Računalne simulacije interakcija dviju domena autotransporterske esteraze EstA iz bakterije Pseudomonas aeruginosa

Mrnjavac, Natalia

University of Zagreb
Faculty of Science
Department of Biology

Natalia Mrnjavac

Molecular Dynamics Study of Functionally Relevant Interdomain and Active Site Interactions in the Autotransporter Esterase EstA from *Pseudomonas aeruginosa*

Graduate Thesis

Zagreb, 2015

Ovaj rad je izrađen na Zavodu za fizikalnu kemiju Prirodoslovno-matematičkog fakulteta pod vodstvom doc. dr. sc. Branimira Bertoše.

Rad je predan na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta radi stjecanja zvanja magistra molekularne biologije.

The thesis was done at the Division of Physical Chemistry of the Faculty of Science under the supervision of Asst. Prof. Dr. Branimir Bertoša.
It has been submitted for assessment to the Department of Biology of the Faculty of Science in order to earn a Master in molecular biology.

Za početak moram zahvaliti svom mentoru, docentu Branimiru Bertoši, na njegovoj podršci, pomoći i susretljivosti od prvog trenutka kada sam postala njegov diplomand. Neizmjerno sam mu zahvalna na tome što mi je ponudio da radim kod njega diplomski rad u trenutku kada mi je priznanje za moj rad i trud tijekom studija uistinu trebalo. Iznimno sam mu zahvalna i na povjerenju koje mi je ukazao kada mi je omogućio sudjelovanje u nastavi i spremanje vježbi za kolegij Strukturna računalna biofizika, i iskreno se nadam da sam to povjerenje opravdala. Zahvaljujem mu što je uvijek bio spreman saslušati i uvažiti moje mišljenje, zbog čega sam se osjećala cijenjeno i motivirano. Zahvalna sam mu i na razumijevanju u teškim trenucima, na tome što je uvijek bio na mojoj strani, i što je vjerovao u mene i moje sposobnosti i onda kada ja nisam. Najiskrenije Vam hvala.

Željela bih spomenuti i druge profesore koji su me na neki način inspirirali, motivirali i obilježili. U prvom redu zahvaljujem profesorici Iti Gruić, mojoj mentorici završnog rada na preddiplomskom studiju. Zahvaljujući njenom načinu rada, njenom doživljavanju i prenošenju biokemije studentima sam se u to područje nepovratno zaljubila. Bila sam jako sretna zbog svake prilike koju sam imala za slušati njena predavanja i raditi pod njenim vodstvom.

Zahvaljujem i profesoru Davoru Kovačeviću, koji me cijenio i bio susretljiv prema meni od kada sam bila njegov student na Fizikalnoj kemiji, pa do razdoblja kada sam na Zavodu izrađivala diplomski rad. Htjela bih zahvaliti i profesoru Mladenu Kučiniću, mom profesoru zoologije na prvom semestru, čije su priče rasplamsale moju ljubav prema biologiji i fascinaciju živim svijetom. Zahvaljujući njemu sam počela gledati na svijet kao biolog. Ne smijem zaboraviti i docenta Bojana Lazara, s kojim dijelim ogromnu ljubav prema morskim kornjačama.

Moram zahvaliti i svojim gimnazijskim profesorima, naročito Dini Linić, Bruni Kučan, Ratku Duševiću, Melbi Blažić Grubelić i Dragici Paškvan. Oni su me učili i odgajali u okolini gdje se cijenilo znanje, sposobnost i predanost radu. Takav sustav vrijednosti će me cijeli život pratiti, a to dugujem i njima. Uz to, pripremili su me na najbolji mogući način za budućnost, naoružavši me znanjem. Ne mogu izostaviti i profesorice Smiljanu Orlić i Lauru Mužić koje su me još u osnovnoj školi poticale i motivirale.

Velike zahvale idu i mojim prijateljima, bez kojih ne bih sigurno bila ovdje gdje jesam. Moram spomenuti svoje najstarije prijateljice, Saru i Antonellu, i svoje gimnazijsko društvo – hvala vam što postojite, unatoč kilometrima i životnim okolnostima. Zahvaljujem i svojim zagrebačkim prijateljima, kolegama, i jednoj jedinoj sestrični, odnosno svima koji su na ovaj ili onaj način podijelili sa mnom dio mog zagrebačkog života i bili mi podrška, svjesno ili nesvjesno, kada mi je podrška trebala.

Posebno zahvaljujem Mateji, svojoj prvoj stanskoj cimerici, koja je više od ikog drugog dijelila sa mnom muke pisanja diplomskog rada. Hvala ti što me smiruješ, uveseljavaš, i dijeliš sa mnom život u našem malom sretnom balonu.

Moram posebno zahvaliti i Jeleni, svojoj cimerici na preddiplomskom studiju. Kada sam dolazila u Zagreb nisam se mogla nadati boljoj cimerici i prijateljici. Zahvaljujući tebi su sve nesigurnosti samostalnog života, nostalgije, strahovi, domski, fakultetski i svakojaki životni problemi bili nekako lakši. Zahvaljujem ti na svim podijeljenim snovima, maštanjima, osmijesima i suzama.

Ove zahvale ne bi mogle proći bez jednog velikog hvala koje dugujem svojoj kolegici s predavanja, partnerici u seminarskim radovima i praktikumima, cimerici tijekom jedne i pol nezaboravne godine, i odličnoj prijateljici Luciji. Hvala ti, Lu, za sve što si dijelila sa mnom tijekom studija, od nervoze pred prve ispite kad smo bile tek brucoši, neprospavanih noći rješavajući zadatke i pripremajući seminare, do svih šala i smijeha na praktikumima. Hvala ti što si bila tu da me smiriš kada sam paničarila, i da me tako lako odobrovoljiš kada sam bila tužna. Tvoj smisao za humor koji obožavam i tvoje tako iskreno prijateljstvo su učinili da mi studentske godine budu jednostavno toliko bolje. Od srca ti hvala.

Na kraju, hvala mojoj obitelji, a iznad svega - mojim roditeljima. Hvala vam zbog vaše beskrajne ljubavi i truda. Zbog vječite brige i podrške koju ste mi dali tijekom cijelog mog života. Zbog vrijednosti kojima ste me učili težiti. Zbog toga što ste vjerovali u mene i moje sposobnosti u svakom trenutku mog putovanja. Za mnoge duge razgovore i osjećaj da vam se uvijek mogu obratiti kada sam izgubljena. Oduvijek ste bili moja sigurna luka, i kada bih posvetila ovaj rad nekome, zasigurno bih ga posvetila vama.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek Diplomski rad

Računalne simulacije interakcija dviju domena autotransporterske esteraze EstA iz bakterije *Pseudomonas aeruginosa*

Natalia Mrnjavac
Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Enzim EstA je funkcionalno GDSL esteraza. Putuje kroz vanjsku membranu bakterije *Pseudomonas aeruginosa* mehanizmom Va ili autotransporterskim mehanizmom kod kojeg se katalitička domena prenosi u izvanstanični prostor uz pomoć β bačve (autotransporterske domene). U slučaju EstA β bačva i na nju vezana katalitička domena ostaje učvršćena u membrani. Biološki supstrat EstA je nepoznat, ali je poznato da je aktivnost ovog enzima povezana s pokretljivošću bakterijske stanice, stvaranjem biofilmova i produkcijom ramnogalakturonana. Kao u ostalih GDSL hidrolaza, i u aktivnom mjestu EstA se nalazi katalitička trijada i oksioanionska šupljina. U ovom radu su identificirane funkcionalno važne aminokiseline i mreža vodikovih veza u aktivnom mjestu. Uz to, opisane su hidrofobne interakcije i vodikove veze između domena. U odcijepljenoj katalitičkoj domeni je zamijećeno otvaranje aktivnog mjesta kako bi se omogućilo pristajanje tetraedarskog intermedijera, dok je u cjelovitom enzimu EstA s vezanim intermedijerom zamijećena promjena u strukturi heliksa 6 u aktivnom mjestu. Svi rezultati su temeljeni na 100 ns dugim simulacijama molekulske dinamike odcijepljene katalitičke domene i cjelovitog enzima EstA, u oba slučaja sa i bez vezanog tetraedarskog intermedijera.

(82 stranice, 33 slike, 4 tablice, 87 literaturnih navoda, jezik izvornika: engleski)

Rad je pohranjen u Središnjoj biološkoj knjižnici
Ključne riječi: GDSL hidrolaze, katalitička domena, autotransporterska domena, molekulska dinamika
Voditelj: Doc. dr. sc. Branimir Bertoša
Ocjenitelji: Doc. dr. sc. Branimir Bertoša
Izv. prof. dr. sc. Ita Gruić Sovulj
Doc. dr. sc. Ivana Ivančić Baće
Rad prihvaćen: 2. travnja 2015.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Graduate Thesis

Molecular Dynamics Study of Functionally Relevant Interdomain and Active Site Interactions in the Autotransporter Esterase EstA from *Pseudomonas aeruginosa*

Natalia Mrnjavac
Rooseveltov trg 6, 10000 Zagreb, Hrvatska

The enzyme EstA is functionally a GDSL esterase. It is transferred through the outer membrane of *Pseudomonas aeruginosa* by the type Va or autotransporter mechanism, where the transfer of the catalytic domain (the passenger) to the cell exterior is aided by the β barrel domain (the autotransporter). In EstA the barrel remains membrane embedded with the passenger bound to it. The physiological substrate of EstA is unknown, although its activity is known to be related to bacterial cell motility, biofilm formation and rhamnogalacturonan production. As a GDSL hydrolase, the active site of EstA contains a catalytic triad and oxyanion hole. Relevant active site residues, including an active site hydrogen bond network, are described in this work. In addition, interdomain hydrogen bonds and hydrophobic interactions are characterised. Active site opening to fit the tetrahedral intermediate was observed in the isolated passenger domain, while a structural perturbation of the active site helix 6 was noticed when the tetrahedral intermediate was bound in full-length EstA. All results are based on 100 ns long molecular dynamics simulations of the passenger domain of EstA and full-length membrane embedded EstA, both with and without a bound tetrahedral intermediate.

(82 pages, 33 figures, 4 tables, 87 references, original in: English)

Thesis deposited in the Central Biological Library
Key words: GDSL hydrolases, passenger domain, autotransporter domain, molecular dynamics
Supervisor: Dr. Branimir Bertoša, Asst. Prof.
Reviewers: Dr. Branimir Bertoša, Asst. Prof.
Dr. Ita Gruić Sovulj, Assoc. Prof.
Dr. Ivana Ivančić Baće, Asst. Prof.
Thesis accepted: 2th April 2015

Table of Contents

Abbreviations

AT – autotransporter

MD – molecular dynamics

OM – outer membrane

PDB ID – Protein Data Bank Identifier

POPE – 1-palmitoyl-2-oleoyl-phosphatidylethanolamine

RMSF – root mean square fluctuation

RMSD – root mean square deviation

1 Introduction

1.1 Type V Bacterial Protein Secretion – the Autotransporter mechanism

1.1.1 Autotransporter Structure and Subdivision

Out of the six major protein secretion pathways recorded in Gram-negative bacteria, the type V mechanism, also known as the autotransporter pathway, is the simplest and most widespread (Scott-Tucker and Henderson 2009, Pukatzki et al. 2006, Pallen et al. 2003, Henderson et al. 2000). It was named that way because at the time it seemed that these proteins needed no additional factors or energy to be translocated to the extracellular milieu (Jose et al. 1995). Autotransporter proteins (ATs) are functionally esterases, proteases or adhesins often acting as virulence factors in pathogenic bacteria (Yen et al. 2008, Wells et al. 2007, Henderson and Nataro 2001, Wilhelm et al. 1999, Loveless and Saier 1997).

Before undergoing posttranslational modifications, all ATs share a few common structural features: an N-terminal signal sequence that targets them for transport through the bacterial inner membrane, a functional passenger domain that is secreted to the outside of the cell, and a C-terminal β barrel, known as the autotransporter domain, which is inserted into the bacterial outer membrane (OM). The passenger domain in most cases adopts a β helical structure, which is characterised by two or three series of parallel β sheets connected to each other by loops, which can be short, but can also include entire protein domains (Kajava and Steven 2006, Henderson et al. 1998, Pohlner et al. 1987). According to some authors, the type V secretion system can be subdivided into three categories termed Va, Vb and Vc (Cotter et al. 2005, Henderson et al. 2004, Jacob-Dubuisson et al. 2001, Henderson et al. 2000). Classical ATs, often referred to simply as *autotransporters*, are part of the Va group. The Vb group is also named *the two-partner system* where the passenger and AT domains are translated as two separate polypeptide chains. The Vc group consists of ATs that are inserted into the OM as trimers in order to form a functional β barrel. Other authors tend

to count in two more categories, i. e. the Vd and the Ve mechanism (Grijpstra et al. 2013, Celik et al. 2012, Salacha et al. 2010). The Vd secretion mechanism has been so far reported only for the patatin-like protein PlpD, and it is unusual for the AT domain resembles the barrel domain of the two-partner system, but in this case it is on the same polypeptide chain as the passenger. The Ve group, on the other hand, includes ATs with an inverted primary structure, with the passenger domain following the AT domain.

1.1.2 Secretion Mechanism of Va Autotransporters

In 1987 Pohlner et al. proposed the first model for OM translocation of an AT protein. They were working on the IgA protease from *Neisseria gonorrhoeae* and proposed a model where the autotransporter, or as they named it - *helper* domain, incorporates in the OM to form a pore which is then used by the protease domain for transport to the extracellular milieu. The protease domain would fold to an active conformation during this transport, and would then be autoproteolitically cleaved and released to the extracellular space.

It is generally accepted that ATs cross the inner membrane via the Sec apparatus followed by cleavage of the signal sequence by a periplasmic signal peptidase (Izard and Kendall 1994, Brundage et al. 1990, Driessen and Wickner 1990, Pohlner et al. 1987). In the periplasm some ATs have been shown to interact with chaperones such as SurA, Skp and DegP, presumably to prevent aggregation and misfolding, or even to aid in the folding of the AT β barrel (Volokhina et al 2011, Ruiz-Perez et al. 2009, Purdy et al. 2007). However, the folding and mode of insertion of the barrel into the OM is still unclear. Actually, the entire OM translocation step has been much disputed. It would seem ATs are not as autonomous in their transport to the cell exterior as they first appeared to be. Except for the aforementioned chaperones, ATs almost certainly require a factor homologous to the *E.coli* Bam complex (such as the newly discovered Tam complex in *E.coli*), which has been found to be crucial for transport of proteins across the bacterial OM (Selkrig et al. 2012, Jain and Goldberg 2007, Volhoux et al. 2003). ATs have been experimentally shown to interact with the Bam complex (Ieva and Bernstein 2009, Sauri et al. 2009). Primary structure comparison proved the existence

of a ubiquitous signal most often found at the C-terminus of OM barrel proteins which is probably related to Bam interaction (Celik et al. 2012). However, the mechanism adopted by the Bam complex itself to transport proteins across the OM is also unclear. The role of the Bam complex during autotransporter translocation could be at least threefold: it could have a role in inserting the AT domain into the OM, but it could also aid in maintaining an open conformation of the AT domain until the passenger is translocated, or even become the pore through which the passenger domain would cross the membrane (Sauri et al. 2011, Bernstein 2007). At the moment the model which implies a concerted action of the AT β barrel and the Bam complex seems like the most plausible one (Figure 1).

However, there is still the open question as to how exactly the protein passes through the pore. Junker and colleagues showed that the carboxyl terminus of the passenger domain crosses the OM before the N-terminus (2009). Furthermore, the β barrel creates a transmembrane pore of about 10 Å in diameter, which is just wide enough to fit two extended polypeptide chains (Oomen et al. 2004, Loveless and Saier 1997). This goes in favour of the hairpin transport model which was proposed back in 1987 by Pohlner and colleagues and is still widely accepted. According to this model, the C-terminus of the passenger domain would enter the pore first, form a hairpin on the other side of the membrane and allow the sliding of the rest of the polypeptide chain of the passenger domain towards the extracellular matrix. The passenger domain would fold while exiting the pore and this vectorial folding would provide the driving force for transport (Junker et al. 2006). However, many experimental data is ambiguous. There has been evidence of transport of fully folded and even glycosylated proteins by the autotransporter mechanism, which most certainly wouldn't fit in the β barrel pore (Sherlock et al. 2006, Brandon and Golberg 2001, Veiga et al. 1999).

After transport, the passenger domain of an autotransporter protein will have one of three faiths – it may stay attached to the AT domain, get cleaved and released into the extracellular environment, or be cleaved, but remain bound to its AT domain. Cleavage may be inter- or intramolecular, the latter being proteolytic or mediated by specific residues in the passenger domain or in the α-helical domain linker (Barnard et

al. 2012, Charbonneau et al. 2009, Fink et al. 2001, Egile et al. 1997, Shere et al. 1997).

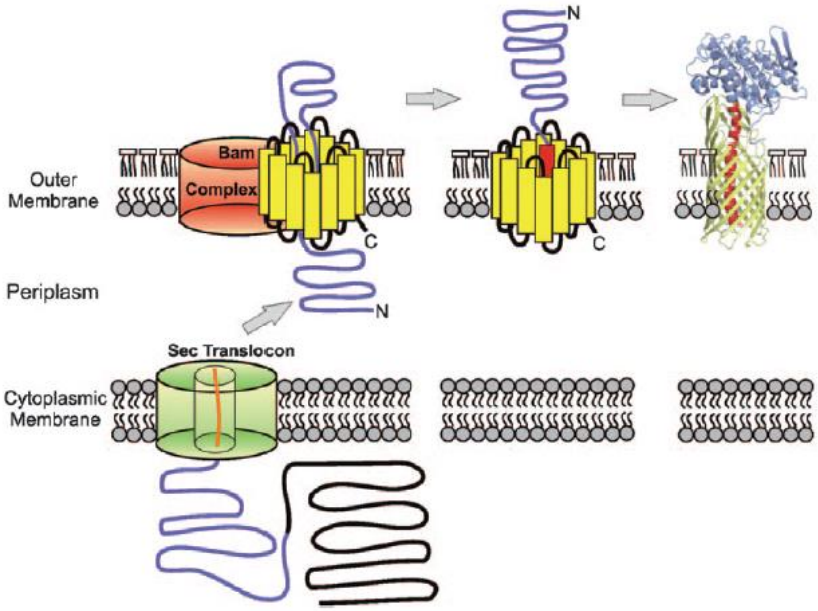


Figure 1. Type Va (autotransporter) secretion mechanism. The protein is transferred through the bacterial inner membrane by the Sec apparatus, after which the signal sequence is cleaved. Transfer through the OM is most likely carried out by a Bam homologue and the AT β barrel in a concerted fashion. In case of EstA the passenger domain remains covalently bound to the barrel domain after transfer (reproduced from Wilhelm et al. 2011).

1.2 The SGNH Hydrolase Protein Family

1.2.1 Tertiary Structure and Function

Hydrolases with an α/β fold were first identified by Ollis and colleagues in 1992. They found that a catalytic triad composed in most cases of Ser, His and Asp was present in all members of this structural superfamily. In 1995 Upton and Buckley reported a new family of lipolytic enzymes, the GDSL hydrolases. They were so named because of the characteristic GDSL motif, where S denotes the serine residue of the catalytic triad. In the same paper the researchers reported five conserved blocks in the primary structure of members of this family. In 1999 Arpigny and Jaeger classified the GDSL hydrolases as the second out of eight classes of bacterial lipolytic enzymes they described in their review article. A subgroup of this class was later recognized as SGNH

hydrolases based on the conserved residues in the common blocks found by primary structure alignments, with S standing for the catalytic serine in block I, G for the oxyanion hole glycine of block II, N for the asparagine of the oxyanion hole belonging to block III and H for the catalytic histidine of block V (Akoh et al. 2004, Yu-Chih et al. 2003, Mølgaard et al. 2000).

SGNH hydrolases are peculiar in their tertiary structure, with a fold that differs from the common α/β fold. Normally α/β hydrolases consist of 8 β strands connected by α helices (henceforth referred to simply as *helices*), and the catalytic triad is located on loops, with the one carrying the nucleophile, i. e. the nuclephilic elbow, being the most conserved feature of the fold (Ollis et al. 1992). SGNH hydrolases, on the other hand, display a so-called α/β/α fold, meaning they have a core composed of 4-5 β strands sandwiched by helices on both sides, and no nucleophilic elbow in the active site (Akoh et al. 2004, Mølgaard et al. 2000). In contrast to the GXSXG motif typical of α/β hydrolases, SGNH hydrolases have the aforementioned GDSL motif found much closer to the N-terminus (Akoh et al. 2004, Cho and Cronan 1993).

SGNH hydrolases are found mostly in bacteria and plants, with functions ranging from proteases and lipases to arylesterases, carbohydrate esterases, thioesterases and acyltransferases (Akoh et al. 2004). They can play a role in bacterial virulence or have effects on plant morphogenesis and defence mechanisms (Leščić Ašler et al. 2010).

Some GDSL esterases belong to the type V bacterial secretion system, and they constitute the passenger domain of ATs serving as bacterial virulence factors (e.g. EstA from *Pseudomonas aeruginosa*, EstP from *Pseudomonas putida*, ApeE from *Salmonella enterica*, Lip-1 from *Xenorhabdus luminescens*, etc.) (Wilhelm et al. 2011, Loveless and Saier 1997). They have some common distinct characteristics: they are not generally secreted into the extracellular medium, Lip-1 being an exception to this rule, they possess lipolytic activity, hence could in theory hydrolyse membrane lipids, and their tertiary structure is typical for SGNH hydrolases, but very unusual for autotransporter passenger domains which are usually β-helical (Wilhelm et al. 2011, Kajava and Steven 2006, Wang and Dowds 1993) (Figure 2).

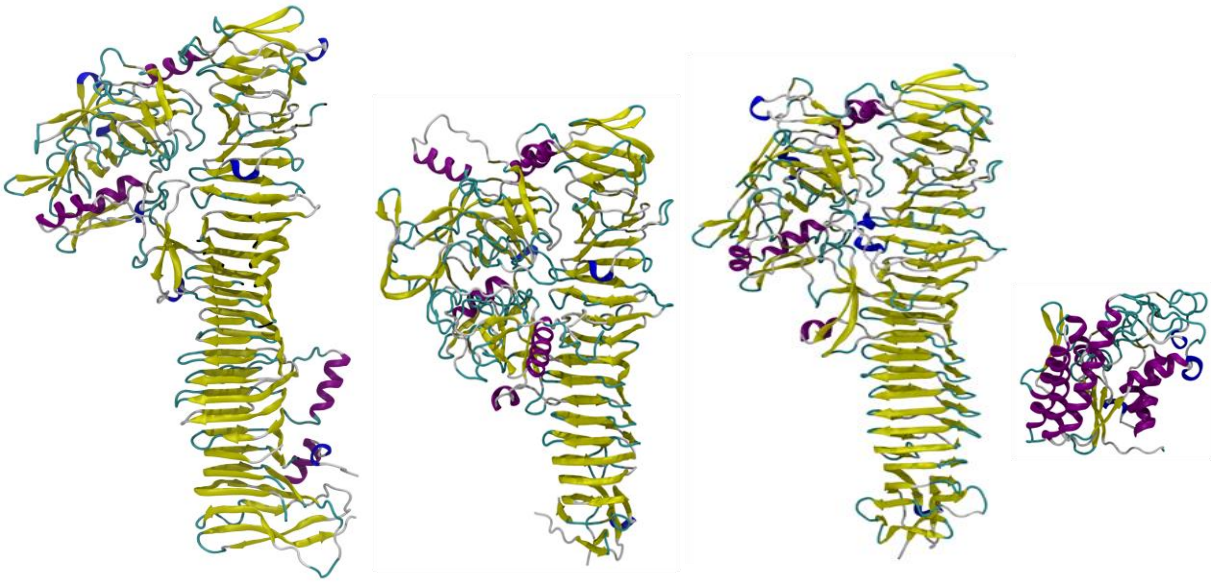


Figure 2. Passenger domains of some autotransporter proteins. The passenger domains shown are those of haemoglobin protease - Hbp from *Escherichia coli* (PDB ID: 1WXR; Otto et al. 2005), immunoglobulin A1 protease - IgAP from *Haemophilus influenzae* (PDB ID: 3H09; Johnson et al. 2009), adhesion and penetration protein - Hap from *H. influenzae* (PDB ID: 3SYJ; Meng et al. 2011), esterase A - EstA from *Pseudomonas aeruginosa* (PDB ID: 3KVN; van den Berg 2010). The passenger domain of EstA isn't β helical like the majority of ATs; it's smaller and displays the α/β/α fold typical of GDSL hydrolases.

1.2.2 Active Site Characteristics

Generally, esterases and lipases display broad substrate specificity, and SGNH hydrolases are no exception (Akoh et al. 2004, Fojan et al. 2000). TAP, the thioesterase I/protease I/lysophospholipase I from *Escherichia coli* and most studied SGNH hydrolase to date, is maybe the best example of a hydrolase displaying high substrate promiscuity (Yu-Chih et al. 2003). EstA and EstP from *Pseudomonas* species also display this property, although to a lower extent (Leščić Ašler et al. 2010). In addition, TAP was found to have a rigid core composed of β strands and α helices, and flexible loops which account for a flexible substrate binding pocket and active site (Yu-Chih et al. 2003, Yao-Te et al. 2001). Conformational changes due to loop movement

were also previously reported for SGHN hydrolases, especially for the so-called switch loop (Kovačić et al. 2013, Akoh 2004).

The catalytic triad composed of Ser/His/Asp and the oxyanion hole made up of Gly/Asn or Ser/Gly/Asn have a central place in the active site. Another characteristic feature of the SGNH family is the placement of the catalytic His and Asp only two residues apart, while they are found on different loops in other hydrolases (Mølgaard et al. 2000). The catalytic triad and the oxyanion hole residues are connected to each other by hydrogen bonding with two bridging conserved water molecules, as reported by Yu-Chih and colleagues (2003). Except for these bonds which would have the role of positioning the catalytic residues, they also found hydrogen bonding between different structural blocks of the SGNH hydrolases, as well as intra-block hydrogen bonds which would partake in stabilising secondary structures. Besides this specific hydrogen bond network, another peculiar activity-related feature of SGNH hydrolases was reported – a non-active site tryptophan residue was found to hydrogen bond with an active site leucine which is broadly conserved in bacteria and adjacent to the catalytic serine. This tryptophan was also found to exert hydrophobic interactions with other active site residues and somehow influence catalytic activity (Li-Chiun et al. 2009).

1.3 EstA - a GDSL Autotransporter

1.3.1 Structure and Function

EstA is an SGNH esterase found in the OM of *Pseudomonas aeruginosa*, a Gram-negative bacterium that commonly inhabits the soil, but can be an opportunistic plant and animal pathogen (Wilhelm et al. 1999). The natural substrate(s) and precise physiological function of this enzyme are not known, although it has been found that it preferentially hydrolyses substrates with short-chain fatty acids *in vitro*, and can also exhibit phospholipase B activity (Wilhelm et al. 2011, Leščić Ašler et al. 2010). In 2007 a study was conducted by Wilhelm and colleagues trying to elucidate the role of EstA. They found that, while it is not essential for growth under laboratory conditions, EstA inhibits all three types of cell motility observed in *P. aeruginosa*, namely swimming,

twitching and swarming, at the same time hindering rhamnolipid synthesis or transport and preventing the formation of normal bacterial biofilms.

EstA was predicted to be a GDSL autotransporter with a catalytic triad consisting of Ser14, His289 and Asp286 (Wilhelm et al. 1999) (Figure 3a). This was later confirmed by the crystal structure, which was solved in 2010 at 2.5 Å resolution and became the first solved structure of a full-length AT (van den Berg). The structure revealed the position of the active site on the apical surface of the passenger, at the entrance of a large hydrophobic pocket. Moreover, EstA was found to consist of two distinct domains of approximately 30 kDa each (Figure 3b). The β barrel domain which crosses the OM is composed of 12 antiparallel β strands, with an additional three-stranded β sheet named sheet D which is located on the upper part of the barrel towards the passenger domain. This additional sheet can be found in the structure of the recently solved AT domain of AIDA-1 from entheropathogenic *Escherichia coli* (PDB ID: 4MEE; Gawarzewski et al. 2014), however it is not present in other solved AT domain structures, such as NalP from *Neisseria meningitidis* (PDB ID: 1UYN; Oomen et al. 2004), EspP from enterohaemorrhagic *E. coli* (PDB ID: 2QOM; Barnard et al. 2007), Hbp from pathogenic *E. coli* (PDB ID: 3AEH; Tajima et al., to be published) or BrkA from *Bordetella pertussis* (PDB ID: 3QQ2; Zhai et al. 2011) (Figure 4).

The passenger domain of EstA is directly linked to the β barrel by a curved central α helix which spans the membrane through the lumen of the barrel and is attached to the first β strand on the periplasmic side. Being a typical SGNH hydrolase, the passenger domain of EstA differs significantly from other autotransporter passengers – it is much smaller and adopts an α/β/α fold. However, van den Berg suggested that a sequential folding mechanism for the EstA passenger is plausible based on the crystal structure, and might provide the driving force for OM translocation (2010). The fold consists of a four-stranded parallel β sheet surrounded by helices, which are connected by a big amount of loop structure. In addition, another β sheet, termed sheet B and composed of two antiparallel strands, can be found in the EstA passenger, but not in smaller SGNH hydrolases such as the thioesterase I/protease I/lysophospholipase L_1 (known as TAP) from *E. coli* (PDB ID: 1IVN; Yu-Chih et al. 2003), or the rhamnogalacturonan acetylesterase from *Aspergillus aculeatus* (PDB ID: 1DEO;

Mølgaard et al. 2000). Another region worth mentioning is block IIIa, which is found in all AT SGNH hydrolases, and in EstA consists of amino acids 238-258 (Leščić Ašler et al. 2010). It is far from the active site and includes two cysteines connected by a disulphide bridge (Figure 3a).

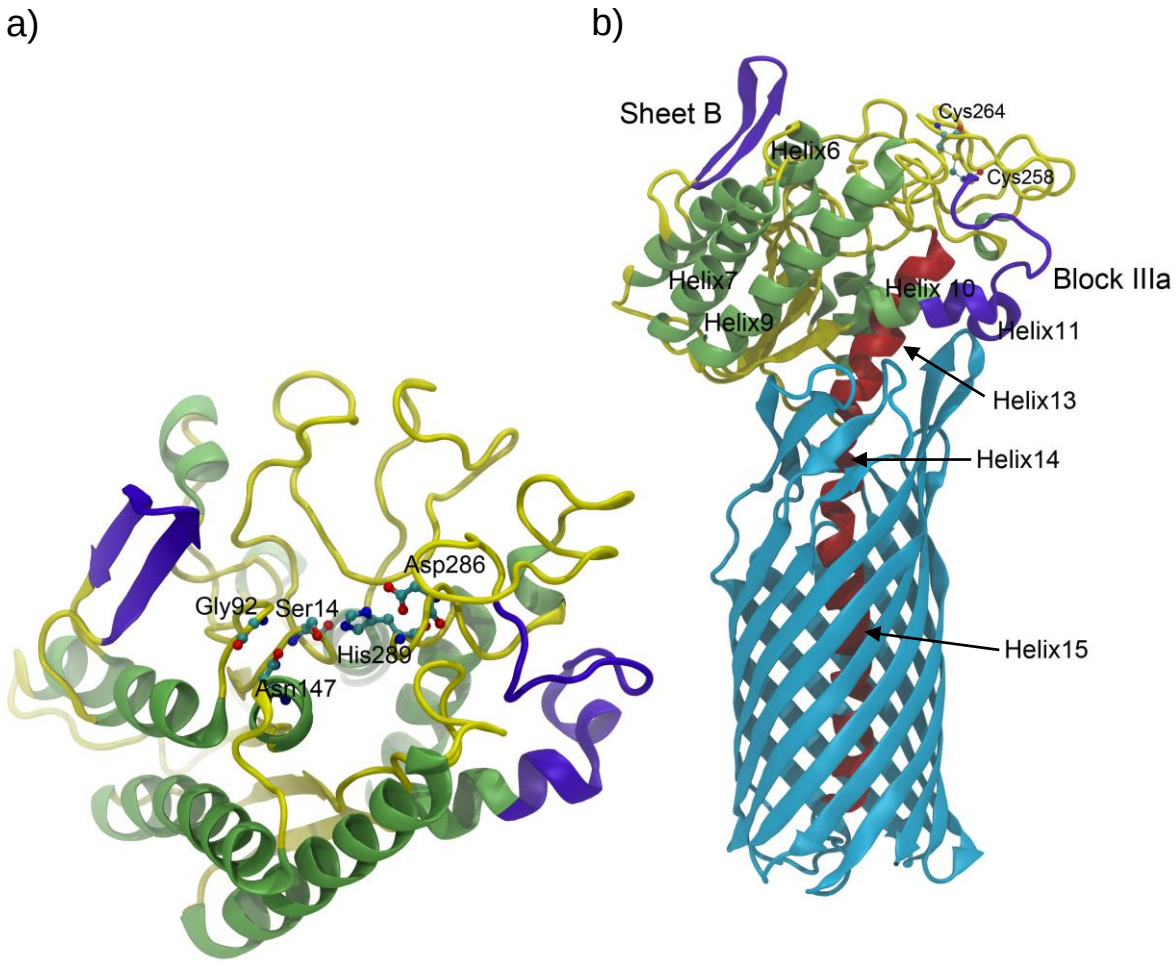


Figure 3. EstA crystal structure (PDB ID: 3KVN). a) Residues of the catalytic triad (Ser14-His289-Asp286) and the oxyanion hole (Gly92-Asn147) are shown in ball-and-stick representation. b) The AT domain is shown in blue. The central membrane-spanning helix, which comprises helices 13, 14 and 15, is in red. Other helices of the passenger domain are coloured lime. Sheet B and block IIIa, atypical regions for GDSL hydrolases, are coloured violet. Cys258 and Cys264 that are engaged in a disulphide bond are shown in ball-and-stick representation.

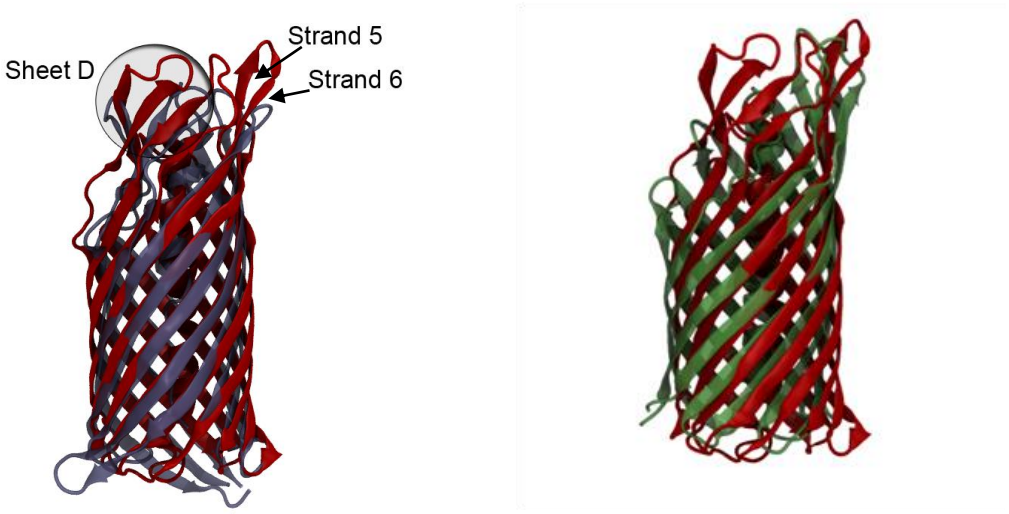


Figure 4. Autotransporter domain of EstA (red) superimposed with the AIDA-I (violet) and NalP (green) autotransporter domains. Sheet D is an additional secondary structure present in EstA and AIDA-I, but absent in NalP and other ATs.

1.3.2 Research Goals

Experimental data of EstA activity measured *in vitro* surprisingly showed that the isolated passenger domain exerts higher catalytic activity than the full-length protein (F. Kovačić*, personal comm.*). The same phenomenon was observed in EstP, a homologue of EstA from *Pseudomonas putida*. In 2010 Leščić Ašler et al. found that the isolated passenger domain of EstP showed five to ten times higher catalytic activity than the full-length protein for all tested substrates. That was very unusual, since it is generally accepted that enzymes have evolved to optimally catalyse the turnover of their substrates. However, an isolated passenger domain, which seemed to be much more efficient, was never detected *in vivo*. Moreover, a hybrid enzyme was constructed from the passenger domain of EstA and the AT domain of EstP, and vice versa. While the enzyme composed of the EstP passenger and EstA AT domain exhibited lower catalytic activity than EstP, as expected for an artificial construct, the hybrid consisting of EstA passenger and EstP barrel domain showed three to seven times higher activity than full-length EstA for all substrates (Table 1). When comparing the kinetic parameters for the turnover of *p*-nitrophenyl butyrate, the K_m value of full-length EstA was higher than the K_m of the hybrid enzyme consisting of EstA passenger and EstP barrel domain,

which points to more effective substrate approach or binding in the latter case (Table 2). These results were quite unexpected. Before these measurements, when contemplating the activity of full-length enzymes and isolated passenger domains, one could argue that interactions of the passenger with the membrane might cause the difference in activity. However, activity measurements of hybrid enzymes showed that the C-terminal domain itself affects the passenger domain activity.

Table 1. Results of activity measurements for EstA, EstP and their hybrid enzymes (data taken from Leščić Ašler et al. 2010).

Substrate	$EstP_N$[a]	EstP[b]	EstA[c]	$EstP_N$-$EstA_C$[d]	$EstA_N$-$EstP_C$[e]
p-nitrophenyl acetate (C2)	498.0	55.9	3.0	5.4	4.9
p-nitrophenyl propionate (C3)	162.0	25.5	8.3	3.9	47.5
p-nitrophenyl butyrate (C4)	928.3	105.2	75.1	19.2	161.0
p-nitrophenyl valerate (C5)	502.5	86.4	32.4	15.8	85.8
p-nitrophenyl caproate (C6)	736.0	125.1	26.6	21.4	113.7
p-nitrophenyl caprylate (C8)	281.3	24.5	6.0	6.3	28.4
p-nitrophenyl caprate (C10)	1262.9	125.6	9.5	17.6	18.2
p-nitrophenyl laurate (C12)	281.3	28.3	7.0	6.4	72.6
p-nitrophenyl myristate (C14)	23.1	5.6	6.0	2,1	89.3
p-nitrophenyl palmitate (C16)	4.1	0.8	4.2	0.5	71.8
p-nitrophenyl stearate (C18)	1.4	0.1	4.0	0.1	50.8

Activities are expressed in U mg^{-1}.

[a] EstP N-terminal (passenger) domain.

[b] EstP full-length enzyme.

[c] EstA full-length enzyme.

[d] Hybrid enzyme made up of the N-terminal (passenger) domain of EstP and C-terminal (AT) domain of EstA.

[e] Hybrid enzyme made up of the N-terminal (passenger) domain of EstA and C-terminal (AT) domain of EstP.

Table 2. Kinetic parameters for the turnover of *p*-nitrophenyl butyrate (data taken from Leščić Ašler et al. 2010).

Enzyme	V_m [U mg^{-1}]	K_m [mM]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} M^{-1}]
EstP$_N$	735	0.3	345	1.2 x 10^6
EstP	85	0.3	91	3.3 x 10^5
EstA	220	0.7	247	3.4 x 10^5
EstP$_N$-EstA$_c$	11	0.5	12	2.5 x 10^4
EstA$_N$-EstP$_c$	43	0.5	47	9.4 x 10^4

When the 3D structure of EstA was solved, van den Berg suggested there might be many interactions between the passenger and AT domain (2010). In the same year Leščić Ašler et al. speculated that these putative interactions might alter the arrangement of residues in the active site and in its vicinity, resulting in a difference in activity. The aim of the research presented in this thesis was to characterise the interdomain interactions and to find the differences in the structural and dynamical properties of the active site between the isolated passenger domain of EstA and the full-length enzyme. Such results could point towards the cause of the unexpected activity increase detected experimentally in the isolated EstA passenger.

1.4 Molecular Dynamics Simulations

Molecular dynamics (MD) is a computational technique that allows prediction of the motion of atoms in the system, i. e. its dynamics, based on Newtonian classical mechanics. In MD simulations forces on every particle in the system are computed by deriving the potential energy function of the system, and consequently accelerations can be calculated based on Newton's second law of motion. These calculations are solved for every time step, yielding new velocities for each particle every time. The result is a trajectory file containing information on coordinates of every particle in the system as a function of time. The idea is to generate an ensemble of structures that will be representative of a real system made up of a great number of molecules, in accordance with the ergodic hypothesis.

While experiments give information on macroscopic properties, which are actually an average over all molecules in the system, MD simulations can be used to study atomistic features and properties of the system. Therefore, MD simulations can be used to complement experiments in order to gain additional insight into the system, help with the interpretation of experimental results, and to reduce their number. In some cases, simulations can replace experiments when they are either impossible to carry out or too expensive or dangerous.

MD simulations rely on force fields. Force field is a term used to denote a set of parameters and the function which calculates the potential energy of the system using these parameters. Parameters must be transferable, meaning they can be used to model similar systems. The potential energy depends on the geometry and conformation of the molecule, namely on internal coordinates such as bond length, angles, torsion angles and non-bonded interactions (Leach 2001). An example of a simple functional form of a force field that takes into account only the most basic internal coordinates of the molecule is given below.

$$V(r^N) = \sum_{bonds} \frac{k_s}{2}(l - l_0)^2 + \sum_{angles} \frac{k_b}{2}(\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n\omega - \gamma)] + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \right\} \quad (1)$$

$V(r^N)$ is the potential energy which is a function of the coordinates of all particles in the system. The first and second term describe the changes in bond lengths and angles, and are modelled as a harmonic potential which gives the increase in energy as the value deviates from the reference value. The third term is a torsional potential which models energy changes when atoms rotate around a bond. The last term describes non-bonded interactions. They are calculated between each pair of atoms *i* and *j* in different molecules, and between each pair of atoms *i* and *j* separated by at least three bonds when the atoms are in the same molecule. The first term is the Lennard-Jones potential which models van der Waals interactions, while the second one is Coulomb's law used to model electrostatic interactions. Some force fields can have additional terms (Leach 2001).

Force fields differ both in the parameters that are used and in the functional form that is applied to calculate the potential energy of the system. General and specialized force fields exist, the former including parameters for a large variety of atom types, the latter comprising optimized parameters for a group of compounds or molecules. The most commonly used force fields for biological molecules are AMBER, CHARMM, GROMOS and OPLS.

When performing an MD simulation one must carefully choose many simulation conditions, the most important being: how to treat the system in terms of degrees of freedom (use an atomistic or a coarse-grained model), how to solvate the system (implicitly or explicitly), how to treat boundary conditions and to control the temperature and the pressure of the system, and how to sample the conformational space (defined by the choice of sampling algorithm).

An MD production run is often preceded by an energy minimization where unfavourable interactions that increase the energy of the system are relaxed, which allows the system to reach the nearest local minimum on the potential energy landscape. Afterwards, one or more equilibration simulations are performed in order to stabilise all relevant parameters such as temperature, pressure, solvent density, velocities, etc.

3D coordinates of protein structures are an important prerequisite for MD simulations. They are mostly obtained by X-ray diffraction, but a crystal structure shows one moment in time, one of the thousands possible conformations, or sometimes not even, but a constrained structure of a protein in a crystal, which can be far from its physiological state. Since much of a protein's function might be due to its dynamic properties, MD is a valuable technique, and will be even more so in the years to come, with the development of better force fields and stronger supercomputers. This work presents an effort to get a better insight into the interdomain communication and functioning of the EstA autotransporter taking into account its dynamic properties.

2 Materials and Methods

2.1 Simulations of the Passenger Domain of EstA

2.1.1 System preparation

The protein was studied *in silico* using 100 ns long molecular dynamics simulations. Four different systems were simulated: the isolated passenger domain of EstA and the full-length EstA enzyme, both with and without 4-hydroxyphenyl octanoate bound as a tetrahedral intermediate in the active site. Simulations were partially performed at the Ruđer Bošković Institute, courtesy of Dr. Mario Vazdar of the Laboratory for Physical-Organic Chemistry, and at the Forschungszentrum Jülich in Germany, courtesy of Prof. Dr. Karl-Erich Jäger.

The crystal structure of EstA obtained from the Protein Data Bank (PDB ID: 3KVN) was used to prepare the systems for simulation (van den Berg, 2010). The unit cell is composed of two identical EstA protein chains (the RMSD value between them is 0.18 Å), so chain A was chosen to prepare the systems. The N-terminal tag (MHHHHHHHLE-) and the detergent molecules present in the crystal structure were removed. Water molecules within 3 Å from the protein heavy atoms were retained for simulations.

In order to isolate the passenger domain, residues 307-622 were deleted, leaving residues 1-306 which, according to van den Berg (2010), constitute the passenger domain. Parameters from ff99SB Amber force field (Hornak et al. 2006) and GAFF general force field (Wang et al. 2004) were used for parametrisation. In case of the passenger domain with the tetrahedral intermediate in the active site, the intermediate was positioned based on the placement of the detergent molecule in the active site of the crystal structure, on one hand, and on the principle of maximising favourable electrostatic interactions, on the other. A water molecule found in the active site was deleted because of its placement in the oxyanion hole in the crystal structure. When 4-hydroxyphenyl octanoate is placed in the active site, the carbonyl oxygen from its carboxyl group positions itself in the oxyanion hole (Figure 5). The substrate in

tetrahedral form was covalently bound to Ser14 from the catalytic triad. His289, also part of the catalytic triad, was doubly protonated.

Polar hydrogen atoms were added to both the bound and unbound passenger simulation systems by the WHAT IF web server (Vriend 1990) which protonates residues in a way to obtain an optimal hydrogen-bonding network, while non-polar hydrogens were added by the *tleap* program of the Amber 12 program package (Case et al. 2012). The protein was placed in the centre of a cubic box, the size of which was between 70 and 80 Å in all directions for both prepared systems. The box was filled with TIP3P water molecules (Jorgensen et al. 1983) (Figure 6). Periodic boundary conditions (PBC) were applied. The system was electrostatically neutralised by addition of 9 sodium ions. Both prepared systems consisting of the isolated EstA passenger domain comprised about 39000 atoms.

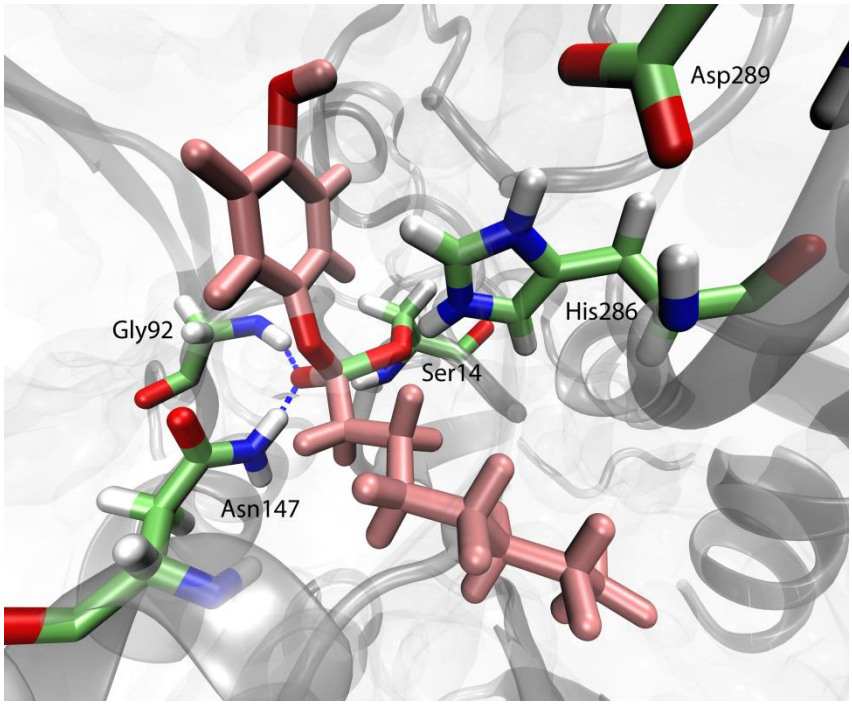


Figure 5. Active site of EstA passenger domain with bound tetrahedral intermediate. 4-hydroxyphenyl octanoate is depicted in pink and covalently bound to Ser14 of the catalytic triad. Its carbonyl oxygen is stabilised by hydrogen bonding with the backbone nitrogen of Gly92 and the amide nitrogen of Asn147, which together shape the oxyanion hole.

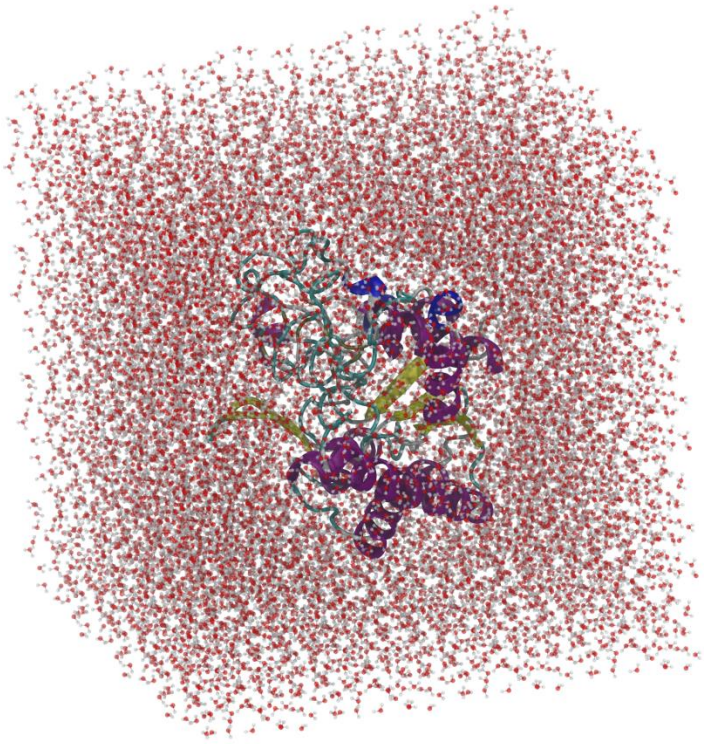


Figure 6. EstA passenger domain in a cubic simulation box solvated with TIP3P water molecules.

2.1.2 Minimisation, Equilibration and MD Simulations

Minimisations and equilibration simulations of the passenger domain of EstA were performed by the Sander program of the Amber 12 program package. For the passenger domain with tetrahedral intermediate the Pmemd implementation of the Sander program was used. Four steps of minimisations consisting of 1000 cycles each were carried out in both cases, each step aimed at optimising the geometry of a different part of the system. This was achieved by restraining motion of a particular part of the system using positional restraints with a force constant of 100 kcal mol^{-1} $Å^{-2}$. In the first minimisation step all protein atoms and substrate atoms were restrained, leaving the water molecules and ions to relax. In the second step protein hydrogen atoms were allowed to move freely as well, and in the system where the tetrahedral intermediate was present, it was no longer restrained. Restraint was put only on the

protein backbone in the third minimisation step. The last minimisation step was carried out without restraints.

After obtaining a structure close to a local energy minimum for both systems, two short equilibration simulations were carried out. During 300 ps of simulation each system was heated up to 310 K at constant volume. This temperature was chosen because the bacterium *P. aeruginosa*, which the protein is isolated from, is among others a human pathogen, and it reproduces at this temperature in the human body. Afterwards, 200 ps of equilibration simulation at constant pressure was carried out for both systems.

MD simulations of EstA passenger domain and passenger domain with bound tetrahedral intermediate were carried out in the Gromacs program package (Berendsen et al. 1995). The first 20 ns in both cases were conducted with Gromacs 4.6.3, and simulations were extended to 100 ns with Gromacs 4.6.5. Simulations were conducted with a time step of 1 fs. The used cut-off scheme was Verlet in combination with Particle-Mesh Ewald (PME) electrostatics; the real space Coulomb cut-off was set to 1 nm, while the Lennard-Jones cut-off was 1.4 nm. For temperature coupling a Nose-Hoover thermostat with time constant 0.5 ps was used (Hoover 1985, Nosé 1984), while isotropic pressure coupling was performed with a Parrinello-Rahman barostat, a time constant for coupling of 10 ps and 1.013 bar as reference pressure (Nosé and Klein 1983, Parrinello and Rahman 1981).

2.2 Simulations of EstA in a Lipid Bilayer

2.2.1 System preparation

The EstA structure file from the Protein Data Bank was adapted similarly as described for the passenger domain in the previous chapter. In this case the full-length chain A composed of 622 amino acids without the N-terminal tag, detergent and water molecules was used.

1-palmitoyl-2-oleoyl-phosphatidylethanolamine (POPE) was chosen to build the lipid bilayer because it is by far the most common lipid molecule in the outer membrane of *Pseudomonas* (Nikaido and Hancock 1986). POPE parameters were downloaded

from people.su.se/~jjm/Stockholm_Lipids (Jämbeck and Lyubartsev 2012b). Water molecules were removed from the POPE coordinate file and the bilayer was extended by doubling its length in the *x*-axis direction. The general lipid force field forcefield.ff (Jämbeck and Lyubartsev, 2012a) was used, together with the Amber99SB protein force field (same as ff99SB used for the passenger domain). The simulation systems of full-length EstA in lipid bilayer were set up in Gromacs.

To prepare the system consisting of full-length EstA with bound intermediate, the position of 4-hydroxyphenyl octanoate in the active site was taken from the previously prepared passenger domain structure. Hydrogen atoms were added to the protein structure by the Gromacs sofware, which protonates amino acid side chains based on their ionization state at pH 7, except His for which protonation is based on an optimal hydrogen bonding conformation. Both the bound and unbound enzyme were inserted into the lipid bilayer in VMD (Visual Molecular Dynamics; Humphrey et al. 1996) based on information from the OPM (Orientations of Proteins in Membranes) database (Lomize et al. 2006). Phospholipid molecules within 0.8 Å of the protein were removed. In both systems a box of about 120 x 60 x 170 Å was generated, with the lipid bilayer spanning the whole length and width of it. An explicit solvent model was used and the box was filled with TIP3P water molecules (Figure 7). Water molecules overlapping with the bilayer were removed. In order to neutralize the systems, 24 water molecules were replaced with sodium ions. This added up to a total of about 120000 atoms in the full-length EstA systems.

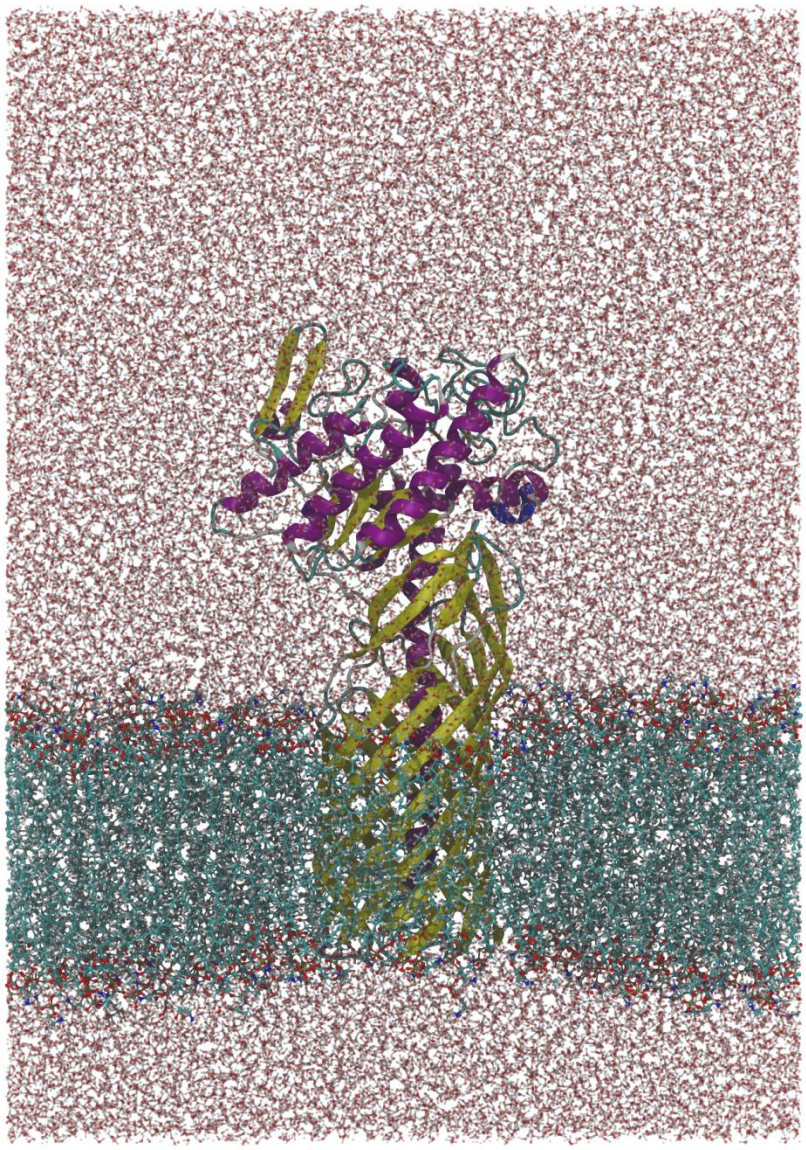


Figure 7. Full-length EstA embedded in a POPE bilayer in a simulation box solvated with TIP3P water molecules.

2.2.2 Minimisation, Equilibration and MD Simulations

Energy minimisation and equilibration simulations were carried out with Gromacs 4.6.2. Both prepared full-length systems were subjected to a steepest descent minimization with the maximum number of steps set to 50000. However, energy minimisation was completed sooner, when the force criterion was met, i.e. the force on every atom was below 1000 kJ mol^{-1} nm^{-1}. The maximum step size was 0.01 nm.

Electrostatics were treated with PME and short-range interactions cut-offs were 1.2 nm. Periodic boundary conditions were applied in all three directions.

The first 500 ps of equilibration simulations were carried out under a canonical (NVT) ensemble with all protein heavy atoms position restrained. The time step was 2 fs. Constraint was put on all bonds with the LINCS algorithm (Hess et al. 1997). PME treatment of electrostatic interactions was used with same cut-offs as in the energy minimisation. Temperature coupling was done with the V-rescale algorithm, treating the protein, or protein and intermediate, the phospholipid bilayer and the solvent as three independent coupling groups (Bussi et al. 2007). The time constant for temperature coupling was 0.1 ps and the reference temperature was set to 310 K in all coupling groups.

An equilibration simulation under the isothermal-isobaric (NPT) ensemble was carried out following the NVT equilibration. Parameters for electrostatic interactions treatment, as well as constraints and restraints were identical to the previous equilibration. The Verlet cut-off scheme was used, and the equilibration simulation was conducted for 1 ns with a time step of 1 fs. The Nose-Hoover thermostat was used for temperature coupling, with the same coupling groups as described above. The time constant for temperature coupling was set to 0.5 ps and the reference temperature was 310 K. Semiisotropic pressure coupling was done using the Parrinello-Rahman ensemble, a 5 ps time constant and a reference pressure of 1 bar.

MD simulations of full-length EstA and full-length EstA with tetrahedral intermediate bound in the active site were conducted with Gromacs 4.6.2 for 30 ns and were afterwards extended to 100 ns with Gromacs 4.6.5. The time step was set to 1 fs. Simulations were performed without constraints. Electrostatics were treated as in the equilibration simulations, along with the Verlet cut-off scheme. Temperature coupling conditions were as in the previous equilibration simulation, while for pressure coupling a 2 ps time constant was used in order to maintain the pressure of 1 bar.

2.3 Analysis

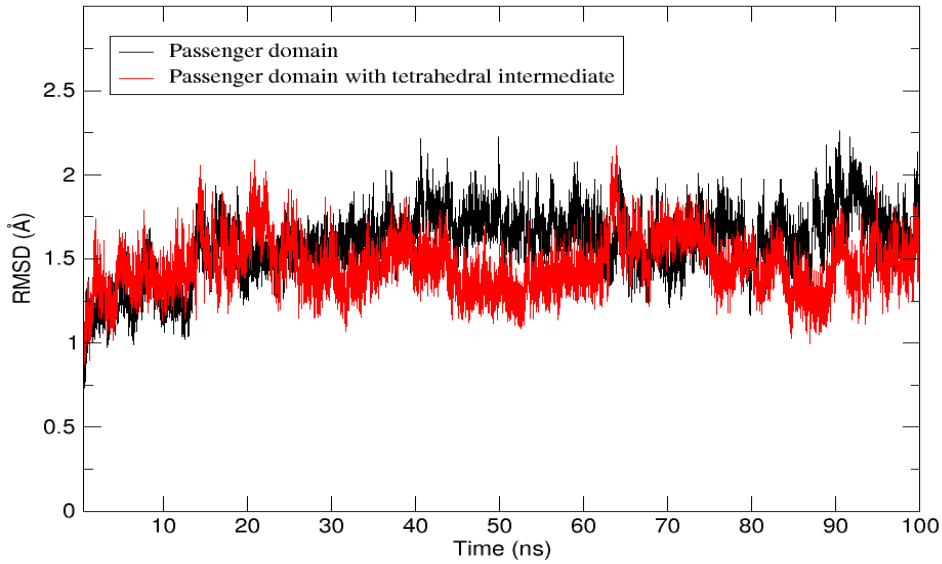
Trajectories were visualized and analyzed using VMD and the tools available in the Gromacs program package.

3 Results

3.1 Structure and Dynamics Overview

All systems reached equilibrium in the simulations, as confirmed by the RMSD plots of the production runs of all four systems (Figure 8). However, the systems simulated without tetrahedral intermediate bound in the active site seemed to take more time to converge towards an equilibrium conformation, possibly because the starting crystal structure had a detergent bound in the active site. The average RMSD value is largest for the simulation of the isolated passenger domain (1.65 Å) and smallest for the passenger domain with tetrahedral intermediate (1.47 Å); the average RMSD for the full-length enzyme with intermediate is very close to this value as well (1.49 Å), and the full-length enzyme without intermediate bound in the active site lies somewhere in between (1.6 Å). The largest difference between minimum and maximum RMSD values was found in the simulation of the passenger domain with tetrahedral intermediate (1.18 Å), whilst the smallest fluctuation between minimum and maximum was in the simulation of full-length EstA without intermediate in the active site (0.94 Å). Equilibrated simulations were considered for these calculations, i.e. trajectories from 20 ns onwards.

a)



b)

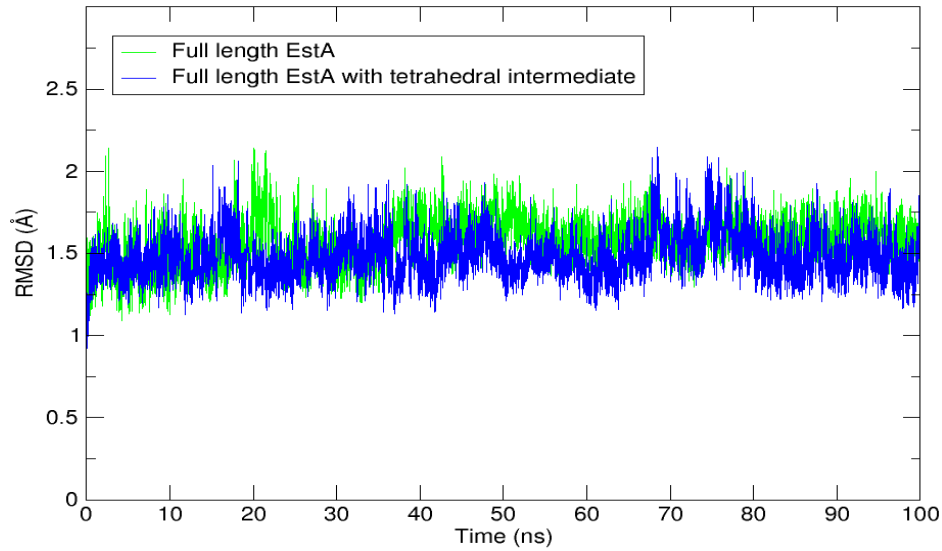


Figure 8. Root mean square deviation plots of EstA MD simulations. a) EstA passenger domain with and without tetrahedral intermediate. b) Full-length EstA with and without tetrahedral intermediate.

Cluster analysis was carried out on whole trajectories using the *gromos* clustering algorithm. As opposed to the RMSD calculations in the previous paragraph which were based on Cα atoms in relation to the equilibrated protein structure, in the case of clustering all structure pairs were compared and grouped based on the calculated RMSD values for the catalytic domain. Using the same cut-off value of 2 Å for all simulations, the conformations were grouped into 5 clusters for the passenger domain simulation, 4 clusters for the passenger domain with bound intermediate, 3 clusters for the full-length enzyme simulation, and only one cluster in the case of full-length EstA with bound intermediate. Hence full-length EstA with tetrahedral intermediate shows the smallest catalytic domain structural variability throughout the simulation, whilst the isolated passenger domain simulation exhibits the biggest variability. Overall, in both simulations of membrane bound full-length protein there was more similarity among the sampled structures.

When the cluster centroids (representative structures of each cluster) were superimposed for all systems, the largest RMSD differences were found in several protein regions: part of helices 10 and 11, the bottom and in some cases top part of helix 9 with the loop following the helix, part of helix 5 and its adjacent loop, and helix 3 and the adjoining unstructured region (Figure 9). These differences were mostly consistent in all simulated systems, with the exception of helices 10 and 11, which fluctuate considerably only in the simulations of isolated passenger domain. When the centroids of the most populated cluster of each system are superimposed, a difference in the active site helix 6 is conspicuous, in addition to the previously mentioned flexible regions - in the simulation of the full-length enzyme with tetrahedral intermediate the bottom part of helix 6 and the adjacent turn adopt a different conformation (Figure 10).

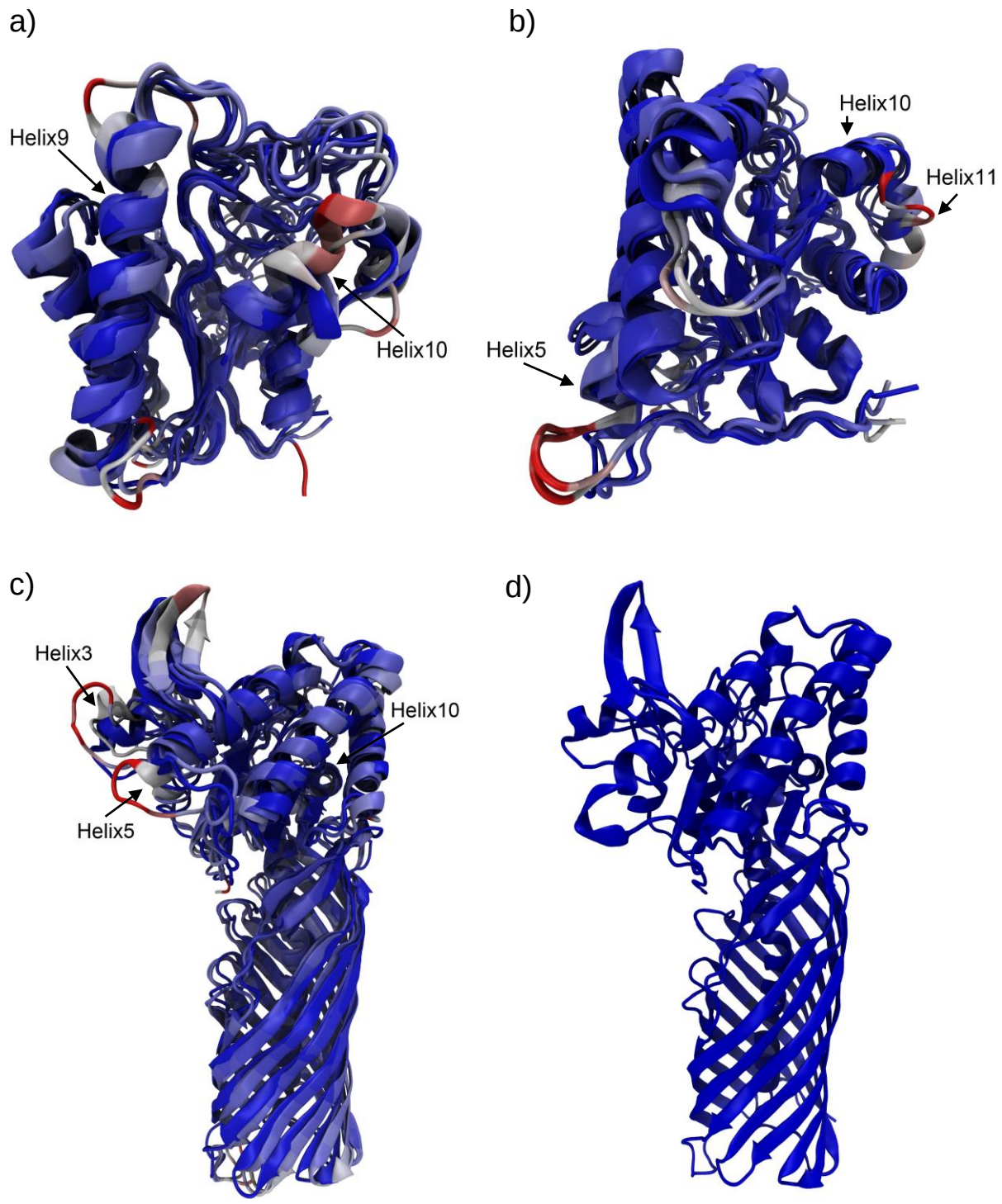


Figure 9. Superimposed centroids of each cluster for all four systems. Structures are coloured by RMSD (blue – lower, red – higher). a) Simulation of isolated passenger domain (5 clusters). b) Simulation of isolated passenger domain with intermediate (4 clusters). c) Simulation of full-length protein (3 clusters). d) Simulation of full-length protein with intermediate (1 cluster). The cut-off for cluster analysis was set to 2 Å.

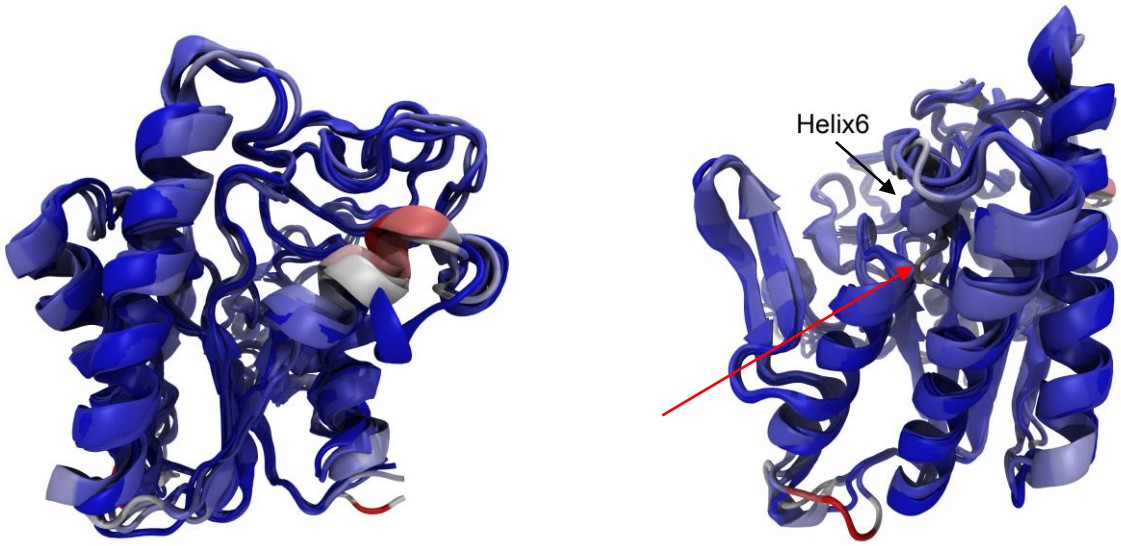


Figure 10. Superposition of the dominant cluster centroids of all four simulation systems (only passenger domains). Colouring is based on RMSD as in the previous figure. The turn beneath helix 6 is structured differently in the system of full-length enzyme with tetrahedral intermediate (red arrow).

The root mean square fluctuation (RMSF) per residue averaged for the regions pointed out by the cluster analysis was higher than the average RMSF for the whole protein in the great majority of cases. The highest peaks were associated with the isolated passenger domain without intermediate, with the most prominent flexible regions being helix 9 and the following loop, as well as helix 5 with the adjoining loop (average RMSF of 2.28 and 2.09 Å respectively, in contrast with a protein average of 1.12 Å). This is consistent with the cluster analysis unveiling the biggest number of clusters for this simulation system. The average RMSF values of helix 9 with its adjoining loop and helix 10 are higher in the isolated passenger domain systems. On the other hand, helix 3 and its adjacent loop tend to be more flexible in the full-length enzyme simulations (Figure 11).

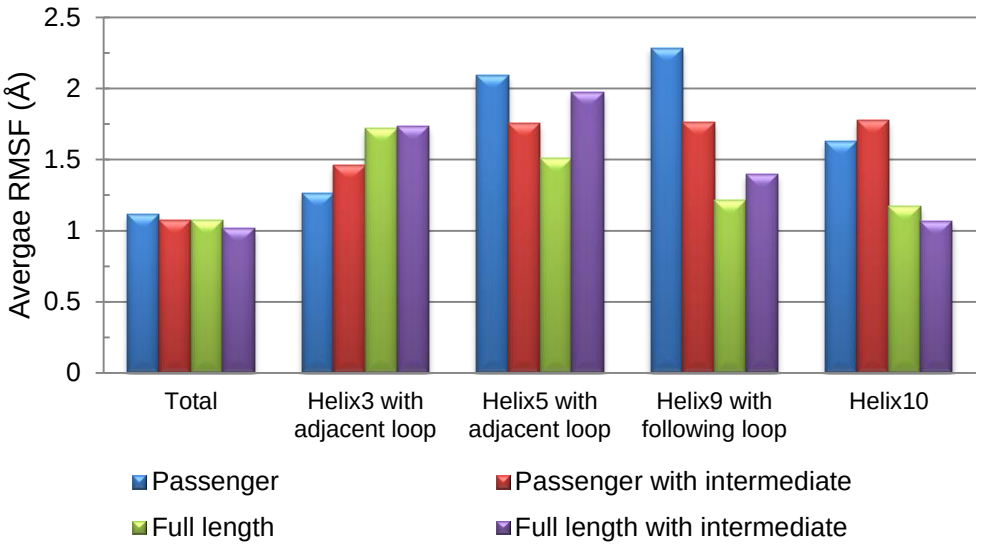


Figure 11. Average RMSF values of flexible regions found by superposition of cluster centroids.

3.2 Interdomain Interactions

3.2.1 Hydrogen Bonds

Hydrogen bonds between the passenger and autotransporter domain of EstA can be divided into three major groups or interaction clusters, i.e. domain-domain interface regions into which interactions can be grouped (Figure 12).

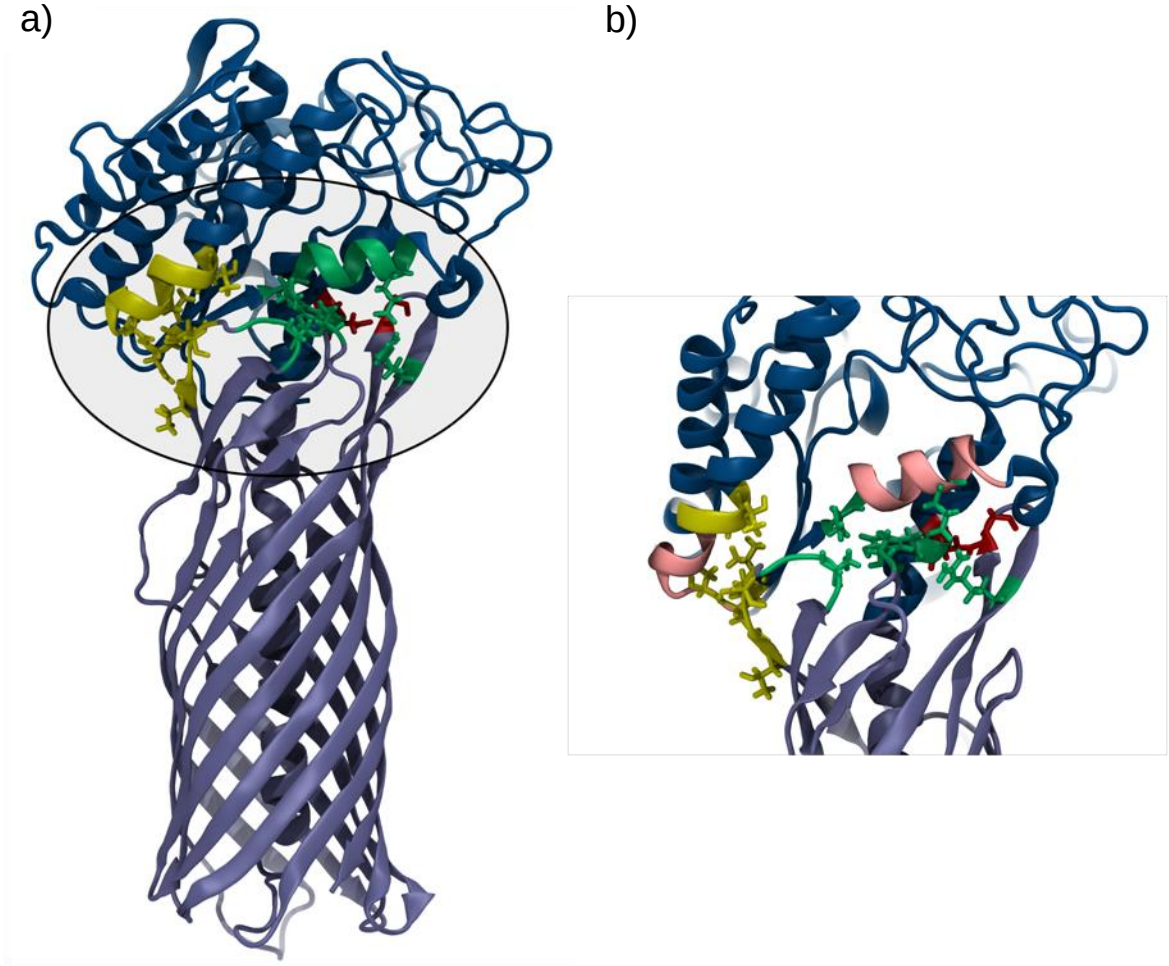


Figure 12. Interdomain hydrogen bonding regions in EstA. a) The passenger domain is coloured dark blue, the AT domain is icy blue. The first interaction region (sheet D cluster) is shown in yellow, the second one (helix 10&13 cluster) in green, and the third one (a single interaction between the loop connecting strands 5 and 6 and helix 13) is coloured in red. b) A slightly rotated representation, with colouring as in a). Domain-domain interface regions with a diminished RMSF in the simulations of full-length EstA are shown in pink.

The first interaction cluster includes interactions between residues of the AT domain located on sheet D and residues of the passenger located on the bottom part of helix 9 and the last strand of the parallel β sheet (Figure 13a). The RMSF per residue for the bottom part of helix 9 and the loop connecting it to the aforementioned β strand (i.e. amino acids 220-226) is larger in the simulations of the passenger domain, when

there are no interdomain interactions. When comparing the two simulations of the passenger domain, the flexibility of this region is larger when there is no tetrahedral intermediate bound. The interactions that form this cluster are conserved in both simulations of full-length EstA (with and without bound tetrahedral intermediate). However, while the hydrogen bond between Thr570 and Asn226 is the same in both cases, Asn567 interacts much more frequently with Val227 when no tetrahedral intermediate is bound to the enzyme (64.3% of the simulation time when a 3.5 Å cut off distance between donor and acceptor and a 30^{o} acceptor – donor - hydrogen angle are set as hydrogen bonding criteria), and more often with Thr217 when the intermediate is bound (71.5% of the simulation time); in the latter case hydrogen bonding between Asn567 and Val227 never occurs. As opposed to full-length EstA without bound intermediate, where interactions of Asn567 with Leu220 and Ala225 occur during the rest of the simulation, in the system of full-length protein with bound intermediate hydrogen bonding with Leu220 is negligible, leaving the Asn567-Ala225 bond to compensate for the broken Asn567-Thr217 interaction when necessary (Figure 14).

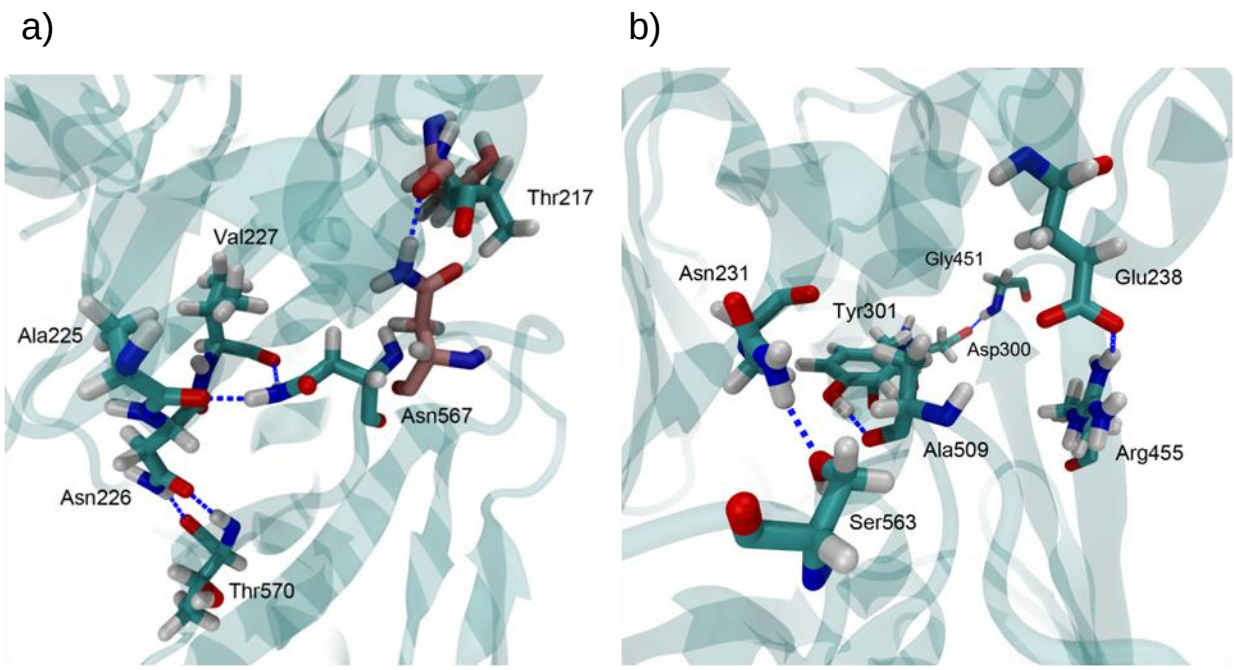
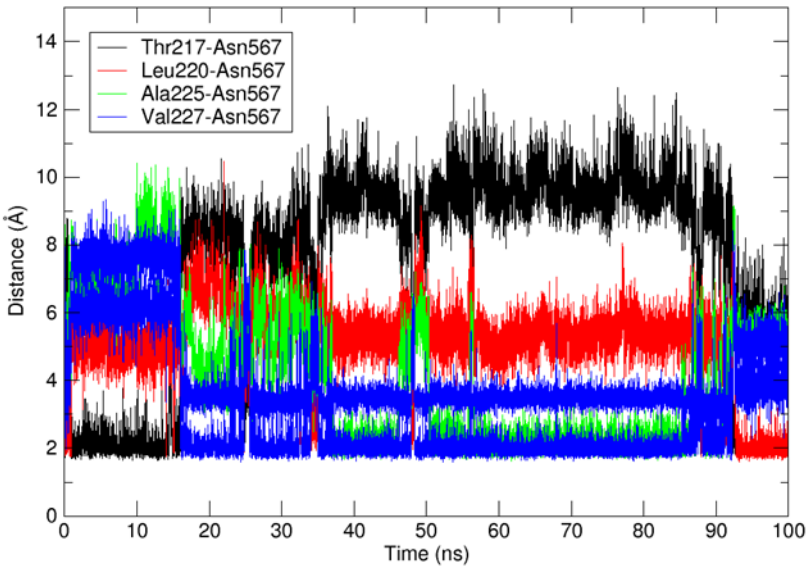


Figure 13. Close-up of interdomain hydrogen bonding in EstA. The structure of full-length EstA without bound intermediate at 75 ns is shown (ribbons and cyan coloured residues). Residue numbers ≥307 belong to the AT domain. a) Interdomain hydrogen bonds between residues on sheet D of the transmembrane barrel and the bottom of helix 9 and a β strand of the passenger domain. Pink residues are from the superimposed structure of full-length EstA with bound intermediate at 75 ns, showing the interaction that differs among the two systems. b) Hydrogen bonds belonging to the other two interaction regions, in which residues of helix 10 and the upper part of the membrane spanning helix of the passenger domain interact with residues from two loops and strands of the AT domain.

a)



b)

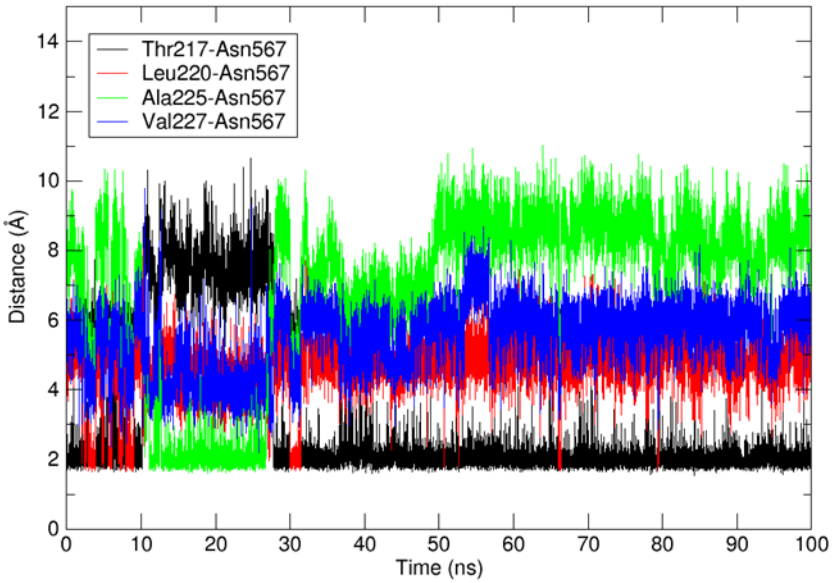


Figure 14. Interdomain hydrogen bonding between Asn567 and residues of the passenger domain during 100 ns of simulation. a) Plot of said interactions in full-length EstA without intermediate in active site. Asn567 hydrogen bonds mostly with Val227, as well as Thr217, Leu220 and Ala225. b) Plot of interactions in full-length EstA with bound tetrahedral intermediate. Asn567 hydrogen bonds mostly with Thr217, and to a smaller amount with Ala225.

Interdomain hydrogen bonding also occurs between residues of helices 10 and 13 (i.e. the upper part of the membrane-spanning helix) of the passenger and residues of the AT domain located on loops adjacent to sheet D and a prolonged strand of the β barrel (strand 6) (Figure 13b). In this case the interactions affect the flexibility of a participating region as well - the RMSF of the majority of helix 10 (amino acids 233-242) is higher in both simulations of the passenger domain (the average RMSF is 1.63 Å in the passenger, 1.77 in the passenger with intermediate, 1.20 in the full-length enzyme and 1.06 in the full-length enzyme with intermediate). All interactions included in this cluster are either less occurring or more fluctuant in the simulation of the full-length protein with tetrahedral intermediate. The most stable interaction is the hydrogen bond between Tyr301 of the membrane-spanning helix (helix 13) and Ala509, which breaks only shortly around 50 ns in the system with tetrahedral intermediate. On the other hand, the interaction between Glu238 and Arg455 breaks several times in this simulation system, mainly due to Arg455 fluctuations. The higher RMSF per residue of the 506-509 region in the simulation of full-length EstA with tetrahedral intermediate can at least partially account for the lower stability of interactions in this system.

The last group of hydrogen-bonded residues between domains consists of an isolated interaction between amino acid Gly451 (and Gly452 to a much smaller amount) on the turn connecting the two prolonged strands of the β barrel (namely strands 5 and 6) and residue Asp300 of the membrane-spanning central helix (helix 13). The flexibility of the membrane-spanning α helix doesn't differ between systems. As opposed to the interactions in the prior hydrogen bonding cluster, in this case the interaction is more perseverant in the system with bound intermediate (Figure 13b).

As for the interactions subdivision, the first described interaction cluster can be much more easily discerned than the other two. The above subdivision of these two clusters is based on the relatedness of the hydrogen bonds they engage into. Another possible subdivision would be based on the structural moieties in which the interacting amino acids are found; in this case one group would contain all the interactions of helix 10 with the AT domain, namely Asn231-Ser563 and Glu238-Arg455, while the last cluster would comprise all the interactions of helix 13, i.e. the upper part of the membrane spanning helix, namely Tyr301-Ala509 and Asp300-Gly451. However, for

the purpose of this work the first and functionally more logical subdivision has been chosen. An overview of the interactions grouped into three clusters is presented in Table 3.

Occasional interdomain hydrogen bonds are also formed between N-terminal residues during the simulations, namely between the amino group of Ala1 and the hydroxyl group of Ser362, and the backbones of Ser3 and Asp364. These contacts are interchanged with hydrogen bonds between these N-terminal amino acids and residues of the passenger domain. As a consequence of these interactions the N terminus is far less flexible when the protein is simulated with its transmembrane domain.

Table 3. Overview of hydrogen bonding interactions between the passenger and autotransporter domain of EstA.

Hydrogen bonding interaction region	Donor residue : Acceptor residue[a,b]	Donor atom/group – Acceptor atom/group[c]	Comparison of occurrence during simulation (protein vs. protein with intermediate)[d]
Sheet D	Asn226 : Thr570*/**	-NH2 — O_{bb}-	=
	Asn226 : Thr570	-NH2 — OH	
	Thr570 : Asn226	-NH_{bb} — O=C-	=
	Asn567 : Thr217**	-NH2 — O_{bb}-	=
	^{X}Asn567 : Leu220	-NH2 — O_{bb}-	
	Asn567 : Ala225	-NH2 — O_{bb}-	
	^{X}Asn567 : Val227*	-NH2 — O_{bb}-	
Helices 10 & 13	Asn231 : Ser563	-NH2 — OH	>
	Arg455 : Glu238	-NH2 — $^{-}$OOC-	>
	Tyr301 : Ala509	-OH — O_{bb}-	>
Turn between β strands 5 & 6	Gly451 : Asp300	-NH_{bb} — $^{-}$OOC-	<

[a] Interactions which alternate in the simulation and joined represent a single continuous interacting entity are grouped and considered together. In such a group, an interaction which stands out for being dominant during most of the simulation time is marked with an * for full-length EstA, ** for full-length EstA with bound tetrahedral intermediate, and */** for an interaction dominant in both systems.

[b] Hydrogen bonds which don't occur in EstA with bound intermediate are marked with an ^{X} preceding the name of the donor amino acid (interactions which would be found only in EstA with intermediate haven't been encountered).

[c] The index $_{bb}$ specifies a backbone atom, other atoms/groups belong to residue side chains.

[d] Only hydrogen bonds which are conserved throughout one or both simulations of full-length EstA are relevant and considered; differences between the two systems are reported, where > means greater occurrence in protein without intermediate, < means greater occurrence in protein with intermediate, and = means equal hydrogen bond occurrence in both simulations.

3.2.2 Hydrophobic Interactions

Hydrophobic interactions are not strictly defined between two atoms or even residues, as opposed to hydrogen bonds. A 5 Å cut-off was used to determine the existence of hydrophobic interactions. Interdomain hydrophobic interactions in EstA form one large hydrophobic cluster, with participating residues from almost all of the above encountered domain interface regions – sheet D and its attached loops, the last two strands of the passenger domain β sheet and the bottom part of helix 9, the prolonged strand 5 of the β barrel, helix 10 and the membrane spanning helix 13 (Figure 15a).

The amino acids engaging into interdomain hydrophobic interactions as part of the interaction region referred to previously as sheet D cluster are Leu186, Ile228 and Pro229 on the passenger domain β sheet, Thr217 on helix 9 and, on the transmembrane domain, Leu564 and Phe569, located on the first strand and connected loop of sheet D, respectively. The remaining two interaction regions described in the previous chapter cannot be discerned when considering hydrophobic interactions. Residue Leu234 from helix 10 interacts with Ala509 located on a loop adjacent to sheet D on the AT domain. Leu235, also located on helix 10, seems to have an important role, connecting helix 10 to the membrane spanning central helix through its interaction with Leu297, to the loop on sheet D through interactions with Ala509 and Thr510, and to the prolonged strand 5 of the transmembrane domain by interacting with Leu450. Other residues involved in hydrophobic interdomain interactions as part of this cluster are Leu512, Leu561 and Leu305, as well as Phe246 on the opposite side, which is the only residue from helix 11 (a small 3_{10} helix following helix 10) participating in interdomain interactions (Figure 15b).

Tyr301 has probably a central role in keeping the cluster compact, and achieves this by interacting with residues from all interface regions. Moreover, it interacts with Leu230 which, along with Ile232 from helix 10 and Ile298 of the central helix, engage in hydrophobic interactions with Trp185, a residue associated with the active site (Figure 15b). Interestingly, although the flexibility of most of helix 10 depends on the presence of the AT domain, the RMSF of residues 230 and 232 seems to be influenced by other factors. These two residues show highest RMSF in full-length EstA without

intermediate, and lowest RMSF in full-length EstA with bound intermediate in the active site. As Trp185 shows the same RMSF pattern, the flexibility or rigidity of these residues is probably interdependent and mediated by hydrophobic interactions.

Differences in hydrophobic interactions between the two full-length EstA simulations are generally consistent with the differences already reported for hydrogen bonding. Most interactions that include residues 509 and 510, which are located on a loop with higher flexibility in the system with tetrahedral intermediate, are consequently less stable in this system. On the other hand, interactions of residues on the turn between strand 5 and 6 of the barrel and helix 13 residues are more stable in the system with tetrahedral intermediate.

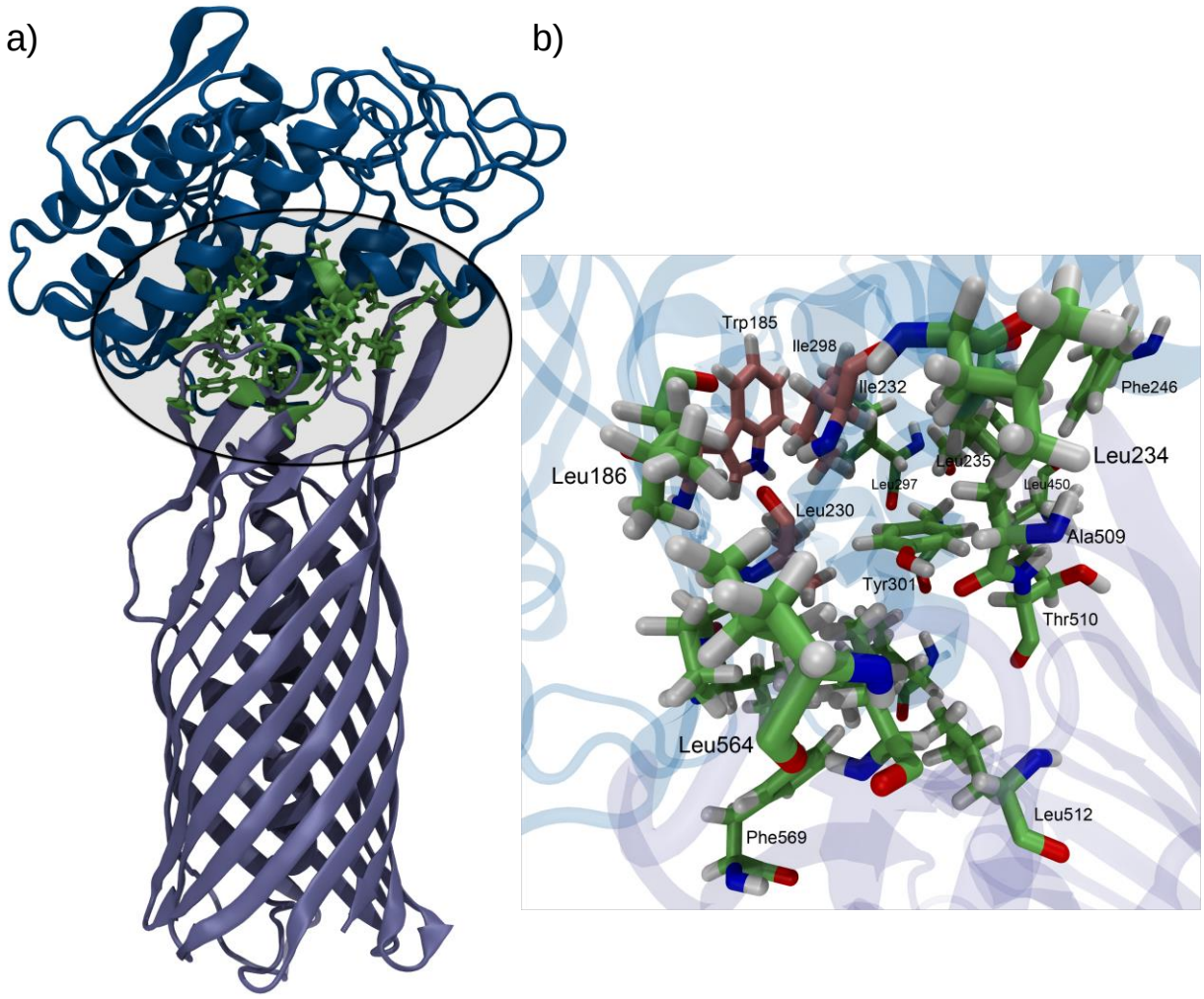


Figure 15. Interdomain hydrophobic cluster in EstA. a) The passenger domain is represented in dark blue, the AT domain in icy blue. The regions involved in hydrophobic interactions and the associated residues are shown in green. b) Residue Trp185 associated with the active site and residues 230, 232 and 298 which interact with it are shown in pink. Tyr301 has a central role interacting with Leu230, Leu235, Leu305 (not visible), Leu450 and Thr510.

3.3 Active Site Geometry and Hydrogen Bond Network

3.3.1 Active Site Architecture and Substrate Binding

The catalytic triad of EstA is composed of Ser14 on one loop, and Asp286 and His289 on another (Figures 3b and 5). The hydrogen bonding of the catalytic triad residues was conserved in both simulations with tetrahedral intermediate in the active site. It was, however, not conserved in the systems without the intermediate in the

active site, where there is no hydrogen bond between Ser14 and His289. This reflects on the distance between these two loops, which is stable and constant in the systems with a formed catalytic triad, but smaller and fluctuant in the systems without the intermediate in the active site. Fluctuations are the largest in the full-length enzyme without intermediate in the active site. In the systems with no tetrahedral intermediate bound in the active site, Ser14 engages into hydrogen bonding with Gly92 and occasionally with Asn147. When the intermediate is bound in the active site, these residues form the oxyanion hole which stabilises the transition state oxyanion. Ser14 also contributes to oxyanion stabilisation with its backbone nitrogen. In addition to the two loops with the catalytic triad residues, Gly92 of the oxyanion hole and Phe34 which stabilises the intermediate by stacking are found each on a different loop. Another important structural element in the active site is helix 6, which carries Asn147 of the oxyanion hole. The RMSF values of the catalytic triad and oxyanion hole are lower in the passenger with intermediate than in the full-length enzyme with intermediate. This is especially prominent in the oxyanion hole residues (Figure 16). While this seems to have little or no effect on the oxyanion stabilisation by Gly92, hydrogen bonding between Asn147 and the tetrahedral intermediate oxyanion is more stable, i.e. with smaller length fluctuation, in the active site of the isolated passenger domain (Figure 17).

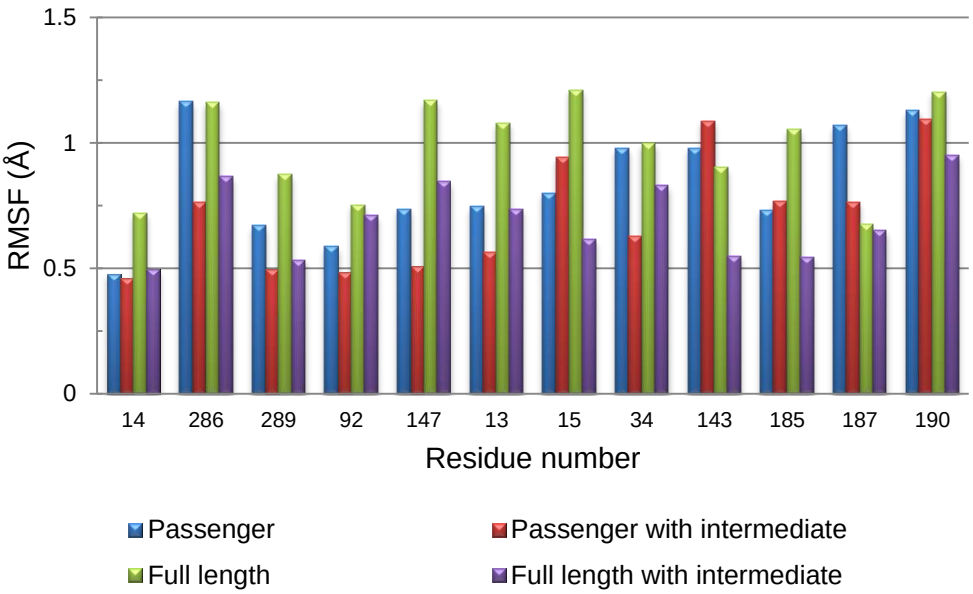


Figure 16. RMSF of chosen relevant active site residues. Residues 14, 286 and 289 form the catalytic triad, while residues 92 and 147 shape the oxyanion hole.

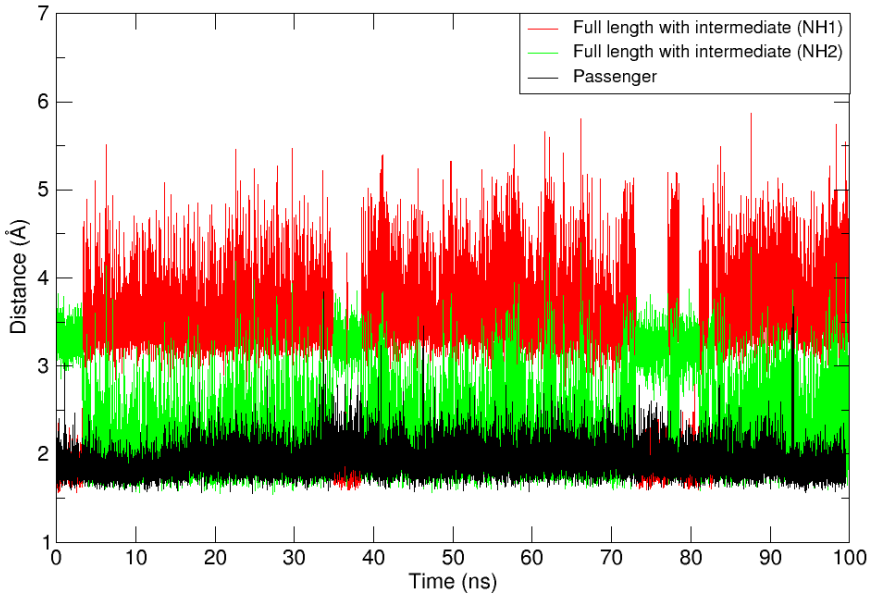


Figure 17. Hydrogen bond between Asn147 of the oxyanion hole and the tetrahedral intermediate. In the full-length enzyme both hydrogens of the amide group engage in hydrogen bonding at different times in the simulation.

The remainder of tetrahedral intermediate – enzyme interactions is hydrophobic in nature. Relevant residues for hydrophobic substrate stabilisation are Leu15, Leu150, Val288, Leu187, Pro188, Leu190, and Phe34 (Figure 18). The average RMSF of the tetrahedral intermediate is slightly larger in the passenger domain in comparison with the full-length enzyme (1.25 vs. 0.80 Å). The cause of this lies in the flexibility of part of the tetrahedral intermediate that includes the phenol moiety, and of its final 2-3 carbon chain atoms. Only one position is allowed in the full-length enzyme active site, while in the isolated passenger two orientations are possible for both groups, and they are not interdependent (Figure 18). Flexibility of the part of the tetrahedral intermediate that includes the phenol moiety is probably inhibited by Leu150 in the full-length enzyme, while the inhibition of motion of the last few carbons might be related to residues Leu187 and Leu190 which are also part of a non-active site hydrophobic cluster including residues from helix 6 (Phe149), helix 9 and helix 7 (described in the following chapters). The interactions of Phe34 and Ser14 as oxyanion stabiliser with the tetrahedral intermediate are influenced by the flexibility of the phenyl part of the intermediate. Therefore in the isolated passenger the distance plots show two leaps during 100 ns of simulation, the first when the part of the intermediate that includes the phenol moiety adopts a flexed position around 15 ns, and the second one at about 90 ns when it returns to the starting position. Interactions of the intermediate with Leu187, Pro188 and Leu190, on the other hand, depend on the position of the chain end of the intermediate. They are conserved in the full-length enzyme, but are broken in the isolated passenger whenever the chain end changes orientation (in the beginning, end and transiently in the middle of the simulation).

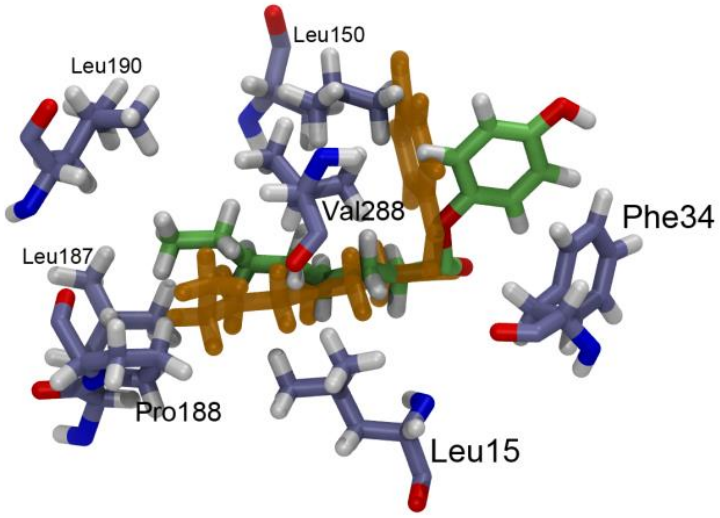


Figure 18. Tetrahedral intermediate stabilisation by hydrophobic interactions in the active site. All residues and the intermediate in light green are from the full-length enzyme. The intermediate coloured in ochre is from the simulation of passenger with intermediate, showing the alternative positioning of the part of the intermediate that includes the phenol moiety and of the chain end, which sometimes occur during the simulation of this system.

3.3.2 Active Site Interactions and Hydrogen Bond Network

Hydrogen bonds in the active site of EstA keep the site's integrity and stabilise catalytic residues and groups in the positions necessary for catalysis. Hydrogen bonds which are conserved through all the systems shape the active site loops and position them in relation to one another. The bond between the backbone of Val90 and Tyr93 shapes the loop with the oxyanion hole residue Gly92. The bonds between the backbone of Asp13 and the backbones of Ser16 and Asp17 shape the loop with the catalytic Ser14 (with the exception of the full-length enzyme without intermediate in the active site where the bond Asp13 – Asp17 is intermittent). The loop itself is spatially positioned with two hydrogen bonds, each extending from one side of it towards the adjacent loop with Gly92; these hydrogen bonds are Gly12 – Val90 and Asp17 – Gly91 (Figure 19).

The remaining hydrogen bonds span the active site and their presence and stability depend on the presence of the tetrahedral intermediate and of the AT domain. Conserved water molecules throughout the whole simulation are important to stabilise the active site when the tetrahedral intermediate is bound. In the simulation of the

passenger without intermediate, water molecules are present periodically in the active site, while there are no active site water molecules in the full-length enzyme without intermediate. An overview of all interactions shaping the active site and structuring it for catalysis is given in Table 4.

3.3.3 Hydrogen Bonding in Active Site without Tetrahedral Intermediate

The full-length enzyme without intermediate has the most fluctuant active site with few conserved hydrogen bonds compared to the other three simulated systems. Even the aforementioned structurally important bonds are fluctuant and in some cases intermittent in this system, but their combination is such that at every time the general secondary and tertiary structures are conserved. In this system Ser14 is stabilised by Gly92 of the oxyanion hole. Asp13 fluctuates considerably (Figure 15) and it is stabilised by hydrogen bonding with the backbone of Leu15, whilst when it switches position it hydrogen bonds with the backbone with Gly92 and Tyr93 of the adjacent loop. In addition, Asp13 is occasionally stabilised by hydrogen bonding with Asn147, although both residues are very flexible and the bond is therefore very unstable. The described main part of the active site, which includes residues around the nucleophile and oxyanion hole, is connected with Trp185 in the remote part of the active site by hydrophobic interactions between said tryptophan and mainly Leu15. Trp185 engages into close hydrogen bonding with Tyr141, which is conserved except for a few short interruptions during the simulation.

In the passenger domain simulation without bound intermediate Ser14 and His289 are too far away to engage in hydrogen bonding, although their distance is mostly constant. That is because Ser14 is fixed in position thanks to very close hydrogen bonding with Gly92, and it also forms a more fluctuant hydrogen bond with Asn147. Asp13 forms hydrogen bonds with Asn147 and Leu15 which are more stable than in the full-length enzyme without bound intermediate described above. At the same time, there is no bond between Asp13 and the loop which carries the oxyanion hole residue Gly92 because of smaller fluctuation of Asp13 and Asn147 and their different positioning (the side chain of Asp13 is oriented towards the active site centre, and never towards the Gly92 loop). In addition, in this system water molecules occasionally enter and form

hydrogen bonds in the main part of the active site. These bonds however are not conserved throughout the simulation. There can be one, two or three water molecules hydrogen bonding in the main part of the active site, and when they are present they have a bridging role, connecting the side chain of Asp13 with the backbone of residues 145-148 from helix 6 and the backbone of Tyr93 of the Gly92 loop. When present, these water molecules allow for a compact cluster to form in the upper main part of the active site, connecting helix 6, the Ser14 loop and the adjacent Gly92 loop. As for the region connecting the main and remote part of the active site, Asp13 hydrogen bonds closely with Thr143 in the first 37 ns of simulation, after which Thr143 switches from facing upwards (towards Asp13 and the main part of the active site) to facing downwards (towards Trp185 and the remote region). However, in this new position it does not engage into close interactions. Trp185 is connected to the main active site region by hydrophobic interactions mainly with Leu15, as in the previously described full-length enzyme system. However, in this case Trp185 hydrogen bonds with Tyr141 only during the first 20ns of simulation. A water molecule is present occasionally in this part of the active site as well, connecting Trp185 with Gly144 from helix 6.

3.3.4 Hydrogen Bonding in Active Site with Tetrahedral Intermediate

When the tetrahedral intermediate is bound in the active site, the catalytic triad residues are engaged into tight hydrogen bonding to one another. Ser14 is covalently bound to 4-hydroxyphenyl octanoate, and Gly92 and Asn147 stabilise the oxyanion through electrostatic interactions. The active site is more rigid and interactions are more stable in comparison with the enzyme without tetrahedral intermediate in the active site. However, there are significant differences between the passenger domain and the full-length enzyme during the simulations that include the tetrahedral intermediate.

In the passenger domain with bound intermediate Asp13 has the lowest RMSF of all simulated systems and only in this system it doesn’t rotate, but interacts with Asn147, which is also very rigid in this system (Figure 16). Consequently, this interaction is realised with the same oxygen atom from the carboxylic group of Asp13 throughout the simulation. The hydrogen bond is not very tight and breaks two times during the simulation, similarly as in the passenger without bound intermediate. The

side chain of Asp13 is also stabilised by Leu15, as in the systems without intermediate. However, two or three conserved water molecules are crucial to keep the main part of the active site compact and functional. They are constantly present during the simulation, serving the same role as previously described for the passenger without intermediate, although in that case they were present only occasionally. There are only four water molecules which interchange and occupy three or sometimes two positions in the upper main part of the active site, and they connect residues 144-147 from helix 6 with Asp13 and Tyr93, forming a hydrogen bonding cluster. The connections to the remote part of the active site with Trp185 are similar as in the passenger without intermediate in the active site. The hydrogen bond of Asp13 and Thr143 is again intermittent, and the main connection path is through hydrophobic interactions of Leu15 and Trp185, which are more stable than in the previously described systems. In addition, as seen for the passenger without intermediate, a water molecule occasionally connects Gly144 from helix 6 to Trp185 by hydrogen bonding. This water molecule is always one of the four molecules which are associated with the upper main part of the active site. The hydrogen bond of Trp185 with Tyr141 is fluctuant and exhibits a few larger interruptions; therefore it cannot be considered relevant for enzyme activity or active site stability in this system. The described hydrogen bond network is shown in Figure 19a.

In the passenger with bound intermediate the active site hydrogen bonds were similar as in the passenger without bound intermediate, only in some cases more stable, tighter, certain residues were more rigid, and the water molecules were conserved. The full-length enzyme with intermediate, on the other hand, differs considerably in terms of hydrogen bonds (Figure 19b). Helix 6 has in part lost its ordered α helicity in this system, and in the bottom part of the helix a kink was introduced, bringing residues Gly145 and Gly146 closer to Asp13. They engage into stable hydrogen bonding, while there is no hydrogen bonding of Asp13 with Asn147. Asn147 has a different position than in the isolated passenger with intermediate, stabilising the oxyanion from above, rather than from below. Asp13, on the other hand, has retained a similar position as in the other systems, therefore it hydrogen bonds with Leu15. In addition, Asp13 engages into tight hydrogen bonding with Thr143, which is

not the case in any other simulated system. Thr143 is the first residue of the remote active site region and has the lowest RMSF in this system, as is the case for all other residues of this region (Figure 16). Moreover, Thr143 forms a conserved, although fluctuant hydrogen bond with Ser16. It also engages into tight and stable hydrophobic interactions with Trp185. In the other systems the distance between these two residues was very fluctuant with sharp changes throughout the simulation. Another connection with the main part of the active site goes through hydrophobic contacts of Leu15 and Trp185, with this interaction being more stable and tight than in the other systems. As for Trp185, it engages in a stable and close hydrogen bond with Tyr141 during the whole simulation. Although there are no water molecules in the main part of the active site like in the previous system, there are three conserved water molecules in the remote part of the active site. These water molecules, always the same three, form a water molecule chain during the whole simulation and connect residue Asp13 with Gly144 from helix 6 and Trp185. In the last few nanoseconds the bond with Asp13 breaks and two water molecules maintain the structure and compactness of the remote part of the active site, connecting Gly144 and Trp185, while the third water molecule leaves the active site. Interestingly, in the isolated passenger with bound intermediate the compactness of the upper main part of the active site seemed to be relevant, while in the full-length enzyme with intermediate interactions are organized in order to keep the remote part around Trp185 rigid and well connected to the main part.

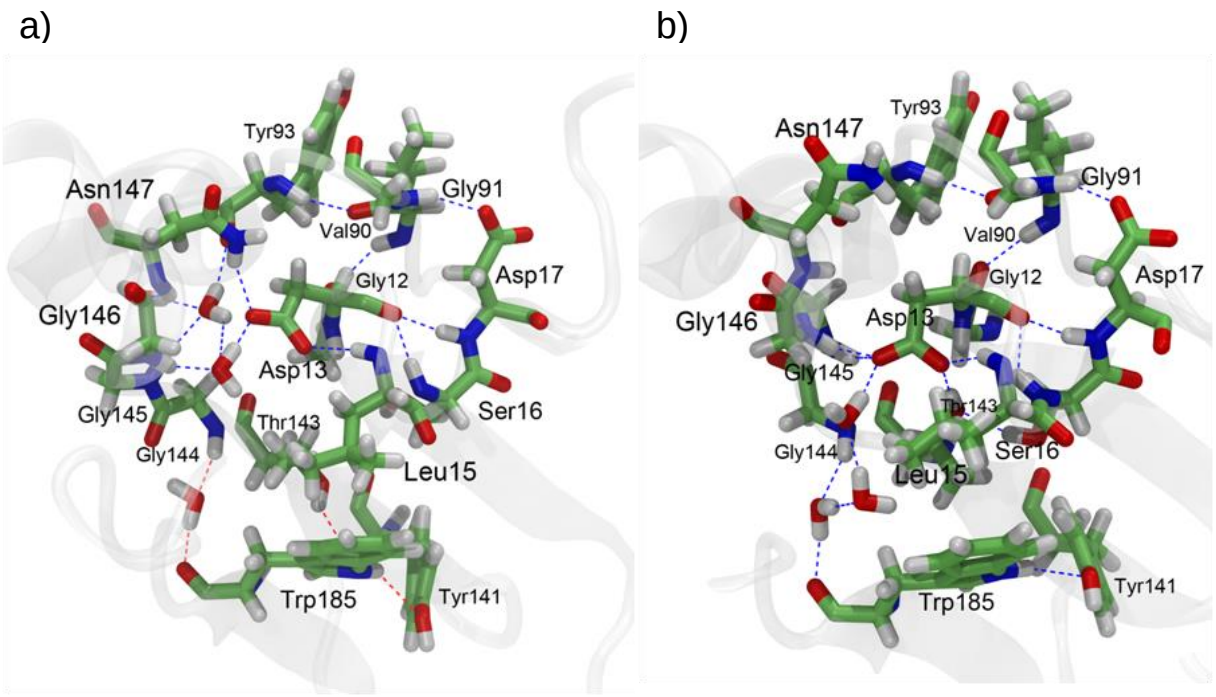


Figure 19. Active site hydrogen bonding in systems with tetrahedral intermediate (intermediate not shown). Frames are taken at 50 ns. a) Isolated passenger with tetrahedral intermediate. Three conserved water molecules in the main upper part of the active site are important in stabilising it for catalysis. Hydrogen bonds conserved throughout the simulation are depicted in blue, intermittent bonds that happened to be present in this frame are shown in red. The same bonding pattern, with some difference in bond conservation and occurrence, is present in the simulations without intermediate in the active site. b) Full-length EstA with tetrahedral intermediate. Three conserved water molecules connect the main and the remote part of the active site. All shown bonds are conserved throughout the simulation. Interactions 90-93, 13-16, 13-17, 12-90, 17-91 keep the loop with Ser14 and the loop with Gly92 structured and in position relative to one another in all systems.

Table 4. Overview of active site interactions in four different simulated systems.

	Interaction[a]	Passenger[b]	Full-length[b]	Passenger with intermediate[b]	Full-length with intermediate[b]
Catalytic triad	Ser14 – His289 – Asp286	-	-	✓	✓
Hydrogen Bonds positioning the loops	Val90 – Tyr93	✓	✓	✓	✓
	Asp13 – Ser16	✓	✓	✓	✓
	Asp13 – Ser17	✓	i.	✓	✓
	Gly12 – Val90	✓	✓	✓	✓
	Asp17 – Gly91	✓	✓	✓	✓
Hydrogen bonds in active site interior	Ser14 – Gly92	✓	✓	-	-
	Asp13 – Asn147	i.	i.	i.	-
	Asp13 – Gly145	-	-	-	✓
	Asp13 – Gly146	-	-	-	✓
	Asp13 – Leu15	✓	i.	✓	✓
	Asp13 – Gly92	-	i.	-	-
	Asp13 – Tyr93	-	i.	-	-
	Asp13 – Thr143	i.	i.	i.	✓
	Thr143 – Ser16	-	-	-	✓
	Trp185 – Tyr141	i.	i.	i.	✓
Hydrophobic interactions	Leu15 – Trp185 (hydrophobic)	✓	✓	✓	✓
	Thr143 – Trp185 (hydrophobic)	-	-	-	✓
Hydrogen bonds with conserved water molecules	Asp13/Tyr92/Gly144, Gly145, Gly146, Asn147 – H_2O cluster	i.	-	✓	-
	Asp13/Gly144/Trp185 – H_2O cluster	i. (only 144-H_2O-185)	-	i. (only 144-H_2O-185)	✓

[a] Hydrogen bond if not stated otherwise.

[b] ✓...present; -...absent; i...intermittent.

3.3.5 Active Site Opening in the Isolated Passenger

As opposed to the radius of gyration of the entire passenger domain which is quite similar between systems, with some additional fluctuation in the simulation of the passenger without bound intermediate, the radius of gyration of the active site exhibits some differences between systems (Figure 20, Supplementary Figure 1). The active site was defined with six amino acids which were chosen based on their location. Residues from each of the key structural motifs within the active site were picked – Leu15 from the loop containing the catalytic Ser14, Gly92 from the adjacent loop, Asn147 and Leu150 from helix 6, Val288 from the loop with the other two residues of the catalytic triad, and Phe34, chosen because of its position at the entrance of the

active site. The radius of gyration showed that fluctuations of the active site are quite conspicuous in the full-length enzyme without bound intermediate. When the intermediate is bound to the full-length EstA enzyme, the rigidness of the active site increases. However, the isolated passenger domain exhibits a peculiarity - the radius of gyration of the active site of the passenger without intermediate is smallest when comparing all systems, while it is biggest in the passenger with intermediate. This points to an active site opening mechanism for substrate fitting in the isolated passenger which doesn't seem to happen in the full-length enzyme.

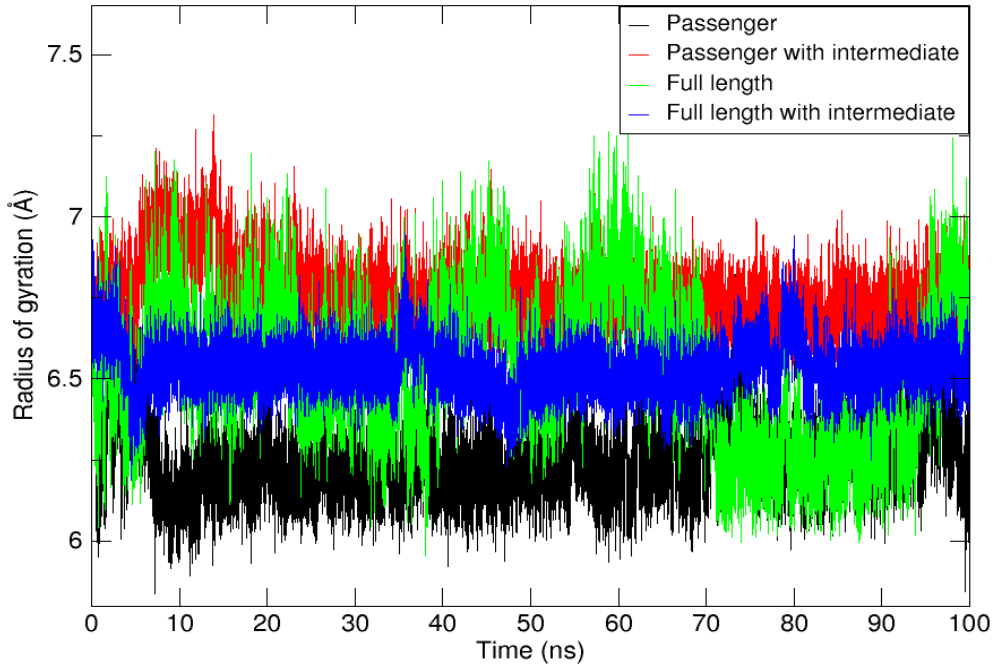


Figure 20. Radius of gyration of EstA active site in four different simulation systems. The active site was defined by amino acids 15, 34, 92, 147, 150 and 288.

This is in accordance with the measured average helical tilt of the active site helix 6 which was largest in the passenger domain simulation without intermediate bound in the active site (these results will be presented in detail in the next chapter). In order to investigate what caused such difference in the radius of gyration further distance measuring was carried out. The fluctuations of all measured active site

distances were largest in the full-length protein without bound intermediate, showing that the active site is definitely most flexible in this case. The measured distance between the centre of mass of the upper part of the central helix (helix 13) and helix 6 was the smallest in the passenger domain system without tetrahedral intermediate. This distance was the same in both passenger with intermediate and full-length with intermediate (Supplementary Figure 2). The same pattern seen in the active site radius of gyration was noticed when comparing distances between helix 6 and the loop with residues Asp286 and His289 of the catalytic triad (Figure 21). These groups are found at opposite sides of the active site, and they are closest in the passenger without bound intermediate, furthest away in the passenger with bound intermediate, while the distance is in between for the full-length systems, again supporting active site opening upon substrate binding for the isolated passenger. That is, the active site of the isolated passenger without bound intermediate is in a closed conformation, and it switches to an open conformation upon substrate binding. The distance between the loop carrying the nucleophile Ser14 and the active site helix 6 shows the same pattern of distance for the two passenger domain systems as the active site radius of gyration. However, in the full-length enzyme with intermediate this distance is short as in the passenger without bound intermediate (Figure 22). To sum up, the difference in the active site radius of gyration among the systems is due to different helix 6 positioning.

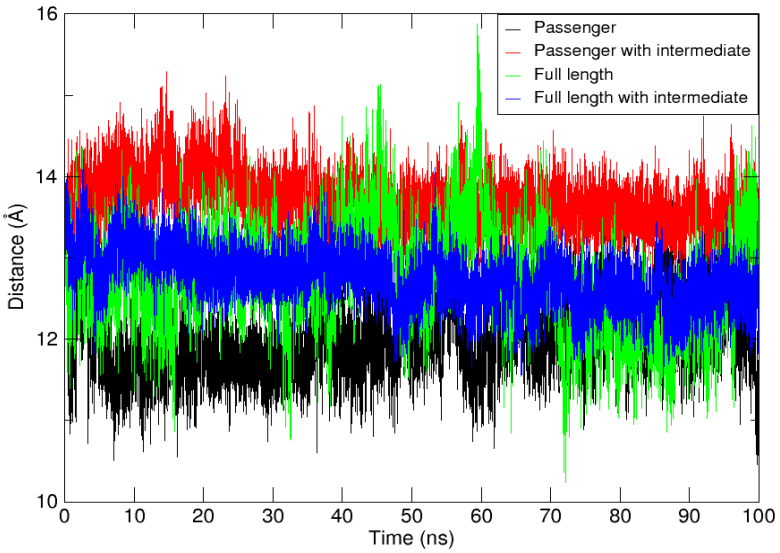


Figure 21. Distance between helix 6 and the loop containing Asp286 and His289 of the catalytic triad.

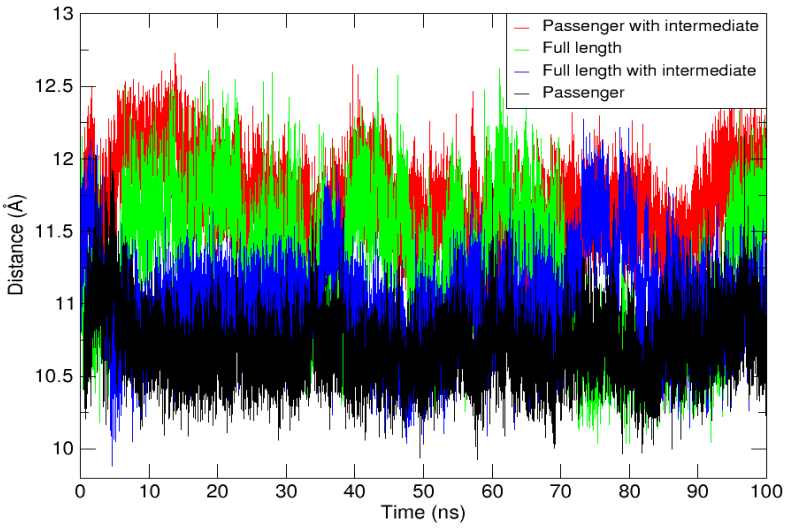


Figure 22. Distance between helix 6 and the loop containing Ser14 of the catalytic triad.

Superimposed structures of different systems show that the position of residue Asn147 of the oxyanion hole is the most likely cause of active site opening in the isolated passenger when a substrate/intermediate is bound. This residue would clash

with the substrate/intermediate if it stayed in the same position as in the passenger without bound intermediate upon substrate binding (Figure 23). In addition to Asn147, helix 6 motion influences Leu150 positioning. This residue is involved in hydrophobic substrate/intermediate stabilisation, and in this system adopts a position which allows flexibility of part of the tetrahedral intermediate that includes the phenol moiety, a motion that is inhibited in the full-length enzyme with intermediate where no active site opening occurs.

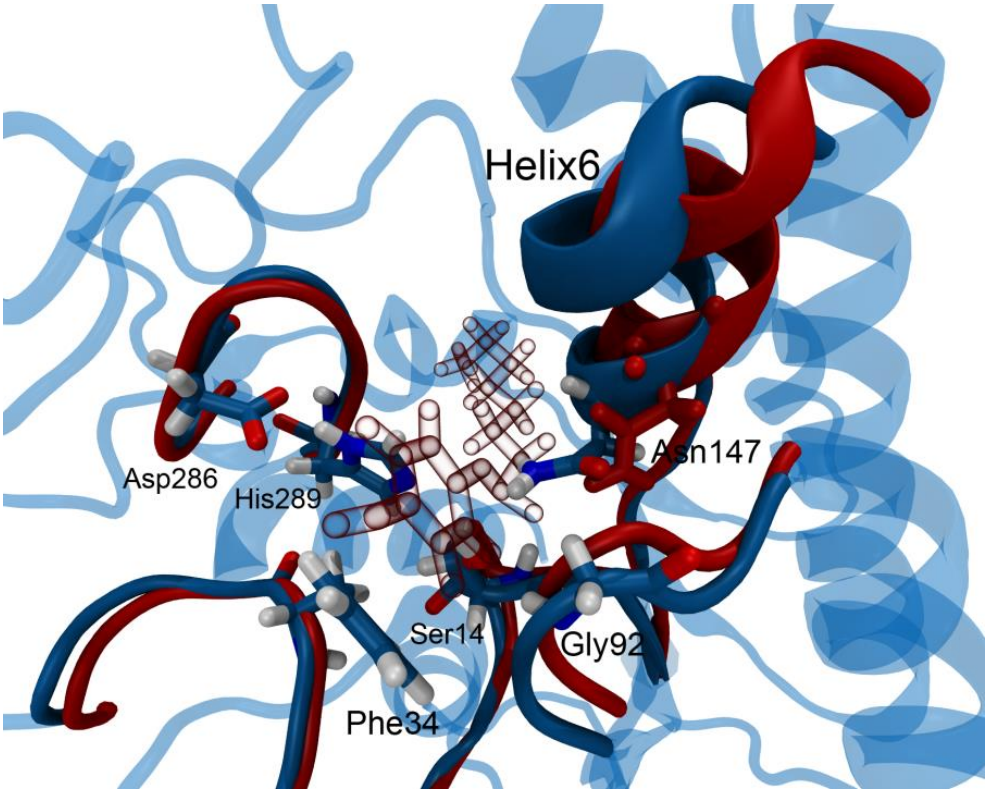


Figure 23. Active site comparison of the passenger domain systems: closed conformation of the passenger domain without bound intermediate (blue) and open conformation of the passenger with intermediate (red). The loops carrying residues that were used for radius of gyration calculation are shown. The catalytic triad (14, 286, 289) and oxyanion hole residues (92, 147) are depicted as they are in the passenger without intermediate. Asn147 in both systems is represented, showing how helix 6 motion relieves the steric effect upon substrate/intermediate binding. A red glassy structure of the tetrahedral intermediate is shown.

3.3.6 Helix 6 Structural Change in Full-length EstA

The active site of full-length EstA is much more flexible when no intermediate is bound, and it perceivably becomes more rigid when the substrate/intermediate binds. The full-length enzyme appears to be unable to adapt its active site the same way it is achieved in the isolated passenger, so the structure of helix 6 itself is forced to change when the intermediate is bound (Figure 24). The majority of the hydrogen bonds which normally define this α helix tend to break, and occasionally new hydrogen bonds are formed between amino acids that are 3 residues apart. In addition, a changed hydrogen bonding profile accounts for the kink introduced in the short unstructured region preceding helix 6, which changes the position of Gly145 and Gly146 in relation to the other active site residues. These changes in the helix are correlated with a change in position of Asn147, which is turned upwards in comparison to the other simulated systems. Its RMSF value is larger than in both passenger domain simulations (0.85 Å as opposed to 0.74 Å in the passenger without intermediate and 0.50 Å in the passenger with intermediate). All of the aforementioned might influence the role of Asn147 in oxyanion stabilisation.

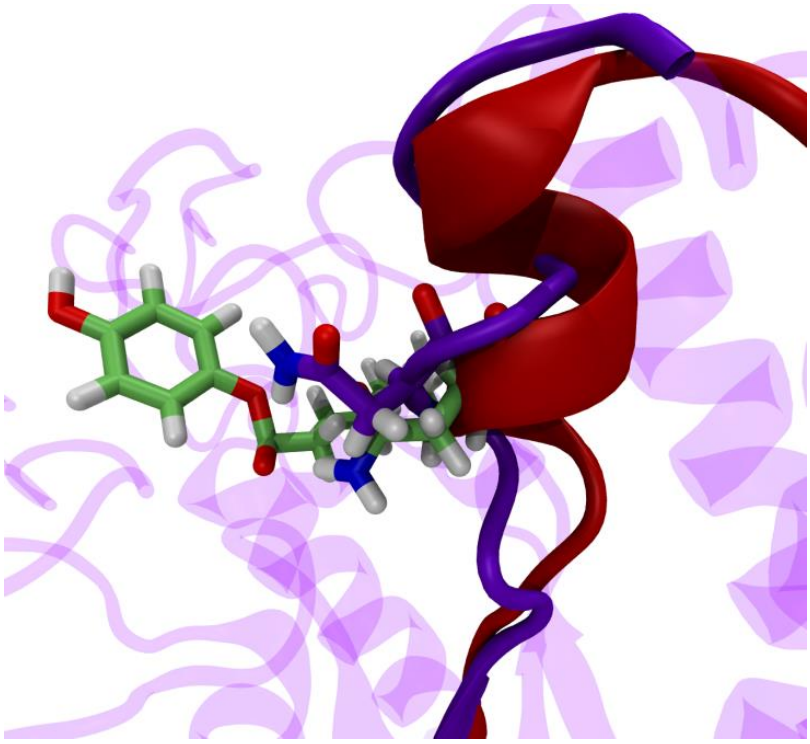


Figure 24. Helix 6 structural change in full-length EstA with bound tetrahedral intermediate. The passenger with bound intermediate is in red and the full-length enzyme with intermediate in violet. The tetrahedral intermediate is shown in light green. In the passenger system Asn147 of the oxyanion hole faces downwards, while in the full-length enzyme it is oriented upwards.

3.4 Domain Interface – Active Site Communication

3.4.1 Helical Tilt

A path which could potentially transfer the information of AT domain presence from the interdomain interface to the active site goes through helix 9, whose bottom part interacts with the transmembrane domain by hydrogen bonding, and helix 7, which is connected by a loop to the active site helix 6. A difference in the helical tilt, especially in helices 6 and 9, was noticed among different simulated systems (Figure 25). The tilt per residue in relation to the average helix axis was calculated for helices 9, 7 and 6, and averaged for the whole simulation. It is important to stress that the average axes are calculated in each simulated system, hence they differ in different simulations. Therefore, although the helical tilt value doesn't give an absolute measure of helix displacement, it does serve as an indicator of the amount of local tilting.

Figure 26 clearly shows that helix 9 in the passenger domain simulation experiences the biggest amount of axis tilting. Helix 9 tilting is most prominent in its bottom part in the simulations of the passenger without transmembrane domain, i.e. when this part of the helix doesn't engage into interdomain interactions. Helix 9 tilting is the smallest in the full-length enzyme simulation. On the other hand, it is nearly equal in both simulations with tetrahedral intermediate, although the bottom part of the helix tilts more strongly in the passenger with intermediate (residues 218-223), and the upper part in the full-length enzyme with bound intermediate (residues 202 to 208). As for helix 7, the overall tilt is the largest in the case of full-length enzyme with tetrahedral intermediate, with the upper part of the helix, i.e. residues 157-163, especially standing out (Figure 27). The helix 7 tilt of the other simulated systems is nearly equal when averaged, with the isolated passenger showing the largest tilting in the bottom part of the helix, and the remaining systems in the central part. Helix 6 tilting was largest in the passenger domain simulation, followed by the full-length enzyme with tetrahedral intermediate. The tilt for the latter, however, may be due to a less stabilised helical structure in this system. Tilting was lowest in the simulation of the isolated passenger with bound intermediate (Figure 28).

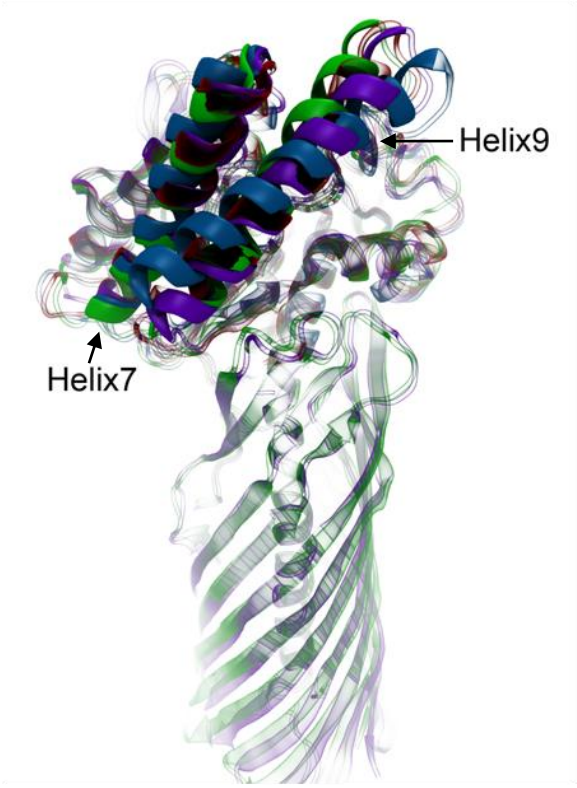


Figure 25. Superposition of structures showing axis tilt of helices 7 and 9. The passenger domain structure was taken at 92 ns and the remaining systems at 100 ns. The passenger domain system is shown in blue, the passenger with intermediate in red, the full-length enzyme in green and the full-length enzyme with intermediate in violet.

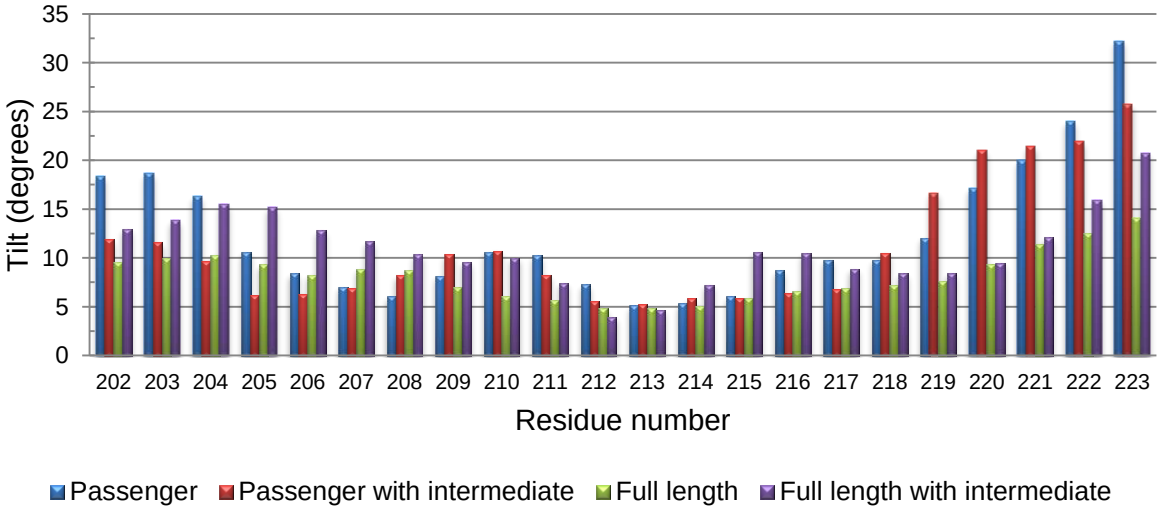


Figure 26. Average tilt per residue of helix 9 (the first and the last residue are not included because of internal properties of the g_helixorient algorithm used).

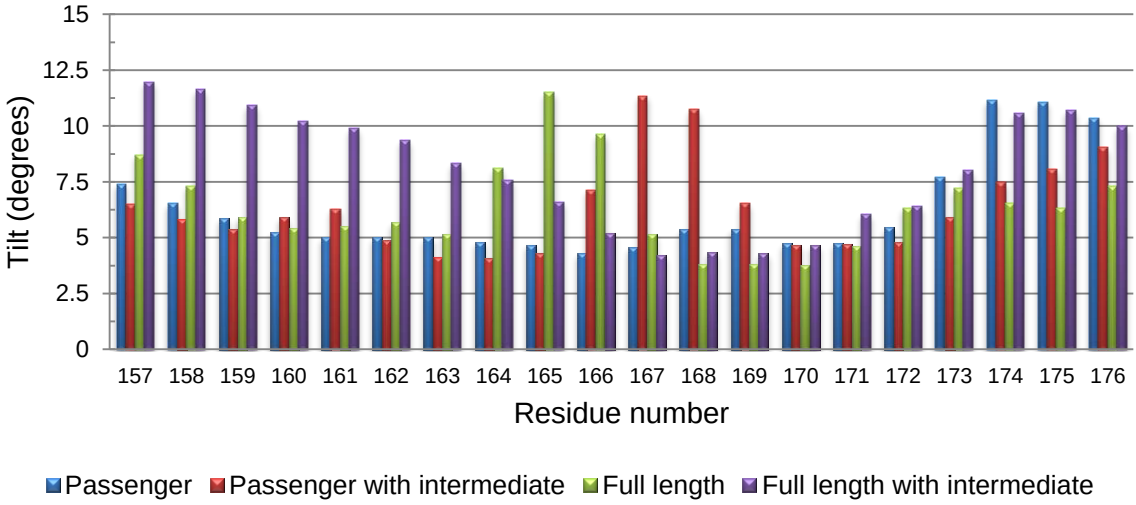


Figure 27. Average tilt per residue of helix 7.

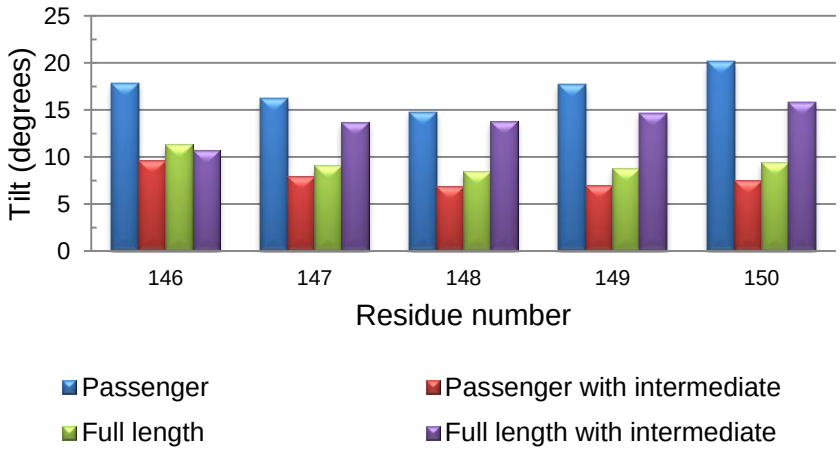


Figure 28. Average tilt per residue of helix 6.

3.4.2 Interhelical Interactions

Distances between helices 9, 7 and 6 were monitored during the simulations. Although there wasn't a big difference in the distance of the centres of mass of helices 9 and 7, the distance measured in the full-length enzyme with no intermediate was smaller than in the other systems, which can be correlated with the small axis tilt of helix 9 (Supplementary Figure 3). The passenger with intermediate bound, on the other hand, showed the most stable distance, i.e. with least abrupt changes. On the amino acid level, this is reflected in the hydrogen bonding of Gln161 from helix 7 with Glu215 and

Gln219 from helix 9. Other hydrogen bonding interactions between helices 7 and 9 are formed periodically, are not conserved throughout the simulation and show no obvious simulation system dependence. On the other hand, the hydrogen bonds of Gln161 with Glu215 and Gln219 are strong and conserved in the passenger domain with intermediate (Figure 29, Supplementary Figure 4). This reflects on the RMSF, which is lowest in this system for all three involved residues. In the other systems these bonds show large fluctuations, at times above 8 Å, and are formed only occasionally, especially in the full-length enzyme with intermediate.

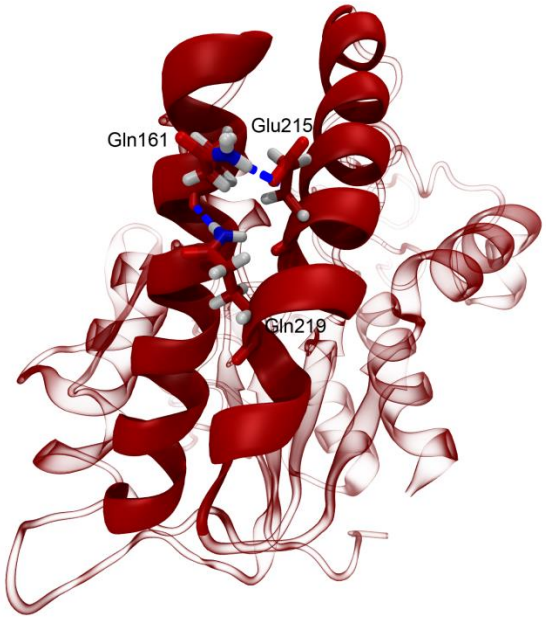


Figure 29. Hydrogen bonding of Gln161 from helix 7 with Gln215 and Gln219 from helix 9. The interaction is conserved throughout the simulation of passenger domain with intermediate.

The distance between the centres of mass of helix 6 and the upper half of helix 7 was biggest in the passenger domain and smallest in the passenger domain with intermediate, which is consistent with the results reported in the previous chapters regarding helix 6. When helix 6 is bent towards the active site in the passenger without bound intermediate, which results in a closed active site conformation, it is also furthest away from the adjacent helix 7. A crucial interaction between these two helices which may be involved in the regulation of the active site openness is the hydrogen bond formed between Asp148 and Arg166 or Arg153 (Figure 30). Asp148 from helix 6 forms

a conserved hydrogen bond with Arg166 from helix 7 in the passenger with intermediate. In this case Arg 166 is always facing upwards, towards Asp148. In the other simulations this arginine is not involved in a conserved bond and has higher side chain fluctuation (RMSF is 0.6 Å in passenger with intermediate, 1.4 Å in passenger without intermediate simulation, 1.8 Å in full-length EstA without intermediate and 2.2 Å in full-length with intermediate). In the simulation of full-length enzyme with intermediate Asp148 is involved in a conserved hydrogen bond with another arginine, Arg153 located on the loop connecting helices 6 and 7. The position of this residue is such that this hydrogen bond cannot hold helix 6 in a position allowing an open active site conformation and adjustment to the bound substrate, as is the case for Asp148-Arg166. Moreover, the RMSF of Arg153 is highest in the full-length enzyme with intermediate (2.25 Å). In addition to the described hydrogen bonds with arginines, Asp148 hydrogen bonds with Thr95 located at the beginning of helix 4. This interaction is conserved in all systems, except for the last part of the simulation of full-length enzyme with intermediate, where the hydrogen bond breaks. This agrees with the RMSF of Thr95, which is smallest in the passenger with intermediate, and largest in the full-length enzyme with intermediate.

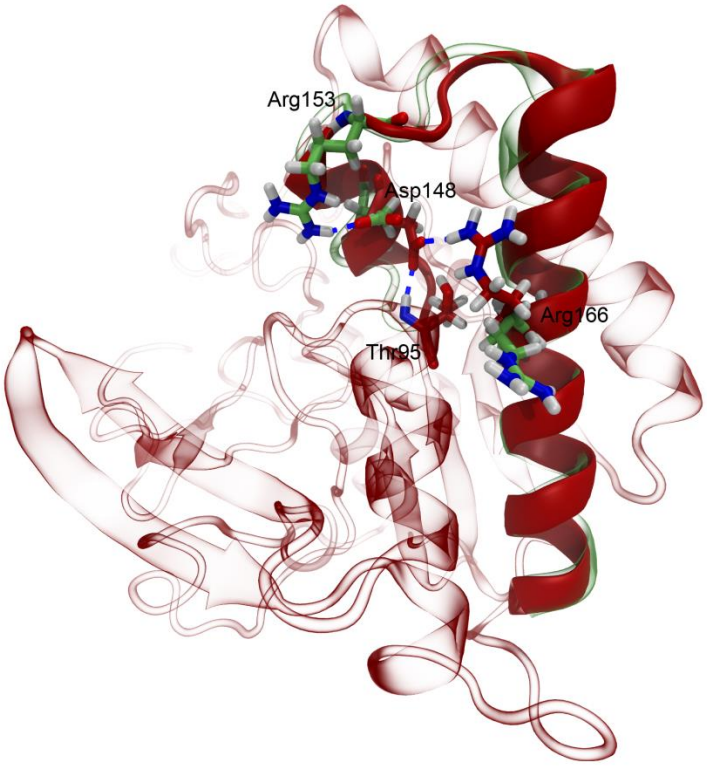


Figure 30. Hydrogen bonding between helices 6 and 7. Residues and structures from the simulation of passenger domain with intermediate are shown in red. In this case Arg166 forms a conserved hydrogen bond with Asp148, which also bonds with Thr95. Residues and structures from the simulation of full-length enzyme with intermediate are shown in green. Asp148 hydrogen bonds with Arg153 during the entire simulation. Arg166 is facing downwards and doesn't participate in said interaction. Asp148 bonds with Thr95 through most of the simulation, except for the end. The superimposed structures were obtained after 100 ns of simulation.

The distance between the centres of mass of helix 6 and helix 9 is similar in all systems and revolving around 13 Å. Communication between these helices is achieved strictly through hydrophobic interactions, where Phe149 is the only participating residue from helix 6 and, consequently, the key residue of the cluster. Amino acids associated with this hydrophobic cluster are found on the loop connecting helix 6 to helix 7, as well as on helix 7, helix 9 and the adjacent unstructured region which comes before it in terms of residue number, but spatially runs parallel with it and engages in hydrogen bonding with said helix. Interestingly, residue 149 is positioned facing downwards through the entire simulation of the full-length enzyme with intermediate, while it's facing upwards during almost the whole simulation of the full-length enzyme without

intermediate in the active site (about 96% of the simulation time). In the passenger domain systems residue 149 switches between the upwards and downwards position (Figures 31 and 32). The position of Phe149 in the full-length enzyme with intermediate is probably correlated with the structural change of helix 6 in this system. The RMSF of Phe149 is in accordance with the above mentioned results – it's higher in the passenger domain systems (with the highest being 1.7 Å in the passenger with intermediate), while it is lowest in the full-length enzyme with intermediate, namely 0.94 Å. Interestingly, in terms of RMSF Phe149 is the exception in a region where fluctuations of other residues are higher in the full-length enzyme simulations than in the passenger domain simulations (namely, region 147 – 169, comprising the upper part of helix 6, the upper half of helix 7 and the loop between them). The Phe149 hydrophobic cluster includes residues which take part in tetrahedral intermediate stabilisation, such as Leu187 and Leu190, and therefore provide a direct path of communication with the active site.

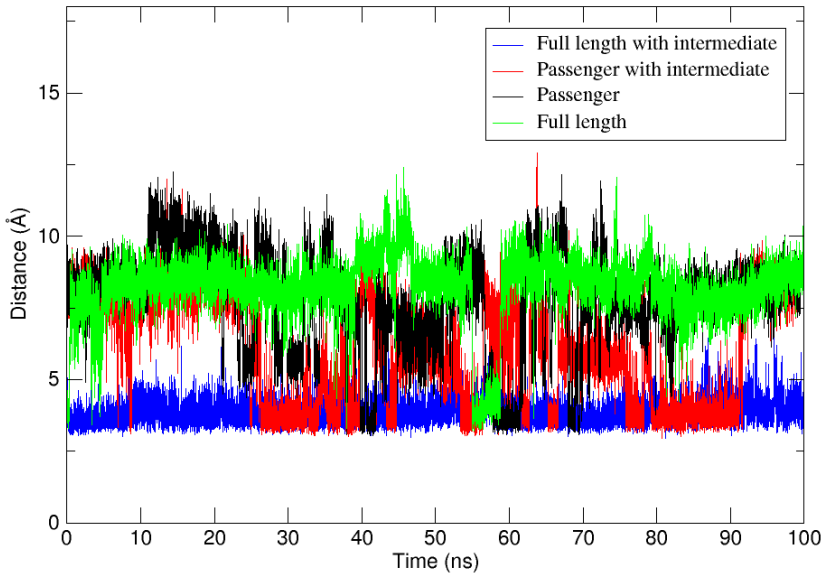


Figure 31. Distance between Phe149 and Leu187, reflecting the position of residue 149 in the various systems. In the full-length with intermediate it's always turned downwards and close to Leu187, in the full-length enzyme it is mostly facing upwards, while in the passenger domain simulations it switches periodically between positions.

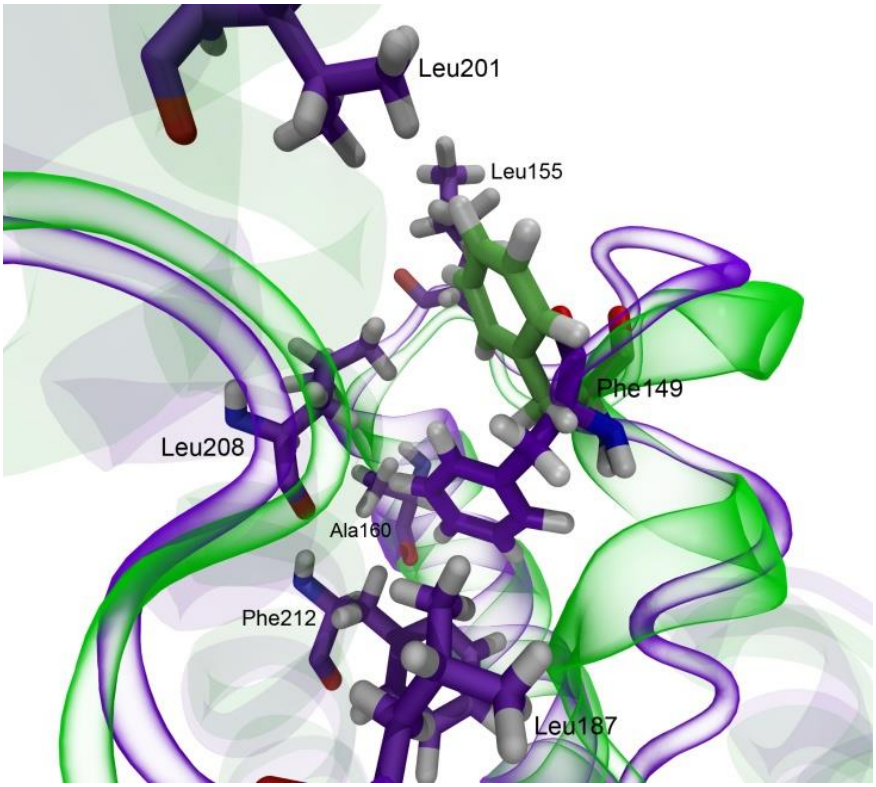


Figure 32. Residues involved in the hydrophobic cluster connecting helices 6 and 9. Residues and structures of the full-length protein without intermediate are shown in green. Violet is used to depict residues and structures of the full-length protein with intermediate. The upwards facing orientation of Phe149 in one case, and downwards facing in the other, is evident, as is the structural change of helix 6 in the full-length enzyme with intermediate. The positions of the other residues of the cluster are from the structure of full-length protein with intermediate.

3.4.3 The Trp185 – Helix 10 Path

Except for the above described path including helices 9, 7 and 6, another possible connection between the interdomain interface and active site was found. It involves helix 10 which, except for interacting with the AT domain, also engages into hydrophobic contacts with Trp185 of the active site through residues Leu230 and Ile232 (Figure 33). Interestingly, the whole region extending from the end of helix 9 to the end of helix 10, namely from residue 220 to 243, is characterized by a higher RMSF in the passenger domain simulations compared to the full-length systems. However, Leu230 and Ile 232 are exceptions, having lowest RMSF in the full-length enzyme with intermediate (0.64 and 0.7 Å, respectively), but much higher, or even highest RMSF in the full-length EstA without intermediate in the active site (0.99 and 1.1 Å). The RMSF

pattern is the same for Trp185 (0.54 and 1.05 Å for full-length EstA with and without bound intermediate, respectively), showing that probably the rigidity or flexibility of Trp185 is transmitted to residues 230 and 232 through hydrophobic interactions, or vice versa. Other residues which are part of this hydrophobic patch are Pro188, which participates in substrate/intermediate stabilisation in the active site, and Leu298 from the central membrane spanning helix.

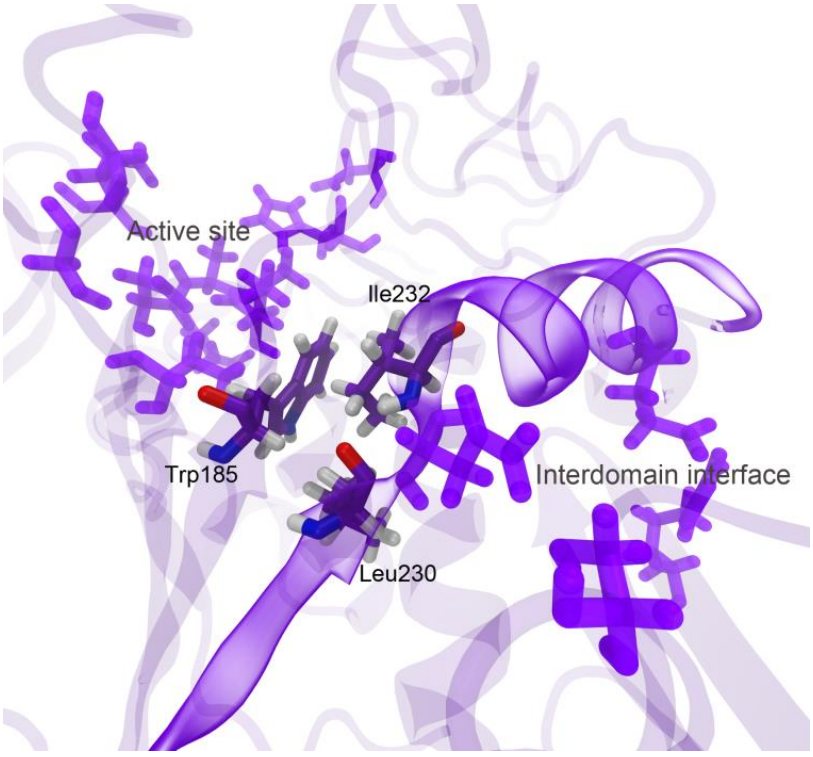


Figure 33. Hydrophobic interactions between Trp185 in the active site and Leu230 and Ile232 at the bottom of helix 10. Interactions are shown for the full-length enzyme with tetrahedral intermediate after 100 ns of simulation. Active site residues and residues involved in interdomain hydrogen bonding on helix 10 and the AT domain are completely coloured purple.

4 Discussion

In 2010 both the solved crystal structure of EstA and kinetic experiments pointed towards the existence and possible importance of interactions between the passenger and the AT domain (Leščić Ašler 2010, van den Berg 2010). These interactions were characterised in this thesis. Hydrogen bonds were divided into three interaction clusters. Hydrophobic interactions couldn't be strictly grouped, but the regions involved were the same as for hydrogen bonds. Interestingly, most of the interacting regions were not the usual structural parts of a GDSL hydrolase or an AT β barrel, rather it seems like specific protein structures were developed or conserved for interdomain interactions. Sheet D, for once, is an interacting region of the β barrel, but it's not found in most ATs (with the exception of AIDA-I, which also remains tightly bound to its AT domain, although after proteolytic processing) (Gawarzewski et al. 2014). Another interaction region of the β barrel consists of its strands 5 and 6 which are atypically elongated in EstA, so they can extend up to the passenger domain and participate in interdomain contacts. The third peculiar region is block IIIa, a sequence of residues with similar physicochemical characteristics that is found only in GDSL hydrolases that are ATs (Leščić Ašler et al. 2010). In EstA this region comprises the second half of helix 10 and the loop which follows, including the small helix 11.

The first interaction cluster found in EstA includes residues of sheet D and the bottom part of helix 9 and the adjacent loop and β strand. This cluster exhibited some differences between the bound and unbound full-length enzyme in terms of hydrogen bonds (namely, Asn567 hydrogen bonds mostly with Val227, but also with Thr217, Leu220 and Ala225 in the enzyme without bound intermediate, while in the complex with the tetrahedral intermediate Asn567 – Thr217 is the dominant interaction, which is shortly replaced by Asn567 – Ala225). The observed difference may be somehow related to helix 9 tilting, which was lowest for the unbound full-length enzyme. However, the measured tilt was highest for the isolated passenger and similar for both systems with intermediate in the active site, showing that substrate/intermediate presence also influences helix 9 tilt. As expected, the interacting region which consists of helix 9 and the adjacent unstructured loop had considerably lower RMSF in the full-length systems.

The other two clusters didn't show such differences in individual interactions. In the second cluster connecting helices 10 and 13 to loops of sheet D and a prolonged β strand, the stability of hydrogen bonds and hydrophobic interactions was diminished in the full-length enzyme with intermediate. This may be due to hydrophobic interactions of helices 10 and 13 with the active site residue Trp185. Although this hydrophobic cluster doesn't considerably affect the rigidity of helix 10 (only a small RMSF difference), it may make it less prone to position itself as to optimise contacts with the transmembrane domain. The RMSF of the entire region extending from the end of helix 9 to the end of helix 10 is larger in the isolated passenger domain systems, where there are no interdomain interactions to stabilise these regions. However, residues 230 and 232 involved in hydrophobic interactions with Trp185 exhibit highest RMSF in the unbound full-length enzyme, showing that the high active site flexibility and fluctuation in this system is clearly transferred to these residues located on helix 10. In the full-length enzyme with bound intermediate the region around Trp185 is kept rigid and well connected to the catalytic core region. In order to achieve this, it looks like two conditions need to be met: a substrate or intermediate must bind in the active site making the part with the catalytic core (the nucleophile and the oxyanion hole) more ordered and rigid, and interdomain interactions must keep helix 10 structured. This allows helix 10 to hydrophobically interact with the lower part of the active site, namely with Trp185, inducing rigidness from this side. Since helix 10 is very fluctuant and occasionally tends to partly lose its secondary structure in the passenger domain systems, its residues cannot participate in a compact hydrophobic cluster, so there is no active site stabilisation through helix 10 and Trp185 in these two systems. As expected, the isolated passenger domains were found to adopt more varied conformations than the same domains in the full-length systems. The conformations sampled by the passenger domain simulation without bound intermediate were grouped into five clusters. With the same cut-off, the conformations of the passenger with intermediate were grouped into four, showing that the intermediate induces rigidity in the structure. The structures adopted by the passenger domain in the unbound full-length enzyme were grouped into three clusters, while all the conformations adopted in the full-length with intermediate were grouped into one cluster, showing that in this case as well the

presence of the intermediate rigidifies the structure. The results of the cluster analysis prove that the isolated passenger domain is more structurally flexible than the passenger domain in the full-length enzyme, where it is anchored to the AT domain through the central membrane-spanning α helix and connected to it by means of the described interdomain interactions. Although the active site of the unbound full-length enzyme is very fluctuant, the connection with the AT domain makes up for this flexibility, resulting in the conformations adopted by this system's passenger domain being grouped in less clusters than the isolated passengers.

The active site of EstA hadn't been characterised in detail before, except for the identification of the catalytic residues. The catalytic triad was correctly predicted and later confirmed (van den Berg 2010, Wilhelm et al. 1999). MD simulations performed as part of this work showed that the bond Ser14 - His289 is not conserved in the active site without bound tetrahedral intermediate. In this case Ser14 tends to hydrogen bond with residues of the oxyanion hole, i.e. Gly92 and Asn147. These residues were also previously predicted based on alignment with other members of the SGNH superfamily (Leščić Ašler et al. 2010, Mølgaard et al. 2000). In a study based on the SGNH hydrolase TAP from *E. coli* Yu-Chih and colleagues confirmed a notion postulated by Mølgaard and co-workers in 2000, that is that the enzyme TAP, and probably the whole protein family, stabilises the oxyanion with three amino acids instead of the two residues typical for serine hydrolases (2003). They speculated this feature may compensate for the lack of a hydrogen bond of the catalytic histidine with a backbone carbonyl group. Both of these finding were confirmed for EstA in this work – a three residue oxyanion hole (in addition to the above mentioned residues, the backbone nitrogen of the catalytic Ser14 participates in oxyanion stabilisation), and no hydrogen bond of the imidazole ring of His289 with the protein backbone. In addition, in the same paper Yu-Chih and co-workers described an active site hydrogen bond network in TAP, which they supposed was conserved in the SGNH hydrolase superfamily. The authors grouped the active site hydrogen bonds into three categories: inter-block, intra-block and water-mediated hydrogen bonds. This was also confirmed for EstA. The bonds of the inter-block group are Gly12 – Val90 and Asp17 – Gly91, and they keep the two loops carrying the nucleophile and the oxyanion hole Gly92 positioned in the active site. Intra-

block hydrogen bonds keeping these loops structured are Asp13 – Ser16, Asp13 – Ser17 and Val90 – Tyr93. These bonds are found at the boundary of the active site. Water-mediated hydrogen bonds were conserved only when the tetrahedral intermediate was bound in the active site, which then had to be additionally stabilised and catalytically important residues had to be precisely positioned. Since the hydrogen bond network in TAP was characterised in the crystal structure, which cannot show the conservation and occurrence of the bonds, it is possible that in this enzyme as well water-mediated bonds in the active site are transient, and become relevant only when a substrate/intermediate is bound. In addition to the three hydrogen bond groups defined earlier, in this thesis another group of hydrogen bonds is proposed for EstA, which would comprise internal active site hydrogen bonding. Hydrogen bonds in this group could be included in the inter-block and intra-block groups, but they were found to exhibit some differences depending on AT domain presence and tetrahedral intermediate binding. Therefore, they are not crucial in shaping the general active site geometry, but are important for protein function. An overview of active site hydrogen bonds is given in Table 4.

In the active site of full-length EstA without tetrahedral intermediate bound, all internal active site noncovalent bonds occur only periodically. Even the hydrogen bond between the side chain of Asp13 and the backbone of Leu15 is intermittent, while it's conserved in the other systems. In addition, this hydrogen bond was previously reported as important in stabilising the nucleophile loop in SGNH hydrolases, with the aspartate being conserved in the SGNH family (Mølgaard et al. 2000). In the unbound isolated passenger the active site was found to be less prone to fluctuate, as proven by the active site radius of gyration, the distances between measured interacting regions and the RMSF of active site residues. In this case most internal active site bonds were intermittent, but still more stable and conserved than in the full-length enzyme without bound intermediate (these hydrogen bonds include Asp13 – Asn147, Asp13 – Asn143 and Trp185 – Tyr141 in the remote region of the active site). On the other hand, relevant differences in the internal active site hydrogen bond network were detected between the two systems with bound intermediate, i.e. the isolated passenger and the full-length enzyme. In case of the passenger, in addition to the bonds found in the

systems without tetrahedral intermediate, two or three conserved water molecules were found to mediate hydrogen bonding between the nucleophile loop, the loop with the oxyanion hole residue Gly92, and active site helix 6 (which carries the oxyanion hole residue Asn147). These interactions had the important role of forming a compact catalytic core around the substrate/intermediate. Interestingly, in the full-length enzyme with bound intermediate there were no water-mediated hydrogen bonds in this region. That is mainly because two new hydrogen bonds (Asp13 – Gly145 and Asp13 – Gly146) were formed directly between the nucleophile loop and helix 6. These bonds were conserved in this system due to an introduced kink in the bottom part of helix 6 which brought the interacting regions closer together, bypassing the need for water-mediated hydrogen bonding. Asn147 was also switched in position in the full-length system with bound intermediate, stabilising the oxyanion from above rather than below, which was the case in the isolated passenger (Figure 24). Oxyanion stabilisation was weaker because Asn147 was more fluctuant per se, and consequently so were the interactions it engaged in (Figure 17). Gly92, the other oxyanion hole residue, was also more fluctuant. Moreover, in this system it seemed to be much more important than in the other systems to stabilise the remote region of the active site and its connection to the main part which includes the catalytic core. The hydrogen bond Trp185 – Tyr141 in the remote part of the active site is conserved only in this system, and so are the hydrogen bonds between the nucleophile loop and Thr143, a residue located centrally, between the catalytic and remote part of the active site. This threonin is, in turn, connected by hydrophobic interactions to Trp185, together with residues of the nucleophile loop. Moreover, conserved water-mediated hydrogen bonds are also employed in connecting these active site regions, i.e. the nucleophile loop (Asp13) and helix 6 (Gly144) to Trp185. This shows that in the full-length enzyme with intermediate active site stabilisation by internal hydrogen bonding was moved further from the catalytic core, and the compactness of the region leading to and around Trp185 was important. Although the details and reasons for this are not clear, this is most probably related to the stabilisation of this remote active site region by hydrophobic interactions with residues from helix 10, which in turn engages into interdomain interactions. It should be noted that Trp185 might be equivalent to Trp23 in TAP, which was found to

play an important role in substrate binding, accommodation of varied substrates and substrate turnover (Li-Chiun et al. 2009). An important role for this tryptophane was afterwards also confirmed in another GDSL hydrolase from *P. aeruginosa*, TesA (Petrović 2013). In these enzymes, however, this residue is directly connected to the nucleophile loop by hydrogen bonding, which is not the case in EstA. Moreover, in the isolated passenger domain of EstA, which is a GDSL hydrolase like TAP and TesA (which are not ATs), the connection between the catalytic core and the Trp185 region is poor. Its measured catalytic activity, on the other hand, was higher than the full-length enzyme (F. Kovačić, *personal comm.*). Therefore experiments and/or simulations with site directed mutagenesis should be done to study the role of this residue in EstA.

In addition to hydrogen bonds, hydrophobic interactions have a crucial role in substrate/intermediate stabilisation. It was previously known from the crystal structure that the catalytic residues of EstA lie at the entrance of a hydrophobic cleft (van den Berg 2010). Indeed, residues Leu15, Leu150, Val288, Leu187, Pro188, Leu190 and Phe34 were found to form a hydrophobic substrate/intermediate binding crevice. Interestingly, the intermediate was less fluctuant in the full-length enzyme, while the phenol moiety and the end of the hydrocarbon chain were allowed some motion in the isolated passenger. Flexion of part of the intermediate including the phenol moiety was inhibited in the full-length enzyme due to different positioning of helix 6 residues (such as Asn147 and Leu150) because of a structural change in the helix, while chain end fluctuations were restrained by certain residues of the hydrophobic cleft such as Leu187 and Leu190. These residues are found on an unstructured region parallel to helix 9, and are part of a bigger hydrophobic cluster including helix 9, helix 7, helix 6 and the so-called switch loop between them. Interestingly, a structurally analogous region to the unstructured region adjacent to helix 9 of EstA was found to be important for substrate accommodation in TAP, along with the switch loop (Yu-Chih et al. 2005). In addition, the hydrophobic cluster this region participates in was shown to influence substrate preference in TAP (Ya-Lin et al. 2012). The simulations carried out on EstA cannot unveil these effects, but it was found that these same regions (i.e. the region adjacent to helix 9 and the switch loop) were important in EstA as well. The hydrophobic cluster including the unstructured region adjacent to helix 9 is most stable and compact in the

full-length enzyme with intermediate. The central residue of this cluster, Phe149 from helix 6, is differently positioned depending on tetrahedral intermediate presence in the full-length enzyme, but not in the isolated passenger. Phe149 may be fixed in the upwards orientation in the full-length enzyme without intermediate due to too little tilting of helix 9 in this system, which doesn't leave enough space for big phenylalanine side chain fluctuations. In the full-length enzyme with intermediate bound this residue is always facing downwards, forming a compact cluster that could inhibit helix 6 motion and the consequent active site opening which was noticed in the isolated passenger. In addition, because of the position and steric effect of Phe149, this residue might be the one inducing the structural change seen in helix 6. However, the other way around is also a possibility, i.e. that the structural change of helix 6 allows Phe149 to adopt this position. This may be more plausible, and would mean that the required direct hydrogen bonding of helix 6 residues (Gly145 and Gly146) and Asp13 in the active site introduced a kink in the bottom part of helix 6, which allowed a change in position of Phe149 when the substrate/intermediate bound. Direct hydrogen bonding between helix 6 and the nucleophile loop may be required because a limited number of water molecules may be allowed in the active site due to its hydrophobicity. Since in this system water-mediated connection to the remote part of the active site seems to be important, as mentioned previously, catalytic core stabilisation may have been obtained by helix 6 structure disruption.

Residues from helix 7 and the switch loop between helices 6 and 7 also participate in the above mentioned hydrophobic cluster. However, this region is even more interesting because it carries the amino acids which are probably responsible for the active site opening, a feature observed only in the isolated passenger. Distance measuring between key structures in the active site, active site radius of gyration and structure superposition all point to the conclusion that the isolated passenger has a closed active site conformation when no intermediate is bound. Such conformation couldn't fit the intermediate because of steric clashes. The passenger with bound intermediate therefore exhibits an open active site conformation, with the biggest active site radius of gyration amongst all systems. This points to an induced fit mechanism in the isolated passenger (Figure 23). Induced fit was previously suggested as a possible

mechanism employed by GDSL hydrolases to accommodate different substrates in their flexible active sites (Akoh 2004). However, this study suggests that induced fit in the EstA isolated passenger is a general prerequisite for substrate binding. Moreover, it doesn't depend on active site flexibility, but rather on specific hydrogen bonding that occurs between particular structures. The crucial hydrogen bond is Asp148 (helix 6) – Arg166 (helix 7), which is conserved only in the simulation of passenger with bound tetrahedral intermediate and presumably serves to keep helix 6 in an open conformation. The distance between helix 7 and helix 6 is smallest in this system, implying that helix 7 doesn't undergo motion towards and from the active site to regulate active site opening, but the motion is restricted to helix 6. This is probably achieved by anchoring helix 7 to helix 9 by hydrogen bonding, which is conserved throughout the simulation only in this system. This interaction may be impossible to form in the full-length system, potentially because of a difference in bending or some other property of the region of helix 9 involved in the hydrogen bond, which is due to the interdomain interactions this helix engages in. Anyhow, this option should be further explored. The full-length enzyme's active site, on the other hand, doesn't show any difference in the level of openness in the bound and unbound state, only considerable fluctuations when unbound. It is not excluded that the active site of the full-length enzyme does undergo some kind of induced fit to accommodate different substrates, but no active site opening analogous to the one found in the isolated passenger was detected. A possible inhibitor of such opening may be the hydrophobic cluster which includes Phe149 from helix 6 and seems to be especially compact in the full-length enzyme with intermediate. The active site radius of gyration for the full-length enzyme is in between the closed and open conformations of the isolated passenger. While in the isolated passenger helix 6 motion optimally positions residues such as Asn147 and Leu150 in relation to the substrate/intermediate, this is probably achieved by the already mentioned helix 6 structural change in the full-length enzyme with bound intermediate. It's possible that the inability of the full-length enzyme to exhibit induced fitting by active site opening is another reason for the observed structural change of helix 6 as an alternative way to position relevant active site residues. Asp148 forms a conserved hydrogen bond with Arg153 from the switch loop in this system. This interaction may be made possible by

the structure of helix 6, and its presence inhibits any kind of Asp148 – Arg166 hydrogen bonding, even transient. This is probably why the biggest amount of local tilting per residue was found in the upper part of helix 7 in this system, and also a high RMSF for the whole switch loop and great part of helix 7. Interestingly, Arg153 is the most fluctuant residue of the loop, particularly in the full-length enzyme with intermediate, i.e. when it engages in hydrogen bonding. Not only due to its localization on a flexible loop, but also due to its position in relation to helix 6, Arg153 and its hydrogen bond with Asp148 cannot mediate helix 6 motion.

To sum up, the precise path by which the information on interdomain contacts is transferred to the active site remains elusive. However, it seems most likely that the very fluctuant active site of the full-length enzyme is stabilised by binding the intermediate in the apposite crevice and by increasing the rigidity of the remote part around Trp185 (which is connected to the catalytic core) through hydrophobic interactions with helix 10. This is enabled by the interdomain interactions helix 10 participates in. In addition, interdomain contacts of helix 9 could be signaled to the active site through the hydrophobic cluster that includes Phe149, and on the other hand through helix 7 and helix 6, by inhibiting their hydrogen bonding and active site opening in the full-length system. The question that sparked this research was what causes the difference in activity experimentally observed in the isolated passenger and the full-length enzyme (F.Kovačić, *personal comm.*). MD simulations gave a few possible explanations. The full-length enzyme may be less catalytically active because of its inability to undergo induced fit by helix 6 motion. This, among other things, causes worse oxyanion stabilisation due to different Asn147 positioning (in addition, a higher RMSF of Gly92 was observed in this system). Another possible reason is the absence of conserved water molecules in the upper part of the active site, termed the catalytic core, which are present in the isolated passenger with bound tetrahedral intermediate. In addition, substrate binding in an extremely fluctuant active site, like in the unbound full-length enzyme, might be more difficult and slower. On the other hand, the tetrahedral intermediate is less fluctuant in the full-length enzyme than in the isolated passenger. It is not clear whether this is a pro for the catalytic activity of the isolated passenger or the full-length enzyme. A better stabilised intermediate might be an

advantage for the full-length enzyme; however, a higher energy structure in the isolated passenger may also be beneficial, since this structure has yet to undergo another enzymatic reaction. The stabilisation of the remote part of the active site including Trp185 by closer contacts and conserved water molecules in the full-length enzyme seems possibly beneficial for enzyme catalysis in this system. Consequently, no precise conclusions could be drawn regarding the reasons for an activity increase in the isolated passenger, although there were evident differences in the active sites.

With that said, it must be kept in mind that the measured activity differences between the isolated passenger and full-length enzyme have to be regarded with caution. The isolated passenger is a soluble protein, and was isolated as such. However, the EstA passenger domain is thought to fold in a vectorial fashion, directed by the membrane bound β barrel, as assumed for other ATs as well (van den Berg 2010, Junker et al. 2009, Junker et al. 2006). Therefore, there is a possibility that the isolated passenger domain and the passenger when bound to the β barrel are not equally folded, and that the observed activity difference is a consequence thereof. In addition, all proteins were purified from inclusion bodies. Since this purification protocol requires refolding, even the full-length enzyme structure may differ from the one present in the cell OM. The crystal structure on which all MD simulations were based was derived from the full-length enzyme isolated from the *P.aeruginosa* OM. It would be best if kinetics measurements were carried out on EstA purified this way, which would allow working with the biologically relevant conformation (fold). Since purification of the passenger domain after β barrel directed folding is a problem, it would be very interesting to carry out MD simulations of the EstA-EstP hybrid enzyme once the EstP structure is solved. This way the difference in activity observed in this hybrid enzyme, which is most probably based on interdomain interactions as well, could be understood. In addition, in order to better understand or prove the findings presented in this work, it would be necessary to carry out MD simulations and kinetics measurements of EstA with mutations in some putative key residues. Mutations in residues with an important role in interdomain interactions should be done, such as Asn226, Asn567, Asp300, Tyr301, Asn231 and Glu238, and their combinations, in order to block an entire interaction cluster. Possible targets for mutations to perturb hydrophobic interdomain

contacts are Leu235, Phe246, Tyr301, Leu450, Ala509 and Leu564. As for the active site, mutation of Asp13, Leu15, Asn147 and Trp185 might give some useful insight. In addition, Phe149, Arg153, Arg166, as well as Leu230 and Ile232 should be mutated to investigate these residues as possible routes through which information of interdomain contacts could be transferred in the protein. Mutations should be done in both isolated passenger and full-length enzyme, when possible.

Another fact that should be taken into account is that the physiological substrate(s) of EstA is unknown. Since GDSL ATs usually remain covalently bound to the cell surface and have hydrolytic activity, it was assumed they might hydrolyse membrane lipids (Wilhelm et al. 2011). From the observed effects of *estA* gene deletion on *P. aeruginosa* cells no conclusion on the specific function and substrate of EstA could be drawn (Wilhelm et al. 2007). It is therefore possible that full-length membrane embedded EstA is most active towards its biological substrate, and the isolated passenger shows higher activity for substrates used in kinetics assays. The membrane embedded enzyme may have an active site less adaptable to various substrates, but may still exhibit high activity towards its cognate substrate.

The evolution of GDSL ATs is also worth mentioning. While the β barrels are highly conserved among all AT proteins, passenger domains can be very diverse. This suggests that evolution of different ATs has happened by multiple independent recombination events between β barrels and passenger domains (Henderson et al. 1998). According to this theory, such a recombination event between a soluble GDSL hydrolase and an AT β barrel would have resulted in the formation of GDSL ATs. If not enough evolutionary time has passed or selective pressure happened in the meantime, the GDSL domain might still be more active as a soluble protein. It may have not completely adapted to being bound to the AT domain, especially if its activity in the full-length enzyme, even if diminished, was still enough for cell functioning. What is more, it may even be that the bound passenger domain exerts exactly the amount of activity that is needed in *P. aeruginosa*, and is an adaptation to the needs of the cell, while more effective catalysis would result in unnecessarily too fast product accumulation.

Another appealing theory is that EstA may undergo a conformational change which would include breaking of interdomain contacts and thereby enhancing catalytic

activity. That is, in certain conditions when EstA activity would be needed, the passenger domain would be able to adopt a conformation which would be more similar to the isolated passenger, which exhibits higher activity. EstA was found to be active in biofilm formation, which occurs in human lung tissue in cystic fibrosis patients, and is the main reason why these bacteria are able to survive and thrive in the human organism (Wilhelm et al. 2007). Conditions which would induce this activity may be varied, from higher temperature or smaller oxygen concentrations, to host derived signals, nutritional cues, sub-inhibitory antibiotics concentrations and others (Wei and Ma 2013). Except for the aforementioned mutations of residues involved in interdomain interactions in EstA, it would be interesting to carry out a coarse-grained MD simulation which may allow visualizing such conformational change of the passenger domain bound to the AT barrel. If EstA has to remain membrane bound due to its function, maybe because its substrate is membrane bound or even a membrane lipid, this might be a way of enzyme activation which bypasses the usual proteolytic cleaving in AT proteins.

5 Conclusions

The passenger and AT domain of EstA engage into interdomain non-covalent interactions, namely hydrogen bonds and hydrophobic interactions. They are grouped into three interaction clusters. Interacting residues are located on specific structures not usually found in GDSL hydrolases (block IIIa) or AT β barrels (sheet D and the prolonged strands 5 and 6).

Important active site residues include the catalytic triad, a three-residue oxyanion hole (backbone of Ser14, Gly92 and Asn147), hydrophobic residues involved in tetrahedral intermediate stabilisation, and residues participating in the active site hydrogen bond network. Three types of hydrogen bonds form in the active site as part of the hydrogen bond network: hydrogen bonds positioning the loops (i.e. the nucleophile loop and the loop carrying Gly92), hydrogen bonds in the active site interior and water-mediated hydrogen bonds (which increase the compactness and stability of the active site only when the tetrahedral intermediate is bound). Water-mediated hydrogen bonds stabilise the upper part of the active site which contains the catalytic core in the isolated passenger, while in the full-length enzyme they stabilise the remote part of the active site, connecting the catalytic core to the region around Trp185.

In the full-length enzyme with intermediate a structural change of helix 6 occurs which allows it to hydrogen bond to the nucleophile loop and position Asn147 and other active site residues. In this case the active site is also stabilised by hydrophobic interactions of Trp185 with residues of helix 10, which in turn engages into interdomain interactions. The isolated passenger domain, on the other hand, exhibits an induced fit mechanism mediated by helix 6 motion, with a closed active conformation when no intermediate is present, and an open conformation when the intermediate is bound. The induced fit in the isolated passenger is mediated by hydrogen bonds that form between helices 9, 7 and 6 when the tetrahedral intermediate is bound.

6 References

Akoh CC, Lee G-C, Liaw Y-C, Huang T-H, Shaw J-F (2004) GDSL family of serine esterases/lipases. *Progress in Lipid Research*, **43**, 534–552.

Arpigny JL, Jaeger KE (1999) Bacterial lipolytic enzymes: classification and properties. *The Biochemical Journal*, **343 Pt 1**, 177–183.

Barnard TJ, Gumbart J, Peterson JH *et al.* (2012) Molecular basis for the activation of a catalytic asparagine residue in a self-cleaving bacterial autotransporter. *Journal of Molecular Biology*, **415**, 128–142.

Barnard TJ, Dautin N, Lukacik P, Bernstein HD, Buchanan SK (2007) Autotransporter structure reveals intra-barrel cleavage followed by conformational changes. *Nature Structural & Molecular Biology*, **14**, 1214–1220.

Berendsen HJC, Spoel DVD, Drunen RV (1995) Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm*, **91**, 43–56.

van den Berg B (2010) Crystal structure of a full-length autotransporter. *Journal of molecular biology*, **396**, 627–633.

Bernstein HD (2007) Are bacterial “autotransporters” really transporters? *Trends in Microbiology*, **15**, 441–447.

Brandon LD, Goldberg MB (2001) Periplasmic Transit and Disulfide Bond Formation of the Autotransported Shigella Protein IcsA. *Journal of Bacteriology*, **183**, 951–958.

Brundage L, Hendrick JP, Schiebel E, Driessen AJ, Wickner W (1990) The purified E. coli integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell*, **62**, 649–657.

Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, **126**, 014101.

Case DA, Darden TA, Cheatham TE *et al*. (2012) AMBER 12. University of California, San Francisco.

Celik N, Webb CT, Leyton DL *et al.* (2012) A Bioinformatic Strategy for the Detection, Classification and Analysis of Bacterial Autotransporters. *PLoS ONE*, **7**, e43245.

Charbonneau M-E, Janvore J, Mourez M (2009) Autoprocessing of the Escherichia coli AIDA-I autotransporter: a new mechanism involving acidic residues in the junction region. *The Journal of Biological Chemistry*, **284**, 17340–17351.

Cho H, Cronan JE (1993) Escherichia coli thioesterase I, molecular cloning and sequencing of the structural gene and identification as a periplasmic enzyme. *The Journal of Biological Chemistry*, **268**, 9238–9245.

Cotter SE, Surana NK, St Geme JW (2005) Trimeric autotransporters: a distinct subfamily of autotransporter proteins. *Trends in Microbiology*, **13**, 199–205.

Driessen AJ, Wickner W (1990) Solubilization and functional reconstitution of the protein-translocation enzymes of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 3107–3111.

Egile C, d' Hauteville H, Parsot C, Sansonetti PJ (1997) SopA, the outer membrane protease responsible for polar localization of IcsA in Shigella flexneri. *Molecular Microbiology*, **23**, 1063–1073.

Fink DL, Cope LD, Hansen EJ, Geme JW (2001) The Hemophilus influenzae Hap autotransporter is a chymotrypsin clan serine protease and undergoes autoproteolysis via an intermolecular mechanism. *The Journal of Biological Chemistry*, **276**, 39492–39500.

Fojan P, Jonson PH, Petersen MT, Petersen SB (2000) What distinguishes an esterase from a lipase: a novel structural approach. *Biochimie*, **82**, 1033–1041.

Gawarzewski I, DiMaio F, Winterer E *et al.* (2014) Crystal structure of the transport unit of the autotransporter adhesin involved in diffuse adherence from Escherichia coli. *Journal of Structural Biology*, **187**, 20–29.

Grijpstra J, Arenas J, Rutten L, Tommassen J (2013) Autotransporter secretion: varying on a theme. *Research in Microbiology*, **164**, 562–582.

Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D (2004) Type V protein secretion pathway: the autotransporter story. *Microbiology and molecular biology reviews: MMBR*, **68**, 692–744.

Henderson IR, Nataro JP (2001) Virulence functions of autotransporter proteins. *Infection and Immunity*, **69**, 1231–1243.

Henderson IR, Cappello R, Nataro JP (2000) Autotransporter proteins, evolution and redefining protein secretion. *Trends in Microbiology*, **8**, 529–532.

Henderson IR, Navarro-Garcia F, Nataro JP (1998) The great escape: structure and function of the autotransporter proteins. *Trends in Microbiology*, **6**, 370–378.

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, **18**, 1463–1472.

Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, **31**, 1695–1697.

Hornak V, Abel R, Okur A *et al.* (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *Journal of Molecular Graphics*, **14**, 33–38, 27–28.

Ieva R, Bernstein HD (2009) Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19120–19125.

Izard JW, Kendall DA (1994) Signal peptides: exquisitely designed transport promoters. *Molecular Microbiology*, **13**, 765–773.

Jacob-Dubuisson F, Locht C, Antoine R (2001) Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Molecular Microbiology*, **40**, 306–313.

Jain S, Goldberg MB (2007) Requirement for YaeT in the outer membrane assembly of autotransporter proteins. *Journal of Bacteriology*, **189**, 5393–5398.

Jämbeck JPM, Lyubartsev AP (2012a) Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids. *The Journal of Physical Chemistry B*, **116**, 3164–3179.

Jämbeck JPM, Lyubartsev AP (2012b) An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. *Journal of Chemical Theory and Computation*, **8**, 2938–2948.

Johnson TA, Qiu J, Plaut AG, Holyoak T (2009) Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of haemophilus influenzae immunoglobulin A1 protease. *Journal of Molecular Biology*, **389**, 559–574.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, **79**, 926–935.

Jose J, Jähnig F, Meyer TF (1995) Common structural features of IgA1 protease-like outer membrane protein autotransporters. *Molecular Microbiology*, **18**, 378–380.

Junker M, Besingi RN, Clark PL (2009) Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Molecular Microbiology*, **71**, 1323–1332.

Junker M, Schuster CC, McDonnell AV *et al.* (2006) Pertactin β-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 4918–4923.

Kajava AV, Steven AC (2006) The turn of the screw: variations of the abundant beta-solenoid motif in passenger domains of Type V secretory proteins. *Journal of Structural Biology*, **155**, 306–315.

Kovačić F, Granzin J, Wilhelm S *et al.* (2013) Structural and Functional Characterisation of TesA - A Novel Lysophospholipase A from Pseudomonas aeruginosa. *PLoS ONE*, **8**, e69125.

Leach AR (2001) *Molecular Modeling: Principles and Applications*. Pearson Education Limited, Essex, Englad.

Leščić Ašler I, Ivić N, Kovačić F *et al.* (2010) Probing enzyme promiscuity of SGNH hydrolases. *Chembiochem: A European Journal of Chemical Biology*, **11**, 2158–2167.

Li-Chiun L, Yi-Li C, Hong-Hwa C, Ya-Lin L, Jei-Fu S (2009) Functional role of a non-active site residue Trp(23) on the enzyme activity of Escherichia coli thioesterase I/protease I/lysophospholipase L(1). *Biochimica Et Biophysica Acta*, **1794**, 1467–1473.

Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. *Bioinformatics (Oxford, England)*, **22**, 623–625.

Loveless BJ, Saier MH (1997) A novel family of channel-forming, autotransporting, bacterial virulence factors. *Molecular Membrane Biology*, **14**, 113–123.

Meng G, Spahich N, Kenjale R, Waksman G, St Geme JW (2011) Crystal structure of the Haemophilus influenzae Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. *The EMBO journal*, **30**, 3864–3874.

Mølgaard A, Kauppinen S, Larsen S (2000) Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases. *Structure (London, England: 1993)*, **8**, 373–383.

Nikaido H, Hancock REW (1986) Outer membrane permeability of *Pseudomonas aeruginosa*. In: *The Bacteria - A Treatise on Structure and Function, Volume X: The Biology of Pseudomonas*. (ed. Sokatch JR), pp. 145–193. Academic Press Inc., Orlando, Florida.

Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, **52**, 255–268.

Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, **50**, 1055–1076.

Ollis DL, Cheah E, Cygler M *et al.* (1992) The alpha/beta hydrolase fold. *Protein Engineering*, **5**, 197–211.

Oomen CJ, van Ulsen P, van Gelder P *et al.* (2004) Structure of the translocator domain of a bacterial autotransporter. *The EMBO journal*, **23**, 1257–1266.

Otto BR, Sijbrandi R, Luirink J *et al.* (2005) Crystal Structure of Hemoglobin Protease, a Heme Binding Autotransporter Protein from Pathogenic Escherichia coli. *Journal of Biological Chemistry*, **280**, 17339–17345.

Pallen MJ, Chaudhuri RR, Henderson IR (2003) Genomic analysis of secretion systems. *Current Opinion in Microbiology*, **6**, 519–527.

Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, **52**, 7182–7190.

Petrović S (2013) Istraživanje supstratne specifičnosti acil-CoA tioesteraze I iz bakterije *Pseudomonas aeruginosa* korištenjem simulacije molekularne dinamike (Graduate Thesis). University of Zagreb, Faculty of Science.

Pohlner J, Halter R, Beyreuther K, Meyer TF (1987) Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA protease. *Nature*, **325**, 458–462.

Pukatzki S, Ma AT, Sturtevant D *et al.* (2006) Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model

system. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1528–1533.

Purdy GE, Fisher CR, Payne SM (2007) IcsA Surface Presentation in Shigella flexneri Requires the Periplasmic Chaperones DegP, Skp, and SurA. *Journal of Bacteriology*, **189**, 5566–5573.

Ruiz-Perez F, Henderson IR, Leyton DL *et al.* (2009) Roles of Periplasmic Chaperone Proteins in the Biogenesis of Serine Protease Autotransporters of Enterobacteriaceae. *Journal of Bacteriology*, **191**, 6571–6583.

Salacha R, Kovacić F, Brochier-Armanet C *et al.* (2010) The Pseudomonas aeruginosa patatin-like protein PlpD is the archetype of a novel Type V secretion system. *Environmental Microbiology*, **12**, 1498–1512.

Saurí A, Oreshkova N, Soprova Z *et al.* (2011) Autotransporter β-domains have a specific function in protein secretion beyond outer-membrane targeting. *Journal of Molecular Biology*, **412**, 553–567.

Sauri A, Soprova Z, Wickström D *et al.* (2009) The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. *Microbiology (Reading, England)*, **155**, 3982–3991.

Scott-Tucker A, Henderson IR (2009) Type V secretion. In: *Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis*. (ed. Wooldridge K), pp. 139–158. Caister Academic Press, Norfolk, UK.

Selkrig J, Mosbahi K, Webb CT *et al.* (2012) Discovery of an archetypal protein transport system in bacterial outer membranes. *Nature Structural & Molecular Biology*, **19**, 506–510, S1.

Shere KD, Sallustio S, Manessis A, D'Aversa TG, Goldberg MB (1997) Disruption of IcsP, the major Shigella protease that cleaves IcsA, accelerates actin-based motility. *Molecular Microbiology*, **25**, 451–462.

Sherlock O, Dobrindt U, Jensen JB, Munk Vejborg R, Klemm P (2006) Glycosylation of the Self-Recognizing Escherichia coli Ag43 Autotransporter Protein. *Journal of Bacteriology*, **188**, 1798–1807.

Upton C, Buckley JT (1995) A new family of lipolytic enzymes? *Trends in Biochemical Sciences*, **20**, 178–179.

Veiga E, de Lorenzo V, Fernández LA (1999) Probing secretion and translocation of a beta-autotransporter using a reporter single-chain Fv as a cognate passenger domain. *Molecular Microbiology*, **33**, 1232–1243.

Volokhina EB, Grijpstra J, Stork M *et al.* (2011) Role of the Periplasmic Chaperones Skp, SurA, and DegQ in Outer Membrane Protein Biogenesis in Neisseria meningitidis. *Journal of Bacteriology*, **193**, 1612–1621.

Voulhoux R, Bos MP, Geurtsen J, Mols M, Tommassen J (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science (New York, N.Y.)*, **299**, 262–265.

Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *Journal of Molecular Graphics*, **8**, 52–56, 29.

Wang H, Dowds BC (1993) Phase variation in Xenorhabdus luminescens: cloning and sequencing of the lipase gene and analysis of its expression in primary and secondary phases of the bacterium. *Journal of Bacteriology*, **175**, 1665–1673.

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *Journal of Computational Chemistry*, **25**, 1157–1174.

Wei Q, Ma LZ (2013) Biofilm matrix and its regulation in Pseudomonas aeruginosa. *International Journal of Molecular Sciences*, **14**, 20983–21005.

Wells TJ, Tree JJ, Ulett GC, Schembri MA (2007) Autotransporter proteins: novel targets at the bacterial cell surface. *FEMS microbiology letters*, **274**, 163–172.

Wilhelm S, Gdynia A, Tielen P, Rosenau F, Jaeger K-E (2007) The autotransporter esterase EstA of Pseudomonas aeruginosa is required for rhamnolipid production, cell motility, and biofilm formation. *Journal of bacteriology*, **189**, 6695–6703.

Wilhelm S, Rosenau F, Kolmar H, Jaeger K-E (2011) Autotransporters with GDSL passenger domains: molecular physiology and biotechnological applications. *Chembiochem: A European Journal of Chemical Biology*, **12**, 1476–1485.

Wilhelm S, Tommassen J, Jaeger KE (1999) A novel lipolytic enzyme located in the outer membrane of Pseudomonas aeruginosa. *Journal of Bacteriology*, **181**, 6977–6986.

Ya-Lin L, Li-Chiun L (2012) Multifunctional enzyme thioesterase I/protease I/lysophospholipase L1 of Escherichia coli shows exquisite structure for its substrate preferences. *Biocatalysis and Agricultural Biotechnology*, **1**, 95–104.

Yao-Te H, Yen-Chywan L, Vitaliy Ya G, Tai-Huang H (2001) Backbone dynamics of Escherichia coli thioesterase/protease I: evidence of a flexible active-site environment for a serine protease1. *Journal of Molecular Biology*, **307**, 1075–1090.

Yen YT, Kostakioti M, Henderson IR, Stathopoulos C (2008) Common themes and variations in serine protease autotransporters. *Trends in Microbiology*, **16**, 370–379.

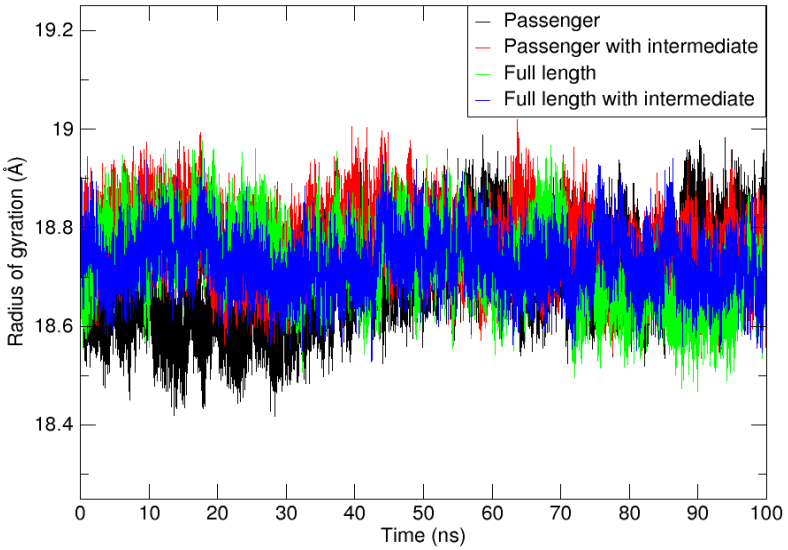
Yu-Chih L, Su-Chang L, Jei-Fu S, Yen-Chywan L (2005) Substrate Specificities of Escherichia coli Thioesterase I/Protease I/Lysophospholipase L1 Are Governed by Its Switch Loop Movement. *Biochemistry*, **44**, 1971–1979.

Yu-Chih L, Su-Chang L (2003) Crystal structure of Escherichia coli thioesterase I/protease I/lysophospholipase L1: consensus sequence blocks constitute the catalytic center of SGNH-hydrolases through a conserved hydrogen bond network. *Journal of molecular biology*, **330**, 539–551.

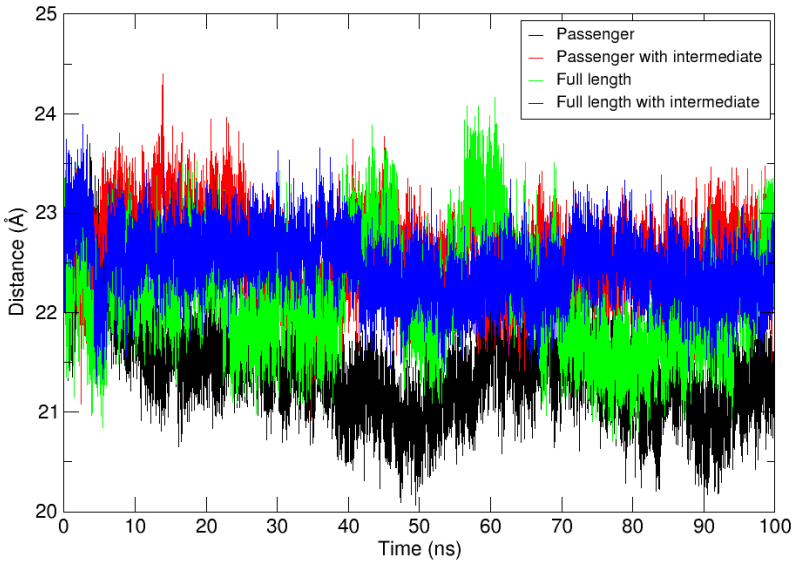
Zhai Y, Zhang K, Huo Y *et al.* (2011) Autotransporter passenger domain secretion requires a hydrophobic cavity at the extracellular entrance of the β-domain pore. *The Biochemical Journal*, **435**, 577–587.

people.su.se/~jjm/Stockholm_Lipids

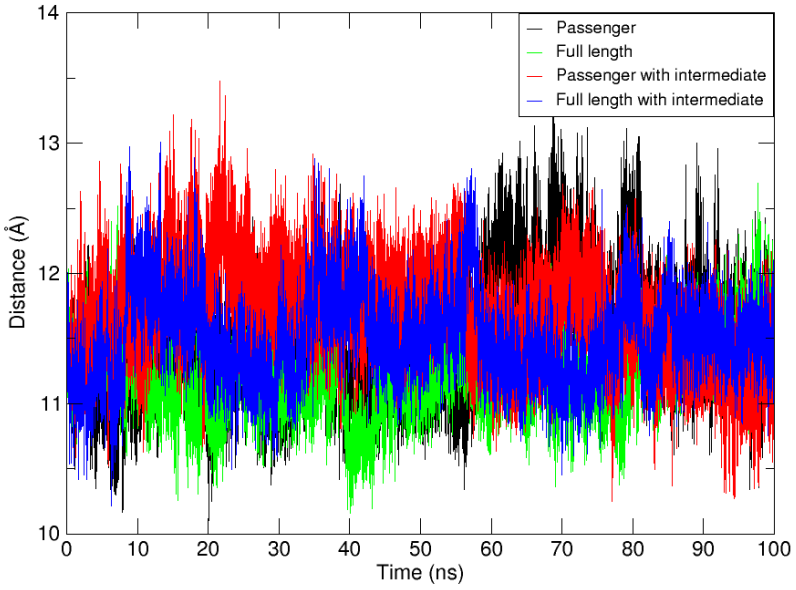
7 Supplementary Material



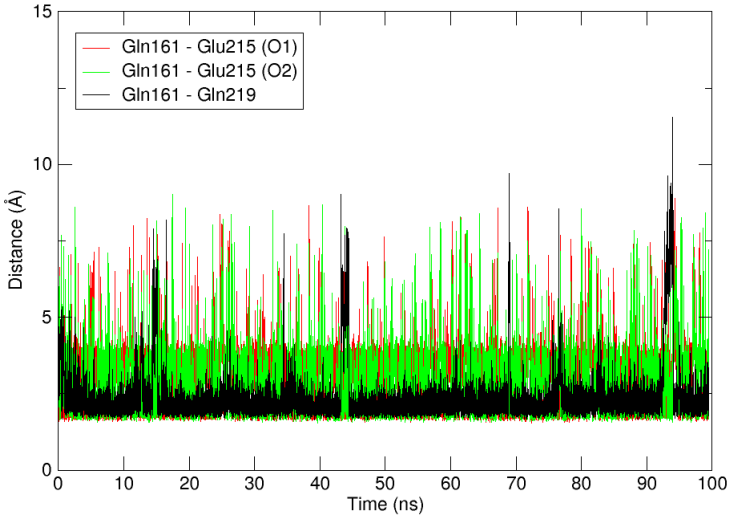
Supplementary Figure 1. Radius of gyration of the passenger domain in all four simulation systems.



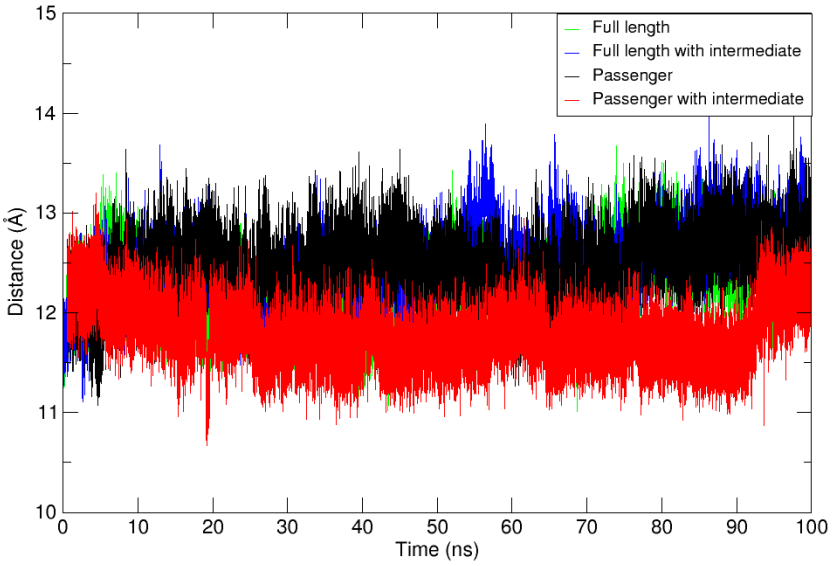
Supplementary Figure 2. Distance between the centres of mass of helix 13 (i.e. the upper part of the central membrane spanning helix) and helix 6 of the active site.



Supplementary Figure 3. Distance between the centres of mass of helices 7 and 9 for all four simulation systems.



Supplementary Figure 4. Hydrogen bonding interaction between helices 7 and 9 in the passenger domain with intermediate, not present as continuous bond in other systems. Hydrogen bonding takes place between the amide $–NH_2$ of Gln161 and the carboxyl group O of Glu215 (the bond with both oxygens is shown), as well as between the backbone O of Gln161 and the amide $–NH_2$ of Gln219.



Supplementary Figure 5. Distance between the centres of mass of helices 6 and 7 for all four simulation systems.

Biography

I was born on November 15th 1990 in Rijeka. There I enrolled in the mathematical and scientific programme in the secondary school "Prva sušačka hrvatska gimnazija". I graduated in 2009 and was chosen as valedictorian student. As a secondary school student I became a recipient of the City of Rijeka scholarship, which I continued to be as a University student. During my secondary education I took part in various competitions on county and national level. Except for national competitions in Croatian and Latin language, I took part in the Biology national competition in 2009. In the same year I won the 1st place in the county competition (Primorsko-goranska) in both Biology and Chemistry.

In June 2009 I attended the 6th ISABS Conference on Human Genome Project Based Applications in Forensic Science, Anthropology and Individualized Medicine (Split), as an award of the Minister of Science and Education for students who took part in the Biology national competition. In the same year I entered the undergraduate programme in Molecular Biology at the Faculty of Science, University of Zagreb, being among the 10% best students according to entrance exams results at the Department of Biology.

During my undergraduate study I rediscovered myself as not only a molecular biologist. I developed an interest for zoology and conservation and became first a member, and in 2011 also the leader of the Dragonfly Section of the Biology Students' Association BIUS. Except for regular field trips, in 2010 I had a poster presentation with the Dragonfly Section at the 1st European Odonatology Conference in Portugal. In 2012 I took part in a volunteering project, working in the Sea Turtle Rescue Center on the small island of Linosa (Italy). As for Department activities, I participated in the organization of the event Biology Night 2010-2012 and was part of the group of students which were awarded the Rector's award for it in the academic year 2011/2012. In addition, in 2010/2011 I was a student laboratory assistant in the course Zoology.

On the other hand, I discovered a new passion for Biochemistry. In 2010 I attended the 10th Congress of the Croatian Society of Biochemistry and Molecular Biology: The Secret Life of Biomolecules (Opatija). I graduated in 2012 with the seminar paper

Catalytic RNAs – molecular fossils of the RNA world, under the supervision of Assoc. Prof. Dr. Ita Gruić.

In 2012 I enrolled in the Molecular Biology graduate programme at the Faculty of Science, University of Zagreb, as the 1st ranked student on the list of applicants. On my second year I was awarded the Dean's award for best student of the Molecular Biology graduate programme. I focused my interest on courses such as Molecular Biophysics, Catalytic Mechanisms in Biological Systems, Bioinformatics, Protein Crystallography, RNA Biology, Structural Computational Biophysics, etc. In 2014 I spent three months working at the Institute for Molecular Enzyme Technology (IMET) at the Forschungszentrum in Jülich, Germany, as an Erasmus Programme scholarship recipient. In the same year I attended the 12th Greta Pifat-Mrzljak International School of Biophysics (Primošten). In 2015 I worked on creating and writing the exercises for the course Structural Computational Biophysics, and these later on became part of an official script of the Department of Chemistry, which I am a co-author of, along with my graduate thesis supervisor, Asst. Prof. Dr. Branimir Bertoša. I also worked as a student assistant in the course Structural Computational Biophysics, where I had the chance to work pretty independently, assisting the students with computational exercises.