



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

2015년 2월
박사학위논문

Crystal structures of bacterial proteins, PenA and CagL, related to human pathogenicity

조선대학교 대학원

생물신소재학과

최 진 명

Crystal structures of bacterial proteins, PenA and CagL, related to human pathogenicity

인간 병원성균의 단백질인

PenA와 CagL의 결정구조 연구

2015년 2월 25일

조선대학교 대학원

생물신소재학과

최 진 명

Crystal structures of bacterial proteins, PenA and CagL, related to human pathogenicity

지도교수 이 성 행

이 논문을 이학 박사학위신청 논문으로 제출함

2014년 10월

조선대학교 대학원

생물신소재학과

최 진 명

최진명의 박사학위논문을 인준함

위원장 조선대학교 교 수 유 호 진 (인)

위 원 조선대학교 부교수 이 성 행 (인)

위 원 조선대학교 부교수 방 일 수 (인)

위 원 조선대학교 조교수 김 은 애 (인)

위 원 전남대학교 부교수 임 영 준 (인)

2014년 12월

조선대학교 대학원

CONTENTS

List of Tables	i
List of Figures	ii
List of Abbreviations	iv
Abstract (in Korean)	1
I . Introduction	4
II . Materials and Methods	11
1. Plasmid constructs	11
2. Site-directed mutagenesis of CagL	11
3. Transformation and protein overexpression	11
4. Protein purification by Ni-NTA agarose	12
5. Dialysis and 6 x His-tag cutting by human α -thrombin	13
6. Ion exchange and size exclusion chromatography	13
7. Crystallization	13
8. X-ray crystallographic data collection, structure, determination, and refinement	14
9. Analytical size exclusion chromatography	16
10. Analytical ultracentrifugation (AUC)	16
III. Results	17
i. Structural study of class A β-lactamase PenA in <i>Burkholderia thailandensis</i>	18
1. Purification of recombinant proteins to PenA-WT and mutants	19
2. Preliminary crystallographic results of PenAs	21

3. Overall structure of PenA-WT and comparison with other β -lactamases	24
4. Reconstitution of the functional omega loop by insertion of repetitive residues in PenA-TR10 and PenA-TR11	28
5. The increase of binding cavity by insertion of repetitive residues in PenA-TR10 and PenA-TR11	29
6. Binding of a ceftazidime analog induces regrouping of the active site ...	32
7. Comparison of active site between PenA-WT and PenA-N136D	34
8. Structural change by CBA binding in PenA-N136D	35
9. Comparison of active sites between PenA-WT and PenA-C69Y	37
10. Structural change by CBA binding in PenA-C69Y	37
ii. Structural study of CagL, a component of the T4SS pilus assembly in <i>Helicobacter pylori</i> K74 strain	41
1. Overexpression and purification of K74 CagL	42
2. Preliminary crystallographic results of K74 CagL	43
3. Overall structure of K74 CagL	45
4. Identification of the oligomeric state of K74 CagL	48
5. Comparison of both K74 CagL and 26695 CagL	50
IV. Discussion	57
V. References	62
VI. Appendix	75
VII. Abstract (in English)	115
VIII. Acknowledgements	118

List of Tables

1. Data collection and refinement statistics of PenA (WT, TR10, TR10-CBA, TR11 and TR11-CBA)	39
2. Data collection and refinement statistics of PenA (N136D, N136D-CBA, C69Y, and C69Y-CBA)	40
3. Data collection and refinement statistics of K74 CagL	53
4. Primer sequence for site mutagenesis of K74 CagL	54
5. Constructs by site mutagenesis on K74 CagL	56

List of Figures

1. General hydrolysis pathway of active site serine β -lactamase and structure representation of cefotaxime, ceftazidime and ceftazidime-like glycyl boronic acid (CBA)	7
2. Model for the assembled T4SS machinery and its role in HP-induced cell signaling .	10
3. Result of SEC and analysis by SDS-PAGE to PenAs	19
4. Crystal and X-ray diffraction images of PenAs	22
5. Overall structure of PenA-WT and structural alignment of PenA-WT with other β -lactamases (TEM-1, SHV-1, and CTX-M-9)	26
6. Structural alignment of conserved residues and catalytic water with other β -lactamases, and water-mediated hydrogen bond network forming active site of PenA-WT	27
7. Representation of tandem repeats inserted on Ω loop in PenA-TR10 and PenA-TR11	28
8. Conformation change of PenA-TR10 and PenA-TR11 compared to PenA-WT	31
9. Conformation change in PenA-TR10 and PenA-TR11 by CBA binding	33
10. Conformation change of PenA-N136D and CBA-bound PenA-N136D compared to PenA-WT	36
11. Conformation change of PenA-C69Y and CBA-bound PenA-C69Y compared to PenA-WT	38
12. Analysis of purified Se-Met CagL by 12% SDS-PAGE	42

13. Crystal and X-ray diffraction images of Se-Met CagL	44
14. Overall structure of K74 CagL	46
15. Intermolecular interaction for architecture on K74 CagL	47
16. SEC of site-mutated CagL proteins	49
17. Analysis of oligomeric state by analytical SEC and AUC	49
18. Sequence alignment of K74 CagL and 26695 CagL.....	51
19. Comparison of structural difference by superimposition of K74 CagL and 26695 CagL	52

List of Abbreviations

SHV-1	Sulfhydryl reagent variable
ESBL	Extended spectrum β -lactamase
TR	Tandem repeat
CBA	Ceftazidime-like boronic acid
T4SS	Type 4 secretion system
Se-Met	Seleno-Methionine
CagPAI	Cytotoxic-associated genes pathogenicity island
SEC	Size exclusion chromatography
AUC	Analytical ultracentrifugation
IPTG	Isopropyl- β -D thiogalactopyranoside
FPLC	Fast protein liquid chromatography
PEG	Polyethyleneglycol

국문 초록

인간 병원성균의 단백질인 PenA과 CagL의 결정구조 연구

최 진 명

지도 교수: 이 성 행

생물신소재학과

조선대학교 일반대학원

Burkholderia thailandensis 내에서 A 부류 β -락타마아제 PenA의 구조 연구

일반적으로 박테리아 내에서 항생제에 대한 방어를 위해 생성 되는 β -락타마아제는 항생제의 β -락탐링 가수분해를 통해 비활성화 시킴으로써 항생제에 대한 저항성이 나타난다. 대부분의 박테리아는 생존을 위해 β -락타마아제 내에서 아미노산의 삭제, 대용, 반복 단위 삽입과 같은 변이에 의한 구조적인 변화를 통해 다양한 항생제에 대한 저항성을 통해 진화되어왔다. 여러 종류의 β -락타마아제 중에서, PenA는 *B. thailandensis* 내에 형성되며 페니실린계에 대한 저항성을 보이는 A 클래스 β -락타마아제 중 하나이다. 잘 알려져 있는 SHV-1 (sulfhydryl reagent variable)과 TEM-1과 같은 β -락타마아제와 함께, PenA 또한 아미노산의 대용과 삭제를 통해 Extended spectrum β -lactamase (ESBL)로 변화가 일어날수 있을 뿐만 아니라 Tandem repeat (TR) 복제 변이에 의해 가역적으로 두 상의 기질영역을 스위치 할 수 있는 능력이 발견되었다. 이러한 변이는 공통적으로 락타마아제 내의 어떤 변화에 의해 ceftazidime과 같은 3세대 세팔로스포린에 대한 저항성이 증가하는 것으로 여겨지고

있다. 비록 ESB�로 되는 점 변이에 관한 연구는 이미 다른 β -락타마아제로부터 알려져 있지만 ‘phase variation’을 하는 긴 반복적인 아미노산의 삽입 (Tandem repeat)에 의한 정확한 구조적 기전 연구는 현재 불분명 하다. 그리하여, PenA-WT과 ESB�이 되는 점 변이 (N136D 와 C69Y) 뿐만 아니라 기질의 스위치가 가능한 TR 변이 (TR10 과 TR11) 구조들에 대해 규명하기 위해 시도되었다. 각각에 해당되는 단백질을 발현, 정제하였고, ceftazidime analogue 인 ceftazidime-like boronic acid (CBA)를 포함하지 않은 구조뿐만 아니라 포함하는 구조와 함께 높은 해상도에서 각각의 x-ray 결정 데이터를 얻었다. 구조분석 결과, PenA-WT과 변이들 사이에서 다른 양상의 활성 면의 구조변화가 나타나지만, 공통적으로 활성 면의 구조 변화에 의한 cavity 변화를 통해 bulky한 CBA의 결합을 확인하였다. 이런 변화는 3세대 세팔로스포린인 ceftazidime과 같은 항생제에 대한 결합 활성도 뿐만 아니라 가수분해 활성도에 대한 촉매작용이 증가한다. 결과적으로 활성 면에서 오메가루프의 변화는 기질 영역의 변화에 매우 중요한 역할을 하고 다양한 항생제에 대한 저항성 증가를 유도하게 되지만, TR에서 원래 기질에 대한 활성도 변화를 위한 구조적인 증거는 여전히 불분명하게 남아있다. 결과적으로, 새로운 TR에 의한 구조와 CBA가 결합된 구조에 대한 연구를 통해 항생제 저항성에 대한 새로운 기전을 확인할 수 있고, 더 나아가 새로운 치료제 개발을 위한 정보를 제공할 수 있다.

한국형 헬리코박터 필로리, K74 내에서 타입 4 분비 체계(T4SS) 섬모 집합체의 구성요소인 CagL의 구조 연구.

헬리코박터 필로리 균체 내에 존재하고 40kb DNA에 해당하는 CagPAI에 의해 암호화되는 T4SS의 구성성분중의 일부는 숙주세포의 수용체와의 상호작용과 숙주세포 안으로 CagA 를 운반하는 중요한 역할을 담당한다. CagL은 타입 4 분비 체계

를 통해 숙주세포 안으로 CagA를 이동시키는 동안 유동적인 RGD domain을 통해 상호작용한다. 그러므로, 병원균-숙주세포의 상호작용에 대한 기전을 조사하기 위해 한국형 *H.pylori* K74 균으로부터 native 와 methionine 위치에 Seleno-Methionine으로 치환된 CagL 재조합 단백질을 발현, 정제하고 구조를 밝히기 위해 결정화시켰다. Se-Met 과 native CagL 결정은 2.9 Å와 2.8 Å 해상도로 회절 되었고, 이들은 동일한 정방정계인 스페이스 그룹 P4₃2₁2에 속하며 하나의 비대칭단위에서 이량체로 구조화되었다. 솔루션 상에서 CagL의 중합상태는 analytical ultracentrifugation (AUC)를 통해 단량체로 확인되었다. 단량체 CagL의 구조는 4개 긴 α-나선과 2개 짧은 α-나선의 나선묶음으로 구성된다. 인테그린과 상호작용하는 것으로 알려진 CagL의 RGD motif는 보존된 단백질 가수분해 효소에 민감하고 RGD helper 배열 사이에 놓인 길고 단단한 α2에서 발견되었다. 이러한 사실은 인테그린 결합을 위해 RGD 부분의 풀림의 가능성을 가리킨다. 그러나 26695 CagL과 비교시, K74 CagL α1-α2 loop은 잘 정의 되었고 두 개의 salt bridge가 두 나선에서 발견된 사실을 통해 추상적, 유동적인 부분을 포함하는 RGD motif는 K74 CagL에서 더욱 안정하다고 제안할 수 있다. 또한, 세포 내에서 CagL은 CagI와 CagH와 함께 복합체와 pilus 형성에 관여한다고 알려져 있기 때문에, 어떻게 CagL이 CagI 와 CagH과 함께 복합체를 형성하고 숙주 수용체인 인테그린과 함께 상호작용하는 기전에 대해 연구할 필요가 있다.

I . Introduction

i . Structural study of class A β -lactamase PenA in *Burkholderia thailandensis*

In bacteria, cell wall plays a very important role in maintaining cell integrity. Gram-positive bacteria consist of a thick layer of cell wall as well as a layer of cytoplasmic membrane. But, gram-negative bacteria are comprised of an outer membrane, periplasmic space, and a layer of cytoplasmic membrane. In prokaryotes, the conserved cell wall components could serve as a signal to host immunologic or pathologic responses [1-4]. Bacteria cell walls are made of peptidoglycan, which is essential for survival of bacteria. β -lactam antibiotics including penicillin inhibit cell wall biosynthesis by preventing the cross-linking of peptidoglycan and consequently induce to cell wall rupture and cell death. However, bacteria have evolved a defense system to evade this and thereby causing antibiotic resistance

The β -lactamases in bacteria, which can hydrolyze antibiotics, are usually produced for inactivating the antibiotic property [5]. These β -lactamases are homologically divided into 4 classes (A-D). Class A, C, and D act as a serine-based mechanism, whereas class B use metal cofactor such as zinc for hydrolysis [6].

In general, β -lactamases have five regions for forming cavity wall that plays a key role in binding and hydrolysis of substrate [7]. 1) the 70SXXK73 motif, in which Ser70 and Lys73 are involved in the proton shuttle during catalytic process [8], 2) in the SDN loop (130SDN132), Ser130 also participates in the proton shuttle and is critical for the activation of lactamase [9], 3) the 234KTG236 on β 3-strand has been demonstrated to be critical for binding and hydrolysis of β -lactams and β -lactamase inhibitor through hydrogen bond, 4) Ω loop (164-179) contains two conserved residues, Glu166 and Asn170, that seem essential in positioning the conserved water molecule very close to the active serine, and 5) the conserved Tyr105 residue in 102-110 loop is known to be important in the substrate stabilization and discrimination by an induced conformational change when substrate comes into the cavity [10, 11].

In active site, hydrolytic mechanism of antibiotics by β -lactamases forms a high-energy tetrahedral acylation intermediate by nucleophilic attack of serine on the carbonyl of the β -lactam antibiotics. This intermediate changes to lower-energy covalent acyl-enzyme and the C-N bond is broken. An activated water molecule attacks the covalent complex and leads to deacylation of the intermediates. Hydrolysis between the β -lactam carbonyl and the oxygen of the serine regenerates the active enzyme and releases the inactive β -lactam [8] (Fig.1A). In short, β -lactamases usually deduce hydrolysis of β -lactam through these acylation and deacylation intermediates for resistance to antibiotics.

Until now, the hydrolysis mechanism of antibiotics is well-known in β -lactamases, TEM and SHV family. Both β -lactamases, which were detected in clinical isolates of *E.coli* and *K.pneumoniae*, have resistance to penicillin and narrow spectrum cephalosporin. However, due to the exposure of newly developed cephalosporin containing oxy-imino side chain in environments, bacteria quickly fosters a mechanism of resistance that can hydrolyze third generation cephalosporin such as ceftazidime. To date, at least 29 TEM- and 11 SHV- related enzymes are reported to have increased hydrolytic activity against oxi-imino β -lactams [12-15]. These spectrum change of antibiotics mainly appears due to mutation of β -lactamase including point mutation and insertion of repeat units [16-18]. In addition to, the structural study for hydrolysis of antibiotics in TEM and SHV β -lactamases conducted by x-ray crystallographic study with inhibitors including ceftazidime-like glycyl boronic acid (CBA) [19-21]. Positions of residue in enzymes, which has from one to four amino acid substitution for extended spectrum β -lactamases (ESBLs) in mutants of TEM and SHV, located in close proximity to the active site (179, 205, 238, and 240 in SHV family and 104, 164, 237, 238, and 240 in TEM family) [8].

β -lactamases are encode in plasmid in case of TEM and SHV as well as on the chromosome like other enzymes including PenA in *B.pseudomallei* and GES-1 in *K.pneumoniae* [12]. The *Burkholderia* genus refers to a group of virtually ubiquitous Gram-

negative, motile, and aerobic rod shaped bacteria. Among genus of *Burkholderia*, *B. mallei*, *B. pseudomallei*, and *B. cepacia* are pathogenic members; *B. mallei*, responsible for glanders, a disease that occurs mostly in horses and related animals; *B. pseudomallei*, causative agent of melioidosis; and *Burkholderia cepacia*, an important pathogen of pulmonary infections in people with cystic fibrosis (CF). A few cases in *B. pseudomallei* and Bcc reported about resistance to ceftazidime by mutation including single amino acid substitutions, Pro167Ser and Cys69Tyr [22-25]. Also, structures to PenA and PenI, which corresponds to β -lactamase of *B. cepacia* complex and *B. pseudomallei*, are recently determined [7].

The β -lactamase PenA (NCBI reference sequence: YP_439646.1) using this study, which is expressed in *B. thailandensis* belongs to class A, can be used as good model to study substrate expansion due to resistance to amoxicillin and closely ortholog genes in the pathogenic species *B. pseudomallei*, *B. mallei*, Bcc [22-24]. Clones that had resistance to ceftazidime were selected by a large-scale mutant selection and were identified as result of point and TR mutations [26]. According to the reasons described above, point mutants including C69Y and N136D observe spectrum change to ceftazidime but failed reversion to original antibiotic. However, TR mutants were reported to switch reversibly between two phases of β -lactamase substrate spectrum [27]. But, exact structural mechanism by TR mutation remains unclear. Therefore, PenA-WT and mutants including point mutation (N136D and C69Y) and TR mutation (TR10 and TR11) were expressed, purified and crystallized to study the structural mechanism that cause change in substrate spectrum at active site. As a result, omega loop in both TR10 and TR11 mutant showed disruption by duplicated insertion of amino acid (TR10: 10 a.a, TR11: 20 a.a) and TR11 yielded more serious change than TR10 mutant. Other mutants, PenA-N136D and PenA-C69Y show change of active site by point mutation. All mutants are commonly bound with CBA due to the increased cavity of active site although the structure of omega loop changes differently, suggesting that enzyme might have increased volume of cavity by different structural change of Ω loop at active site for the bulky substrate to enter and bind.

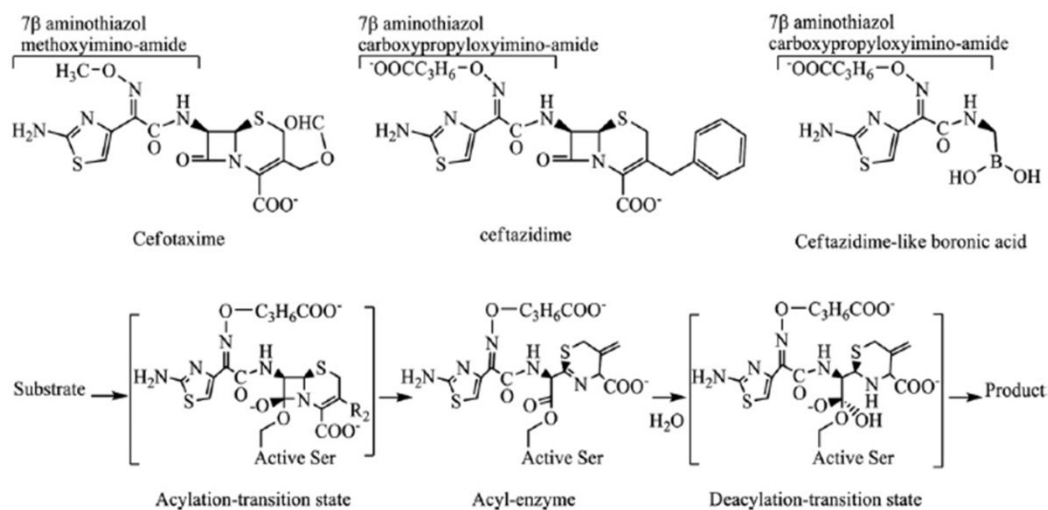


Figure 1. General hydrolysis pathway of active site serine β -lactamases and structure representation of cefotaxime, ceftazidime and ceftazidime-like glycol boronic acid (CBA) [28]

ii. Structural Study of CagL, a component of the T4SS pilus assembly in *Helicobacter pylori* K74 strain.

Helicobacter pylori, which is a helix-shaped gram-negative and microaerophilic bacteria, is a specific human pathogen found in the stomach that harbors more than 50% of the world's population [29]. It is named as a class 1 carcinogen by the world health organization [30]. People infected with *H.pylori* have an increased risk of developing gastric diseases including peptic ulcers, gastric adenocarcinoma, and gastric lymphoma [31-34]. The encoded CagPAI is the virulent gene in *H.pylori* strains and is associated with a higher rate of disease than its lacking strains [35-37]. Especially, *H.pylori* of East Asia was reported to have a higher risk of gastric cancer compared to western type due to the increased evidence of CagA phosphorylation and IL-8 production [38].

Secretion system of bacteria has a very important mechanism for the transport or translation of effector molecules into host cell. Gram negative bacteria have at least a six specialized secretion systems (Type1-6). Among 6 secretion system, type 4 secretion system (T4SS) is mainly a conjugation system found in gram-negative bacteria for transport of macromolecules into host cell [39, 40]. One of the most studied type 4 secretion system is the VirB system of *A.tumefaciens*, which is a transport nucleoprotein complexes into plant cell that causes crown gall disease [41]. The VirB secretion apparatus consists of twelve VirB proteins that are comprised of core complex (VirB3, VirB6, VirB7, VirB8, VirB9, and VirB10), periplasmic or extracellular pilus structure (VirB2, VirB5, and VirB7), two inner-membrane-associated ATPases (VirB4 and VirB11), a peptidoglycan hydrolyase (VirB1), and coupling protein (VirD4). Among VirB proteins, VirB5 was predicted to function very similar to CagL although it has low sequence identity. In similar with VirB5 system, T4SS of *H.pylori* also produced 27 proteins for assembly of secretion apparatus and effector protein by using a Cag (cytotoxic-associated genes) pathogenicity island encoded 40kb stretch of DNA. Among twenty-seven proteins, cag type 4 secretion apparatus consists of 14 essential components that

it is only display only four significant sequence similarities by comparing with components of the *A.tumefaciens* VirB system [42, 43].

The T4SS in *H.pylori* also forms pilus for delivery of oncoprotein such as CagA into the host [44, 45]. Mechanism for delivery of CagA starts from interaction between adhesion protein of *H.pylori* and receptor of epithelial cells. It binds to and mediates contact with surface receptor integrin of host cell for CagA translocation and interleukin-8 secretion.

CagL, which plays a bridge role for interaction with host cell, is a pathogenicity-associated factor from *H.pylori* [46]. It is mainly localized to several bacterial subcellular sites including a soluble bacterial fraction, the bacterial surface, and pili [46-49]. It lacks homologs in other bacterial species. It is reported that CagL can bind and recognize $\alpha 5\beta 1$ integrin as well as $\alpha v\beta 3$ and $\alpha v\beta 5$ [46, 50]. Although integrin is generally known to interact with the domain containing RGD motif like FN, CagL containing RGD motif remains unclear. One group has demonstrated that CagL is RGD-dependent for binding with receptor that plays an important role for initiation between *H.pylori* and epithelial cells [51, 52]. In contrast, other group suggested that translocation of CagA is RGD-independent for integrin binding through IL-8 production [53]. Although RGD-dependent or independent binding with integrin is argued until now, search for CagL structure has a very important meaning as initiation step for interaction between *H.pylori* and host cell. Recently, CagL variants of western type 26695, CagL^{KKQEK} and CagL^{meth}, exhibit structures that are comprised of an elongated four-helix bundle [54]. Unlike fibronectin, strikingly it was found that RGD motif was located at the center of second helix surface not same as the predicted structure of VirB5 protein, TraC. Also, 26695 CagL suggested that N-terminal region of $\alpha 2$ may be flexible for interaction with integrin. CagI and CagH, components comprised of cagPAI, also demonstrated about possibility of assembly since both the proteins were co-purified with CagL [55]. CagH, CagI, and CagL have a conserved c-terminal S/T-K-I/V-I-V-K hexapeptide motif. Deletion of this motif inhibited not only functional activity of T4SS but also formation of pilus assembly [55].

In this study, to identify CagL structure, CagL recombinant protein from *H. pylori* Korean strain k74 was purified to be soluble, crystallized and diffracted by x-ray. As a result, overall structure of k74 CagL monomer resembles with that of 26695 CagL variants structure that forms the elongated four helices and two short helices [54]. RGD motif was also located on the second long helix and was exposed on the surface. CagL identified molecular weight (~26kDa) of CagL-WT by analytical size exclusion chromatography and analytical ultracentrifugation (AUC), suggesting that exists to soluble monomer form in solution.

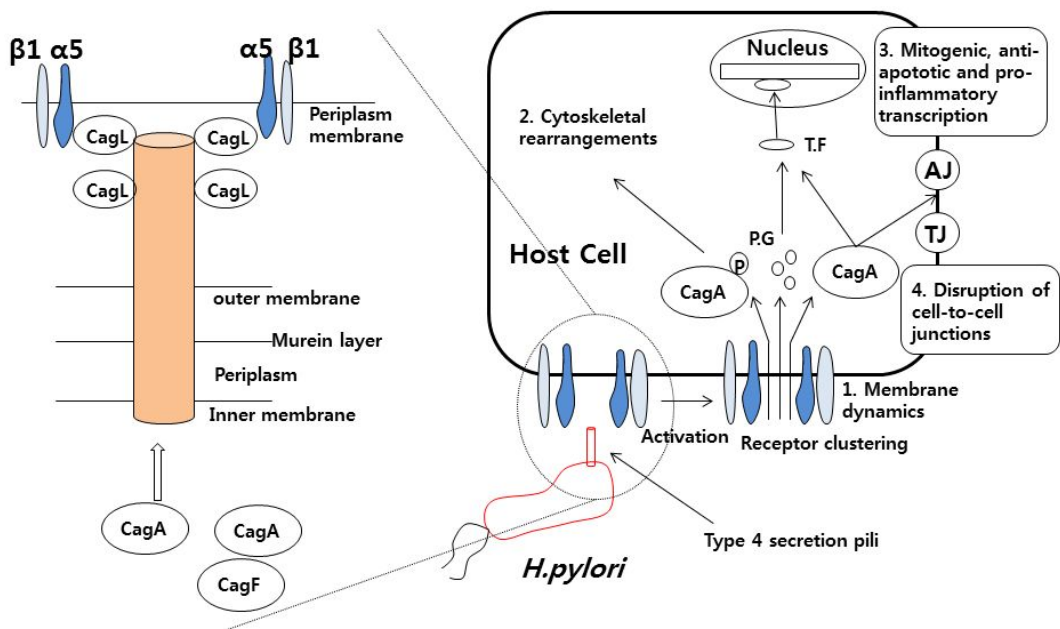


Figure 2. Model for the assembled T4SS machinery and its role in HP-induced cell signaling [56]

II. Materials and Methods

1. Plasmid constructs

PenA plasmids used five kinds including PenA-WT, tandem repeat mutants (TR10 and TR11), and point mutants (N136D and C69Y). These plasmids cloned into protein expression vector pET28a (+) from *B.thailandensis* were provided by lab of the professor Kim (Korea University, Seoul, Korea)

CagL plasmid cloned into pET28a (+) from *H.pylori* K74 strain was provided by lab of the professor Cha (Yonsei University, Seoul, Korea)

2. Site-directed mutagenesis of CagL

Site-directed mutagenesis was performed using CagL plasmid that was provided as template. To generate CagL mutants, PCR was reacted with primers described in table 4 by using Muta-Direct Site-Directed Mutagenesis Kit (iNtRON, biotech). Site-mutated DNA was identified by sequencing analysis (Macrogen, Korea) and used for protein purification.

3. Transformation and protein overexpression

Plasmids of pET28a (+)-PenAs (WT, TR10, TR11, N136D and C69Y) were transformed into *E.coli* BL21 (DE3) on solid LB media plate containing 100 µg/ml kanamycin for recombinant protein expression. The transformed cells grown in 1.5 L LB media at 37°C (200 rpm) with 100 µg/ml kanamycin until the OD₆₀₀ reached 0.8. And then, the grown cells were overexpressed protein during 4 hours (37°C, 200 rpm) or overnight (18°C, 200 rpm) with 0.5 mM isopropyl-β-D thiogalactopyranoside (IPTG). Overexpressed cells were harvested by centrifugation (5000 g, 20 min, and 4°C) and restored pellet at -20°C.

Plasmids of k74 strain CagL-WT and site-directed mutants were transformed into *E.coli* Rosetta2 on solid LB media containing 100 µg/ml kanamycin and 50 µg/ml chloramphenicol. Transformed cells grown in 1.5L LB medium with kanamycin (100 µg/ml) and

chloramphenicol (50 µg/ml) at 37°C (200 rpm) until OD₆₀₀ reached 0.8. Recombinant protein of CagL was induced by the addition to 1mM IPTG with 1% MeOH at 18°C (200 rpm) for overnight. Cells were collected by centrifugation (5000 x g, 20 min, and 4°C) and stored at -20°C.

Selenomethionine-incorporated CagL used for substitution of methionine according to protocol described by Guerrero, et al [57]. Transformed cells were grown in 1L LB to an OD₆₀₀ 0.5-0.8 and harvested. Collected cells were resuspended and washed by sterilized water in order to remove the residual LB media. After removal of sterilized water by centrifugation, cells were suspended and cultured in 1L freshly prepared Se-Met Minimal Medium (1g NH₄Cl, 3g KH₂PO₄, 6g Na₂HPO₄·7H₂O, 2% w/v glucose, 0.03% w/v MgSO₄, 10 mg thiamine, 10 mg Fe₂(SO₄)₃, pH 7.4, 100 µg/ml kanamycin, and 50 mg L-seleno-methionine) with 1 mM IPTG at 20°C for overnight.

4. Protein purification by Ni-NTA agarose

The frozen cells (pellet to 1.5 L cultured cell) of two kinds (PenA and CagL) were slowly thawed and resuspended in 40 mL lysis buffer containing 20 mM Tris pH 7.5, 500 mM NaCl, 10 mM Imidazol, 1 mM PMSF, 1 mM DTT, 1 mM MgCl₂, and a little DNase 1. Resuspended cells were disrupted by sonication at 25% amplitude and pulse (20 sec, 20 sec) for 5 minute on ice and the disrupted cell was removed by centrifugation at 4°C (2000 x g, 30 min). The acquired supernatant was passed at 1 ml/min by econo-pump into column (Bio-Rad) packed with Ni-NTA agarose (Quiagen, USA). After binding, column was washed to 2 ml/min by washing buffer (20 mM Tris pH 7.5, 500 mM NaCl and 20 mM imidazole) and 6 x His-tagged protein was eluted by elution buffer (20 mM Tris pH 7.5, 500 mM NaCl, and 250 mM NaCl). Eluted protein was verified by color change with 1 x bradford solution

5. Dialysis and 6 x His-tag cutting by human α -thrombin

The eluted soluble protein was dialyzed by using a dialysis membrane MWCO 10 KD (Spectrum Laboratories, Inc. USA) in buffer (20 mM Tris pH 7.5, 150 mM NaCl) at 4°C for overnight. The human α -thrombin (HTI, USA) used at 5000 x dilution with 2 mM CaCl_2 for cleavage reaction with 6 x His-tagged protein at RT for 30 min-2 hours.

6. Ion exchange and size exclusion chromatography (SEC)

For cation exchange chromatography of PenA, protein was dialyzed to low salt buffer (20 mM Tris pH 7.5, 20mM NaCl) by using same membrane (MWCO 10 KD) for overnight. Dialyzed protein was subsequently injected to onto pre-equilibrated cation exchange column MonoS (GE healthcare, USA) and eluted to a linear gradient of 20 mM-1M NaCl by using fast protein liquid chromatography (FPLC). Eluted proteins at major peaks were collected and further purified by SEC pre-equilibrated (20 mM Tris pH 7.5, 50 mM NaCl) Hi-Load 16/60 Superdex 200 column (GE healthcare, USA) for crystallization

CagL directly performed by SEC pre-equilibrated final buffer (20 mM Tris pH 7.5, 50 mM NaCl) and collected at major peak. Fractionized peak region was identified by 12% SDS-PAGE.

7. Crystallization

The purified proteins were concentrated to 15 mg/ml in 20 mM Tris pH 7.5 and 50 mM NaCl by using vivaspin 10K (Satorius, Germany). In order to crystallize proteins, the 1 μ l of each protein mixed with 1 μ l of well solutions was crystallized by hanging drop method on 24well VDX plate (Hampton Research, USA) by using Hampton Research kits (Index, SaltRx, PEG/Ion, PEG/Ion 2, Crystal Screen, Crystal Screen 2, and Crystal Screen Lite. USA), Emerald Biosystems kits (Wizard 1, 2, 3 and 4. USA) and Home-made solution for screening at 20°C and 4°C.

PenA-WT crystals formed from four condition containing three Hampton solution (Index

70: 0.2M NaCl, 0.1M Tris pH 8.5, 25% PEG 3350, Index 90: 0.2M Na-Formate, 20% w/v PEG 3350, and SaltRX 2-45: 0.1 M Bis-Tris Propane pH 7.0, 4 M Ammonium acetate) and one Wizard 3-42 (0.1M Tris pH 8.5, 0.1M Li-sulfate, 30% PEG 4000). Improved PenA-WT crystals induced from condition containing 0.1 M Tris pH 8.5, 0.1 M Li-sulfate, 30% (w/v) PEG 4000, and 15% glycerol at 20°C. Crystals of PenA-TR10 and PenA-TR11 initially formed from index 95 (Hampton) at 20°C and 4°C. Improvement of PenA-TR10 revealed a well solution containing 30% (w/v) PEGMME 2000, 0.1 M K-thiocyanate, and 10% glycerol at 4°C after a week. Improvement of PenA-TR11 also formed from a well solution containing 0.1 M Tris pH 7.5, 30 % (w/v) PEGMME 2000, and 0.1M K-thiocyanate within 3 days at 20°C. PenA-N136D and PenA-C69Y appeared from condition containing 0.1 M Na-acetate pH 5.0, 25% PEG 3350, and 0.2 M NaCl at 20°C and index 41 (0.2 M Sodium acetate trihydrate and 20% w/v PEG 3350) at 4°C, respectively.

Se-Met CagL crystals appeared to Crystal Screen Lite No.40 of Hampton solution (0.1 M Na-citrate pH 5.6, 10% iso-propanol, and 10% PEG 4000) after 3-4 days and grown to maximum size until a week. To more improvement of crystals, Se-Met and native CagL protein investigated improvement of crystals by the addition of 96 additive screening solution (Hampton, USA) with initial solution (Crystal Screen Lite No.40) and the best crystal from additive No.81 (3% dioxane) was obtained.

8. X-ray crystallographic data collection, structure determination, and refinement

The PenA crystals were soaked in cryo-protectant (mother-liquor with glycerol 10-20%) for 30 seconds and were flash-frozen in liquid nitrogen. The X-ray diffraction data were collected at a 1.5 Å resolution from PenA-WT on the PAL-7A beam line and at a 1.5 Å, 1.6 Å resolution from PenA-TR10 and PenA-TR11 on the PAL-5C beam line at the Pohang Accelerator Laboratory (Pohang, Republic of Korea). To obtain crystals of PenA-TR10 and PenA-TR11 that were formed complex with CBA (ceftazidime-like glycyboronic acid), both

crystals were soaked with CBA (2 mM) in a cryo-protectant containing 5% dimethylsulfoxide, 30% (w/v) PEGMME 2000, 0.1 M K-thiocyanate, and 10% glycerol for overnight at 4°C, and the crystals were then flash-frozen. Crystals of PenA-TR10 and PenA-TR11 soaked with CBA were diffracted to 1.6 Å and 1.7 Å on the PAL-5C beam line, respectively. PenA-N136D and PenA-C69Y also obtained data at 1.4 Å and 1.3 Å resolution on the PAL5C. Crystals of PenA-N136D and PenA-C69Y with 2 mM CBA in a cryo-protectant also were diffracted at 1.4 Å resolution on the PAL7A and PAL5C, respectively. All data was processed with the program HKL2000 (HKL Research, Inc.). Results of processing are shown in Table 1, 2.

Model building and refinement were performed with the COOT [58] and REFMAC5 program in the CCP4 suite [59], and figures prepared with the PyMol program. Molecular replacement result for the PenA-WT were obtained with the program MOLREP (R-factor: 34.9%) using CTX-M-9 (PDB code 1YLJ) as a search model without the water and ions. The well-refined PenA-WT PDB was used as an initial structure for determination for PenA mutants as a search model. In particular, the CBA PDB file was created using the PHENIX program LIGANDFIT [60] and fitted into Fo-Fc density that was consistent with the ligand during the building of PenA-CBA in COOT. The refinements for all of the models that could be expected at those resolutions converged to reasonable values using the PHENIX program (Table 1, 2).

Selenomethionine substituted CagL was also soaked in cryo-protectant (well solution with 20% glycerol) and were frozen in liquid nitrogen. Crystal of Se-Met CagL diffracted at 2.9 Å resolution on the PAL-5C beam line with single wavelength 0.97943. The oscillation ranges were 360° with oscillations per frame being 1° and 10 second exposure time and distance measured at 450 mm. The result of indexing, integration and scaling by HKL2000, Se-Met CagL was belonged to the tetragonal a space group of P4₃2₁2 (Table 3).

Initial structure of Se-met CagL was obtained from the anomalous signal of a Se-Met substituted crystal by using the single anomalous dispersion (SAD) method. An absorption

peak wavelength of Se atoms was 0.97943 and identified by energy scanning. Initial model of CagL was performed by autosol in PHENIX Program for SAD phasing. Native data of CagL also obtained at 2.8 Å and initial structure was determined by using Se-Met CagL structure as a search model. Further model building and refinement is completely performed by Coot and PHENIX (Table 3).

9. Analytical size exclusion chromatography

To analyze molecular weight of CagL, analytical size exclusion chromatography using HiLoad 16/600 Superdex 200 pg (GE Healthcare) was used. Purified CagL and mutants (~2 mg) were passed at a flow rate of 1 ml/min into pre-equilibrated column in 50 mM Phosphate, pH 7.2 and 150 mM NaCl. A mixture (1.5 ~ 2 mg/each protein in total 0.5 ml) of standard proteins (GE Healthcare LMW calibration kit) including conalbumin (75 kDa), Ovalbumin (44 kDa), carbonic anhydrase (29 kDa), Ribonuclease A (13.7 kDa), and Aprotinin (6.5 kDa) are prepared and used for a standard curve.

10. Sedimentation equilibrium of CagL using analytical ultracentrifugation (AUC)

Sedimentation equilibrium experiments were performed in analytical ultracentrifuge equipped with absorbance optics and an An60Ti rotor in a Beckman XL-A (Germany). Purified CagL protein was used to 24 μM in buffer containing 20 mM Tris, 50 mM NaCl, and pH 7.5. This study was carried at 25000 rpm during 72 hour at 4°C. Absorbance profiles were acquired at 280 nm between radius 6.75 and 7.15cm. Molecular weight of CagL protein was calculated by analysis using raw data files with program SEDPHAT ver.10.58d.

III. Results

Part I

Structural study of Class A β -Lactamase PenA in *Burkholderia thailandensis*

1. Purification of recombinant proteins to PenA-WT and mutants

Recombinant proteins of PenA-WT and mutants (TR10, TR11, N136D, and C69Y) cloned in pET28a (+) were overexpressed in soluble form by using *E. coli* BL21 (DE3) for protein expression. Each protein purified by using Ni-NTA agarose resin. 6 x His-tagged fragments of eluted proteins were removed by treatment of human α -thrombin with 2mM CaCl_2 . To obtain high purity of protein, cationic exchange and size exclusion chromatography were performed as describe in materials and methods (Fig. 3B). Final buffer was changed to 20 mM Tris pH 7.5 and 50 mM NaCl for crystallization. Size of each protein was analyzed by 12% SDS-PAGE (PenA-WT: 28 kDa, TR10: 29 kDa, TR11: 30 kDa, N136D: 28 kDa, and C69Y: 28 kDa) (Fig. 3A).

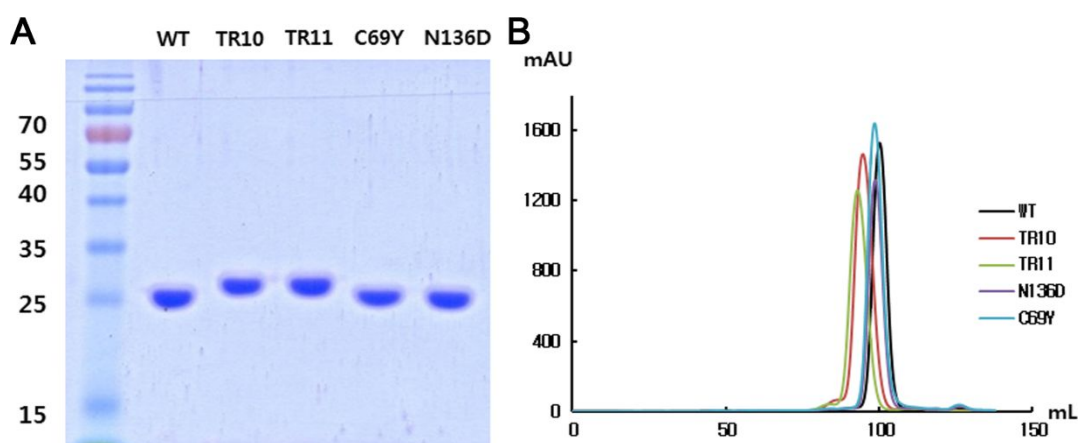


Figure 3. Result of SEC and analysis by SDS-PAGE to PenA proteins (WT, TR10, TR11, N136D, and C69Y)

A. 5 kinds of PenA were conducted by gel filtration using Hi-Load 16/600 Superdex 200 column. Each protein represents to different color line; black (WT), brown (TR10), green (TR11), purple (N136D), blue (C69Y). B. Purified final PenA proteins was analyzed by electrophoresis with 12% SDS-PAGE

2. Preliminary crystallographic results of PenAs

To crystallize proteins of PenA-WT and mutants, each protein was concentrated until ~10 mg/ml and used for crystallization by using commercial screening solution. After 3-5 days, initial crystals of each protein appeared from some of the condition and grown to big size in a week. PenA-WT appeared to very thin rod-like shape (Fig. 4A-a) and plate shape at four conditions including three of Hampton (Index 70, 90, and SaltRX2-45) and one of Wizard (3-42) as described in materials and methods. Crystal of PenA-WT was further improved by condition containing 0.1 M Tris pH 8.5, 0.1 M Li-sulfate, 30% PEG 4000, 15% glycerol at 20 °C (setup condition from Wizard 3-42) and x-ray diffraction data was obtained as described in Table 1. Initial crystals of PenA-TR10 and PenA-TR11, which revealed to transparent stick shape and cluster of sticks (Fig. 4A-b, c), formed from Index 95 (30% PEGMME 2000, 0.1 M K-thiocyanate) at 4°C and 20°C, respectively. Crystals were subjected to x-rays and the diffraction data from setup condition containing 10-15% glycerol with initial condition (Index 95) was collected. PenA-N136D also show stick shape and was obtained from 0.1 M Na-acetate, pH 5.0, 25% PEG 3350, and 0.2 M NaCl at 20°C. PenA-C69Y obtained crystals of rod-like shape from PEG/Ion 27 (0.2 M Sodium acetate trihydrate and 20% w/v PEG 3350) at 4°C (Fig. 4A-d, e).

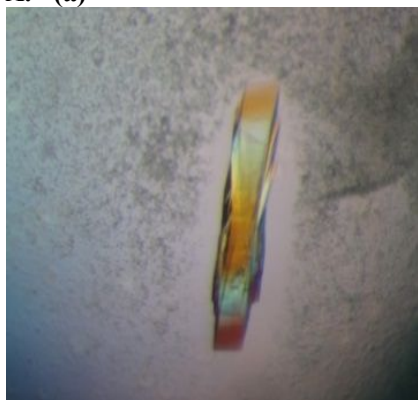
To obtain x-ray crystallographic data to PenA-WT and mutants, each crystal soaked by 10-20% glycerol containing well solution and was flash-frozen in liquid nitrogen. Each PenA crystals were diffracted at high resolution (WT: 1.5Å, TR10: 1.5Å, TR10-CBA: 1.6Å, TR11: 1.6Å, TR11-CBA: 1.7Å, N136D: 1.4Å, N136D-CBA: 1.4Å, C69Y: 1.3Å, and C69Y-CBA: 1.4Å). Diffracted images to each crystal are shown in Fig. 4B.

As a result of processing, space group of PenA-WT belongs to orthorhombic $P2_12_12$ and contains one molecule in asymmetry unit. PenA-TR10 and complex with CBA belonged to different space that TR10 is monoclinic $P2_1$ containing two molecules in an asymmetry unit and complex with CBA is orthorhombic $P2_12_12$ containing one molecule. PenA-TR11 and

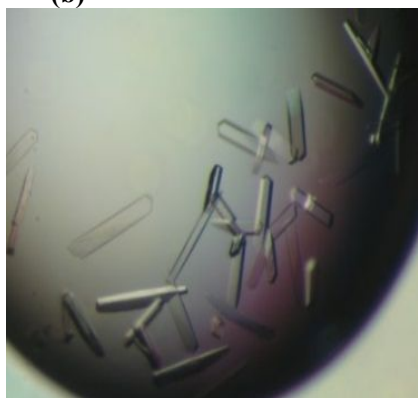
complex with CBA belong to same orthorhombic $P2_12_12$ containing one molecule in an asymmetry unit. N136D and complex with CBA belong to monoclinic $P2_1$ containing two molecules in an asymmetry unit. C69Y and complex with CBA belong to orthorhombic $P2_12_12_1$ containing one molecule in an asymmetry unit. Space group and molecule number to each crystal are different, suggesting that there might be a difference of crystal packing by a variety of conditions in solution. The initial structures of PenA-WT were solved by molecular replacement using CTX-M-9 (PDB: 1YLJ) as a search model and resolved to R_{work} (34%)/ R_{free} (36%) by Refmac5. The final structures were completed to R_{work} (15.8%)/ R_{free} (18%) of by iterative cycles of refinement and model building.

Initial structure of TR mutants (TR10 and TR11) and point mutants (N136D and C69Y) resolved using PenA-WT PDB as a search model, and then, complete structure was determined by Coot and PHENIX program. Detailed data collection and refinement of PenA-WT and mutants described on Table 1, 2.

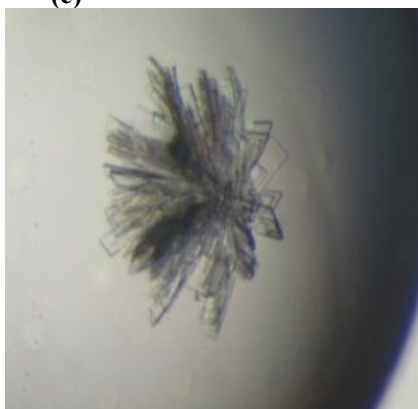
A. (a)



(b)

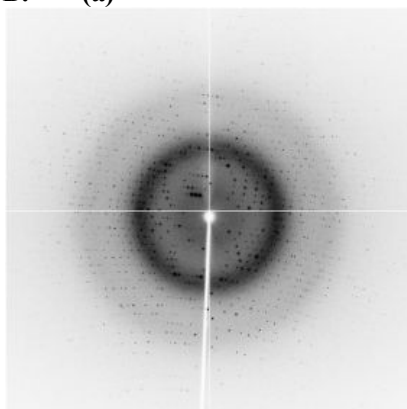


(c)

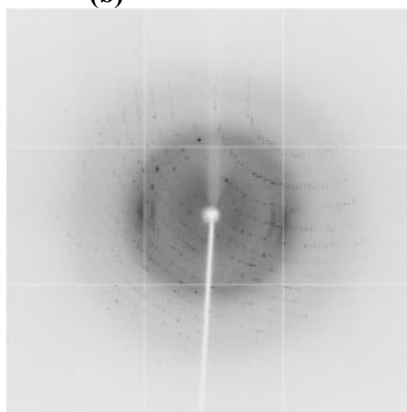


(d)

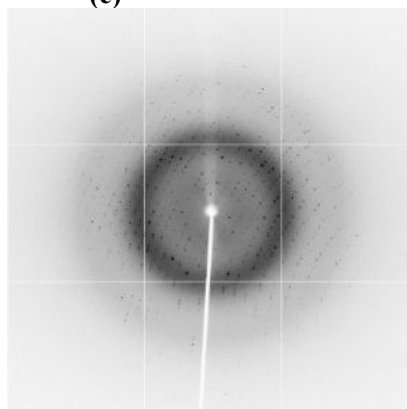
B. (a)



(b)



(c)



(d)

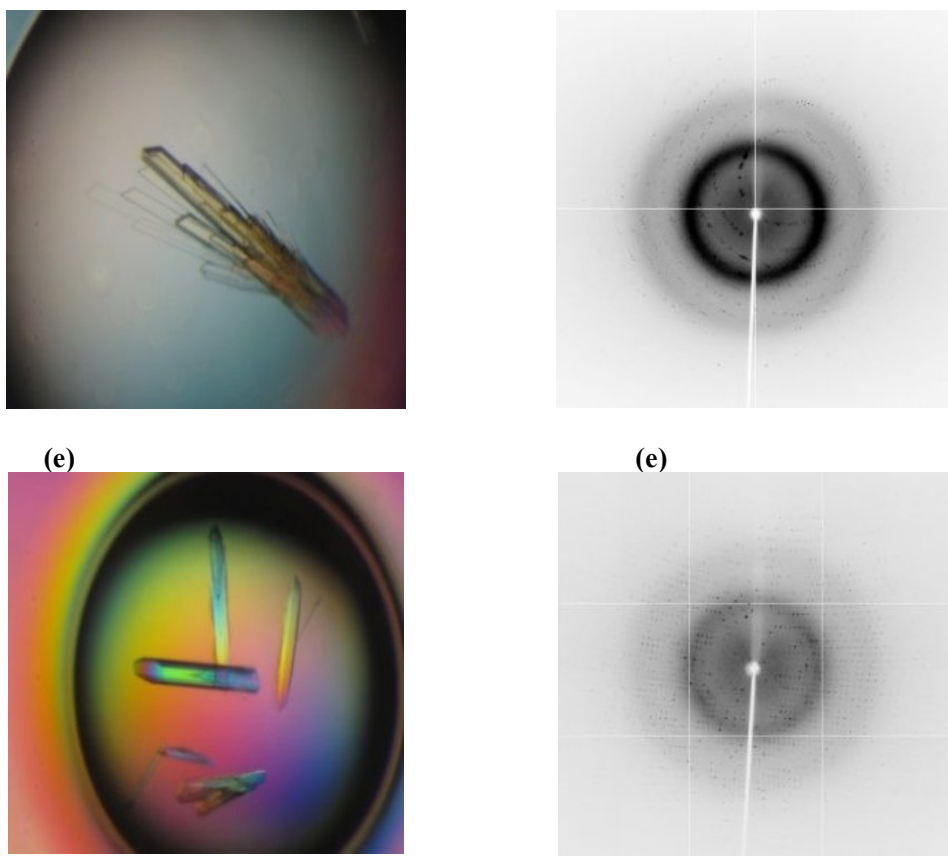


Figure 4. Crystal and x-ray diffraction images of PenA-(a) WT, (b) TR10, (c) TR11, (d) N136D and (e) C69Y.

A. (a) ~ (e) Crystals of PenAs completely grew by hanging drop on well solution condition in 1-2 weeks as described in materials and methods.

B. Crystals of PenA-WT(a) and PenA-TR10 (b) diffracted by parameters including distance (160 mm), exposure time (1 sec), ω (1°) and 360° . (C) Crystal of PenA-TR11 diffracted by same parameters as PenA-WT except 200 mm distance. (d) Crystal of PenA-N136D diffracted by parameters (distance: 150 mm, exposure time: 1 sec, ω : 1° , total 300). (e) Crystal of PenA-C69Y diffracted by parameters (distance: 150 mm, exposure time: 1 sec, ω : 1° , total 180).

3. Overall structure of PenA-WT and comparison with other β -lactamases

The overall fold of PenA-WT resembles with those of other class A β -lactamase, which presents α/β and α domain to an active site containing conserved residues to cooperate substrate recognition and catalytic activity [28] (Fig. 5A). Calculated RMSD value of C α position between PenA and other class A β -lactamases including SHV-1, TEM-1, and CTX-M-9 showed the structural similarities with a range of 1.5 to 2.4 Å. Especially, the structures of an ESBL, PenI from *B. pseudomallei* and a carbapenemase, PenA from *B. cepacia complex* have lower RMSD values (0.36Å, 0.56Å) with our PenA than those representative β -lactamases, meaning that our PenA has higher structural similarity to those from same genus *Burkholderia*. Furthermore, the residues in catalytic cavities between PenA-WT, SHV-1, TEM-1, and CTX-M-9 seem to superimpose well (Fig. 5B). In general, class A β -lactamases form active site by five regions [7]; 1) the 70SXXK73 motif, 2) in the SDN loop (130SDN132), 3) the 234KTG236 on β 3 strand, 4) Ω loop (164-179) and 5) the conserved Tyr105 residues in 102-110 loop.

In the previous studies, β -lactamases including SHV-1, TEM-1, and CTX-M weakly hydrolyze third generation cephalosporin ceftazidime [21]. However, the PenA-WT do not hydrolyze cephalosporin antibiotic such as ceftazidime, suggesting that active site is likely to have some of structural difference for substrate binding. Although essential residues involved in the catalytic activity among the class A β -lactamases are highly conserved, PenA-WT structure revealed two major differences in the active site; new interaction between residue Arg104 and Thr167 and positions of water molecules.

Firstly, the interaction between Arg104 and Thr167 may provide a force restricting the substrate spectrum of PenA to Penicillins compared with other Class A β -lactamases such as SHV-1 and TEM-1. Asp, Glu, or Lys falls in the position 104 with Thr or Pro in the position of 167 in most of the members, which produce no such significant interaction between loop 102-110 and upper Ω loop (164-179) (Fig. 6A). This lack of the interaction in turn releases the

Tyr105 further way from substrate binding cleft, which might increase the cavity size and subsequently let those enzymes be able to accommodate larger substrate like ceftazidime. However, the formation of the interaction in PenA-WT bringing Tyr105 down to narrow the cavity size with respect to the size of other members since the interaction tends to hold the loops tighter, implying not only more rigid nature of the 102-110 loop via the interaction limits the substrate spectrum of PenA, but breakdown of the interaction by some ways also possibly induces substrate spectrum or specificity changes to some extent.

Secondly, our structure of PenA-WT active site revealed one extra water molecule (called W5) that has not been found in any other structure of Class A β -lactamases so far. Usually, 3-4 water molecules always including catalytic water (W1) form hydrogen bond to each other and connect surrounding residues (Ser70, Asn132, Glu166, Asn170, and Thr237) (Fig. 6B). Among the water molecules, the catalytic water (W1) is known to be key water for catalysis and is responsible for the deacylation of substrate [8]. The new water molecule (W5) in our structure is located closer to the surface of the cavity but further from W1 than other water molecules. However, W5 absolutely connects to W1 by indirect hydrogen bonding with W3, which in turn interact with W1. W5 is also likely to be stabilized in WT active site architecture by interaction with two conserved residues, Asn132 in SDN loop and Asn170 in Ω loop (Fig. 6B). W5 has not been observed from active sites of broader substrate spectrum enzymes including SHV-1 and CTX-M-9 whose catalytic cavities are wider opened. Moreover, W5 provides additional interaction with CBA in the insertion mutants making new interactions with substrate, indicating that it keeps the cavity integrity. From these observations, W5 is likely to reside in the position to stabilize the compact conformation of the active site and maintain a strict substrate specificity comparing with other members. Together, those two new structural characters may play important roles in bypassing clinical trials by the pathogen and can be used as a good model for study of extended spectrum to third generation cephalosporin such as ceftazidime.

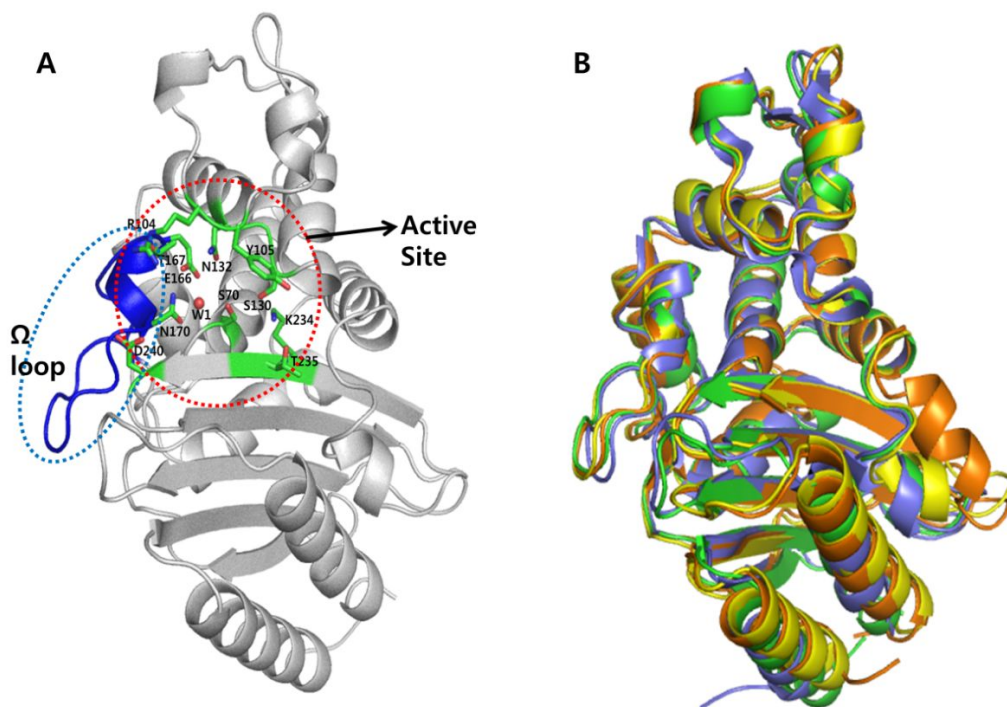


Figure 5. Overall structure of PenA-WT and structural alignment of PenA-WT with other β -lactamases (TEM-1, SHV-1, and CTX-M-9).

A. Blue and red dash lines represent Ω loop region and active site with conserved residues and catalytic water, respectively. **B.** Alignment of four kind enzyme reveals to different color cartoon (WT: green, SHV-1: orange, TEM-1: yellow, CTX-M-9: slate)

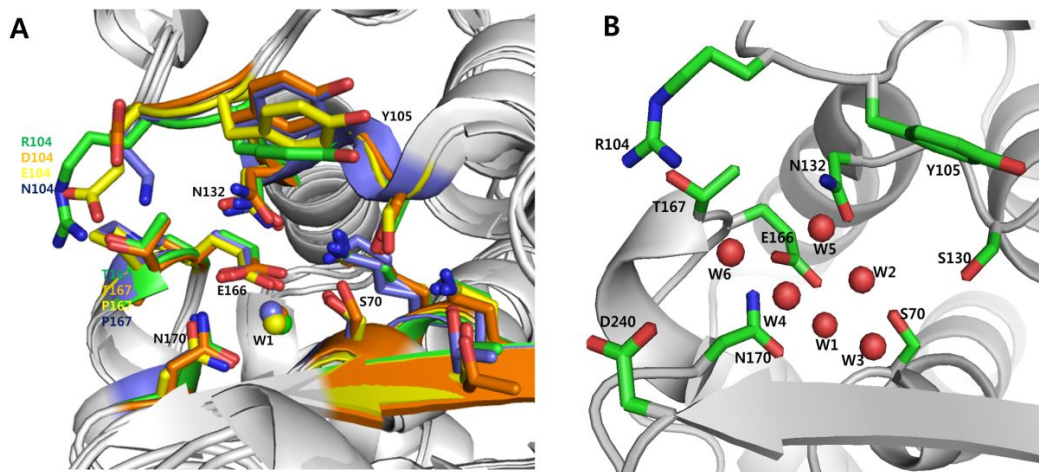


Figure 6. Structural alignment to conserved residues and catalytic water with other β -lactamases, and water-mediated hydrogen bond network forming active site of PenA-WT.

A. Residues and catalytic water to each lactamases represent as different color stick (green: PenA-WT, orange: SHV-1, Yellow: TEM-1, slate: CTX-M-9). B. W1: catalytic water. W5 only contains unique water molecule for tight narrow cavity located on surface of PenA-WT. W2-W4 and W6 mediate hydrogen bond network in active site.

4. Reconstitution of the functional omega loop by insertion of repetitive residues in PenA-TR10 and PenA-TR11

The structures for PenA-TR10 and PenA-TR11 had a highly disordered region, where mapping of the residues was not practicable (Fig. 7). These regions in each structure included the entire first repeat unit and the first residue in the second unit, and in the case of TR11, which has a longer repeat, nine additional residues in the upstream were also disordered (Fig. 8). The rest of the TRs in each structure, which is the second repeat unit except for the first residue, replaced the dislocated residues. It is reasonable to postulate that other duplications may be reformed in the similar way by extruding the extra residues from the omega loop region. It is remarkable that this stress-absorbing capacity significantly increases the chance to take diverse mutations, and thereby a variety of changes in the active site.

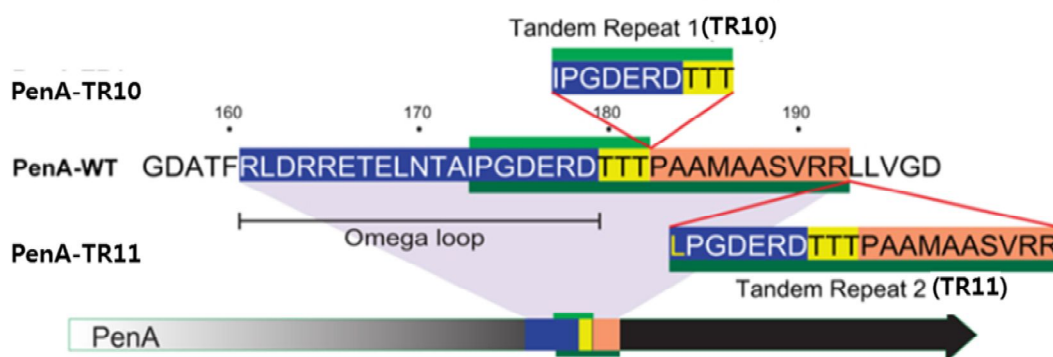


Figure 7. Representation of tandem repeat inserted on Ω loop in PenA-TR10 and PenA-TR11 [27]. Number of amino acid inserted TR10 and TR11 are 10 and 20 amino acids. TR10 was inserted to repetitive 10 amino acid between Thr182 and Pro183. TR11 was also inserted to repetitive 20 amino acid except for first amino acid I→L between Arg192 and Leu193.

5. The increase of binding cavity by insertion of repetitive residues in PenA-TR10 and PenA-TR11

The reorganization of the omega loop affected the architecture in the active site. One of the most notable changes by insertion of tandem repeat was the disruption of a short α -helix (Glu166-Asn170) in the omega loop (Fig. 8B). In the structure of PenA-TR10, the disruption of the α -helix led to the weakness of a hydrogen bond that is normally present between Thr167 and Arg104, which tightly packs the active site cavity (Fig. 8A, B). As a result, the active site cavity was expanded, while the side chains of Asn170 and Asp240 were redirected to point outward (Fig. 8A, B). Asn170 exhibited a particularly two movement, swinging away 180° from the active site cavity by 9.63 \AA and rotation of side chain (Fig 8A, B). Consequently, the salt bridge between Asn170 and Glu166 and the hydrogen bond between the residue and the catalytic water were disrupted, further widening and destabilizing the active site cavity. As a result of the increased instability in the cavity, another water molecule entered and occupied the expanded space, thereby reestablishing the hydrogen bond network (Fig. 8C). Asn170 and Asp240 represent the omega loop and the $\beta 3$ strand structure, respectively, which are both located in the wall of the active site cavity. In the structure of PenA-TR11, residues Glu166, Thr167, and Asn170 were not mapped (Fig. 8C), suggesting that the salt bridges involving these residues were also disrupted. The Asp240 in both PenA-TR10 and PenA-TR11 was closely superimposed on top of that in PenA-WT despite its reorientation, demonstrating that dislocation of the omega loop was the primary cause for the widened active site cavity.

A few studies have shown that a conserved residue Tyr105 is involved in substrate recognition and stabilization of the enzyme-substrate complex through the planar hydroxyl benzyl ring, which stacks with the thiazole ring in cephalosporins [11]. The structure of PenA-TR10 showed that Tyr105 was pushed away from the active site cavity by 1.17 \AA relative to the corresponding position in PenA-WT. Tyr105 was reoriented, which allowed for better substrate access to the active site cavity (Fig. 8A). Such an alteration of Tyr105 appeared to be caused by

the movement of Arg104, which fell away from Thr167 in the omega loop upon disruption of the hydrogen bond between them (Fig. 8A).

The W5 looks like it has very important role for maintenance of stable cavity by interaction with both Asn132 (2.99 Å) and Asn170 (3.92 Å). However, moving or removal of W2 by broken interaction with Asn170 by TR insertion can be unstable and cause increase of cavity size.

The Ω loop of PenA-TR11, as a result of severe disruption of omega loop by TR duplication, disordered to the wide range than that of PenA-TR10 like above mentioned. Change of structural conformation of PenA-TR11 revealed a similar with that of PenA-TR10. That is, the broken interaction between Arg104 and Thr167, movement of Arg104 to right direction, rising of Tyr105 (1.0 Å), and the bend of Asp240 to downward are quite similar to change of penA-TR10. Along with these changes, the catalytic water is moved to outside (1.49 Å) and the new water comes into near Asn170 and forms interaction with catalytic water (Fig. 8C).

These changes of Ω loop structure in each structure are likely to characterize a bulky substrate accessing like ESB� through expansion of catalytic cavity for hydrolytic activity.

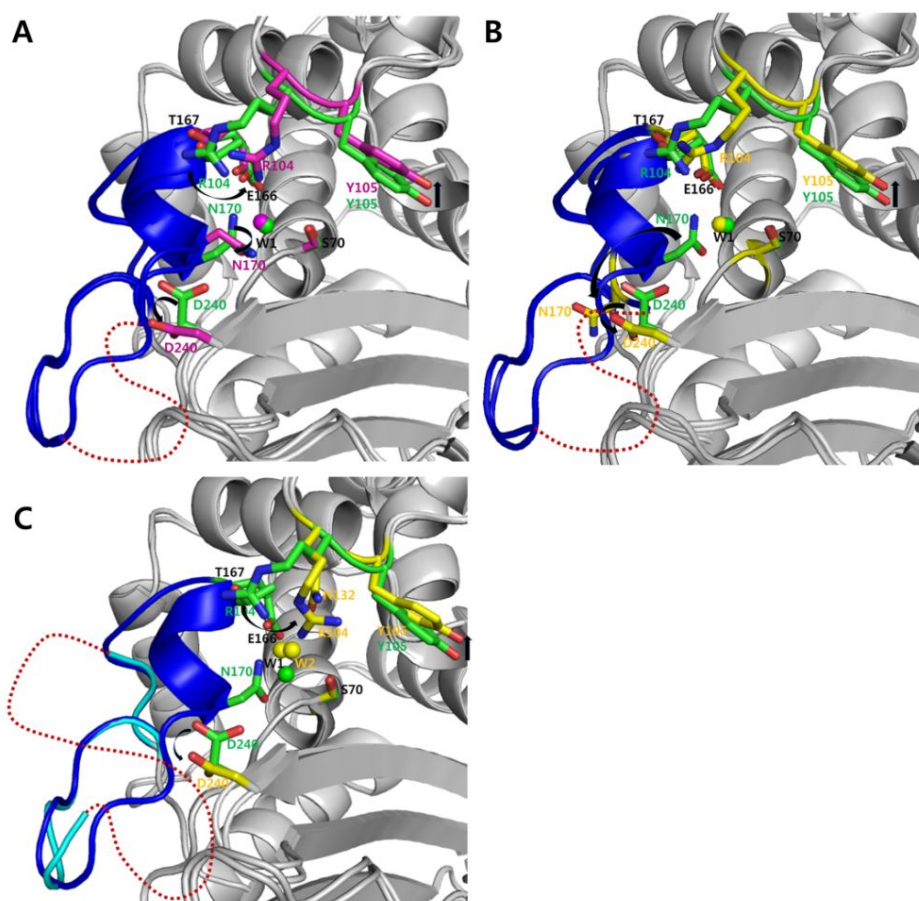


Figure 8. Conformation change of PenA-TR10 and PenA-TR11 compared to PenA-WT (Red dash line shows disorder region of Ω loop). A and B are represent two forms by the inserted amino acid in TR10. A form shows that Asn170 is rotated and interaction between Arg104 and Thr167 is broken (WT: green, TR10A: magenta). B form can show dynamic movement of Asn170 to outside by 180° swing out. Interaction of Arg104 with Thr167 also is weak and Asp240 is rotated to outside (WT: green, TR10B: yellow). C represents conformation change of active site in TR11. TR11 also is disordered more wide region than TR10. TR11 cannot show Asn170 (WT: green, TR11: yellow).

6. Binding of a ceftazidime analog induces regrouping of the active site.

To determine how binding of a substrate to PenA-TRs affects the active site architecture, Structures of PenA-TR10 and PenA-TR11 fitted with ceftazidime-like glycyI boronic acid (CBA) are analyzed. In the structure of PenA-TR10, when CBA occupied the active site cavity, Arg104 was pushed further apart (approximately 5.22 Å) from Thr167, and consequently, Tyr105 moved away farther (approximately 2.24 Å), stabilizing and locking CBA in position via thiazole ring stacking (Fig. 9A). Restoration of the α -helix simultaneously occurred in the omega loop, repositioning Asn170 and Glu166 and thereby reestablishing the interaction between the two residues through catalytic water (Fig. 9A). The Asp240 also restored to original position. All of these interconnected conformational changes apparently resulted in induced-fit docking between PenA-TRs and CBA. The CBA interacts with some residues of active site for catalytic activity in similar to inhibitor binding in SHV-1. O¹² of CBA maintain stably to planar interact with Asn132 (2.92 Å) and newly input water (2.66 Å), and O^{b2} of CBA interact with Ser130 (2.86 Å), and O^{b1} of CBA interact with N and Thr237 (2.87 Å, 2.74 Å) (Fig. 9A). Three water molecules interacted with Ser70 and are replaced by covalent bond of CBA and new water formed hydrogen bond with O¹² of CBA and Asn170 (3.79 Å) for stable conformation.

In contrast with PenA-TR10-CBA, The CBA in PenA-TR11 occupy space unlike PenA-TR10-CBA in respect to position and stability, in that the bulky C7 β side chain of CBA interact with some residues more flexible (B factor=42.6 Å²) than PenA-TR10-CBA (B factor=18.7 Å²) near short helix of omega loop of PenA-WT. When the CBA come into catalytic cavity, the Tyr105 position firstly moves up (0.78 Å) than PenA-TR11, and five water molecules for hydrogen bond network removed and CBA occupy to short helix position of Ω loop (Fig. 9B). Secondly, although the Asp240 is reoriented into for tight packing of CBA at PenA-TR10, it is not recovered to original position after binding of CBA at PenA-TR11. With these conformation change, O¹² of CBA interact with both Asn132 (3.22 Å) and N Arg104 (2.47 Å),

and N¹⁸ of CBA-thiazol ring interact with Asp240 (3.21 Å). Arg104 also interacts with O₁₇ (3.09 Å) besides O¹² of CBA (Fig. 9B). As a result, the CBA in PenA-TR11 occupy to different position unlike PenA-TR10. This fact means that CBA in PenA-TR11 is flexible more than PenA-TR10 and looks like possible change of CBA position in cavity according to extension of flexibility of Ω loop

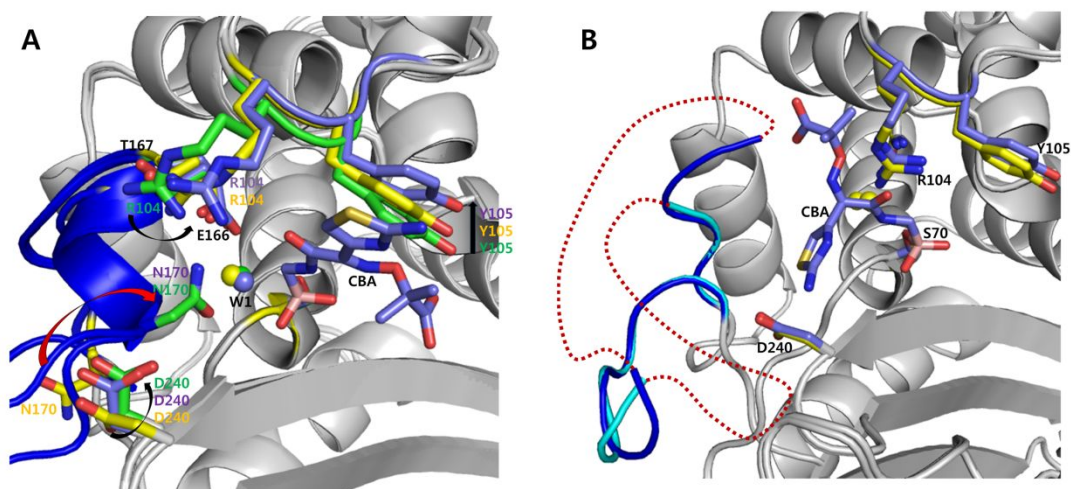


Figure 9. Conformation change in PenA-TR10 and PenA-TR11 by CBA binding.

A. Structures of PenA-WT (green), TR10 (Yellow), and TR10-CBA (slate) are superimposed. Arg104 in TR10 with CBA move to right direction than TR10 and Asn170 and Asp240 return to original position. Thiazole ring of CBA packs with Tyr105 and catalytic water do not change. B. Structures of PenA-TR11 (yellow) and TR11-CBA (slate) also are superimposed. CBA occupied on position of Ω loop short helix with Ser70.

7. Comparison of active site between PenA-WT and PenA-N136D

To investigate conformation change appearing PenA-N136D with PenA-WT, recombination protein of PenA-N136D purified, crystallized, and determined as described in materials and method. Detailed information of N136D data collection and refinement was shown in Table 2

RMSD value of C α position between PenA-WT and N136D is very low value (0.48), indicating that overall structure of N136D resembles with PenA-WT. When PenA-WT superposed on PenA-N136D, one of major difference is change of interaction between Arg104 on 102-110 loop and Thr167 on Ω loop at active site. In PenA-WT, Asp101 and Asn136 maintain conformation by hydrogen bond intermediating water molecule. However, when Asn136 mutated to Asp, position of water moves outside and breaks the hydrogen bond (Fig. 10A). This change results in rising of loop containing Asp101 by loosen interaction and breaking interaction between Arg104 and Thr167. Consequently, these changes suggest that the broken hydrogen bond by water-moving between Asn136 and Asp101 might induce space for change of substrate spectrum.

8. Structural change by CBA binding in PenA-N136D

To identify structural mechanism for ESBL, Crystal of N136D with CBA was diffracted and determined. Final structure of PenA-N136D-CBA also contains two molecules in asymmetry unit. At structure of N136D with CBA, binding of CBA revealed to two conformations (Fig. 10B, C). First conformation (A chain) can show docking conformation by forming of covalent bond of CBA with Ser70 and hydrogen bond with residues including Ser130, Asn132, Glu166, Asn170, Asp240, and Thr237. Catalytic water removed by OH of boron moving to position of catalytic water, meaning that might be in deacylated state by catalytic water (Fig. 10B). Second conformation (B chain) also appears that CBA covalently interact with Ser70 and forms hydrogen bond with residues including Ser130, Asn132, Thr237, and Asp240. Catalytic water maintains interaction with Glu166 and Asn170, meaning that it might be acylated intermediate state by catalytic water (Fig. 10C). Taken together, PenA-N136D observed change of active site that can induce binding of substrate for extended spectrum. Also, PenA-N136D with CBA also identified mechanism that hydrolyze substrates through change of two states such as intermediate of acylation and deacylation state.

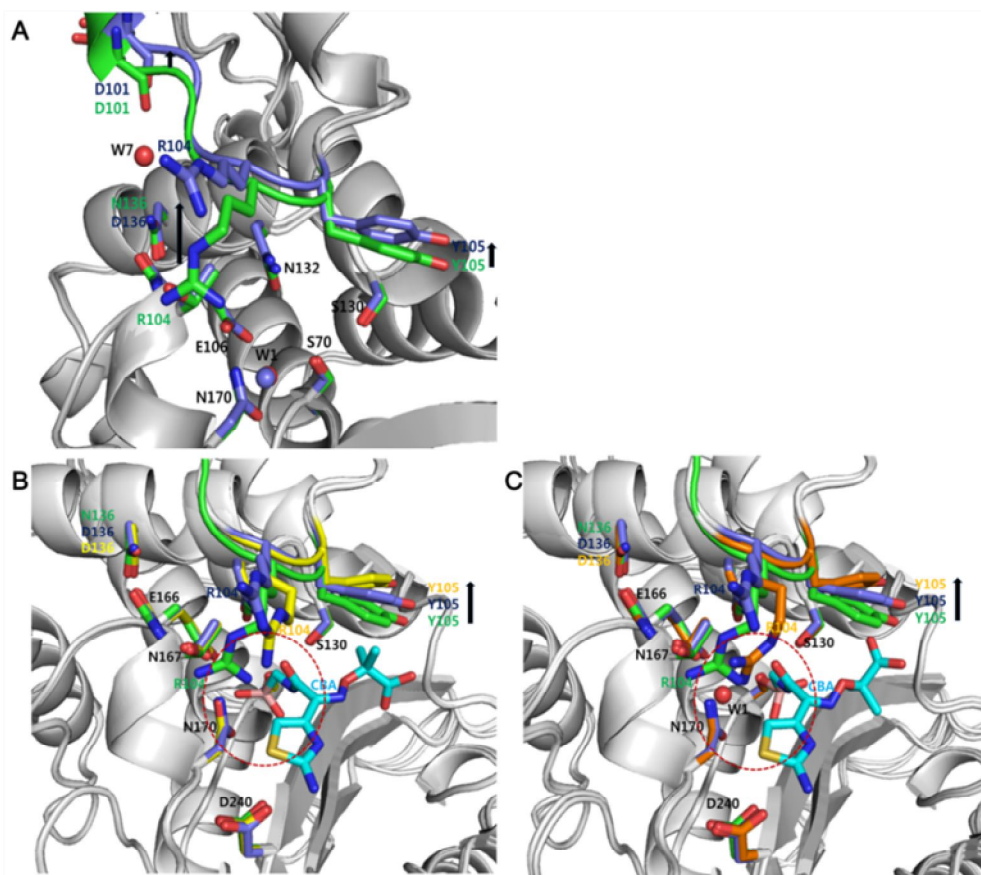


Figure 10. Conformation change of PenA-N136D and CBA-bound PenA-N136D compared to PenA-WT

A. Change of cavity by PenA-N136. Superimposition of PenA-N136D (slate) with PenA-WT (green) can show that removal of water molecule (W7) by Asn→Asp on position 136 is broken water-mediated two hydrogen bond of Asn136-W7-Asp101 on PenA-WT. Result of the broken hydrogen bond induces unstable loop containing 101 to 105 and is up to Arg104 and Tyr105. B is deacylation transition state conformation (yellow) that one of the boronic acid oxygen of CBA boron is occupied instead of catalytic water. C is acylation transition state conformation (orange) containing catalytic water.

9. Comparison of active site between PenA-WT and PenA-C69Y

To investigate conformation change appearing PenA-C69Y with PenA-WT, protein of PenA-C69Y purified, crystallized and determined by previously described method. Data collection and refinement were described in Table 2.

Although RMSD value of Ca position between PenA-WT and C69Y has a high value (1.52 Å) than N136D, overall structure of C69Y resembles with PenA-WT. Change of active site by C69Y can show some differences when compared with WT. Firstly, 45° rotation of Asn170 side chain which interact with catalytic water and Glu166 breaks salt bridge with Glu166, meaning that it induces the loosen triad interaction (Fig. 11A). Secondly, expanded side chain of C69Y induces change of conformation of loop connecting N-terminal of β 3 strand, resulting in the increased cavity for substrate input (Fig. 11A). Thirdly, interaction between Arg104 and Thr167 also broken for binding of extended spectrum (Fig. 11A). These changes suggest that C69Y can induce the increase of cavity for space and input of bulky substrate

10. Structural change by CBA binding in PenA-C69Y

To identify structural mechanism for ESBL, Crystal of C69Y with CBA diffracted at high resolution and initial structure determined. Final structure of PenA-C69Y with CBA belonged to space group $P2_12_12_1$ containing one molecule in the asymmetric unit as shown to Table 2. CBA in active site covalently interacts with Ser70 and forms hydrogen bond with residues including Ser130, Asn132, Thr237, and Asp240 (Fig. 11B). Asn170 reoriented and maintains interaction with Glu166 and catalytic water. Also, Tyr105 is upper for input of CBA by broken force between Arg104 and Thr167, resulting in acylated-intermediate complex. Taken together, PenA-C69Y induces the change of cavity by point-mutated residue for hydrolysis of extended spectrum and complex with CBA also identify mechanism that hydrolyzes substrates.

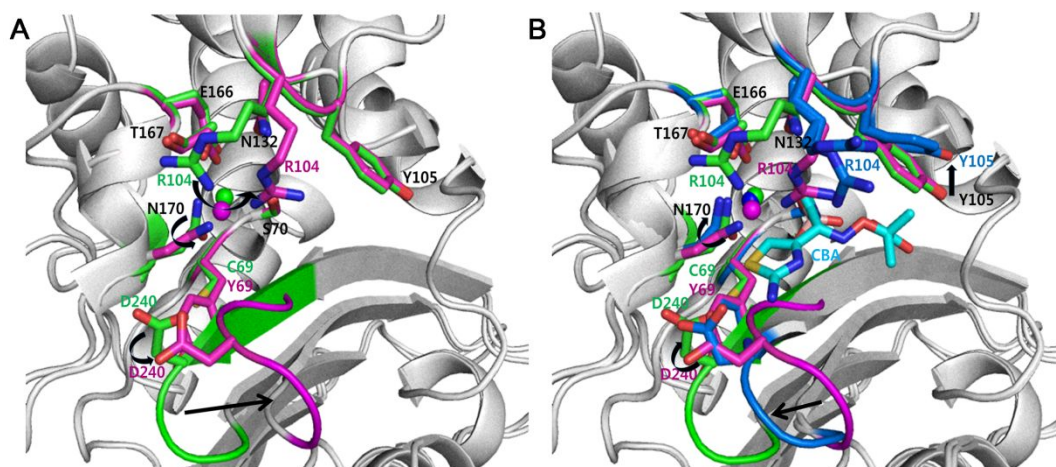


Figure 11. Conformation change of PenA-C69Y and CBA-bound PenA-C69Y compared to PenA-WT

A. Superimposition of PenA-WT (green) and PenA-C69Y (magenta). Arg104 moves in the right direction by the broken interaction with Thr167 and Asn170 which is rotated to $\sim 90^\circ$.

The loop containing Asp240 is moved outward to increase the cavity.

B. Superimposition of Pen-WT (green), PenA-C69Y (magenta), and PenA-C69Y (blue) with CBA (cyan). Asn170 and Asp240 are re-orientated to original direction or position and CBA binds to acylation transition state conformation. Tyr105 is upper than PenA-C69Y.

Table 1. Data collection and refinement statistics of PenAs (WT, TR10, TR10_CBA, TR11, and TR11_CBA)

Parameter	WT	TR10	TR10_CBA	TR11	TR11_CBA
Data collection					
Wavelength (Å)	0.97934	0.97951	0.97951	0.97954	1.00000
Resolution (Å)	50-1.5 (1.53-1.5)	50-1.5 (1.53-1.5)	50-1.6 (1.63-1.6)	50-1.6 (1.63-1.6)	50-1.7 (1.73-1.7)
Space group	P2 ₁ 2 ₁ 2	P2 ₁	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Unit cell (Å)	a=75.933 b=81.637 c=36.834 α=β=γ=90	a=37.031 b=92.551 c=68.866 α=γ=90, β=92.611	a=70.059 b=92.881 c=34.993 α=β=γ=90	a=69.486 b=97.242 c=34.004 α=β=γ=90	a=70.269 b=97.448 c=34.222 α=β=γ=90
Total reflections	535224	363038	286864	442064	205607
Unique reflections	37006 (3611)	69853 (6880)	30933 (3008)	31195 (1532)	26287 (1274)
R _{merge} (%)	5.4 (12.5)	6.1 (20.7)	7.2 (31.5)	7.4 (33.2)	11.4 (30.1)
Completeness (%)	98.73 (97.78)	99.8 (98.89)	99.96 (99.73)	99.9 (100)	99.9 (100)
Redundancy	14.4 (14.5)	5.2 (5.1)	9.3 (9.4)	14.2(14.2)	7.5 (7.6)
Average I/σ (I)	54.47 (31.13)	26 (10.21)	28.96 (9.7)	55.54 (12.26)	28.66 (7.69)
Refinement					
R _{work} / R _{free} (%)	16.26/18.98	16.96/19.72	16.41/19.99	16.03/18.44	16.41/19.99
Residues/Water	268/342	533/569	268/254	255/267	257/277
CBA			1		1
RMSD					
Angle (°)	1.05	1.07	1.17	1.109	1.148
Length (Å)	0.006	0.006	0.008	0.006	0.007
Average B-factors (Å ²)	19.4	14.7	7.1	16.20	24.20
Ramachandran plot					
Most favored regions (%)	98	97.1	98	97.61	97.63
Allowed regions (%)	2	2.51	2	1.99	2.37
Outliers	0	0.39	0	0.40	0

Values in parentheses correspond to highest resolution shell.

†R-merge= $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl) \rangle$ is the average intensity of symmetry-related observations.

Table 2. Data collection and refinement statistics of PenAs (N136D, N136D_CBA, C69Y, C69Y_CBA)

Parameter	N136D	N136D_CBA	C69Y	C69Y_CBA
Data collection				
Wavelength (Å)	0.97934	0.97857	0.97951	0.97954
Resolution (Å)	50-1.4 (1423-1.4)	50-1.4 (1.42-1.4)	50-1.3 (1.32-1.30)	50-1.4 (1.42-1.4)
Space group	P2 ₁	P2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell (Å)	a=34.986 b=92.723 c=68.878 α=γ=90, β=92.719	a=34.883 b=92.393 c=68.841 α=γ=90, β=94.265	a=38.477 b=52.981 c=122.731 α=β=γ=90	a=38.660 b=53.238 c=122.981 α=β=γ=90
Total reflections	647757	516034	413760	618880
Unique reflections	86131 (4254)	84356 (4189)	62299 (3048)	50752 (2479)
R _{merge} (%)	8.0 (27.1)	8.9 (28.7)	5.3 (9.8)	7.8 (25.6)
Completeness (%)	99.9 (100)	98.6 (99.2)	99.3 (98.4)	99.6 (99.4)
Redundancy	7.5 (7.4)	6.1 (6.0)	6.6 (6.5)	12.2 (12.1)
Average I/σ (I)	40.23 (10.44)	36.43 (7.38)	49.47 (25.74)	43.78 (18.45)
Refinement				
R _{work} / R _{free} (%)	15.48/17.82	16.06/18.25	15.60/17.71	15.57/18.13
Residues/water	540/786	540/632	270/481	270/395
CBA		2		1
RMSD				
Angle (°)	1.148	1.169	1.116	1.126
Length (Å)	0.006	0.006	0.006	0.006
Average B-factors (Å ²)	12.294	13.159	11.327	14.241
Ramachandran plot				
Most favored regions (%)	97.33	97.9	97.35	97.73
Allowed regions (%)	2.48	2.1	2.65	2.27
Outliers	0.19	0	0	0.00

Values in parentheses correspond to highest resolution shell.

†R-merge= $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl) \rangle$ is the average intensity of symmetry-related observations.

Part II

Structural Study of CagL, a Component of the T4SS Pilus Assembly in *Helicobacter pylori* K74 strain.

1. Overexpression and purification of K74 CagL

To obtain recombinant CagL-WT and site-mutated protein, plasmids was transformed into *E. coli* Rosetta2 with antibiotic. Transformed cells were induced with 1 mM IPTG during overnight at 18°C, 200 rpm for overexpression. Harvested cells were completely suspended in lysis buffer, and then, were purified to pure proteins by Ni-NTA binding affinity and size exclusion chromatography as described to materials and methods. Final buffer for crystallization and AUC is used to 20 mM Tris pH 7.5 and 50 mM NaCl. Purified protein was analyzed by 12% SDS-PAGE for size (24.6 kDa) and purity (Fig. 12). Site-mutated proteins were also shown to same pattern.

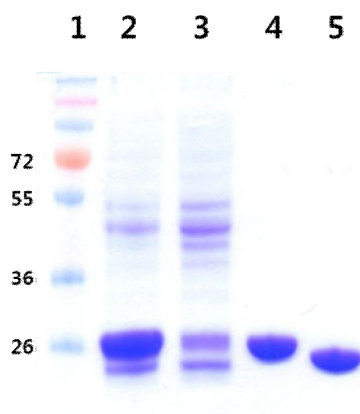


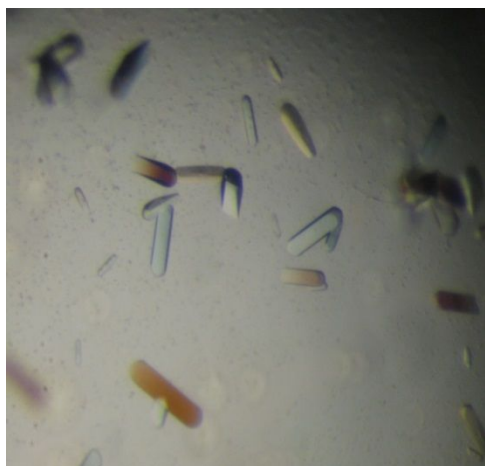
Figure 12. Analysis of purified Se-Met CagL by 12% SDS-PAGE. Se-Met CagL was purified by Ni-NTA affinity and SEC, and was analyzed by 12% SDS-PAGE. Lane 1: Marker (kDa), Lane 2: Supernatant, Lane 3: Flow through, Lane 4: 6 x His tagged CagL, Lane 5: CagL digested by human α -thrombin

2. Preliminary crystallographic results of K74 CagL

To obtain CagL crystal substituted Se-Met, purified CagL was concentrated ~10 mg/ml and crystallized to hanging drop by using commercial solutions at 20°C and 4°C. Initial crystals of Se-Met CagL appeared from only one condition (Crystal Screen Lite No 40 including 0.1 M Na-acetate pH 5.6, 10% isopropanol, and 10% PEG 4000) and looks like a small and thin rod-like shape (Fig. 13). The best crystals obtained from additive screen No.81 and used for x-ray diffraction.

To obtain x-ray crystallographic data to Se-Met CagL, Crystal of Se-Met CagL soaked with 20% glycerol diffracted at 2.9 Å resolution. Diffracted photo of crystal represents as Fig. 13. Diffracted data was collected, integrated, and scaled by the HKL2000 program. As a result of processing, detailed information of Se-Met CagL described as Table 3. The initial structures of Se-Met CagL were solved to R_{work} (29%)/ R_{free} (33%) by Autosol in PHNIX program for SAD phasing. The positions of 6 out of found 10 Se atoms in initial structure substituted at methionine position with 0.36 of FOM. To obtain more improved structure, native crystal diffracted at 2.8 Å resolution and determined by molecular replacement and Refmac5 using Se-Met CagL PDB as a search model. The final structure was completed to R_{work} (20.12%)/ R_{free} (25.67%) by iterative cycles of refinement and model building.

(A)



(B)

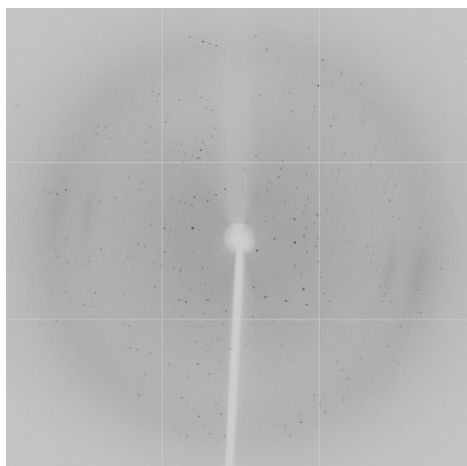


Figure 13. Crystal and x-ray diffraction images of Se-Met CagL.

A. Crystals of Se-Met CagL appeared from 3-4 days and completely grown until one to two weeks by hang drop in well solution condition containing 0.1 M Na-acetate pH 5.6, 10% isopropanol, 10% PEG 4000 and 3% dioxane at 20°C. B. Protein crystal of Se-Met CagL was diffracted at maximum 2.9 Å. The oscillation ranges were 360° with oscillations per frame being 1° and 10 second exposure time and distance measured at 450 mm.

3. Overall structure of K74 CagL

To examine structure to adhesion protein CagL comprising Type 4 secretion system, both Se-Met substituted and native CagL (21-237) were crystallized and determined to same tetragonal shape that contains two molecules in the asymmetric unit and belonged to space group $P4_32_12$ (Table 3). RMSD value of C α between each monomer is 0.59, meaning that each other very similar. Interface of both monomer looks like dimer form by interaction opposing $\alpha 5$ - $\alpha 6$ including hydrogen bond (Thr175-Glu219, Ser177-Arg223), salt bridge (Lys187-Gln220, Arg194-Glu209) using PISA program[61] and hydrophobic interaction (Leu156-Leu160, Leu198-Leu205, Leu190-Leu216) using program VMD 1.9 [62] (Fig. 14A).

Monomer of CagL consists of four long helices ($\alpha 1$, $\alpha 2$, $\alpha 5$, and $\alpha 6$) like bundle and two short helices ($\alpha 3$, $\alpha 4$) between long $\alpha 2$ and $\alpha 5$ (Fig. 14B). Structure of CagL is totally defined well but some regions were disordered. Three residues (21-23) of N-terminus and six residues (232-237) of C-terminus in both monomers not show and one of two monomers (A chain) not fit in electron density map from Lys52 to Asp58, indicating that these regions were likely to be very dynamic. Architecture of CagL was maintained by a variety of intermolecular interaction at following regions. $\alpha 1$ and $\alpha 2$ forms salt bridge at two regions. 1) Lys35 of $\alpha 1$ interacts to triad with Glu87 (3.0 Å, 3.6 Å) and Glu90 of $\alpha 2$ (3.4 Å, 3.9 Å). 2) Arg76 of RGD motif and Asp46 of $\alpha 1$ (3.64 Å, 3.11 Å). $\alpha 2$ and $\alpha 5$ forms hydrogen bond at two regions; 1) Asp78 and Asn166 (2.72 Å, 2.40 Å). 2) Asn85 and Asn89 interact with Ser159 to 3.33 Å and 3.16 Å (3.29 and 3.25). $\alpha 5$ and $\alpha 6$ forms hydrogen bond and salt bridge at three position; 1) Hydrogen bond can shows Asp132 with Tyr225 and Arg229 (2.03 Å (2.39 Å) and 3.80 Å (2.99 Å)) as well as Ser177 and Asn186 (2.90 Å, 2.90 Å). 2) Lys149 with Asp212 and Glu215 interacts with salt bridges (2.7 Å (3.38 Å) and 3.44 Å (3.49 Å). As described above, long helix form helix bundle by salt bridge and hydrogen bond for conformation maintenance. Short helix also interact between residues as followed; π - π interaction of Tyr103 and Tyr113 between $\alpha 2$ and $\alpha 3$, disulfide bond of Cys128 and Cys139 between $\alpha 4$ and $\alpha 5$, and hydrogen bond of Leu119 and

Glu140 (Fig. 15). Some of conserved aromatic residues (Phe86, Phe91, Phe92, Phe100, Tyr163, His200, Phe204, and Tyr206) forms hydrophobic core region that is likely to maintain stable conformation between $\alpha 2$ and $\alpha 6$.

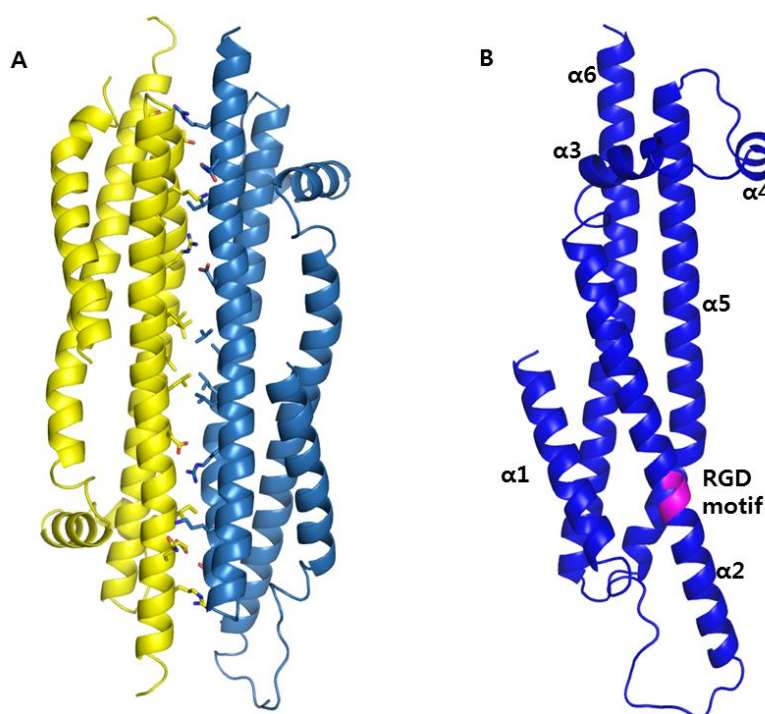


Figure 14. Overall structure of K74 CagL

A. CagL contains two molecules in the asymmetric unit. $\alpha 5$ and $\alpha 6$ of each molecule are opposite to each other that is likely to interact by salt bridge (Lys187-Gln220, Arg194-Glu209), hydrogen bond (Thr175-Glu219, Ser177-Arg223), and hydrophobic interaction (Leu156-Leu160, Leu198-Leu205, Leu190-Leu216). B. Monomer structure of CagL. Monomer of CagL consists of four elongate helices ($\alpha 1$, $\alpha 2$, $\alpha 5$, and $\alpha 6$) and two short helices ($\alpha 3$, $\alpha 4$). RGD motif represents to magenta color.

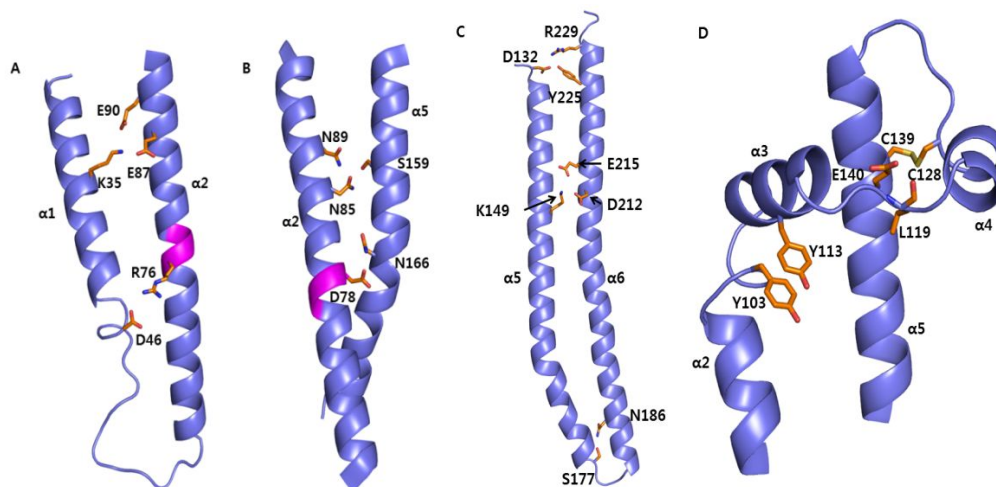


Figure 15. Intermolecular interaction for architecture on K74 CagL

A. Interaction between $\alpha 1$ and $\alpha 2$. Salt bridges of Lys35-Glu (87 and 90) and Asp46-Arg76. B. Interaction between $\alpha 2$ and $\alpha 5$. Hydrogen bonds of Asn (85 and 89)-Ser159 and Asp78-Asn166. C. Interaction between $\alpha 5$ and $\alpha 6$. 1) Two regions of triad interaction at Asp132-two residues (Tyr225 and Arg229) and Lys149-two residues (Asp212 and Glu215). 2) Hydrogen bond between Ser177 and Asn186. D. The increased stability of short helices. This region is maintained by π - π interaction of Tyr103-Tyr113, disulfide bond of Cys128-Cys139, and hydrogen bond of Leu119-Glu140.

4. Identification of oligomeric state of K74 CagL

To identify whether CagL exists as dimer in solution, CagL was firstly investigated oligomer state by analytical size exclusion chromatography. Molecular weight of CagL protein was calculated by standard curve using LMW kits and can show high peak at a region of dimer size (49.5 kDa) (Fig. 17A). So, to identify interface of both monomer forming dimer, CagL was investigated where interaction of two monomers formed structurally. Interface of CagL identified some pairs of residues that formed by a variety of bonding; hydrogen bond (Thr175-Glu219, Ser177-Arg223), salt bridge (Lys187-Gln220, Arg194-Glu209), and hydrophobic interaction (Leu156-Leu160, Leu198-Leu205, Leu190-Leu216). And then, to identify residue relative to dimerization, residues of regions described above were changed by site-mutagenesis and some of them are analyzed by size exclusion (Table 5). Although all mutants not investigated, all mutant proteins identified that were eluted on same position of WT (Fig. 16).

To examine more exactly the oligomer state of CagL, CagL protein was calculated molecular weight by sedimentation equilibrium using analytical ultracentrifugation (AUC) (Fig. 17B). Molecular weight of CagL by AUC is calculated to 25.99 kDa. This value is similar with molecular weight of monomer CagL. Therefore, although CagL by analytical size exclusion was calculated to dimer size, CagL suggests that it is not dimer but monomeric state as elongated molecule like TraC and 26695 CagL.

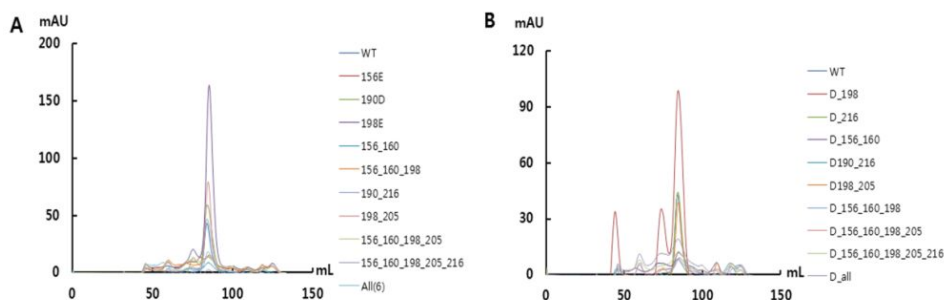


Figure 16. SEC of site-mutated CagL proteins (Table 4). A. Total ten CagL mutants containing WT were eluted by SEC using Hi-Load 16/60 Superdex 200 Column. B. Total ten CagL mutants containing WT including double mutation were eluted by SEC.

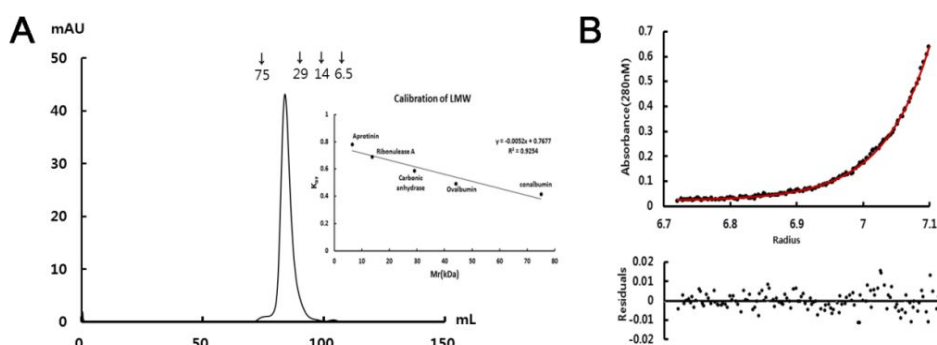


Figure 17. Analysis of oligomeric state by analytical size exclusion chromatography and analytical ultracentrifugation. A. CagL (~2 mg) using Hi-Load 16/60 Superdex 200 column for analytical size exclusion chromatography. To calculate molecular weight, five standard proteins (each 2-3 mg) using LMW kits are mixed in PBS buffer (50 mM phosphate, 150 mM NaCl, pH 7.2) and flow 500 μ l to 1 ml/min for calculation. Molecular weights of standard proteins are conalbumin (75.0 kDa), Ovalbumin (44.0 kDa), carbonic anhydrase (29.0 kDa), Ribonuclease A (13.7 kDa), and Aprotinin (6.5 kDa). B. Sedimentation equilibrium of CagL. Concentration of CagL used 25 μ M relevant OD₂₈₀ 0.25. The condition is carried at 25000 rpm during 72 hour at 4°C. All spots are fitted on curve at radius of between 6.7 and 7.1 cm. Molecular weight of CagL was automatically calculated to 26.0 kDa by program SEDPHAT ver.10.58d.

5. Comparison of both K74 CagL and 26695 CagL

To identify difference of K74 CagL and other proteins (26695 CagL and TraC), sequence alignment of K74 CagL with 26695 CagL is shown in Figure 18. Sequence homology of K74 CagL against 26695 CagL shows almost consensus (Identity: 96%/homology: 99%) except for eight positions but TraC shows low consensus (Identity: 27%/homology: 54%) as already reported previously [54].

To investigate structural difference to CagL of K74 and 26695, three CagL structures including k74 CagL and 26695 CagL (CagL^{meth}: lysine methylation and CagL^{KKQEK}: surface entropy reduction mutations) are superimposed. RMSD value of C α between CagL and CagL^{meth}, or CagL^{KKQEK} is 1.39 and 3.30, indicating that K74 CagL has a high similarity to CagL^{meth} than CagL^{KKQEK} (Fig. 18A). In 26695 CagL structure, structure of CagL^{meth} already was proposed more closed to native state than that of cagL^{KKQEK} although mutant formed by lysine methylation [54]. As a result of comparison of structure between K74 CagL and CagL^{meth}, both structures can show a great resemblance unlike CagL^{KKQEK}. Firstly, both structures formed linear α 2 helix through salt bridge and hydrogen bond between Asp46-Arg76 and Asp78-Asn166. Secondly, triad interaction of Lys149 with Asp212 and Glu215 at K74 CagL is also similar to CagL^{meth}.

As described above, CagL of K74 and CagL of 26695 are totally resembled to each other but also have a little difference. Specially, one of a significant difference is a disordered loop of α 1- α 2. K74 CagL was defined well at one region of two loop of α 1- α 2 though high value of B-factor (Fig. 19). C-terminus of α 1 is distorted and revealed shorter helix about one turn less than that of 26695 CagL. Movement of Phe49 into cleft of α 2 and α 5 is likely to increase flexibility of α 1- α 2 loop than 26695 (Fig. 19C). Studies about α 1- α 2 reported that Tyr58/Glu59 mutation of 26695 CagL prevent delivery of CagA into host cell [63], meaning that loop of α 1- α 2 might play a very important role for interaction with receptor of host cell. Next is triad interaction of Lys35 with Glu87 and Glu90 by salt bridges in K74 CagL but Gln35 in 26695

only forms hydrogen bond with Glu87, suggesting that this interaction may less flexible of $\alpha 1$ and $\alpha 2$ (Fig. 19B).



Figure 18. Sequence alignment of K74 CagL and 26695 CagL. Sequence alignment of three proteins K74 CagL, 26695 CagL, and Trac was performed by clustalX2. Secondary structure of K74 CagL was represented to upper aligned sequence.

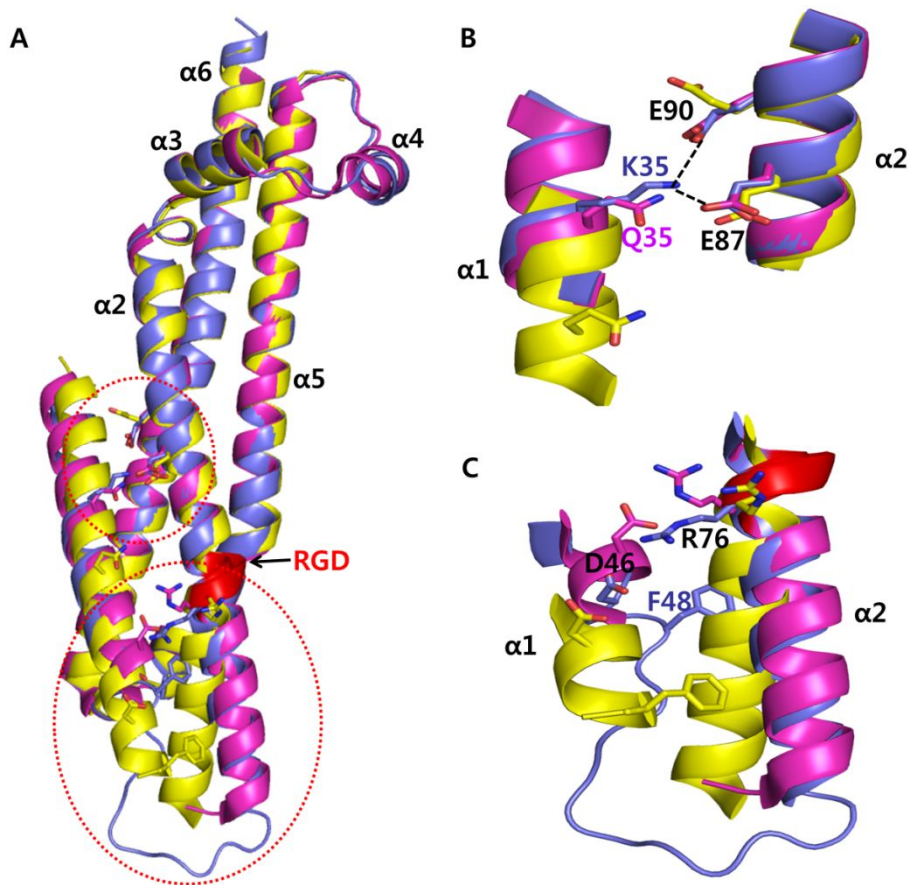


Figure 19. Comparison of structural difference by superimposition K74 CagL and 26695 CagL. A. Structural alignment of K74 CagL (slate), CagL^{meth} (magenta) and CagL^{KKQEK} (yellow) of 26695. Difference of aligned three CagL was represented by red circle dash line. B. Flexible region between $\alpha1$ helix and $\alpha2$ helix was shown.

Table 3. Data collection and refinement statistics of K74 CagL

	Se-Met_CagL	Native_CagL
Diffraction statistics		
Beam line		PAL-5C
Wavelength (Å)	0.97943	0.97938
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell parameters		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	97.761, 97.761, 144.838	97.616, 97.616, 144.242
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Data resolution (Å)	50.0-2.9 (2.95-2.90)	50.0-2.8 (2.9-2.8)
Completeness (%)	100	98.5 (94.6)
Redundancy	15.0 (13.1)	6.0 (3.5)
Total reflection	243436	106426
Unique reflections	16234 (775)	17703 (1639)
R-merge† (%)	22.9 (33.7)	11 (33)
Average I/σ	5.4 (3.5)	14 (2.64)
Matthew's coefficient (Å ³ Da ⁻¹)	2.36	2.32
Solvent content (%)	48	46.95
Refinement		
R _{work} / R _{free} (%)		20.12/25.67
Protein residues/water		410/82
RMSD		
Angle (°)		1.072
Length (Å)		0.009
Average B-factors (Å ²)		19.751
Ramachandran plot		
Most favored Regions (%)		95.29
Allowed regions (%)		4.47
Outliers		0.25

Values in parentheses correspond to highest resolution shell.

†R-merge= $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl) \rangle$ is the average intensity of symmetry-related observations.

Table 4. Primer sequence for site mutagenesis of K74 CagL

Region	residue	Primer sequence(5'-3')
S177	S → K	5' tccttaatcactgctcaagcagacgctttaga 3' ↓ Forward: 5' tccttaatcactgctcaagcagacgctttaga 3' Reverse: 5' tctacaagcgctctcttagcagtgattaagga 3'
	S → A	Forward: 5' tccttaatcactgctcagcagacgctttaga 3' Reverse: 5' tctacaagcgctctcttagcagtgattaagga 3'
	S → E	Forward: 5' tccttaatcactgctgaacagacgctttaga 3' Reverse: 5' tctacaagcgctctcttagcagtgattaagga 3'
R194	R → D	5' ttcattaaaggaatcagaagctcatgttagc 3' ↓ Forward 5' ttcattaaaggaatcacagctcatgttagc 3' Reverse 5' gctaacatgagcttctgattcctttaatgaa 3'
	R → K	Forward 5' ttcattaaaggaatcaaaagctcatgttagc 3' Reverse 5' gctaacatgagcttctgattcctttaatgaa 3'
	R → A	Forward 5' ttcattaaaggaatcgaagctcatgttagc 3' Reverse 5' gctaacatgagcttctgattcctttaatgaa 3'
E209	E → D	5' ggtctttttaattatttagaagtgaggacgcattag 3' ↓ Forward 5' ggtctttttaattatttagaagtgaggacgcattag3' Reverse 5' ctaatgcgtccaacttctctaaataatttaaaagacc3'
	E → K	Forward 5' ggtctttttaattatttagaagtgaggacgcattag3' Reverse 5' ctaatgcgtccaacttctctaaataatttaaaagacc3'
	E → A	Forward 5' ggtctttttaattatttagcgaagtgaggacgcattag3' Reverse 5' ctaatgcgtccaacttctctaaataatttaaaagacc3'
Q220	Q → R	5' gaaatatccctagaaacaagcaagcggcaatac 3' ↓ Forward 5' gaaatatccctagaaacgaagcaagcggcaatac 3' Reverse 5' gtattgccgcttgcttcttctagggatatttc 3'
	Q → E	Forward 5' gaaatatccctagaaacaagcaagcggcaatac 3' Reverse 5' gtattgccgcttgcttcttctagggatatttc 3'
	Q → L	Forward 5' gaaatatccctagaaactaagcaagcggcaatac 3' Reverse 5' gtattgccgcttgcttcttctagggatatttc 3'
R223	R → K	5' ctagaacaaagcaagcggcaatacctacaag 3' ↓ Forward 5' ctagaacaaagcaagaagcaatacctacaag 3' Reverse 5' ctgttaggtattgcttcttctgttcttag 3'
	R → L	Forward 5' ctagaacaaagcaagctgcaatacctacaag 3' Reverse 5' ctgttaggtattgcagcttcttctgttcttag 3'
	R → D	Forward 5' ctagaacaaagcaagcaccaatacctacaag 3' Reverse 5' ctgttaggtattgcttcttctgttcttcttag 3'
	R → A	Forward 5' ctagaacaaagcaaggccaatacctacaag 3' Reverse 5' ctgttaggtattgcgccttcttctgttcttag 3'
L156E	L → E	5' tataagagagcgttagaacttctctaa3' ↓ Forward 5' tataagagagcgtgaagaacttctctaa Reverse 5' tttagagaagttcttctacgctctcttata

L160E	L→E	<p>5' tttagaaacttctctaaggcgatcaga</p> <p>↓</p> <p>Forward 5' tttagaaacttctgaaaaggcgatcaga</p> <p>Reverse 5' tctgatacgccctttcagaagtttctaaa</p>
I190D	I→D	<p>5' gaataaaaaatttcattaagggaatcagaa</p> <p>↓</p> <p>Forward 5' gaataaaaaatttcgataaagggaatcagaa</p> <p>Reverse 5' ttctgattcctttatcgaaattttattc</p>
I216D	I→D	<p>5' ggacgcattagaaataccctagaacaaa</p> <p>↓</p> <p>Forward 5' ggacgcattagaa gactccctagaacaaa</p> <p>Reverse 5' ttgttctaggga gcttctaatgcgtcc</p>
L198E	L→E	<p>5' cagaaaagctcatgttagctcacgacaagg</p> <p>↓</p> <p>Forward 5' cagaaaagctcatggagctcacgacaagg</p> <p>Reverse 5' ccttgcgtgagcttc catgagctttctg</p>
L205E	L→E	<p>5' cgacaaggctcttttaaatatttagaga</p> <p>↓</p> <p>Forward 5' cgacaaggctctttgaaaattatttagaga</p> <p>Reverse 5' tctctaataattttc aaagacctgtcg</p>

Table 5. Constructs by site mutagenesis of K74 CagL

A) Hydrophobic interaction region	Site mutation
156, 190, 198	(L→E), (I→D), (L→E)
156-160, 190-216, 198-205	(L→E, L→E), (I→D, I→D), (L→E, L→E)
156-160-198	(L→E, L→E, L→E)
156-160-198-205	(L→E, L→E, L→E, L→E)
156-160-198-205-216	(L→E, L→E, L→E, L→E, I→D)
156-160-198-205-216-190	(L→E, L→E, L→E, L→E, I→D, I→D)
B) Hydrogen and hydrophobic interaction region	Site mutation
177-194-156	(S→K, R→D, L→E)
177-194-216	(S→K, R→D, I→D)
177-194-198	(S→K, R→D, L→E)
177-194-156-160	(S→K, R→D, L→E, L→E)
177-194-190-216	(S→K, R→D, I→D, I→D)
177-194-198-205	(S→K, R→D, L→E, L→E)
177-194-156-160-198	(S→K, R→D, L→E, L→E, L→E)
177-194-156-160-198-205	(S→K, R→D, L→E, L→E, L→E, L→E)
177-194-156-160-198-205-216	(S→K, R→D, L→E, L→E, L→E, L→E, I→D)
177-194-156-160-198-205-216-190	(S→K, R→D, L→E, L→E, L→E, L→E, I→D, I→D)
C) hydrogen bond, ion bond region	Site mutation
177, 194, 209, 220, 223	(S→K, A, E), (R→D, K), (E→D, K) (Q→R, E, L) (R→K, L, D)
177-194. 194-223	(S→K, R→D), (R→D, R→D)

IV. Discussion

Pathogenicity is the potential capacity causing dangerous diseases from microbes or viruses. This is characterized as ability evolved during struggle for their survival.

One of a variety of mechanism causing pathogenicity is a protective mechanism that increases resistance to antibiotics by β -lactamase in bacteria. Antibiotics hydrolysis by β -lactamase is a very important mechanism for survival. Bacteria mainly evolved for adaption to antibiotics by a variety of method such as point mutations and tandem repeats insertion of amino acid in β -lactamase. Previous studied well-known β -lactamases, TEM, SHV, and CTX-M family, already described that the adaption to antibiotics in enzyme can induce change of substrate spectrum by point mutation on omega loop [16-18]. SHV-16 and others are also reported that insertion of repeat unit was changed in substrate spectrum [64, 65].

PenA, which has a resistance to amoxicillin but not ceftazidime, was already known to a similar ortholog family to human pathogen *B.pseudomallei*, *B.mallei*, Bcc. Some of mutants by large scale screening demonstrate that increase in resistance to antibiotics resulting in the increased MIC value for ceftazidime (CAZ) [26, 27]. Also, above results suggested that point mutants characterized for ESBL but TR mutants can reversely switching between two substrate phases.

In this study, Pen-WT as standard structure was determined at high resolution. It is likely to possess a narrow cavity by tight interaction of Arg104 and Thr167 and position of Tyr105 than other β -lactamase (Fig. 6), causing the difference of substrate spectrum which is no resistance to third generation antibiotic, ceftazidime.

As a result, determining of TR mutants, Ω loop region of active site in TR10 and TR11 structure was found to seriously disordered. In the case of TR10, almost all of 11-12 amino acids, which contain entire first TR region with first amino acid of second TR on omega loop, disappeared. TR11 was more disordered than TR10 that cannot show entire omega loop region

containing 28 amino acid of first TR region with second first residue. This suggests that both mutants may be in preparation for binding of substrate through disorder of omega loop by TR insertion. In CBA-bound TR mutants, one of the differences is antibiotic binding position for hydrolysis in active site. CBA in TR10 positions flatly that thiazole ring stabilizes by interaction with Tyr105 like that of SHV-1. But, CBA in TR11 binds unstably by interacting with thiazole ring in Asp240 in wide active site than TR10. Consequently, TRs were concluded that a temporal structural relaxation by insertion in the active site cavity allowed the access of bulky substrates to the cavity and re-packed after docking for the subsequent catalytic reactions.

Structural mechanism of pointed mutants including N136D and C69Y with CBA also were investigated by crystallization. As a result of structural study in pointed mutants, both mutants exhibit commonly the broken interaction between Arg104 and Thr167. In N136D structure, substituted Asp136 induces upward of the loosen loop containing Arg104 and Tyr105 by the broken interaction of water between substituted Asp136 and Asp101. In CBA-soaked N136D, CBA enters into the increased active site by mutation, and then, two conformation revealed acylation and deacylation form for CBA binding at Ser70, meaning that two conformations reveals reaction for hydrolysis of antibiotic.

In PenA-C69Y, change of cysteine to tyrosine at position 69 induces change of active site by the broken interaction between Arg104 and Thr167. In CBA-bound PenA-C69Y, CBA binds to acylation form through the increased cavity by mutation. CBA binding for hydrolysis in active site of both structures (N136D and C69Y) bind to similar with that of TEM-1 and CTX-M-9 that thiazole ring of CBA interacts with the side chain of Asp240. In other words, change for ESBL in PenA-N136D and PenA-C69Y retains space for entrance of substrate through rising of Tyr105 by broken interaction between both residues, Arg104 and Thr167.

In summary, structural difference of TR mutants (TR10 and TR11) and point mutants (N136D and C69Y) for extended spectrum is the change of active site for increase of cavity volume. That is, TRs allow to extended substrate spectrum by induction of temporal

perturbation of Ω loop but point mutants allow to ESBL through the increased cavity by loosening loop containing Tyr105. This difference suggests that TRs might be a possible mechanism for reversely switching of spectrum between two phases of substrate unlike pointed mutants. Furthermore, these results provide information about how the design of the β -lactamases with the omega loop as part of the binding cavity is effective in their adaptation to new drugs.

The other mechanism causing pathogenicity is translocation of effector molecule by contact of pathogen-host cell.

Helicobacter pylori is one of well-known pathogen causing diseases including gastritis, peptic ulcers and gastric adenocarcinoma in stomach. it uses type 4 secretion system composing of a variety proteins encoded Cag PAI gene and induces signaling pathway by translocation of oncoprotein such as cagA into host cell [66]. CagL, which is specific adhesion protein among cag proteins, is known to binding partner for interaction with host cell receptor, integrin. CagI and CagH, components of assembly forming pilus of *H.pylori*, also reported about complex formation with CagL [55].

Structure of K74 CagL in this study identified preferentially as using Se-Met substituted crystal. Unlike 26695, K74 CagL is high soluble than 26695 in solution, indicating that obtained protein to soluble form not refolding. Also, pI value difference of both K74 (7.90) and 26695 (5.31) CagL exhibited by sequence difference of $\alpha 1$ - $\alpha 2$ loop.

CagL from *H.pylori* K74 strain was determined and demonstrated structural differences in CagLs. Although the structures of the two CagLs between CagLs (CagL 26695 and K74) are similar as shown above, sequence polymorphism in conjunction with structures may provide insights for biological functions of CagL during T4SS-mediated interaction with host cell. It has been believed that the RGD sequence is supposed to adopt partly flexible and loopy structure [51], and in turn suitable to interact with integrin on the target cell membrane [67]. However, the RGD sequence in CagL is located on the long rigid $\alpha 2$ helix in both CagL

structures. Therefore, the unprecedented location of the RGD in the rigid helix 2 on both of CagL structures (Fig. 19) raises a question whether the motif or protein itself might not be involved in the direct contact with host cell or other part of the protein could be responsible for the interaction. Since both the cases are tightly associated with the conformation of the RGD sequence, the plasticity of $\alpha 1$ and $\alpha 2$ helices is also considered to be important. The previous report suggested that the loose interaction of $\alpha 1$ helix with adjacent helices including $\alpha 2$, resulting in the bending of $\alpha 2$ and subsequent local unwinding around RGD sequence [54]. As shown in CagL K74 structure, an additional salt bridge by K35 is likely to help tighter interaction between $\alpha 1$ and $\alpha 2$. Furthermore, D46 in K74 CagL that make salt bridge with R76 in RGD resides in the loop, meanwhile D46 in 26695 CagL was found in the $\alpha 1$ helix (Fig. 19), which may also affect the formation of the relatively well-defined loop connecting $\alpha 1$ to $\alpha 2$ in our structure (Fig.19). This correlates with the fact that our protein was comparatively more stable against long-term storage than refolded 26695 CagL that was auto-degradable and protease sensitive in the same region. Those observations may indicate that the two helices are packed tighter and less flexible around the RGD motif comparing to the region of 26695. In this regard, the RGD motif is likely to remain rigid and might not be the main sequence for the docking to host cell membrane in the case of CagL. In agreement, the complex structure of talin with integrin showed the possibility of non-RGD sequence participation in the interaction, in which the interface of F2-F3 domain of talin did not contain RGD sequence [68]. In addition, RGD-independent translocation of CagA into host cell by T4SS indicates that CagL may not be the necessary component for the host-pathogen recognition [53]. Therefore, it is possible that other region of the protein may involve in the integrin interaction. As demonstrated, the structural analysis of CagL in respect to the host interaction via RGD sequence is not clear.

In addition to CagL, because CagI and CagH are also known to interaction with integrin through formation of complex with CagL, purification of recombinant protein CagI and CagH are attempted for identification whether it interacts with each other, until now. However,

according as both proteins expressed to insoluble form, protein of soluble form for analysis failed.

Taken together, in order to elucidate precise role of canonical RGD motif on T4SS-mediated *Hp* carcinogenesis, the complex structure of CagL with integrin is necessary.

V. References

1. Goldman, W.E., D.G. Klapper, and J.B. Baseman, *Detection, isolation, and analysis of a released Bordetella pertussis product toxic to cultured tracheal cells*. Infection and immunity, 1982. 36(2): p. 782-94.
2. Melly, M.A., Z.A. McGee, and R.S. Rosenthal, *Ability of monomeric peptidoglycan fragments from Neisseria gonorrhoeae to damage human fallopian-tube mucosa*. The Journal of infectious diseases, 1984. 149(3): p. 378-86.
3. Viala, J., et al., *Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island*. Nature immunology, 2004. 5(11): p. 1166-74.
4. Zeng, X. and J. Lin, *Beta-lactamase induction and cell wall metabolism in Gram-negative bacteria*. Frontiers in microbiology, 2013. 4: p. 128.
5. Neu, H.C., *Effect of beta-lactamase location in Escherichia coli on penicillin synergy*. Applied microbiology, 1969. 17(6): p. 783-6.
6. Ambler, R.P., *The structure of beta-lactamases*. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 1980. 289(1036): p. 321-31.
7. Papp-Wallace, K.M., et al., *Insights into beta-lactamases from Burkholderia species, two phylogenetically related yet distinct resistance determinants*. The Journal of biological chemistry, 2013. 288(26): p. 19090-102.
8. Matagne, A., J. Lamotte-Brasseur, and J.M. Frere, *Catalytic properties of class A beta-lactamases: efficiency and diversity*. The Biochemical journal, 1998. 330 (Pt 2): p. 581-98.
9. Drawz, S.M. and R.A. Bonomo, *Three decades of beta-lactamase inhibitors*. Clinical microbiology reviews, 2010. 23(1): p. 160-201.
10. Petit, A., et al., *Multiple substitutions at position 104 of beta-lactamase TEM-1: assessing the role of this residue in substrate specificity*. The Biochemical journal,

1995. 305 (Pt 1): p. 33-40.
11. Doucet, N., P.Y. De Wals, and J.N. Pelletier, *Site-saturation mutagenesis of Tyr-105 reveals its importance in substrate stabilization and discrimination in TEM-1 beta-lactamase*. The Journal of biological chemistry, 2004. 279(44): p. 46295-303.
 12. Bush, K., G.A. Jacoby, and A.A. Medeiros, *A functional classification scheme for beta-lactamases and its correlation with molecular structure*. Antimicrobial agents and chemotherapy, 1995. 39(6): p. 1211-33.
 13. Labia, R., et al., *The kinetics of SHV-2 plasmid-mediated beta-lactamase compared to those of the parent enzyme from which it is derived*. Drugs under experimental and clinical research, 1988. 14(5): p. 335-9.
 14. Sougakoff, W., et al., *Plasmid-mediated resistance to third-generation cephalosporins caused by point mutations in TEM-type penicillinase genes*. Reviews of infectious diseases, 1988. 10(4): p. 879-84.
 15. Collatz, E., et al., *Substitution of serine for arginine in position 162 of TEM-type beta-lactamases extends the substrate profile of mutant enzymes, TEM-7 and TEM-101, to ceftazidime and aztreonam*. Gene, 1989. 78(2): p. 349-54.
 16. Paterson, D.L. and R.A. Bonomo, *Extended-spectrum beta-lactamases: a clinical update*. Clinical microbiology reviews, 2005. 18(4): p. 657-86.
 17. Queenan, A.M., et al., *Interactions of ceftobiprole with beta-lactamases from molecular classes A to D*. Antimicrobial agents and chemotherapy, 2007. 51(9): p. 3089-95.
 18. Anderson, S.D. and J.G. Gums, *Ceftobiprole: an extended-spectrum anti-methicillin-resistant Staphylococcus aureus cephalosporin*. The Annals of pharmacotherapy, 2008. 42(6): p. 806-16.
 19. Knox, J.R., *Extended-spectrum and inhibitor-resistant TEM-type beta-lactamases: mutations, specificity, and three-dimensional structure*. Antimicrobial agents and

- chemotherapy, 1995. 39(12): p. 2593-601.
20. Chen, Y., B. Shoichet, and R. Bonnet, *Structure, function, and inhibition along the reaction coordinate of CTX-M beta-lactamases*. Journal of the American Chemical Society, 2005. 127(15): p. 5423-34.
 21. Brittain, D.R., et al., *Competing S(N)2 and E2 reaction pathways for hexachlorocyclohexane degradation in the gas phase, solution and enzymes*. Chemical communications, 2011. 47(3): p. 976-8.
 22. Tribuddharat, C., et al., *Burkholderia pseudomallei class a beta-lactamase mutations that confer selective resistance against ceftazidime or clavulanic acid inhibition*. Antimicrobial agents and chemotherapy, 2003. 47(7): p. 2082-7.
 23. Poirel, L., et al., *Naturally occurring Class A ss-lactamases from the Burkholderia cepacia complex*. Antimicrobial agents and chemotherapy, 2009. 53(3): p. 876-82.
 24. Sam, I.C., K.H. See, and S.D. Puthuchear, *Variations in ceftazidime and amoxicillin-clavulanate susceptibilities within a clonal infection of Burkholderia pseudomallei*. Journal of clinical microbiology, 2009. 47(5): p. 1556-8.
 25. Sarovich, D.S., et al., *Characterization of ceftazidime resistance mechanisms in clinical isolates of Burkholderia pseudomallei from Australia*. PloS one, 2012. 7(2): p. e30789.
 26. Yi, H., et al., *Twelve positions in a beta-lactamase that can expand its substrate spectrum with a single amino acid substitution*. PloS one, 2012. 7(5): p. e37585.
 27. Yi, H., et al., *The tandem repeats enabling reversible switching between the two phases of beta-lactamase substrate spectrum*. PLoS genetics, 2014. 10(9): p. e1004640.
 28. Delmas, J., et al., *Structure and dynamics of CTX-M enzymes reveal insights into substrate accommodation by extended-spectrum beta-lactamases*. Journal of molecular biology, 2008. 375(1): p. 192-201.
 29. Brown, L.M., *Helicobacter pylori: epidemiology and routes of transmission*.

- Epidemiologic reviews, 2000. 22(2): p. 283-97.
30. Pandey, R., et al., *Helicobacter pylori and gastric cancer*. Asian Pacific journal of cancer prevention : APJCP, 2010. 11(3): p. 583-8.
 31. Amieva, M.R. and E.M. El-Omar, *Host-bacterial interactions in Helicobacter pylori infection*. Gastroenterology, 2008. 134(1): p. 306-23.
 32. Blaser, M.J. and J.C. Atherton, *Helicobacter pylori persistence: biology and disease*. The Journal of clinical investigation, 2004. 113(3): p. 321-33.
 33. Cover, T.L. and M.J. Blaser, *Helicobacter pylori in health and disease*. Gastroenterology, 2009. 136(6): p. 1863-73.
 34. Suerbaum, S. and P. Michetti, *Helicobacter pylori infection*. The New England journal of medicine, 2002. 347(15): p. 1175-86.
 35. Blaser, M.J., et al., *Infection with Helicobacter pylori strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach*. Cancer research, 1995. 55(10): p. 2111-5.
 36. Kuipers, E.J., et al., *Helicobacter pylori and atrophic gastritis: importance of the cagA status*. Journal of the National Cancer Institute, 1995. 87(23): p. 1777-80.
 37. Parsonnet, J., et al., *Risk for gastric cancer in people with CagA positive or CagA negative Helicobacter pylori infection*. Gut, 1997. 40(3): p. 297-301.
 38. Argent, R.H., et al., *Differences in Helicobacter pylori CagA tyrosine phosphorylation motif patterns between western and East Asian strains, and influences on interleukin-8 secretion*. Journal of medical microbiology, 2008. 57(Pt 9): p. 1062-7.
 39. Backert, S. and T.F. Meyer, *Type IV secretion systems and their effectors in bacterial pathogenesis*. Current opinion in microbiology, 2006. 9(2): p. 207-17.
 40. Cascales, E. and P.J. Christie, *The versatile bacterial type IV secretion systems*. Nature reviews. Microbiology, 2003. 1(2): p. 137-49.
 41. Christie, P.J., *Type IV secretion: the Agrobacterium VirB/D4 and related conjugation*

- systems. Biochimica et biophysica acta, 2004. 1694(1-3): p. 219-34.
42. Fischer, W., *Assembly and molecular mode of action of the Helicobacter pylori Cag type IV secretion apparatus*. The FEBS journal, 2011. 278(8): p. 1203-12.
 43. Terradot, L. and G. Waksman, *Architecture of the Helicobacter pylori Cag-type IV secretion system*. The FEBS journal, 2011. 278(8): p. 1213-22.
 44. Akopyants, N.S., et al., *Analyses of the cag pathogenicity island of Helicobacter pylori*. Molecular microbiology, 1998. 28(1): p. 37-53.
 45. Censini, S., et al., *cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors*. Proceedings of the National Academy of Sciences of the United States of America, 1996. 93(25): p. 14648-53.
 46. Kwok, T., et al., *Helicobacter exploits integrin for type IV secretion and kinase activation*. Nature, 2007. 449(7164): p. 862-6.
 47. Rohde, M., et al., *A novel sheathed surface organelle of the Helicobacter pylori cag type IV secretion system*. Molecular microbiology, 2003. 49(1): p. 219-34.
 48. Kutter, S., et al., *Protein subassemblies of the Helicobacter pylori Cag type IV secretion system revealed by localization and interaction studies*. Journal of bacteriology, 2008. 190(6): p. 2161-71.
 49. Tegtmeyer, N., et al., *A small fibronectin-mimicking protein from bacteria induces cell spreading and focal adhesion formation*. The Journal of biological chemistry, 2010. 285(30): p. 23515-26.
 50. Wiedemann, T., et al., *Helicobacter pylori CagL dependent induction of gastrin expression via a novel alphavbeta5-integrin-integrin linked kinase signalling complex*. Gut, 2012. 61(7): p. 986-96.
 51. Conradi, J., et al., *Cyclic RGD peptides interfere with binding of the Helicobacter pylori protein CagL to integrins alpha(V)beta (3) and alpha (5)beta (1)*. Amino Acids, 2012. 43(1): p. 219-32.

52. Conradi, J., et al., *An RGD helper sequence in CagL of Helicobacter pylori assists in interactions with integrins and injection of CagA*. *Frontiers in cellular and infection microbiology*, 2012. 2: p. 70.
53. Jimenez-Soto, L.F., et al., *Helicobacter pylori type IV secretion apparatus exploits beta1 integrin in a novel RGD-independent manner*. *PLoS pathogens*, 2009. 5(12): p. e1000684.
54. Barden, S., et al., *A helical RGD motif promoting cell adhesion: crystal structures of the Helicobacter pylori type IV secretion system pilus protein CagL*. *Structure*, 2013. 21(11): p. 1931-41.
55. Shaffer, C.L., et al., *Helicobacter pylori exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface*. *PLoS pathogens*, 2011. 7(9): p. e1002237.
56. Backert, S. and M. Selbach, *Role of type IV secretion in Helicobacter pylori pathogenesis*. *Cellular microbiology*, 2008. 10(8): p. 1573-81.
57. Guerrero, S.A., et al., *Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems*. *Applied microbiology and biotechnology*, 2001. 56(5-6): p. 718-23.
58. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. *Acta crystallographica. Section D, Biological crystallography*, 2004. 60(Pt 12 Pt 1): p. 2126-32.
59. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method*. *Acta crystallographica. Section D, Biological crystallography*, 1997. 53(Pt 3): p. 240-55.
60. Adams, P.D., et al., *PHENIX: building new software for automated crystallographic structure determination*. *Acta crystallographica. Section D, Biological crystallography*, 2002. 58(Pt 11): p. 1948-54.

61. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*. Journal of molecular biology, 2007. 372(3): p. 774-97.
62. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. Journal of molecular graphics, 1996. 14(1): p. 33-8, 27-8.
63. Tegtmeyer, N., et al., *Helicobacter pylori CagL Y58/E59 mutation turns-off type IV secretion-dependent delivery of CagA into host cells*. PloS one, 2014. 9(6): p. e97782.
64. Arpin, C., et al., *SHV-16, a beta-lactamase with a pentapeptide duplication in the omega loop*. Antimicrobial agents and chemotherapy, 2001. 45(9): p. 2480-5.
65. Nukaga, M., et al., *Molecular evolution of a class C beta-lactamase extending its substrate specificity*. The Journal of biological chemistry, 1995. 270(11): p. 5729-35.
66. Odenbreit, S., et al., *Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion*. Science, 2000. 287(5457): p. 1497-500.
67. Gilcrease, M.Z., *Integrin signaling in epithelial cells*. Cancer letters, 2007. 247(1): p. 1-25.
68. Anthis, N.J., et al., *The structure of an integrin/talin complex reveals the basis of inside-out signal transduction*. The EMBO journal, 2009. 28(22): p. 3623-32.
69. Vieille, C. and G.J. Zeikus, *Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability*. Microbiology and molecular biology reviews : MMBR, 2001. 65(1): p. 1-43.
70. Cheetham, P.S.J. and A.N. Wootton, *Bioconversion of D-galactose to D-tagatose*. Enzyme Microb. Technol., 1993. 15: p. 105-108.
71. Sa-Nogueira, I. and H. de Lencastre, *Cloning and characterization of araA, araB, and araD, the structural genes for L-arabinose utilization in Bacillus subtilis*. J Bacteriol, 1989. 171(7): p. 4088-91.
72. Sa-Nogueira, I., et al., *The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence, genetic organization and expression*. Microbiology, 1997. 143 (Pt 3): p.

- 957-69.
73. Schleif, R., *Regulation of the L-arabinose operon of Escherichia coli*. Trends Genet, 2000. 16(12): p. 559-65.
 74. Levin, G.V., *Tagatose, the new GRAS sweetener and health product*. J Med Food, 2002. 5(1): p. 23-36.
 75. Lee, D.W., et al., *Characterization of a thermostable L-arabinose (D-galactose) isomerase from the hyperthermophilic eubacterium Thermotoga maritima*. Appl Environ Microbiol, 2004. 70(3): p. 1397-404.
 76. Oh, D.K., *Tagatose: properties, applications, and biotechnological processes*. Appl Microbiol Biotechnol, 2007. 76(1): p. 1-8.
 77. Lu, Y., G.V. Levin, and T.W. Donner, *Tagatose, a new antidiabetic and obesity control drug*. Diabetes Obes Metab, 2008. 10(2): p. 109-34.
 78. Kim, P., *Current studies on biological tagatose production using L-arabinose isomerase: a review and future perspective*. Appl Microbiol Biotechnol, 2004. 65(3): p. 243-9.
 79. Roh, H.J., et al., *Bioconversion of D-galactose into D-tagatose by expression of L-arabinose isomerase*. Biotechnol Appl Biochem, 2000. 31 (Pt 1): p. 1-4.
 80. Yamanaka, K., *L-arabinose isomerase from Lactobacillus gayonii*. Methods Enzymol., 1975. 41: p. 458-461.
 81. Tewari, Y.B., D.K. Steckler, and R.N. Goldberg, *Thermodynamics of the conversion of aqueous xylose to xylulose*. Biophys Chem, 1985. 22(3): p. 181-5.
 82. Hong, Y.H., et al., *Production of D-tagatose at high temperatures using immobilized Escherichia coli cells expressing L-arabinose isomerase from Thermotoga neapolitana*. Biotechnol Lett, 2007. 29(4): p. 569-74.
 83. Jorgensen, F., O.C. Hansen, and P. Stougaard, *Enzymatic conversion of D-galactose to D-tagatose: heterologous expression and characterisation of a thermostable L-*

- arabinose isomerase from Thermoanaerobacter mathranii*. Appl Microbiol Biotechnol, 2004. 64(6): p. 816-22.
84. Hung, X.G., et al., *Characterization of a thermophilic L-arabinose isomerase from Thermoanaerobacterium saccharolyticum NTOU1*. Biochemical Engineering Journal, 2014. 83: p. 121-128.
 85. Lim, B.C., H.J. Kim, and D.K. Oh, *Tagatose production with pH control in a stirred tank reactor containing immobilized L-arabinose rom Thermotoga neapolitana*. Appl Biochem Biotechnol, 2008. 149(3): p. 245-53.
 86. Rhimi, M., et al., *The acid tolerant L-arabinose isomerase from the food grade Lactobacillus sakei 23K is an attractive D-tagatose producer*. Bioresour Technol, 2010. 101(23): p. 9171-7.
 87. Liang, M., et al., *Bioconversion of D-galactose to D-tagatose: continuous packed bed reaction with an immobilized thermostable L-arabinose isomerase and efficient purification by selective microbial degradation*. Appl Microbiol Biotechnol, 2012. 93(4): p. 1469-74.
 88. Kim, B.C., et al., *Cloning, expression and characterization of L-arabinose isomerase from Thermotoga neapolitana: bioconversion of D-galactose to D-tagatose using the enzyme*. FEMS microbiology letters, 2002. 212(1): p. 121-6.
 89. Lee, D.W., et al., *Distinct metal dependence for catalytic and structural functions in the L-arabinose isomerases from the mesophilic Bacillus halodurans and the thermophilic Geobacillus stearothermophilus*. Arch. Biochem. Biophys., 2005. 434(2): p. 333-343.
 90. Lee, D.W., et al., *A thermodynamic study of mesophilic, thermophilic, and hyperthermophilic L-arabinose isomerases: the effects of divalent metal ions on protein stability at elevated temperatures*. FEBS Lett., 2005. 579(5): p. 1261-1266.
 91. Hong, Y.H., et al., *Creation of metal-independent hyperthermophilic L-arabinose*

- isomerase by homologous recombination. J Agric Food Chem*, 2011. 59: p. 12939-12947.
92. Lee, S.J., et al., *Homologous Alkalophilic and Acidophilic L-Arabinose isomerases reveal region-specific contributions to the pH dependence of activity and stability. Appl Environ Microbiol*, 2012. 78(24): p. 8813-6.
 93. Manjasetty, B.A. and M.R. Chance, *Crystal structure of Escherichia coli L-arabinose isomerase (ECAI), the putative target of biological tagatose production. J Mol Biol*, 2006. 360(2): p. 297-309.
 94. Patrick, J.W. and N. Lee, *Purification and properties of an L-arabinose isomerase from Escherichia coli. J. Biol. Chem.*, 1968. 243(16): p. 4312-4318.
 95. Patrick, J.W. and N. Lee, *Subunit structure of L-arabinose isomerase from Escherichia coli. J Biol Chem*, 1969. 244(16): p. 4277-83.
 96. Seemann, J.E. and G.E. Schulz, *Structure and mechanism of L-fucose isomerase from Escherichia coli. J Mol Biol*, 1997. 273(1): p. 256-68.
 97. Lee, Y.J., et al., *Structural insights into conserved L-arabinose metabolic enzymes reveal the substrate binding site of a thermophilic L-arabinose isomerase. FEBS Lett*, 2014. 588(6): p. 1064-70.
 98. Nazina, T.N., et al., *Taxonomic study of aerobic thermophilic bacilli: descriptions of Geobacillus subterraneus gen. nov., sp. nov. and Geobacillus uzenensis sp. nov. from petroleum reservoirs and transfer of Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermodenitrificans to Geobacillus as the new combinations G. stearothermophilus, G. th. Int J Syst Evol Microbiol*, 2001. 51(Pt 2): p. 433-46.
 99. Takami, H., et al., *Thermoadaptation trait revealed by the genome sequence of thermophilic Geobacillus kaustophilus. Nucleic Acids Res*, 2004. 32(21): p. 6292-303.
 100. Lee, D.W., et al., *Proteomic analysis of acetylation in thermophilic Geobacillus*

- kaustophilus*. Proteomics, 2013. 13(15): p. 2278-82.
101. Cao, T.P., et al., *Crystallization and preliminary X-ray crystallographic analysis of L-arabinose isomerase from thermophilic Geobacillus kaustophilus*. Acta crystallographica. Section F, Structural biology communications, 2014. 70(Pt 1): p. 108-12.
 102. Lee, D.W., et al., *Characterization of a thermostable L-arabinose (D-galactose) isomerase from the hyperthermophilic eubacterium Thermotoga maritima*. Applied and environmental microbiology, 2004. 70(3): p. 1397-404.
 103. Jorgensen, F., O.C. Hansen, and P. Stougaard, *Enzymatic conversion of D-galactose to D-tagatose: heterologous expression and characterisation of a thermostable L-arabinose isomerase from Thermoanaerobacter mathranii*. Applied microbiology and biotechnology, 2004. 64(6): p. 816-22.
 104. Manjasetty, B.A. and M.R. Chance, *Crystal structure of Escherichia coli L-arabinose isomerase (ECAI), the putative target of biological tagatose production*. Journal of molecular biology, 2006. 360(2): p. 297-309.
 105. Seemann, J.E. and G.E. Schulz, *Structure and mechanism of L-fucose isomerase from Escherichia coli*. Journal of molecular biology, 1997. 273(1): p. 256-68.
 106. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D*. Nucleic acids research, 2010. 38(Web Server issue): p. W545-9.
 107. Dundas, J., et al., *CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues*. Nucleic acids research, 2006. 34(Web Server issue): p. W116-8.
 108. Lee, Y.J., et al., *Structural insights into conserved L-arabinose metabolic enzymes reveal the substrate binding site of a thermophilic L-arabinose isomerase*. FEBS letters, 2014. 588(6): p. 1064-70.
 109. Nagorski, R.W. and J.P. Richard, *Mechanistic imperatives for aldose-ketose*

- isomerization in water: specific, general base- and metal ion-catalyzed isomerization of glyceraldehyde with proton and hydride transfer. Journal of the American Chemical Society, 2001. 123(5): p. 794-802.*
110. Lee, D.W., et al., *Distinct metal dependence for catalytic and structural functions in the L-arabinose isomerases from the mesophilic Bacillus halodurans and the thermophilic Geobacillus stearothermophilus. Archives of biochemistry and biophysics, 2005. 434(2): p. 333-43.*
 111. Vogt, G., S. Woell, and P. Argos, *Protein thermal stability, hydrogen bonds, and ion pairs. Journal of molecular biology, 1997. 269(4): p. 631-43.*
 112. Taylor, T.J. and Vaisman, II, *Discrimination of thermophilic and mesophilic proteins. BMC structural biology, 2010. 10 Suppl 1: p. S5.*
 113. Maes, D., et al., *The crystal structure of triosephosphate isomerase (TIM) from Thermotoga maritima: a comparative thermostability structural analysis of ten different TIM structures. Proteins, 1999. 37(3): p. 441-53.*
 114. Kumar, S., C.J. Tsai, and R. Nussinov, *Factors enhancing protein thermostability. Protein engineering, 2000. 13(3): p. 179-91.*
 115. Vetriani, C., A.L. Reysenbach, and J. Dore, *Recovery and phylogenetic analysis of archaeal rRNA sequences from continental shelf sediments. FEMS microbiology letters, 1998. 161(1): p. 83-8.*
 116. Jaenicke, R., *Glyceraldehyde-3-phosphate dehydrogenase from Thermotoga maritima: strategies of protein stabilization. FEMS microbiology reviews, 1996. 18(2-3): p. 215-24.*
 117. Privalov, P.L. and S.J. Gill, *Stability of protein structure and hydrophobic interaction. Advances in protein chemistry, 1988. 39: p. 191-234.*
 118. van Bueren, A.L., et al., *Three-dimensional structure of a thermophilic family GH11 xylanase from Thermobifida fusca. Acta crystallographica. Section F, Structural*

- biology and crystallization communications, 2012. 68(Pt 2): p. 141-4.
119. Shirley, B.A., et al., *Contribution of hydrogen bonding to the conformational stability of ribonuclease T1*. Biochemistry, 1992. 31(3): p. 725-32.
 120. Izumori, K., Y. Ueda, and K. Yamanaka, *Pentose metabolism in Mycobacterium smegmatis: comparison of L-arabinose isomerases induced by L-arabinose and D-galactose*. Journal of bacteriology, 1978. 133(1): p. 413-4.
 121. Livesey, G. and J.C. Brown, *D-tagatose is a bulk sweetener with zero energy determined in rats*. The Journal of nutrition, 1996. 126(6): p. 1601-9.
 122. Espinosa, I. and L. Fogelfeld, *Tagatose: from a sweetener to a new diabetic medication?* Expert opinion on investigational drugs, 2010. 19(2): p. 285-94.
 123. Oh, D.K., *Tagatose: properties, applications, and biotechnological processes*. Applied microbiology and biotechnology, 2007. 76(1): p. 1-8.
 124. Kim, H.J., et al., *A feasible enzymatic process for D-tagatose production by an immobilized thermostable L-arabinose isomerase in a packed-bed bioreactor*. Biotechnology progress, 2003. 19(2): p. 400-4.

VI. Appendix

국문초록

Geobacillus kaustophilus 의 호열성 L - arabinose isomerase

(AI) 구조 연구

최 진 명

지도 교수: 이 성 행

생물신소재학과

조선대학교 일반대학원

*G. kaustophilus*는 높은 온도 (60℃) 에서 자랄 수 있는 호열성과 호기성 박테리아로 알려져 있다. L-arabinose isomerase (AI)는 L-arabinose를 L-ribulose로 전환할 뿐만 아니라 D-갈락토오스를 이용하여 D-타가토스의 생성을 위해 또한 이용될 수 있다. D-타가토스는 당 대용으로 낮은 열량을 함유하는 다이어트 음식으로의 이용과 함께 타입 2 당뇨병과 비만의 치료에도 이용되는 것으로 알려져 있다. 갈락토스를 이용한 타가토스 전환시, 중온성 AI와 달리 호열성 AI는 촉매작용을 위해 뚜렷한 메탈 의존성과 상승된 온도에서 열 안정성을 보인다는 점에서 주목되어 왔다. 그러나, 호열성 AI가 어떻게 분자수준에서 다른 기질 선호와 메탈 요구를 보이는지는 아직까지 분명하지 않게 남아있다. 그러므로, 이러한 의문을 해결하기 위해 apo-, holo-, 및 L-arabitol 결합된 호열성 *G. kaustophilus* AI의 결정 구조를 2.4Å, 2.3Å, 그리고 2.25Å으로 나타내었다. L-arabitol이 결합된 *G. kaustophilus* AI의 구조에서 기질

결합 포켓은 한 단량체의 N-말단 영역과 접촉하는 이웃 단량체의 C-말단에 의해 형성된다. 놀랍게도, Mn^{2+} 과 L-arabitol 이 결합된 *G. kaustophilus* AI에서 기질 결합면 근처의 몇몇 잔기의 중요한 이동을 보였다. 이러한 구조적인 데이터를 통해 상승된 온도에서 이성질화 반응은 메탈에 의해 조절되는 기질 결합 모델을 제안하였다. 더욱이, *E.coli*의 AI와의 밀접한 전체 구조 유사성에도 불구하고 *G. kaustophilus* AI는 여러 가지 면에서 내열성뿐만 아니라 기질 특이성의 혼란을 촉진하는 active site로 진화되었다.

ABSTRACT

Crystal structure of thermophilic L-arabinose isomerase from *Geobacillus kaustophilus*.

Choi, Jin Myung

Advisor: Prof. Lee, Sung Haeng, Ph.D.

Department of Bio-materials,

Graduate School of Chosun University

Geobacillus kaustophilus is a thermophilic bacteria that can grow at aerobic condition and high temperature, 60°C. L-Arabinose isomerase (AI) in bacteria that is known to catalyze interconversion of L-arabinose and L-ribulose can be also used for the production of D-tagatose by using D-galactose. D-tagatose is used for not only diet food of a low calorie but also treatment of type II diabetes and obesity. Unlike mesophilic AIs, thermophilic AIs showed the distinct metal dependence for their catalysis and thermostability at elevated temperatures. However, it still remains unclear how thermophilic AIs showed different substrate preferences and metal requirements at molecular levels. Therefore, in order to solve these questions, crystals of the apo, holo, and L-arabitol bound hexameric AI from *G. kaustophilus* were diffracted to 2.40 Å, 2.30 Å, and 2.25 Å respectively. The structure of GKAI bound L-arabitol was revealed that the substrate binding pocket is comprised of the N-terminal domain of one subunit and the C-terminal of contact neighboring subunit. Notably, Mn^{2+} and L-arabitol /GKAI underwent significant displacements of several residues near the substrate binding site comparing with apo GKAI. These structural data propose a metal-mediated substrate binding

model for the isomerization reaction at elevated temperatures. Furthermore, detailed inspection indicated that despite the close overall structural similarity to *E. coli* AI, GKAI has evolved distinct active site features that promote the promiscuity of substrate specificity as well as thermostability.

Introduction

Thermophilic bacteria usually can grow well at 50°C or more, a maximum of up to 70°C or more, and a minimum of about 40°C. *Geobacillus kaustophilus* is aerobic, thermophilic bacteria that can grow well at optimal temperature 60°C. These (hyper) thermophilic bacteria are generally known to have high thermostability and optimal activity at high temperature [69].

One of the enzymes relating to sugar metabolism in thermophilic bacteria, L-arabinose isomerases (EC 5.3.1.4), which is encoded by the *araA* gene of the arabinose operon, catalyzes isomerization of L-arabinose to L-ribulose as the first step in L-arabinose catabolism as well as conversion of D-galactose into D-tagatose [70-73]. D-tagatose produced by this isomerization is a rare natural hexoketose as low-calorie sugar with industrial interest. Over the past decade, some of arabinose isomerases were biochemically characterized and widely exploited to the production of D-tagatose not only as a new GRAS sweetener and health product [74-76], but also a new antidiabetic and obesity control drug [77]. Previously, many efforts have been made using microbial AIs to produce D-tagatose from D-galactose derived from lactose hydrolysis [70, 76, 78]. Nevertheless, the overall yield of product formation by mesophilic AIs from *Escherichia coli* (ECAI) and *Lactobacillus gayonii* was insufficient to produce D-tagatose in commercial scales [70, 79, 80]. Alternatively, a biological conversion of D-galactose to D-tagatose by using thermostable AIs from (hyper)thermophiles is advantageous with respect to feasibility of process, low viscosity of substrates, high conversion yields, and minimized microbial contamination [70, 75], because the isomerization of aldose to ketose is an endothermic reaction [75, 81]. In this respect, thermostable enzymes such as *Thermotoga neapolitana* AI [82], *T. maritima* AI [75], *Thermoanaerobacter mathranii* AI [83], and *Thermoanaerobacterium saccharolyticum* AI [84] would be more beneficial than other mesophilic counterparts for the high yields of D-tagatose production [79, 85-87]. Indeed, *T. neapolitana* AI (TNAI) and *T. maritima* AI (TMAI) exhibited isomerization activity of D-

galactose to D-tagatose with high yields of $\geq 60\%$ conversion at 80°C in the presence of divalent metal ions such as Co^{2+} or Mn^{2+} [75, 82, 88]. Notably, it was found that thermostable AIs from (hyper) thermophile are highly metal-dependent for their catalysis, which would be the distinct properties to mesophilic counterparts [89]. In addition, hyperthermophilic AIs exhibited promiscuity of substrate specificity as well as high metal dependence for their structural stability [88, 90].

Subsequently, mutational studies using chimeric AIs demonstrated that the metal dependent catalytic activity and thermostability of hyperthermophilic AIs at elevated temperatures might be ascribed to structurally region-specific evolution [91, 92], based on the fact that the amino acid sequences of (hyper) thermophilic AIs exhibited high similarity ($\geq 70\%$) to those of mesophilic AIs. Previously, the crystal structure of ECAI was determined as a trimeric architecture in the asymmetric unit [93]. Since the subunit structure of ECAI had been characterized as a hexamer with biophysical data and electron microscopy studies [94, 95], its three dimensional structure was constructed as hexamer using the crystal structure of L-fucose isomerase from *E. coli* (ECFI) [93, 96]. Although the primary structure of ECAI showed less than 10% identity to ECFI, their quaternary structures is quite similar [93]. Based on the comparative analysis, the substrate binding region and metal binding ligands of ECAI were proposed [93]. The proposed residues are highly conserved in the amino acid sequences of microbial AIs regardless of their optimal temperatures for catalysis [75, 93]. However, compared to mesophilic AIs that showed little activity for the isomerization of D-galactose and D-tagatose as the hexose substrate, promiscuity of substrate preference for thermophilic AIs that exhibit unusually high activity for D-galactose remains obscure.

With these and related reasons, structural information of thermostable AIs from (hyper) thermophiles is prerequisite for understanding the molecular basis of thermostability and substrate specificity of thermophilic AIs in comparison with mesophilic counterparts. This information will provide insights into catalytic activity and thermostability of thermostable AIs.

Very recently, *Geobacillus stearothermophilus* was attempted to predict the substrate binding site of a homologous AI with no three-dimension structure available using a structural genomics approach [97]. Such a bioinformatics-guided analysis proposed that residues located in the subunits of multimeric AI would be also significant for providing the catalytic cavity. Previously, AI (GKAI) from the thermophilic bacterium *Geobacillus kaustophilus* was chosen as the good model system which can grow optimally at 60°C [98-100]. In light of this, the *araA* gene from *G. kaustophilus* in *E. coli* was cloned, purified to get the recombinant L-arabinose isomerase. Furthermore, crystals of GKAI was determined its three-dimensional structure with and without Mn^{2+} and L-arabitol. Based on these results, the relationship between structure and function for AIs with respect to different substrate preference, metal requirements, and thermostability was discussed, underlying the molecular evolution of sugar isomerases for thermal adaptation at elevated temperatures.

Materials and Methods

1. Plasmid construct

GKAI plasmid cloned into pET28a (+) from *Geobacillus kaustophilus* was provided by lab of the professor Lee (Kyungpook National University, Daegu, Korea)

2. Transformation and protein overexpression

GKAI also was transformed into BL21 (DE3) on solid LB media containing 100 µg/ml kanamycin for protein expression. Transformed cells were cultured in 1.5L LB media with 100 µg/ml kanamycin until OD₆₀₀ value reaches 0.6-0.8. And then, cells were induced with 1 mM IPTG at 37°C for 5 hours and then harvested by centrifugation (5000 x g, 20 min, 4°C)

3. Protein purification by Ni-NTA agarose

The frozen cells (pellet to 1.5 L cultured cell) was slowly thawed and resuspended in 40 mL lysis buffer containing 20 mM Tris pH 7.5, 500 mM NaCl, 10 mM Imidazol, 1 mM PMSF, 1 mM DTT, 1 mM MgCl₂, and a little DNase 1. Resuspended cells were disrupted by sonication at 25% amplitude and pulse (20 sec, 20 sec) for 5 minute on ice and the disrupted cell was removed by centrifugation at 4°C (2000 x g, 30 min). The acquired supernatant was passed at 1 ml/min by econo-pump into column (Bio-Rad) packed with Ni-NTA agarose (Quiagen, USA). After binding, column was washed to 2 ml/min by washing buffer (20 mM Tris pH 7.5, 500 mM NaCl and 20 mM imidazole) and 6 x His-tagged protein was eluted by elution buffer (20 mM Tris pH 7.5, 500 mM NaCl, and 250 mM NaCl). Eluted protein was verified by color change with 1 x bradford solution

4. Dialysis and 6 x His-tag cutting by human α-thrombin

The eluted soluble protein was dialyzed by using a dialysis membrane MWCO 10 KD (Spectrum Laboratories, Inc. USA) in buffer (20 mM Tris pH 7.5, 150 mM NaCl) at 4°C for

overnight. The human α -thrombin (HTI, USA) used at 5000 x dilution with 2 mM CaCl_2 for cleavage reaction with 6 x His-tagged proteins at RT for 30 min-2 hours.

5. Size exclusion chromatography

GKAI directly performed by SEC pre-equilibrated final buffer (20 mM Tris pH 7.5, 50 mM NaCl) and collected at major peak. Fractionized peak region was identified by 12% SDS-PAGE.

6. Crystallization

The purified GKAI was concentrated up to 6~9 mg/ml in 20mM Tris-HCl buffer (pH 7.5 and 50 mM NaCl) with or without 0.5 mM Mn^{2+} . After screening, crystals for the two cases were initially grown from several conditions including Hampton SaltRx 2 No. 9 (1.8 M Na-K phosphate pH 6.9) and Wizard 1 No. 36 (0.1 M imidazol hydrochloric acid pH 8.0, 1 M sodium citrate tribasic) within 2~5 days. Apo-GKAI was crystallized only from the Hampton SaltRx 2 No. 9. Instead, GKAI crystal containing Mn^{2+} appeared from both of the conditions. In order to obtain crystals with better quality and size, further improvement trial was performed by varying the concentrations of salt, precipitants, pH values around the original solution and adding Hampton additive screening kit. The two types of Mn^{2+} -GKAI crystals having a shape of rod with an angle and a large log with a sharp point were typically grown from solutions containing 0.1 M imidazol hydrochloric acid pH 8.0, 1 M sodium citrate tribasic and 1.8 M Na-K phosphate pH 6.9, and 4% (v/v) acetone in a week, respectively. The ligand complex of GKAI was obtained by soaking Mn^{2+} -GKAI crystals with 10 mM of L-arabitol in the well solutions above overnight at 20°C. For data collection, crystals of Mn^{2+} and Mn^{2+} -ligand complex were immersed in a cryo-protectant solution of 30-35% of glycerol, and were flash-cooled in liquid nitrogen. Apo-form of GKAI was also crystallized from Hampton SaltRx 2 No. 9 (1.8 M Na-K phosphate pH 6.9), and the quality of the crystals improved with subtle pH changes pH 7.2 from 6.9. Although crystallized in the similar solution like metal containing GKAI, apo-form GKAI crystal was more stable in cryo-solution containing paratone-N.

7. X-ray crystallographic data collection, structure determination, and refinement

The crystal of GKAI apo-form diffracted to 2.4 Å resolutions and belonged to the space group $P2_12_12_1$. For Mn-GKAI crystals, two diffraction data sets with different space groups (C2 and $P2_12_12_1$) were collected from the two crystallization solutions as described previously [101]. Among them, Mn-GKAI selected 2.3 Å data set for the structure determination of Mn-GKAI, whose space group is C2 (Table 6). Crystal of Mn^{2+} -L-arabinose-GKAI complex was collected to 2.25 Å resolutions with a space group $P2_12_12_1$. These resulted in 45~48% crystal solvent content with Matthews coefficient of 2.25~2.37, from which these crystals contained appropriate six molecules in the asymmetric unit. The diffraction data were indexed and scaled with program HKL2000 (HKL Research) and diffraction statistics are shown in Table 5. All three structures of GKAI were solved by the molecular replacement (MR). All the MR calculations were performed using MOLREP, part of the CCP4 software suit. Model building and refinement of the MR solution obtained were carried out using the program COOT and the program PHENIX or REFMAC.

A MR solution of Mn-GKAI was obtained using the monomeric structure of apo E. coli L-arabinose isomerase monomer structure (ECAI) (PDB code: 2AJT) as a search model. This calculation reached values of 28.37% for R_{work} and 33.74% for R_{free} , thus indicating that the appropriate MR solution was obtained for the six molecules of GKAI. Based on the location of Mn^{2+} in ECAI, the observed Fo-Fc density map revealed the position of Mn^{2+} in the active site of GKAI. The following refinement and model building process further dropped the R_{work}/R_{free} values to 15.56 and 21.29%, respectively (Table 1).

To obtain MR solution for apo- and ligand complex structure of GKAI, one molecule from the Mn^{2+} -GKAI structure was used as a search probe. This MR brought appropriate R_{work}/R_{free} values to 22.36/27.47% for apo-form and finally converged to the values of 16.62/21.80 after several rounds of alternating manual rebuilding processes. The initial structure solution for ligand soaked GKAI was also successfully obtained with the statistics of 21.66/26.46%

($R_{\text{work}}/R_{\text{free}}$). In particular, the L-arabitol PDB file was created using the PHENIX program LIGANDFIT and fitted into Fo-Fc density that was consistent with the ligand during the building of L-arabitol-Mn²⁺-GKAI model building in COOT. Mn²⁺ ion was also found in the electron density map as the same way performed in metal-bound GKAI. From the further refinement, the $R_{\text{work}}/R_{\text{free}}$ of Mn²⁺-L-arabitol-GKAI form converged to the values of 15.45/20.10%. For all of the models, structural validation on the final model was performed and the crystallographic statistics are shown in Table 6.

8. Analytical size exclusion chromatography

To investigate oligomeric state of GKAI, GKAI was also used to same method as described in CagL. A mixture (1.5~2 mg/each protein in total 0.5 ml) of standard proteins (GE Healthcare HMW calibration kit) including tyroglobulin (669 kDa), ferritin (440 kDa) aldolase (158 kDa), conalbumin (75 kDa), and Ovalbumin (44 kDa) are prepared for a standard curve and used for calibration.

Results

1. Overexpression and purification of GKAI

To obtain recombinant L-arabinose isomerase from *Geobacillus kaustophilus*, pET28a (+)-GKAI was transformed into *E. coli* BL21 (DE3) with antibiotic. Purified protein of GKAI obtained by previous published method [101]. Size (~56 kDa) and purity of purified GKAI were analyzed by 12 % SDS-PAGE as shown in Fig. 19. 6 x His-tag uncut and cut of GKAI also revealed (Fig. 1). 6 x His-cleaved protein obtained through size exclusion chromatography by final buffer including 20 mM Tris pH 7.5 and 50 mM NaCl for crystallization.

2. Result of analytical size exclusion chromatography to GKAI

To analyze molecular weight of GKAI, the freshly purified protein was purified and loaded ~2 mg (500μl) by using size exclusion chromatography. Region of high peak revealed on position of 60-70 ml. The native molecular size of GKAI (Monomer: 56 kDa) was estimated as 336 kDa (six monomer) in solution by using HMW kit (GE Healthcare) containing standard proteins as described in Fig. 2. This result indicated that GKAI functions as a hexamer in solution.

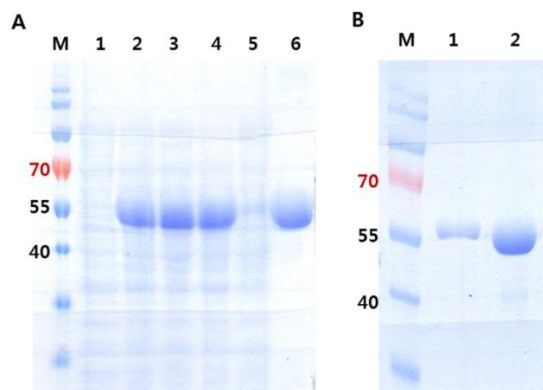


Figure 1. Size and purity of GKAI were analyzed by 12% SDS-PAGE. A. M: Marker, Lane1: uninduced, 2: induced, 3: supernatant, 4: after heat, supernatant, and 5: Flow through, 6: elution, B. M: Marker, Lane 1: uncut, 2: cut

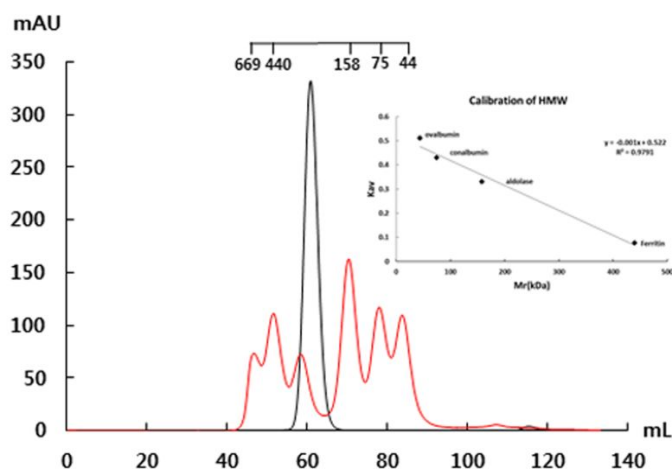


Figure 2. Molecular weight calibration of GKAI by analytical size exclusion using HMW calibration kits. GKAI used about 2 mg for calibration in 50 mM phosphate, pH 7.2, and 150 mM NaCl by 1ml/min. A mixture (1.5 ~ 2 mg/each protein in total 0.5 ml) standard proteins including tyroglobulin (669 kDa), ferritin (440 kDa) aldolase (158 kDa), conalbumin (75 kDa), and ovalbumin (44 kDa) are also used for a standard curve.

3. Overall structure of GKAI

Preliminary results of GKAI were already revealed in previously published journal [101]. Crystals of GKAI were obtained with apo-, metal-, and ligand-bound thermostable AI at resolutions of 2.40, 2.30 and 2.25 Å from *G. kaustophilus*, respectively. In the final structures, almost all of amino acids including metal ions and ligand were well fitted to the electron density map with reliable statistics (Table 1).

In GKAI structures, there are six protein molecules (M1, 2, 3, M1*, 2*, 3*), forming a sandwich-like homo-hexamer (Fig. 3). Unlike the crystal packing of mesophile ECAI, the asymmetric unit contains six monomer with two main symmetries; the 3-fold non-crystallographic symmetry between monomers in the homo-trimer in one side and 2-fold axis between homo-trimers as shown in Fig. 3. This seems to suggest that GKAI functions as a hexamer since the number of molecule in asymmetric unit or symmetry related crystal packing reflects the oligomeric state of protein in solution in some cases. This speculation was proven by biochemical study (i.e., analytical size-exclusion chromatography) (Fig. 2) although many of the previous reports expected thermostable AIs including TMAI and TNAI to form homo-tetramer instead of homo-hexamer that mesophile ECAI adopts [102-104]. GKAI hexamer seems to form as a dimer of trimer by the crystal packing-based *E. coli* AI hexamer formation model. However, since our structural and biochemical evidences correlate, the hexamer is rather homogeneous in solution and not exists in dynamic mixtures of hexamer and trimer in the cell. Also, it is to note that further structural studies on other several AIs from (hyper) thermophiles including *Alicyclobacillus acidocalcarius* and *Alicyclobacillus* sp. TP7 revealed the similar quaternary organization in biochemical and structural properties with those of GKAI.

A cartoon representation of the overall monomer structure of GKAI is shown in Fig. 3D. As expected from sequence homology and secondary structure (Fig. 4), GKAI subunit presents a similar fold with ECAI and ECFI [104, 105]. The r.m.s. deviations for Ca atoms and Z-scores

of DALI search [106] for apo-GKAI are 1.2 Å and 51.9 with apo-ECAI (PDB code 2AJT, 60% identity), and 2.9 Å and 31.8 with apo-ECFI (1FUI, 11% identity), suggesting that GKAI is likely to have similar domain organization with those sugar isomerases.

Indeed, GKAI monomer subunit appears to be divided into three domains; N-terminal domain (N terminus-Leu177), central domain (Lys178-Gly325), and C-terminal domain (Lys326-end) with a composition of 17 α -helices and 18 β -strands in secondary structure elements. A typical Rossmann fold that is structurally similar to nucleotide binding is found in the N-terminal domain (Fig. 3D), where five parallel β -strands (β 1- β 5) alternatively run with six α -helices (α 1- α 3 and α 5- α 6, α 4 is too short and half loose). Although overall secondary structure elements organization in central domain is similar to that of N-terminal, it only contains four parallel β -strands (β 6- β 9) and the way of linking within helices (α 6- α 14) and strands does not appear to be a Rossmann fold. In the C-terminal domain is formed an antiparallel β -barrel that is composed of six β -strands (β 10- β 11 and β 14- β 17). Note that six strands out of nine strands participate β -barrel formation in all GKAI C-terminal domain and only six strands among seven strands do in the corresponding domain of ECAI. The designated three domains of the monomeric GKAI make interfaces with two other monomers, in which cooperatively participate in the formation of active sites in a functional homo-hexamer.

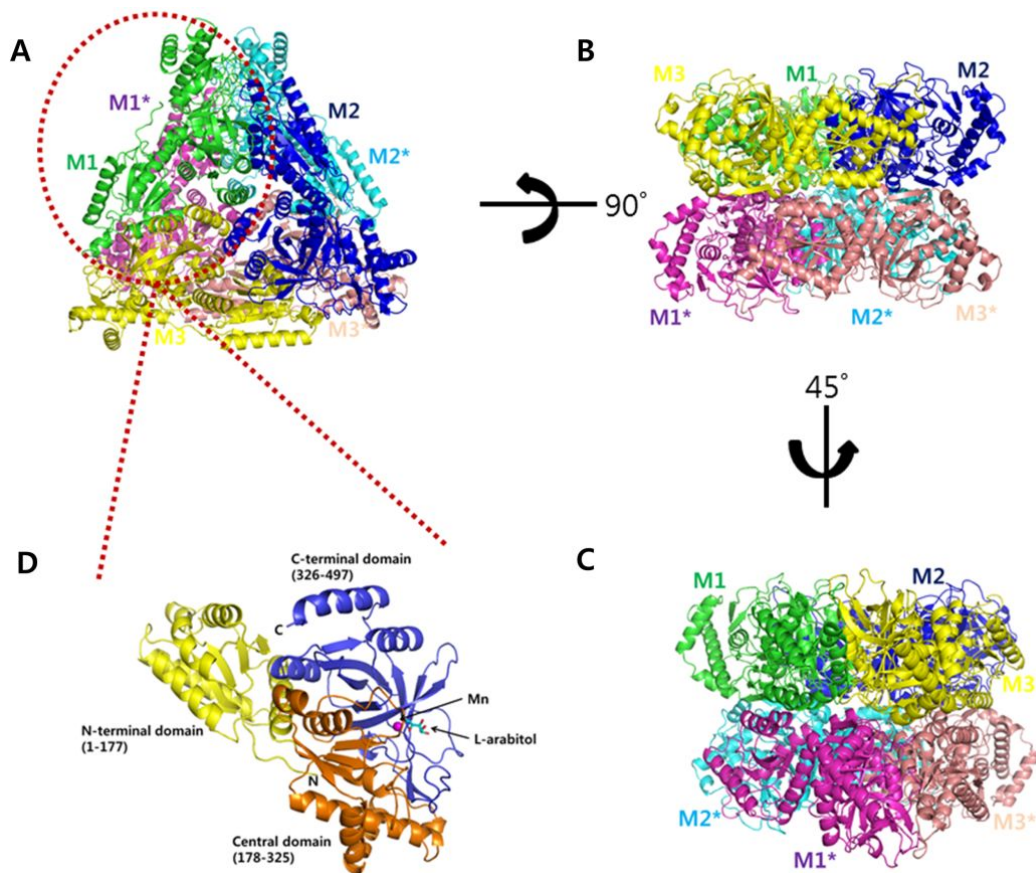


Figure 3. Overall structure of GKAI. **A.** GKAI contains six molecules in the asymmetry unit. Each molecule is represents as followed; M1 (green), M2 (blue), M3 (yellow), M1* (magenta), M2* (cyan), and M3* (light pink). **B.** hexamer GKAI rotated to 90° in x axis for side aspect. **C.** It is rotated 45° from B. **D.** Monomer structure of GKAI. Monomer consists of three domains (N-terminal domain: 1-177, central domain: 178-325, C-terminal domain (326-497) that is comprised of 16 α -helices and 18 β -strands.

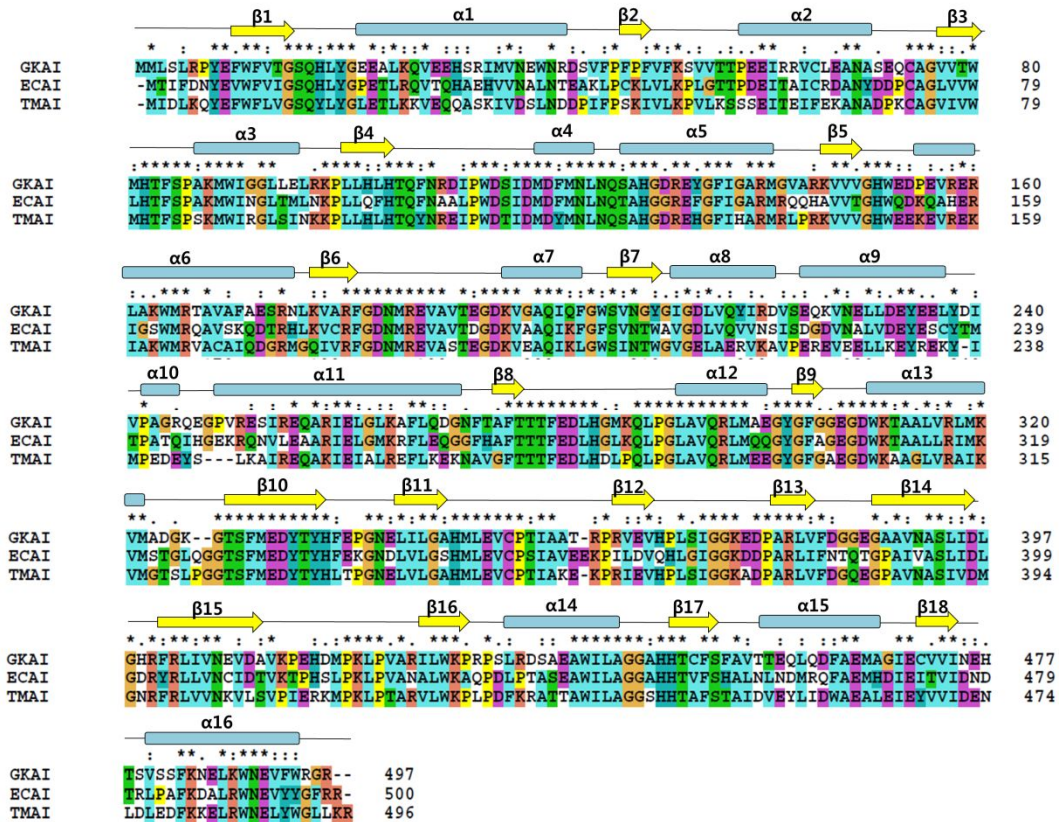


Figure 4. Sequence alignment of GKAI, ECAI, and TMAI.

ECAI (mesophile), GKAI (thermophile), and TMAI (hyperthermophile) are aligned sequence and represent secondary structure for comparison.

4. Interface of inter-subunit interaction

The six molecules in the hexamer are almost identical with R.M.S. deviations for main-chain atoms being 0.27 Å-0.62 Å in apo-GKAI, 0.20 Å-0.30 Å in holo-GKAI, and 0.18 Å-0.25 Å in GKAI/L-arabitol, suggesting that structural integrity of each subunit has been maintained during assembly.

Trimer in one side of the hexamer is made of head-to-tail contact with a monomer (Fig. 5A). Each subunit in the complex contains three parts respectively interacting two monomers in the same side (head-to-tail contact) and one from the other side. For all cases, hydrogen bondings and salt bridges play roles in the interface, but hydrophobic interaction was not found. Residues in C-terminal domain mainly involves in the interface formation with N-terminal domain in the neighboring subunit from same side plate, where residues in loops and β -sheets are found in the interface (Fig. 5 and Table 4). CASTP analysis revealed a major cavity (483.3 Å³) created by the contact between C-terminal domain and N-terminal domain of two subunits of apo-GKAI [107]. The SAS (solvent accessible area) of the surface (386.9 Å²) comprises 17 % of the total SAS (2285 Å²) [61]. This space is surrounded by secondary elements in the domains from adjacent subunits; Loop of β 6- α 7 from central domain, loop of β 12- β 13 and the β 10-strand from C-terminal domain are in contact with the α 4-helix and the loop of β 1- α 1 located in N-terminal domain in neighboring unit. The cavity can be divided into two regions in structure. First part turned out to be an active site, in which most of residues on loops from two subunits tend to coordinate with metal and ligand. However, second region extended from the active shows many of direct interactions between β -strands (part of β -barrel) from C-terminal domain and α -helix from N-terminal domain from each subunit (Fig. 5 and Table 4). Typically, this space is built by a combination of β -barrel and α 15 in C-terminal domain and α 5 from N-terminal domain. Inside the region, there are two important mode of inter-chain interactions were found. The penetration of a long side chain of Arg132 in α 5 of N-terminal domain allows it to make two bonds, which are salt bridge with Asp333 (in β 10) and hydrogen bond

with Tyr336 from C-terminal domain in neighboring monomer. Although away from the region, His105 (in $\beta 4$) in N-terminal domain projecting into the space interacts with Tyr336 via hydrogen bonding. Another residue His129 in $\alpha 5$ (on the same with Arg132 in the helix) is also hydrogen-bonded to Asp333. This “serial alternative or staggered chain interaction” (His105, His129, Arg132: Glu332, Asp333, Tyr336) seems to tighten the two subunits and support the active site architecture that is surrounded by loops (Fig. 5B). This kind of interaction mode is also found in other sugar isomerases including ECAI and ECFI. In addition, those amino acids are well conserved between members. Thus, the interaction might play a common role in the hexamer assembly in AIs.

Besides the head-to-tail interaction, each subunit comes in contact with a two monomers in the other plate of trimer. Since two plates (trimer per plate) are twisted with each other, loop of $\beta 6$ - $\alpha 7$ and $\alpha 7$ in central domain of one subunit place positions on top of the dimer junction (Fig. 5C). In this contact, the residues in the central domain in one subunit interact with two domains in dimer (one from central domain and N-terminal domain from each subunit) (Fig. 5C). Along with contact area between central domains are same residues in the same loop located face-to-face in 3D arrangement (Table 4 and Fig. 5C), where nine hydrogen bonds and four strong salt bridges (Arg181: Asp184, Asp184: Arg181, Lys196: Glu193, and Asp214: His284) were included. Two short helices ($\alpha 7$ and $\alpha 3$) run in parallel, contacting via salt bridge (Gln200: Lys88). Another important interaction interface in hexamer formation was found between N-terminal domains, in which $\alpha 2$ from each subunit runs in parallel and make contact via salt bridges and hydrogen bonding between Arg62, Asn69, and Glu72. Additional contacts were also made farther from the active site forming salt bridges between Lys88: Asn208 and Asp195: Arg140 in pairs.

In previous reports regarding mesophilic AI structures (ECAI and bacillus AI), the interactions between two trimers were not described in detail. Their crystallographic asymmetric unit contains only three monomers with equi-triangular arrangement, so that the

models tend to have limitation in explaining the bonding elements for the hexamer formation. However, our structures clearly show the mode of interactions between each subunit in the hexamer. Especially, the extensive salt bridges in combination with hydrogen bonds describes position around active site and seems to play roles in supporting and maintaining integrity of the site during functional protein assembly.

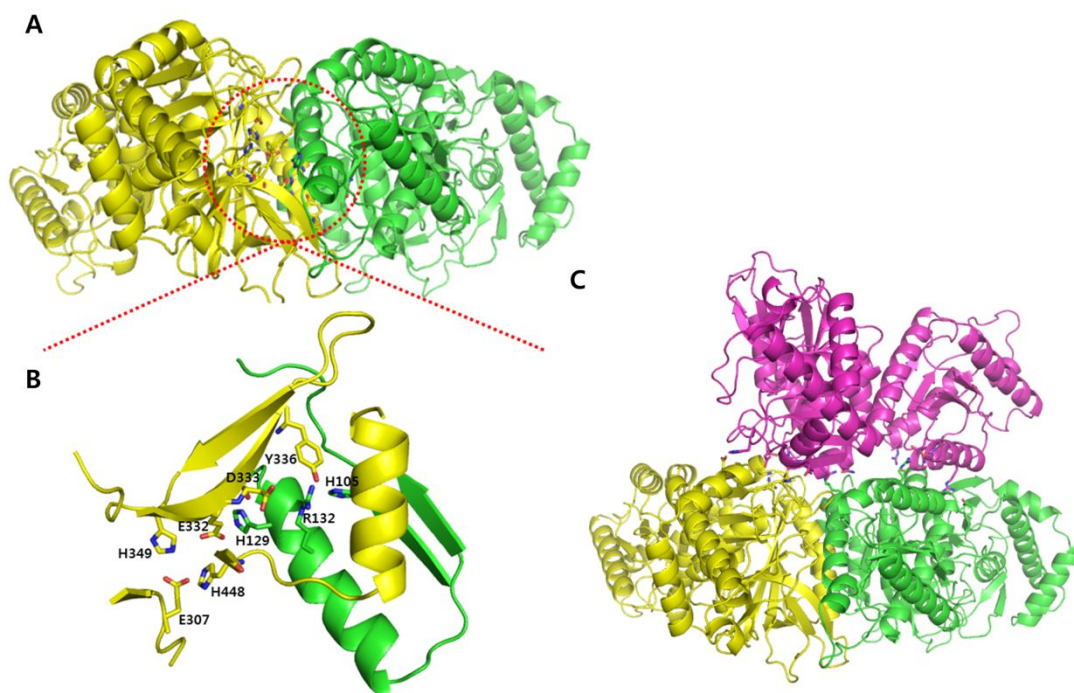


Figure 5. Interaction on the contact face forming cavity and interface between two monomer on one side and monomer on the other side.

A. Region in red dash circle represents residues forming cavity between two subunits. B. extended region from interface of both monomer show as serial alternative or staggered chain interaction. C. Between upper one subunit and down two subunits maintains hexamer assembly by interaction of contact residues

5. Active site

Although the corresponding region with secondary structure is similar in apo-ECAI that was suggested as a putative active site [104], our structures identified a substantial active site of AI with precise coordination between residues and metal and/or ligand. The substrate binding cleft of GKAI is solvent accessible and accommodates conserved residues that appear suited to perform catalytic and/or structural roles, which is consistent with the previously predicted substrate binding region for GSAI [108]. As stated, we separated them into two spaces in parallel inter-subunit contact of apo-GKAI. Among them, the cavity shared between N-terminal domain (in that α -helices ($\alpha 4$, $\alpha 5$), $\beta 1$ - $\alpha 1$ loop, and the $\beta 3$ - $\alpha 3$ loop) from one subunit (M1 subunit) and loops in central and C-terminal domain from the neighboring subunit (M3), is most likely to be a substrate-binding site. The latter part appeared to provide the most of the metal and substrate binding residues on $\beta 6$ - $\alpha 7$, $\beta 9$ - $\alpha 13$ and $\alpha 14$ - $\beta 17$.

The structure of the active site with bound metal ion is shown in Fig. 6C. Electron densities around six inter-subunit clefts successfully assigned divalent cation Mn with strong difference maps followed by the subsequent refinement. Considering the fact that Mn^{2+} was added to the purified apo-GKAI from harsh EDTA treatment for crystal screening, identifications of bound metal ions of GKAI were therefore refined as Mn^{2+} . The bound metal ion is located at the cleft region and coordinated by two of each His and Glu residues in β -strands (Fig. 6B, C). As show in Fig. 6B, the metal ions in GKAI are coordinated in a distorted tetrahedral form with four coordination bonds; they are Glu307, Glu332, His349, and His448. The atoms N ^{ϵ} of His349 and His 448 coordinate with Mn^{2+} within the distance of 2.54 Å and 2.45 Å, and the O ^{$\delta 2$} atoms of Glu307 and Glu332 are 2.43 Å and 2.47 Å, respectively. In addition, very similar metal ligands (i.e., His350, His450, Glu306, and Glu333) corresponding to these residues of GKAI were also found from ECAI (PDB:2AJT) [104], suggesting that these Glu307, Glu332, His349, and His448 from central and C-terminal domains comprise of the metal binding site of GKAI.

Our structures revealed fluctuation of the number of water molecules in active site undergoing catalysis (Fig. 6). The active site harbors in apo-enzyme six molecules in maximum, which are hydrogen bonded to hydrophilic residues including Glu307, His447, and Glu332. Mn ion appeared to replace the water molecule coordinated by four residues (Glu307, Glu332, His349, and His448), leaving hydrogen bond networks almost intact as in apo-form. During the substrate entry into the site, two water molecules were substituted with L-arabitol and remaining three waters appear to stabilize substrate by forming heavy hydrogen bond networks together with Mn ion. Comparison of the water molecule number in the active site of mesophilic AIs (ECAI and ECFI) presents clear differences in the water-mediated bindings. Active site of AIs from Mesophilic bacteria including ECAI, ECFI and Bacillus FI contains less number of water molecules (Table 2). Especially, active site in the structure of substrate-ECAI contains 2-3 less water molecules than those of GKAI, which suggests that GKAI might stabilize substrate binding with more hydrogen bonding network. In this regards, the water molecule content during the catalysis in thermophilic AI might give an insight about the protein stability at elevated temperature.

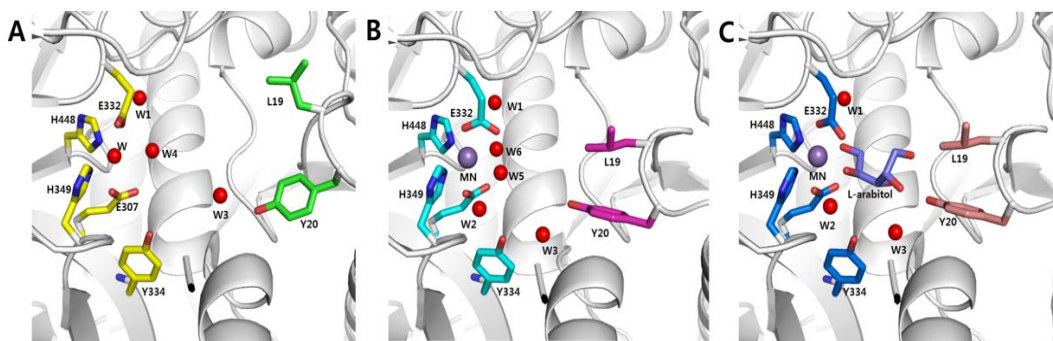


Figure 6. Active site of apo-, Mn-, and L-arabitol bound GKAI.

A. Active site of apo-GKAI. Apo-GKAI contains water instead of metal ion in active site forming four charged residues (Glu307, Glu332, His349, and H448) and contains 3 water molecules in cavity. B. Active site of Mn-GKAI. Active site contain Mn^{2+} ion with number of the increased water molecules. C. Active site of L-arabitol bound GKAI. In active site, L-arabitol interacts with Mn and is surrounded by some of residues. Water molecule decreases more two molecules than Mn-GKAI by entrance of L-arabitol.

6. Substrate-binding site

The structure of the substrate-binding site of GKAI with the bound L-arabitol is shown in Fig. 7A. As described above, each active site is located in a cleft between the C-terminal domain and the central domains of a subunit, adjacent to the other subunit interface (i.e., the N-terminal domain). The O1 atom and O2 of L-arabitol coordinates with Mn^{2+} (2.16 Å) and makes hydrogen bonds with His447 (3.41 Å) and W-1 (3.13 Å). Glu332 is located 2.95 Å from the O1 atom. The O2 atom of L-arabitol coordinates with Mn^{2+} (2.54 Å), and hydrogen bonds with W2 (2.77 Å). This coordination is likely to help fix the substrate in the proper conformation for the hydride or proton shift between the C1 and C2 of substrate [109]. It is likely that W1 and W2 are key water molecules responsible for the hydride or proton shift in the proposed catalytic mechanism, because they possibly make hydrogen bonds with the O1 and O2 of L-arabitol, respectively. W2 is also coordinated by Glu332 (3.79 Å). A water molecule (W3) coordinating to the O4 atom of L-arabitol forms double hydrogen bond of Tyr334 (2.88 Å) with neighbor Tyr20. C5 undergoes hydrophobic interactions with Phe280, Ile371, and *Leu19 (Leu19 from a neighboring subunit). Remarkably, O4 and O5 also make two hydrogen bonds with *Tyr20 (2.6 Å) and *Gln17 (2.94 Å), respectively, strongly indicating that the substrate binding coordination occurs cooperatively through the intersubunit interaction (Fig. 7).

As shown in Fig. 7, the crystal structures of GKAI and its complex with L-arabitol as the substrate analogue revealed that the enzyme undergoes large displacements of several amino acids reflecting conformational changes in the course of catalysis. In the presence of Mn^{2+} , Tyr20 and Leu19 of holo GKAI is proximal to O4 of L-arabitol at 2.08 Å and 3.56 Å, respectively. On the other hand, when the structure of holo GKAI with L-arabitol is superimposed with apo GKAI without L-arabitol, the corresponding Leu19 and Tyr20 residues of apo GKAI is located at 7.02 Å and 5.17 Å away from the C5 and O4 atoms of L-arabitol. This means that GKAI in the presence of Mn^{2+} can recognize substrates with different

configurations of C4, C5, and O5 by using Tyr20, Leu19 and Gln17, and vice versa. However, in the absence of Mn^{2+} , the substrate recognition at the 4 and 5 positions of L-arabitol is thought to be less favourable for apo GKAI than that of L-arabitol for holo GKAI, because the enzyme activity of apo enzyme for L-arabinose is significantly lower than that of holo enzyme. C5 makes hydrophobic interactions and also O5 is in a hydrophobic environment created by **Leu19, *Phe280, and *Ile371, without any hydrogen bond with amino acid residues. The hydroxyl group, O5, in hydrophobic environment increases structural energy of the substrate-enzyme complex, leading low affinity of substrate to the enzyme. In each subunit, each monomer undergoes conformational change. In accordance with the fact that thermostable AIs showed high dependence of metal ions for their catalysis, GKAI exhibited sequential, ordered kinetics: Mn^{2+} binds first, followed by L-arabinose. Previously, it has been reported that in contrast to mesophilic *Bacillus halodurans* AI, thermophilic *Geobacillus stearothermophilus* AI undergoes conformational changes by Mn^{2+} at elevated temperatures [110]. Therefore, GKAI was suggested that the binding of Mn^{2+} to GKAI converts the inactive form of the enzyme into an active form at elevated temperatures.

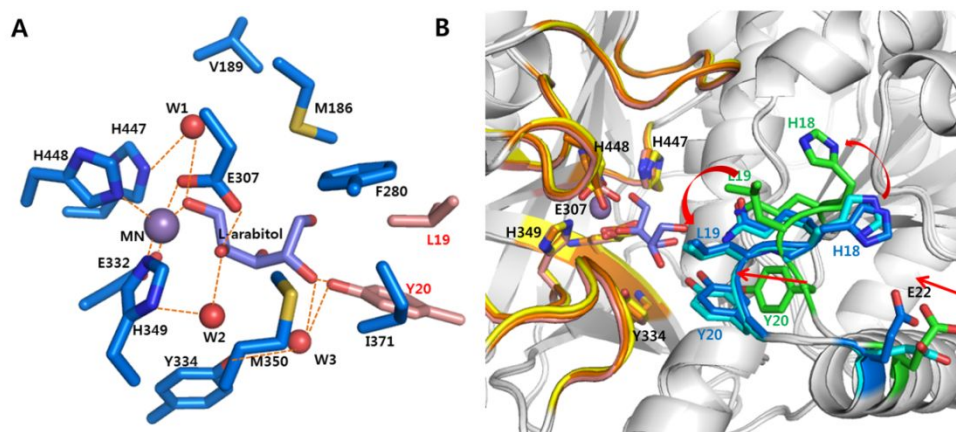


Figure 7. Interaction of Mn and L-arabitol by water molecules and surround residues in active site and residues of active site changed by Mn and L-arabitol binding. A. In addition to Mn coordinated by surrounding residues, L-arabitol also interacts with not only with Glu307, W1, and W2 at one subunit but also Tyr20 and W3 of neighboring subunit. B. Superimposition of three structures (apo-, Mn-, and L-arabitol bound GKAI) can show conformation change by Mn and substrate binding. Loop containing Leu19 and Tyr20 of neighboring subunit move to active site direction by Mn coordination for preparation of substrate binding.

7. Thermostability of thermophilic AI.

GKAI optimally catalyzes its substrate at around $\sim 60^{\circ}\text{C}$ [69], indicating that the thermophilic enzyme maintains its overall architecture (hexameric) and integrity of active site at the elevated temperature. Not to mention catalysis of this protein, the resistance to such a high temperature has become an open question to general protein chemistry field. Although proper answers are not available yet, there are factors that are involved with the thermostability of enzymes from (hyper) thermophiles; amino acid sequence composition and changed environment of active site surface [69, 111, 112]. These two factors are tightly associated with the proportion of charged residues (especially arginine) on interfaces and water-mediated hydrogen bond in active site. In this regards, our structures have great advantages to investigate not only for proving the general principles of thermostability but also for the application of basic knowledge for industry since our structures are the first successful in AI research area in comparison with apo form of mesophilic AI (ECAI) structure. With these in mind, comparative arginine residues compositions were first searched, resulting in that GKAI arginine contents is 2% higher than ECAI and other mesophiles including ECFI and *Sereptococcus* FI etc (Table 3). Therefore, some of non-conserved arginine residues (Fig. 8, sequence alignment) between the two AIs might be possibly located at subunit interface to make extra bonds that are not comparable in mesophilic AI.

Finally, three out of ten Arg residues appeared to contribute to subunit interactions. Arg62 and Arg98 in N-terminal domain of one subunit (naming, M1) make contact with Glu72 and Glu96 in N-terminal domain of the other subunit (naming M1*), resulting in the formation of additional four salt bridges since two trimeric units inversely contact. Another Arg145 (M1) contributes to hydrogen bond with peptide carbonyl oxygen of Leu442 (M3). It has been suggested that more salt bridges/ion pairs (network of salt bridges) and hydrogen bonds are key for an increased thermostability [113, 114].

Interestingly, the breakdown of two inter-subunit salt bridges was enough to lead to an

adverse effect of on the thermostability of hexameric thermophilic glutamate dehydrogenases [115]. Furthermore, The free energy required for stabilization of cellular protein is known to be around ~ 50 kJ/mol, and a single salt bridge contribute ~ 22 kJ/mol to protein folding that remains unaffected even at extreme temperature more than hydrophobic interaction [116, 117]. Based on the reports, GKAI in comparison with ECAI is likely to be more stable at least as amount of ~ 100 kJ/mol free energy of folding, which might be enough for GKAI to survive at the elevated temperature (~ 70 °C).

As stated, there was difference in the number of water molecules in active site; GKAI contains 2-3 more than ECAI while undergo catalytic process. The hydrogen bond in active site has not been identified with sufficient resolutions, leaving the correlation between the role of hydrogen bond and thermostability unclear. However, our structures revealed that the hydrogen bonding network dynamics during catalysis in active site in details with acceptable resolutions. Recently, van Bueren et al., showed that changes of residues in active site accompanied additional water-mediated hydrogen bond between the residues and substrate of thermophile xylanase [118]. This might suggested that the novel hydrogen network formation stabilize the active site during catalysis. In this regards, our structures may provide strong evidences of the role of water molecules for the maintenance of active site integrity at high temperatures. As described, there are 3 additional water-mediated hydrogen bonds in GKAI structure bound L-arabitol. Those bonds mainly interconnect ligand with residues while substrate enters. Although some of those residues are conserved (Leu19 and Tyr20 in GKAI corresponding to Leu16 and Tyr17) in alignment, their side chain only in GKAI indeed underwent significant conformational changes, resulting in the interaction with substrate via hydrogen bond with water. Besides those, GKAI was found to have new water-mediated hydrogen bonds to substrate that are not observed in ECAI. Since an individual hydrogen bond is known to contribute an average of 5.5 kJ/mol (~ 1.3 Kcal/mol) to the stabilization [119], more energy than ECAI seems to stabilize the transition state of substrate in GKAI than mesophilic AI during

catalytic process. Although our case added more example of the potential of hydrogen bond mediated substrate thermal stabilization, further investigation is definitely needed with additional structural studies on thermophilic enzymes.

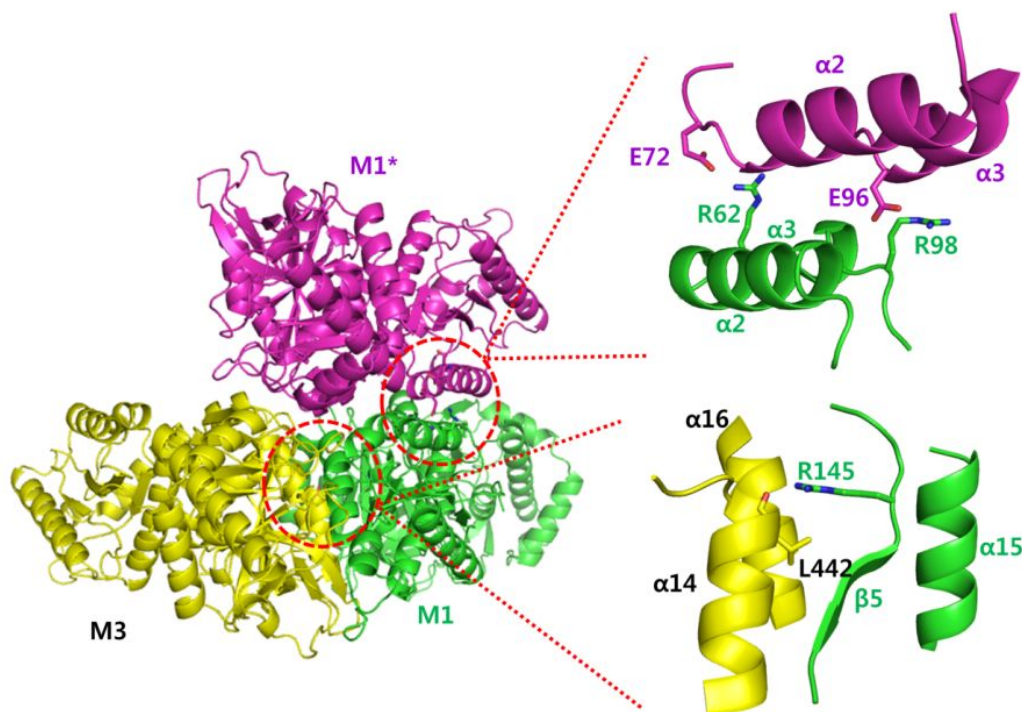


Figure 8. Interface interaction for thermostability non-existing other AI such as ECAI. Arg145 of M1 interacts with Leu442 of M3 on interface. And Arg62 and Arg98 of M1 also forms salt bridges with Glu72 and Glu96 of M1* (Two region represent as red dash line).

8. Different substrate preference

On the basis of structure of GKAI/ L-arabitol, the enzyme-substrate interactions of L-arabinose and D-galactose can be deduced. Since L-arabinose and D-galactose have the same configurations of asymmetric carbon atoms as L-arabitol, substrate-enzyme interactions are thought to be almost identical between them, but enzyme activities for them are significantly different. The structural variation between L-arabitol and L-arabinose is found at position 5, where L-arabinose has a hydroxyl group instead of an aldehyde group and D-galactose has no group. The simple modeling to generate the bound l-mannose by adding a hydroxyl group to L-arabitol at the 6 position indicated that the O6 atom can be located in a staggered conformation without any unusual short contacts with amino acid residues; the torsion angle of C4-C5-C6-O6 is -60° . In the resultant model of the bound L-arabitol, O6 is in hydrophobic environment created by *Phe66, Trp104, and Phe131, and it may decrease the affinity of substrate to enzyme, as described in detail below. In the binding of D-galactose, which has no group at the 6 position, a lack of hydrophobic interactions of C6 with the enzyme found in GKAI /L-arabitol may lead to lower enzyme activity for D-galactose.

Displacements of several amino acid residues restructures of the substrate binding site so that D-galactose as the substrate are more accessible to the mobilization. Due to the structural similarity of its substrates, which may be explained by the sharing of hydroxyl group at C3-C4 in a cis-configuration, AI can also catalyze the isomerization of D-galactose to D-tagatose [120].

Table 1. Statistics for data collection and refinement of GKAI

	Apo-GKAI	Mn-GKAI	L-arabitol GKAI
Data collection			
Beam line	PAL5C,7A		
Space group	P2 ₁ 2 ₁ 2 ₁	C2	P2 ₁ 2 ₁ 2 ₁
Wavelength(Å)	1.00150	0.97951	1.00001
Unit-cell parameters(Å°)	a=116.725, b=140.746, c=215.818 α=β=γ=90°	a=223.659, b=153.120, c=91.239 α=γ=90°, β=103.850°	a=118.514, b=146.263, c=215.677 α=β=γ=90°
Resolution(Å)	50-2.4(2.44-2.4)	50-2.3(2.34-2.3)	50-2.25(2.29-2.25)
Total reflections	982633	629574	2123676
Unique reflection	133329	131054	175982
Completeness (%)	96.2(94.1)	99.5(99.1)	99.2(99.4)
Redundancy	7.4(7.1)	4.8(4.7)	12.1(11.4)
R-merge† (%)	11.5(45.9)	14.4(83.3)	11.2(45.9)
Average I/σ (I)	14.091(3.746)	18.33(2.81)	20.576(5.286)
Solvent content (%)	45.41	45.31	48.23
Matthews coefficient (Å ³ Da ⁻¹)	2.25	2.25	2.37
Refinement			
R _{work} / R _{free} (%)	16.62/21.80	15.56/21.29	15.45/20.10
Residues/Waters	2970/1187	2962/1172	2972/1645
Mn		6	6
L-arabitol			6
RMSD			
Angle (°)	1.094	1.097	1.071
Length (Å)	0.008	0.008	0.007
Average B-factors (Å ²)	8.6	13.80	10.20
Ramachandran plot			
Most favored regions (%)	96.69	97.15	97.33
Allowed regions (%)	3.31	2.64	2.5
outliers	0	0.20	0.17

Values in parentheses correspond to highest resolution shell.

†R-merge= $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl) \rangle$ is the average intensity of symmetry-related observations.

Table 2. Number of water containing in active site at both GKAI and ECAI

Subunit	Apo_ GKAI (ECAI)	Mn_ GKAI (ECAI)	Ara_ GKAI (rib_ ECAI)
M1	6 (3)	7 (4)	3 (2)
M2	6 (3)	5 (2)	3 (2)
M3	4 (1)	5 (3)	3 (0)

Table 3. Comparison of amino acid composition between GKAI and other AI (FI)

Amino acid composition	L-arabinose isomerase from <i>G.kaustophilus</i>	L-arabinose isomerase from <i>E.coli</i>	L-Fucose isomerase from <i>E.coli</i>	D-arabinose isomerase from <i>Bacillus</i>
Ala (A)	40 (8%)	42 (8.4%)	64 (10.8%)	48 (8.1%)
Arg (R)	33 (6.6%)	23 (4.6%)	32 (5.4%)	28 (4.7%)
Asn (N)	17 (3.4%)	22 (4.4%)	24 (4.1%)	30 (5.0%)
Asp (D)	25 (5%)	33 (6.6%)	32 (5.4%)	35 (5.9%)
Cys (C)	5 (1%)	7 (1.4%)	12 (2.0%)	8 (1.3%)
Gln (Q)	16 (3.2%)	26 (5.2%)	21 (3.6%)	13 (2.2%)
Glu (E)	46 (9.3%)	25 (5.0%)	40 (6.8%)	46 (7.7%)
Gly (G)	42 (8.5%)	39 (7.8%)	53 (9.0%)	60 (10.1%)
His (H)	16 (3.2%)	21 (4.2%)	16 (2.7%)	6 (1.0%)
Ile (I)	21 (4.2%)	24 (4.8%)	35 (5.9%)	36 (6.1%)
Leu (L)	41 (8.2%)	48 (9.6%)	41 (6.9%)	39 (6.6%)
Lys (K)	20 (4.0%)	22 (4.4%)	25 (4.2%)	41 (6.9%)
Met (M)	18 (3.6%)	17 (3.4%)	22 (3.7%)	19 (3.2%)
Phe (F)	27 (5.4%)	23 (4.6%)	17 (2.9%)	23 (3.9%)
Pro (P)	21 (4.2%)	20 (4.0%)	23 (3.9%)	27 (4.5%)
Ser (S)	22 (4.4%)	17 (3.4%)	35 (5.9%)	22 (3.7%)
Thr (T)	21 (4.2%)	33 (6.6%)	33 (5.6%)	35 (5.9%)
Trp (W)	13 (2.6%)	11 (2.2%)	14 (2.4%)	11 (1.8%)
Tyr (Y)	10 (2.0%)	11 (2.2%)	17 (2.9%)	27 (4.5%)
Val (V)	43 (8.7%)	36 (7.2%)	35 (5.9%)	41 (6.9%)
Charged residues	71 (-) + 53 (+) (24.9%)	58 (-) + 45 (+) (20.6%)	72 (-) + 57 (+) (21.8%)	81 (-) + 69 (+) (25.2%)
Total	497	500	591	595

Table 4. A. Interaction of apo-GKAI subunits

M1-M2	<p>Hydrogen bond</p> <p>Glu188 (O)-Ala87 (N), Lys88 (N) : 3.52, 2.96, Ala190 (N)-Glu133 (OE2) : 2.84 Val191 (N)-Glu133 (OE1) : 2.94 Thr335 (OG1)-Gln107 (OE1), Asn125 (O) : 3.0, 3.0 Tyr366(OH)-His105 (NE2), Ala128 (O), Arg132 (NH1) : 2.60, 3.83, 2.88 Tyr366 (O)-Gln107 (NE2) : 2.96 Asp333(O)-His129 (ND,ND1) : 3.48, 2.67 Glu438 (OE1)-Arg160 (NH1, NH2) : 3.58, 3.30 Ile441(O), Leu442(O)-Arg145(NH1) : 3.08, 3.26 Arg488 (NH2)-Glu465 (OE1) : 3.40 Phe338 (N)-Arg110 (O) : 2.67 Trp489 (NE1)-Asn490 (OD1) : 2.85</p> <p>Salt bridge</p> <p>Glu332 (OE2)-His129 (ND1, NE2) : 3.67, 3.71 Asp333 (OD2)-Arg132 (NE, NH1, NH2) : 3.88, 3.72, 3.90 His337 (ND1)-Asp111 (OD1) : 3.97 Glu438 (OE1)-Arg160 (NH1,NH2) : 3.30, 3.58 Arg488 (NH2), Lys484 (NZ)-Glu465 (OE1): 3.39, 3.40</p>
M2-M3	<p>Hydrogen bond</p> <p>Glu188(O)-Ala87(N), Lys88(N) : 2.92, 3.46 Ala190 (N)-Glu133 (OE2) : 2.74 Val191(N)-Glu133 (OE1) : 3.06 Asp333 (O)-His129 (ND1) : 2.73 Thr335 (OG1)-Gln107(OE1), Asn125 (O) : 2.94, 2.85 Tyr336(N)-Gln107 (OE1) : 2.99 Tyr336 (OH)-Ala128 (O), His105 (NE2), Arg132 (NH1) : 3.89, 2.70, 2.30 Phe338 (N)-Arg110 (O) : 2.77 Lys419 (NZ)-Ser116(O), Asp118 (OD2) : 2.98, 3.24 Ile441 (O)-Arg145 (NH1) : 3.15 Arg488 (NH2)-Glu465 (OE1) : 3.32 Trp489 (NE1)-Asn490 (OD1) : 2.95 Glu491 (OE2)-Arg497 (NH1) : 3.77 Glu491 (OE1)-Arg497 (NH2) : 3.27</p> <p>Salt bridge</p> <p>Glu332 (OE2)-His129 (ND1, NE2) : 3.32, 3.41 Asp333 (OD1)-Arg132 (NE) : 3.70 Asp333 (OD2)-Arg132 (NE, NH1, NH2) : 3.68, 3.75, 3.50 Lys419 (NZ)-Asp118 (OD2) : 3.24. Lys484 (NZ), Arg488 (NH2)-Glu465 (OE1) : 3.93, 3.32 Glu491 (OE1)-Arg497 (NH2) : 3.27 Glu491 (OE2)-Arg497 (NH1, NH2) : 3.77, 3.97</p>
M3-M1	<p>Hydrogen bond</p> <p>Glu188(O)-Ala87(N), Lys88(N) : 3.41, 2.95 Glu188(OE2)-His18(NE2) : 3.18 Ala190(N)-Glu133(OE2) : 2.71 Val191(N)-Glu133(OE1) : 3.04 Asp333(O)-His129(N, ND1) : 3.45, 2.79 Thr335(OG1)-Gln107(OE1), Asn125(O) : 3.21, 2.98 Tyr336(N)-Gln107(OE1) : 2.92 Tyr336(OH) – His105(NE2),Ala128(O), Arg132(NH2) : 2.48, 3.72, 2.69 Thy336(O) – Gln107(NE2) : 2.96 Phe338(N)-Arg110(O) : 2.83 Glu438(OE1)-Arg160(NH1) : 3.62 Glu438(OE2)-Arg160(NH2) : 3.46 Ile441(O), Leu442(O)-Arg145(NH1) : 3.18, 3.22 Arg488(NH2)-Glu465(OE1) : 3.73 Trp489(NE1)-Asn490(OD1) : 2.90</p> <p>Salt bridge</p> <p>Glu188(OE2)-His88(NE2) : 3.18 Glu332(OE2)-His129(ND1,NE2) : 3.52, 3.58 Asp333(OD2)-Arg132(NE, NH1, NH2) : 3.97, 3.78, 3.85 Arg434(NE, NH1)-Glu156(OE1) : 3.92, 3.99 Glu438(OE1)-Arg160(NH1,NH2) : 3.62, 3.68 Glu438(OE2)-Arg160(NH2) : 3.46 Lys484(NZ), Arg(488) - Glu465(OE1):3.91, 3.73 Glu491(OE1,OE2)- Arg497(NH1) : 3.13, 3.06</p>
M1-M1*	<p>Hydrogen bond</p> <p>Arg62(NH1)-Glu72(OE1) : 2.67 Arg62(NH2)-Asn69(O) : 3.61</p>

	Asn69(O)-Arg62(NH2) : 3.25 Glu72(OE1)-Arg62(NH1) : 3.24 Lys88(NZ)-Val207(O), Asn208(OD1) : 2.87, 2.93 Arg140(NH2)-Asp195(OD2) : 2.79 Gln200(NE2)-Lys88(O) : 3.36 3.11 Salt bridge Arg62(NH1)-Glu72(OE1,OE2) : 2.67, 3.81 Arg98(NH1)-Glu96(OE1) : 3.40 Asp195(OD2)-Arg140(NH2);2.79	Asn69(ND2)-Glu96(OE2) : 3.50 Lys88(O)-Gln200(NE2) : 3.34 Arg98(NH1)-Glu96(OE1) : 3.40 Asp195(OD2)-Arg140(NH2) : 2.79 Val207(O), Asn208(OD1)-Lys88(NZ): 2.90, 3.11 Glu72(OE1, OE2)-Arg62(NH1): 3.24, 3.70 Arg140(NH2)-Asp195(OD2) : 2.79
M3-M1*	Hydrogen bond Arg181(NH1)-Asp184(OD2) : 2.87 Arg187(NH1)-Gly209(O) : 2.72 Lys196(NZ)-Glu193(OE1) : 3.85 Gly209(OD1)-Arg187(NH1) : 2.98 His284(NE2)-Asp214(OD1) : 2.67 Salt bridge Arg181(NH1)-Asp184(OD2) : 2.87 Lys196(NZ)-Glu193(OE1) : 2.87 His284(NE2)-Asp214(OD1) : 2.57	Asp184(OD2)-Arg181(NH1) : 2.71 Glu188(N) - Asn208(OD1) : 2.79 Asn208(OD1)-Glu188(N) : 2.93 Asp214(OD1)-His284(NE2) : 2.57 Asp184(OD2)-Arg181(NH1) : 2.71 Asp214(OD1)-His284(NE2) : 2.57

B. Interaction of holo-GKAI subunits

M1-M2	Hydrogen bond Glu188(O) - Ala87(N), Lys88(N) : 3.55, 2.88, Val191(N) - Glu133(OE1) : 3.24 Thr335(OG1) - Gln107(OE1), Asn125(O) : 3.29, 2.92 Tyr336(OH) - His105(NE2), Arg132(NH2) : 2.83, 2.70 Tyr336(O) - Gln107(NE2) : 3.10 Ile441(O) - Arg145(NH1,NH2) : 2.98, 3.11 Asn485(ND2)-Glu486(OE2) : 3.46 Trp489(NE1) - Asn490(OD1) : 2.93 Salt bridge Asp333(OD1) - Arg132(NE, NH1) : 3.55, 3.42 Asp333(OD2) - Arg132(NE, NH1) : 3.70, 3.92 Arg488(NH1, NH2) - Glu465(OE2) : 3.74, 3.27	Ala190(N) - Glu133(OE2) : 3.01 Asp333(O) - His129(ND1) : 2.91 Phe338(N)-Arg110(O) : 2.69 Ile442(O)-Arg145(NH1):3.05 Arg488(NH2) - Glu465(OE2) : 3.27
M2-M3	Hydrogen bond Glu188(O) - Ala87(N),Lys88(N) : 3.35, 2.84 Val191(N) - Glu133(OE1) : 3.18 Thr335(OG1) - Gln107(OE1), Asn125(O) : 3.04, 2.99 Tyr336(N) - Gln107(OE1) : 2.92 Tyr336(OH) - Ala128(O), His105(NE2), Arg132(NH2) : 3.81, 2.60, 2.67 Phe338(N) - Arg110(O) : 2.67 Arg434(NH2)-Glu156(OE2) : 3.88 Ile441(O) -Arg145(NH2) : 3.18 Arg488(NH2) - Glu465(OE2) : 3.77	Ala190(N) - Glu133(OE2) : 2.78 Asp333(O) - His129(ND1) : 2.87 Arg434(NE)-Glu156(OE1) : 3.24 Glu438(OE1)-Arg160(NH1) : 3.62 Leu442(O)-Arg145(NH1) : 3.20 Trp489(NE1) - Asn490(OD1) : 2.95

	Salt bridge Glu332(OE2) - His129(ND1,NE2) : 3.29, 3.85 Asp333(OD1) - Arg132(NE, NH1, NH2) : 3.55, 3.42, 3.33 Arg434(NE) - Glu156(OE1,OE2) : 3.24, 3.85 Glu438(OE1,OE2) - Arg160(NH1) : 3.62, 3.71 Arg488(NH2) - Glu465(OE2) : 3.77
M3-M1	Hydrogen bond Glu188(O)-Ala87(N), Lys88(N) : 3.55, 3.01 Ala190(N)-Glu133(OE2) : 2.86 Val191(N)-Glu133(OE1) : 3.08 Asp333(O)-His129(N, ND1) : 3.47, 2.73 Thr335(OG1)-Gln107(OE1), Asn125(O) : 3.14, 3.24 Tyr336(N)-Gln107(OE1) : 2.92 Tyr336(OH) - His105(NE2, Arg132(NH2)) : 2.81, 2.67, Thy336(O) - Gln107(NE2) : 2.81 Phe338(N)-Arg110(O) : 2.69 Arg434(NH2)-Glu156(OE2) Ile441(O)-Arg145(NH1,NH2) : 3.27, 3.3.07 Leu442(O)-Arg145(NH1) : 3.03 Arg488(NH2)-Glu465(OE1,OE2) : 3.32, 3.20 Trp489(NE1)-Asn490(OD1) : 2.93 Salt bridge Glu332(OE2)-His129(ND1,NE2) : 3.70, 3.43 Asp333(OD2)-Arg132(NE) : 3.83 Asp333(OD1)-Arg132(NE,NH1,NH2) : 3.61, 3.15, 3.24 Arg434(NE)-Glu156(OE1) : 3.43 Arg434(NH2)-Glu156(OE1,OE2) : 3.43, 2.87 Arg488(NH2)-Glu465(OE1,OE2) : 3.32, 3.20 Glu491(OE1,OE2)- Arg497(NH1) : 3.13, 3.06
M1-M1*	Hydrogen bond Arg62(NH1)-Glu72(OE1) : 3.12 Asn69(O)-Arg62(NH2) : 3.73 Glu72(OE1)-Arg62(NH1) : 3.16 Lys88(O)-Gln200(NE2) : 3.27 Lys88(NZ)-Val207(O), Asn208(OD1) : 3.18, 3.31 Glu96(OE1)-Arg98(NH1) : 3.79 Arg98(NH1)-Glu96(OE1) : 3.19 Arg140(NH2)-Asp195(OD2) : 2.57 Asp195(OD2)-Arg140(NH2) : 2.66 Gln200(NE2)-Lys88(O) : 3.47 Val207(O)-Lys88(NZ) : 3.09 Salt bridge Arg62(NH1)-Glu72(OE1,OE2) : 3.12, 3.81 Arg62(NH2)-Glu72(OE1) : 3.90 Glu72(OE1)-Arg62(NH1,NH2) : 3.16, 4.00 Glu96(OE1,OE2)-Arg98(NH1):3.79, 3.89 Arg98(NH1)-Glu96(OE1) : 3.19 Arg140(NE, NH2)-Asp195(OD2) : 3.94, 2.57 Asp195(OD2)-Arg140(NH2);2.66
M3-M1*	Hydrogen bond Arg181(NH1)-Asp184(OD2) : 2.98 Asp184(OD2)-Arg181(NH1) : 2.91 Arg187(NH1)-Gly209(O) : 2.97 Glu188(N) - Asn208(OD1) : 2.83 Asn208(OD1)-Glu188(N) : 2.80 Gly209(OD1)-Arg187(NH1) : 2.82 Asp214(OD1)-His284(NE2) : 2.68 His284(NE2)-Asp214(OD1) : 2.48 Salt bridge Arg181(NH1)-Asp184(OD2) : 2.98 Asp184(OD2)-Arg181(NH1) : 2.91 Asp214(OD1)-His284(NE2) : 2.68 His284(NE2)-Asp214(OD1) : 2.48 Gly496(O)-Arg495(NE,NH2) : 3.79, 3.61

C. Interaction of L-arabitol bound GKAI subunits

M1-	Hydrogen bond
-----	----------------------

M2	<p>Glu188(O) - Ala87(N), Lys88(N) : 3.39, 2.92, Ala190(N) - Glu133(OE2) : 2.78 Val191(N) - Glu133(OE1) : 3.06 Asp333(O) - His129(ND1) : 2.69 Thr335(OG1) - Gln107(OE1), Asn125(O) : 3.06, 2.95 Tyr336(OH) - His105(NE2), Ala128(O) : 2.88, 3.82 Tyr336(O) - Gln107(NE2) : 2.88 Tyr336(N) - Gln107(OE1) : 3.07 Phe338(N)-Arg110(O) : 2.81 Arg434(NE)-Glu156(OE1) : 3.32 Arg434(NH2)-Glu156(OE2) : 3.02 Ile441(O) - Arg145(NH1,NH2) : 3.29, 3.36 Ile442(O)-Arg145(NH1):3.33 Lys484(NZ)-Glu465(OE1) : 3.44 Arg488(NH2) - Glu465(OE1) : 3.65 Trp489(NE1) - Asn490(OD1) : 2.96</p> <p>Salt bridge Glu332(OE2)-His129(ND1,NE2) : 3.56, 3.65 Asp333(OD1) - Arg132(NE) : 3.43 Asp333(OD2) - Arg132(NE, NH1, NH2) : 3.69, 3.41, 3.45 Lys484(NZ) - Glu465(OE1) : 3.65 Arg488(NH2) - Glu465(OE1) : 3.74, 3.27</p>
M2-M3	<p>Hydrogen bond Glu188(O) - Ala87(N),Lys88(N) : 3.46, 3.00 Glu188(OE1) - Met89(N) : 3.79 Ala190(N) - Glu133(OE2) : 2.60 Val191(N) - Glu133(OE1) : 2.85 Asp333(O) - His129(N,ND1) : 3.57, 2.81 Thr335(OG1) - Gln107(OE1), Asn125(O) : 3.11, 3.06 Tyr336(N) - Gln107(OE1) : 2.88 Tyr336(OH) - His105(NE2),Arg132(NH1) : 2.64, 2.74 Tyr336(O) - Gln107(NE2) : 2.79 Phe338(N) - Arg110(O) : 2.71 Leu420(N) - Ile117(O) : 3.34 Arg434(NE)-Glu156(OE1) : 3.55 Ile441(O) - Arg145(NH1,NH2) : 3.19, 3.20 Leu442(O)-Arg145(NH1) : 3.10 Lys484(NZ), Arg488(NH2) - Glu465(OE1) : 3.37, 3.27 Trp489(NE1) - Asn490(OD1) : 2.96</p> <p>Salt bridge Glu332(OE2) - His129(ND1,NE2) : 3.23, 3.51 Asp333(OD1) - Arg132(NE) : 3.62 Asp333(OD2) - Arg132(NE, NH1, NH2) : 3.78, 3.67, 3.67 His337(ND1)-Asp111(OD1) : 3.95 Arg434(NE) - Glu156(OE1) : 3.55 Arg434(NH2) - Glu156(OE2) : 3.90 Lys484(NZ), Arg488(NH2) - Glu465(OE1) : 3.37, 3.27</p>
M3-M1	<p>Hydrogen bond Glu188(O)-Ala87(N), Lys88(N) : 3.47, 2.97 Ala190(N)-Glu133(OE2) : 2.77 Val191(N)-Glu133(OE1) : 2.86 Asp333(O)-His129(N, ND1) : 3.57, 2.94 Thr335(OG1)-Gln107(OE1), Asn125(O) : 3.05, 3.08 Tyr336(N)-Gln107(OE1) : 2.77 Tyr336(OH) - His105(NE2), Asn125(O), Arg132(NH2) : 2.56, 3.08, 2.83 Thy336(O) - Gln107(NE2) : 2.77 Phe338(N)-Arg110(O) : 2.84 Leu420(N) - Ile117(O) : 3.23 Ile441(O)-Arg145(NH1,NH2) : 3.19, 3.23 Leu442(O)-Arg145(NH1) : 3.25 Lys484(NZ)-Glu465(OE1) : 3.35 Trp489(NE1)-Asn490(OD1) : 2.82</p> <p>Salt bridge Glu332(OE2)-His129(ND1,NE2) : 3.60, 3.82 Asp333(OD1)-Arg132(NE,NH1,NH2) : 3.82, 3.66, 3.62 Lys484(NZ)-Glu465(OE1) : 3.35</p>
M1-M1*	<p>Hydrogen bond Arg62(NH1)-Glu72(OE1) : 3.00 Glu72(OE1)-Arg62(NH1) : 3.45</p>

	<p>Lys88(O)-Gln200(NE2) : 3.35 Leu95(O)-Arg98(NE) : 3.51 Arg98(NH1)-Glu96(OE2) : 3.37 Arg140(NH2)-Asp195(OD2) : 2.89 Gln200(NE2)-Lys88(O) : 3.20</p> <p>Salt bridge Arg62(NH1)-Glu72(OE1) : 3.00 Glu96(OE1)-Arg98(NH1):3.83 Arg140(NH2)-Asp195(OD2) : 2.89</p>	<p>Lys88(NZ)-Val207(O), Asn208(OD1) : 2.94, 3.09 Glu96(OE1)-Arg98(NH1) : 3.83 Arg98(NE)-Leu95(O) : 3.41 Asp195(OD2)-Arg140(NH2) : 2.93 Val207(O), Asn208(OD1)-Lys88(NZ): 2.96, 2.97</p> <p>Glu72(OE1)-Arg62(NH1): 3.45 Arg98(NH1)-Glu96(OE2) : 3.37 Asp195(OD2)-Arg140(NH2);2.93</p>
M3-M1*	<p>Hydrogen bond Arg181(NH1)-Asp184(OD2) : 2.96 Arg187(NH1)-Gly209(O) : 2.89 Glu188(OE2) – Asn208(ND2) : 3.62 Asn208(ND2)-Glu188(OE2) : 3.57 Asp214(OD1)-His284(NE2) : 2.59</p> <p>Salt bridge Arg181(NH1)-Asp184(OD2) : 2.98 Lys196(NZ)-Glu193(OE1) : 3.91 His284(NE2)-Asp214(OD1) : 2.66</p>	<p>Asp184(OD2)-Arg181(NH1) : 3.06 Glu188(N) – Asn208(OD1) : 2.84 Asn208(OD1)-Glu188(N) : 2.89 Gly209(O)-Arg187(NH1) : 2.77 His284(NE2)-Asp214(OD1) : 2.66</p> <p>Asp184(OD2)-Arg181(NH1) : 2.91 Asp214(OD1)-His284(NE2) : 2.59</p>

Discussion

L-arabinose is one of sugar isomerases that can convert not only L-ribulose to L-arabinose but also D-galactose to D-tagatose. Biological method using enzyme is known to benefit for D-tagatose production than chemical method.

D-tagatose as sweetener substituted sugar such as sucrose has been an industrial interest not only diet food of low-calorie but also prescription drug. Furthermore, it can be used for therapy as antidiabetic, antiobestic, and anticancer agent [121-123]. AI of a variety of bacteria including *E. coli* can produce D-tagatose but efficiency of D-tagatose production can appear differently by parameters such as pH, divalent metal ions, and temperature. Generally, (Hyper) thermophilic bacteria has increased yield than mesophilic bacteria by high reaction temperature, metal ion, and pH. Until now, biochemistry study of (hyper) thermophilic AI is already investigated but structural study of AI not yet known except AI of mesophilic *E. coli*.

Herein, AI from *G. kaustophilus* was purified to soluble form and determined to apo-, Mn^{2+} -, and L-arabitol bound AI structures. Structure of GKAI contains six molecules to dimer of trimer. Active site of GKAI is located in cleft between the C-terminal and the central domain of a subunit and N-terminal domain of other subunit. Structures binding metal ion (Mn^{2+}) and L-arabitol also well defined in active site.

GKAI catalyzes substrate optimal temperature 60°C by thermostability of enzyme at high temperature. Although structure of thermophilic AI is similar with that of mesophilic, thermostability can increase by factors such as composition of Arg and number of water-mediated hydrogen bonds. First, Arg composition of GKAI is 2% higher than other AI including ECAI, ECFI, and *Sereptococcus* FI. Specially, non-conserved Arg such Arg62, 98, and 145 on interface of GKAI formed interaction (salt bridge and hydrogen bond) unlike other AI. Second, GKAI show difference in number of water molecules in active site comparing other AI. A number of water in GKAI active site contains two or three more than ECAI. These water-mediated hydrogen bond between residues and substrate in active site suggested that provide

evidence to the increased of thermostability as mentioned in other researches [118]. Thermophilic GSAI increases production of D-tagatose because high temperature (60-65°C) induces the galactose-tagatose equilibrium towards tagatose [124]. Thermophilic GKAI also seem to reaction activity similar with GSAI, suggesting that it can increase conversion efficiency of galactose as changing substrate preference by high temperature.

Consequently, structure of GAKI suggests that it might provide information to a thermostability and change of substrate preference with metal-dependent reaction at high temperature.

ABSTRACT

Crystal structures of bacterial proteins, PenA and CagL, related to human pathogenicity

Choi, Jin Myung

Advisor: Prof. Lee, Sung Haeng, Ph.D.

Department of Bio-materials,

Graduate School of Chosun University

i. Structural study of class A β -lactamase PenA in *Burkholderia thailandensis*

The β -lactamase, which was produced for protection to antibiotics in bacterium, generally reveals antibiotic resistance by hydrolyzing the antibiotics' β -lactam ring. Until present, Almost of bacterium using β -lactamase such as TEM and SHV evolved for survival against a variety of antibiotics through structural change by mutation such as deletion, substitutions, and repeat unit insertion of amino acid. Among β -lactamases, the Pen A used in this study belongs to class A β -lactamase in *B. thailandensis* that it originally exhibits an antibiotic resistance to penicillin such as amoxicillin. In addition to the well-known β -lactamases including SHV-1 and TEM-1, PenA also was found not only change to ESBL by mutating such as substitution and deletion but also switching of two phases reversely by TR duplication, resulting that can increase resistance to third generation cephalosporin such as ceftazidime. Although study about point mutation for ESBL was already known from other β -lactamases, the studies about an accurate structural mechanism by insertion of long repetitive amino acid for inactivation of antibiotics are little. So, structures of PenA-WT and mutants including TR mutants (TR10 and TR11) and

point mutants (N136D and C69Y) are investigated. Each protein are expressed, purified, and obtained x-ray crystallographic data at high resolution (1.4-1.7 Å) without and with ceftazidime-like glycyI boronic acid (CBA), respectively. Result of structure analysis, the major change between PenA-WT and mutants can show commonly change of omega (Ω) loop on active site through conformation change for the increased cavity size. This change increases not only binding affinity to antibiotic such as third generation cephalosporin, ceftazidime but also catalysis of hydrolysis activity. Consequently, change of omega loop at active site is very important role to change of substrate spectrum and induces the increase of resistance to a variety of antibiotic. Using these results, we can suggest that study to architecture of active site with CBA provides information to new drug development.

ii. Structural Study of CagL, a Component of the T4SS Pilus Assembly in *H. pylori* k74 strain.

Type 4 secretion system (T4SS) in virulent *H.pylori* isolates, which are encoded with Cag (cytotoxic-associated gene) pathogenicity island, 40kb DNA, plays a very important role to interact with receptors of host cell and transfer CagA into host cell. CagL is characterized to interact with integrin through its flexible RGD domain during CagA translocation via a type IV secretion system (T4SS) of *Helicobacter pylori* (Hp). Herein, to investigate interaction of pathogen-host cell, the CagL protein from *H.pylori* K74 strain was preferentially purified and was crystallized Se-Met and native CagL protein. Crystals of Se-Met and native CagL were subjected to x-ray diffraction at 2.9 Å and 2.8 Å resolution. It belonged to the same tetragonal space group P4₃2₁2 and a dimer was found in asymmetry unit. Oligomeric state of CagL in solution was identified as monomer by AUC. Overall structure of monomer CagL reveals like a helix bundle consisting of four elongated α -helices and two α -short helices. The RGD motif, which is known to interact with integrin, is found in the long rigid α 2 helix flanked by the

conserved protease sensitive and RGD helper sequences. This indicates the possible unfolding of RGD region for integrin association. However, in comparison with 26695 CagL, the loop of CagL-K74 connecting $\alpha 1$ to $\alpha 2$ was well defined, and two additional salt bridges were found between the helices, suggesting that the putative flexible region harboring RGD motif might be more stable in this CagL variant. Also, because CagL is known to complex formation with CagI and CagH in cell, how CagL forms complex with other cag proteins such as CagI and CagH and how it interacts with host cell receptor integrin will be more study.

VIII. Acknowledgements

박사학위 졸업논문문을 쓰는 동안 대학을 졸업하고 10여년 후에 뒤늦게 다시 학업을 시작할 때 막막했던 기억이 떠올랐습니다. 아무것도 모른 채 전남대 약대에서 시작했던 석사과정과 전공을 바꿔 조선대에서 다시 새로운 내용을 접한 박사과정을 하는 8년여 동안 저한테 힘들었지만 참으로 의미 있는 시간이었습니다.

먼저, 박사 학위과정 동안 비록 교수님의 기대에 제대로 미치지 못했다고 생각이 들지만 저를 박사과정 첫 제자라며 관심과 마음으로 지도해주신 이성행 교수님께 감사의 말씀을 드립니다. 처음 실험실에 들어왔을 때 실험실이 제대로 갖추지 못한 상태에서 시작하여 막막하기도 하였지만, 시간이 지난 후, 지금 생각해보면 좋은 경험이라고 생각이 듭니다. 실험실 생활하는 동안에 비어있던 실험실을 셋팅부터 2번의 이사와 함께 지금의 실험실까지, 변화가 많았던 것으로 기억이 납니다.

처음에 나를 다시 이 길로 들어서게 하고 힘들 때면 술한잔 하며 술친구가 되어준 동기 현중이, 나보다 먼저 학위과정을 끝내고 연구교수와 포닥을 하고 있는 대학 후배들이지만 대학원 선배인, 홍범, 현우, 기운이, 그 구려대 교수이지만 한참 베풀어 열심히 영호성, 뒤늦게 시작하여 많은 공감대를 형성했던 라영이 등 이러한 동기, 선배들 덕분에 그 동안에 많은 의지가 되어 할 수 있었습니다.

지금 생각해보면 박사학위 전공은 저한테 처음에 무지 생소하여 감이 오질 않고 어떻게 해야 될지 막막하기만 하였습니다. 기본적인 클로닝, 단백질 정제, 결정화, 모델링 등 이러한 일들은 여기 와서 처음으로 하는 일들이고 체력적으로 뿐만 아니라 기본적인 이론지식이 없었기 때문에 무지 힘들었습니다. 그러나 이러한 과정 동안에 아래와 같이 많은 도움을 주신 분들 덕분에 많은 도움을 받아 연구를 할 수 있었습니다. 단백질 결정학에 관해 청강을 허락해주신 GIST 엄수현 교수님, 포항 가솔기 연구소에서 프로세싱과 함께 이론적으로 많은 도움을 주신 임영준 교수님, 빔라인 이용시 같이 사용할수 있고 프로세싱에 도움과 항상 물어

볼 때 마다 잘 가르쳐준 윤형섭, 이영진, 안준엽 박사과정 GIST 학생들에게 감사합니다.

그리고 무엇보다도 사회생활을 접고 갑자기 대학원을 다시 다닌다고 해도 묵묵히 지원해주신 부모님께 죄송함과 더불어 진심으로 감사 드립니다.

지금까지의 과정을 토대로 박사과정 졸업이 끝이 아니고 다시 새로운 과정의 시작이라는 생각과 함께 더 나은 결과를 위해 좀더 열심히 하겠습니다.