationen zur Blockierung des Viruseintritts und damit zur Kontrolle der HCV-Infektion liefern. Die extrazellulären Domänen von E2 und SRB1 wurden kloniert und aus verschiedenen eukaryontischen Expressionssystemen exprimiert. Die transient in HEK293-Zellen exprimierten Proteine zeigten eine hohe Aggregation aufgrund von Fehlfaltung. Später wurde HCV sE2 homogen aus einer stabilen CHO- lec 3.8.2.1 Zelllinie exprimiert. Das aufgereinigte SE2 wurde zur Kristallisation eingesetzt, es konnte jedoch kein Kristall gewonnen werden. Etwa zur gleichen Zeit wie diese Experimente wurden zwei E2-Kernstrukturen ohne die hypervariablen Regionen veröffentlicht. Die SRB1-Ektodomäne wurde homogen in Insektenzellen exprimiert, und das Protein konnte zu einem hohen Reinheitsgrad aufgereinigt werden. Aus unbekanntem Grund konnten die Ergebnisse in späteren Experimenten nicht repliziert werden.

Aufklärung der Autoaktivierungsprozess von Legumain

Die Cystein-Protease Legumain ist das Nebenprodukt der HCV-sE2-Aufreinigung. CHO- lec 3.8.2.1 gibt Prolegumain konstitutiv ins Zellkulturmedium ab. Legumain gilt als eine Asparaginyl-Endopeptidase, weil es bei pH 5,8 spezifisch nach Asparagin spaltet. Man geht davon aus, dass Säugerlegumain in der Immunität, Signaltransduktion, Knochenresorption und Tumorwachstum/Metastasierung eine Rolle spielt. Zwei Prolegumain-Strukturen und eine aktive Legumain-Struktur wurden in dieser Untersuchung aufgeklärt. Die Tertiärstruktur von Prolegumain hat drei Domänen: die C-terminale Legumain-Stabilisierungs- und Aktivtätsmodulations-Domäne (legumain stabilization and activity modulation, LSAM), das pH-aktivierte Legumain (AEP) und ein Aktivierungspeptid (AP). Bei der pH-Verschiebung von neutral zu sauer kann Prolegumain an zwei Schnittstellen gespalten werden, was zu einer Freisetzung des AP-Peptids und der C-terminalen LSAM-Domäne führt und so aktives Legumain erzeugt. Die hochaufgelöste Struktur von aktivem Legumain zeigt einen Succinimidrest im aktiven Zentrum. Das Succinimid wird aus einem Aspartatrest zyklisiert und hängt nicht mit Proteinabbau zusammen. Obwohl Kristallisationsplatten mit aktivem

Legumain befüllt wurden, zeigen zwei Kristallformen die Prolegumain-Architektur, in der die gespaltenen Domänen zusammengehalten, aber nicht ligiert werden. In der Studie von Zhao et al. wurde über die Religation von Peptidbindungen *in vitro* berichtet. Die strukturelle Untersuchung von Legumain würde wertvolle Informationen für das Wirkstoffdesign und die Entwicklung eines Impfstoffes gegen Krebs liefern.

Summary

Structural analysis of *Yersinia* pyruvate kinases reveals the basis of allosteric regulation

Foodborne diseases are becoming more prevalent recently. *Yersinia enterocolitica* is one of the pathogenic bacteria that cause the foodborne disease called yersiniosis. The study on the *Yersinia* enterocolitica's metabolic system would be a good strategy to control the infection. Pyruvate kinase, which is one of the glycolytic enzymes, is highly regulated and plays an important role in controlling the metabolic flux. *Yersinia* has two types of pyruvate kinases, YePykF and YePykA, which are allosterically regulated by fructose 1, 6-bisphosphate (F16BP) and AMP, respectively. Non-tagged *Yersinia* pyruvate kinases (PK) YePykF and YePykA were expressed and purified from *E. coli* expression system. Both PKs were crystallized alone and also co-crystallized with their allosteric effectors. The X-ray structures of *Yersinia* PKs revealed architectures similar to the other already reported pyruvate kinases. The comparison between YePykF and YePykF+FBP complex structures revealed conformational changes on both allosteric site and active site. Upon F16BP binding, a loop region in allosteric site changes its conformation, and the B domain rotates towards the A domain by 14.6° on average leading to a partial closure of the active site. YePykA structure was solved only in complex with AMP. Due to the low quality of the dataset, the AMP binding site remains unclear.

Towards structural elucidation of hepatitis C virus (HCV) - host receptor interactions

Hepatitis C virus (HCV) is the major causative pathogen of acute and chronic liver diseases, liver cirrhosis, hepatocellular carcinoma and liver failure. Approximately 160 million people are infected worldwide, with an estimated 130 million chronically infected. No vaccine is available and current therapies are effective against most genotypes but not all. HCV host cell entry requires numerous factors cooperating in a temporally and spatially ordered manner. The interaction between HCV envelope protein E2 and a scavenger receptor class B type I (SRB1) is considered necessary in the virus entry process. The structural analysis of E2 and

SRB1 would provide valuable information for blocking the virus entry and hence control of the HCV infection. Both E2 and SRB1's extracellular domains were cloned and expressed from different eukaryotic expression systems. The proteins transiently expressed with HEK 293 cells showed high aggregation due to misfolding. Later HCV sE2 was homogenously expressed from CHO lec 3.8.2.1 stable cell line. The purified sE2 was set up for crystallization, but no crystals were obtained. Around the same time of these experiments, two E2 core structures without hypervariable regions were published. SRB1 ectodomain was homogenously expressed from insect cells, leading to protein of high purity. Due to unknown reason, these results were not reproducible.

Exploration of the legumain autoactivation process

The cysteine protease legumain turned out to be a byproduct of HCV sE2 purification. CHO lec 3.8.2.1 is constitutively secreting prolegumain to the cell culture medium. Legumain is considered to be an asparaginyl endopeptidase because it specifically cleaves after asparagine at pH 5.8. Mammalian legumain is considered to function in immunity, signal transduction, bone resorption and tumor growth/metastasis. Two prolegumain and one active legumain structures have been solved in this work. The tertiary structure of prolegumain has three domains: the C-terminal legumain stabilization and activity modulation (LSAM) domain; the pH-activated legumain (AEP) and an activation peptide (AP). During pH shift from neutral to acidic, prolegumain can be cleaved at two cleavage sites, causing the release of the AP peptide and C-terminal LSAM domain, and leading to an active legumain. The high-resolution active legumain structure reveals a succinimide residue in the active site. The succinimide is cyclized from an aspartate residue, and not related to protein degradation. Although crystallization plates have been set up with active legumain, two crystal forms show the prolegumain architecture, in which the cleaved domains are held together but not ligated. The re-ligation of peptide bonds in vitro has been reported by others. The structural study on legumain would provide valuable information for drug design and cancer vaccine development.

I Structural analysis of *Yersinia* pyruvate kinases reveals the basis of allosteric regulation

I.1 Introduction

I.1.1 Infection

Infection is the invasion and multiplication of pathogenic microorganisms in a host organism's body tissues, especially that causing local cellular injury by competitive metabolism, toxins, intracellular replication, or antigen-antibody response (Dorland's Medical Dictionary for Health Consumers, 2007). The infectious process involves several factors, and occurs only if each factor is present in a proper sequence. This involves an infectious agent, a reservoir in which the microbes can survive, a portal of exit to get out of the reservoir, a mode of transmission by an intermediate object, a portal of entry to get into a new reservoir and a susceptible host who is unable to resist infection forming a nonreversible chain of infection (Figure I.1). If any link in the chain can be broken, the infection won't spread further.
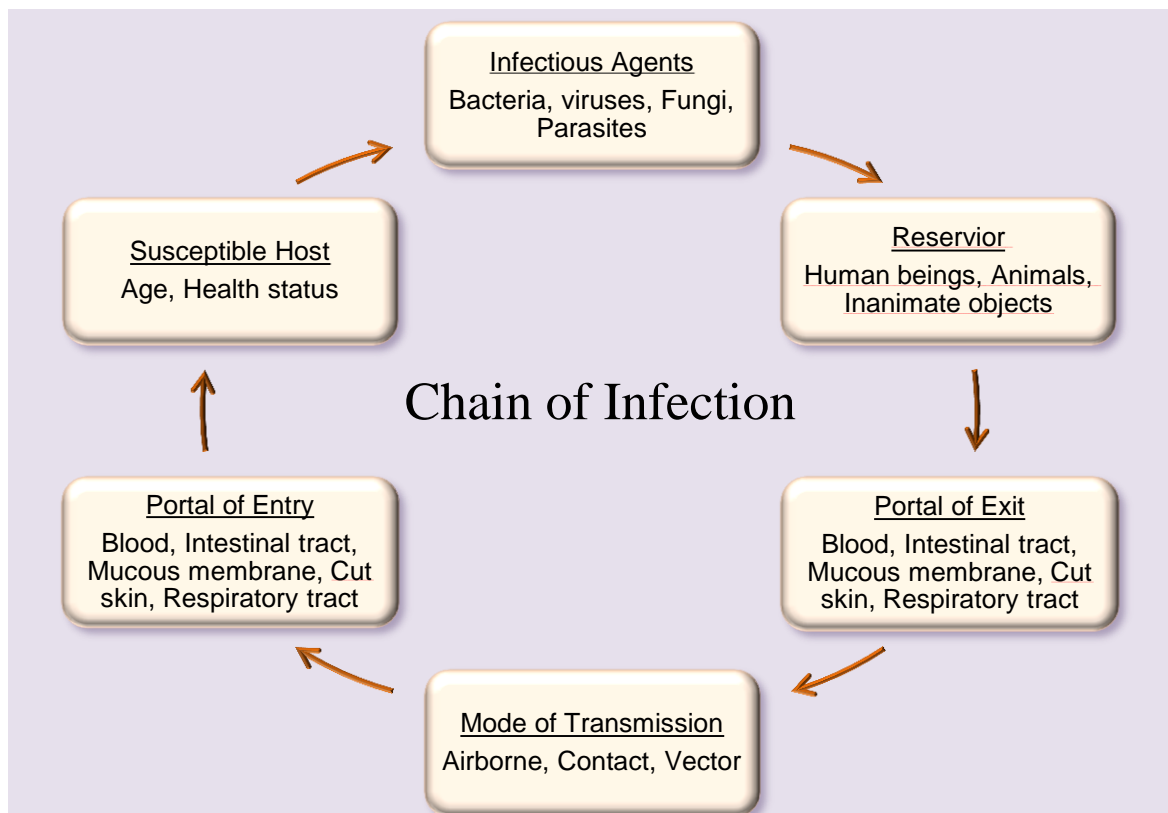

Figure I.1: Chain of Infection. Schematic of the chain of infection according to http://contagions.wordpress.com/2010/11/06/what-is-the-chain-of-infection/.

Infectious agents also called pathogens are microorganisms that include bacteria, viruses, fungi, protozoa and others. Pathogens vary in infectivity and virulence, with differences depending on different architecture of the pathogens and the virulence factors they produced. For initiating a disease, an infectious dose (a sufficient number of pathogens) is also required.

I.1.2 Bacterial infection

Bacterial infection is any type of infection that is caused by pathogenic bacteria. Our bodies contain billions of bacteria, the mass of which is estimated to account for 1-3% of the body weight (MacDougall, 2012). Most bacteria are harmless or even beneficial and are called normal flora. The normal flora can help digesting food, sustaining normal pH of the body, and providing essential nutrients. The normal flora does not cause disease unless the balance is disturbed or the normal living place of the bacteria changed (Yonath, 2004). The World Health Organisation (WHO) reported emerging foodborne diseases were caused by bacterial pathogens (WHO, Fact sheet N° 124). Pathogenic bacteria that cause these diseases include *Yersinia enterocolitica, Salmonella spp.*, *Vibrio cholerae*, *Listeria monocytogenes* and pathogenic *E. coli.* Foodborne diseases are becoming more prevalent for several reasons that include increasing international travel and trade, the storage of large amounts of frozen food, and most importantly, the rise in pathogen resistance against antibiotics.

I.1.3 *Yersinia* infection

The genus *Yersinia* belongs to the family Enterobacteriaceae. These are Gram-negative rod shaped, and facultative anaerobic bacteria. So far, 17 species have been reported in this genus, but only three species are pathogenic for human beings. These are *Yersinia pestis*, *Yesinia pseudotuberculosis* and *Yersinia enterocolitica* [European Food Safety Authority (EFSA), 2013].

Yersinia enterocolitica and *Yersinia pseudotuberculosis* are the most common pathogens of foodborne diseases, which cause the disease called yersiniosis [ECDC Annual epidemiological report (2011)]. In 2011, 7041 confirmed cases of yersiniosis were reported in European countries, with an overall proportion of 2.19 per 100000 persons. In the past, Germany accounted for the highest proportion of 47.9% of all reported cases. *Yersinia*

enterocolitica was the main species in human reported cases (98.4% of all confirmed cases in 2011). Many cases are considered to be associated with the ingestion of uncooked pork (ECDC annual epidemiological report, 2013). After ingestion via contaminated food or water, the enteropathogenic *Yersinia* pass through the intestinal tract and reach the terminal ileum, where they can bind to the microfold cells (M cells) of the small intestine. Via a plasmid encoded Type III secretion system *Yersinia* translocate anti-phagocytic proteins (Yops) into the macrophages and induces apoptosis. The bacteria are released and then may disseminate to mesenteric lymph nodes or even deeper tissues such as the liver and spleen (Figure I.2). Common symptoms are fever, abdominal pain, and diarrhea, which is often bloody (Bottone, 1997). The diagnosis is usually confused with appendicitis. Symptoms occur four to seven days after exposure to the infectious pathogen, and usually last for 1 to 3 weeks or longer. Treatment usually requires aggressive antibiotics that include ciprofloxacin, chloramphenicol, ampicillin, and polymyxin (https://en.wikipedia.org/wiki/Yersiniosis). More than 50% of *Yersinia enterocolitica* infections occur in children younger than five, and about 38% of all infections occur in infants less than one year old. Infections in infants are more dangerous and might develop into bacteremia, where bacteria are present in the blood stream and usually fatal (Bent et al., 2013).
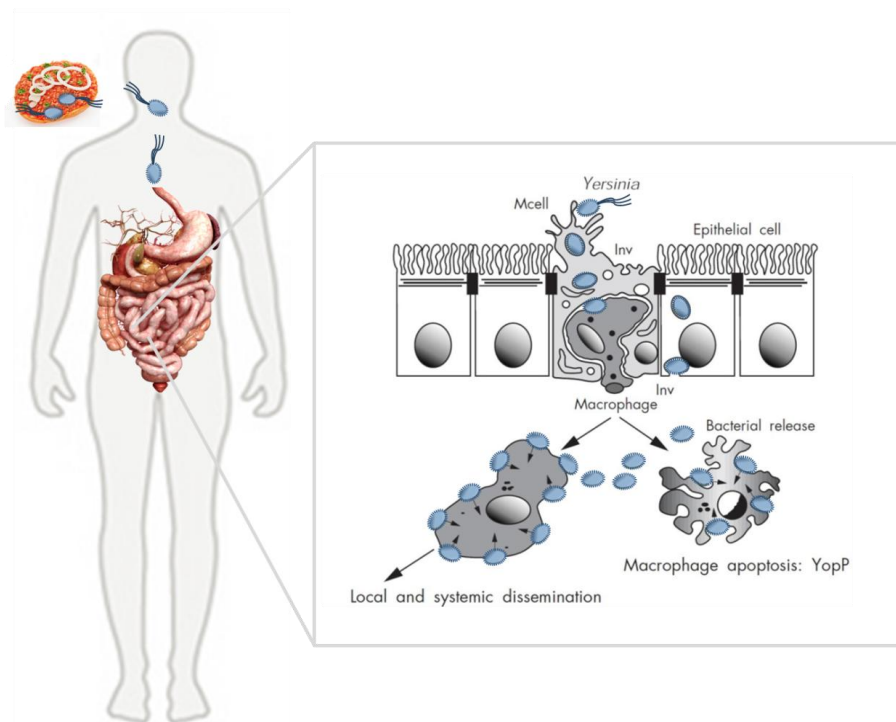

Figure I.2 Infection route of enteropathogenic *Yersinia* species. Figure was drawn according to Sansonetti, P. (2002).

After exposure to host cells, *Yersinia enterocolitica* initiates infection through a protein export apparatus called the Type III secretion system (T3SS) or an injectisome. Several well-known Gram-negative bacterial pathogens, including *Shigella flexneri*, *Pseudomonas aeruginosa*, *Salmonella typhimurium* and *Yersinia* use this system. There are two types of Type III secretion systems (T3SS) present in *Yersinia enterocolitica* biovar 1B, the plasmid encoded Ysc T3SS and the chromosomally encoded Ysa T3SS. The Ysa T3SS was reported to be only been expressed at 26 ℃ in a high nutrient medium with a high concentration of salt (Bent et al., 2013). The injectisome is composed of ~25 proteins and facilitates the bacteria to attach to the cell membrane and to inject effector proteins through this membrane into host cells (Cornelis, 2006). Recent studies combining electron microscopy (EM), X-ray crystallography, and nuclear magnetic resonance (NMR) spectroscopy revealed amazingly detailed models of the injectisome. The injectisome can be divided into 3 parts: a hollow, ~60nm long needle protruding from the bacteria surface, a basal body that spans through two bacteria membranes and the periplasm, and a cytoplasmic part (Figure I.3). The basal body is formed by three multimeric proteins (YscC, D, J in Yersinia spp). YscC secretin forms a 12-15mer barrel-shaped structure that spans the OM and protrudes into the periplasm (Burghout et al., 2004; Hodgkinson et al., 2009; Schraidt and Marlovits, 2011; Spreter et al., 2009). YscJ is a lipoprotein and was proposed to form a 24-subunit ring, which reaches into the inner membrane (Hodgkinson et al., 2009; Silva-Herzog et al., 2008). Ysc has an elastic structure and acts as a connector between the secretin and the IM (Diepold et al., 2010).

According to our collaborators from RKI (Robert Koch Institute), two secreted YscM proteins (YscM1 and YscM2) of T3SS are assumed to interact with pyruvate kinases and activate them. The interaction is not only with *Yersinia* PK but also with the PK from host cells. The assumption is that the interaction between YscM and PKs could activate PK and provide more energy to facilitate infection and more metabolites for bacteria grow, and at the same time, disturb the metabolic system of host cells. In this work, the interaction between YscM and PKs has been explored, but haven't been quantified to a stoichiometry level in vitro.
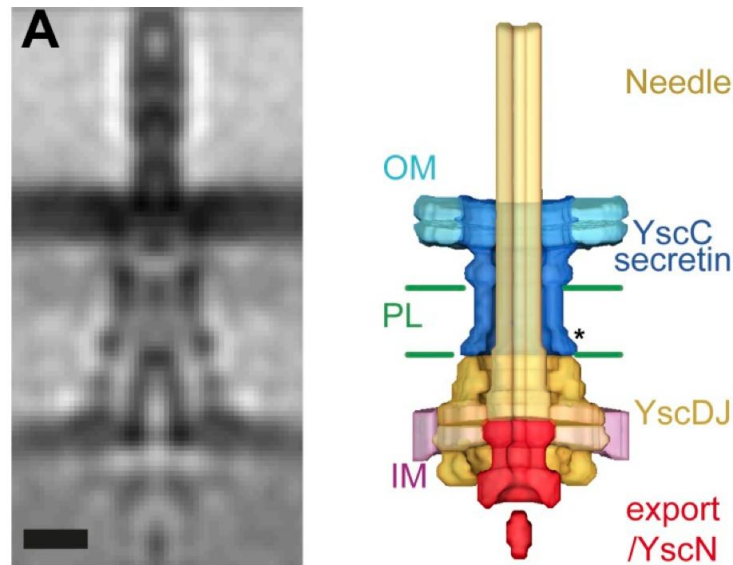
Figure I.3: Structure of the *Y. enterocolitica* injectisome in situ. Slice through the average 3D structure of the injectisome and a model with components. OM–outer membrane, PL–peptidoglycan layers, IM–inner membrane。 * indicates the junction between YscC and YscD (Kudryashev et al., 2013).

I.1.4 Pyruvate kinase

Pyruvate kinase (ATP: pyruvate 2-O-phosphotransferase (EC 2.7.1.40); PK) catalyzes the last step of glycolysis, where a phosphoryl group is transferred from phosphoenolpyruvate (PEP) to ADP to yield pyruvate and ATP (Valentini et al., 2000). The PK catalyzes the reaction in two steps (Figure I.4). The first step, a phosphoryl group is transferred from PEP to ADP, forming an enolate intermediate and yielding an ATP molecule; in a second step the Thr328 and the Mg^{2+} cation increase the acidity of water through hydrogen bonding and charge stabilization, respectively. A specific acid is generated to act on the enolate intermediate causing the protonated enolate intermediate to tautomerize to pyruvate. A hydroxide is produced and can be reprotonated back to the original state (Dombrauckas et al., 2005).
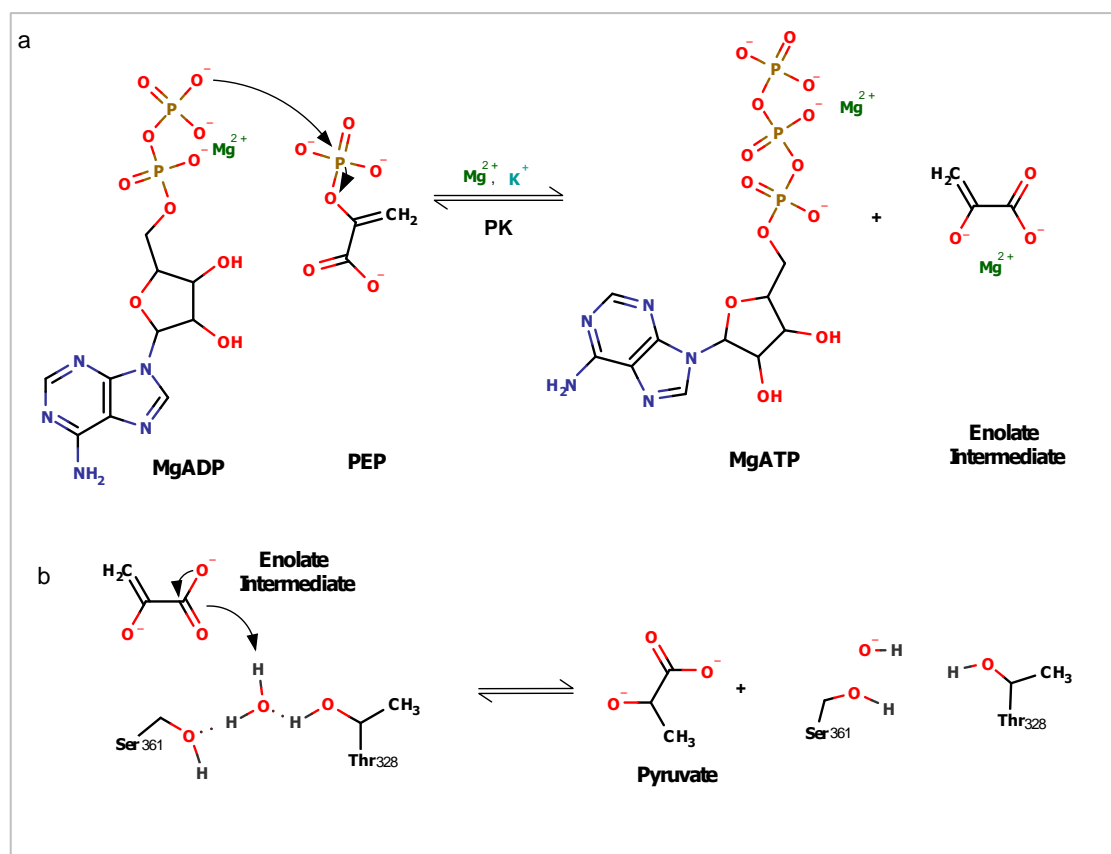
Figure I.4: The reaction mechanism of pyruvate kinase catalyzed reaction. The reaction takes place in two steps. (a) Step 1: the transfer of a phosphate group from PEP to ADP, and the formation of an enolate intermediate. (b) Proposed hydrogen-bonding scheme through a water molecule, which is responsible for the protonation of enolate. Protonated enolate intermediate tautomerizes to pyruvate and a hydroxide is yielded. Adopted from Dombrauckas et al. (2005).

PK catalyzes one of the two steps producing ATP in glycolysis. Under physiological conditions, the reaction is irreversible and highly favored because the energy released from the transition of PEP to pyruvate (ΔG_{PEP} = 58 kJ/mol) is much higher than the synthesis of ATP (ΔG_{ATP} = 29 kJ/mol) (Dombrauckas et al., 2005). PK requires monovalent K^+ and divalent Mg^{2+} or Mn^{2+} cations for its activity (Mattevi et al., 1995). The substrate PEP and product pyruvate are involved in a number of metabolic pathways. As most enzymes of glycolysis, PKs are highly regulated by many different chemicals and play an important role in controlling the metabolic flux. It serves as a switch between the glycolytic and the gluconeogenic pathways, and the glycolytic phospho-metabolites are used as synthetic precursors in cellular proliferation (Bakszt et al., 2010; Mazurek et al., 2005).

Bacteria typically harbor a single type of PK enzyme. However, a few bacteria, such as *Escherichia coli*, *Salmonella* and *Yersinia*, harbor two types of PK isozymes (type I and type II pyruvate kinases). Both PK isozymes can be activated by their substrate PEP and inhibited by the product ATP, which is called homotropic activation. However, both PK isozymes can also be activated by their effectors that bind to different position in the active site, which is called heterotropic activation. The type I pyruvate kinase (PykF) is usually activated by fructose 1, 6-diphosphate (F16BP) or fructose 2, 6-diphosphate (F26BP). The type II pyruvate kinase (PykA) is usually activated by AMP and monophosphorylated sugars (i.e. ribose-5-phosphate) (Hofmann et al., 2013).

In eukaryotes, most PKs resemble the prokaryotic type I isozyme. They are allosterically activated by F16BP or F26BP. In mammals, there are four PK isozymes that are expressed in different tissues or in different life stages. PKL exists in liver, PKR in red blood cells, PKM1 is present in normal differentiated tissues, and PKM2 is found in embryonic cells and other proliferating cells, such as cancer and tumor cells. PKL, PKR and PKM2 isozymes are allosterically regulated by F16BP and PKM1 is constitutively active (Dombrauckas et al., 2005). In mammals, some hereditary diseases are related to mutations of PKs. The most famous one is the hereditary nonspherocytic hemolytic anemia, which is caused by PK deficiency in erythrocytes. Over 130 mutations are known that lead to a lowered production, activity, or stability of PKR. A complete knock out of this enzyme is usually lethal. In severe cases it can be life threating, and the patients might need regular blood transfusions to survive. A transgenic rescue strategy is still ongoing (Kanno et al., 2007).

PKs play an especially important role in certain cancer cells. Tumors occur when cells divide and grow excessively in the body. Oxygen in the tumor microenvironment is consumed and an anaerobic microenvironment is formed. Tumors without sufficient supply of oxygen rely heavily on ATP produced via the glycolytic cycle and hence on the activity of PKs (Dombrauckas et al., 2005). There is a strong correlation between the up-regulation of human PKM2 and incogenesis, and PKM2 is found in all tumors studied so far (Christofk et al., 2008). In contrast, inhibition of PK activity was shown to be beneficial for certain types of cancer cells as it leads to an increased availability of glycolytic metabolites that are necessary for tumor growth (Anastasiou et al., 2012). The decision whether glucose is converted to lactate for maximized production of ATP or used for the synthesis of cellular building blocks is made at the level of PKs, and regulation of PK activity is crucial for tumor

cells to maintain the fine balance between these two opposing requirements (Mazurek, 2011). Hence the regulation of PKs activity might be a strategy for cancer treatment. Potential ways to achieve this could be through the use of small molecules that act as allosteric activators/inhibitors of PKs or the induction of oxidative stress as this is known to lead to PK dissociation and inactivation (Mazurek, 2011).

I.1.4.1 Architecture of pyruvate kinases

The first PK structure has been reported in 1975 (Stuart et al., 1979, Stammers and Muirhead, 1975). Since then, 67 Structures of PK from 14 species (with pairwise sequence identities as low as 24%) have been elucidated in complex with various ligands, and the investigation is still ongoing. PKs have been extensively characterized in *Leishmania mexicana*, *Trypanosoma brucei*, *Felis catus* (Cat) and *Homo sapiens* (Human). The common characteristics of these enzymes are the subunit is around 50 kDa and the oligomeric state is either T or R-state tetramer. For bacterial phosphofructokinase, the T-state means tense state and is usually the inactive state, while the R-state means the relaxed state, and is usually the active state (Perutz, 1989). For convenience, we retain the nomenclature of T-state and R-state to be inactive and active states, respectively.

Every PK subunit has four domains. The first 50 amino acids form an N-terminal domain, which is usually a helix-loop-helix motif, or only one helix in some pyruvate kinases. The N-terminal domain is only present in eukaryotic cells and is absent in prokaryotic cells (Rigden et al., 1999). The A domain, composed of an $(\alpha/\beta)8$ barrel, forms the rigid body of PK. The C domain, usually composed of α and β elements, is the effector binding domain, while the B domain, composed of β-strands and random coils, is usually flexible (Bakszt et al., 2010). The active site is located in a pocket between the A and B domains. Two kinds of interfaces are indicated with green and black dashed lines in Figure I.5. Two subunits form an asymmetric dimer through an adjacent A-domain interface and is called the "large" interface. The tetramer structure formed through an adjacent C-domain interface is called the "small" interface. The PK homotetramer, consisting of a dimer of dimers, forms a ring in which the B domains behave like mobile lids covering both ends of the ring (Figure I.5) (Tadesse et al., 2013).
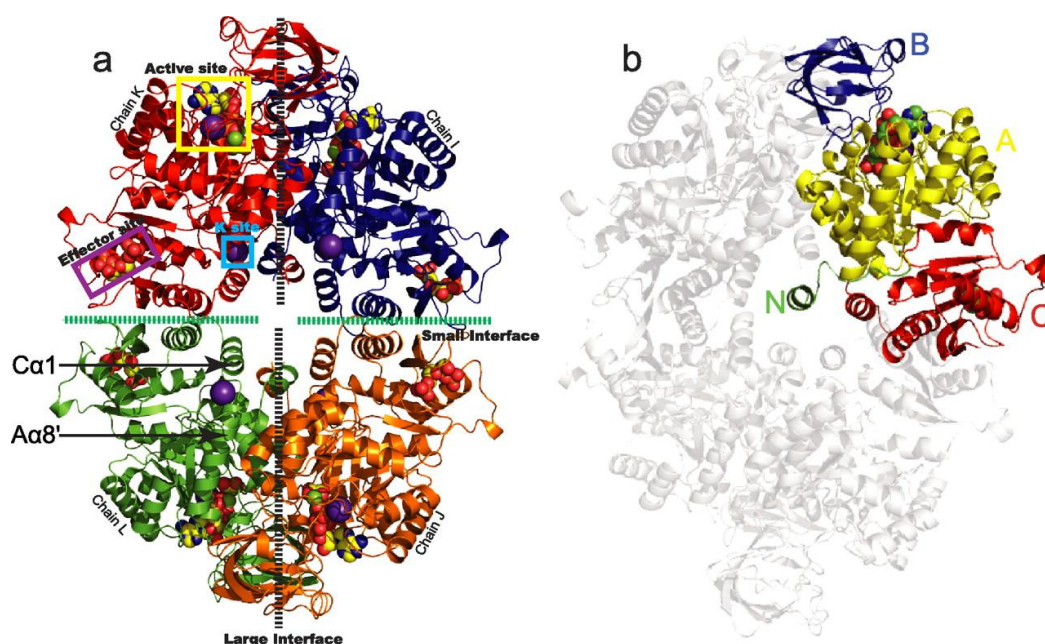
Figure I.5: Structure of the *Lm*PYK ATP OX Fru-2, 6-BP complex showing domain boundaries and tetramer architecture (Morgan et al., 2010). (a) *Lm*PYK ATP OX Fru-2,6-BP tetramer highlighting the active, effector, and potassium (K$^+$) sites. The large and small interfaces between monomers are shown as *dashed lines.* (b) *Lm*PYK ATP OX Fru-2, 6-BP tetramer in which one subunit has been *colored* to show domains; N domain (*green*, residues 1–17), A domain (*yellow*, residues 18–88 and 187-356), B domain (*blue*, residues 89–186), and C domain (*red*, residues 358–498).

Although PKs have many characteristics in common, there are structural differences between different species. These differences have been explored and targeted in research for anti-infectious diseases caused by bacteria and parasites (Nowicki et al., 2008; Zoraghi et al., 2011a; Zoraghi et al., 2011b). Bis-indole alkaloids and IS-130 have been found to be effective inhibitor for methicillin resistant *Staphylococcus aureus* PK. These chemicals have been explored in antibacterial drug research (Axerio-Cilies et al., 2012; Zoraghi et al., 2011b). Chemicals like suramin (Tadesse et al., 2013), 4-[(1,1-dioxo-1,2-benzothiazol-3-yl) sulfanyl] benzoic acid (DBS) (Morgan et al., 2012) have been found to work as inhibitors of *Leishmania Mexicana* PK and have been utilized in anti-parasitic research.

I.1.4.2 *Yersinia enterocolitica* pyruvate kinases

Yersinia enterobacteriaceae harbor both isozymes YePykF and YePykA. YePykF has 470 amino acids with a molecular weight of 50.5 kDa, and is allosterically regulated by F16BP.

YePykA has 480 amino acids with a molecular weight of 51.5 kDa, and is allosterically regulated by AMP and ribose-5 phosphate. The sequence identity of YePykF and YePykA is 36.7% (Figure I.6) and the enzymatic activity of YePykF is higher than that of YePykA (Hofmann et al., 2013).

Figure I.6: Amino acid sequence alignment of YePykA and YePykF. Identical residues are highlighted in red, similar residues are surrounded by a blue box.

I.2 Aim of the work
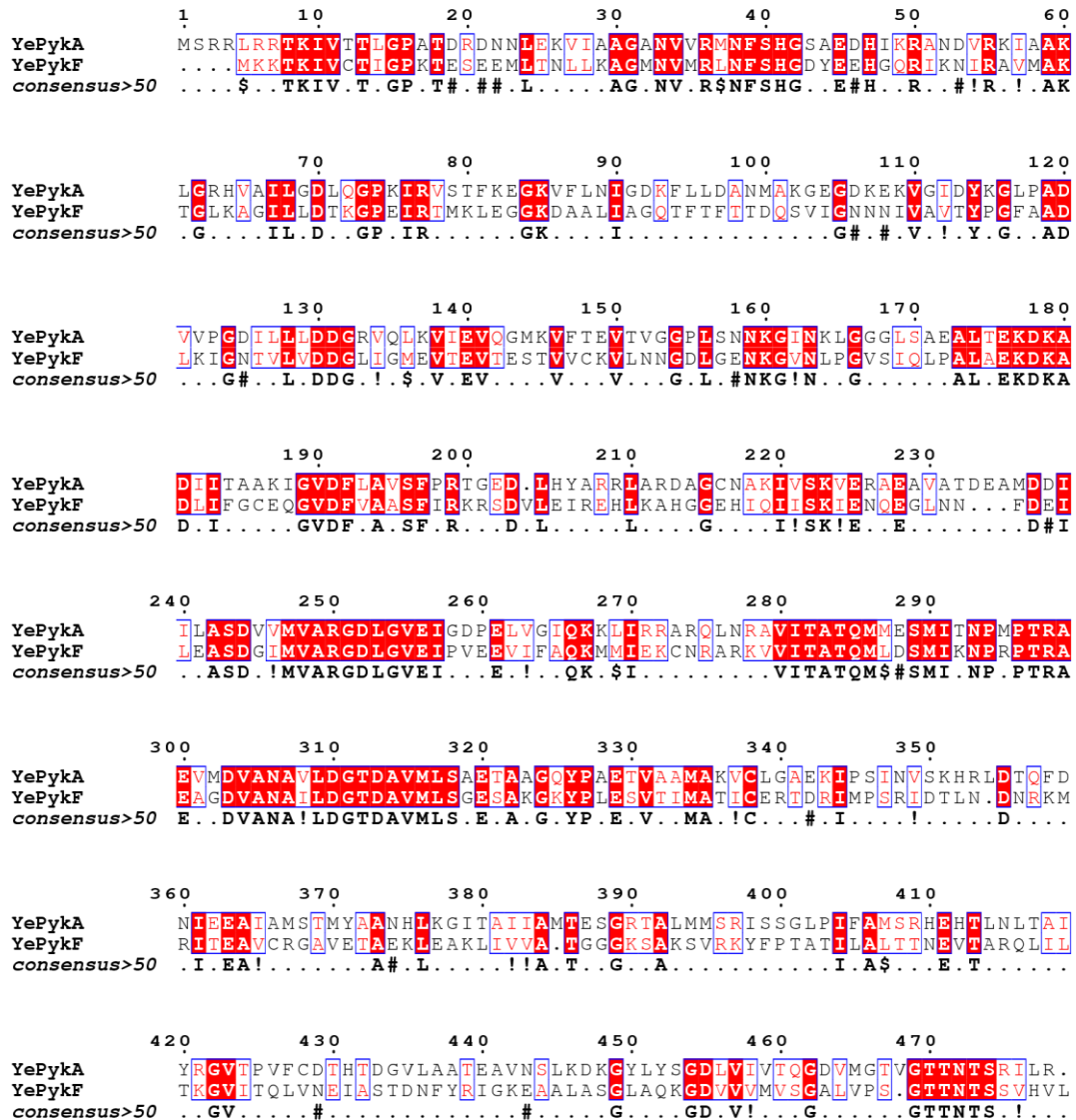
This project is aimed at elucidating the allosteric effect of PK by effector binding. The investigation of the regulation mechanism of PKs will provide information about the effects

of disturbing the balance between energy production and metabolites production of tumor cells, thus providing insights for drug design in cancer treatment. Another aim is to explore the activation mechanism between PK and YscM to further explain the *Yersinia* infection mechanism.

I.3 Results

I.3.1 Purification of pyruvate kinase YePykF and YePykA

YePykF and YePykA of *Yersinia enterocolitica* WA-314 were cloned into a IPTG-inducible pWS vector. Non-tagged YePykF and YePykA were recombinantly expressed in Escherichia coli BL21 (DE3) pLysS competent cells. Both proteins have higher solubility when expressed in 300 K than in 310 K. Larger scale protein production was done by adding 0.1 mM IPTG in 2YT medium at 300 K for 5 hours. Both proteins were purified by combining anion exchange chromatography on a HiPrep16/10 Q XL column and size exclusion chromatography on a Superdex 200 prep grade column. Elutions of YePykF and YePykA from size exclusion chromatography were in accordance with the formation of protein tetramers. Unusually YePykF (50.5 kDa) was eluted significantly and reproducibly before YePykA (51.5 kDa) with a peak at 171 ml compared to 176 ml for YePykA. This could be explained by unspecific interaction of YePykA with the Superdex matrix that delays the elution. The yields for YePykF and YePykA were approx. 183 mg/L and 250 mg/L respectively. The specific activities determined by Dr. Julia Hofmann (Robert Koch Institute) for YePykF and YePykA (108 U/mg and 85 U/mg) are consistent with other pyruvate kinases (Hofmann et al., 2013; Zoraghi et al., 2010).

Protein samples with 10% glycerol have been stored at 193 K for crystallization tests. Before each crystallization test, one frozen protein fraction (10mg) was thawed on ice. Re-purification was performed with anion exchange chromatography on a MonoQ 5/50 GL column and size exclusion chromatography on a Superdex 200 HR 10/30 column. The aim of the re-purification was to remove the putative inhomogeneity that might have arisen during freezing and thawing, and hence increase the protein's propensity for crystallization. YePykF and YePykA show similar behavior in their size exclusion chromatography profile. YePykF and YePykA both eluted around 10 - 15 mL, which corresponds to the size of a protein

tetramer. SDS-PAGE analysis showed protein bands at about 51.5 kDa for YePykA and 50.5 kDa for YePykF, with purity above 98% (Figure I.7 and Figure I.8). Although YePykF and YePykA have identical molecular weights but differ in their pI values. YePykF has a pI value of 6.0, while YePykA has a pI value of 6.8. Accordingly, in anion exchange chromatography, YePykF eluted earlier (44 ml, 23% conductivity) than YePykA (45 ml, 25% conductivity).
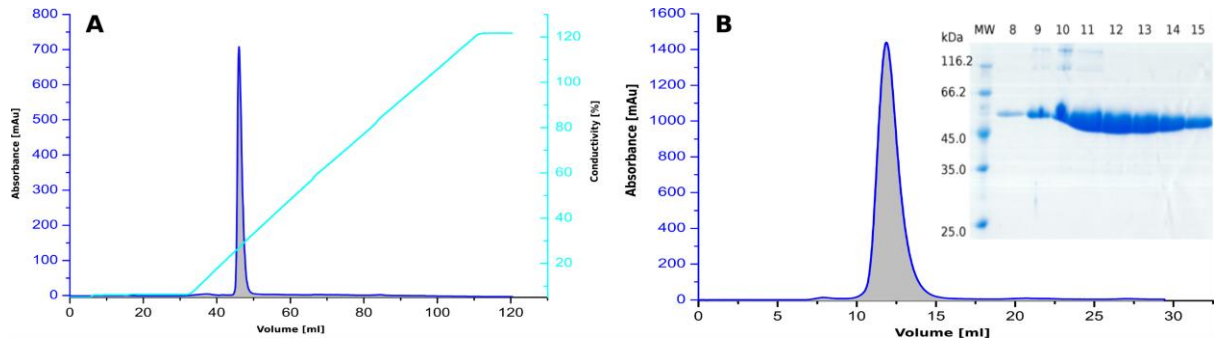
Figure I.7: Chromatograms of the YePykF purification monitored at λ = 280 nm, and SDS-PAGE of size exclusion peak fractions. (**A**) Chromatograms of YePykF MonoQ 5/50 QL anion exchange Chromatography. The blue curve indicates UV absorbance, and the cyan curve indicates the conductivity. (**B**) Chromatograms of YePykF on Superdex 200 10/30 column. Coomassie-stained SDS-PAGE containing every peak-fraction of the YePykF gel filtration, the fraction size is 1ml. MW: Molecular weight marker.
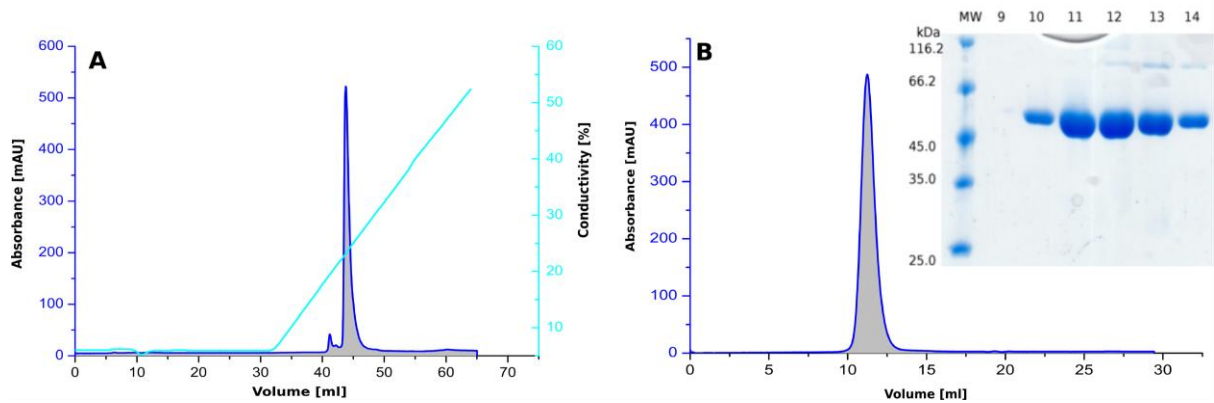
Figure I.8: Chromatograms of the YePykA purification monitored at λ = 280 nm, and SDS-PAGE of size exclusion peak fractions. (**A**) Chromatograms of YePykA MonoQ 5/50 QL anion exchange Chromatography. The blue curve indicates UV absorbance, and the cyan curve indicates the conductivity. (**B**) Chromatograms of YePykA on Superdex 200 10/30 column and SDS-PAGE of peak fractions. Coomassie-stained SDS-PAGE showing peak-fractions of the YePykA gel filtration run. The fraction size is 1ml. MW: Unstained low molecular weight marker.

Thermoshift assay, also called Differential Scanning Fluorimetry (DSF), measures the thermo stability of proteins and protein-ligand complexes using a fluorescent protein binding dye. The assay usually includes a buffer screen and an extra additive screen. The protein stability changes with buffer pH, salt content and cofactors. The first step buffer screen identified a buffer (10 mM Tris, pH 8.0, 100 mM KCl) where YePykF and YePykA were most stable. The second step was based on the buffer condition that confirmed and tested different additives that might stabilize the proteins. YePykF and YePykA were analyzed by thermal shift assay with five different ligands for each. SYPRO orange was employed as fluorescence signal for its non-specific binding ability to a protein hydrophobic surface (Kranz and Schalk-Hihi, 2011; Niesen et al., 2007). YePykF was tested with pyruvate, oxalate, PEP, ADP and F16BP. YePykA was tested with pyruvate, oxalate, PEP, ribose 5-phosphate and AMP. Each 50 µl testing system includes 0.5 mg/ml YePykF or YePykA protein, 5 mM of the test ligand. In negative controls, no ligand was added. The plate was slowly heated up from 283 K to 363 K in 0.5 K increments. The excitation and emission wavelengths were 485 and 575 nm, respectively. The fluorescence signal was analyzed with the CFX Manager software and interpreted as melt curves. F16BP had the greatest impact on the melting temperature of YePykF, increasing it by 12 K, from 319 K 331 K (Figure I.9). For PykA, AMP binding yielded the largest melt temperature jump from 319 K to 328 K. This melting temperature shift indicates the improvement of protein thermo stability by ligand binding. Even further, we could speculate the conformational change of free protein from a loose state to a tight state after ligand binding.
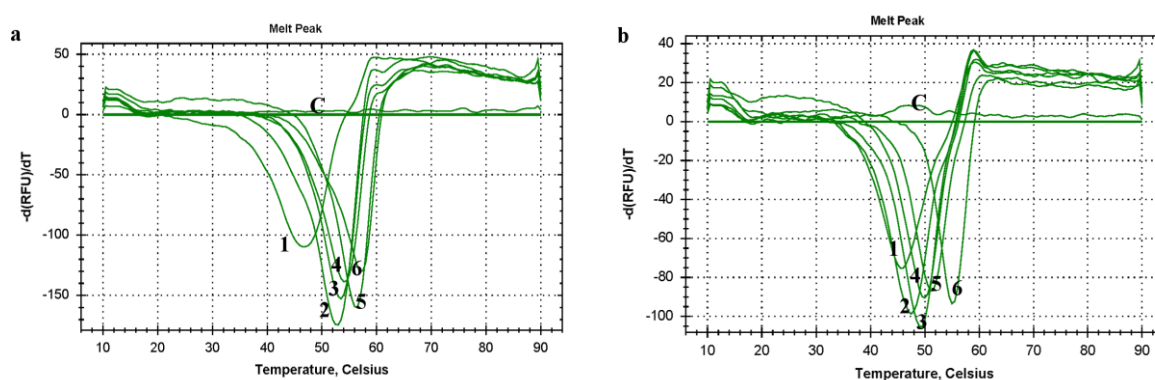
Figure I.9: Thermo shift assay for YePykF and YePykA. Curves with different numbers indicate the different ligands tested. (a) YePykF with different ligands. 1: No ligand; 2: 5 mM pyruvate; 3: 5mM oxalate; 4: 5mM PEP; 5: 5mM ADP; 6: 5mM F16BP; C: buffer control. (b) YePykA with different ligands. 1: No ligand; 2: 5 mM pyruvate; 3: 5 mM oxalate; 4: 5 mM PEP; 5: 5 mM ribose 5-phosphate; 6: 5 mM AMP; C: buffer control.

To analyze the secondary structure of YePykF and YePykA, circular dichroism (CD) spectroscopy was employed. CD spectroscopy is a technique where the CD of molecules is measured over a range of wavelengths. It is extensively used to study chiral molecules of all types and sizes, but it is mostly useful in studying larger biological molecules. The CD spectra exhibited two minima around 208 nm and 222 nm, which is a typical indication of an α-helical conformation. Recorded CD spectra illustrate minor differences between YePykA and YePykF and confirm their overall high content of α-helices. In the presence of the known allosteric activator F16BP, the CD spectrum of YePykF was only marginally changed (Figure I.10). This indicates that F16BP binding didn't change the secondary structure content of YePykF, the conformational change might be rigid body domain movement.
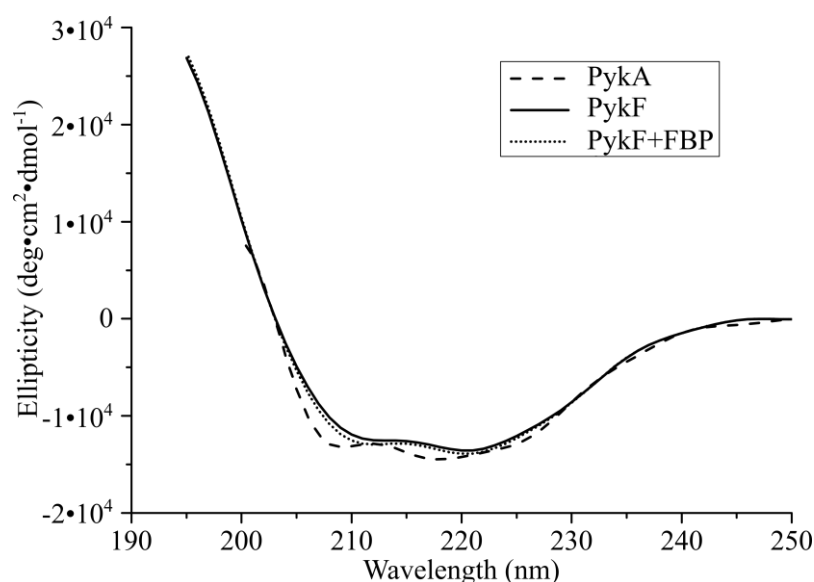
Figure I.10: Structural prediction of PykA and PykF by CD spectroscopy. The far-UV spectra of YePykA (22 μM), YePykF (10 μM) and YePykF+F16BP (YePykF: 9 μM; F16BP: 10 mM) were recorded in 20mM Tris-HCl (pH 8.0) at 283 K using a path length of 1 mm.

I.3.2 Crystallization of YePykF, YePykF+FBP and YePykA

Crystallization tests of purified YePykF and YePykA were set up with the sitting drop vapor diffusion method. Protein and reservoir volume ratio was set to 1:1 (0.2 μl : 0.2 μl) with a total volume of 0.4 μl. 496 different crystallization conditions were carried out with different protein concentrations (5-10 mg/ml). Although the protein is stable at higher concentrations, crystallization at ≥10 mg/ml led to too many tiny crystals, and ≤ 5mg/ml yielded irregular shaped crystals. Hence, 8 mg/ml of protein was used as a standard for crystallization

experiments. The crystallization temperature was varied from 277 K to 293 K (277 K, 285 K, 293 K). YePykF can be stabilized by F16BP, and YePykA can be stabilized by AMP. In order to elucidate the conformation differences between proteins and protein-effector complexes, co-crystallization experiments including their effectors was also performed.

Three dimensional crystals of PykF+F16BP complex crystallized in conditions consisting of 19-25% PEG 6000, 0.2M Tris/HCl, pH 7.6, 0.1M KCl at 293 K. Small three dimensional crystals were observed under the microscope 2-3 hours after the setup of the drop and were tested for X-ray diffraction after 2-3 days (Figure I.11 C and D). Crystal quality was improved by iterative fine screens matrix micro seeding. The best quality crystals were obtained from a condition containing 18% PEG 6000, 0.2M Tris/HCl, pH 7.6, 0.1M KCl at 293 K, with a crystal length of 0.8 mm. In contrast, YePykF did not yield any crystal over the same time period.

Lenticular crystals (Figure I.11 A and B) of YePykF have been obtained from initial screens 6 months after screens were set up at 285 K. The condition contains 1.6 - 2 M ammonium sulfate, 0.1 M Tris/HCl, pH 8.0. The largest crystal has a length of 50 μm, which is much smaller, compare to 0.8 mm of the largest YePykF+FBP crystal. Since the crystallization process was too long, no optimization was performed. Single crystal was fished and cryo-protected with ammonium sulfate as a cryo-protectant.

Non-tagged YePykA was expressed and purified similar to YeykF. In the initial screens, many needle crystals of YePykA and YePykA+AMP were obtained under different conditions. However, these crystals showed no diffraction in X-ray diffraction experiments. Even after several rounds of optimization, their quality did not improve. There was only one condition from a Midas screen, containing pentaerythritol ethoxylate (15/4 EO/OH) as a precipitant, that yielded three-dimensional crystals for YePykA+AMP (Figure I.11 E and F). Pentaerythritol ethoxylate is a branched polymer built on a pentaerythritol backbone. It forms branched polymers and behaves differently from traditional precipitants like MPD and PEGs in size and characteristics. In addition, pentaerythritol ethoxylate polymers also function as cryoprotectant. Crystals grown in a high concentration of this precipitant can be cryo-protected and frozen directly from the crystallization drop (Gulick et al., 2002). After several rounds of fine screens and seeding by the hanging drop method, the best crystal diffracted to 3.3 Å resolution, which allowed us to solve the YePykA+AMP structure. The X-ray data collection and refinement statistics are shown in Table II.1. These crystals are only from

YePykA in complex with AMP. For YePykA alone, only tiny crystals were obtained, which did not diffract to better than 5 Å, leaving the apo YePykA structure still unsolved.
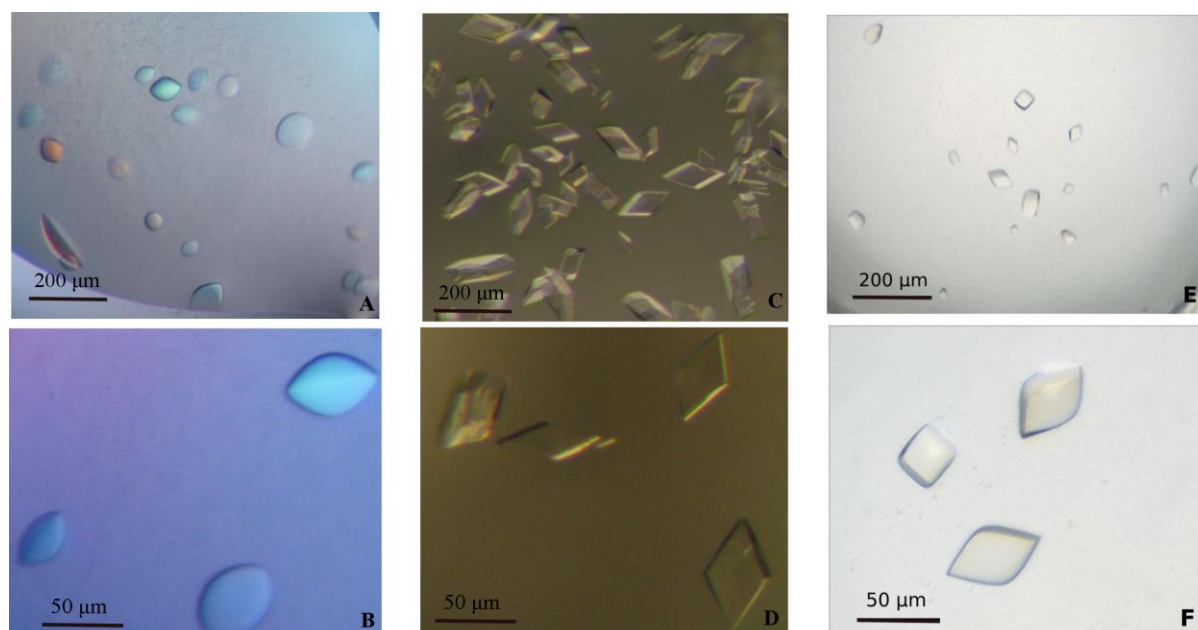
Figure I.11: Crystals of *Yersinia* pyruvate kinases. A and B: Lenticular crystals of YePykF; C and D: Three dimensional crystals of YePykF+F16BP; E and F: three-dimensional crystals of YePykA.

I.3.3 Data collection, model building of YePykF and YePykF+FBP and YePykA+AMP

Complete X-ray diffraction data of all three forms of crystals were collected using synchrotron radiation. Crystals of YePykF+FBP diffracted to 1.9 Å at the beamline 14.2 of BESSY II, Berlin, Germany. The data were processed by the XDS package in space group C121. From sequence alignment result, E.coli PykF has the highest sequence identity (85.7 %) with YePykF in the available PDB entries. E.coli PykF (1E0T) was chosen as search model for molecular replacement. Guided by a sequence alignment of YePykF and E.coli PykF, the program CHAINSAW (CCP4, 1994) was used to trim the search model at non-conserved positions. Matthew's coefficient (Vm = 2.56) revealed there are two molecules per asymmetric unit, and the solvent content is 51.9 %. Based on Vm, Phaser statistics indicated a very high probability of this solution being correct with a log likelihood gain of 3067. Similar to the other reported PK structures, the B domain of YePykF+FBP was also found to be out of electron density. The initial model quality was improved by manual building of B domain in Coot and refinement in Phenix.refine. After inclusion of effector atoms (F16BP)

and water molecules, an R factor value of 19.1% and R_{free} value of 22.4% were reached, which is in a good agreement with the resolution of 1.9 Å.

YePykF crystals diffracted X-ray to 2.7 Å using synchrotron radiation at the beamline P11 of PETRA III, Hamburg, Germany. Initial indexing revealed a hexagonal symmetry with unit cell dimensions of a = b = 263.75 Å and a short c axis of 101.29 Å. The data were processed by XDS with a space group P622, which contains two possible space groups $P6_4 22$ and $P6_2 22$. Initially, molecular replacement was done against PykF+FBP structure, but the B domain didn't fit into the electron density. During structure building, the B domain was found to move as a rigid body. Thus the B domain was separated from the YePykF+FBP structure as an ensemble. Molecular replacement was done against 2 ensembles from the PykF+FBP structure: B domain (residues 71 - 167) and AC domain (residue 1 - 70 and residue 168 - 470). The calculation of a Phaser_MR self-rotation function, taking into account all screw axes, revealed $P6_2 22$ as the correct space group, with a log likelihood gain of 10385 and the translation function Z-score of 30.3. Matthew's coefficient (Vm = 2.52) indicated two molecules per asymmetric unit and the solvent content was 51.2 %. After manual building in Coot and refinement with REFMAC5, the structure model of YePykF was improved with an R factor value of 18.4% and R_{free} value of 21.5%.

X-ray diffraction data of crystals from PykA+AMP were also collected using synchrotron radiation at the beamline P11 of PETRA III, Hamburg, Germany. The crystals diffract beyond 3.1 Å but the dataset could only be scaled to 3.3 Å due to anisotropic problems. The indexing and integration was carried out with the XDS package (Kabsch, 2010). These crystals have a very high Wilson B-factor of 102.4 Å^2, whereas the average values of the PDB structures with similar resolutions is 65.25 Å^2 for 3.25 - 3.5 Å resolution. The Matthew's coefficient (Vm = 2.47) indicated four molecules per asymmetric unit with a solvent content of 50.3 % (Matthews, 1968). Initial indexing revealed a P321 space group, which has three alternative space groups P321, $P3_2 21$ and $P3_1 21$. Calculation of the Phaser_MR self-rotation function solved the space group as $P3_2 21$. Based on sequence alignment and the available PK structure models, YePykF+FBP was used as a molecular replacement model. The trimmed model of YePykF+F16BP was produced by CHAINSAW and depended on the sequence alignment of YePykA and YePykF. Molecular replacement was carried out with Phaser_MR against two ensembles from trimmed YePykF+F16BP (B domain and AC domain) with a log likelihood gain of 13706 and translation function Z-score of 17.5. Refinement was done by

Phenix.refine including refinement of NCS restraints, secondary structure restraints and Ramachandran restraints. The final R factor of 24.7% and R_{free} of 29.6% were reached. Data collection and refinement statistics are summarized in Table I.1.

Table I.1: Data collection and refinement statistics for YePykF, YePykF+F16BP and YePykA+AMP datasets

Data statistics	YePykF	YePykF+F16BP	YePykA+AMP
Beamline	Petra III, P11	BESSY II, 14.2	Petra III, P11
Wavelength (Å)	1.033	0.918	1.033
Space group	P 6_2 2 2	C 1 2 1	P 3_2 2 1
Unit cell constants: a, b, c (Å)	263.8, 263.8, 101.3	168.7, 90.0, 68.3	151.0, 151.0, 154.0
α, β, γ (°)	90, 90, 120	90, 94.67, 90	90, 90, 120
Resolution range (Å)	20.0-2.7 (2.8-2.7)	20.0-1.9 (1.97-1.90)	20.0-3.3 (3.40-3.30)
I/σ (I)	20.0 (2.3)	13.9 (2.7)	26.6 (2.7)
Completeness (%)	100.0 (100.0)	96.5 (97.4)	96.8 (98.0)
Unique reflections	56975 (5602)	77321 (7752)	29956 (2992)
Multiplicity	2.0 (2.0)	1.8 (1.8)	2.0 (2.0)
Rmerge (%)	28.9 (29.7)	27.8 (33.0)	32.0 (32.9)
Wilson B-factor (Å^2)	57.04	25.89	102.4
Solvent content (%)	51.2	51.9	50.3
Refinement statistics			
Mol. PK / a.s.u.	2	2	4
Rwork (%)	18.4 (26.1)	19.1 (26.4)	24.7 (51.8)
Rfree (%)	21.5 (26.5)	22.4 (26.9)	29.6 (54.8)
av. B-factor of all atoms (Å^2)	72.4	35.2	107.3
r.m.s.d. bonds (Å) / angles (°)	0.004/0.83	0.008/1.05	0.010/2.04
Ramachandran favored (%)	96.1	98	96
Ramachandran allowed (%)	3.6	1.89	3.34

(): Values in () stand for the highest resolution shell.

I.3.4 Overall structure of YePykF and YePykF+FBP

The type I pyruvate kinase of *Yesinia* (YePykF) consists of 470 amino acids. Non-tagged YePykF gene was recombinant expressed in *Escherichia coli* BL21 (DE3) pLysS strain.

About 98.5% (463 out of 470) and 95.7% (450 out of 470) of the amino acids could be interpreted in the electron density of YePykF+FBP and YePykF structures, respectively. Both structures have a missing region from Thr345 to Arg350, and the YePykF structure has two more missing regions located from Met280 to Asn287 and Ser312 to Lys317.

YePykF is well structured with mostly α-helices and β-strands, in which the determination of secondary structure elements are based on the DSSP algorithm (Kabsch and Sander, 1983). YePykF contains 15 α-helices and 21 β-strands (Figure S1). Both the N-terminus and C-terminus are composed of β-strands and fitted quite well into the electron density. The missing region in both structures is interpreted as a big loop region by DSSP. The other two missing regions in YePykF are interpreted as small helices, each of which locates between a β-strand and a large helix.

The YePykF tertiary structure can be divided into three distinct domains (Figure I.12 a and b). In the middle of each subunit, there is a rigid A domain (residues 1 - 70 and 168 - 341, cyan), which is composed of an (α/β) 8 barrel and 3 small extra helices (β1, α1, β2, α2, β3, α3, β12, α4, β13, α5', α5, β14, α6', α6, β15, α7, β16, α8, α9). On one side of the subunit, there is a B domain (residues 71 to 167, green), which is composed mostly of 7 β-strands (β4, β5, β6, β7, β8, β9, β10, β11) and random coils. The B domains in the physiological tetramer behave like lids covering the A domain, and point to the outside of the homotetramer. On the opposite side of B domain, there is a C domain (residues 341 - 460, pink), which has a αβαβαβββ topology formed by 4 α-helixes and 5 β-strands (α10, β17, α11, β18, α12, β19, α13, β20, β21). C domains from 4 subunits assemble in the middle of homotetramer, and are responsible for forming of the homotetramer. The catalytic site (active site, violet) is located in a cleft between the B and A domains, in which α6' is critical for the activity. The allosteric effector-binding site is located in a concave of C domain (yellow orange). The effector-binding involves three loops, which are located between α9 and α10, β17 and α11, β20 and β21. Detailed interaction between the effector molecule and the YePykF will be shown in section I.3.5.

The quaternary structures of both YePykF and YePykF+FBP are tetramers (Figure I.12 c). The tetramer organization has already been confirmed by size exclusion chromatography. Both structures have an asymmetric unit as homodimer and the physiological unit as homotetramer. The homodimer and homotetramer form through interactions of residues on the interfaces. The homodimer was formed through interactions of residues on adjacent A

domains (large interface, indicated as blue dashed line). The homotetramer was formed by dimerization of homodimers, through the interactions of residues on adjacent C domains (small interface, indicated as green dashed line). Four chains of homotetramer adopt almost the same conformation.

The YePykF and YePykF+FBP structures form tetramers through interactions of adjacent subunits between A and C domains. Bordering A domains form A-A or the "large" interface, while adjacent C domains constitute C-C or the "small" interface. In *Leishmania Mexicana,* the transition of PK from inactive T-state to active R-state undergoes a symmetrical 6° rigid-body rotation of each of the four subunits. The transition also involves the formation of 8 salt bridges on the small interface (Morgan et al., 2010; Zoraghi et al., 2011b). Based on the structural differences of the small interfaces of methicillin-resistant *Staphylococcus aureus* (MRSA) PK and mammalian orthologs, a series of natural products was selected with antibacterial function (Zoraghi et al., 2011b). Some activators of human PKM2 bind in a pocket on the large interface but not to the allosteric site, and promote the association of PKM2 subunits into stable tetramers (Anastasiou et al., 2012). Therefore the investigation of the interface interaction of YePykF is really necessary to provide information for drug development.

In the YePykF+FBP asymmetric unit, homodimer forms through interactions of residues on the lager interface (between monomers A and B, or monomers C and D). Three α-helices (α6, Arg244-Lys272; α7, Asn291-Gly305; α9, Gly318-Asp336) from chain A and the corresponding three helices from chain B are involved in this interaction (Figure I.13 a). To analyze the dimer and tetramer assembly, the PDBePISA web server (https://www.ebi.ac.uk/pdbe/pisa) was used. PISA suggests that the large interface plays an auxiliary role in dimer formation. In the YePykF large interface, one salt bridge and 16 hydrogen bonds are formed between monomer A and B, 30 residues from each monomer are involved. The interface area is 1112 Å^2 for monomers A and B on average. In YePykF+FBP structure these interactions increased to 8 salt bridges and 17 hydrogen bonds between 39 residues from monomers A and B, 39 residues from each monomer are involved. The interface area is 1455 Å^2 for monomers A and B on average, which is 343 Å^2 larger than YePykF (Table II.2). The solvent free energy gain on interface formation of YePykF+FBP ($\Delta^i G$ = -13.8 kcal/mol) is also higher than that of YePykF ($\Delta^i G$ = -12.6 kcal/mol). The increased number of interactions and interface areas, and also the free energy gain in the

YePykF+FBP structure indicate tighter interaction between two molecules in the asymmetric unit.
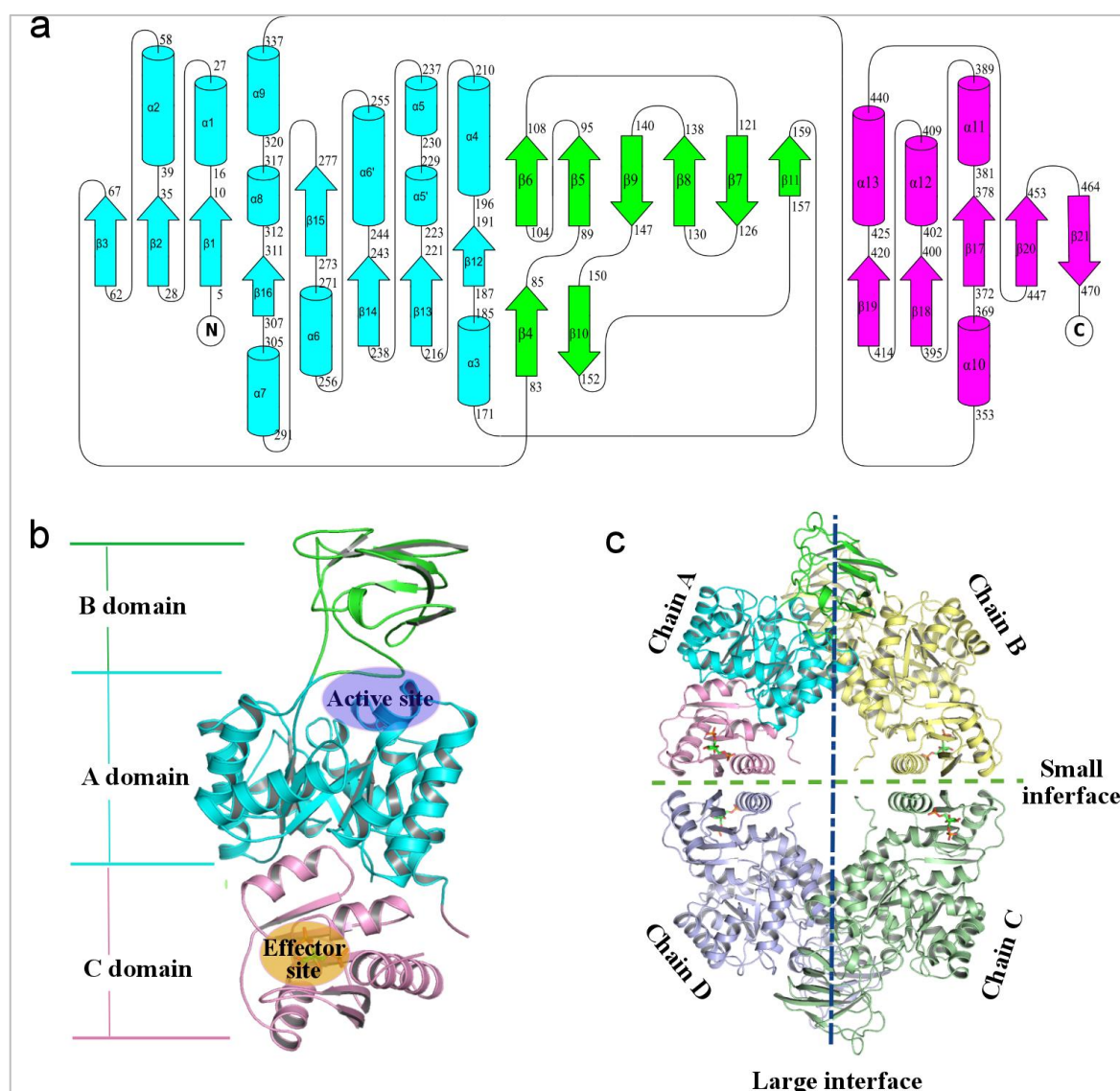
Figure I.12: Overall structure of YePykF. (a) Topology diagram representing the overall folding of YePykF. Secondary structure elements are colored according to domains: A domain - cyan, B domain – green, C domain - pink. (b) A cartoon representation of the YePykF+FBP tertiary structure. The catalytic site (violet), and allosteric site (yellow orange) are indicated. (c) Quaternary structure of YePykF+FBP.

The physiological tetramer forms through interactions of the small interface (between Chain A and Chain D, or between B and Chain C). The water accessible surface of the tetramer is about 72180 Å^2 and about 10300 Å^2 of the surface area is buried in YePykF, whereas about 69430 Å^2 is water accessible and about 11450 Å^2 is buried in the YePykF+FBP tetramer.

Two α-helices (α10, Lys351-Lys372; α13, Ala424-Ser440) and a C terminal loop-β-strand motif (Gly454-Leu470) and the same elements from adjacent molecules are involved in the formation of the small interface (Figure I.14 b).
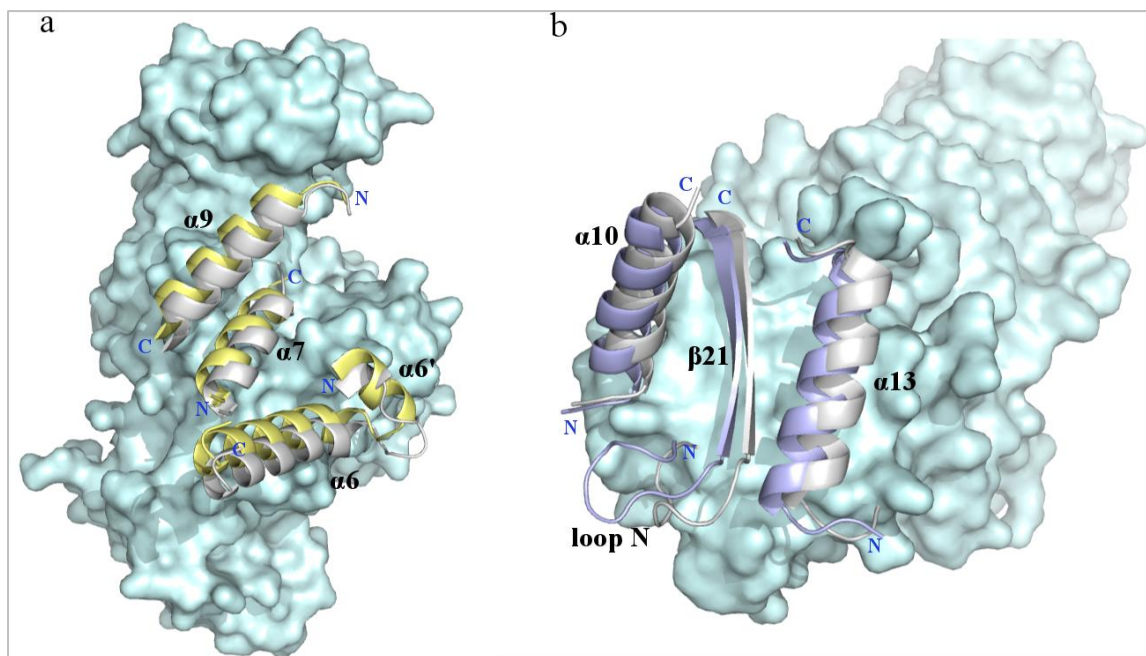
Figure I.13: Interfaces of YePykF. Chain A of YePykF and YePykF+FBP are superposed and shown in surface and colored light green. Secondary elements that involved in the interface interaction are shown as cartoon (a) Large interface of adjacent A domains (PykF+FBP: yellow, PykF: gray); (b) Small interface of adjacent C domains, (PykF+FBP: light blue, PykF: gray). The α-helices and β-strands are numbered according to the protein secondary structure; the N-terminal and the C-terminal of each helix are marked out as blue letters.

In the YePykF structure, 15 hydrogen bonds and one salt bridge are formed by 19 residues from each subunit's C domain and the interface area is 875 $Å^2$ for monomers A and D on average. In YePykF+F16BP, only 8 hydrogen bonds are formed by 14 amino acids, and the interface area 656 $Å^2$ (Table I.2). The solvation free energy gain of YePykF for interface formation (Δ^iG = -6.9 kcal/mol) is lower than that of YePykF+FBP (Δ^iG = -7.4 kcal/mol). The lower solvation free energy gain of YePykF might be the reason for PDBePISA as artifacts of crystal packing. However, the solvation free energy gain upon formation of the tetramer assembly for YePykF ($\Delta^{int}G$ = -190.9 kcal/mol) and YePykF+FBP ($\Delta^{int}G$ = -68.4 kcal/mol) suggests the tetramer is stable in solution. Compare to YePykF, the number of interaction residues and the area of the small interface are decreased in YePykF+F16BP. One explanation for this could be the conformational change of the effector-binding site. FBP and

29

sulfate groups binding sites involve three loops, one of which (loop N, the loop before the N-terminal β21) is involved in interface formation. F16BP binding changes the conformation of loop N and results in less interactions, although there is no hydrogen bond or salt bridge involved in this loop.

Table I.2: Interface properties of YePykF and YePykF+F16BP

Interface	Structure	Chain ID	Interface residues	Interface area (Å2)	Hydrogen bonds	Salt bridges	CSS*
Large interface	YePykF	A B	30	1112	16	1	0.20
	YePyk+FBP	A B	39	1456	17	8	0.46
Small interface	YePykF	A D	29	875	8	4	0
	YePyk+FBP	A D	24	656	8	0	0.19

*: CSS, Complex Formation Significance Score, ranges from 0 to 1 as interface relevance to complex formation increases.

I.3.5 Allosteric site of YePykF

To obtain structural information concerning the allosteric mechanism of YePykF, F16BP was added to YePykF protein in co-crystallization experiments. Three dimensional crystals (Figure I.11) of YePykF+FBP were obtained. X-ray diffraction datasets were collected with synchrotron radiation with the highest resolution limit of 1.9 Å. Molecular replacement was performed against a trimmed model from *E.coli* Pyruvate kinase PykF (PDB code: 1E0T). After manual building in Coot and refinement with Phenix.refine, F16BP electron density is clearly visible in a cleft in the C domain. The F16BP electron density was well defined to the σ level of 3.0 in Fo - Fc omit map and has peak heights greater than 1σ in an unbiased 2Fo - Fc map (Figure I.14 a). F16BP forms salt bridges and hydrogen bonds with neighborhood residues. The 1'- phospho group of F16BP forms direct hydrogen bonds to side chains and backbones with Thr378, Gly379, Lys382 and Ser383. The 6'- phospho group forms direct hydrogen bonds with Arg353 and Thr355. The C2 hydroxyl group of D-fructofuranose forms hydrogen bonds with Thr355 and Thr462. Additionally, some indirect interactions are mediated by water molecule networks involving residues Ala455, Gly380, Gly381 and

Thr461. The distance between the two phosphorous atoms in the YePykF +FBP structure is 8.3 Å.

In the YePykF structure, each monomer in the asymmetric unit binds two sulfate groups (SULF1 and SULF2). This is because YePykF was crystallized in conditions containing a high concentration of ammonium sulfate. As reported previously, the allosteric effector-binding site of YePykF is capable of binding sulfate ions (Morgan et al., 2010). SULF1 occupies a position corresponding to 1- phosphate of F16BP in the YePykF+FBP structure. SULF1 forms hydrogen bonds with hydroxyl groups from Thr378 and Ser383, and extra hydrogen bonds with the nitrogen atoms from Lys382, Gly379, and Ser383. SULF1 also forms hydrophobic interactions with Gly380, Ser459 and Gly460. SULF2 forms bonds that interlink the A and C domains. SULF2 is hydrogen-bonded to the Arg271 from the A domain, and to the Arg353 and Lys382 from the C domain. It also forms hydrophobic interaction with Thr355 from the C domain. The distance between the two sulfur atoms in YePykF structure is 9.2 Å. The interacting residues and distances are shown in Figure I.14 b.

The detailed interactions in the effector-binding site were further analyzed by ligplot. Compared to the sulfate groups in the YePykF structure, F16BP molecules in the YePykF+F16BP structure bind to more amino acids through hydrogen bonds and hydrophobic interactions. 13 hydrogen bonds with 7 residues for F16BP in YePykF+F16BP compare to 10 hydrogen bonds with 6 residues for sulfate groups in YePykF. Hydrophobic interactions with 4 residues of F16BP in YePykF+F16BP compare to 3 residues of sulfate groups in YePykF. Moreover, the distance between the 2 phosphorous atoms of F16BP is 8.3 Å compare to 9.2 Å of the two sulfur atoms (Figure I.15).
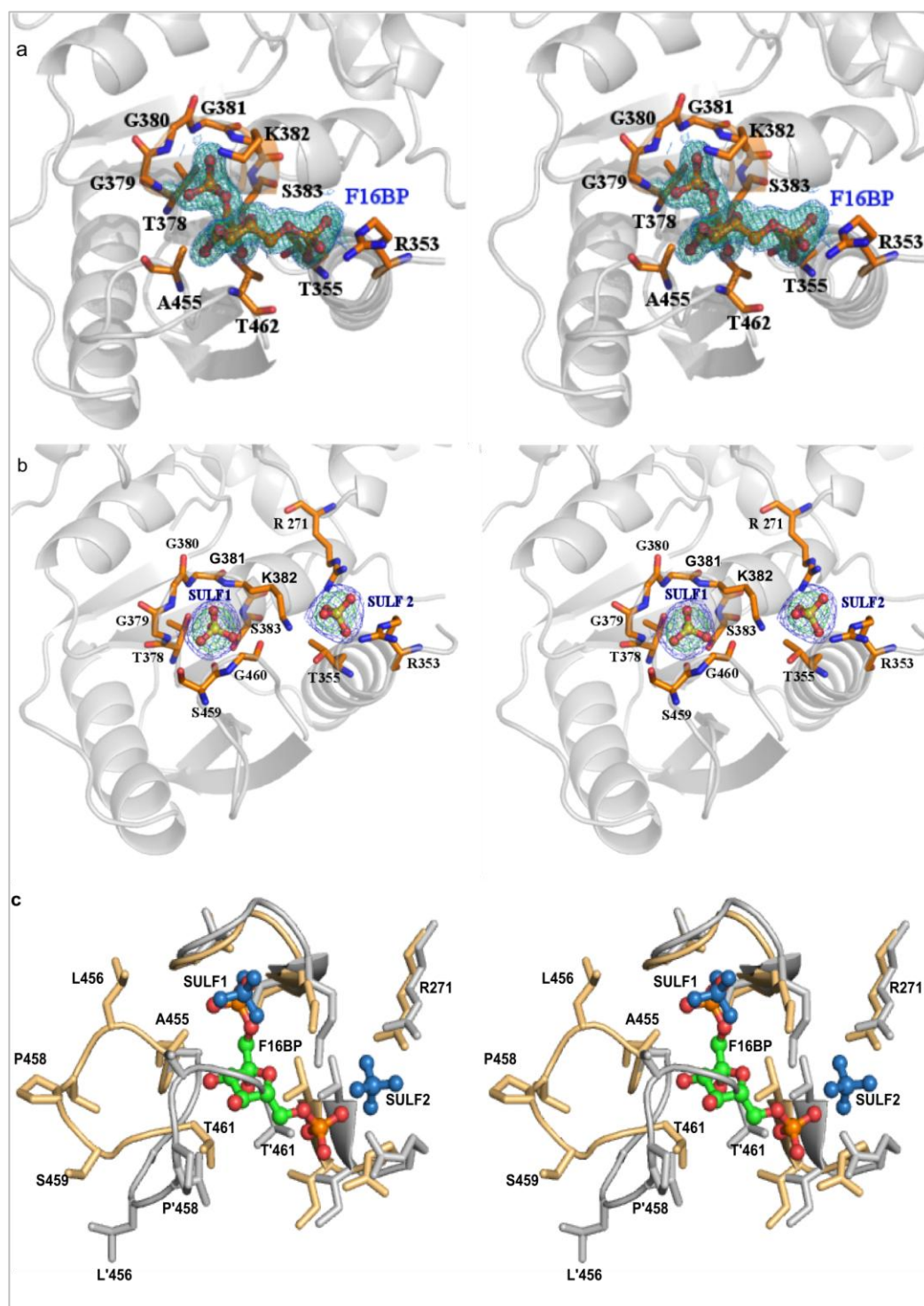
Figure I.14: YePykF Effector binding site. Selected residues are shown as sticks, F16BP and sulfate groups are shown as sticks and spheres. (a) A stereo view of F16BP binding site. F16BP was omitted for calculation of the Fo-Fc omit map (green mesh, 3σ). The 2Fo-Fc map was calculated using the whole YePykF+F16BP refined model (blue mesh, 1σ). (b) A stereo view of sulfate groups binding sites. SULF1 and SULF2 were omitted for calculation of the omit map (green mesh, 3σ). The 2Fo-Fc map was calculated using the whole YePykF refined model (blue mesh, 1σ). (c) Superposition of two effector bindings sites. YePykF is colored grey and YePykF+FBP is colored yellow. Sulfate groups are colored blue and F16BP is colored green and orange.
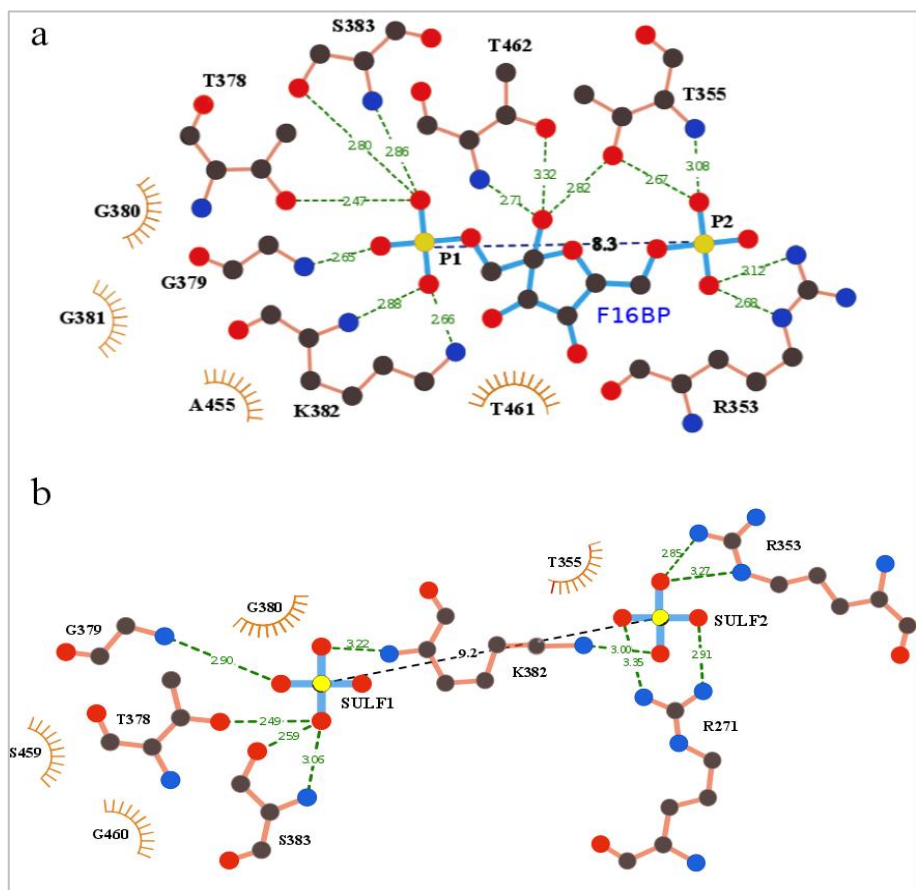
Figure I.15: Schematic representation of interactions in effector site. Black balls represent carbon atoms, red balls represent oxygen atoms, blue balls represent nitrogen atoms, and yellow balls represent phosphorous atoms. Hydrogen bonds and important interactions indicated by dashed lines are colored green and black respectively. Distances are indicated in Angstroms. Hydrophobic contacts are represented by arcs in which spokes radiating towards the contacting ligand atoms are colored brown. (a) Interactions between F16BP and YePykF effector site. (b) Interactions between sulfate groups and the YePykF effector site.

Comparing the two YePykF structures, we could conclude the effector-binding site has a high affinity for negatively charged sulfate and phosphate groups. The effect of F16BP activation is not dependent on phosphate, but on the D-fructofuranose. Superposition of the two effector-binding sites shows a different conformation of a loop region (Gly454 - Asn463). In the YePykF structure, this loop spans two sulfate groups. However, in YePykF+FBP, the D-fructofuranose of F16BP occupies the position of the loop and pushes the loop away (Figure I.14 c). In previously reported structures, only phosphate or sulfate binding in the effector site does not activate PK (Bakszt et al., 2010; Rigden et al., 1999; Zoraghi et al., 2011b), but F16BP and F26BP do. The assumption is that D-fructofuranose of FBP causes a different conformation of the loop region and is responsible for the PK activation.

I.3.6 Structure determination of YePykA+AMP

The type II pyruvate kinase of *Yersinia Enterocolitica* (YePykA) consists of 480 amino acids. Non-tagged YePykA was recombinantly expressed and purified. In the most complete crystal structure model, more than 96.7% of the amino acids could be interpreted in the electron density of the YePykA+AMP structure. However, the completeness of peptide chains varied among monomers in one asymmetric unit. 471, 462, 461 and 463 out of 480 amino acids could be interpreted for chain A, B, C and D, respectively. The first N-terminal methionine is not visible possibly because of its high flexibility or posttranslational modification.

The secondary structure of YePykA was also determined from the DSSP algorithm and is shown in Figure S1. YePykA consists of similar secondary structure elements in specific positions to that of YePykF. Although the sequence identity between YePykF and YePykA is only 36.7%, YePykA subunits still share similar architecture with YePykF (Figure I.16 a). The tertiary structure of YePykA can also be divided into 3 distinct domains: the rigid A domain (residues 1 to 74 and 172 to 348, cyan), composed of an (α/β)8 barrel and small α-helices; the flexible B domain (residues 75 to 171, green), composed of 7 β-strands, one small α-helix and almost 50% random coils; the C domain (residues 349 to 480, light pink), composed of helices and strands forming an αβαβαββ topology. The mainly difference between YePykA and YePykF is located in B domain. Seven β-strands and one small helix were found in YePykA's B domain, whereas 11 β-strands and no helix were found in YePykF's B domain.

The four monomers in the asymmetric unit are arranged as a tetramer similar to type I PKs. Every two monomers interact with each other through hydrogen bonds and salt bridges. The webserver PDBePISA was again used for analyzing the assembly of the asymmetric unit. The water accessible surface of YePykA tetramer is 80969 Å^2 and about 4484 Å^2 of surface area is buried. The analysis revealed the largest interfaces could be found between the monomers A and B as well as C and D (1132 Å^2 on average). The smaller interfaces could be located between monomers A and D as well as B and C (1042 Å^2 on average). The detailed information about the number of residues and bonds, which are involved in the interface formation, are listed in Table II.3. The solvation free energy for the large interfaces ($\Delta^i G$ = -17.9 kJ/mol on average) is higher than that of the small interfaces ($\Delta^i G$ = -11.4 kJ/mol on average). PDBePISA judges all the interfaces as 100% relevant for complex formation and suggests the tetramer as the physiological relevant oligomer.
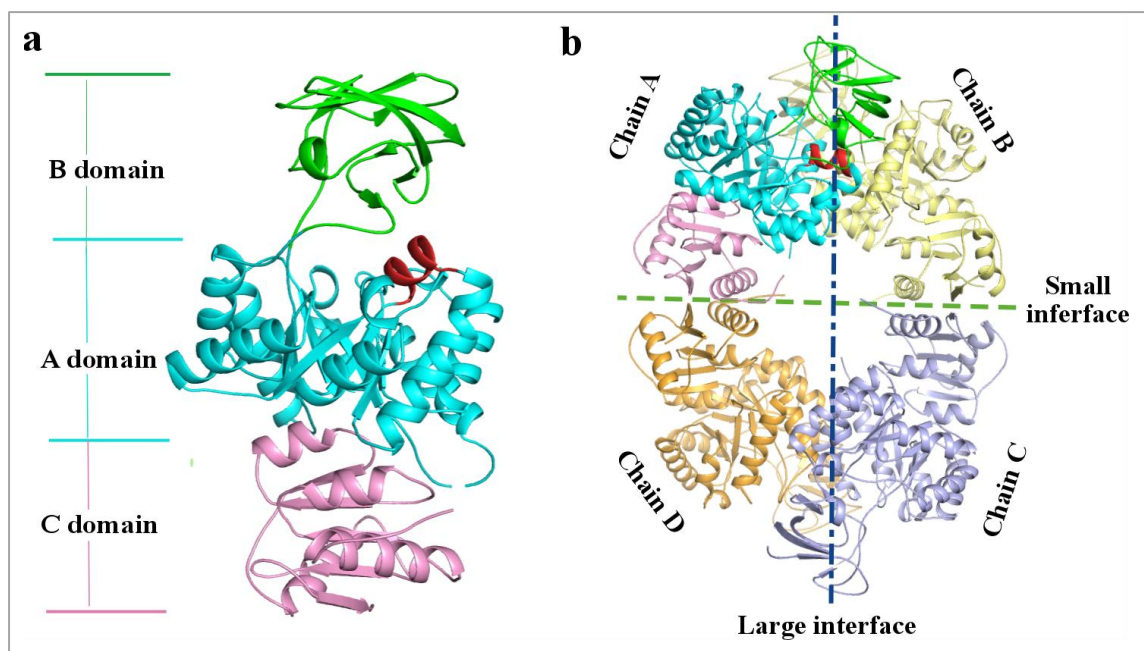
Figure I.16: Structure of YePykA. (a) A cartoon representation of the YePykA monomer. Structural domains are colored differently; A domain – colored cyan, B domain – green, C domain – light pink. The Aα6' helix is colored in red. (b) A cartoon representation of the YePykA tetramer. Chain A is colored according to domains; chain B, chain C and chain D are colored as yellow, light blue and orange, respectively.

Table I.3: Interface properties of YePykA tetramer

Interfaces	Chain ID	Interface residues	Interface area (Å^2)	Salt bridges	Hydrogen bonds	CSS*
Large interface	A	32	1147	3	7	1
	B					
	C	43	1256	6	9	1
	D					
Small interface	A	28	1030	3	11	1
	D					
	B	27	1053	0	11	1
	C					

*CSC: Complex Formation Significance Score, CSC ranges from 0 to 1 as interface relevance to complex formation increases.

I.4 Discussion and outlook

I.4.1 YePykF allosteric mechanism

As was reported before, PKs exist in either the T-state (inactive) or the R-state (active) enzymes. The T-state and R-state enzymes can switch between each other depending if the allosteric effector is present. One exception is the mammalian muscle isoenzymes, which always exists as the R-state and constitutively active in absent of allosteric effectors (Rigden et al., 1999). In the last 20 years, many efforts have been made to explore the mechanism of allosteric regulation. Researchers have tried to compare different states of enzymes from different organisms, and speculated that the allosteric regulation mechanism might be the effector induced B domain closure (Christofk et al., 2008; Dombrauckas et al., 2005; Rigden et al., 1999; Tulloch et al., 2008). For example, Mattevi et al. (1995) compared the T-state *E. coli* PK with the non-allosteric rabbit muscle enzyme PKM1, which is thought to adopt a conformation similar to the active R-state, and revealed the allosteric strategy is the rotations of B and C domains by $17°$ and $15°$, respectively. The comparison between the F16BP-binding human PKM2 (hPKM2) with the T- and R-states of rabbit muscle PKM1, revealed the partial B domain closure after F16BP binding (Dombrauckas et al., 2005). However, these evidences are not based on PK structures from the same organism, and B domain is even reported to adopt different conformations among different subunits in the same PK tetramer (Bakszt et al., 2010). To prove the effector induced B domain closure allosteric strategy, direct evidence about the effector-free and effector-bound PK structures from the same organism are needed. Pyruvate kinase from *Leishmania mexicana* has been studied to a large content. The transition of *Leishmania mexicana* pyruvate kinase from T-state to R-state can be summarized as a simple symmetrical $6°$ rigid body rocking motion of the AC cores (A domain and C domain) in each of the four subunits, and the formation of eight salt bridges across the small interface, but the B domain movement is not reported as allosterically regulated (Morgan et al., 2010). Commonly, there is a significant structural divergence between prokaryotes and eukaryotes, which has hampered a detailed understanding of the allosteric regulation mechanism of prokaryotic PKs.

In this work, *Yesinia* PK structures of both T- and R- states from the same species have been solved. The comparison between them revealed a detailed allosteric regulation mechanism of prokaryotic PKs. To quantify the structural differences upon F16BP binding, I superposed the

Cα atoms of the AC domains from YePykF and YePykF+FBP onto each other, and calculated the Cα r.m.s. deviations of each residue. Superposition of YePykF_chain_A and YePykF+FBP_chain_A shows slight difference in chain A but tremendously conformational change in the B domain of chain B. (Figure I.17 a). The B domain has a maximum Cα r.m.s. deviations of 15 Å and 9.6 Å on average. Superposition of asymmetric dimer of two structures shows similarly conformational changes in both subunits (Figure I.17 b). The average Cα r.m.s. deviations for B domains of two subunits are 5.6 Å and 4.1 Å, compared to the AC domains whose Cα r.m.s. deviations are 0.9 Å and 0.92 Å, respectively. Despite B domain, a loop region (residue 452 to residue 463) also shows large Cα r.m.s. deviations with an average value of 5.3 Å. Checking with the structure, this loop region is exactly involved in F16BP binding. The conformational change of the loop region was already discussed in I.3.5

From the superposition results we could derive that YePykF has an open conformation of B domain and belongs to the T-state, whereas YePykF+F16BP has a closed conformation of B domain and belongs to the R-state. The closing of the B domain narrows the cleft between A and B domains. In order to quantify the B domain movement, the distance between domain B and domain A were measured. Distances of the Cα carbons of residues Asp127 (from domain B) and Val249 (from domain A) were measured to indicate the biggest distance between the domain A and domain B. The distance was 10.9 Å in YePykF, and this value decreased to 6.7 Å in YePykF+FBP (Figure I.17 a). The molecular parameters LWH (length, width, height) were also measured to quantify this B domain movement effect. The orthogonal parameters of YePykF are $137 \times 86 \times 96$ Å^3, and these were decreased to $132 \times 84 \times 86$ Å^3 in YePykF+F16BP structure (Figure I.18), which can be interpreted as B domain movement yields a smaller spatial volume and tighter stacking of monomers.
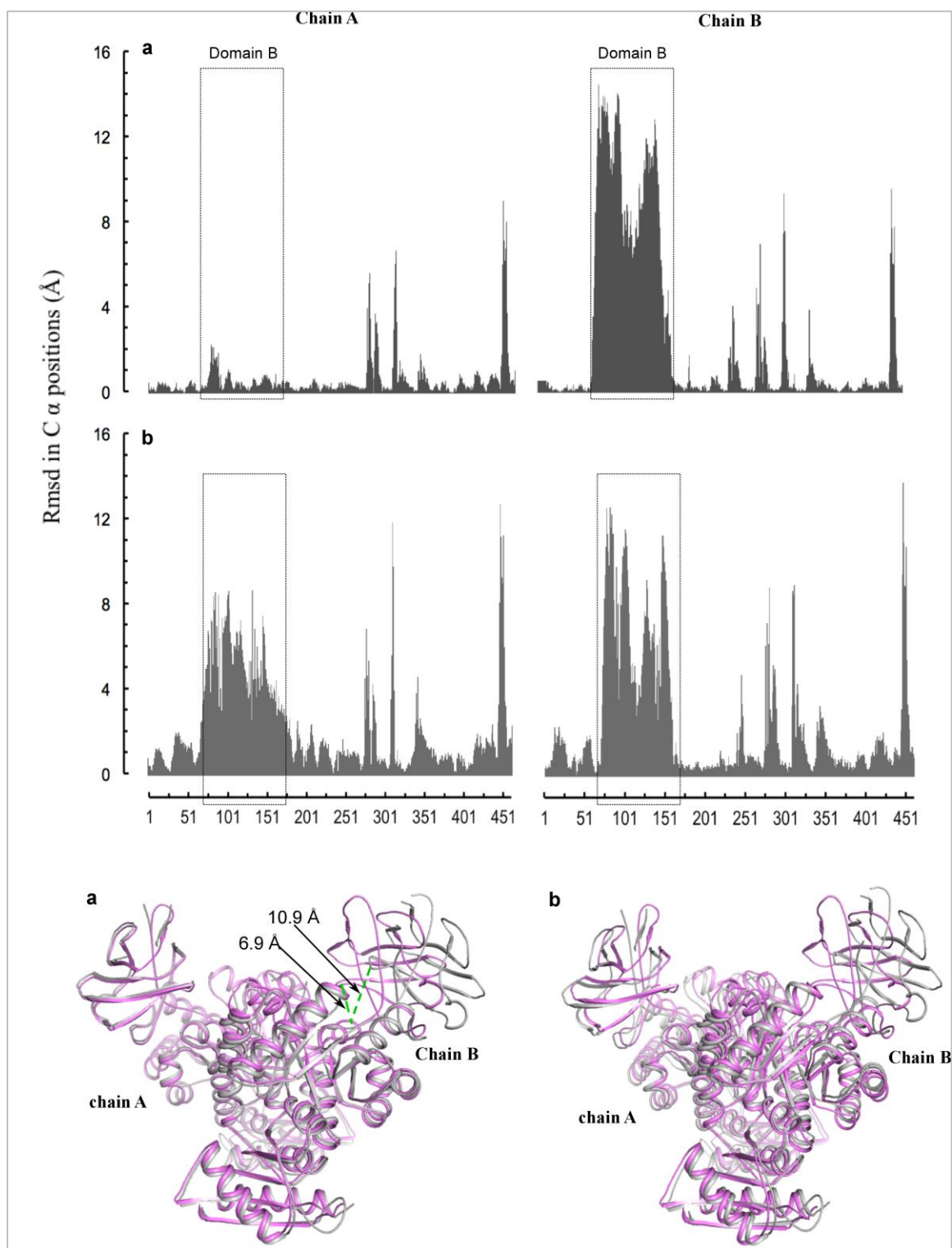
Figure I.17: Superposition of YePykF+F16BP and YePykF. (a) Superposition of the Cα atoms of YePykF+F16BP_chain_A and YePykF_chain_A. (b) Superposition of the Cα atoms of YePykF+FBP and YePykF asymmetric dimer. The diagram shows the Cα r.m.s. deviations of every residue, and the black dashed line boxes indicate B domains. The pictures below show the corresponding superposition of two structures in Pymol, YePykF is colored in gray, YePykF+F16BP is colored in violet, and the distances between Asp127 and Val249 are indicated in green dashed lines.
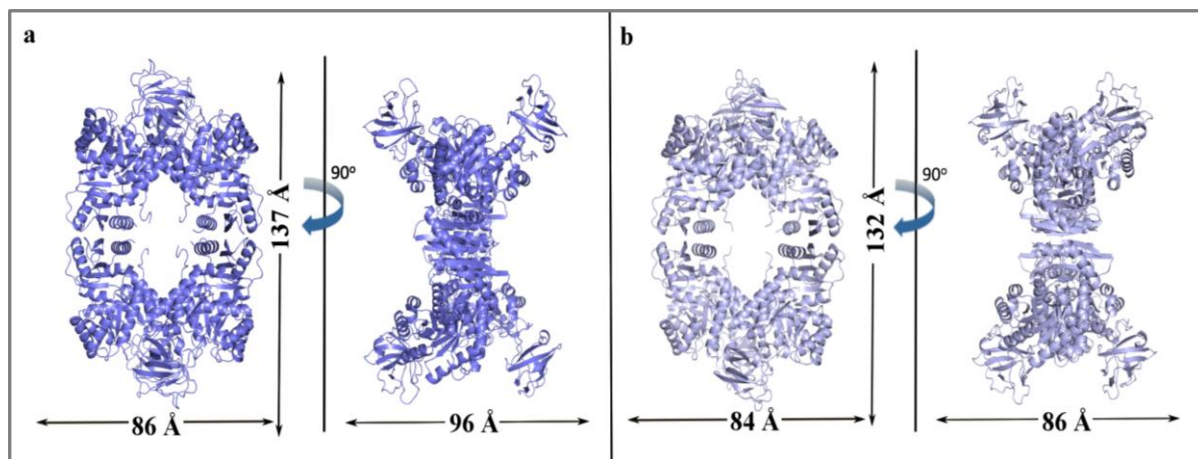
Figure I.18: Orthogonal view of homotetrameric YePykF and YePykF+F16BP. (a) The orthogonal parameters of YePykF are $137 \times 86 \times 96$ Å^3. (b) The orthogonal parameters of YePykF+F16BP are $132 \times 84 \times$ 86 Å3.

To further describe the domain motions, the program DynDom (Hayward and Berendsen, 1998; Hayward and Lee, 2002; Poornam et al., 2009) was employed. The DynDom program measures the rotation degree and translation of one domain of the protein against the protein rigid body. The moving domains and bending residues calculated by DynDom are exactly the B domains and their domain borders. Based on DynDom analysis, the conformational change of YePykF upon F16BP binding can be concluded as follows: B domain pivots around residues 68-71 and residues 165-171 and rotates towards A domain by 9.6° and 19.6° for chain A and chain B respectively (Table I.4). The average value of the B domain rotation is 14.6°. The rotation of B domain changes the volume of the active cleft, and yields a partial closure active site.

Table I.4. DynDom analysis of two chains of YePykF and YePykF+F16BP

Confomer 1	Confomer 2	Moving domain	Bending residues	Rotation angle (°)	Closure (%)	Cα RMSD of AC domain (Å)
PykF chain A	PykF+FBP chain A	-	-	-	-	-
PykF chain B	PykF+FBP chain B	69-170	68-69 165-171	28.8	95.7	1.42
PykF chain A	PykF+FBP chain B	65-186	64 -66 186-187	9.6	95.6	1.36
PykF chain B	PykF+FBP Chain A	70-168	69 -71 167-169	19.5	88	1.47

I.4.2 Catalytic site of YePykF

As is shown before, the catalytic site of YePykF and YePykA should be located in a pocket between A and B domains of each subunit, assuming conservation of the catalytic region of PKs. A two turn Aα6' helix (243 RGDLGVE249), which is considered necessary for enzyme activity, is highly conserved (Cook et al., 2012). According to Morgan et al. 2010, residues (Arg32, Asn34, His37, Arg73, Lys220, Glu222, Gly245, Asp246, Thr278 and Ser312) involved in substrate binding are well conserved in yeast, rabbit, human, and *E. coli* sequences (Figure I.19 a). This conservation is also present in *Yesinia* pyruvate kinases sequence for YePykF and YePykA, and the conserved residues could be located in the concave between A and B domains. Despite Aα6′ helix, all catalytic residues were located on or next to the β-strands core of the (α/β)8 barrel. β-strands Aβ2 and Aβ5 Aβ are directly involved, whereas Aβ3, Aβ5 and Aβ7 associated loops are involved in the enzyme catalytic activity (Figure I.19 b).
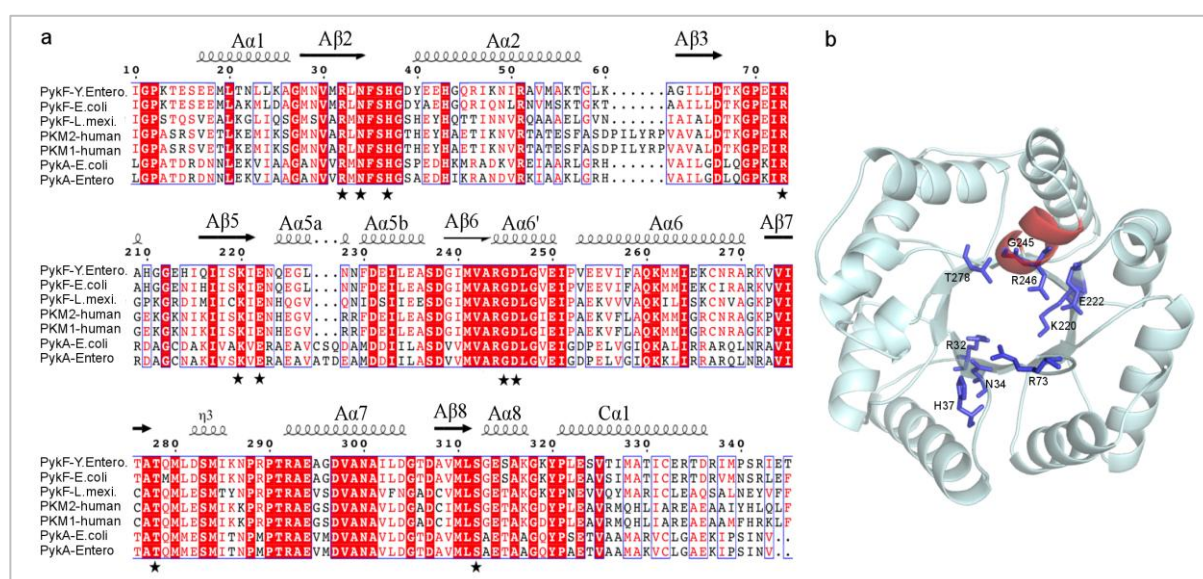
Figure I.19: Catalytic site of YePykF. (a) Structure based sequence alignment of active site associated sequence of pyruvate kinase from different species. Identical residues are highlighted in red and similar residues are surrounded by a blue box. The conserved residues involved in the active site are marked out with black stars. (b) Cartoon view of catalytic residues in YePykF structure. Aα6′ helix is colored in red, and catalytic residues are shown as sticks and colored in blue.

In the case of YePykF chain B for example, B domain pivots around residues 68 - 69 and residues 165 - 171 and rotates towards A domain by 28.8 °. This B domain movement decreased of the biggest distance between domain A (Val249) and domain B (Asp127) from 10.9 Å to 6.7 Å and thus makes the concave between A and B domains smaller. The smaller spatial volume of active site could facilitate the substrate binding and stabilize it. Once the catalytic reaction was performed, the product was released and gives space for new substrate.

Apart from the B domain movement, the Aα6' helix is found unwound in YePykF but helical in YePykF+FBP structure (Figure I.20). The unwinding of Aα6' helix has also been reported in the inactive state of *Cryptosporidium parvum* and *Leishmania Mexicana* PKs (Cook et al., 2012; Morgan et al., 2010). As was discussed, the Aα6' helix is necessary for the enzyme activity. This has led to proposition of a model for the structural rearrangement at the active site (Cook et al., 2012; Morgan et al., 2010; Tulloch et al., 2008) from R- to T-state.
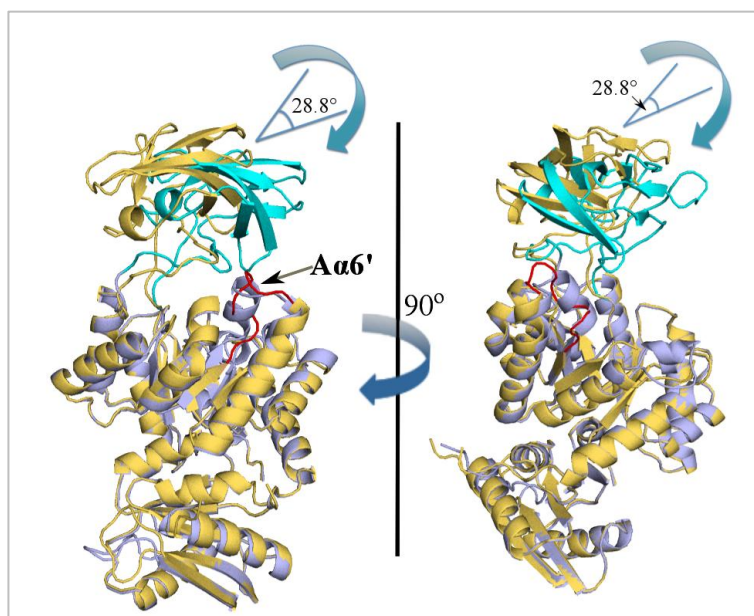
Figure I.20: Comparison of YePykF (T-state) and YePykF+FBP (R-state). Superposition was done with YePykF_chain_B and YePykF+FBP_chain_B. YePykF is colored light yellow, the unwound Aα6' helix is colored red. The B domain (residues 71−167) is shown in cyan for the YePykF +FBP structure, and the AC domain is violet. The Cα atoms r.m.s. derivations value of domains A and C (residues 1 − 70 and 168 − 470) is 1.42 Å. The 28.8 °movement of domain B corresponding to the closure of the active site is highlighted.

The electrostatic surface potential of YePykF was also analyzed to characterize the active site. The electrostatic surface potential was calculated by APBS (Baker et al., 2001), which shows

the active pocket to be highly negatively charged (Figure I.23 a). Since ADP and PEP are negatively charged, the substrates binding must have the aid of metal ions such as K^+ and Mg^{2+}. According to Morgan et al., substrates ADP, PEP and metal ions (K^+, Mg^{2+}) bind to the active site through hydrogen bonds, ionic bonds and mediating water molecules.

The MolProbity (Davis et al., 2007) geometry analysis of YePykF+FBP model shows that two residues (Thr278 and Ser312) that are crucial for PK catalysis exhibit geometry out side of the favored regions of the Ramachandran plot, although they correlate well with the electron density map (Figure I.21). This phenomenon has also been observed in many other PK structures. The abnormal geometry of residues Thr278 and Ser312 is due to restricted geometry, which facilitates the interactions with the substrates (Mattevi et al., 1995; Morgan et al., 2010). However, the abnormal geometry didn't show in YePykF structure. It is speculated that the abnormal geometry of Thr278 and Ser312 is due to the presence of F16BP, although it's not clear how F16BP affects these two amino acids. It is also possible that the absence of the two residues' abnormal geometry in YePykF is due to the low quality of the X-ray diffraction data.
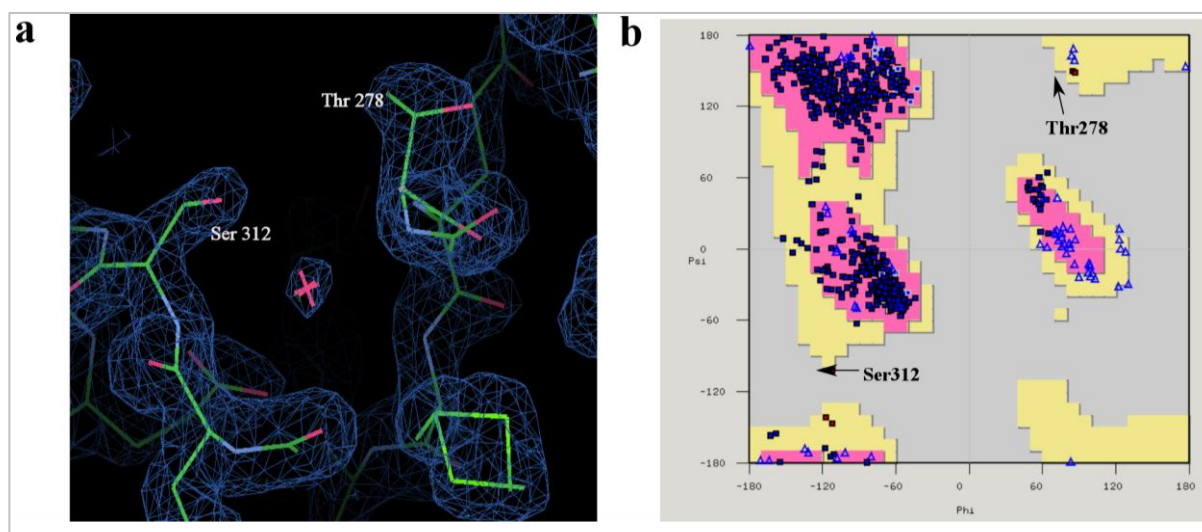
Figure I.21: Geometry of active site of the YePykF+FBP model. (a) Electron density map of Thr278 and Ser312, contoured at 1.5 σ. (b) Ramachandran plot of refined structure model of YePykF+FBP.

I.4.3 Effector binding sites of YePykF and YePykA

In the presence of allosteric effectors, the catalytic ability of PKs is clearly higher than without allosteric effectors (Morgan et al., 2013; Zhong et al., 2013). The residues involved in F16BP binding of YePykF are well conserved in type I *Enterobacteriaceae* PKs, but show less conservation compared to type II PKs and type I PKs in other species (Figure I.22). The electrostatic surface potential shows that the effector-binding position is highly positively charged, which facilitates the binding capacity of negatively charged F16BP molecules (Figure I.23 b). The first binding effect is the conformational change of the binding site. Comparing the effector-binding site of the YePykF+FBP and YePykF structures reveals that the loop N (Gly454 - Asn463) which spans between two sulfate ions in YePykF, changes its conformation upon F16BP binding. The effector-binding site is capable of binding sulfate and phosphate but these ions could not activate PK. We speculate that the activation strategy of F16BP is not depending on phosphate groups, but on the presence of the D-fructofuranose moiety of F16BP. The phosphate groups of F16BP merely facilitate the binding but not involved in the activation of YePykF. From the superposition, we found that the D-fructofuranose in YePykF+FBP structure occupies the position of the loop N and pushes the loop out. The primary conformational change of the loop N might cause a hinge motion and lead to the secondary conformation change of the B domain, which is 40 Å away from this effector-binding site, and thus trigger the activation of the enzyme.
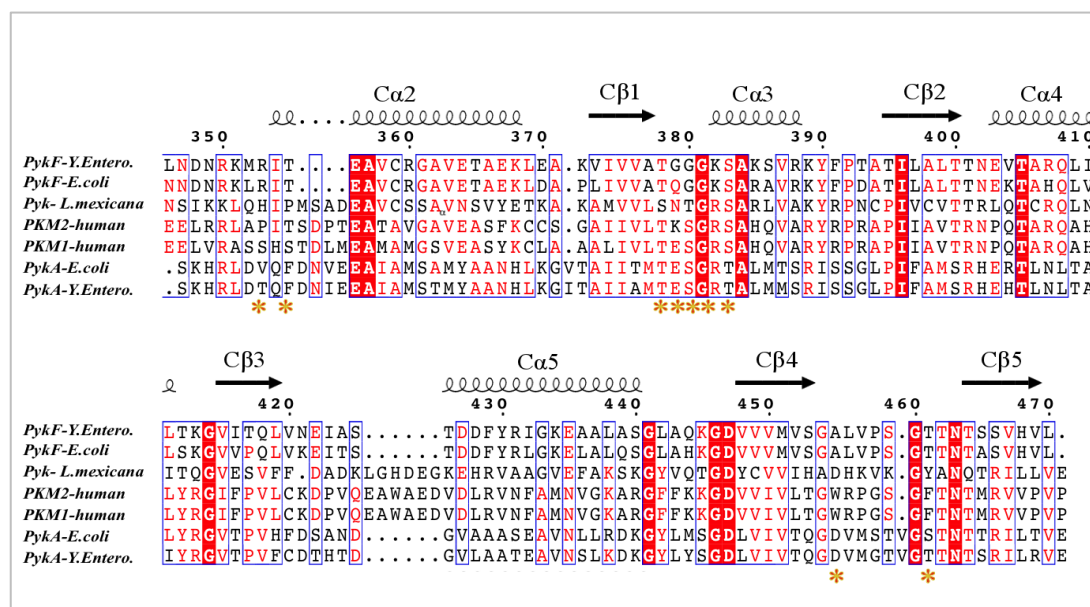
Figure I.22: Structure based sequence alignment of the effector binding site associated sequence of pyruvate kinase from different species. Identical residues are highlighted in red and similar residues surrounded by a blue box. For YePykF, the residues involved in F16BP binding are indicated by asterisks.
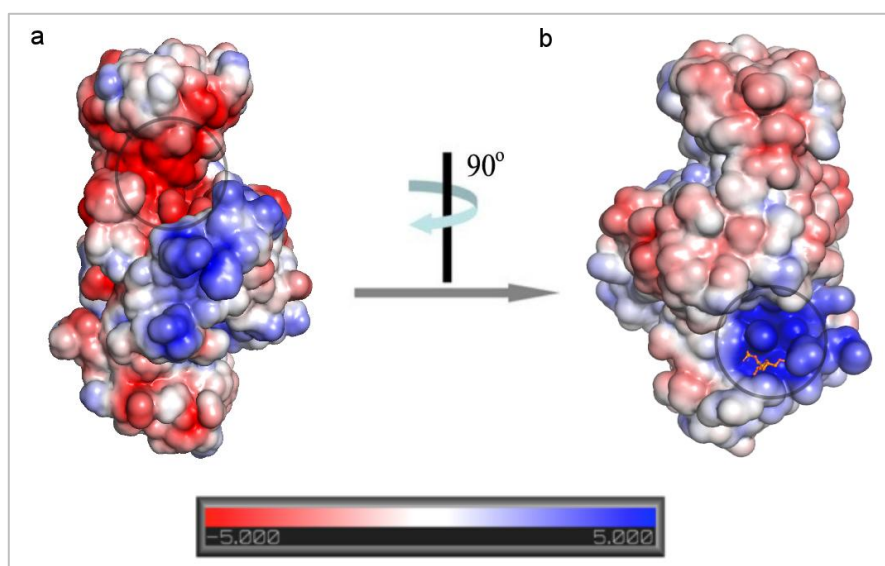
Figure I.23: Electrostatics of YePykF+FBP showing the empty active site and F16BP binding site. Blue: positive; red: negative.

Due to the low resolution X-ray diffraction data of PykA+AMP crystals, no electron density for AMP molecules could be seen in the electron density map. So far, only one type II pyruvate kinase structure from *Geobacillus stearothermophilus* (GstPK, PDB code: 2E28) has been solved at a resolution of 2.4 Å (Suzuki et al., 2008). GstPK has an extra C-terminal sequence (ETCS) composed of 110 residues. The ETCS forms a C' domain, which is next to the C domain. A sulfate ion was found in the C domain. The sulfate-binding site resembles the 6-phosphte group of F16BP-binding site in yeast PK, and is proposed to be the effector-binding site. According to Suzuki et al., the sulfate-binding involves two threonine residues (Thr381 and Thr386) and a histidine residue (His425) (Figure I.24). The sequence identity between YePykA and GstPK is only 31.2%, but the sequence alignment shows that the three residues (Thr386, Thr391, his431) in YePykA are conserved with those in GstPK. These three residues associated loops and helix might be the AMP binding site. However, since little information is known about type II PKs, they could also have a quite different effector-binding site from the type I PKs. In our YePykA structure model, strong unexplained electron density was observed at His375 imidazole group. The His375 might also be the potential AMP binding site. But the probability is very low, because the imidazole group of His375 points out of the monomer and capable of binding metal ions. The strong electron density might due to the binding of unknown metal ions.
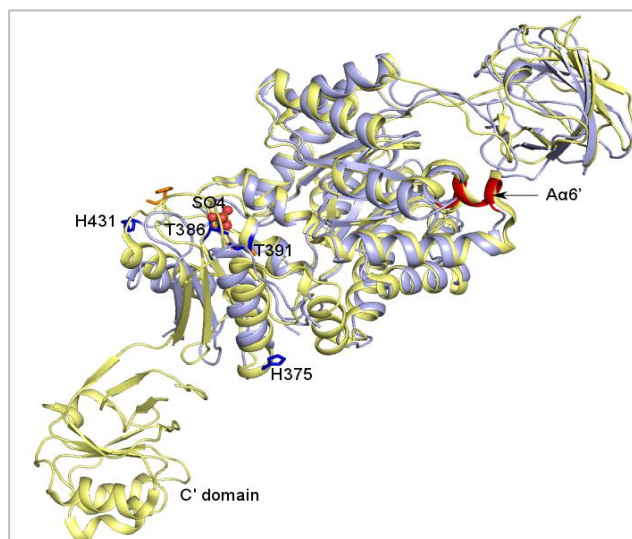
Figure I.24 Superpostion of YePykA and GstPKNC reveals potential effector binding site. YePykA and GstPKNC are shown in cartoon. YePykA is colored in light blue; the predicted active site Aα6′ helix is colored in red; the potential effector binding residues are shown in sticks and colored in blue. GstPKNC is colored in light yellow; the sulfate group is shown in sticks and sphere; the sulfate binding residues are colored in orange.

I.4.4 Allosteric mechanism of YePykF

That substrate-binding causes B domain closure of PKs has been reported for *Leishmania mexicana* (Morgan et al., 2010) and *Trypanosoma brucei* (Zhong et al., 2013). The X-ray structure determination of YePykF and YePykF+F16BP revealed the B domain closes partially after the effector molecules bind to the allosteric site. In summary, the allosteric mechanism of prokaryotic PKs could be proposed as the effector-induced B domain movement, which leads to a partial closure of the active site, and changes the conformation of PK enzyme from T- to R-state (Figure I.25 b). The active site closure is not only depending on the substrate binding, but also on the effector binding to the allosteric site (Figure I.25 d). Allosteric sites usually show less conservation among species than highly conserved active sites. Based on this, the structural information of diverse PK isoforms, especially the allosteric sites, could eventually be utilized for drug design.
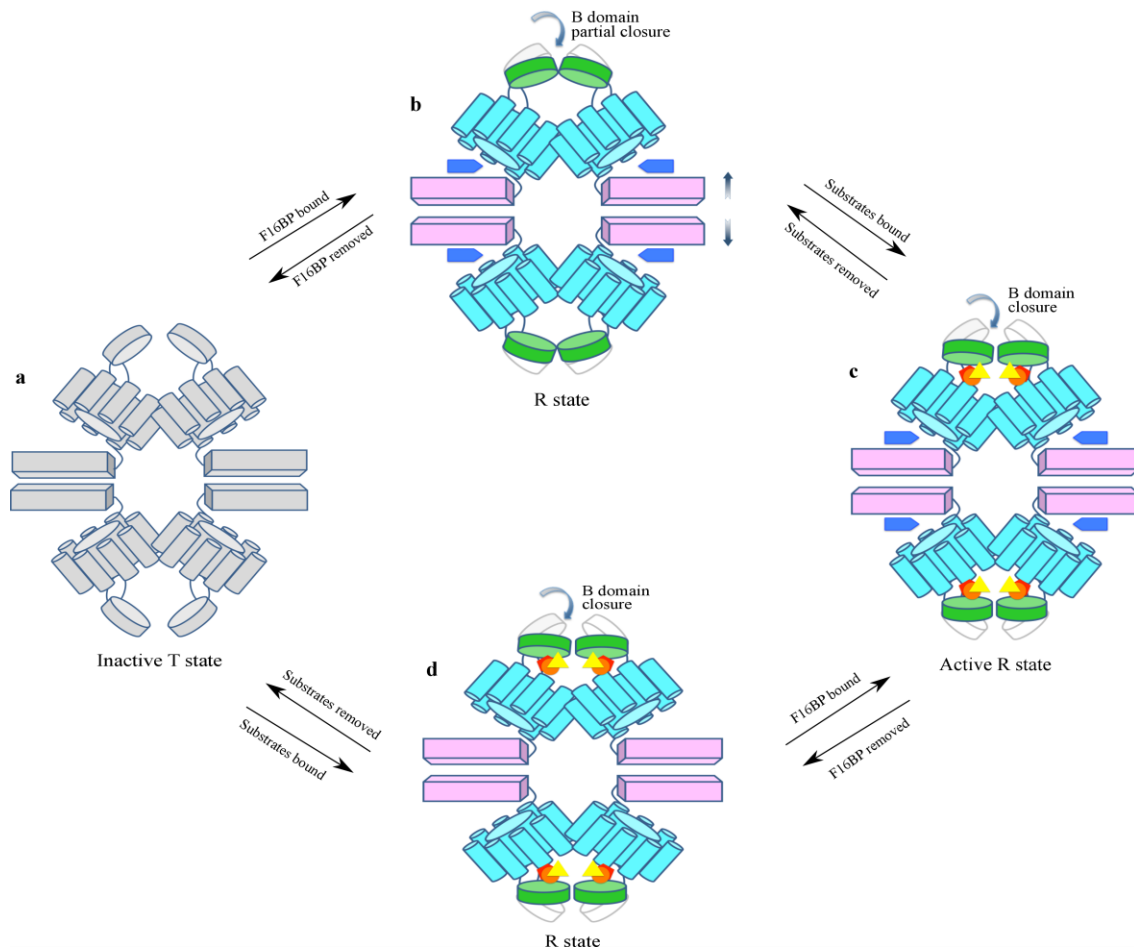
Figure I.25: YePykF crystal structures revealed the allosteric mechanism. (a) Schematic representation of the inactive T state YePykF structure (apoenzyme); (b) Binding of F16BP causes a partial closure of B domain and a broader C-C distance, YePykF changes from T state to R state; (c) Binding of substrates to the YePykF+FBP, theoretically causing a completely closure of B domain; (d) Binding of substrates to the apoenzyme causes a B domain movement and changes the enzyme into R state. (b, c, d) YePykF are colored in different domains, A domain, cyan, B domain, green, C domain, pink.

I.4.5 Catalytic function of YePykF and YePykA

Based on the sequence alignment result of PKs, YePykA conserves the catalytic site similar to YePykF. The Aα6′ is still present in the YePykA+AMP's structure and clearly defined as a two turn helix. The catalytic site of YePykA can be similarly located in the pocket between the B and A domains as in YePykF. As was mentioned before, PykA has lower catalytic ability in all tested conditions (Hofmann et al., 2013; Ponce, 1999; Siddiquee et al., 2004). The reason why *Enterobacteria* need two types of pyruvate kinases is still not clear. By superimposing YePykF+FBP and YePykA+AMP, two structures showed high similarity

(Figure I.26). Two loop regions (residues 83-85 and 150-152, colored pink) in YePykA+AMP structure are defined as β-strands in YePykF+FBP structure. This might due to the low data quality of YePykA+AMP structure. But one well-defined β-strand (residues 162-164, colored yellow) in YePykA+AMP structure is present as a loop in YePykF+FBP structure. This β-strand is located between B and A domains. We speculate the extra β-strand in YePykA+AMP might restrain the flexibility of B domain and also affect the spatial volume of the active site upon substrates binding. Hence the small β-strand might contribute to the lower catalytic ability of YePykA. To prove this, the future research should focus on getting better quality of YePykA and YePykA plus substrates structures.
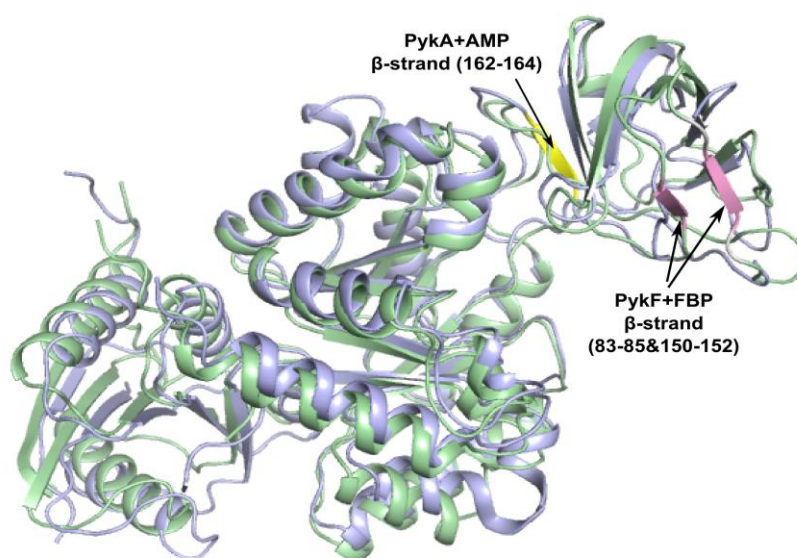
Figure I.26: Superposition of YePykA+AMP and YePykF+FBP structures. Structures are shown in cartoon. YePykA+AMP is colored light blue, the β-strand (residues 162-164) is colored yellow. YePykF+FBP is colored in light green; two β-strands (residues 83-85 and 150-152) are colored pink.

II Towards structural elucidation of hepatitis C virus (HCV) – host receptor interactions

II.1 Introduction

II.1.1 Viral infection

A virus particle (also called virion) is a microscopic infectious entity consisting of only one type of genetic material, either DNA or RNA, surrounded by a protein, lipid or glycoprotein coat. A virus is not capable of replicating itself but requires the use of the machinery and metabolism system of the cells (Figure II.1). Viruses can infect all living forms, including animals, plants, bacteria and archaea (Koonin et al., 2006). Viral infections are more systemic than localized bacterial infection. Antibiotics usually target the bacteria replication machinery and enzymes, systems that viruses do not possess. Hence viral infections cannot be treated by antibiotics.
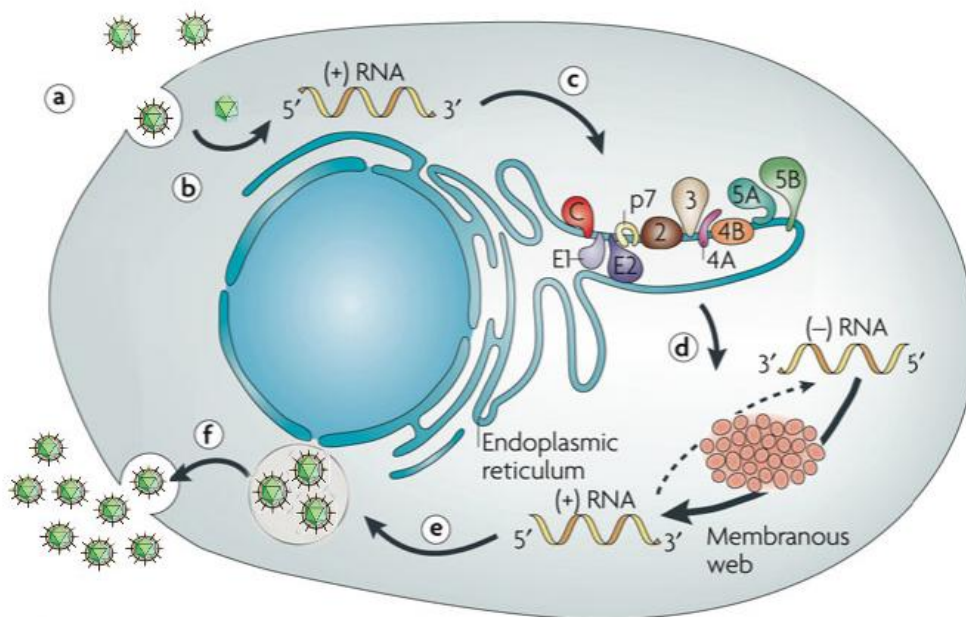
Figure II.1: A schematic diagram of the virus replication cycle (for example, HCV follows the same infectious path). (a) Attachment and endocytosis; (b) uncoating; (c) IRES-mediated translation and polyprotein processing; (d) RNA replication; (e) virion packing; (f) virion maturation and release. Adapted from Moradpour et al. (2007).

II.1.2 Hepatitis C virus infection

Hepatitis C virus (HCV) is the major causative pathogen of acute and chronic liver diseases, liver cirrhosis, hepatocellular carcinoma and liver failure. Approximately 160 million people are infected worldwide, with an estimated 130 million with a chronic infection (Lavanchy, 2009). Among all the cases, ~27% liver cirrhosis and ~25% of hepatocellular carcinomas are correlated with HCV infection, which makes chronic hepatitis C infection a major indicator for liver transplantation. No vaccine is available and current therapies are effective against some genotypes but not all (Simmonds, 2013).

HCV is a member of the genus Hepacivirus within the family *Flaviviridae.* It has a positive-strand RNA genome and lipid associated envelope. Due to the high genomic variability, HCV has seven genotypes (numbered 1-7), most of which have subtypes, named a, b, and so on. The 9.6 kb RNA genome contains one long open reading frame (ORF) flanked by highly structured 5' and 3' non-translated regions (NTRs) (Figure II.2 a). The ORF encodes a polyprotein containing ~ 3000 amino acids. The polyprotein is processed into 10 different structural and non-structural proteins by protease. The structural proteins and P7 are generated by host signal peptidase or signal peptidase mediated cleavages. NS2 and non-structural proteins are processed by the viral cysteine protease NS2-3 and serine-type proteaseNS3-NS4A complex. Structural proteins (core protein, E1, E2), P7 and NS2 are classified as the assembly module. P7 and NS2 play the supporting role for viral particle assembly but are not incorporated into the virus particle. The non-structural proteins NS3, NS4A, NS4B, NS5A and NS5B, which are function for virus RNA replication, are classified as replication module (Figure II.2 a). Every protein is partially embedded into intracellular membranes by one or several transmembrane elements. E1 and E2 form heterodimer, NS5A form homodimer, and P7 forms oligomeric complexes (Figure II.2 b).
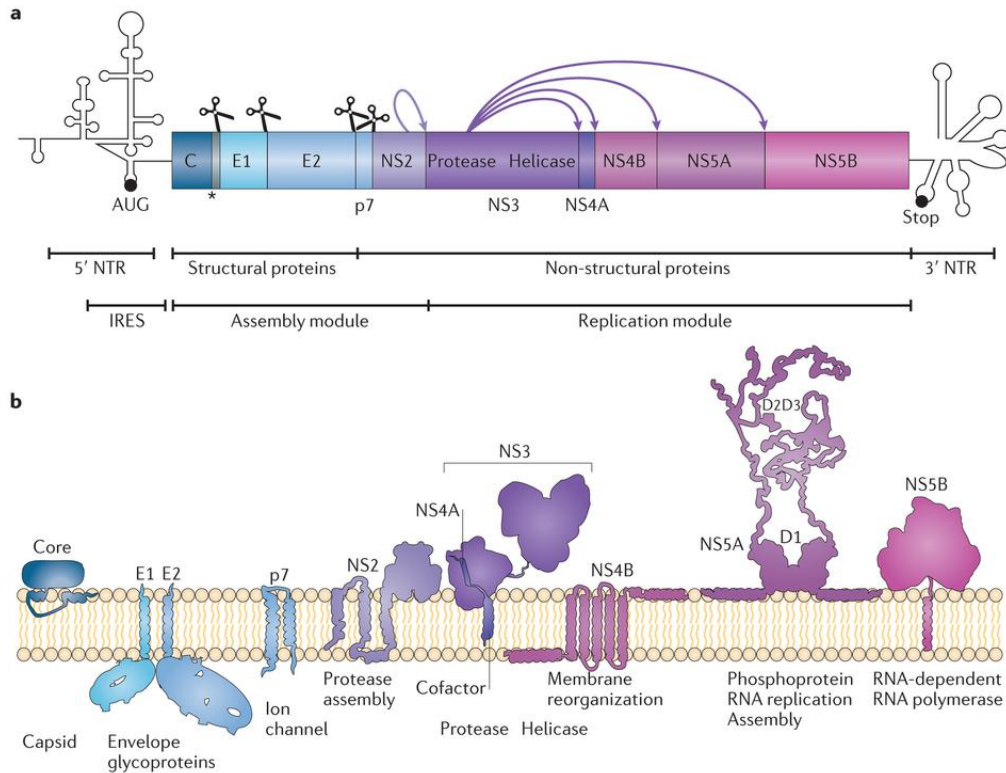
Figure II.2: Hepatitis C virus genome organization and the membrane topology of cleaved viral proteins. (a) The hepatitis C virus (HCV) ORF encoding a polyprotein, and the predicted secondary structures of 5′ and 3′ NTRs. Start and stop codons of the ORF are indicated by black dots. The internal ribosome entry site (IRES) is contained within 5′ NTR. (b) Membrane topology and major functions of the HCV polyprotein cleavage products (Bartenschlager et al., 2013).

The HCV particle (virion) is enveloped by a membrane, with glycoproteins E1 and E2 displayed on the surface. It has been reported to be 40 - 80 nm in diameter, pleomorphic, lacking obvious symmetry or surface features and containing an electron dense core (Figure II.3 a) (Bradley et al., 1985; Catanese et al., 2013b; Gastaminza et al., 2010; He et al., 1987; Merz et al., 2011). HCV particles have been observed to have low buoyant density, much lower than expected for a protein-envelope virion particle. The lower buoyance is due to their association with serum lipoproteins, including apolipoprotein A-I (apoA-I), apoB-48, apoB-100, apoC-I and apoE (Catanese et al., 2013b; Diaz et al., 2006; Felmlee et al., 2010; Kono et al., 2003; Lindenbach, 2013; Merz et al., 2011; Thomssen et al., 1992). This association between virus particles and serum lipoproteins indicates that HCV particles exist as hybrid 'lipoviral particles' (LVPs). The benefit of this association can prevent the virus from antibody neutralization due to the shielding of the envelope protein epitope. Liver cells take

up lipoprotein via receptor-mediated endocytosis, which facilitate virus entry into hepatocytes (Ross-Thriepland et al., 2013). However, the overall architecture of LVPs remains unclear. Currently, there are two models proposed for the structure of LVPs: the transiently interacting model and the hybrid particle model (Figure II.3 b, c).
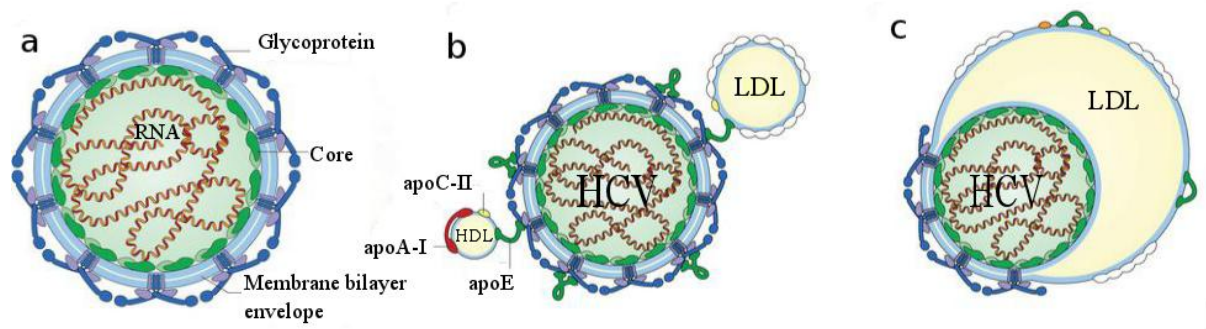
Figure II.3: Models of hepatitis C virus (HCV) lipoviral particle (LVP) structure. (a) Model of a hepatitis C virus (HCV) particle, showing the bilayer membrane envelope, E1–E2 surface glycoproteins, and the nucleocapsid containing core protein and viral RNA. (b) The transiently interacting model for lipoviral particle (LVP) structure. (c) The hybrid particle model for LVP structure, illustrating an HCV particle sharing an envelope with a low-density lipoprotein (LDL) particle, adapted from Lindenbach and Rice, et al. (2013).

HCV entry is the first step of the HCV infection process. It requires the cooperation of numerous factors in a temporally and spatially ordered manner (Figure II.4). A series of host membrane proteins are involved in this entry process; these are HSPG (heparan sulphate proteoglycans), LDLR (low-density-lipoprotein receptor), SRB1 (scavenger receptor class B type I), CD81, EGFR (epidermal growth factor receptor), CLDN1, OCLN and so on. Lipoprotein-associated HCV particles attach to the surface of hepatocytes via low affinity interaction between virion-associated apoE and its host acceptors, LDLR and glycosaminoglycans (GAGs) that are presented on HSPGs (Agnello et al., 1999; Albecka et al., 2012; Germi et al., 2002; Monazahian et al., 1999). The high density lipoprotein receptor SRB1 also serves as a HCV receptor; its cholesterol transfer activity might help virus particle dissociate from their LVPs, and its interaction with HCV E2 glycoprotein might change the conformation of E2 for exposing the CD81-binding determinants (Bankwitz et al., 2010; Dao Thi et al., 2012; Scarselli et al., 2002; Zahid et al., 2013). The critical step in HCV entry is the lateral movement of CD81-bound HCV particles to tight junctions and the CD81-HCV complex interacts with claudin-1 (CLDN1). Several signal transduction pathways involving EGFR and downstream RAS GTPase signaling, as well as RHO GTPases regulate this cell surface transportation (Brazzoli et al., 2008; Diao et al., 2012; Farquhar et al., 2012;

Lupberger et al., 2011; Zona et al., 2013). The interaction of HCV-CD81 with CLDN1 triggers the clathrin-mediated endocytosis and translocates HCV particles into hepatocytes (Farquhar et al., 2012). Although occludin (OCLN) is also reported to be essential for HCV entry, its specific role is not clear at the moment (Ploss et al., 2009; Sourisseau et al., 2013).
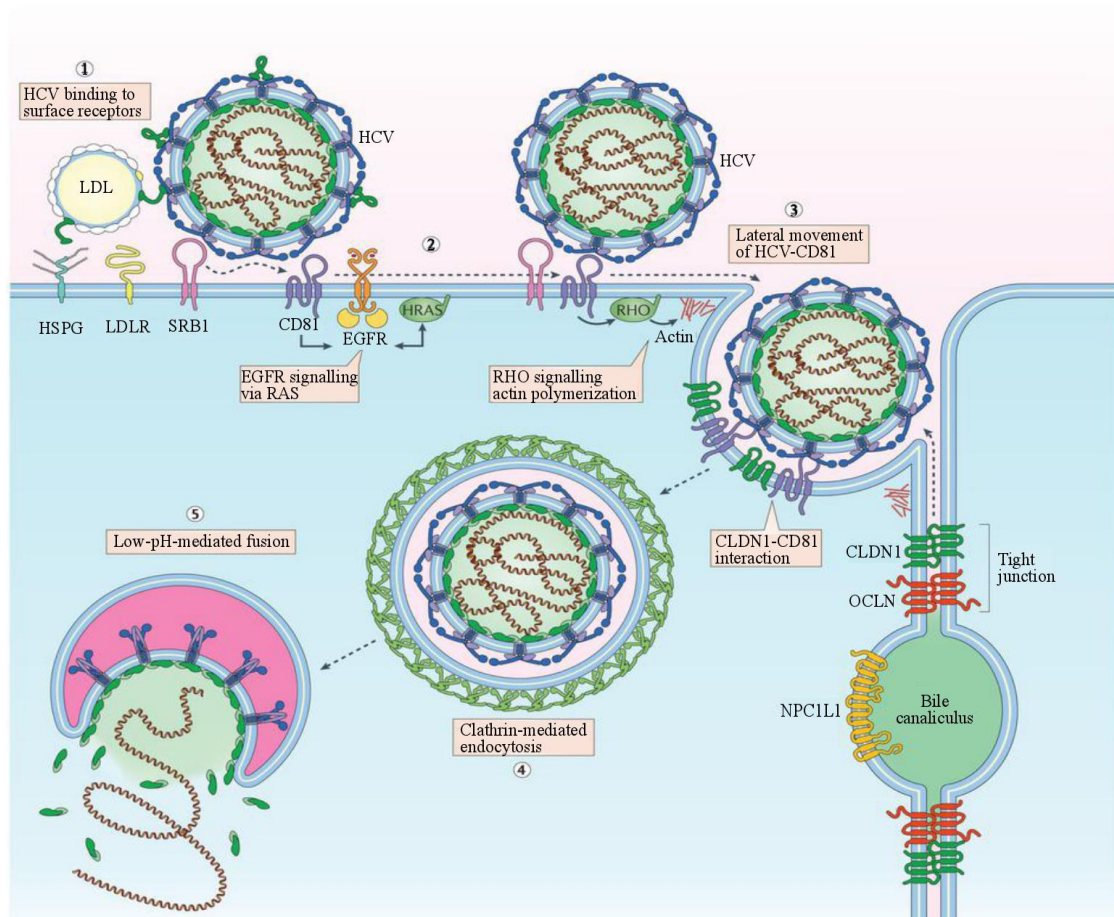
Figure II.4: An integrated model of HCV entry. Step 1: low affinity interaction of LVPs with HSPGs and LDLR and SRB1. The interaction with SRB1 might change the conformation of HCV E2 glycoprotein, exposing the binding site to CD81; Step 2: the interaction of HCV E2 with CD81 activates signal transduction through EGFR, RAS GTPases and RHO GTPases; Step 3: Signal transduction promotes HCV-CD81 movement to the tight junction; Step 4: Clathrin mediated endocytosis; Step 5: Low pH induces HCV fusion in the endosomal compartment. NPC1L1, Niemann - Pick C1-like 1; OCLN, occludin (Lindenbach and Rice, 2013).

II.1.3 HCV envelope proteins

HCV envelope glycoprotein 1 (E1) and envelope glycoprotein 2 (E2) are type I transmembrane proteins with a highly glycosylated ectodomain and a C-terminal hydrophobic anchor. HCV E1 and E2 glycoproteins form a noncovalent heterodimer through their C-terminal transmembrane domains, which is essential for interaction with receptors and neutralizing mAbs (Drummer and Poumbourios, 2004; Lavie et al., 2007). E2 is composed of an ectodomain (ETD, residues 384 - 661) and a C-terminal transmembrane domain (residues 715 - 746), between which the stem region (residues 662 - 714) is located. E2 has 9 to11 N-linked glycosylation sites and 18 conserved cysteine residues. Despite these conservations, HCV E2 contains 3 hypervariable regions: hypervariable region 1 (HVR1, residues 384 - 411), hypervariable region 2 (HVR2, residues 460 - 485) and the intergenotypic variable region (HVR3 or igVR, residues 570 - 580) (Figure II.5). The deletion of the hypervariable regions has been shown to have no effect on the overall folding of the protein and the binding with CD1 and mAbs (Kato et al., 1992; Krey et al., 2010; McCaffrey et al., 2007; McCaffrey et al., 2011; Weiner et al., 1991).
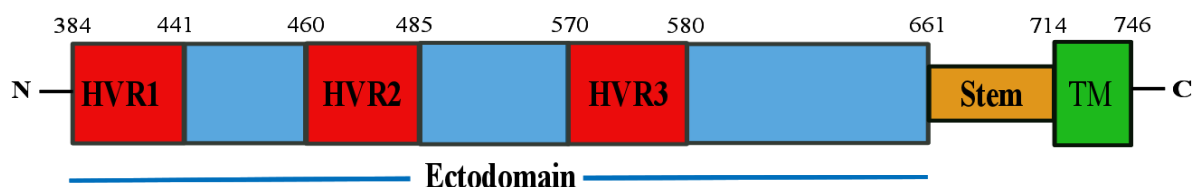
Figure II.5: Schematic representation of HCV E2 glycoprotein domain organization. HVR1, HVR2 and HVR3 are indicated as red frames. The TM domain is colored in green; the stem region is colored in orange. The borders of each region are indicated as amino acids numbers.

II.1.4 HCV human receptor SRB1

Scavenger receptor class B type I (SRB1) is a 509 amino acids glycoprotein and belongs to the CD36 superfamily. It is composed of a larger extracellular domain, which is highly glycosylated and contains disulfide bonds, and two transmembrane domains at both N- and C- termini (Figure II.6) (Calvo and Vega, 1993; Krieger, 2001; Zeisel et al., 2007). SRB1 is primarily expressed in liver and non-placental steroidogenic tissues. Its normal function is to

bind high-density lipoproteins (HDLs), very-low-density lipoproteins (VLDLs) and oxidized forms of low-density lipoproteins (LDLs), and mediates the cholesterol uptake from these lipoproteins (Acton et al., 1996). SRB1 was also identified as a co-receptor of HCV particles based on the evidence of its interaction with recombinant HCV E2 on HepG2 cells in the absence of CD81, and promotes the cell-to-cell transmission of HCV virus (Bartosch et al., 2003; Catanese et al., 2013a; Scarselli et al., 2002). The function of SRB1 during HCV entry has been divided into 3 steps. First, low affinity attachment of LVPs via lipoproteins occurs. Second, lipid transfer activity of SRB1 dissociates HCV particle and lipoproteins. Lastly, the direct interaction of SRB1 and E2 yields the determinants for CD81 binding (Lindenbach and Rice, 2013; Zahid et al., 2013). Previous reports show recombinant soluble E2 devoid of HVR1 exhibits decreased interaction with SRB1, and the entry of HCVpp lacking HVR1 no longer inhibited by SRB1-specific antiserum (Bankwitz et al., 2010; Scarselli et al., 2002). These results indicate the E2 potential interaction position for SRB1 might locate on the N-terminal HVR1 region.
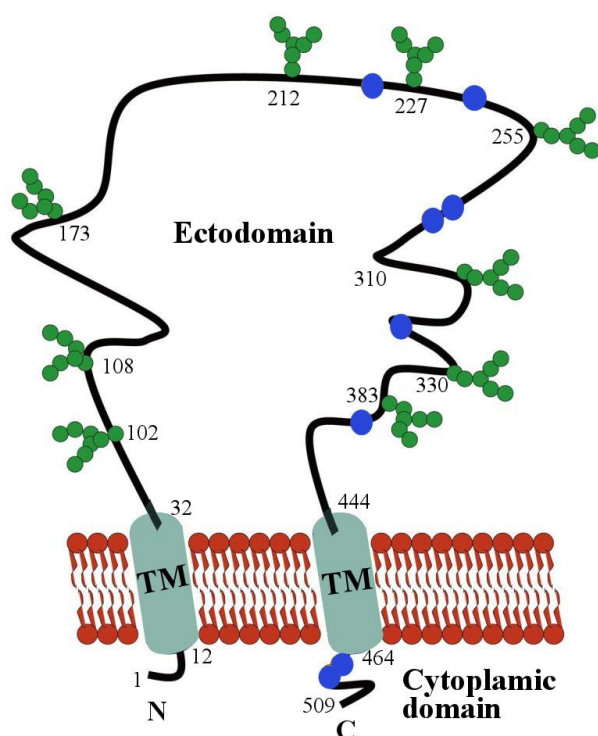
Figure II.6: Schematic representation of the scavenger receptor class B type I protein. Structural elements include a large ectodomain, two transmembrane domains, and short intracellular N-terminal and C-terminal domains. Glycosylation sites are colored in green, cysteine residues are colored in blue, and the transmembrane domains are colored in cyan. Numbers are according to the amino acids numbering of SRB1. The Figure was modified from Hoekstra et al. (Hoekstra et al., 2010).

II.2 Aim of the work

HCV E2 and SRB1 interaction is the first step in the protein-receptor interaction during HCV entry. The structures of SRB1 and HCV E2 were still unknown when this project was initiated. Furthermore, the interaction mechanism between SRB1 and E2 is still unsolved. The exploration of such knowledge will push forward our understanding of HCV entry process, and will assist in developing HCV vaccines and new entry inhibitors.

II.3 Results

II.3.1 HCV sE2 expression, purification and crystallization

In the HCV envelope glycoprotein E2-ectodomain (sE2), N-linked glycans constitute nearly 50% of the molecular weight. Folding of HCV E2 is a slow process and requires the assistance of ER chaperon machinery such as calnexin (Choukhi et al., 1998). Because of the high glycosylation and intermolecular disulfide bonds, a homogenous E2 protein cannot be obtained by normal expression strategies for crystallization. A high proportion of the expressed E2 forms aggregates through intermolecular disulfide bonds, which is common and has been proposed to be a physiologically relevant pathway of the HCV lifecycle (Dubuisson, 2000; Michalak et al., 1997; Whidby et al., 2009). In order to get soluble, homogenous soluble E2 (sE2, residues 384-661) protein, three expression strategies (transient expression, baculovirus-insect cell expression and stable cell line) were performed simultaneously.

HCV sE2 was fused with a His$_6$ tag and cloned into a pcDNA 3.1-tpa eukaryotic expression vector and transfected into a HEK293 6E cell line with PEI (polyethylenimine). Together with transfection, 5μM kifunensine (α-mannosidases inhibitor) can be added to the culture to decrease the complexity of glycosylation. In the presence of kifunensine, protein glycosylation was decreased to the high mannose type. The high mannose type glycosylated protein can be truncated efficiently by endoglycosidases with only a single N-acetylglucosamine (GlcNAc) left on each N-glycosylation site. Two glycosylation patterns (high mannose type and fully glycosylation pattern) of HCV sE2 have been expressed depending on the addition of kifunensine. HCV sE2 was purified from the supernatant of HEK 293 6E suspension culture with Ni-NTA affinity chromatography. The purification products were analyzed with SDS-PAGE. Due to glycosylation, the protein band appears

larger than its theoretical size. The theoretical size of sE2 is 31 kDa, but it appears to be around 55 kDa on SDS-PAGE (Figure II.7 a). The fully glycosylated sE2 appears smeary on SDS-PAGE, this is probably because of heterogeneous glycosylation. To verify sE2 expression, a Western blot experiment using anti-His antibodies was performed. After staining, a band corresponding to that on SDS-APGE was visualized on the Western blot membrane (Figure II.7 b). High mannose type sE2 was also sent for N-terminal sequencing and this confirmed the first 13 amino acids of sE2. The high mannose sE2 protein has been deglycosylated with EndoHf and yielded two bands between 35 - 40 kDa on SDS-PAGE (Figure II.7 a). We speculate that these two bands are caused by heterogeneous glycosylation. The deglycosylated protein was then assessed by MALDI - TOF - mass spectrometry. The results revealed five partially used glycosylation sites among the 13 predicted ones. These partially occupied glycosylation sites are at Asn532 (90%), Asn576 (96%) and Asn623 (30%) and Asn (64%). The deglycosylated protein was further purified by cation exchange chromatography and then applied onto a size exclusion column. However, the protein shows considerate aggregation on size exclusion chromatography (Figure II.8 a). To check if the aggregation is caused by deglycosylation, another size exclusion chromatography was performed with non-deglycosylated protein (high mannose sE2). Unfortunately, it shows similar behavior to the deglycosylated sample (Figure II.8 b).
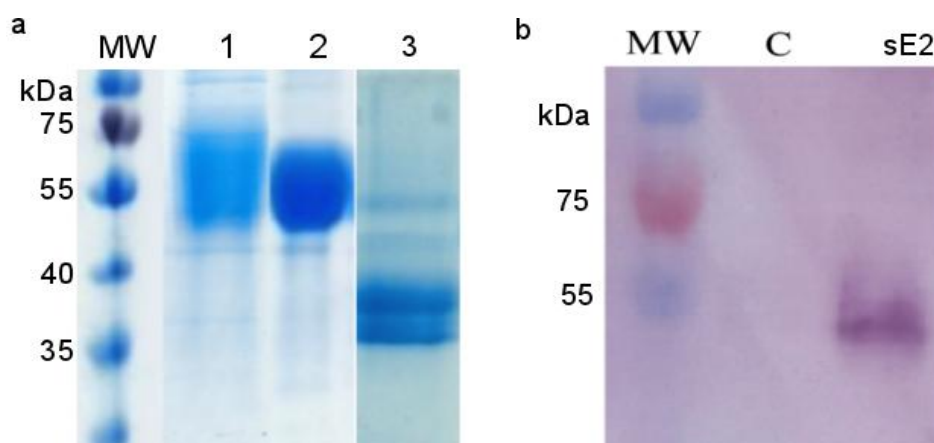
Figure II.7: Transiently expressed HCV sE2. (a) SDS-PAGE analysis of transiently expressed sE2 after affinity chromatography. 1: fully glycosylated sE2; 2: high mannose type sE2; 3: high mannose sE2 deglycosylated with EndoHf. (b) Western Blot detection of HCV E2 using anti-His antibody. C: control sample of non-transfected HEK 293 6E cell supernatant; sE2: high mannose type of sE2.
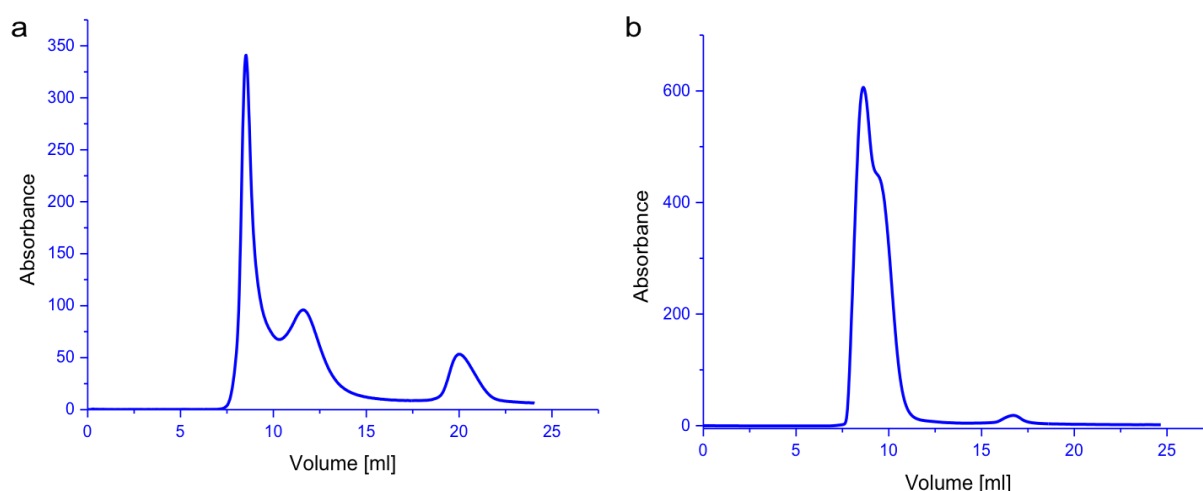
Figure II.8: Size exclusion chromatography of sE2 shows considerate aggregation. (a) Size exclusion chromatography of deglycosylated sE2; (b) Size exclusion chromatography of non-deglycosylated sE2.

To investigate the reason for aggregation, non-reducing SDS-PAGE experiments were performed. Non-reducing SDS-PAGE showed three major bands, with estimated sizes corresponding to monomer, dimer and aggregates, while the reduced SDS-PAGE showed only one band corresponding to a monomer (Figure II.9 a). Fraser J et al. reported two reduced cysteines (Cys597 and Cys620) in E2 protein are essential for virus entry (Fraser et al., 2011). The free cysteines might form intermolecular disulfide bonds and cause protein aggregation. To verify this, a Blue Native PAGE experiment was performed. Prior to electrophoresis, 0.1-1.5 mM DTT was added to protein samples and incubated at 4 °C for 3 hours. However, the present of different concentrations of DTT did not solve the aggregation problem as the smeary and broad bands still indicated high aggregation (Figure II.9 b). In addition, two reduced cysteines (Cys597 and Cys620) were mutated into serines to avoid the intermolecular disulfide bonds (data not shown). Unfortunately, this did not help either. Taken together these results, the transient expressed HCV sE2 form aggregation, and the aggregation problem can not be solved by adding reducing agent or mutation, thus homogeneous sE2 can not be obtained by this method and the crystallization experiment could not be performed.
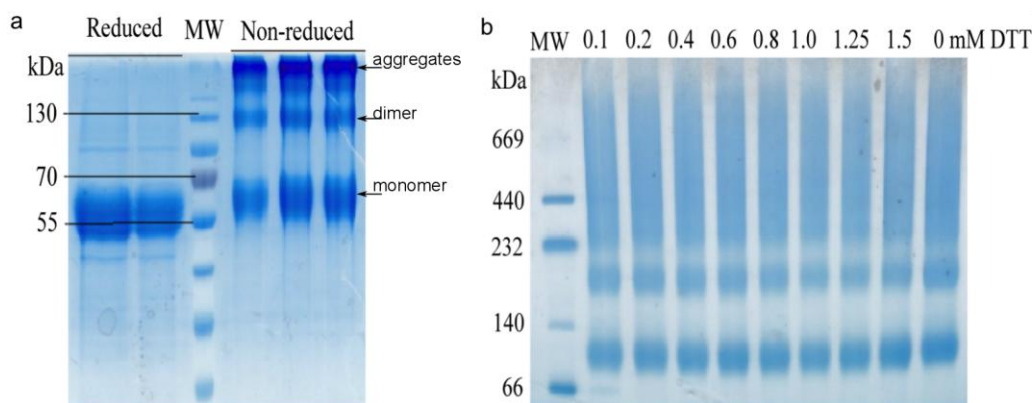
Figure II.9: Non-reduced SDS-PAGE and Blue Native PAGE analysis of E2 aggregation. (a) SDS-PAGE of reduced sE2 and non-reduced E2 ETD; MW: high molecular weight prestained protein marker (Thermo scientific). (b) Blue Native PAGE of E2 ETD in the presence of serial concentrations of DTT (0 - 1.5 mM); MW: high molecular weight Native PAGE marker.

At the same time, a stable cell line expressing sE2 was constructed with the help of Dr. Joop van den Heuvel. The stable cell line was derived from a CHO Lec3.2.8.1 cell line. CHO Lec3.2.8.1 synthesizes glycoproteins with truncated carbohydrates, which have enabled the crystallization of many glycoproteins. Its protein products are glycosylated with high mannose type N-linked glycans (Wilke et al., 2010).

His$_6$-tagged sE2 was secreted by a stable cell line. The successful construction of this stable cell line was confirmed by Western blot using anti-His antibodies. sE2 has been purified with Ni-NTA affinity chromatography and followed by cation exchange chromatography. The purified protein was later applied onto a Superdex 200 10/30 column. Fortunately, sE2 behaves as a monomeric protein on size exclusion chromatography (~ 50 kDa, Figure II.10 a). The protein has been further confirmed by N-terminal sequencing. The peak fractions of size exclusion chromatography were collected and deglycosylated with EndoHf. A further size exclusion chromatography was performed with Superdex 75 10/30 column (higher resolution) to remove EndoHf and glycans. Deglycosylated sE2 stays as a monomer and appears as a homogenous protein (~ 35 kDa) on SDS-PAGE. Deglycosylated sE2 was concentrated and used for crystallization experiments. Unfortunately, crystallization setups did not yield any crystals with commercial screens, even after several rounds.
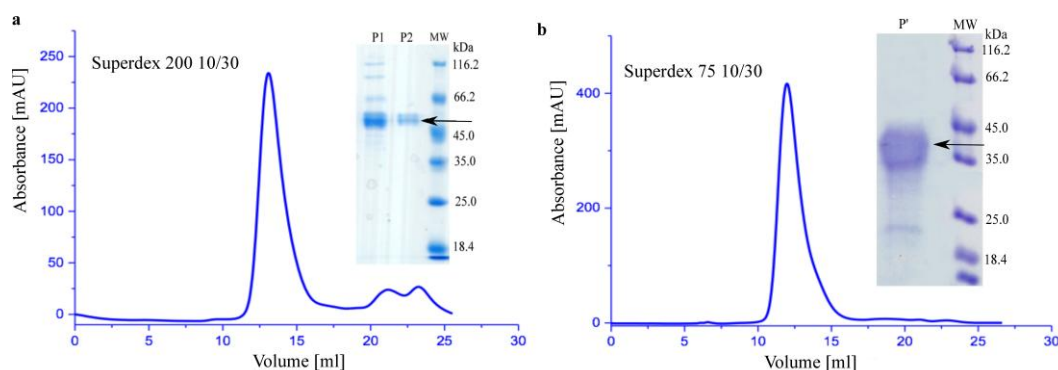
Figure II.10: Size exclusion chromatography of HCV sE2 expressed by a CHO lec 3.8.2.1 stable cell line. (a) Size exclusion chromatography with Superdex 200 10/30 column of sE2 and SDS-PAGE result, P1: protein sample after affinity purification; P2: protein with peak fraction of size exclusion chromatography. (b) Size exclusion chromatography with Superdex 75 10/30 of deglycosylated sE2 and SDS-PAGE result. P': peak fraction of size exclusion chromatography; MW: unstained low molecular weight protein marker. Black arrows indicate bands of sE2 protein.

A baculovirus-insect cell system was also used for His_6-tagged sE2 expression. A recombinant baculovirus carrying the sE2 gene was successfully constructed. However, no expression was detected in the baculovirus infected insect cells (data not shown).

II.3.2 SRB1 expression and characterization

Scavenger receptor class B type I (SRB1) has been identified as one of the HCV receptors (Scarselli et al., 2002). It is a transmembrane protein that belongs to the CD 36 family. The extracellular domain is predicted to contain 9 N-glycosylation sites and three intramolecular disulfide bonds (Calvo and Vega, 1993; Krieger, 2001; Zeisel et al., 2007). Here the extracellular domain of SRB1 (residues 33 - 433) was cloned and expressed by two different strategies: transient expression and baculovirus-insect cell expression systems.

In the initial transient expression tests with HEK 293 6E cells, SRB1was fused with a His_6-tag and cloned into the pFlpBtM II vector. But almost no expression was detected from the transfected cell supernatant. A pYD11 vector that included a C-terminal Fc coding sequence was selected to improve the protein yield. The SRB1 extracellular domain was inserted into the pYD11 vector, and transiently transfected into the HEK 293 6E cell line. Kifunensine was added to the culture to reduce the glycosylation complexity. Fc fusion SRB1 (SRB1-Fc) was purified with a protein A affinity column. SRB1-Fc has a theoretical size of 75 kDa, but due

to glycosylation, it shows a band around 110 kDa on SDS-PAGE (Figure II.11 a). HRP-conjugated goat anti-human IgG (Fc specific) antibodies were used for detecting SRB1-Fc expression in the Western blot experiment. The corresponding bands were visualized on a PVDF membrane after 10s of exposure time with a LAS 3000 imager system (Figure II.11 b). SRB1-Fc expression was further verified by N-terminal sequencing in house. Compare to low expression of His$_6$-tagged SRB1, Fc fusion SRB1 has a very high yield of about 7 mg per liter culture. However, it has the same aggregation problem as HCV sE2 protein. Due to high aggregation, the Fc fragment couldn't be cleaved off by PreScission protease.
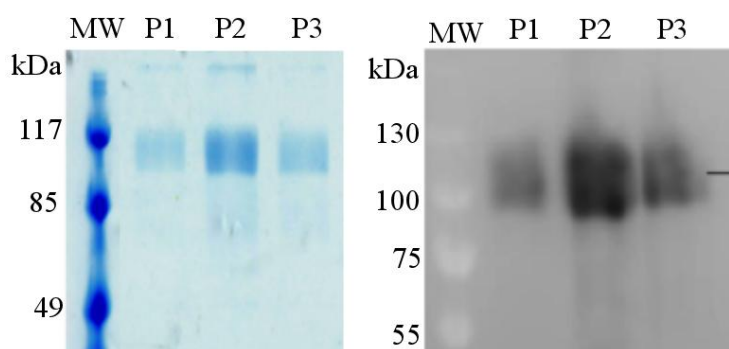
Figure II.11: SDS-PAGE and Western Blot SRB1-Fc. (a) SDS-PAGE of SRB1-Fc after affinity purification; (b) Western Blot of SRB1-Fc using HRP- conjugated goat anti-human IgG (Fc specific) antibodies.

Another attempt was to express and crystallize SRB1 single domains. Limited proteolysis was performed to define domain borders with purified SRB1-Fc. Different concentrations of pepsin, trypsin and chymotrypsin were incubated together with SRB1-Fc at RT for 30 min. Limited proteolysis yielded 3 cleavage sites inside SRB1-Fc. One common cleavage site is the border between SRB1 and Fc fragment. The other two cleavage sites divided the SRB1 extracellular domain further into 3 domains: domain 1, residues 1 - 306; domain 2, residues 307 - 352; domain 3, residues 353 - 433 (Figure II.12). Domain 1 has also been cloned into the pYD11 vector and transiently expressed similar to the full-length SRB1 ectodomain. Unfortunately, the purified product of domain 1 still shows strong aggregation (Data not shown).
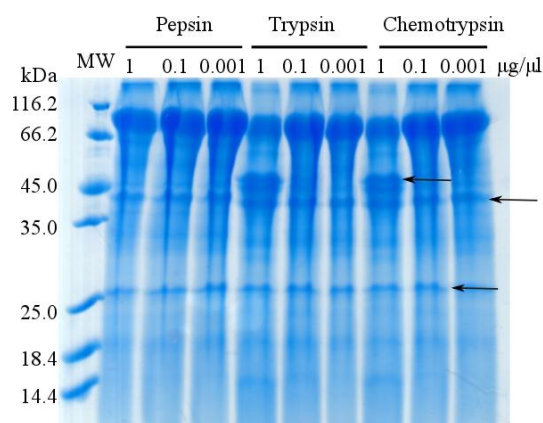
Figure II.12: Limited proteolysis of transiently expressed SRB1-Fc. Limited proteolysis of SRB1-Fc has been performed with 3 different concentrations (1 μg/μl, 0.1 μg/μl and 0.001 μg/μl) of pepsin, trypsin and chymotrypsin at RT for 30 min. Black arrows indicate the protein band cut for N-terminal sequencing.

Another expression strategy for SRB1 uses the baculovirus-insect cell system. Human SRB1 (residues 33 - 433) was cloned into the pFlpBtM II donor vector. Recombinant bacmids were generated using the Tn7 transposition method in bacmids of the EMBacY (MBY) system, which includes a YFP-gene as a maker for infection monitoring (Meyer et al., 2013; Trowitzsch et al., 2010). Sf21 was infected with P1 (passage one) baculovirus. Three days after infection, the cell culture was harvested and cells were lysed for protein purification. N-terminally strep-tagged SRB1 (strep-SRB1) ectodomain has been expressed and purified with a strep-tactin affinity column (Figure II.13 a). After affinity column, the protein was further purified with an anion exchange column. The purification product was visualized as a single band with a size of around 55 kDa on SDS-PAGE. Mass-spectroscopy and N-terminal sequencing confirmed strep-SRB1 expression and purification. Strep-SRB1 behaves as a monomer in size exclusion chromatography. However, the expression and purification experiments only succeeded once. During later infection trials with the same recombinant baculovirus, only a very small amount of protein was obtained. Due to insufficient amount of material, crystallization of SRB1 protein was not pursued.
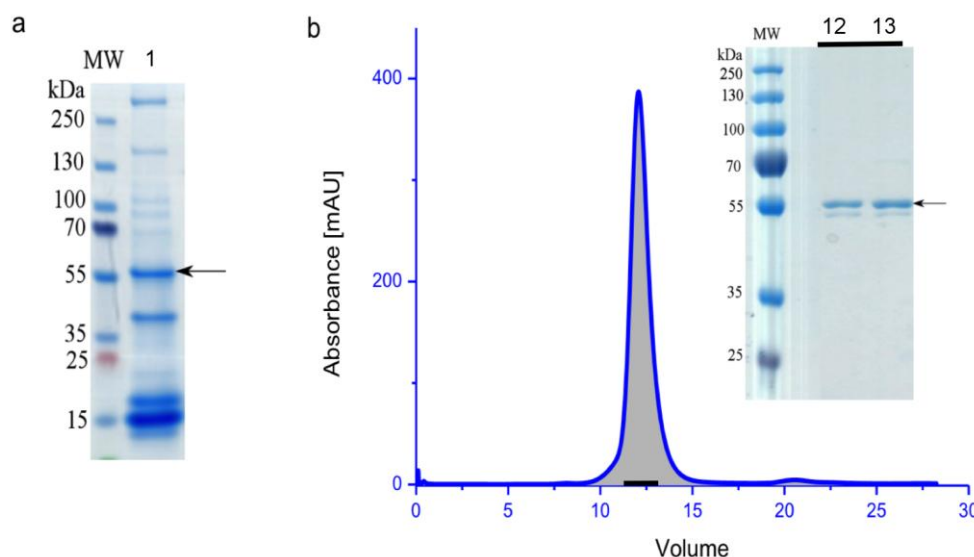
Figure II.13: Purification of strep-SRB1 from insect cells. (a) SDS-PAGE analysis of affinity chromatography purification products, 1: protein sample purified with strep-tactin affinity column. (b) Size exclusion of strep-SRB1 with Superdex 200 10/30 column, and the SDS-PAGE analysis of peak fractions. 12,13: peak fractions from size exclusion chromatography; MW: prestained high molecular weight marker.

II.4 Discussion and outlook

II.4.1 HCV envelope protein E2

A stable cell line expressing HCV E2 ectodomain (384-656, sE2) has been successfully constructed. The protein purified from this expression system shows no aggregation, which indicates the properly fold protein. However, the purified and deglycosylated protein didn't yield any crystals after many crystallization trials. Crystallization is a thermodynamic process, which is mainly driven by the entropy gain from the released water molecules upon crystallization. If the protein is too flexible, this gain is compensated by an entropy loss from the restricted movement of protein side chains and loops. As the result, the protein does not crystallize. If the flexible loops are removed in the first place, the protein might become crystallizable (Derewenda, 2004; Derewenda and Vekilov, 2006; Hayward and Lee, 2002). Hence, more attention should be put on the protein constructs. While we were struggling with getting sE2 crystallized, two structures of HCV E2 core domain structures have been solved. One E2 core structure is using engineered E2 core constructs to avoid flexible hypervariable loops (HVR1 and HVR2), and is solved in cooperation of neutralizing antibody AR3C [PDB code: 4MWF (Kong et al., 2013)]. Another E2 core structure is using shorter constructs

(residues 456 -656), and the Fab fragment of antibody 2A12 was used for co-crystallization [PDB code: 4NX3 (Khan et al., 2014)]. The difference between our sE2 protein and E2 core is that the latter lacks the N-terminal hypervariable regions 1 and 2 (HVR1 and HVR2, about 80 amino acids). HVR1 and HVR2 were proved to be flexible by limited proteolysis and hydrogen deuterium exchange (Khan et al., 2014), methods that we didn't perform with our sE2 protein. In our HCV sE2 constructs, both flexible HVR1 and HVR2 are included, which is probably the reason why we didn't get crystals. Further, they used antibodies or antibody fragments to facilitate crystallization. Antibodies and antibody fragments have been utilized a lot as the crystallization "chaperones" to aid the structural determination of otherwise 'uncrystallizable' or 'undruggable' target proteins. Another advantage for antibody mediated crystallization is that it can provide model-based phasing information (Griffin and Lawson, 2011), since at the time we started there is no molecular replacement model available. Because we couldn't get hands on such antibodies, this method has not been put into practice.

The expression strategy that Khan et al. and Kong et al. used is transient expression with adherent monolayer HEK cell lines (HEK293T-GnTI- and HEK293T) in supplement with kifunensine. Similar transient expression strategy has been used in our experiments; the only difference is that we used the suspension cell culture. Although the suspension cell culture could produce protein in large amount, the proteins always showed high aggregation, which is probably due to misfolding. Since HCV E2 folding is a slow process, which involves complicated posttranslational modifications such as glycosylation and forming disulfide bridges. It seems that the lower cell density and better nutrient accessibility in a monolayer cell culture is beneficial for the production of properly folded E2 protein.

Both HCV E2 core structures are consistent with each other except that Kong et al.'s structure includes a N-terminal engineered fragment and three extra mapped disulfide bridges. The E2 core domain has a globular fold, consisting of mostly β-strands and random coils. Two pairs of β-strands are antiparallelly arranged in the middle, flanked by short helices, loops and β-strands (Figure II.14). Despite seven pairs of disulfide bridges, which cross-link loops, strands and helix, the E2 core domain consists of nearly 62% disordered structure. Further, six glycosylation sites are all located in loops and the glycans are highly flexible. Before the HCV E2 core structure has been solved, it had been speculated to be similar to the class II fusion protein of flavivirus. All class II fusion proteins have an elongated structure, consisting of mainly β-strands, and exist as homo- or heterodimers with membrane fusion

interface in neutral pH. Upon receptor binding or pH decrease in the environment, the protein undergoes self-rearrangement into stable trimers, exposing the fusion peptide and resulting in viral and host membrane fusion (El Omari et al., 2013; Li et al., 2013; Vaney and Rey, 2011; White et al., 2008). Due to the lack of extended organization and fusion peptide, HCV E2 protein seems not belonging to the class II fusion proteins (Khan et al., 2014; Kong et al., 2013).
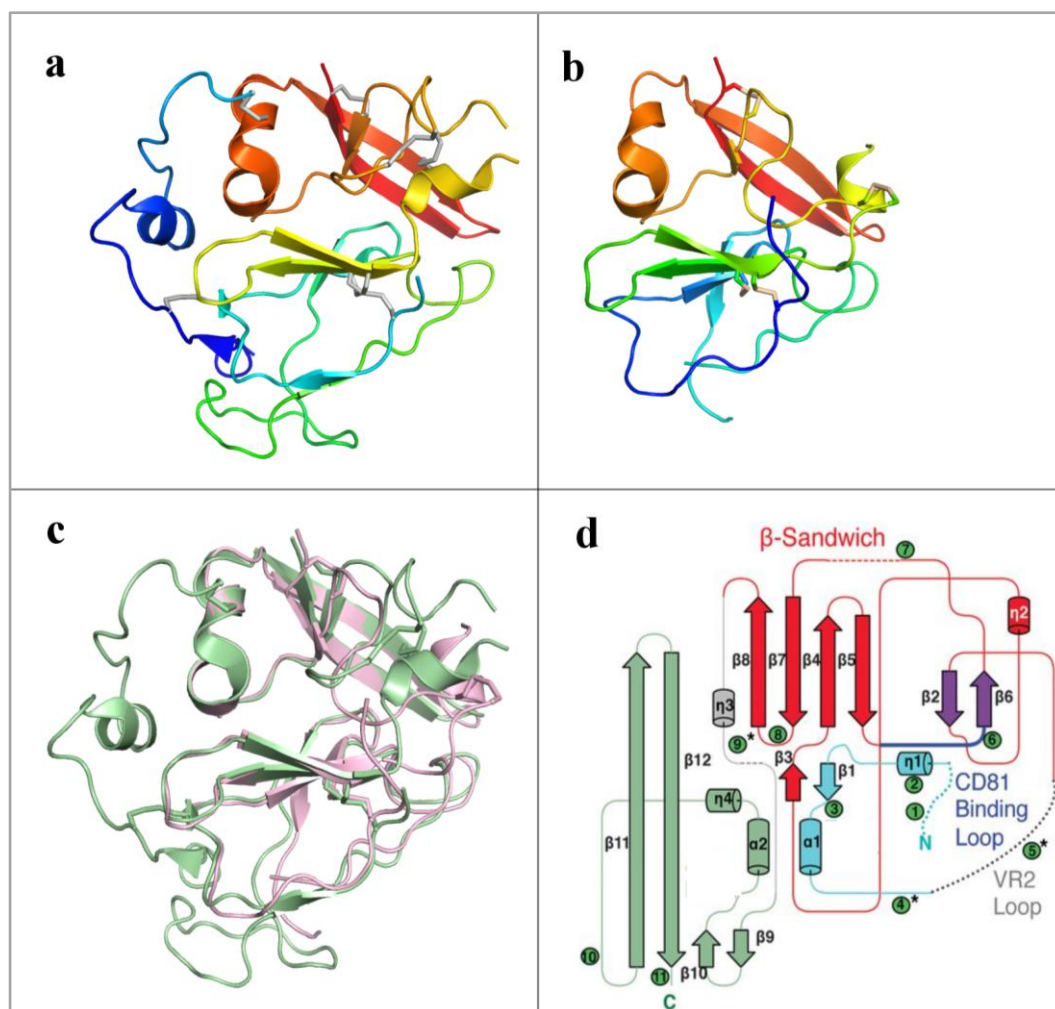
Figure II.14: Cartoon view of HCV E2 core domain structures from 4MWF and 4NX3, and superposition. (a) Cartoon view of E2 core domain from 4MWF; (b) E2 core domain from 4NX3, both (b) and (c) are colored from N terminus (blue) to C terminus (red), disulfide bridges are colored in gray; (c) Superposition of two E2 core structures, 4MWF is colored in light green, 4NX3 is colored in light pink; (d) Topology of HCV E2 core structure, adopted from Kong et al. (2013).

The determination of the HCV E2 core structure is a big progress in the HCV research. For further understanding of the HCV entry process, HCV E1 structure and E1E2 complex structure determination has been put on the agenda. The mapping of E2 broadly neutralizing epitopes provides new opportunities for HCV vaccine development(Khan et al., 2014). The structural information of HCV E2 core will finally lead to new drug design.

II.4.2 Scavenger receptor class B type I

SRB1 is expressed only in limited tissues, such as liver and steroidogenic tissues, where it mediates cholesteryl ester uptake from high-density lipoprotein (HDL) (Silver, 2004). With the assistance of Fc fragment of IgG, SRB1 was highly expressed in transiently transfected HEK293 6E cell culture. Similar to HCV E2, HCV human receptor SRB1 is highly glycosylated and the transiently expressed protein had a high proportion of aggregates. In the baculovirus-insect cell expression system, homogenous SRB1 ectodomain has been obtained for the first time in small-scale culture. However, due to the unknown reason, almost no protein has been obtained in the larger scale infection trials, even with reconstructed recombinant baculovirus.

Since monolayer cells could produce properly folded protein, using monolayer cell culture for transient transfection and expression might solve the aggregation problem. The transient expression strategy is more efficient in protein production than in insect cells. The later exploration of SRB1 structure should consider monolayer cell culture for the transient transfection and expression. Because Fc fragment could assist protein folding and improve protein solubility, the Fc fusion SRB1 constructs could be kept in future. It is also necessary to optimize the ectodomain constructs, for example, knock out flexible loops and crystallize separate domains. Another strategy for co-crystallization with HCV E2 core protein could also be tried.

III Exploration of the legumain autoactivation process

III.1 Introduction

Legumain (EC=3.4.22.34) belongs to the cysteine proteases clan CD, family C13 (Chen et al., 1998). Legumain is structurally related to caspases and is considered to be an asparaginyl endopeptidase because it specifically cleaves after asparagine at pH 5.8 (Chen et al., 2000; Liu et al., 2012). It has also been reported to cleave after aspartate residues at lower pH (pH 3.0) (Chen et al., 1997; Halfon et al., 1998). Legumain is well conserved in plants, invertebrate parasites and mammals. Mammals have been found to show ubiquitous legumain expression with the highest levels in kidney, placenta, spleen, liver and testis (Yamane et al., 2002). Mammalian legumain is considered to function in immunity and signal transduction. It has also been proposed to be essential in processing microbial antigens for class II histocompatibility complex (MHC II) presentation (Manoury et al., 1998) and in processing of cathepsin B, L and proMMP-2 into their active forms (Chen et al., 2001; Maehr et al., 2005; Shirahama-Noda et al., 2003; Smith et al., 2012). Legumain has also been expressed at a high level in bone marrow. It blocks the later stages of osteoclast formation and regulates bone resorption (Choi et al., 1999). Recently, legumain has been found to be up-regulated in a number of diseases, such as atherosclerosis, stroke, and cancer. Overexpression of legumain was demonstrated to promote cell migration, invasion, and angiogenesis (Edgington et al., 2013; Liu et al., 2003; Luo et al., 2006; Murthy et al., 2005; Ohno et al., 2013). Within the tumor microenvironment, cells that express high levels of legumain are tumor-associated macrophages (TAMs), but not tumor cells themselves (Luo et al., 2006). Based on this property, legumain has been selected as an anti-tumor target for anti-tumor vaccines in future cancer treatment (Lewen et al., 2008; Luo et al., 2006).

The Chinese hamster legumain gene encodes a pre-protein of 438 amino acids (Xu et al., 2011). It shares 83.1% sequence identity with human legumain. After cleavage of N-terminal signal peptide (residues 1 - 20) an inactive prolegumain (residues 21 - 438, ~56 kDa) is produced. In the cytoplasmic and nucleic environment, prolegumain could perform proteolytic maturation into its active form asparagine endopeptidase (AEP) by pH shift. 10 residues at the N-terminus was further cleaved off, an activation peptide (AP, residues 297 - 337) was further cleaved and released, and a less active legumain ACP (asparaginyl carboxypeptidase, ~ 50 kDa) was produced. In significantly lower pH (lysosome and endo/lysosome), highly active AEP (residues 31-291, ~35kDa) was produced through release

of the C-terminal legumain stabilization and activity modulation (LSAM, residues 337 - 436). In an extracellular environment, prolegumain could be loaded onto αVβ3 receptors, where the first proteolytic activation to ACP takes place. A second proteolytic activation could be performed with integrin stabilized ACP into integrin stabilized AEP. The intracellular and extracellular legumain could participate in different signal pathways, antigen processing and other protein processing (Dall and Brandstetter, 2013). The multiple functions and multiple locations of legumain are shown in Figure III.1.

Interestingly, it has been discovered recently that the proteolytic activation of AEP is a reversible process. Partially activated AEP could be reversed back to its proenzyme form when the protein buffer pH shifted from acidic to neutral (Zhao et al., 2014). Legumain activity can be inhibited by cystatin C, E, M and ovocystatin (Alvarez-Fernandez et al., 1999), which inhibit through competitive binding of asparagine peptide to the active site (Zhao et al., 2014).
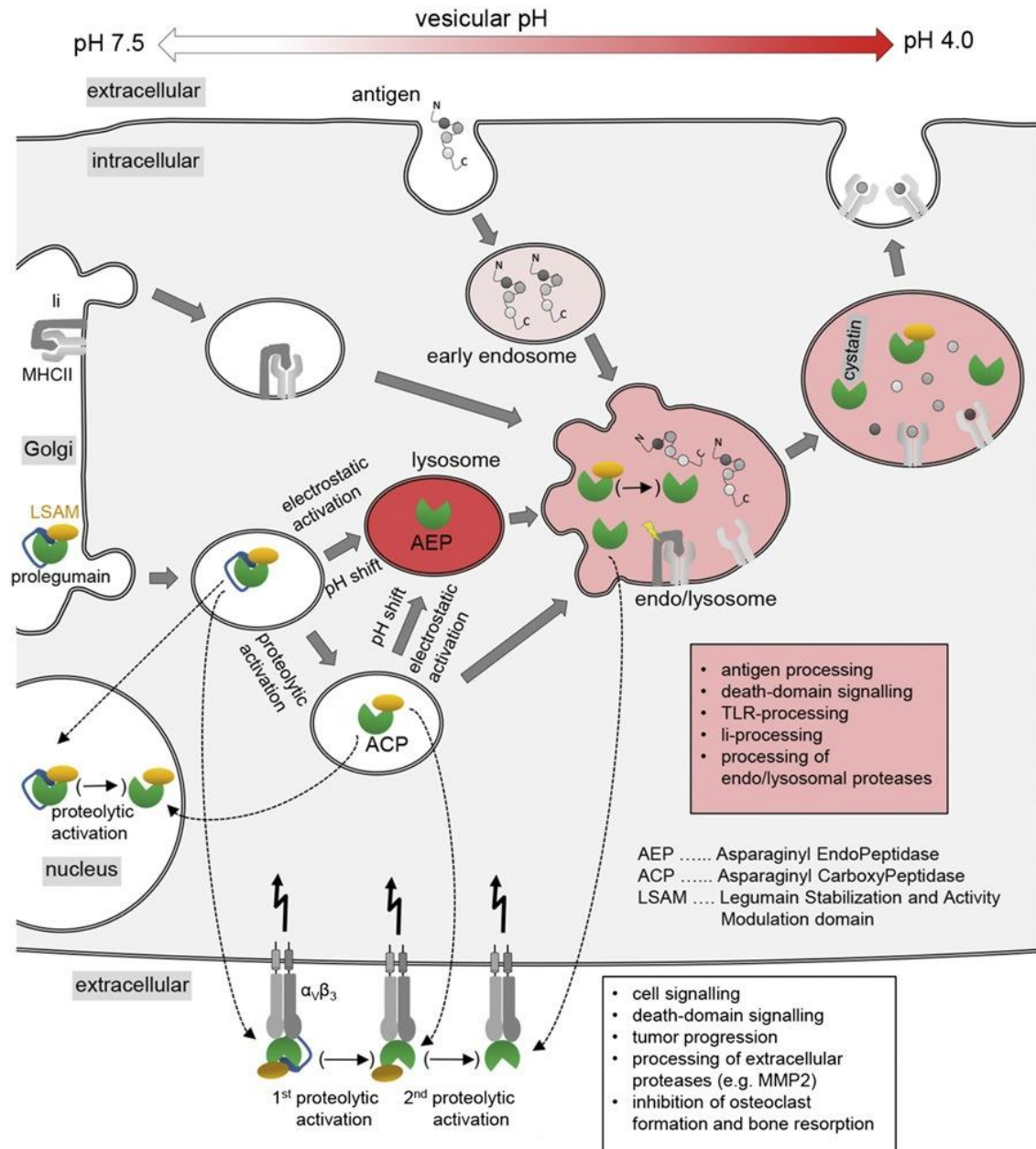
Figure III.1: Multiple functions and multiple locations of legumain. After its posttranslational modifications within the Golgi apparatus and vesiculation, prolegumain can be either released into cytosol or loaded onto $\alpha_V\beta_3$ integrin receptors. In an intracellular environment, transfer/fusion of prolegumain with lysosomal compartments will result in electrostatic release, and subsequent cleavage into AEP. In an extracellular environment, integrin stabilized AEP could be achieved by steps of proteolytic activation, as is necessary for its extracellular tasks (white box). Figure adapted from Dall and Brandstetter (2013).

III.2 Results

III.2.1 Legumain structure determination

During our experiment of constructing stable cell lines expressing HCV sE2 with the CHO lec 3.8.2.1 cell line, we found it constitutively secreting large amount of prolegumain into the culture media. The secreted prolegumain has a high binding ability for Ni-NTA, protein A and strep-tactin resins. Prolegumain was purified from CHO lec 3.8.2.1 cell supernatant with a series of purification steps (Ni-NTA affinity, anion exchange and size exclusion chromatographies). The size of prolegumain after anion exchange chromatography appeared to be around 57 kDa. After size exclusion chromatography with a pH 5.4 buffer, a smaller band appeared which indicated a portion of prolegumain undergoes self-cleavage. Prolegumain was deglycosylated in a pH 5.4 buffer, in which deglycosylation and auto-cleavage happened simultaneously. After deglycosylation, another size exclusion chromatography was used to remove impurities. Finally, a protein of 37 kDa was yielded in situ, which is the active legumain (asparaginyl endopeptidase, AEP) (Figure III.2 a). Glycosylated HCV-sE2 high mannose protein has a size around 55 kDa, and the deglycosylated protein has a size about 35 kDa. In the beginning, the prolegumain and legumain were mistaken for HCV sE2 protein, but later the N-terminal protein sequencing rectified this. The 37 kDa legumain was used for crystallization experiments, and resulted in rod-shaped crystals (Figure III.2 b). Crystals were obtained from three conditions: Condition 1 contained 0.03M $MgCl_2$, 0.03M $CaCl_2$, 10% PEG 20 000, 20% PEG550 MME and 0.1 M MOPS/HEPES buffer, pH 7.5; condition 2 had similar ingredients with condition 1 but had a buffer pH of 8.0; condition 3 contained 0.1 M NaCl, 23% PEG 3350 and 0.1 M MOPS buffer, pH 6.5. Crystals were named legumain$_{pH7.5}$, legumain$_{pH8.0}$ and legumain$_{pH6.5}$ respectively according to the pH of crystallization.

Complete X-ray diffraction data were collected at wavelengths of 1.033191 Å at beamline P11 of PETRAPETRA III Hamburg, Germany. Data were processed with the XDS package (Kabsch, 2010). Three different space groups $P2_12_12_1$, $P3_221$ and $P6_1$ were obtained from legumain$_{pH7.5}$, legumain$_{pH8.0}$ and legumain$_{pH6.5}$ crystals respectively. Molecular replacement (MR) was carried out with Phaser_MR against human legumain (PDB code: 4AWA) (Dall and Brandstetter, 2013). Prior to molecular replacement, non-conserved amino acid side chains of search models were pruned with Chainsaw. The log likelihood gain were 1524,

1579 and 3572 for datasets legumain$_{pH7.5}$, legumain$_{pH8.0}$ and legumain$_{pH6.5}$, respectively. However, additional unmodeled electron densities were observed for legumain$_{pH7.5}$ and legumain$_{pH8.0}$ during manual building with Coot. These two datasets were again using MR against human prolegumain (PDB code: 4FGU). Surprisingly, the log likelihood gain values were more than doubled, and the electron density maps were properly interpreted. These two crystal forms (legumain$_{pH7.5}$ and legumain$_{pH8.0}$) were shown to contain the different protein prolegumain, which is the zymogen of active legumain. Matthew's coefficients indicate each crystal form consists of only one molecule per asymmetric unit. The solvent contents of the three crystal forms are between 31% and 37%. Legumain$_{pH7.5}$ crystals have a Wilson B-factor of 30.3 Å whereas legumain$_{pH8.0}$ crystals have a higher value of 40.4 Å. Leguamin$_{pH6.5}$ crystals have a Wilson B-factor which is lower than 20 Å. Legumain$_{pH7.5}$ and legumain$_{pH8.0}$ have similar highest resolution limit of 2.4 Å and 2.3 Å, respectively. The crystals of Leguamin$_{pH6.5}$ diffracted beyond 1.5 Å resolution. The initial models for the three crystal forms were improved by iterative steps of manual building in Coot and refinement in Phenix.refine. Refinement was stopped after converging values of R_{work} and R_{free} were reached. The values of R_{free} / R_{work} are within the good range for datasets resolution and applied refinement strategies (Tickle et al., 1998). More than 97 % of the residues lie in the most favored region of the Ramachandran diagram in all three structural models. Data collection and refinement statistics for the three crystal forms are summarized in Table III.1.
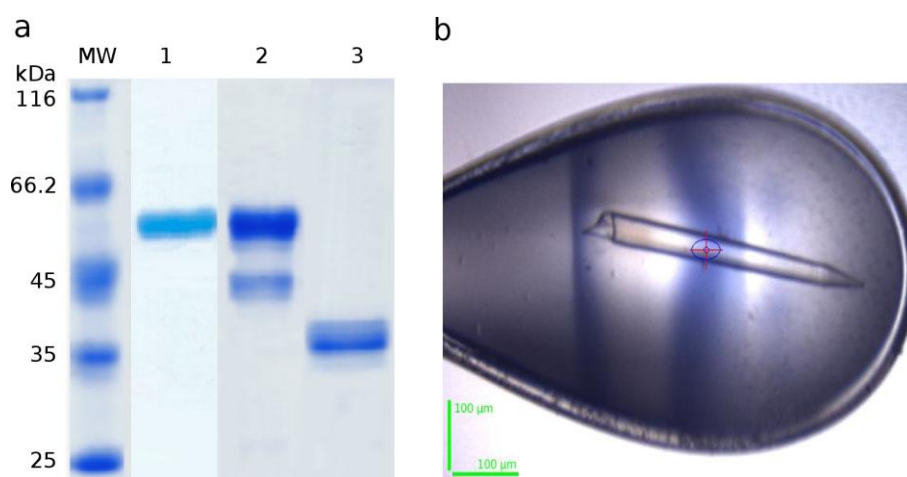
Figure III.2: Purification and crystallization of legumain. (a) SDS-PAGE analysis of purified prolegumain and its deglycosylation and auto-activation. 1: Prolegumain after anion exchange chromatography; 2: size exclusion chromatography in pH 5.4 buffer indicates partial auto-cleavage; 3: Deglycosylated and auto-activated legumain; MW: low molecular weight protein marker. (b) Crystal of legumain at pH6.5.

Table III.1: Data collection and refinement statistics of legumain crystals

Data statistics	legumain$_{pH7.5}$	legumain$_{pH8.0}$	legumain$_{pH6.5}$
Space group	P 2_1 2_1 2_1	P 3_2 2 1	P 6_1
Cell constants a, b, c (Å)	38.3, 60.3, 154.0	50.8, 50.8, 253.6	43.6, 43.6, 209.6
α, β, γ (°)	90, 90, 90	90, 90, 120	90, 90, 120
Mol. / a.s.u.	1	1	1
Wilson B-factor (Å^2)	30.3	40.4	19.2
Solvent content (%)	31.3	35.4	36.5
Resolution range (Å)	20.0-2.4 (2.49-2.40)	20.0-2.3 (2.38-2.30)	20.0-1.5 (1.55-1.50)
I/σ(I)	11.3 (2.4)	16.5 (2.8)	13.0 (2.18)
Completeness overall (%)	100.0 (99.9)	100.0 (100.0)	100.0 (100.0)
Unique Reflections	14574 (1401)	17828 (1717)	35874 (3570)
Multiplicity	11.8 (12.0)	17.7 (17.8)	6.6 (5.0)
Rmerge (%)	22.6 (78.4)	2.6 (23.8)	3.3 (36.3)
Wavelength (Å)	1.033191	1.033191	1.581827
Refinement Statistics			
R$_{work}$ (%)	20.7 (27.5)	20.2 (27.9)	17.6 (28.3)
R$_{free}$ (%)	25.0 (33.0)	24.0 (28.0)	20.6 (30.0)
av. B-factor of protein atoms(Å^2)	35.1	49.4	32.7
r.m.s.d. bonds (Å) / angles (°)	0.003 / 0.78	0.003 / 0.79	0.012 /1.48
Ramachandran favored (%)	98.0	97.0	98.0
Ramachandran allowed (%)	1.47	2.48	1.62

(): Values in () correspond to the highest resolution shell.

III.2.2 Prolegumain structures

Although we intended to set up crystallization screens with cleaved legumain, two crystal forms from legumain$_{pH\ 7.5}$ and legumain$_{pH\ 8.0}$ were interpreted to include the C-terminal cap, and the protein content is Chinese hamster prolegumain (PDB code: 4FGU). Similar to human prolegumain, the Chinese hamster prolegumain also possesses three domains: an AEP core domain (cyan, Ala26 - Met286), an activation peptide (AP, orange, Lys287 - Leu331) and a legumain stabilization and activity modulation domain (LSAM, slate, Val332-Ser431) (Figure III.3). The AEP core domain is composed of 6 well-aligned β-strands flanked by five major α-helices and loops. The LSAM domain is composed of five α-helices, which have a death domain like fold (Dall and Brandstetter, 2013), and two pairs of disulfide bridges for stabilizing the helices. The LSAM domain behaves like a cap covering the top of AEP, which prevents the substrate from accessing the active site. The AP fragment includes a N-terminal loop and a C-terminal helix. In prolegumain, AP binds to the active site in a substrate-like way and blocks access of true substrate to the active site (Dall and Brandstetter, 2013).
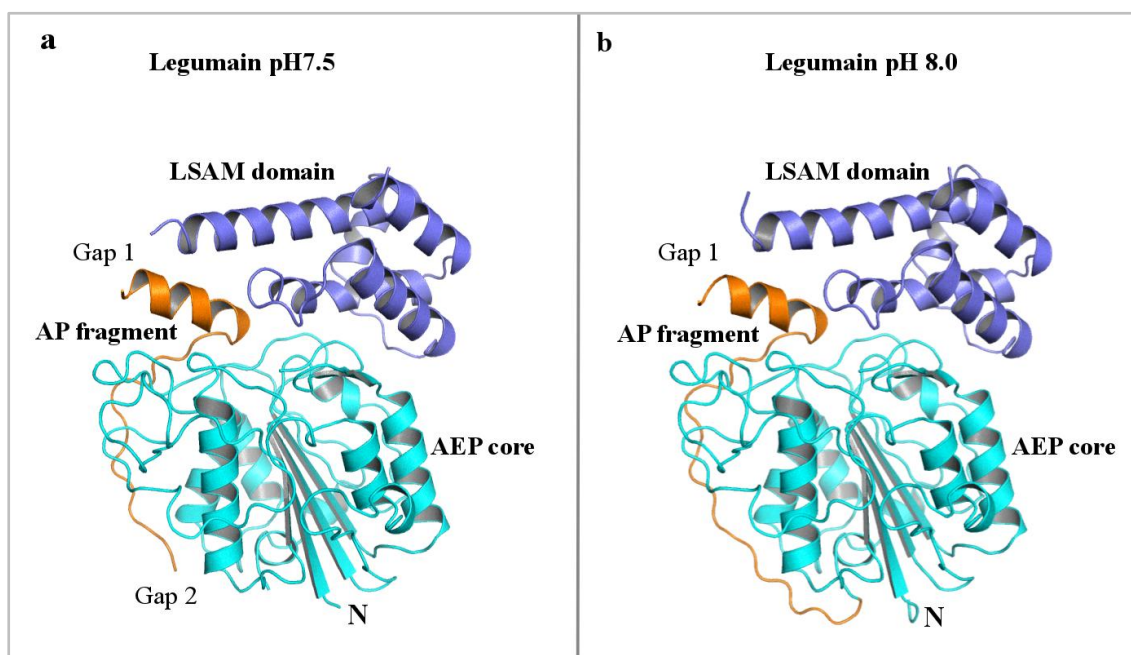
Figure III.3: Organization of legumain$_{pH7.5}$ and legumain$_{pH8.0}$. Structural domains are colored differently, LSAM domain: slate; AP fragment: orange; AEP core: cyan. The gaps in structural models are marked as Gap 1 and Gap 2.
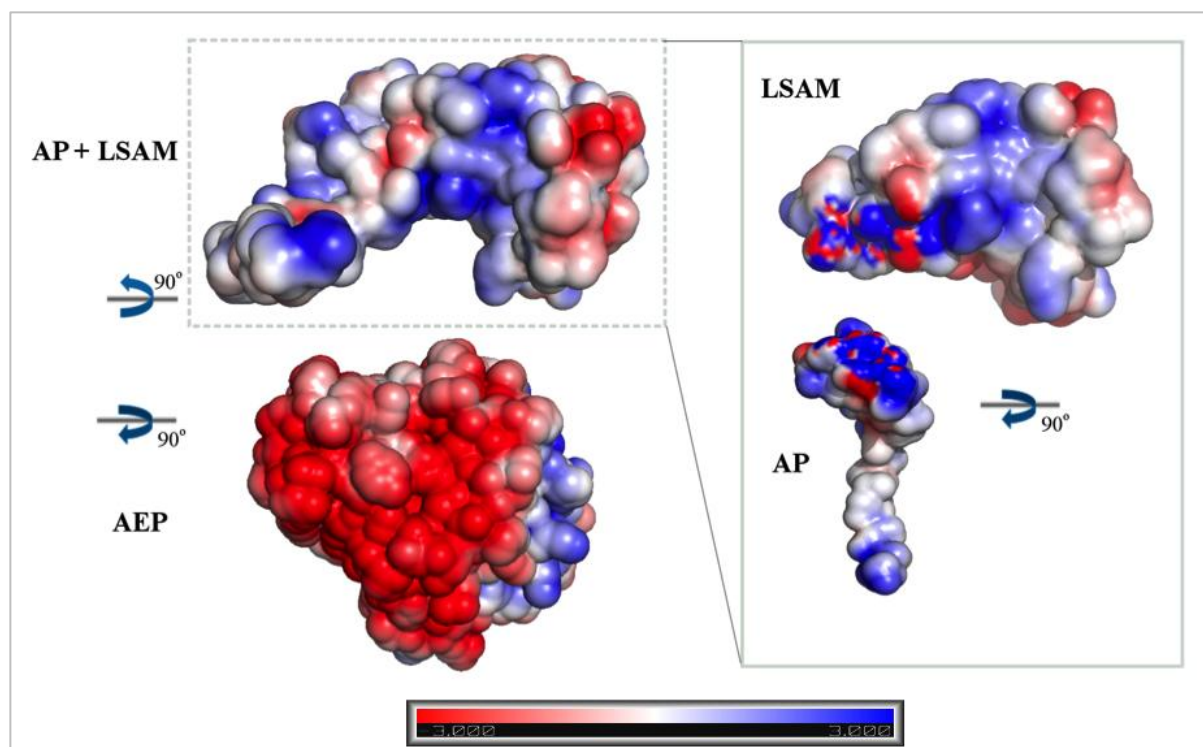
Figure III.4: Electrostatics of legumain domains. Blue represents positive potential, and red represents negative potential. The active site surface of the AEP core domain is highly negative charged and colored in red. The electrostatic surface potentials are mapped onto Van-der-Waals surfaces (red: negative, blue: positive, white: neutral).

Compared to the full-length human prolegumain structure, two gaps have been found in the Chinese hamster legumain$_{pH7.5}$ structure (Figure III.3 a). Gap 1 (Leu324 - Leu336) contains the β-cleavage site at Asn328 - Asp329 and Gap 2 (Met282 - Lys292) contains the putative α-cleavage site (Asp308). The gaps are also the domain borders between the AEP domain, AP fragment and LSAM domain. The difference of legumain$_{pH7.5}$ from legumain$_{pH8.0}$ is that Gap 2 is not present in the former. Based on the reverse autoproteolytic theory (Zhao et al., 2014), legumain$_{pH7.5}$ and legumain$_{pH8.0}$ are two intermediates before complete reversal back to full-length prolegumain. This could indicate that the reversible autoproteolytic process is performed in steps during pH shifts from acidic pH to basic pH. During an increase in pH, the three domains of legumain are attached together through electrostatic interactions, ligation of the AP fragment is followed by ligation of LSAM to finally produce the prolegumain. The LSAM domain and AP fragment interact with the AEP core domain through numerous hydrogen bonds and salt bridges. Electrostatic analysis shows legumain is a highly charged molecule; the surface potential charge of AEP core domain at the active site is extremely

negative. Under neutral and basic conditions, the surface potential charge of LSAM domain and AP fragment could complement the negative charge on the active site surface and keep the protein intact. Under acidic conditions, the active site surface is protonated and the LSAM domain and AP fragment would dissociate (Figure III.4). Without the ligation between domains in legumain$_{pH7.5}$, three domains could still hold together in the inactive form as prolegumain. The LSAM domain and AP fragment behave like a cap that covers the active surface of the AEP core and prevents the substrate from accessing the active site in neutral and basic pH.

III.2.3 Legumain$_{pH6.5}$ structure

A high-resolution dataset with the lowest resolution limit of 1.5 Å was obtained for leguamin$_{pH6.5}$, which allowed us to view the active site in detail. The crystal structure of Chinese hamster active legumain$_{pH6.5}$ showed similar architecture to the human AEP structure (Dall and Brandstetter, 2013). There are six β-strands (β1, β2, β3, β6, β7, β8) in a parallel arrangement in the center, surrounding by five major α-helices (α1, α2, α4, α7, α8). The residues that are critical for the AEP activity are located on α1, a loop between β3 and β4, and a loop between α5 and β6. Legumain is also N-glycosylated where all three glycosylation sites are located in the AEP core. Two glycosylation sites NAG2 and NAG3 (linked to Asn172 and Asn 277 respectively) clearly showed a 100% occupancy in electron density map during refinement, while NAG1 (linked to Asn 96) showed a lower occupancy (Figure III.5).

III.2.4 A Succnimide and a S-oxy cysteine in the active site

The catalytic site sits in a pocket on the surface of the AEP core domain and is covered by the LSAM domain in prolegumain. Asn42, His148 and C189 are reported as key residues for legumain activity (Dall and Brandstetter, 2013). Correspondingly, Asn44, His150 and Cys191 are key residues for mouse AEP activity. In addition, Arg46, Glu189, Ser217, Ser218 and Asp233 are also reported to be essential for mouse AEP activity. In Chinese hamster legumain, Asn47, His153 and Cys194 are critical residues for legumain activity. In our 1.5Å resolution structure, we found Asp152, which is located next to His153, was cyclized to a succinimide (SNN) residue, and Cys194 (cysteine) was oxidized to S-oxy cysteine (CSX)

(Figure III.6 a and b). The detailed structure of the active site will be discussed in the next chapter.
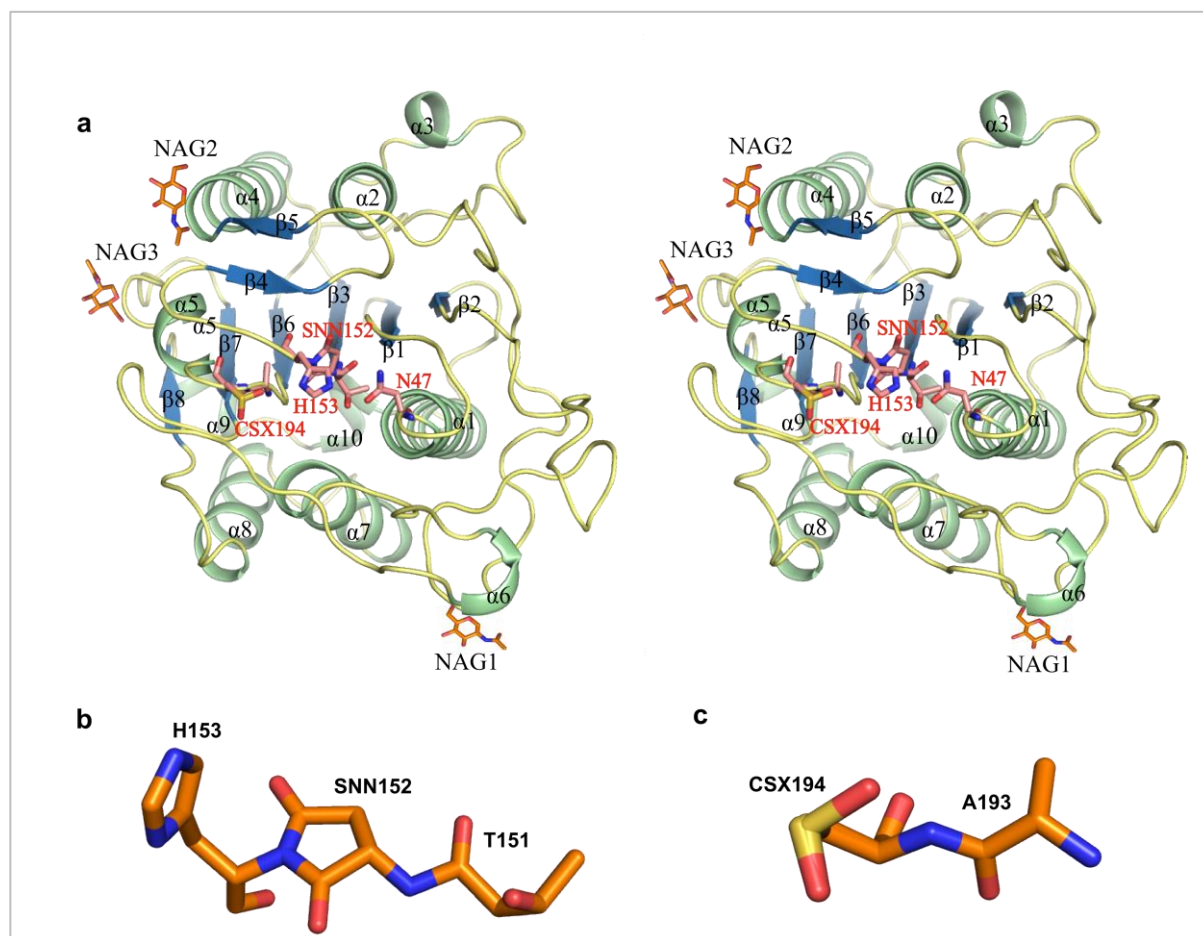
Figure III.5: Stereo view of legumain$_{pH6.5}$ structure and special residues. (a) Stereo view of legumain$_{pH6.5}$ α-helices and β-strands are numbered; key residues for the AEP activity are shown as sticks. α-Helices are colored palegreen, β-strands are colored skyblue and loops are colored light orange; (b) close view of residue succinimide (SNN152); (c) close view of S-oxy cysteine (CSX194).

III.3 Discussion and outlook

During purification of HCV sE2 from CHO stable cell line supernatant, a protein yielded the same band (~55 kDa) as HCV sE2 on SDS-PAGE was co-purified together with sE2 with affinity chromatography. In the acidic deglycosylation buffer (pH 5.4), the protein decreases its size to 35 kDa, which is similar to size of deglycosylated sE2 protein (~ 33 kDa). This protein was initially mistaken as HCV sE2, but later it was proved to be Chinese hamster legumain with N-terminal protein sequencing method. The 55 kDa protein is the zymogen of legumain and is called prolegumain. In deglycosylation buffer, prolegumain performs self-cleavage and deglycosylation, which removes the C-terminal fragment and N-glycans. After self-cleavage of C-terminal fragment, the protein turns into active asparaginyl endopeptidase legumain. Legumain was further purified with a size exclusion chromatography and gave out rod-shape crystals in crystallization screens. So to speak, the Legumain crystals are byproducts of HCV E2 project.

Two forms of prolegumain structures were obtained, both of which are similar except the α-cleavage site (KRK^{291}) and β- cleavage site ($N^{325}D^{326}$). The superposition of legumain$_{pH7.5}$ to human prolegumain (4AWA) and mouse prolegumain (4NOM) showed a slight inconsistency in the AEP domains (Figure III.7). The main differences are the LSAM domains, which shift slightly in three structures. The r.m.s. derivations between legumain$_{pH7.5}$ and 4AWA is 2.2 Å, and the value is 2.0 Å with 4NOM. One active legumain (AEP) structure was also solved with very high resolution at 1.5 Å. An aspartate residue next to His153 was cyclized into a succinimide. Checking the electron density maps of legumain active sites from structures published by Dall & Brandstetter (2013). (Figure III.6 b) and Zhao et al. (2014) (Figure III.6 c), we found that aspartate does not fit properly into the electron density where succinimide should fit. Succinimide was reported to be the intermediate from aspartyl and asparaginyl peptides during spontaneous protein degradation (Geiger and Clarke, 1987; Stephenson and Clarke, 1989). Since the cyclic succinimide was found in all legumain structures and seems to be stable in physiological conditions, the presence of succinimide here should not contribute to AEP degradation. It is assumed that the formation of succnimide is due to the highly negative charged and hydrophobic environment, which results in dehydration and cyclization of aspartate.
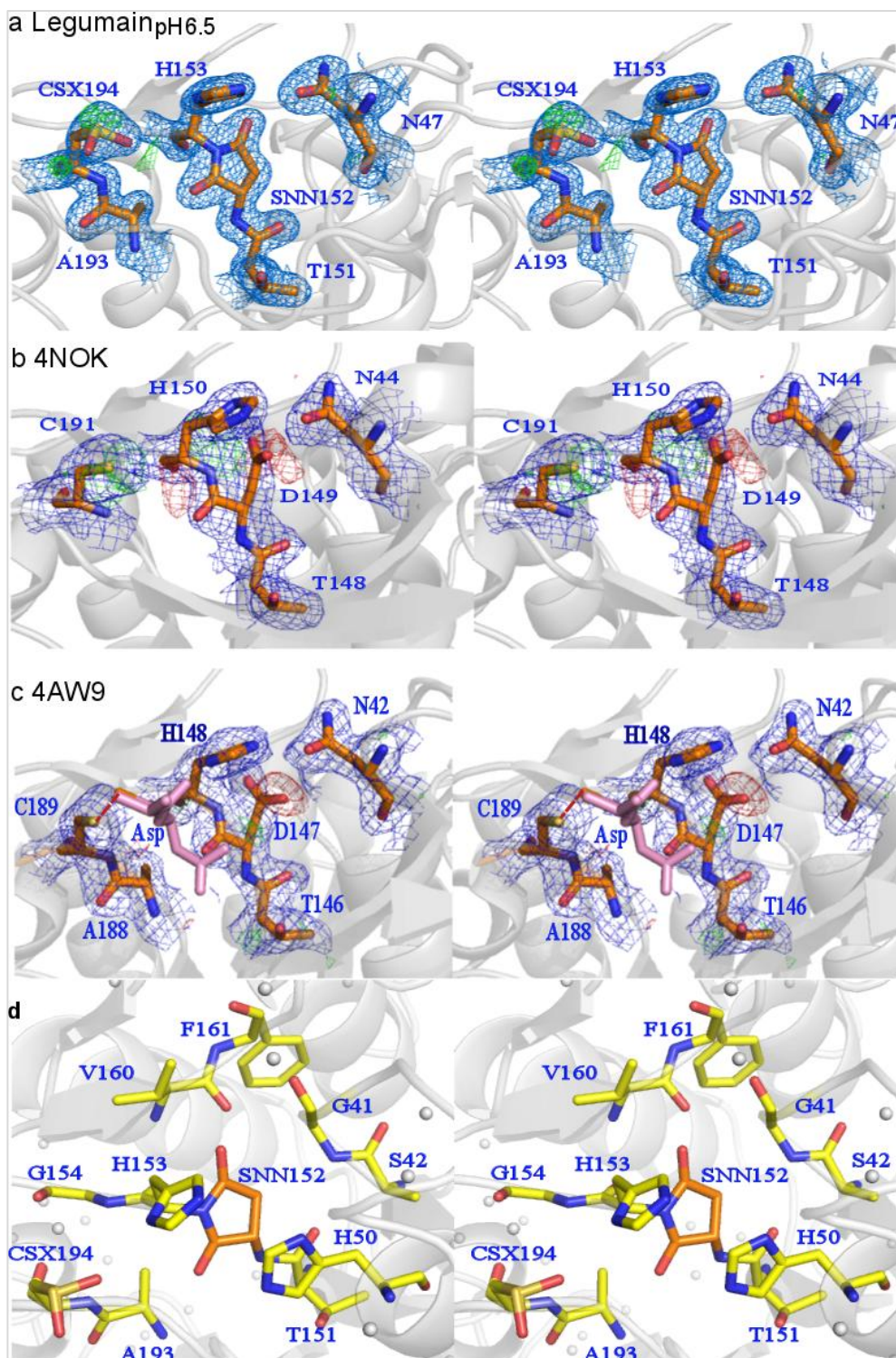
Figure III.6: Key residues in the active site of AEP from different organisms. Stereo views of legumain active sites, key residues are shown as sticks, the refined electron density map 2Fo-Fc (contoured at 1σ) is shown as a blue mesh, Fo-Fc map (contoured at 3σ) is shown as a green mesh. (a) Active site of legumain$_{pH6.5}$; (b) active site of mouse legumain 4NOK; (c) Human legumain 4AW9 with ligand binding, ligand is shown in pink sticks; (d) The hydrophobic environment of SNN152. SNN152 is shown in orange and blue sticks, the surrounding residues are shown in yellow and blue sticks.

81

The succinimide in the active site provides a more rigid conformation, which is necessary for enzyme specificity, than an aspartate residue.

Further, in mouse legumain structure (4NOK), cysteine residue has a positive triangle feature in the Fo-Fc map, which might be a S-oxy cysteine (CSX) with alternate conformations. In our structure, we built two alternate conformations of S-oxy cysteine (CSX) in the Cys194 position. Since the reaction from cysteine to CSX is irreversible, CSX is impossible to participate in the protease reaction, so the oxidization of cysteine might be an artifact of the crystallization process.

One thing that makes people confused is the α- and β- cleavage sites of prolegumain. In the legumain$_{pH7.5}$ structure model, both cleavage sites are open, which could be explained as electrostatic interaction on the domains' interface. In acidic pH the three domains expel each other, but in neutral and basic pH, these three domains attach together and go back to its original position and shield the active site. In legumain$_{pH8.0}$ structure, the α-cleavage site is still ligated. This could due to the incomplete cleavage before crystallization experiment. Another explanation from Zhao, *et al.* is the reversible autoproteolysis. During pH shift from acidic to neutral, three domains assemble together to form an inactive prolegumain before "healing" of peptide bonds. That of the α-cleavage site follows the healing of the β-cleavage site. As is commonly known, the peptide bond forming needs energy, but in the pure protein system, no energy is involved. Also the SDS-PAGE experiment couldn't confirm the split fragments is ligated by peptide bonds, which might be ligated by other covalent bonds. The reversible auto proteolysis theory needs more evidence to prove the "healing" of peptide bonds.

The prolegumain structures that have been solved in this work could still be the evidence for the auto maturation process from preprolegumain to active legumain. The auto maturation process could be summarized as in Figure III.7 b. The preprolegumain was translated on ribosome. N-terminal signal peptide (residues 1 - 20) was cleaved off by posttranslational modification, and an inactive prolegumain was translocated into lysosome or secreted into cytosol. In slightly acidic pH, the N-terminal was further cleaved by 10 residues (residues 21 - 29), and the C-terminal LSAM (residues 337 - 436) domain was self-cleaved off. In even lower pH or with longer incubation time, the AP fragment was further cleaved off (residues 297 - 337), yielding a highly active legumain AEP. The structural information of legumain

and its auto activation mechanism could be used for vaccine development and drug design in anti-cancer research.
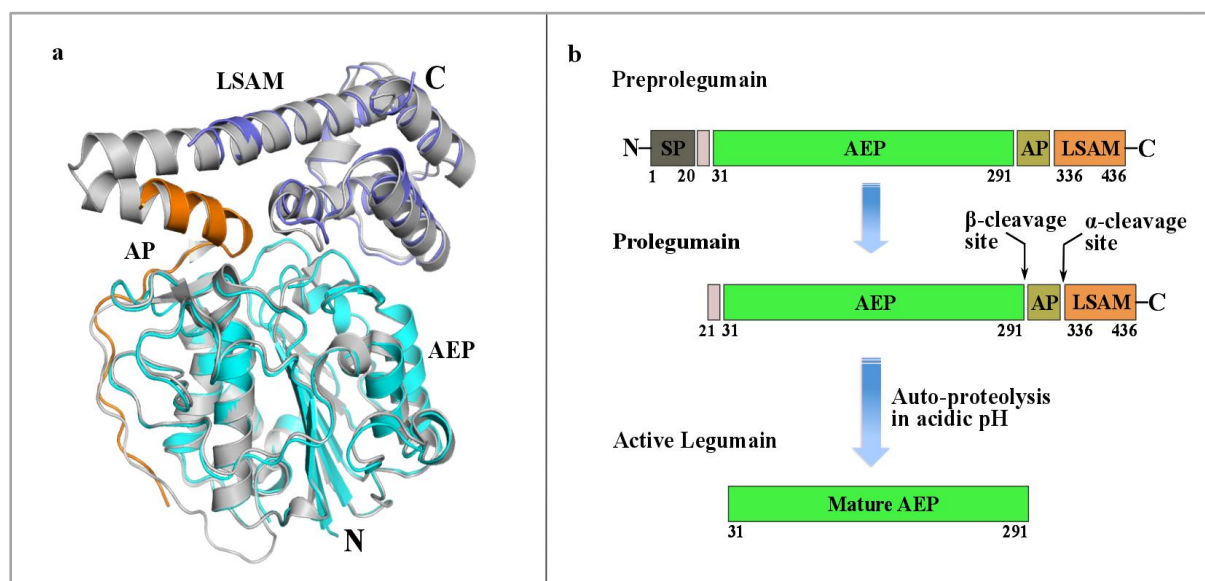
Figure III.7: Superposition of the mouse prolegumain (PDB code: 4NOK) and legumain$_{pH7.5}$ and illustration of auto proteolytic process. (a) Superposition of mouse prolegumain and legumain$_{pH7.5}$. Mouse prolegumain is colored in gray. Legumain$_{pH7.5}$ is colored in domains. (b) Illustration of maturation process of preprolegumain into legumain.

IV Materials and Methods

IV.1 General background knowledge and methods for crystallography

IV.1.1 Vapor diffusion

In order to obtain X-ray diffraction map for X-ray crystallography, a well ordered protein crystal is needed. Crystal formation is principally an energetically favored process: a crystalline structure was formed from a fluid, Entropy is gained due to the loss of the ordered hydration shell surrounding the protein molecules upon nucleation and further adsorption of protein molecules to the seed, and the free energy in the local environment decreases. The crystallization process mechanically consists of two processes: nucleation and crystal growth. For both, the supersaturation state is the prerequisite, and it can be reached through shift of protein concentration and precipitant concentration according to the phase diagram (Figure IV.1).
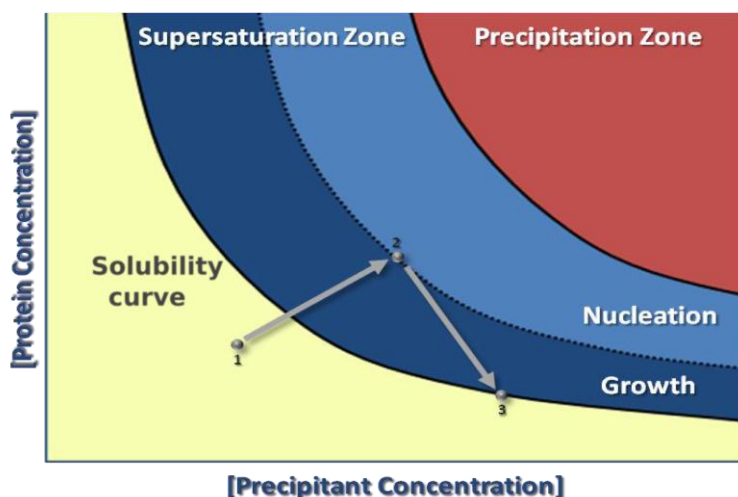
Figure IV.1: A schematic phase diagram for protein crystallization. A solubility curve is drawn in black line between yellow and dark blue region. The dots from 1 to 3 indicate the crystallization process: upon the increase of protein concentration or precipitant concentration, protein molecules reach supersaturation zone and start nucleation; crystals start growing on nucleus and stop growing when protein in solution reaches solubility curve. (Asherie, 2004; Lukat, 2011).

Nucleation is the process where the solute molecules scattered in solvent start to gather into clusters, which keep stable under the current operating conditions for a short instant. These stable clusters constitute the nuclei (Figure IV.1, point 2) for more protein molecules adsorption and lead to crystal growth. During crystal growing, the protein concentration in solution decreases, after reaching the solubility curve, the crystal stops growing (Figure IV.1, point 3).

The most common artificial method for producing crystals is called vapor diffusion, which is usually performed as hanging drop or sitting drop in laboratory (MacPherson, 1982). The reservoir liquid contains precipitants such as polyethylene glycols, alcohols and salts that reduce the solubility of protein. A drop of purified protein solution is mixed with a drop of reservoir solution on a cover slide. The slide is inverted over the well and sealed with grease to form a closed environment with reservoir solution. The protein concentration in the drop increases due to vapor leaving in order to reach equilibrium with its surroundings (Figure IV.2). During this process, the protein concentration will be brought into supersaturation state and crystals might appear.

protein 2µl +
reservoir 2µl   coverslip & grease

H_2O

reservoir ~ 1ml
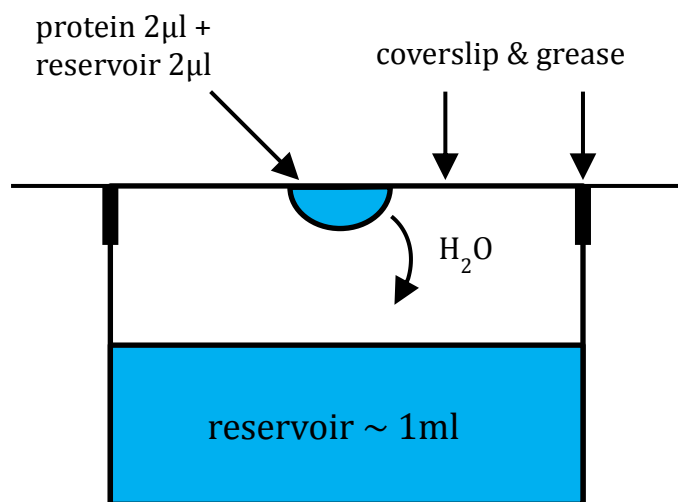
Figure IV.2: Schematic representation of the vapor diffusion method. A hang drop method was shown as an example

.

IV.1.2 Cryocrystallography

As X-rays have short wavelength electromagnetic spectrum (mostly from 0.01 to 10 nm) and carry high photon energies, X-ray photons could ionize atoms and disrupt molecular bonds easily. Due to this radiation damage, the diffraction quality of protein crystals will decay upon X-ray exposure. Even more damages are caused by water radicals, which could diffuse

and contact with protein molecules and thus destroy the regular crystal lattice. To minimize these radiation damages and ensure high quality diffraction data, a proper cryo-protectant has to be found. Crystals protected by cryo-protectant are immobilized by flash freezing in liquid nitrogen, and kept cooled by a flush of nitrogen gas stream at 100 K during measurement.

IV.1.3 Protein Quantitation

For quick determination of protein concentrations, a Nanodrop Protein A280 program was used. The molar weight and molar extinction coefficient are important for the concentration calculation. Both parameters were given by Vector NTI 10.0 software (?). The molar extinction coefficient is determined at a wavelength of $\lambda = 280$ nm, at which protein molecules show a specific absorption peak according to their content of aromatic amino acid side chains. According to Lambert-Beer's law (Beer, 1852), the absorption A of a protein molecule at a specific wavelength is proportional to its concentration c in the cuvette. The proportionality factors are the path length of the cuvette d and the extinction coefficient ε.

$$A_{280} = \varepsilon_{280} \cdot c \cdot d$$

Given the measured value of protein absorption at 280nm (A_{280}), the protein extinction coefficient (ε_{280}), and the known path length of the cuvette d, the protein concentration can be determined.

IV.1.4 R factor

In crystallography, the R factor (R_{work}) is an evaluation of agreement between the atomic model and the experiment X-ray diffraction data. It is defined by the following equation:

$$R = \frac{\sum \left| |F_{obs}| - |F_{calc}| \right|}{\sum |F_{obs}|}$$

where F is the structure factor and the sum extends over all the experimental measurements and the ideal calculated values. The structure factor is positively related to the intensity of the reflection. To avoid bias during refinement process, the R_{free} value is introduced. Before refinement, about 10% of the observations are removed from the data set. Then the

refinement is performed with the left 90%. The R_{free} is calculated to see the agreement between the predicted model and the 10% dataset that were not used in refinement (Brunger, 1992).

IV.1.5 Graphical representations

Graphical representations of the protein structures were made using the program PYMOL (Schrödinger LLC, 2009). Molecular surface potentials were calculated using DELPHI (Honig & Nicholls, 1995; Rocchia *et al.*, 2001). Sequence alignments were produced using CLUSTALW (Thompson *et al.*, 2002) and plotted using ESPRIPT (Gouet *et al.*, 1999). Chemical structures were drawn using CHEMSKETCH (Advanced Chemistry Development, Inc., 2006).

IV.2 *Yersinia* glycolytic enzymes

IV.2.1 Cloning and Expression

YePykF and YePykA of *Yersinia enterocolitica* WA-314 were cloned according to a modified protocol published by our collaborator earlier (Hofmann et al., 2013). *Ye*PykF and *Ye*PykA gene were cloned into pWS plasmid and later transformed into *Escherichia coli* BL21 (DE3) pLysS competent cells. Positive colonies were selected and cultured in 2YT medium at 310k, 150 rpm overnight. The overnight culture was inoculated into 2 Liters of fresh 2YT medium with the percentage of 5%. The culture was shaking at 150rpm, 310k until OD_{600nm} reached 0.6. Expression was induced by addition of 0.1mM of IPTG. After 5 hours of incubation at 300k, the culture was harvested by centrifugation at 6000 rpm for 10min and the pallet was resuspended in resuspension buffer [10 mM Tris pH 8.5, 50 mM KCl, 5 mM DTT, Complete Protease Inhibitor Cocktail (Roche Diagnostics GmbH)]. To disrupt cells, the resuspension was passaged through EmulsiFlex-C3 homogenizer (Avestin) three times. After centrifuging the cell lysate at 20 000g for 20min, the supernatant was filtrated through a 0.2 μm sterile filters (Sartorius, Germany).

IV.2.2 Protein purification

Non-tagged YePykF and YePykA were purified according to the protocol of our collaborator (Hofmann et al., 2013). A HiPrep16/10 Q XL column (GE Healthcare) was employed and equilibrated with 5 column volumes (CV) of buffer A (10 mM Tris-HCl pH 8.5, 50 mM KCl, 1 mM DTT) on ÄKTA purifier system (GE Healthcare). After applying the supernatant, the column was washed with 10 CV of buffer A and eluted by a 20 CV linear gradient from 0% to 25% buffer B (10 mM Tris-HCl pH 8.5, 1 M KCl, 1 mM DTT). The eluted fractions were collected and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). PykF, which has a PI of 6 based on the primary sequence, was eluted in a range of 160-200 mM KCl. PykA with a PI of 6.8 was eluted in a range of 130-185 mM KCl. The fractions including PykF and PykA were pooled and concentrated with a Vivaspin 20 filter (Sartorius Stedim Biotech) with a cut-off of 30 kDa.

A Superdex^{TM} 200 HR 10/30 column (GE Healthcare) was equilibrated with buffer C (10 mM Tris-HCl pH 8, 100 mM KCl, 1 mM DTT). 0.5 mL of concentrated protein (2% of column volume) was injected and loaded onto the column. The elution was carried on with the same buffer for 1.5 CV at a flow rate of 0.5 mL/min. The elutions were fractionated and analyzed by SDS-PAGE. Elute fractions including pure YePykF and YePykA were pooled and concentrated by Vivaspin 2 filter (Sartorius Stedim Biotech) with the cut-off of 30 kDa. The concentration of protein was determined with Nanodrop Protein A280 program as described before. Both proteins were adjusted to a concentration about 8 mg/ml for the crystallization test.

IV.2.3 Thermoshift assay

A 50 µl testing system was set up in 96 well PCR plate. The system included 20mM Tris-HCl, pH 8.0, 100mM KCl, 10x SYBRO Orange (Invitrogen), 0.5 mg/ml YePykF or YePykA protein, 1 mM of the test ligand. In negative controls, no ligand was added. Common binding partners like pyruvate, oxalate and PEP were tested for both proteins. Additionally, ADP and F16BP were tested for YePykF, ribose 5-phosphate and AMP were tested for YePykA. The plate was sealed with optical quality sealer. The plate was slowly heated up from 283 K to 363 K by 0.5 K increment. The excitation and emission wavelengths were 485 and 575 nm, respectively. The fluorescence signal was analyzed with CFX Manager software.

IV.2.4 Circular dichroism (CD) spectroscopy

CD data of YePykF and YePykA were acquired on a Jasco J-815 CD spectrometer (Jasco, Inc) in house. Since F16BP was studied having the largest melting temperature shift and leads to the most stable conformation, YePykF+F16BP was also tested by CD spectroscopy. PykA (22 μM), PykF (10 μM), and PykF in the presence of F16BP (10 mM) were equilibrated in 20 mM Tris-HCl (pH 8.0) and measured at 283 K in a 0.1-cm path-length cuvette. Spectra were recorded in the 190-260 nm wavelength range with 1 nm increments (20 nm / min), 10 s averaging time, and 1 nm bandwidth for 10 repeats. The mean residue molar ellipticity was calculated by

$$[\Theta] = \Theta \times 100 \times M / C \times l \times n$$

where Θ is the ellipticity in degrees, l the optical path in cm, C the protein concentration in mg/ml, M is the protein's molecular mass, n the number of residues in the protein, and $[\Theta]$ the mean residue molar ellipticity in deg•cm^2•dmol^{-1}. The baseline-corrected spectra were used for protein secondary structure analysis.

IV.2.5 Crystallization of YePykF, YePykF + FBP and YePykA + AMP

Prior to crystallization setup, an ultracentrifugation was performed with freshly purified protein at 50 000 g for 20 min, at 277 k. Sparse matrix sampling screens were set up with Honeybee system (Digilab, Inc., Marlborough, USA) in 96-well IntelliPlates (Art Robbins Instruments, Sunnyvale, USA) with sitting drop method. Both YePykF and YePykA were buffer exchanged into a buffer contains10mM Tris-HCl, pH 8.0, 100mM KCl, with concentration of 7.8 mg/ml and 8.2 mg/ml respectively. Protein and reservoir volume ratio was set to 1:1 (0.2 μl : 0.2 μl) with a total volume of 0.4 μl. The sealed crystallization plates were incubated in 293 K and 285 K and later screened with Formulatrix Rock Imager (Formulatrix, Waltham, USA).

In order to optimize crystal size and quality, a fine screen was set up. A hanging drop method was employed to yield lager drops with a volume ratio of 1μl to 1 μl. 24-well hanging drop Cryschem 24-1 SBS plates (Hampton Research, Aliso Viejo, USA) were used for designing the finescreen experiments based on the best initial hit conditions. Two parameters of the

crystallization condition were varied along X-axis and Y-Axis of the 24 well plates, while the other parameters were kept unchanged.

Iterative matrix seeding was employed for PykF+FBP crystals and PykA+AMP crystals. A good quality crystal was selected from finescreen, and a seed - bead (Hampton Research, Aliso Viejo, USA) was used for smashing the crystal. First, 50 μl of reservoir solution was filled into a 1.5 ml Eppendorf tube including seed - bead. Secondly, a good quality crystal was fished out and washed once in 2 μl reservoir drop and later transferred to a new drop, which was then pipetted into the 1.5 ml Eppendorf tube containing the seed head. Thirdly, the 1.5 ml Eppendorf tube including seed - bead and crystal and reservoir buffer was vortexed 5 times and cooled down on ice. Every cycle include 30 seconds vortex and 30 seconds incubation on ice. Lastly, the 1.5 ml Eppendorf tube including smashed seeds was centrifuged at 12 000 g for 3 min to remove large particles. The seeds solution can be diluted in series and stored in -80 ℃ for a seeds stock. A 10 times serial dilution of seeds stock was made (0.1 μl of seeds solution was mixed with 0.5 μl of protein and 0.4 μl of reservoir solution) and used for seeding experiment. The plates were incubated in 293 K or 285 K and crystals could be harvested in 2 - 4 days.

IV.2.5.1 YePykF Crystallization

Lenticular Crystals of *Ye*PykF appeared at 285 k 6 months after screen set up. The condition contains 1.6 - 2 M ammonium sulfate, 0.1 M Tris/HCl, pH 8.0. Because the crystallization process took too long time, no optimization was performed. A single apo YePykF crystal was cryo-protected in 3M ammonium sulfate, 10% glycerol, 0.1 M Tris/HCl, pH 8.0, and was quickly dipped into liquid nitrogen for later data collection experiment.

IV.2.5.2 YePykF+F16BP Crystallization

In order to co-crystallize *Ye*PykF and its effector F16BP (10mM fructose 1, 6-diphosphate) was added to 8mg/mL PykF right before crystallization set up. Crystals of PykF and F16BP complex appeared in conditions consisting of 19-25% PEG 6000, 0.2M Tris/HCl, pH 7.6, 0.1M KCl at 293K. Tiny crystals can be observed under microscope 2-3 hours after plate has been setup. Finscreens were set up with pH shifted along X - axis from 7.0 to 8.0, while polyethylene glycol 6000 (PEG 6000) shifted along Y - axis from 18 % - 21%. 3 rounds of

matrix micro-seeding were performed, the reservoir solutions were set up with pH at 7.6, polyethylene glycol 6000 (PEG 6000) shifted along Y - axis of 24 - well plate from 16.5 % - 18%, with 0.5% as a step. A single crystal was cryo-protected in a solution containing 30% PEG 6000, 0.2M Tris/HCl, pH 7.6, supplemented with 20% glycerol and was fast frozen in liquid nitrogen for later X-ray diffraction test.

IV.2.5.3 YePykA+AMP Crystallization

YePykA (8mg/ml) was mixed with its activator AMP (10mM) immediately before crystallization set up. 3 dimensional crystals appeared in conditions containing 35% Pentaerythritol ethoxylate (15/4 EO/OH), 0.2M $CaCl_2$, 0.1M Tris/HCl, pH 8.0, at 285K. Finescreen was performed with pH shifted along X - axis from 6.0 to 9.0, and Pentaerythritol ethoxylate (15/4 EO/OH), Shifted along Y- axis from 29% to 34%. Matrix microseeding was performed with 26 % to 29 % Pentaerythritol ethoxylate (15/4 EO/OH). A single crystal was cryo-protected with 35% Pentaerythritol ethoxylate (15/4 EO/OH), 10% glycerol, 0.2M $CaCl_2$, 0.1M Tris / HCl, pH 8.0 and was quickly frozen in liquid nitrogen for later X-ray data collection.

IV.2.6 Structure determination

IV.2.6.1 Data collection and processing

X-ray diffraction data of PykF + FBP crystals were collected at beamline 14.2 of BESSY II (Helmholtz centrum Berlin, Germany) for 360 ° with a CCD detector at a wavelength of λ = 0.918 Å. X-ray diffraction data of PykF and PykA + AMP crystals were collected at beam line P11 Petra III/DESY, Hamburg for 360 ° with a Pilatus 6M detector, at a wavelength of λ = 1.033 Å. Data were processed with the XDS package (Kabsch, 2010), additional symmetry elements was evaluated by POINTLESS (Evans, 2006), unit cell content was analyzed by Matthews Coefficient (Matthews, 1968) from CCP4 software suite.

IV.2.6.2 Molecular replacement

Molecular replacement of *Ye*PykF + F16BP was carried out with Phaser_MR (Adams et al., 2010) against the monomer structure of *E. coli* PK (PDB ID: 1E0T). Non-conserved amino acids side chains of the search model were pruned with Chainsaw (Winn et al., 2011) prior to molecular replacement. The sequence identity of YePykF and E. coli PykF is 85.7 %. MR of YePykF was carried out against YePykF+F16BP structure. Molecular replacement was carried out against 2 ensembles. Ensemble 1 is domain B (from amino acid 71 to 167), ensemble 2 is YePykF+F16BP structure without B domain. MR of *Ye*PykA was carried out against *Ye*PykF-F16BP (36.7% sequence identity) structure as search model. MR model was produced by Chainsaw. Molecular replacement was also carried out against 2 ensembles. Ensemble 1 is the MR model domain B (residue 71 to 167), and ensemble 2 is MR model without domain B.

IV.2.6.3 Model building and refinement

The initial model was improved by iterative steps of manual building in Coot (Emsley and Cowtan, 2004; Emsley et al., 2010) and refinement in Phenix.refine (Adams et al., 2010). Rigid body refinement and restrained refinement of PykA+AMP was done using REFMAC5 (Murshudov et al. 2011). Where appropriate, ligands (F26BP, SO4 groups), metal ions and water molecules were added to the structure. Atomic displacement factors (B) were treated as being isotropic. The presence of anisotropic domain movement was acknowledged by carrying out TLS refinement. Refinement was stopped after converging values of R_{work} and R_{free} were reached. The PK activity assays and kinetics study have been done by our collaborator from Robert Koch Institute.

IV.3 HCV and human receptor interactions

IV.3.1 Cloning and expression of HCV sE2 (HCV sE2)

IV.3.1.1 Transient expression of HCV sE2 in HEK 293 cell line

HCV (strain JFH1) soluble E2 (sE2, residues 384-661) was fused with a His_6-tag and cloned into pcDNA 3.1-tpa eukaryotic expression vector. One-day prior transfection, HEK293-6E

cells were cultured in in F17 medium supplemented with 25 mg/L G418, 1 g/L Pluronic F86 and 7.5 mM L-glutamine, cells were adjusted to a concentration of 1.0×10^6 c/ml, with a culture volume of 500 ml. On the second day, cells were transfected with an optimized polyethylenimine PEI transfection procedure(Jager et al., 2013). 500 ng pcDNA 3.1 tpa-sE2 plasmids mixed with 1250ng of PEI (polyethylenimine) in F17 medium, after incubation at room temperature for 15 - 30 min, the mixture were added to the cells. Together with transfection, 10μM kifunensine (α-mannosidases inhibitor) was also added to the culture. Cultures were fed 48 h post transfection with 0.5% trypton TN1 plus one additional volume of fresh F17 medium (kifunensine was also added up to 10μM). 4.5 g/ml glucose was added to the culture 72 h post transfection. 5 days after transfection, the cell supernatant was harvested and buffer changed with PBS for later Western Blot detection and affinity purification.

IV.3.1.2 Stable cell line construction of HCV sE2

HCV E2 ectodmain was cloned into pFlpBtM II vector (Genebank ID: KC991095) with restriction enzymes NheI and HindIII. For stable transfection, 1.5×10^6 cells in 2 ml medium were transfected with pFlpBtM II-sE2 plasmids by using the U-24 program of Amaxa nucleofection device according to the guidelines (Lonza, Co- logne, Germany). 24 hours after transfection, the medium was changed and the cells were seeded into six-well plates for culturing at 37℃，5% CO2, 110 rpm on an Incutec K15-500 linear shaker. In the following days, the cells were expanded and seeded in 8-cm cell culture dishes, white colonies were pipetted out and limited diluted into 96-well plate for cell line purifying. The cell culture supernatant was then analyzed by SDS-PAGE and Western Blot. Conditioned medium (20 L) was produced with verified stable cell line by continuous perfusion with a 2.5 L bioreactor. Cell supernatant was diafiltrated against PBS buffer with a filter cutoff of 10 kDa, supplemented with a "complete" protease inhibitor cocktail (Roche).

IV.3.1.3 Purification of sE2

Initial purification of both supernatants from transient transfection and stable cell line are achieved by Ni-NTA affinity purification (Qiagen). After immobilization of the protein, the column was washed with wash buffer (100mM K2HPO4/KH2PO4, pH 8.0, 300mM NaCl)

and eluted with stepwise elution buffer (wash buffer supplemented with 25, 50, 75, 250mM imidazole). The protein was then buffer changed into 50mM Tris-HCl, pH 8.0, with a Viva spin 20 filter (cut off at 10kDa). Cation exchange chromatography (MonoS HR10/10, GE, 0-1M NaCl, 30 column volumes) was performed, followed by a size exclusion chromatography (Superdex 200 HR 10/300, GE) with a buffer containing 50mM citric acid, pH5.4, 200mM NaCl. The purification results were analyzed with SDS-PAGE and Western Blot. For Western Blot test, anti-His$_6$-tag mouse monoclonal antibody (Novagen, dilution 1:1000) was used as the first antibody and AP-conjugated Anti-Mouse IgG (H+L) (Promega, dilution 1:7500) was used as the second antibody, BCIP and NBT were added into staining solution. For analyzing sE2 characters, non-reduced SDS-PAGE (loading buffer without BME and no boiling) and Blue Native PAGE (NativePAGE™ Novex® Bis-Tris gels, 4-16%, Invitrogen) were also performed.

IV.3.1.4 Deglycosylation and crystallization

HCV sE2 was deglycosylated by EndoHf (NEB) the deglycosylation reaction system was prepared according to the user manual with a protein: EndoHf ratio 70 to 1. The reaction mixture was incubated at 37 ℃ for 2 h or at 4℃ over night. The deglycosylated product was then analyzed by SDS-PAGE and followed by another size exclusion chromatography with (Superdex 75 10/30, GE) a buffer containing 50mM citric acid, pH 5.4, 200 mM NaCl. Purified protein has been concentrated with a Viva Spin 2 filter (10kDa cut off) to 5mg/ml, commercial screens were set up with Honeybee dispenser.

IV.3.2 Cloning, expression and purification of human SRB1 ETD

IV.3.2.1 Transient expression of Fc fusion SRB1

Human SRB1 ETD (33-420) was cloned into pYD11 vector, which containing a PreScission protease cleavage site and a C-terminal hFc (Fc fragment of human IgG) coding sequence. Transient transfection protocol with HEK 293 cells was the same as II.C.1.1.

IV.3.2.2 Insect cell expression system for SRB1

Human SRB1 ETD (residues 33 - 433) was cloned into pFlpBtM II donor vector, recombinant bacmids were generated using the Tn7 transposition method in bacmids of EMBacY (MBY) system, which includes a YFP-gene as a maker for infection monitoring (Meyer et al., 2013; Trowitzsch et al., 2010). Sf21 (DSMZ #ACC 119) suspension cultures were cultivated in ExCell420 (SAFC) on orbital shakers at 100 rpm at 27 ℃ using a 2.5 cm orbit. 1 hour before transfection, 0.5×10^6 cells/well were seeded into 6-well-plates. For each transfection, 10 μl Superfect (Qiagen) and 2 μg isolated bacmids were mixed in 100 μl ExCell420 medium. After incubation at RT for 20 min, the mixture was added on to adherent cells. 5 hours after infection, the transfection medium was replaced with fresh medium. The cell supernatant was harvested 3-5 days post transfection by monitoring YFP expression, the virus titer was determined by plaque assay. For larger scale protein expression, Sf21 suspension cultures were prepared with an initial cell density of 0.5×10^6 cells/mL. The cells were infected with a MOI between1-3 or 10% of P1 virus stock. The cells were harvested 3-4 days post transfection by low speed centrifugation. Cell pallets were suspended in lysis buffer (50 mM Na-Phosphate, 300 mM NaCl, 5 mM imidazol, 0,5% NP40, 3 mM β-mercaptoethanol supplemented with 10 μg DNaseI, and a "Complete" protease inhibitor cocktail (Roche)), the supernatant was centrifuged at an ultra-speed 30 000 g for 1 hour.

IV.3.2.3 Purification of SRB1 ETD

Protein A Affinity Chromatography was used for isolation of Fc fusion constructs, strep-tactin affinity Chromatography (Jena Bioscience) was performed with insect cells' supernatant. Anion exchange chromatography (MonoQ HR10/10, GE, 0-1M NaCl, 30 column volumes) was performed and followed by a size exclusion chromatography (50mM Tris/HCl, pH 7.6, 200mM NaCl). Analysis of protein expression and purification were performed by SDS-PAGE and Western Blot. For Western Blot experiment, HRP- conjugated goat anti-human IgG (Fc specific) antibodies were diluted 1 to 10 000 times and incubated at 37℃ for 1h. Lumi-light Western Blotting substrate (3ml substrate A and 3ml substrate B) was added on to Polyvinylidene fluoride (PVDF) membrane, after incubation at RT for 5min, the membrane was screened with LAS-3000 imager system (Fujifilm).

IV.3.2.4 Functional analysis

HCV infection competition tests were done in Twincore, Hannover, by Dr. Dorothea Bankwitz. Hun 7.5 cells were seeded in 96-well plate 24 h prior to infection. Pre-incubation with Jc1, which contains Renilla-Luc as a reporter was added into Huh 7.5 cell culture. Inoculation of 100 µl virus and test proteins, and incubate for 4 h. 72 h later, luciferase activity was assayed according to (Koutsoudakis et al., 2006).

IV.3.2.5 Limited proteolysis

Limited proteolysis was performed using purified protein from transient transfection. A series (1:10, 1:100, 1:1000) of each protease (trypsin, chymotrypsin and pepsin) were diluted in dilution buffer (10mM HEPES and 500mM NaCl). 100 µg purified Fc-SRB1 protein was mixed with 10µl of each diluted protease (correspond to 1µg, 0.1µg, 0.01µg for each protease), incubate at RT for 30 min. Analyze the digest effect on SDS-PAGE, load protease as control. The protein bands from SDS-PAGE were blotted onto a PVDF membrane, stained with Ponceau S and send for N- terminal sequencing.

IV.3.3 Prolegumain and legumain

IV.3.3.1 Expression and purification of prolegumain

During our experiment of constructing stable cell lines with CHO lec 3.8.2.1, we found it constitutively secreting large amount of prolegumain into the culture media. The secreted prolegumain has the binding ability to Ni-NTA, protein A and strep-tactin resins. Conditioned medium (20 L) was produced with CHO lec 3.8.2.1cell line by continuous perfusion with a 2.5 L bioreactor. Cell supernatant was diafiltrated against Buffer A (50mM $NaHPO_4$, pH 7.6, 300mM NaCl) with a filter cutoff of 10 kDa. The concentrated cell supernatant was then loaded on a 5 ml Ni-NTA column, washed with 100ml buffer B (50mM $NaHPO_4$, pH 7.6, 300mM NaCl, 30mM Imidazole), and eluted with 50ml Buffer C (50mM $NaHPO_4$, pH 7.6, 300mM NaCl, 300mM Imidazole). Anion exchange chromatography and size exclusion chromatography were employed for further purification. Prolegumain has been verified by Mass Spectroscopy and N-terminal protein sequencing.

IV.3.3.2 Deglycosylation of prolegumain and its self-cleavage into legumain

Deglycosylation of prolegumain (1mg) with EndoHf (20mU) was carried out in sodium acetate buffer at pH 5.4, overnight at 37 ℃. After deglycosylation, another size exclusion chromatography was used to remove EndoHf and glycan residues. During deglycosylation at pH 5.4, an auto-cleavage into legumain also carried on at the same time.

IV.3.3.3 Crystallization

5mg/ml of purified legumain were used for crystallization tests. Crystallization screens were set up with commercial screens with Honeybee dispenser. Rod-shape crystals were observed at different conditions from initial screens. Three crystallization conditions have been chosen for fine screens. Condition1 contains 0.03M $MgCl_2$, 0.03M $CaCl_2$, 10% PEG 20 000, 20% PEG550 MME and 0.1 M MOPS/HEPES buffer, pH 7.5. Condition 2 has similar ingredients with condition 1 but has a buffer pH of 8.5. Condition 3 contains 0.1 M NaCl, 23% PEG 3350 and 0.1 M MOPS buffer, pH 6.5. The crystal was fished and cryo-protected and quickly dipped into liquid nitrogen for X-ray diffraction data collection.

IV.3.3.4 Data Collection, Processing, and Structure determination

The X-ray diffraction data were collected at beamline P11 of PETRAPETRA III Hamburg, Germany) at wavelengths of 1.033191 Å. Data were processed with XDS package (Kabsch, 2010) Molecular replacement was carried out with Phaser_MR against human prolegumain (PDB code: 4FGU) and legumain (PDB code: 4AWA). Prior to molecular replacement, non-conserved amino acids side chains of search models were pruned with Chainsaw. The initial model was improved by iterative steps of manual building in Coot and refinement in Phenix.refine. Refinement was stopped after converging values of R_{work} and R_{free} were reached.

References

Acton, S., Rigotti, A., Landschulz, K.T., Xu, S., Hobbs, H.H., and Krieger, M. (1996). Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. Science *271*, 518-520.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Bio.l Crystallogr. *66*, 213-221.

Agnello, V., Abel, G., Elfahal, M., Knight, G.B., and Zhang, Q.X. (1999). Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. Proc. Natl. Acad. Sci. USA *96*, 12766-12771.

Albecka, A., Belouzard, S., Op de Beeck, A., Descamps, V., Goueslain, L., Bertrand-Michel, J., Terce, F., Duverlie, G., Rouille, Y., and Dubuisson, J. (2012). Role of low-density lipoprotein receptor in the hepatitis C virus life cycle. Hepatology *55*, 998-1007.

Alvarez-Fernandez, M., Barrett, A.J., Gerhartz, B., Dando, P.M., Ni, J., and Abrahamson, M. (1999). Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. J. Biol. Chem. *274*, 19195-19203.

Anastasiou, D., Yu, Y., Israelsen, W.J., Jiang, J.K., Boxer, M.B., Hong, B.S., Tempel, W., Dimov, S., Shen, M., Jha, A.*, et al.* (2012). Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. Nat. Chem. Biol. *8*, 839-847.

Asherie, N. (2004). Protein crystallization and phase diagrams. Methods *34*, 266-272.

Axerio-Cilies, P., See, R.H., Zoraghi, R., Worral, L., Lian, T., Stoynov, N., Jiang, J., Kaur, S., Jackson, L., Gong, H.*, et al.* (2012). Cheminformatics-driven discovery of selective, nanomolar inhibitors for staphylococcal pyruvate kinase. ACS Chem. Biol. *7*, 350-359.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037-10041.

Bakszt, R., Wernimont, A., Allali-Hassani, A., Mok, M.W., Hills, T., Hui, R., and Pizarro, J.C. (2010). The crystal structure of Toxoplasma gondii pyruvate kinase 1. PLoS One *5*, e12736.

Bankwitz, D., Steinmann, E., Bitzegeio, J., Ciesek, S., Friesland, M., Herrmann, E., Zeisel, M.B., Baumert, T.F., Keck, Z.Y., Foung, S.K*., et al.* (2010). Hepatitis C virus hypervariable region 1 modulates receptor interactions, conceals the CD81 binding site, and protects conserved neutralizing epitopes. J. Virol. *84*, 5751-5763.

Bartenschlager, R., Lohmann, V., and Penin, F. (2013). The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. Nat. Rev. Microbiol. *11*, 482-496.

Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A., and Cosset, F.L. (2003). Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. J. Biol. Chem. *278*, 41624-41630.

Beer, B.J. (1852). The beer-lambert-bouguer law. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. Annalen der Physik und Chemie, *86*:78-88.

Bent, Z.W., Branda, S.S., and Young, G.M. (2013). The Yersinia enterocolitica Ysa type III secretion system is expressed during infections both in vitro and in vivo. Microbiologyopen *2*, 962-975.

Bottone, E.J. (1997). Yersinia enterocolitica: the charisma continues. Clin. Microbiol. Rev. *10*, 257-276.

Bradley, D.W., McCaustland, K.A., Cook, E.H., Schable, C.A., Ebert, J.W., and Maynard, J.E. (1985). Posttransfusion non-A, non-B hepatitis in chimpanzees. Physicochemical evidence that the tubule-forming agent is a small, enveloped virus. Gastroenterology *88*, 773-779.

Brazzoli, M., Bianchi, A., Filippini, S., Weiner, A., Zhu, Q., Pizza, M., and Crotta, S. (2008). CD81 is a central regulator of cellular events required for hepatitis C virus infection of human hepatocytes. J. Virol. *82*, 8316-8329.

Brunger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472-475.

Burghout, P., van Boxtel, R., Van Gelder, P., Ringler, P., Muller, S.A., Tommassen, J., and Koster, M. (2004). Structure and electrophysiological properties of the YscC secretin from the type III secretion system of Yersinia enterocolitica. J. Bacteriol. *186*, 4645-4654.

Calvo, D., and Vega, M.A. (1993). Identification, primary structure, and distribution of CLA-1, a novel member of the CD36/LIMPII gene family. J. Biol. Chem. *268*, 18929-18935.

Catanese, M.T., Loureiro, J., Jones, C.T., Dorner, M., von Hahn, T., and Rice, C.M. (2013a). Different requirements for scavenger receptor class B type I in hepatitis C virus cell-free versus cell-to-cell transmission. J. Virol. *87*, 8282-8293.

Catanese, M.T., Uryu, K., Kopp, M., Edwards, T.J., Andrus, L., Rice, W.J., Silvestry, M., Kuhn, R.J., and Rice, C.M. (2013b). Ultrastructural analysis of hepatitis C virus particles. Proc. Natl. Acad. Sci. USA *110*, 9505-9510.

Chen, J.M., Dando, P.M., Rawlings, N.D., Brown, M.A., Young, N.E., Stevens, R.A., Hewitt, E., Watts, C., and Barrett, A.J. (1997). Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase. J. Biol. Chem. *272*, 8090-8098.

Chen, J.M., Dando, P.M., Stevens, R.A., Fortunato, M., and Barrett, A.J. (1998). Cloning and expression of mouse legumain, a lysosomal endopeptidase. Biochem. J. *335 (Pt 1)*, 111-117.

Chen, J.M., Fortunato, M., and Barrett, A.J. (2000). Activation of human prolegumain by cleavage at a C-terminal asparagine residue. Biochem. J. *352 Pt 2*, 327-334.

Chen, J.M., Fortunato, M., Stevens, R.A., and Barrett, A.J. (2001). Activation of progelatinase A by mammalian legumain, a recently discovered cysteine proteinase. Biol. Chem. *382*, 777-783.

Choi, S.J., Reddy, S.V., Devlin, R.D., Menaa, C., Chung, H., Boyce, B.F., and Roodman, G.D. (1999). Identification of human asparaginyl endopeptidase (legumain) as an inhibitor of osteoclast formation and bone resorption. J. Biol. Chem. *274*, 27747-27753.

Choukhi, A., Ung, S., Wychowski, C., and Dubuisson, J. (1998). Involvement of endoplasmic reticulum chaperones in the folding of hepatitis C virus glycoproteins. J. Virol. *72*, 3851-3858.

Christofk, H.R., Vander Heiden, M.G., Harris, M.H., Ramanathan, A., Gerszten, R.E., Wei, R., Fleming, M.D., Schreiber, S.L., and Cantley, L.C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature *452*, 230-233.

Cook, W.J., Senkovich, O., Aleem, K., and Chattopadhyay, D. (2012). Crystal structure of Cryptosporidium parvum pyruvate kinase. PLoS One *7*, e46875.

Cornelis, G.R. (2006). The type III secretion injectisome. Nat. Rev. Microbiol. *4*, 811-825.

Dall, E., and Brandstetter, H. (2013). Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. Proc. Natl. Acad. Sci. USA *110*, 10940-10945.

Dao Thi, V.L., Granier, C., Zeisel, M.B., Guerin, M., Mancip, J., Granio, O., Penin, F., Lavillette, D., Bartenschlager, R., Baumert, T.F.*, et al.* (2012). Characterization of hepatitis C virus particle subpopulations reveals multiple usage of the scavenger receptor BI for entry steps. J. Biol. Chem. *287*, 31242-31257.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S.*, et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic. Acids. Res. *35*, W375-383.

Derewenda, Z.S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. Methods *34*, 354-363.

Derewenda, Z.S., and Vekilov, P.G. (2006). Entropy and surface engineering in protein crystallization. Acta Crystallogr. D Bio.l Crystallogr. *62*, 116-124.

Diao, J., Pantua, H., Ngu, H., Komuves, L., Diehl, L., Schaefer, G., and Kapadia, S.B. (2012). Hepatitis C virus induces epidermal growth factor receptor activation via CD81 binding for viral internalization and entry. J. Virol. *86*, 10935-10949.

Diaz, O., Delers, F., Maynard, M., Demignot, S., Zoulim, F., Chambaz, J., Trepo, C., Lotteau, V., and Andre, P. (2006). Preferential association of Hepatitis C virus with apolipoprotein B48-containing lipoproteins. J. Gen. Virol. *87*, 2983-2991.

References

Diepold, A., Amstutz, M., Abel, S., Sorg, I., Jenal, U., and Cornelis, G.R. (2010). Deciphering the assembly of the Yersinia type III secretion injectisome. EMBO J. *29*, 1928-1940.

Dombrauckas, J.D., Santarsiero, B.D., and Mesecar, A.D. (2005). Structural basis for tumor pyruvate kinase M2 allosteric regulation and catalysis. Biochemistry *44*, 9417-9429.

Drummer, H.E., and Poumbourios, P. (2004). Hepatitis C virus glycoprotein E2 contains a membrane-proximal heptad repeat sequence that is essential for E1E2 glycoprotein heterodimerization and viral entry. J. Biol. Chem. *279*, 30066-30072.

Dubuisson, J. (2000). Folding, assembly and subcellular localization of hepatitis C virus glycoproteins. Curr. Top. Microbiol. Immunol. *242*, 135-148.

Edgington, L.E., Verdoes, M., Ortega, A., Withana, N.P., Lee, J., Syed, S., Bachmann, M.H., Blum, G., and Bogyo, M. (2013). Functional imaging of legumain in cancer using a new quenched activity-based probe. J. Am. Chem .Soc. *135*, 174-182.

El Omari, K., Iourin, O., Harlos, K., Grimes, J.M., and Stuart, D.I. (2013). Structure of a pestivirus envelope glycoprotein E2 clarifies its role in cell entry. Cell Rep. *3*, 30-35.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Bio.l Crystallogr. *60*, 2126-2132.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Bio.l Crystallogr. *66*, 486-501.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Bio.l Crystallogr. *62*, 72-82.

Farquhar, M.J., Hu, K., Harris, H.J., Davis, C., Brimacombe, C.L., Fletcher, S.J., Baumert, T.F., Rappoport, J.Z., Balfe, P., and McKeating, J.A. (2012). Hepatitis C virus induces CD81 and claudin-1 endocytosis. J. Virol. *86*, 4305-4316.

Felmlee, D.J., Sheridan, D.A., Bridge, S.H., Nielsen, S.U., Milne, R.W., Packard, C.J., Caslake, M.J., McLauchlan, J., Toms, G.L., Neely, R.D.*, et al.* (2010). Intravascular transfer contributes to postprandial increase in numbers of very-low-density hepatitis C virus particles. Gastroenterology *139*, 1774-1783, 1783 e1771-1776.

Fraser, J., Boo, I., Poumbourios, P., and Drummer, H.E. (2011). Hepatitis C virus (HCV) envelope glycoproteins E1 and E2 contain reduced cysteine residues essential for virus entry. J. Biol. Chem. *286*, 31984-31992.

Gastaminza, P., Dryden, K.A., Boyd, B., Wood, M.R., Law, M., Yeager, M., and Chisari, F.V. (2010). Ultrastructural and biophysical characterization of hepatitis C virus particles produced in cell culture. J. Virol. *84*, 10999-11009.

Geiger, T., and Clarke, S. (1987). Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. J. Biol. Chem. *262*, 785-794.

Germi, R., Crance, J.M., Garin, D., Guimet, J., Lortat-Jacob, H., Ruigrok, R.W., Zarski, J.P., and Drouet, E. (2002). Cellular glycosaminoglycans and low density lipoprotein receptor are involved in hepatitis C virus adsorption. J. Med. Virol. *68*, 206-215.

Griffin, L., and Lawson, A. (2011). Antibody fragments as tools in crystallography. Clin. Exp. Immunol. *165*, 285-291.

Gulick, A.M., Horswill, A.R., Thoden, J.B., Escalante-Semerena, J.C., and Rayment, I. (2002). Pentaerythritol propoxylate: a new crystallization agent and cryoprotectant induces crystal growth of 2-methylcitrate dehydratase. Acta Crystallogr. D Bio.l Crystallogr. *58*, 306-309.

Halfon, S., Patel, S., Vega, F., Zurawski, S., and Zurawski, G. (1998). Autocatalytic activation of human legumain at aspartic acid residues. FEBS Lett. *438*, 114-118.

Hayward, S., and Berendsen, H.J. (1998). Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins *30*, 144-154.

Hayward, S., and Lee, R.A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. J. Mol. Graph. Model. *21*, 181-183.

He, L.F., Alling, D., Popkin, T., Shapiro, M., Alter, H.J., and Purcell, R.H. (1987). Determining the size of non-A, non-B hepatitis virus by filtration. J Infect Dis *156*, 636-640.

Heinz, D.W. (2013). Secrets of a secretin. Structure *21*, 2098-2099.

Hodgkinson, J.L., Horsley, A., Stabat, D., Simon, M., Johnson, S., da Fonseca, P.C., Morris, E.P., Wall, J.S., Lea, S.M., and Blocker, A.J. (2009). Three-dimensional reconstruction of the Shigella T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. Nat. Struct. Mol. Biol. *16*, 477-485.

Hoekstra, M., Van Berkel, T.J., and Van Eck, M. (2010). Scavenger receptor BI: a multi-purpose player in cholesterol and steroid metabolism. World J. Gastroenterol. *16*, 5916-5924.

Hofmann, J., Heider, C., Li, W., Krausze, J., Roessle, M., and Wilharm, G. (2013). Recombinant production of Yersinia enterocolitica pyruvate kinase isoenzymes PykA and PykF. Protein Expr. Purif. *88*, 243-247.

Jager, V., Bussow, K., Wagner, A., Weber, S., Hust, M., Frenzel, A., and Schirrmann, T. (2013). High level transient production of recombinant antibodies and antibody fusion proteins in HEK293 cells. BMC Biotechnol. *13*, 52.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr. D Bio.l Crystallogr. *66*, 133-144.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577-2637.

Kanno, H., Utsugisawa, T., Aizawa, S., Koizumi, T., Aisaki, K., Hamada, T., Ogura, H., and Fujii, H. (2007). Transgenic rescue of hemolytic anemia due to red blood cell pyruvate kinase deficiency. Haematologica *92*, 731-737.

Kato, N., Ootsuyama, Y., Ohkoshi, S., Nakazawa, T., Sekiya, H., Hijikata, M., and Shimotohno, K. (1992). Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. Biochem. Biophys. Res. Commun. *189*, 119-127.

Khan, A.G., Whidby, J., Miller, M.T., Scarborough, H., Zatorski, A.V., Cygan, A., Price, A.A., Yost, S.A., Bohannon, C.D., Jacob, J.*, et al.* (2014). Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. Nature. *509*,381-384.

Kong, L., Giang, E., Nieusma, T., Kadam, R.U., Cogburn, K.E., Hua, Y., Dai, X., Stanfield, R.L., Burton, D.R., Ward, A.B.*, et al.* (2013). Hepatitis C virus E2 envelope glycoprotein core structure. Science *342*, 1090-1094.

Kono, Y., Hayashida, K., Tanaka, H., Ishibashi, H., and Harada, M. (2003). High-density lipoprotein binding rate differs greatly between genotypes 1b and 2a/2b of hepatitis C virus. J. Med. Virol. *70*, 42-48.

Koonin, E.V., Senkevich, T.G., and Dolja, V.V. (2006). The ancient Virus World and evolution of cells. Biol. Direct *1*, 29.

Koutsoudakis, G., Kaul, A., Steinmann, E., Kallis, S., Lohmann, V., Pietschmann, T., and Bartenschlager, R. (2006). Characterization of the early steps of hepatitis C virus infection by using luciferase reporter viruses. J. Virol. *80*, 5308-5320.

Kranz, J.K., and Schalk-Hihi, C. (2011). Protein thermal shifts to identify low molecular weight fragments. Methods Enzymol. *493*, 277-298.

Krey, T., d'Alayer, J., Kikuti, C.M., Saulnier, A., Damier-Piolle, L., Petitpas, I., Johansson, D.X., Tawar, R.G., Baron, B., Robert, B.*, et al.* (2010). The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. PLoS pathogens. *6*, e1000762.

Krieger, M. (2001). Scavenger receptor class B type I is a multiligand HDL receptor that influences diverse physiologic systems. J. Clin. Invest. *108*, 793-797.

Kudryashev, M., Stenta, M., Schmelz, S., Amstutz, M., Wiesand, U., Castano-Diez, D., Degiacomi, M.T., Munnich, S., Bleck, C.K., Kowal, J.*, et al.* (2013). In situ structural analysis of the Yersinia enterocolitica injectisome. Elife *2*, e00792.

Lavanchy, D. (2009). The global burden of hepatitis C. Liver Int. *29 Suppl 1*, 74-81.

Lavie, M., Goffard, A., and Dubuisson, J. (2007). Assembly of a functional HCV glycoprotein heterodimer. Current issues in molecular biology *9*, 71-86.

Lewen, S., Zhou, H., Hu, H.D., Cheng, T., Markowitz, D., Reisfeld, R.A., Xiang, R., and Luo, Y. (2008). A Legumain-based minigene vaccine targets the tumor stroma and suppresses breast cancer growth and angiogenesis. Cancer Immunol. Immunother. *57*, 507-515.

References

Li, Y., Wang, J., Kanai, R., and Modis, Y. (2013). Crystal structure of glycoprotein E2 from bovine viral diarrhea virus. Proc. Natl. Acad. Sci. USA *110*, 6805-6810.

Lindenbach, B.D. (2013). Virion assembly and release. Curr. Top. Microbiol. Immunol. *369*, 199-218.

Lindenbach, B.D., and Rice, C.M. (2013). The ins and outs of hepatitis C virus entry and assembly. Nat. Rev. Microbiol. *11*, 688-700.

Liu, C., Sun, C., Huang, H., Janda, K., and Edgington, T. (2003). Overexpression of legumain in tumors is significant for invasion/metastasis and a candidate enzymatic target for prodrug therapy. Cancer Res. *63*, 2957-2964.

Liu, Y., Bajjuri, K.M., Liu, C., and Sinha, S.C. (2012). Targeting cell surface alpha(v)beta(3) integrin increases therapeutic efficacies of a legumain protease-activated auristatin prodrug. Mol. Pharm. *9*, 168-175.

Lukat, P. (2011). Structural investigation on the Fe (II)/α-ketoglutarate dependent dioxygenase PlaO1 from Streptomyces sp. TÜ6071 (PhD thesis, University of Freiburg), pp. 55-56.

Luo, Y., Zhou, H., Krueger, J., Kaplan, C., Lee, S.H., Dolman, C., Markowitz, D., Wu, W., Liu, C., Reisfeld, R.A.*, et al.* (2006). Targeting tumor-associated macrophages as a novel strategy against breast cancer. J. Clin. Invest. *116*, 2132-2141.

Lupberger, J., Zeisel, M.B., Xiao, F., Thumann, C., Fofana, I., Zona, L., Davis, C., Mee, C.J., Turek, M., Gorke, S.*, et al.* (2011). EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. Nat. Med. *17*, 589-595.

MacPherson, A. (1982). Preparation and analysis of protein crystals (New York). 93-97.

McPherson A. (2001). A comparison of salts for the crystallization of macromolecules. Protein Sci, 10:418-422.

Maehr, R., Hang, H.C., Mintern, J.D., Kim, Y.M., Cuvillier, A., Nishimura, M., Yamada, K., Shirahama-Noda, K., Hara-Nishimura, I., and Ploegh, H.L. (2005). Asparagine endopeptidase is not essential for class II MHC antigen presentation but is required for processing of cathepsin L in mice. J. Immunol. *174*, 7066-7074.

Manoury, B., Hewitt, E.W., Morrice, N., Dando, P.M., Barrett, A.J., and Watts, C. (1998). An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. Nature *396*, 695-699.

Mattevi, A., Valentini, G., Rizzi, M., Speranza, M.L., Bolognesi, M., and Coda, A. (1995). Crystal structure of Escherichia coli pyruvate kinase type I: molecular basis of the allosteric transition. Structure (London, England : 1993) *3*, 729-741.

Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491-497.

Mazurek, S. (2011). Pyruvate kinase type M2: A key regulator of the metabolic budget system in tumor cells. The International Journal of Biochemistry & Cell Biology *43*, 969-980.

Mazurek, S., Boschek, C.B., Hugo, F., and Eigenbrodt, E. (2005). Pyruvate kinase type M2 and its role in tumor growth and spreading. Semin. Cancer Biol. *15*, 300-308.

McCaffrey, K., Boo, I., Poumbourios, P., and Drummer, H.E. (2007). Expression and characterization of a minimal hepatitis C virus glycoprotein E2 core domain that retains CD81 binding. Journal of virology *81*, 9584-9590.

McCaffrey, K., Gouklani, H., Boo, I., Poumbourios, P., and Drummer, H.E. (2011). The variable regions of hepatitis C virus glycoprotein E2 have an essential structural role in glycoprotein assembly and virion infectivity. J. Gen. Virol. *92*, 112-121.

Merz, A., Long, G., Hiet, M.S., Brugger, B., Chlanda, P., Andre, P., Wieland, F., Krijnse-Locker, J., and Bartenschlager, R. (2011). Biochemical and morphological properties of hepatitis C virus particles and determination of their lipidome. J. Biol. Chem. *286*, 3018-3032.

Meyer, S., Lorenz, C., Baser, B., Wordehoff, M., Jager, V., and van den Heuvel, J. (2013). Multi-host expression system for recombinant production of challenging proteins. PLoS One *8*, e68674.

Michalak, J.P., Wychowski, C., Choukhi, A., Meunier, J.C., Ung, S., Rice, C.M., and Dubuisson, J. (1997). Characterization of truncated forms of hepatitis C virus glycoproteins. J. Gen. Virol. *78 (Pt 9)*, 2299-2306.

References

Monazahian, M., Bohme, I., Bonk, S., Koch, A., Scholz, C., Grethe, S., and Thomssen, R. (1999). Low density lipoprotein receptor as a candidate receptor for hepatitis C virus. J. Med. Virol. *57*, 223-229.

Morgan, H.P., McNae, I.W., Nowicki, M.W., Hannaert, V., Michels, P.A., Fothergill-Gilmore, L.A., and Walkinshaw, M.D. (2010). Allosteric mechanism of pyruvate kinase from Leishmania mexicana uses a rock and lock model. J. Biol. Chem. *285*, 12892-12898.

Morgan, H.P., O'Reilly, F.J., Wear, M.A., O'Neill, J.R., Fothergill-Gilmore, L.A., Hupp, T., and Walkinshaw, M.D. (2013). M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation. Proceedings of the National Academy of Sciences *110*, 5881-5886.

Morgan, H.P., Walsh, M.J., Blackburn, E.A., Wear, M.A., Boxer, M.B., Shen, M., Veith, H., McNae, I.W., Nowicki, M.W., Michels, P.A.*, et al.* (2012). A new family of covalent inhibitors block nucleotide binding to the active site of pyruvate kinase. Biochem. J. *448*, 67-72.

Murthy, R.V., Arbman, G., Gao, J., Roodman, G.D., and Sun, X.F. (2005). Legumain expression in relation to clinicopathologic and biological variables in colorectal cancer. Clin. Cancer. Res. *11*, 2293-2299.

Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. *2*, 2212-2221.

Nowicki, M.W., Tulloch, L.B., Worralll, L., McNae, I.W., Hannaert, V., Michels, P.A., Fothergill-Gilmore, L.A., Walkinshaw, M.D., and Turner, N.J. (2008). Design, synthesis and trypanocidal activity of lead compounds based on inhibitors of parasite glycolysis. Bioorg. Med. Chem. *16*, 5050-5061.

Ohno, Y., Nakashima, J., Izumi, M., Ohori, M., Hashimoto, T., and Tachibana, M. (2013). Association of legumain expression pattern with prostate cancer invasiveness and aggressiveness. World J. Urol. *31*, 359-364.

Perutz, M.F. (1989). Mechanisms of cooperativity and allosteric regulation in proteins. Q. Rev. Biophys. *22*, 139-237.

Ploss, A., Evans, M.J., Gaysinskaya, V.A., Panis, M., You, H., de Jong, Y.P., and Rice, C.M. (2009). Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. Nature *457*, 882-886.

Ponce, E. (1999). Effect of growth rate reduction and genetic modifications on acetate accumulation and biomass yields in Escherichia coli. J. Biosci. Bioeng. *87*, 775-780.

Poornam, G.P., Matsumoto, A., Ishida, H., and Hayward, S. (2009). A method for the analysis of domain movements in large biomolecular complexes. Proteins *76*, 201-212.

Rigden, D.J., Phillips, S.E., Michels, P.A., and Fothergill-Gilmore, L.A. (1999). The structure of pyruvate kinase from Leishmania mexicana reveals details of the allosteric transition and unusual effector specificity. J. Mol. Biol. *291*, 615-635.

Ross-Thriepland, D., Amako, Y., and Harris, M. (2013). The C terminus of NS5A domain II is a key determinant of hepatitis C virus genome replication, but is not required for virion assembly and release. J. Gen. Virol. *94*, 1009-1018.

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R.M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R., and Vitelli, A. (2002). The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. The EMBO journal *21*, 5017-5025.

Schraidt, O., and Marlovits, T.C. (2011). Three-dimensional model of Salmonella's needle complex at subnanometer resolution. Science *331*, 1192-1195.

Shirahama-Noda, K., Yamamoto, A., Sugihara, K., Hashimoto, N., Asano, M., Nishimura, M., and Hara-Nishimura, I. (2003). Biosynthetic processing of cathepsins and lysosomal degradation are abolished in asparaginyl endopeptidase-deficient mice. J. Biol. Chem. *278*, 33194-33199.

Siddiquee, K.A., Arauzo-Bravo, M.J., and Shimizu, K. (2004). Effect of a pyruvate kinase (pykF-gene) knockout mutation on the control of gene expression and metabolic fluxes in Escherichia coli. FEMS Microbiol. Lett. *235*, 25-33.

Silva-Herzog, E., Ferracci, F., Jackson, M.W., Joseph, S.S., and Plano, G.V. (2008). Membrane localization and topology of the Yersinia pestis YscJ lipoprotein. Microbiology *154*, 593-607.

References

Silver, D.L. (2004). SR-BI and protein-protein interactions in hepatic high density lipoprotein metabolism. Rev. Endocr. Metab. Dis. *5*, 327-333.

Simmonds, P. (2013). The origin of hepatitis C virus. Curr. Top. Microbiol. Immunol. *369*, 1-15.

Smith, R., Johansen, H.T., Nilsen, H., Haugen, M.H., Pettersen, S.J., Maelandsmo, G.M., Abrahamson, M., and Solberg, R. (2012). Intra- and extracellular regulation of activity and processing of legumain by cystatin E/M. Biochimie *94*, 2590-2599.

Sourisseau, M., Michta, M.L., Zony, C., Israelow, B., Hopcraft, S.E., Narbus, C.M., Parra Martin, A., and Evans, M.J. (2013). Temporal analysis of hepatitis C virus cell entry with occludin directed blocking antibodies. PLoS Pathog. *9*, e1003244.

Spreter, T., Yip, C.K., Sanowar, S., Andre, I., Kimbrough, T.G., Vuckovic, M., Pfuetzner, R.A., Deng, W., Yu, A.C., Finlay, B.B.*, et al.* (2009). A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. Nat. Struct. Mol. Biol. *16*, 468-476.

Stephenson, R.C., and Clarke, S. (1989). Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. J. Biol. Chem. *264*, 6164-6170.

Suzuki, K., Ito, S., Shimizu-Ibuka, A., and Sakai, H. (2008). Crystal structure of pyruvate kinase from Geobacillus stearothermophilus. J. Biochem. *144*, 305-312.

Tadesse, D.A., Bahnson, P.B., Funk, J.A., Morrow, W.E., Abley, M.J., Ponte, V.A., Thakur, S., Wittum, T., DeGraves, F.J., Rajala-Schultz, P.J.*, et al.* (2013). Yersinia enterocolitica of porcine origin: carriage of virulence genes and genotypic diversity. Foodborne Pathog. Dis. *10*, 80-86.

Thomssen, R., Bonk, S., Propfe, C., Heermann, K.H., Kochel, H.G., and Uy, A. (1992). Association of hepatitis C virus in human sera with beta-lipoprotein. Med. Microbiol. Immunol. *181*, 293-300.

Tickle, I.J., Laskowski, R.A., and Moss, D.S. (1998). Rfree and the rfree ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement. Acta Crystallogr. D Bio.l Crystallogr. *54*, 547-557.

Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., and Berger, I. (2010). New baculovirus expression tools for recombinant protein complex production. J. Struct. Biol. *172*, 45-54.

Tulloch, L.B., Morgan, H.P., Hannaert, V., Michels, P.A., Fothergill-Gilmore, L.A., and Walkinshaw, M.D. (2008). Sulphate removal induces a major conformational change in Leishmania mexicana pyruvate kinase in the crystalline state. J. Mol. Biol. *383*, 615-626.

Valentini, G., Chiarelli, L., Fortin, R., Speranza, M.L., Galizzi, A., and Mattevi, A. (2000). The Allosteric Regulation of Pyruvate Kinase: A site-directed mutagenesis study. Journal of Biological Chemistry *275*, 18145-18152.

Vaney, M.C., and Rey, F.A. (2011). Class II enveloped viruses. Cell Microbiol. *13*, 1451-1459.

Weiner, A.J., Brauer, M.J., Rosenblatt, J., Richman, K.H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q.L., Houghton, M.*, et al.* (1991). Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology *180*, 842-848.

Whidby, J., Mateu, G., Scarborough, H., Demeler, B., Grakoui, A., and Marcotrigiano, J. (2009). Blocking hepatitis C virus infection with recombinant form of envelope protein 2 ectodomain. J. Virol. *83*, 11078-11089.

White, J.M., Delos, S.E., Brecher, M., and Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. Crit. Rev. Biochem. Mol. Biol. *43*, 189-219.

Wilke, S., Krausze, J., Gossen, M., Groebe, L., Jager, V., Gherardi, E., van den Heuvel, J., and Bussow, K. (2010). Glycoprotein production for structure analysis with stable, glycosylation mutant CHO cell lines established by fluorescence-activated cell sorting. Protein Sci. *19*, 1264-1271.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Bio.l Crystallogr. *67*, 235-242.

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S.*, et al.* (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat. Biotechnol. *29*, 735-741.

Yamane, T., Takeuchi, K., Yamamoto, Y., Li, Y.H., Fujiwara, M., Nishi, K., Takahashi, S., and Ohkubo, I. (2002). Legumain from bovine kidney: its purification, molecular cloning, immunohistochemical localization and degradation of annexin II and vitamin D-binding protein. Biochim. Biophys. Acta *1596*, 108-120.

Yonath, A., Bashan, A. (2004). Ribosomal crystallography: initiation, peptide bond formation, and amino acid polymerization are hampered by antibiotics. Annu. Rev. Microbiol. *58*, 233-251.

Zahid, M.N., Turek, M., Xiao, F., Thi, V.L., Guerin, M., Fofana, I., Bachellier, P., Thompson, J., Delang, L., Neyts, J.*, et al.* (2013). The postbinding activity of scavenger receptor class B type I mediates initiation of hepatitis C virus infection and viral dissemination. Hepatology *57*, 492-504.

Zeisel, M.B., Koutsoudakis, G., Schnober, E.K., Haberstroh, A., Blum, H.E., Cosset, F.-L., Wakita, T., Jaeck, D., Doffoel, M., Royer, C.*, et al.* (2007). Scavenger receptor class B type I is a key host factor for hepatitis C virus infection required for an entry step closely linked to CD81. Hepatology (Baltimore, Md.) *46*, 1722-1731.

Zhao, L., Hua, T., Crowley, C., Ru, H., Ni, X., Shaw, N., Jiao, L., Ding, W., Qu, L., Hung, L.W.*, et al.* (2014). Structural analysis of asparaginyl endopeptidase reveals the activation mechanism and a reversible intermediate maturation stage. Cell Res. *24*, 344-358.

Zhong, W., Morgan, H.P., McNae, I.W., Michels, P.A., Fothergill-Gilmore, L.A., and Walkinshaw, M.D. (2013). `In crystallo' substrate binding triggers major domain movements and reveals magnesium as a co-activator of Trypanosoma brucei pyruvate kinase. Acta Crystallogr. D Bio.l Crystallogr. *69*, 1768-1779.

Zona, L., Lupberger, J., Sidahmed-Adrar, N., Thumann, C., Harris, H.J., Barnes, A., Florentin, J., Tawar, R.G., Xiao, F., Turek, M.*, et al.* (2013). HRas signal transduction promotes hepatitis C virus cell entry by triggering assembly of the host tetraspanin receptor complex. Cell Host Microbe *13*, 302-313.

Zoraghi, R., See, R.H., Gong, H., Lian, T., Swayze, R., Finlay, B.B., Brunham, R.C., McMaster, W.R., and Reiner, N.E. (2010). Functional analysis, overexpression, and kinetic characterization of pyruvate kinase from methicillin-resistant Staphylococcus aureus. Biochemistry *49*, 7733-7747.

Zoraghi, R., See, R.H., Axerio-Cilies, P., Kumar, N.S., Gong, H., Moreau, A., Hsing, M., Kaur, S., Swayze, R.D., Worrall, L.*, et al.* (2011a). Identification of Pyruvate Kinase in Methicillin-Resistant Staphylococcus aureus as a Novel Antimicrobial Drug Target. Antimicrobial Agents and Chemotherapy *55*, 2042-2053.

Zoraghi, R., Worrall, L., See, R.H., Strangman, W., Popplewell, W.L., Gong, H., Samaai, T., Swayze, R.D., Kaur, S., Vuckovic, M.*, et al.* (2011b). Methicillin-resistant Staphylococcus aureus (MRSA) pyruvate kinase as a target for bis-indole alkaloids with antibacterial activities. J. Biol. Chem. *286*, 44716-44725.
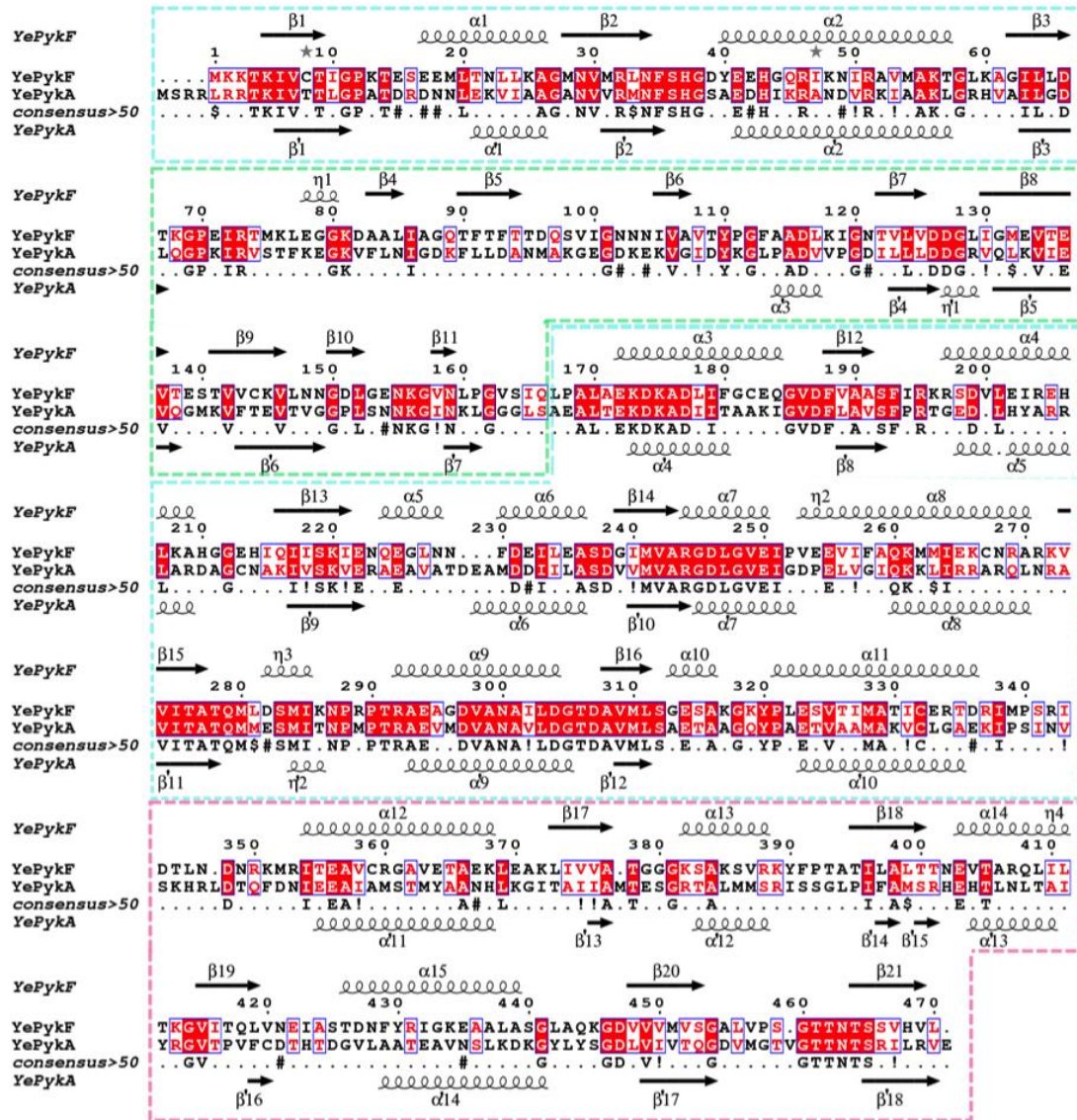
Appendix

Figure S1: Structure based sequence alignment of YePykF and YePykA. Domains are separated by dashed lines. A domain: cyan; B domain: green; C domain: pink.
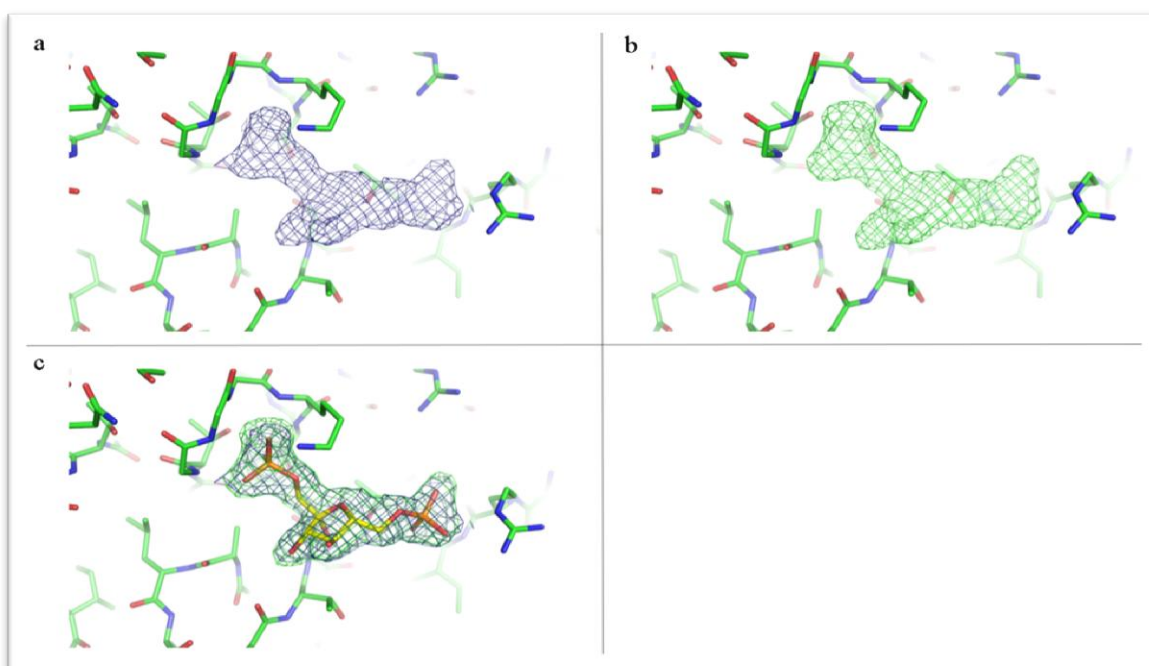
Figure S2: **Electron density maps of YePykF ligand F16BP.** (a) 2Fo-Fc map contoured at 1σ, blue mesh. (b) Fo-Fc map contoured at 3σ, green mesh. (c) The fit of F16BP into electron density maps.

Danksagung

First of all, I am extraordinarily grateful to my supervisor Prof. Dr. Dirk. W. Heinz for giving me such a good chance to work under excellent conditions with such fascinating projects. During my thesis, you gave me valuable advices with your experience and knowledge. With your support, I had the opportunity in attending many international conferences and expanded my knowledge in structure biology.

I wish to express my sincerely thanks to Dr. Jörn Krauße for always helping me in the lab and in front of the computer. Without your supervision, I wouldn't know how to do crystallization and how to solve the structure. Your patient explanation and wisdom removed hundreds of my puzzles. Many thanks for giving me suggestions and correcting my mistakes during my thesis writing.

I would like to thank Prof. Dr. Martin Korte for taking over the chairman of my thesis defense.

Many thanks for Prof. Dr. Dieter Jahn for taking over the responsibility as my supervisor in the university and also the examiner of my defense.

I am indebted to Prof. Dr. Thomas Pietschmann for the supervision and collaboration in the HCV project. Thanks for encouraging me all the time with the interesting topics. I would also like to thank Dr. Dorothea Bankwitz for the functional experiments.

I wish to express my deepest gratitude to Dr. Joop van den Heuvel for his excellent competence in protein expression. Thanks a lot for giving me valuable suggestions and generous support in the lab.

I would like to thank our Dr. Gottfried Wilharm, Christine Heider and Dr. Julia Hofmann for the collaboration in the project of "Yersinia glycolytic enzymes".

I am especially grateful to Dr. Victor Wray for sparing me his valuable time in reading my thesis and correcting my grammar mistakes.

I owe my deepest gratitude to my friend Millica Bajagic. She helped me so much in the initial correcting of my thesis. Thanks so much for always staying with me.

Thanks go to Dr. Peer Lukat and Dr. Joachim Reichelt for teaching me in using softwares and solving computer problems.

Thanks also go to Dr. Christian Kambach for his supervision in the beginning of my PhD. Thanks for teaching me how to format scientific report.

I am grateful to Dr. Andrea Scrima, Dr. Stefan Schmelz and Thomas Heidler for discussing with me about my problems.

It is my honor to work with the whole SFPR group, the friendly working atmosphere made my work smoothfully. Thank you all for the support in the lab and also in my life.

Finally, I would like to thank my parents for supporting me and giving me the maximum freedom to pursue my interest. Feifei, thanks for everything you have done for me in the past four years.

Curriculum vitae

Personal Details

Wei Li

Born on 21.04.1984, in Shandong, China

Single

Chinese

School Education

1991-1996	Elementary School in Xintai, Shandong
1996-2000	Middle School in Xintai, Shandong
2000-2004	High School in Tai'an, Shandong

Academic Education

2003-2007	Bachelor of Science, Bioscience, Shandong University, China
2007-2010	Master of Agricultural Science, Preventive Veterinary Medicine, Chinese Academy of Agricultural Sciences，China
2010-2014	Ph.D. Structure Biology, Technische Universität Braunschweig, Germany

Work Experience

2010	Scientist, Pharmaron, Beijing, China