Heinrich Heine Universität Düsseldorf

Engineering of an organic solvent-tolerant esterase by modifying surface charges following simulation-based predictions

Master Thesis

Submitted in Partial Fulfillment of the
Requirements for the Academic Degree of
Master of Science in Biochemistry

at the
Faculty of Mathematics and Natural Sciences
of Heinrich-Heine University Düsseldorf

presented by
Lara Scharbert

Köln, April 20, 2021

Principle Investigator
Prof. Dr. Birgit Strodel

Institute of Biological Information Processing (IBI-7)
at Forschungszentrum Jülich

and

Institute of Theoretical and Computational Chemistry
of Heinrich-Heine University Düsseldorf

Second Reviewer
Prof. Dr. Karl-Erich Jaeger

Institute of Molecular Enzyme Technology
of Heinrich-Heine-University Düsseldorf
at Forschungszentrum Jülich

Supervisor
Anna Jäckering

Institute of Biological Information Processing (IBI-7)
at Forschungszentrum Jülich

and

Institute of Theoretical and Computational Chemistry
of Heinrich-Heine University Düsseldorf

Supervisor
Dr. Alexander Bollinger

Institute of Molecular Enzyme Technology
of Heinrich-Heine-University Düsseldorf
at Forschungszentrum Jülich

Abstract

The ability of enzymes to catalyze reactions in the presence of organic solvents has great industrial potential. However, organic solvent molecules can strip water molecules off the protein surface and enter the hydrophobic core, resulting in enzyme denaturation. No efficient, generally applicable strategy for enzyme-stability engineering in polar organic solvents has been reported to date. In this study, a mutagenesis strategy based on molecular dynamics (MD) simulations was developed, with the aim to strengthen the hydration shell around an enzyme to protect it from attack by organic solvents. This strategy is tested for the two esterases, PT35 and ED30. The former is considerably stable against denaturation and inactivation by organic solvents and is characterized by a high overall negative charge, whereas the structurally related ED30 is unstable in organic solvents. By removing negatively charged residues on the surface of PT35 and inserting negatively charged residues on the surface of ED30, PT35 becomes less stable and ED30 more stable. The selection of which of the surface residues to mutate was guided by various computational analyses, including MD simulations. The best in silico predictions, along with the wild-type enzymes, were biochemically characterized in the wet lab. A more acetonitrile resistant single mutant, K337E, was identified for ED30, while the D439K single mutant of PT35 became more sensitive towards acetonitrile. Interestingly, K337E was also found to be more thermally stable and D439K to be less thermally stable, which supports the assumption of a positive correlation between thermal and organic solvent tolerance. Non-additional mutation effects were observed for the triple-mutants K124D/R211E/K337E of ED30 and E199R/E328K/D439K of PT35. Analysis of the MD trajectories indicates that the formation and disruption of salt bridges contributes to the increased and decreased tolerance of the mutants, respectively. In summary, the MD-guided mutagenesis approach for engineering more organic solvent-tolerant enzymes looks promising and should therefore be continued in future studies.

Contents

List of Figures

List of Tables

Abbrevations

ACN acetonitrile

Amp100 100 μg/ml ampicillin

AI auto induction

APBS adaptive Poisson-Boltzmann solver

BLAST basic local alignment search tool

CD consensus design

CTAC cetyltrimethylammonium chloride

CHAPS 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate

COBALT constraint-based multiple alignment tool

CS conservation score

C-score confidence score

cumRDF cumulative radial distribution function

CV column volumes

DCLM double cubic lattice method

DMSO dimethyl sulfoxide

E. coli *Escherichia coli*

ED30 EstDL30

EDTA ethylenediaminetetraacetic acid

epPCR error-prone polymerase chain reaction

EPS electrostatic potential surface

$\Delta G^{\ddagger}$ free energy barrier

ΔG free energy

$\Delta\Delta G$ relative change in free energy

H-bonds hydrogen bonds

HPC High-performance computing

IMAC immobilized metal ion affinity chromatography

I-TASSER iterative threading assembly refinement

JURECA Jülich research on exascale cluster architectures

KPi potassium phosphate

LB lysogeny broth

LINCS linear constraint solver

LJ Lennard-Jones

maxSASA maximum solvent-accessible surface area

MC Monte Carlo

MD molecular dynamics

nanoDSF nano differential scanning fluorimetry

Ni-NTA nickel nitrilotriacetic acid

OD optical density

PBC periodic boundary conditions

PCR polymerase chain reaction

pI isoelectric point

PME particle mesh Ewald

***p*NPB** 4-nitrophenyl butyrate

PT35 Paes_TB035

QC QuikChange

RDF radial distribution function

rSASA relative solvent-accessible surface area

Rgyr radius of gyration

RMSD root mean square deviation

RMSF root mean square fluctuation

SASA solvent-accessible surface area

SDS sodium dodecyl sulfate

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC secretory

SEM standard error of the mean

SLIC sequence- and ligation-independent cloning

SP signal peptide

TAT twin-arginine translocase

Tris Tris(hydroxymethyl)aminomethane

Triton X-100 2-[4-(2,4,4-trimethylpentan-2-yl)phenoxy]ethanol

vdW van der Waals

VS virtual sites

1 Introduction

1.1 Biotechnological application of lipolytic enzymes

Lipolytic enzymes comprise esterases (EC 3.1.1.1, carboxylesterases) and true lipases (EC 3.1.1.3, triacylglycerol acyl hydrolases). Both esterases and lipases catalyze the hydrolysis of ester-containing molecules such as triacylglycerols (figure 1.1).[1,2]

However, lipases mainly convert long-chain fatty acid substrates and can moreover be distinguished from esterases by their ability to be activated by the lipid-water interface.[3] A strong increase of lipase activity is observed, when the hydrophobic lipid substrate starts to form an emulsion. This *interfacial activation* can be explained by a lid-like α-helical structure, which uncovers the active site of the lipase by contact with the resulting lipid-water interface.[4] In contrast to lipases, esterases usually hydrolyze partially water-soluble short-chain fatty acid substrates and exhibit classical Michaelis-Menten kinetics, since their activity is not enhanced by the formation of a substrate-water emulsion.[2,4]

Lipolytic enzymes share some similar structural characteristics as they belong to the superfamily of α/β hydrolases.[5] These enzymes possess an α/β-hydrolase fold, which comprises one central β-sheet of mostly parallel β-strands connected by α-helices (figure 1.2a).[6,7] The active site, which catalyzes the hydrolysis of ester bonds, as described in figure 1.2b in more detail, is usually formed by the three residues Ser, His and Asp (or Glu), known as the catalytic triad. The Ser residue is usually part of the pentapeptide Gly-Xaa-Ser-Xaa-Gly (GXSXG), often found in the form of GESAG.[2]

With the aim of facilitating the identification and characterization of novel bacterial lipolytic enzymes, Arpigny and Jaeger (1999)[2] defined 8 different families (I-VIII) on the basis of conserved sequence motifs and biochemical properties of 53 proteins. The largest familiy (family I), which comprises true lipases, was further divided into 6 sub-families. Over the years, several novel families have been proposed, some of which were included in a recent update that expanded the system to 19 families and 8 true lipase sub-families.[8] Members of the same family do not only share conserved motifs and positions like those of the catalytic triad, but also exhibit similar substrate preferences and molecular weights. Esterases belonging to family VII, for instance, are rather large bacterial esterases of about 55 kDa and share sequence homology with eukaryotic acetylcholine

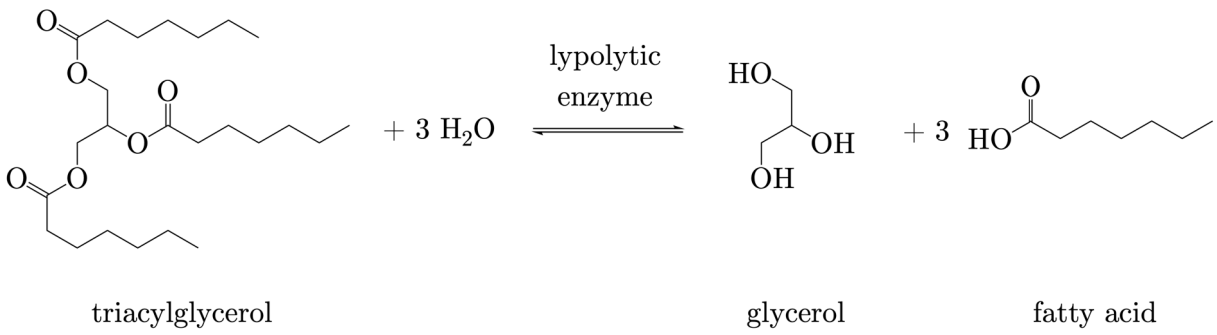

Figure 1.1: Hydrolysis or synthesis of triacylglycerol substrate by lipolytic enzymes. The hydrolysis of triacylglycerol catalyzed by lipolytic enzymes such as esterases and lipases yields glycerol and fatty acids.

esterases and liver carboxylesterases.[2,9] They primarily hydrolyze *p*-nitrophenyl esters and triglycerides with short fatty acid chain length and also biotechnologically relevant substrates.[8]

The application of enzymes in industrial processes has many environmental and economic benefits.[10] Enzymes display sustainable, highly selective catalysts, which avoid the use of scarce precious metals and the need for functional-group activation and (de)protection steps required in conventional organic synthesis and, moreover, function under mild conditions in water. Thus, enzyme catalyzed syntheses are often shorter, more selective, produce purer products and less waste.[11,12]

Due to their broad substrate spectra and their tolerance towards organic solvents, high temperatures and wide pH ranges, as well as their ability to catalyze highly regio- and stereoselective conversions, lipolytic enzymes are one of the most attractive enzymes in industrial applications.[13,14] Lipolytic enzymes are used in a wide variety of products and processes including food, paper, detergent, textile, biodiesel, waste water treatment, degradation of synthetic plastics, analysis of blood triglycerides, and synthesis of pharmaceuticals and fine chemistry.[15–17]

Some of these previously mentioned processes require enzymes that remain active and stable under harsh conditions including a broad range of temperatures, pH, detergents or organic solvents.[18,19] Using enzymes in organic solvents has many advantages, e.g. improved solubility of hydrophobic compounds, suppression of undesirable, water dependent side reactions, prevention of microbial contamination and, with respect to lipolytic enzymes, the shift of thermodynamic equilibrium from hydrolysis toward synthesis.[15,20]

The main disadvantage is the inactivation or denaturation of enzymes by organic solvents.[23] In the past, process conditions were optimized to overcome the limitations of the available enzyme using physical and chemical methods such as immobilization, chemical modification or ionic liquid coating of the biocatalyst, or the modification of the solvent

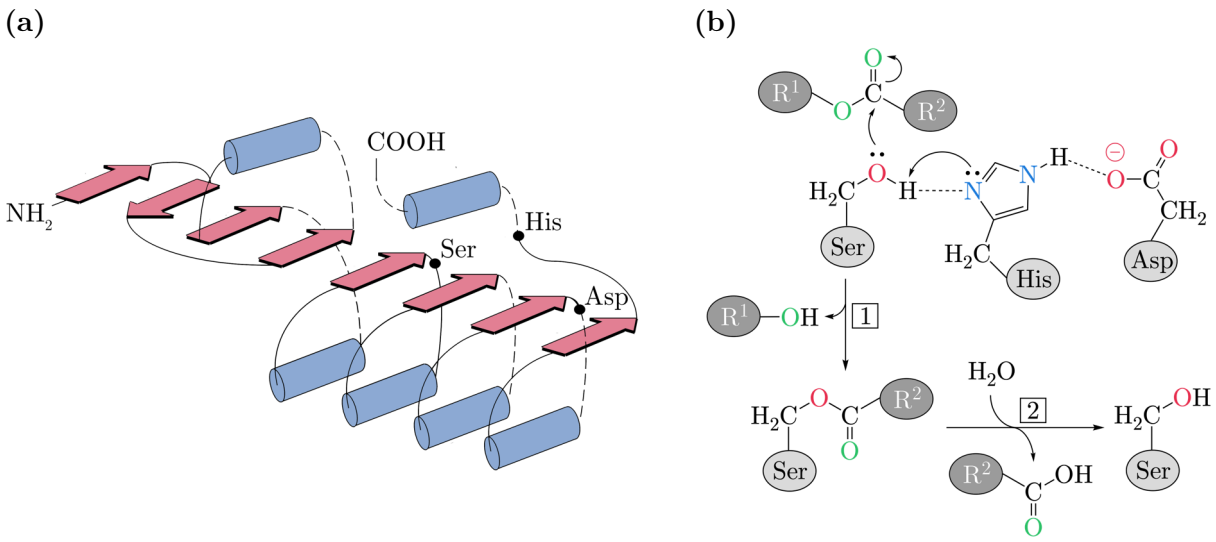

Figure 1.2: Structural and functional properties of lipolytic enzymes. (a) The secondary structure of the 'canonical' α/β hydrolase fold (modified version from Nardini *et al.*, 1999). The α-helices and β-strands are colored blue and red, respectively. The black dots mark the position of the catalytic triad, represented by the residues Ser, His and Asp. Broken lines indicate potential insertions, which can comprise a few amino acid residues up to a complete extra domain, provided that the catalytic functionality is retained.[6] (b) Simplified representation of the mechanism of the catalytic triad for the hydrolysis of an ester substrate. Residues that belong to the catalytic triad (Ser, His and Asp) are shaded light gray, those belonging to the substrate are shown in darker shade. At first, the acidic residue (Asp) polarizes His, which then withdraws the hydroxy proton from serine. Via its oxyanion, the serine can now nucleophilically attack the substrate's carbonyl carbon, releasing an alcohol as the first product. In the second step, the carbonyl carbon of the previously formed acyl-serine complex is now attacked by a water oxygen acting as a nucleophile. The second product, a carboxylic acid, is released and the initial state of the catalytic triad is restored.[21,22]

environment using additives or surfactants. Meanwhile, screening for enzymes that already intrinsically have the desired properties or that have been genetically modified to withstand harsh industrial conditions has gained popularity.[15,24]

Extremophilic organisms, which thrive in environments of extreme temperature (as low as –12 °C and as high as 122 °C)[25,26], pressure (as high as 100 MPa)[27], ionic strength (as high as 5.1 M NaCl)[28], and pH (low as 1.1 and high as 12.5)[29,30], have found to offer enzymes that are adapted to work under the extreme conditions of many industrial processes.[31,32] For example, lipolytic enzymes from (halo)alkaliphiles are of interest for their application potentials in laundry detergents and finishing of fabrics,[33,34] while the use of lipases from thermophiles can enhance lipolytic reactions due to the improved solubility and reduced viscosity of lipidic substrates at high temperatures.[35]

Due to the complex, often unknown growth conditions in their natural habitat, the vast majority of microorganisms cannot be cultured. Until recently, this limitation meant

that most genomes, and hence a variety of enzymes from the microbial world were not accessible.[36,37] Metagenomics, therefore, offers a powerful culture-independent approach to search for novel industrial enzymes by comprehensively analyzing microbial genome mixtures from various environments.[38,39]

Several organic solvent tolerant esterases and lipases were identified from cultured microorganisms or metagenomes obtained by sampling various environments, including soil,[40–42] compost,[18,43] salt and soda lake,[33,44,45] activated sludge,[46,47] polar regions,[48–50] hot springs,[51–53] and oil-contaminated sand or soil.[54–56] Most of these enzymes additionally showed high thermostability or tolerance to high salt concentrations, which supports the assumption of a correlation between thermal, salt and organic solvent tolerance.[57,58]

To circumvent the effort of screening for novel lipolytic enzymes, already existing and well established enzymes can be genetically modified using directed evolution or rational design to obtain mutants that remain active and stable in organic solvents.[24] However, those methods are still time-consuming and expensive.[23] A generally applicable guide for targeted improvement of the tolerance of enzymes to organic solvents is thus in high demand.

1.2 Engineering of organic solvent tolerant enzymes

1.2.1 Stability of enzymes

There are two types of enzyme stability, the thermodynamic and the kinetic stability. Many studies focus on thermodynamic stability by investigating the propensity of enzymes to undergo reversible unfolding. Thermodynamic stability can be defined on the basis of equilibrium thermodynamics as the free energy (ΔG) of folding/unfolding. Common quantities related to ΔG are the melting temperature T_m and its chemical equivalent $C_{1/2}$, the concentration of denaturant required to unfold half of the protein. These parameters can be determined using experimental techniques such as circular dichroism spectroscopy or differential scanning fluorimetry.[59,60]

Kinetic stability, on the other hand, is considered as the length of time an enzyme remains active before it irreversibly denatures. A common measure of kinetic stability is the half-life ($t_{1/2}$) of denaturation or the time required for the enzyme to lose half of its original activity under specific conditions. Other measures include the deactivation constant (K_d), the optimum operating temperature (T_{opt}) and T_{50}, the temperature at which, after a specific incubation period, the original activity is reduced by half.[59]

Thermodynamic and kinetic contribution to enzyme stability can be described using a

simple equilibrium reaction between two protein conformations (figure 1.3). The natively folded state (N) and the unfolded state (U) are separated by a higher energy transition state (TS). Kinetically stable proteins usually possess a high TS, which is associated with extremely slow unfolding rates. In a sense, kinetically stable proteins are trapped in their native conformations by a higher free energy barrier ($\Delta G^{\ddagger}$), even though unfolding is thermodynamically favorable ($\Delta G < 0$) under denaturing solvent conditions.[61]

Organic solvents have the ability to denature and inactivate enzymes by destructing their structural integrity. Hydrophobic interactions are the main contributors to the globular conformation of a protein.[62] These interactions take place between hydrophobic domains, which are folded inside the molecules. When organic solvents are introduced, the polarity of the medium surrounding the enzyme is reduced, prompting the hydrophobic core of the protein to disperse and to unfold.[63] Moreover, to keep their conformational flexibility and enzymatic activity, enzymes require water bound to the enzyme surface. Those crucial water molecules are stripped by and replaced with organic molecules, opening the pathway for organic solvents into the protein core resulting in denaturation of the enzyme.[64] It is further known that hydrophilic solvents possess a higher ability to strip water off the enzyme surface than hydrophobic solvents.[15,65] Thus, the interaction between surface residues and water molecules is assumed to play a key role in preventing the replacement of water by polar organic solvents.[66]

1.2.2 Strategies to improve enzyme stability in organic solvents

In the last years, directed evolution has been the main protein engineering approach used to enhance the organic solvent stability of enzymes.[67–72] A directed evolution approach involves the generation of a large gene variant library by random mutagenesis of the target enzyme and subsequent identification of positive mutants using an appropriate selection or screening method. These steps are repeated several times until the resulting enzyme satisfies the requirements.[73,74] To obtain genetic variety, mutations can be introduced site-specifically by randomizing one codon or randomly across the entire target gene using saturation mutagenesis[75] and error-prone polymerase chain reaction (epPCR)[76] methods, respectively.[77,78] Enzyme variants possessing the desired properties are identified either by selection techniques, often performed using agar plates containing selective media, or by screening methods. For example, microtiter plates are popular tools to screen for positive variants by colorimetric or fluorimetric assays.[74,79]

Various attempts have been made to advance directed evolution, including the combination of directed evolution and rational protein design in so-called semirational design approaches,[80–82] the refinement of selection technologies and high-throughput screening,[83–85] or the application of computer-based methods such as machine learning.[86]

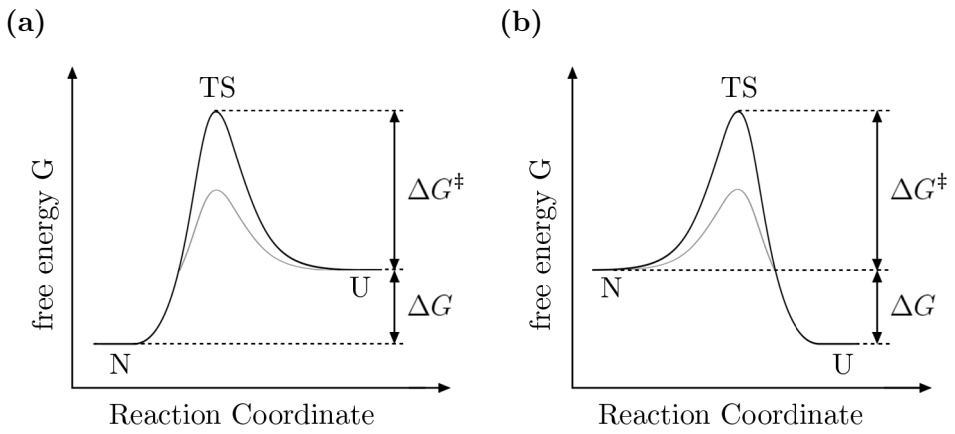

Figure 1.3: Protein unfolding illustrated by a free-energy diagram. The unfolding from the native state (N) through the transition state (TS) to the unfolded state (U) under (a) native and (b) denaturing conditions is shown. Kinetically stable proteins possessing a higher free energy barrier ($\Delta G^{\ddagger}$) are compared to normal proteins (represented in gray). While ΔG relates to thermodynamic stability, $\Delta G^{\ddagger}$ relates to kinetic stability.

However, directed evolution is still very time consuming due to the screening of large libraries (10^3 – 10^6 clones) in multiple rounds of mutagenesis cycles.[87]

By contrast, in rational design approaches only a few mutants are usually generated by site-directed mutagenesis following computer-based predictions. This approach requires knowledge of the enzyme's function and structure obtained, for example, from crystal structures or homology models.[88] As mentioned before, a correlation between thermostability and organic solvent tolerance is presumed.[57,58] This is supported by several reports of mutations that stabilized proteins not only at high temperatures but also in the presence of organic solvents.[89–93] Hence, it is expected that methods that have improved the thermostability of an enzyme can also be used to design an enzyme that is tolerant to organic solvents.

A common approach to stabilize an enzyme is the introduction of disulfide bonds into the protein structure, which has increased the thermostability of several enzymes in the past.[94–97] By engineering disulfide bonds, researchers were also able to improve the tolerance of enzymes towards organic solvents.[98–100] In addition, there is strong evidence that hydrogen bonds and salt bridges are important structural features for protein stability.[101–104] Kawata *et al.* were able to improve the organic solvent stability of a lipase by introducing mutations that were found to enhance the packing of the hydrophobic core by forming salt bridges and hydrogen bonds.[105] However, the contribution of individual hydrogen bonds to protein stability is highly dependent on their environment. Even if the amino acids are properly positioned for bonding, there is no guarantee that their introduction will have stabilizing effects on the overall protein.[101,106–108] Thus, the prediction

of stabilizing effects by the introduction of mutations forming hydrogen bonds or salt bridges is limited.

Energy calculations, machine learning, and phylogenetic analyses have proven to be reliable and effective computational methods for predicting the impact of site-directed mutations on protein stability.[60] The consensus design (CD) method, for instance, is a phylogenetic analysis tool for identifying conserved residues that are assumed to have a stabilizing effect on the protein structure.[109,110] Dror *et al.* were able to generate a lipase with 66-fold improved tolerance to high concentrations of methanol by introducing consensus residues into the protein sequence.[111]

For the prediction of energetic changes caused by mutations, the calculation of the relative change in free energy ($\Delta\Delta G$) upon mutation is a commonly used method of stability predictors.[60] A force field combining a set of parameters and energy functions representing bonded and unbonded interactions is thereby used to estimate the potential energy of the protein structure in a molecular system.[112] The most accurate stability predictions can be obtained from molecular dynamics (MD) or Monte Carlo (MC) simulations using the free energy perturbation method. However, due to the large computational cost, this method is limited to a small number of mutations.[113]

There are also less computationally intensive and simulation-independent stability predictors, which use statistical (SEEF), empirical (EEEF) or physical effective energy functions (PEEF).[114] One of the most popular tools for predicting $\Delta\Delta G$ is FoldX.[115,116] It uses a knowledge-based potential, which relies on statistical analysis of different properties from protein databases. The prediction of $\Delta\Delta G$ is further the basis of many recently developed machine learning methods, which use statistical models, that are build by recognizing patterns in a training data set.[117–121]

However, for the engineering of enzymes in organic solvents in particular, structural changes and interactions between enzyme and solvent can provide relevant information about the stability of the enzyme. MD simulation is the method of choice to investigate the motion of a protein and its surrounding solvent on the atomic level. Park *et al.*, for example, used MD simulations of a lipase in organic solvent to develop an enzyme stabilization strategy. By analyzing the root mean square deviation (RMSD) and the solvent-accessible surface area (SASA), they were able to predict the solvent affecting site on the enzyme surface and design variants with enhanced stability in methanol based on this information. The stabilization strategy by Park *et al.* is based on the assumption that enzyme stability in organic solvents can be improved by lowering the structural fluctuation of surface residues and thereby improve hydrogen bond interactions between surface residues and water molecules.[122]

Cui *et al.* tested another hypothesis based on the observation that salt bridges on the

protein surface are stabilized in polar organic solvents, which may impair the enzyme activity by reduced structural flexibility.[123–125] They were able to generate a lipase with improved organic solvent resistance by disrupting those unfavorable salt bridges identified by MD simulations and replacing them with residues that exhibit improved flexibility and enhanced shells.[123]

Despite conflicting hypotheses of Park *et al.* and Cui *et al.*, both studies underlined the relevance of surface residue hydration in organic solvent tolerant enzymes.[122,123] This is, moreover, supported by several directed evolution and rational design studies reporting the stabilization of enzymes in polar organic solvents by introducing polar or charged amino acids, which facilitate the formation of hydrogen bonds.[123,126–128]

Little attention has so far been paid to the fact that lipolytic enzymes with enhanced stability in presence of organic solvents usually contain a large excess of negatively charged residues. Comprehensive sequence and functional analyses of lipolytic enzymes further revealed that the lower the isoelectric point (pI) of the protein, the more active and stable the enzyme is in presence of higher organic solvent concentrations.[19] In particular, negatively charged residues such as glutamic and aspartic acid are known to have strong water-binding properties, while positively charged groups bind water more weakly.[129,130] Moreover, a correlation between increased negative surface charge and increased protein solubility was reported.[130]

Surface charge engineering has previously been used to improve the stability, activity and solubility of proteins.[131] For example, by introducing 8 mutations to the *Bacillus subtilis* lipase A (BsLA), an extremely negatively charged mutant with highly improved thermostability was obtained.[132] However, enhanced organic solvent tolerance through engineering of negative surface charges has not been reported to date.

1.3 A stable esterase and its unstable counterpart

After the Prestige oil spill in November 2002, the Gram-negative bacterium *Pseudomonas aestusnigri* VGXO14 was isolated from crude oil-contaminated sand samples.[133] Due to its hydrocarbon-rich environment, which also contains organic solvents, and its genome sequence harboring a number of esterase-encoding genes, the bacterium was assumed to encode and produce organic solvent resistant esterases.[134,135]

Bollinger *et al.* identified 13 novel esterases by activity- and sequence-based screening of the *P. aestusnigri* genome, among which Paes_TB035 (PT35) was found to be highly tolerant to several water miscible organic solvents. After 3 h incubation of cell lysate in 80 % of acetonitrile (ACN), 1,4-dioxane and methanol, PT35 showed 33 %, 58 % and 64 %

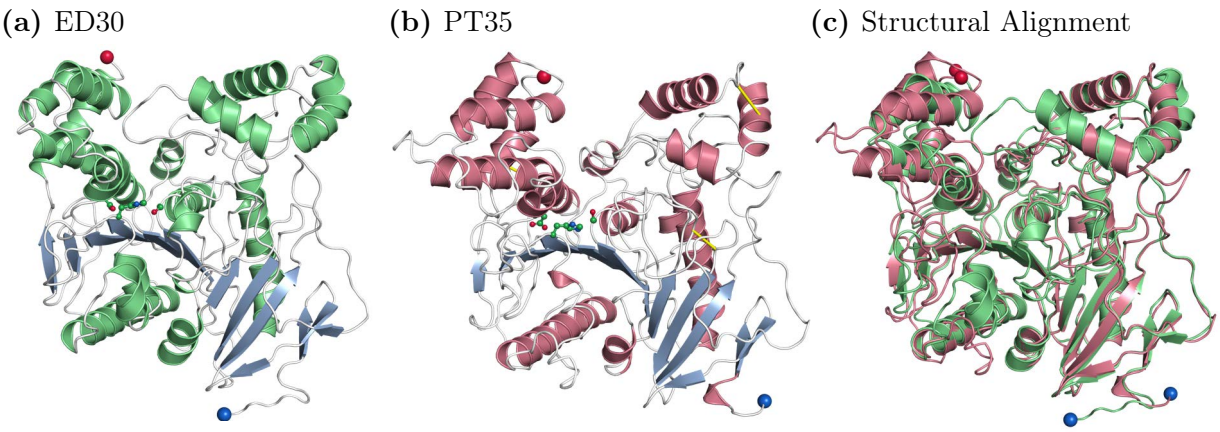

Figure 1.4: Homology models of ED30 and PT35. Homology models of (a) ED30 and (b) PT35 with the secondary structure elements colored as followed: helices (light green or light red), β-sheet (light blue), coil (white). The catalytic triad consisting of S195, E319 and H412 (ED30) or S183, E301 and H418 (PT35) is shown as a ball and stick representation. The atoms of the catalytic triad are colored in green (carbon), red (oxygen) and blue (nitrogen). Disulfide bonds predicted for PT35 are shown as yellow sticks. (c) Comparison of ED30 (light green) and PT35 (light red) via structural alignment. The N-terminus and C-terminus are highlighted as blue and red spheres, respectively.

residual activity, respectively. In the presence of dimethyl sulfoxide (DMSO), the activity of PT35 was even increased, resulting in about 264 % activity after 24 h.[54]

PT35 was further biochemically characterized by M. Bulka using purified enzyme.[136] Here, PT35 was incubated in 10 %, 30 % and 50 % of ACN, 1,4-dioxane, methanol and DMSO and retained activity in the higher concentrations even for one week. Moreover, PT35 was found to prefer short-chain acyl esters, indicated a moderate thermostability ($T_{50} \approx 48\,°C$) and a pH optimum of 10, which suggests that PT35 is a mesophile and alkaline esterase. Besides organic solvent tolerance, PT35 remained also stable and active in presence of several detergents (sodium dodecyl sulfate (SDS), Triton X-100 and CHAPS), monovalent cations (Cs^+, K^+, Na^+, Rb^+), and high NaCl concentrations. Inhibition, on the other hand, was observed for the positively charged detergent cetyltrimethylammonium chloride (CTAC), several divalent cations and the chelating agent ethylenediaminetetraacetic acid (EDTA). Due to its remarkable SDS-resistance, PT35 is assumed to exhibit high kinetic stability. The tolerance towards SDS is often associated with a rigid protein structure due to a high content of β-sheets, and is therefore considered a common property of kinetically stable proteins.[61]

On the basis of computational analyses, the 564 residue and 58.64 kDa PT35 was classified as a member of the family VII and predicted to be a SEC-secreting lipoprotein with an N-terminal 39-residue signal peptide. In addition, PT35 was found to be highly negatively charged due to its excess of negatively charged residues (68, D and E) compared to

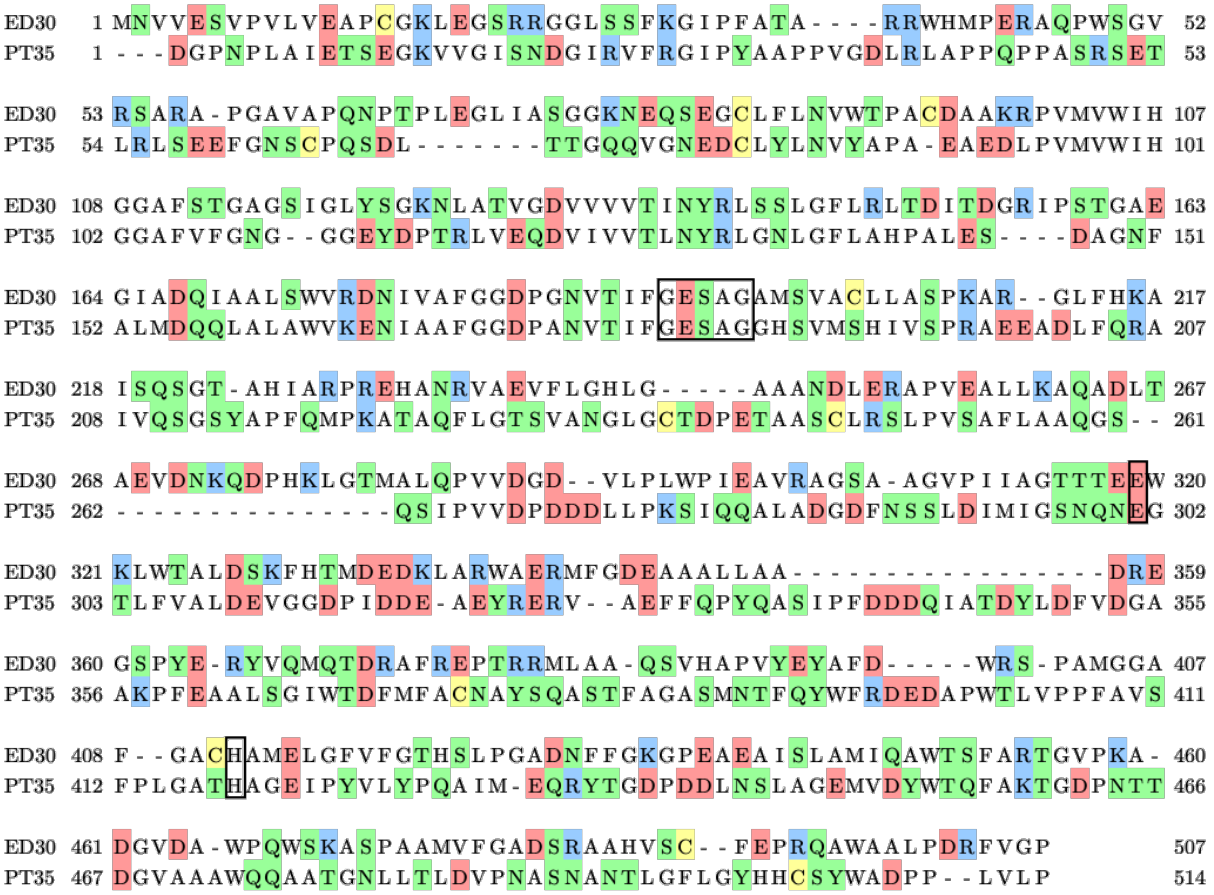

Figure 1.5: Sequence alignment of ED30 and PT35. The alignment was performed with BLAST.[137,138] Polar residues are colored in green, cysteines in yellow, positively charged residues in blue and negatively charged residues in red. Residues belonging to the catalytic triad and the GESAG pentapeptide are framed.

positively charged residues (23, K and R), which results in a low pI of 4.1. This property of PT35 was supposed to play a certain role in its resistance to salts and SDS.[136]

A homology model of PT35 was generated by J. Loschwitz and is used for the computational part of this study (figure 1.4). This model lacks the first 50 N-terminal residues, as a higher model quality was possible without them. The resulting protein has 514 amino acids and a molecular weight of about 55 kDa. For this model, a negative net charge of –48 and a pI of 4.02 was computed. In addition, an α/β hydrolase fold, a catalytic triad comprising S183, E301 and H418 with serine being part of the GESAG pentapeptide and three stabilizing disulfide-bridges (C64-C80, C237-C246, C374-C503) were predicted.[136,139]

In order to investigate the structural properties causing the remarkable stability of PT35, the related but unstable esterase EstDL30 (ED30) was chosen for comparative computational studies.[136,139]

The ED30 gene was previously isolated from an alluvial soil metagenomic library and

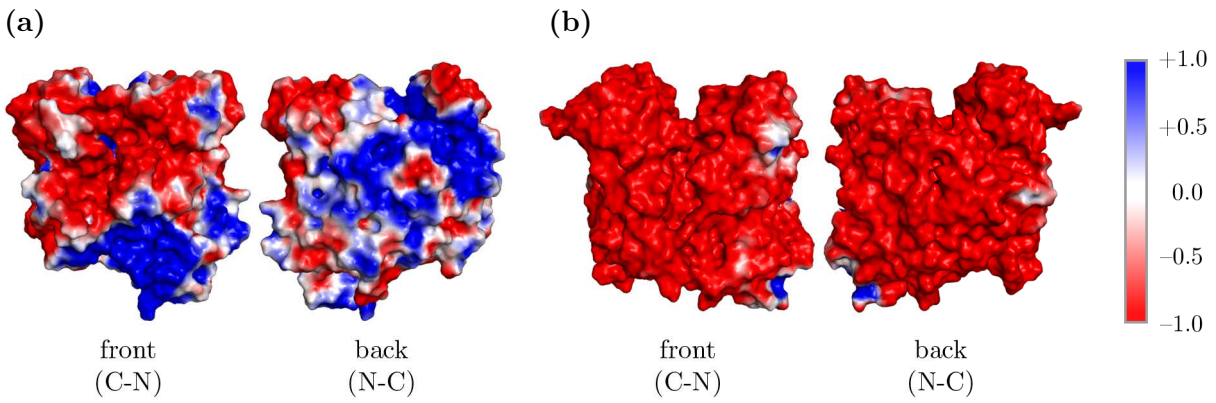

Figure 1.6: Electrostatic potential surface of ED30 and PT35. The electrostatic potential surface (EPS) was computed for (a) ED30 and (b) PT35 using APBS in PyMOL and is visualized in colors ranging from blue (positive, 1) to white (uncharged, 0) and red (negative, –1). Two orientations are presented for both enzymes with the N-terminus on the right and the C-terminus on the left site (C–N) as well as its backside, where the N-terminus is on the left and the C-terminus on the right (N–C).

encodes a 507-residue esterase of family VII with a molecular weight of about 54 kDa.[140] Since no crystal structure of ED30 is yet available, a homology model was generated by J. Loschwitz and is used in the computational part this study (figure 1.4). Based on this model, a catalytic triad formed by S195, E319 and H412, where serine is part of a GESAG pentapeptide, was predicted. Just like PT35, ED30 contains 6 cysteines, which, however, are not predicted to form disulfide bridges. In contrast to PT35, a less negative net charge of –6 and a pI of 6.1 was computed for ED30, which is due to its similar content of positively and negatively charged residues.

Previous biochemical studies found that ED30, just like PT35, preferentially converts short-chain acyl esters and can be considered as an alkaline and mesophile esterase with an optimal temperature of 40 °C.[140] However, ED30 exhibits its maximum activity at pH 8, which is slightly more acidic than the pH optimum of PT35 (pH 10). A moderate tolerance toward the organic solvents methanol and DMSO, EDTA and most divalent cations was shown for ED30. While activation was observed by Fe^{2+}, the activity of ED30 was strongly decreased in presence of Cu^{2+} and Zn^{2+}, SDS and ACN (acetonitrile).[140]

The differences between PT35 and ED30 in respect to their organic solvent tolerance become clear when comparing their activity in distinct water miscible organic solvents in detail. While PT35 was found to be highly stable in 50 % ACN with a residual activity of about 60 % after one week, ED30 showed only 22 % residual activity in 15 % ACN and was almost completely inactivated in 30 % ACN after 40 min incubation. After incubation in 30 % DMSO and methanol, PT35 was fully active in both organic solvents even after one week, while ED30 showed residual activities of 40 % and 63 %, respectively, after

40 min.[136,140]

Both esterases were compared by a pairwise alignment using the basic local alignment search tool (BLAST)[137,138], which revealed 33 % sequence identity and 45 % similarity (figure 1.5).[136] Additional structural comparison of the homology models of PT35 and ED30 indicated a similar α/β hydrolase fold (figure 1.4). A striking difference between PT35 and ED30 is the large charge difference (–42), which was visualized in the electrostatic potential surface (EPS) (figure 1.6). While the surface of ED30 exhibits negative and positive potentials, PT35 has an almost completely negatively charged surface.[139]

In previous studies by A. Jäckering, MD simulations of PT35 and ED30 were performed in presence of organic solvents (methanol, ACN), SDS and the substrate 4-nitrophenyl butyrate (*p*NPB). SDS was found to preferentially interact with positive or neutral surface potential areas while avoiding negative regions. In addition, stronger interactions between SDS and ED30 than between SDS and PT35 were observed, while interaction of SDS with negatively charged residues, especially with those of PT35, was disfavored. These findings were associated with the hypothesis that the negatively charged surface of PT35 repels SDS molecules, which may prevents the enzyme to denaturate.[141] Similar observations were made for the interactions of organic solvents with PT35 and ED30. PT35 and in particular its negatively charged residues were found to be less favored by organic solvents than ED30, which was especially observed for ACN. These findings agreed with the hypothesis that negatively charged residues contribute to the maintenance of the protein hydration shell, which is required for the structural integrity of the enzyme in solvent mixtures including organic solvents.[64,130] Therefore, it was hypothesized that the excess of negatively charged residues combined with the extremely negatively charged surface of PT35 play a key role in the resistance of PT35 not only to SDS but also to organic solvents.[139]

1.4 Aim of the work

Enzymes capable of catalyzing reactions in organic solvents offer many industrially attractive advantages. However, native enzymes generally exhibit low activity and stability in the presence of organic solvents, limiting the efficiency of the reaction process.[23] Several attempts have been made to overcome this limitation by physically, chemically or genetically modifying the target enzyme.[15,20] However, no efficient, generally applicable strategy for enzyme stability engineering in polar organic solvents has been reported to date.

The aim of this work is to test the hypothesis that organic solvent resistance of enzymes can be altered by modifying their surface charge and thus hydration shell. Therefore, two

esterases, the stable PT35 and the unstable ED30 are re-engineered in a rational design approach. The mutagenesis strategy is based on computational analyses, such as the determination of residue conservation, the solvent accessible surface area for the identification of exposed residues, changes in free energy for the estimatation of thermodynamic stability, changes in the electrostatic potential (ESP) upon mutation of charged residues, and MD simulations to estimate the structural stability and hydration shell of the substitutions. Surface potential hot spots are identified by screening the areas with the largest charge difference between the EPS of PT35 and ED30. By inserting charged or polar residues according to certain substitution principles into these hot spots, the surface of PT35 becomes more positive and that of ED30 more negative.

The most favorable substitutions are combined to obtain triple and 12-fold mutants (called multi-mutants) of PT35 and ED30. For both ED30 and PT35, one triple and one multi-mutant was investigated in comprehensive MD simulations and produced for experimental stability studies. ACN, as one of the most important water miscible solvents in industrial processes, was chosen for studying the stability of PT35 and ED30 variants in MD simulations and in experimental approaches.

The results of the work will contribute to the development of rational protein design methods using surface charge and hydration shell engineering, which could significantly facilitate the production of various industrially relevant enzymes with improved stability in presence of polar organic solvents.

2 Material and methods

2.1 Computational Section

For the computational part of this study, the software listed in table 2.1 was applied.

Table 2.1: Software and its application in the computational part of this study.

SOFTWARE	APPLICATION
Adaptive Poisson-Boltzmann solver (APBS)[142] PyMOL[143] plugin	calculation of the electrostatic potential surface (EPS)
FoldX 4.0 YASARA Plugin[115,144] in YASARA 20.10.4[145]	calculation of the relative change in free energy ($\Delta\Delta G$)
GROMACS 2018.6[146,147]	MD simulation and analysis
Jalview 2.11.1.3[148]	analysis and visualization of pairwise and multiple sequential alignments
PyMOL 2.3[143]	in silico mutagenesis and visualization of enzyme structures
Python	plotting of data

High-performance computing (HPC) clusters used to compute MD simulations and analyses were provided by the institutions listed in table 2.2.

Table 2.2: HPC clusters used for MD simulations and analyses.

CLUSTER	FACILITY
CHEOPS	Regional Computing Center at the University of Cologne
HILBERT	Centre for Information and Media Technology at the Heinrich-Heine University
JURECA	Jülich Supercomputing Centre at the Forschungszentrum Jülich[149]

2.1.1 *In silico* generation of single, triple and multi-mutants

Since no qualitatively sufficient crystal structures are yet available for any of the two esterases, the homology models of PT35 and ED30 provided by J. Loschwitz were used as templates to prepare single, triple and multi-mutants. The homology modeling was done with the iterative threading assembly refinement (I-TASSER) webserver[150–152]. Based on the best results and the highest confidence score (C-score), a PT35 model lacking the first 50 N-terminal residues due to a higher C-score of 1.26 without these residues and a ED30 model with a C-score of 0.33 were selected. In addition, 500 ns MD simulations of the wild type PT35 and ED30 in water performed by A. Jäckering were provided.

In order to obtain a more positively charged PT35 and a more negatively charged ED30 surface, negatively charged residues (E, D), positively charged residues (R, K), or polar residues (Q, N, (S)) were substituted by one of the other residue types. Substitutions were introduced into the wild type structure using the PyMOL[143] mutagenesis tool, where the side-chain rotamers producing minimal steric clashes were chosen.

Mutation candidates were determined by structural examination of both homology models (section 2.1.3.1). Single mutants were generated by introducing compatible substitutions to the wild type structure of PT35 and ED30. A compatible substitution and the respective single mutant were evaluated on the basis of (i) residue conservation (section 2.1.3.2), (ii) residue exposure (section 2.1.3.3), (iii) change of electrostatic potential surface (EPS) (section 2.1.3.4) and (iv) the relative change in free energy ($\Delta\Delta G$) (section 2.1.3.5).

The most favorable single mutants of ED30 (R23N, K78Q, K99D, K124D, R211E, K278Q, R302Q, K329E, K337E, R340D, R373D, R381N, R400Q, Q447E, K471N) and PT35 (N20R, D68N, Q73R, D93K, E121Q, E199R, D268R, Q279R, D314N, E328K, Q331R, N375R, Q379R, Q429K, D439K) were selected for investigations using MD simulations (section 2.1.2.3). Subsequently, single mutants were further assessed with respect to their (v) stability (section 2.1.3.5) and (vi) hydration shell (section 2.1.3.6).

Triple and multi-mutants were generated by combining 3 and 12 substitutions, respectively. Those combinants showing the most pronounced alteration and distribution of charges in the EPS were selected for more comprehensive studies. Therefore, one triple mutant (ED30_tmut: K124D/R211E/K337E, PT35_tmut: E199R/E328K/D439K) and one multi-mutant (ED30_mmut: R23N/K78Q/K99D/K124D/R211E/K329E/K337E/R373D/R381N/R400Q/Q447E/K471N, PT35_mmut: N20R/D93K/E121Q/E199R/D268R/D314N/E328K/Q331R/N375R/Q379R/Q429K/D439K) were simulated in water and a water-ACN mixture for 500 ns (section 2.1.2.3). In addition to the most favorable triple mutant, 5 alternative triple mutants of both ED30 (S122D/K278Q/K329D, K99D/R211E/R302Q, R211E/K329D/Q447E, R23N/R400Q/K471N, K278Q/K337E/

R340D) and PT35 (N20R/Q73R/D314N, D93K/E199R/D268N, Q73R/E121Q/E199R, N20R/N375R/Q379R, E328K/D268R/N375R) were generated and simulated in water (section 2.1.2.3).

2.1.2 Molecular dynamics simulations

MD simulations provide an opportunity to study the dynamical properties of a system on the molecular scale. This method is based on the numerical solution of the classical equations of motion, also known as Newton's second law of motion. It can be defined as in equation 2.1, with the mass m_i, the acceleration a_i and f_i describing the force acting on atom i, which is derived from the potential energy $U(R)$. Here, $R = (r_1, r_2, ...r_N)$ represents the set of $3N$ atomic coordinates for a system of size N.[153]

$$m_i a_i = f_i \qquad f_i = -\frac{\partial U(R)}{\partial r_i} \tag{2.1}$$

2.1.2.1 The force field

The parameters and the mathematical function required for calculating $U(R)$ are combined in the force field. Equations 2.2 and 2.3 express the energy function of common atomistic force fields with energy terms representing bonded and nonbonded interactions.[112]

$$U(R)_{bonded} = \frac{1}{2}\sum_{bonds} k_b(b-b_0)^2 + \frac{1}{2}\sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi\left[1+cos(n\phi-\delta)\right] \tag{2.2}$$

$$U(R)_{nonbonded} = \sum_{LJ} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{Coulomb}\frac{q_i q_j}{r_{ij}} \tag{2.3}$$

The energy of bonded interactions comprise harmonic potentials for vibrations of bond lengths b and angles θ, and periodic functions for rotations around bonds measured by dihedral angles ϕ (equation 2.2). The harmonic potentials are a result of squared deviations from the equilibrium values b_0 and θ_0 combined with the force constants k_b and k_θ. A periodic cosine function with the force constant k_ϕ, the periodicity n and the phase angle δ describes the potential associated with the rotation around bonds, where the dihedral angle is determined by three covalent bonds, hence four atomic positions.[153,154]

Nonbonded interactions involving van der Waals and electrostatic interactions are represented by Lennard-Jones (LJ) potentials and Coulomb's law, respectively (equation 2.3). The LJ equation with the prefactor ϵ_{ij} models attractive dispersion (r^{-6}) and repulsive Pauli exclusion (r^{-12}) interactions. If the interatomic distance r_{ij} between atom i and j coincide with σ_{ij}, the LJ potential is minimal. The second part of equation 2.3 is the

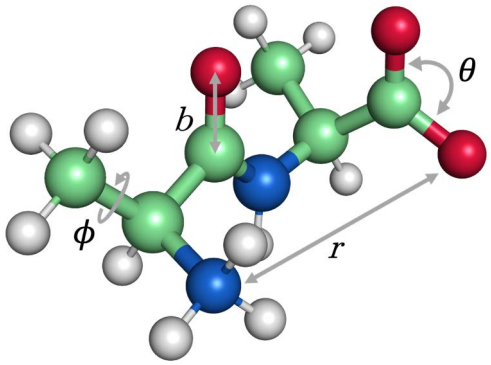

Figure 2.1: Molecular interactions modeled by the force field schematically shown for alanine dipeptide. Bonded interactions involve bond length b vibration, angle θ bending, and rotation around torsion angle ϕ. The distance between two atoms of a nonbonded (LJ or Coulomb) interaction is displayed by r.

Coulomb's law where q_i and q_j represent the charges of atoms i and j.[153,155]

To avoid surface artifacts, the unit cell is surrounded with replicas of itself, which is referred to as periodic boundary conditions (PBC). Thus, the number of interactions within the system is infinite. To minimize the computational costs, which are mainly caused by the calculation of nonbonded potentials, cutoffs are required. While simple cutoffs can be used for LJ interactions due to their fast decay, Coulomb interactions in periodic systems require an alternative technique due to their slow decay. An appropriate method for the calculation of Coulomb interactions is the particle mesh Ewald (PME) method, which is based on the decomposition of the electrostatic potential into short-ranged and long-ranged contributions.[156] While the short-ranged part is calculated by direct summation of forces, long-ranged contribution needs to be approximated using a Fast Fourier transformation.[112,157]

The most frequently used biomolecular protein force fields include Amber[158], CHARMM[159], GROMOS[160] and OPLS-AA[161] force fields. All four families of force fields are based on the general energy functions in equation 2.2 and equation 2.3. In this study, the AMBER99SB*-ILDNP force field[162–164] was used. The Amber force field was originally developed by Kollman *et al.* and published for the first time in 1981.[165] Since then, several AMBER force fields have been developed, each containing a set of parameters appropriate for specific types of molecules. The ff99 force field was developed for organic and biological molecules using restrained electrostatic potential charges.[166] Due to several reported limitations in the ff99 force field, such as overstabilization of helical conformations, a new parameter set, ff99SB, was compiled with improved ϕ and ψ dihedral terms.[167] Further improvement of the side-chain torsion parameters for the residues Ile, Leu, Asp, and Asn resulted in ff99SB-ILDN and optimization of the Pro torsion parameters yielded the ILDNP force field.[163,164] Due to these adaptations and improvements,

the AMBER99SB*-ILDNP force field is particularly suitable for simulations of protein structures.

2.1.2.2 The numerical integration of the equations of motion

Besides the force field parameters and functional form of the potential $U(R)$, an MD simulation requires atomic coordinates of the starting structure and random initial velocities to calculate the forces f_i acting on each atom. A trajectory of molecular motion is obtained by numerically integrating the equations of motion (equation 2.1) and repeating the previous steps with the newly computed coordinates and velocities, using successive time steps δt.[112] The individual steps of an MD simulation are shown in figure 2.2 as a flowchart. There are several techniques to solve the equation of motion (equation 2.1). The Verlet algorithm, introduced by Verlet in 1967,[168] and its variants are the most commonly used methods in MD programs.[169] The original Verlet algorithm is based on a Taylor expansion as described in equation 2.4 with the positions $r(t)$, the velocities $v(t)$ and accelerations $a(t)$.

$$\begin{aligned} r(t+\delta t) &= r(t) + \delta t v(t) + \tfrac{1}{2}\delta t^2 a(t) + ... \\ r(t-\delta t) &= r(t) - \delta t v(t) + \tfrac{1}{2}\delta t^2 a(t) + ... \end{aligned} \tag{2.4}$$

The addition of both equations in 2.4 gives the equation 2.5.

$$r(t+\delta t) = 2r(t) + \delta t^2 a(t) - r(t-\delta t) \tag{2.5}$$

Since the method uses information from the previous step $r(t-\delta t)$, the Verlet algorithm is time-reversible. Here, the velocities are useful to calculate the kinetic energy, and hence the total energy, but they are not needed to compute the trajectories.[169]

In contrast, the Leap Frog algorithm,[170] which is derived from the Verlet algorithm (equation 2.5), uses velocities to calculate new positions by evaluating velocities at half time steps as written in equation 2.6.

$$\begin{aligned} v(t+\tfrac{1}{2}\delta t) &= v(t+\tfrac{1}{2}\delta t) + \delta t a(t) \\ r(t+\delta t) &= r(t) + \delta t v(t+\tfrac{1}{2}\delta t) \end{aligned} \tag{2.6}$$

As the name implicates, the velocities leap over the coordinates to give the next half time step values $v(t-\frac{1}{2}\delta t)$. Thus, velocities and positions are not defined at the same time. Since kinetic and potential energy are therefore also determined at separate times, the total energy cannot be calculated directly by the Leap Frog algorithm.[171]

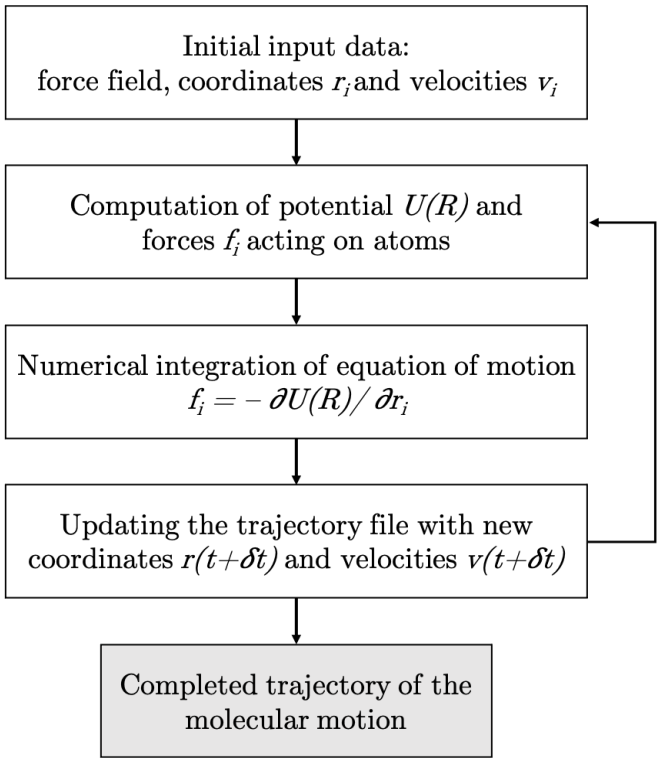

Figure 2.2: Simplified flowchart of a molecular dynamics simulation.

However, for computing positions and velocities at equal times, the so-called Velocity Verlet algorithm[172] was developed. The equations 2.7 are derived from the original implementation of the Verlet algorithm (equation 2.4) and resemble the Leap Frog algorithm (equation 2.6).

$$\begin{aligned} r(t+\delta t) &= r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) \\ v(t+\delta t) &= v(t) + \frac{1}{2}\delta t \left[a(t) + a(t+\delta t)\right] \end{aligned} \tag{2.7}$$

The Velocity Verlet algorithm can calculate the new velocities only after the new positions $r(t+\delta t)$, hence the new accelerations $a(t+\delta t)$ with $a = \frac{f}{m}$ (equation 2.1), have been calculated.[169,171]

To minimize the number of integration steps and thus the computational effort, a long time step δt should be used. The time step size is limited by the time scale of the fastest vibrational motions to be simulated. In most simulations, the bond vibrations are not of interest, hence they can be removed entirely by applying bond constraint algorithms such as SHAKE[173] or LINCS[174], which allow the time step to be extended from 1 fs to 2 fs.[112,169]

The time step can further be increased by restricting the hydrogen degrees of freedom. In this so-called virtual sites (VS) method, hydrogen atoms are treated as *dummy* atoms, which allows a time step of 4 fs.[175]

2.1.2.3 Conduction of MD simulations

All MD simulations performed in this work were realized with GROMACS 2018.6[146,147] using the following procedure.

Setup. The AMBER99SB*-ILDNP[162–164] force field was used for all simulations of PT35 and ED30 mutants. The protonation state of ionizable residues was defined by default (K and R: protonated, D and E: unprotonated, H: neutral) which corresponds to the experimental conditions at pH 7.2 and pH 8.0, respectively. In order to obtain similar simulation box sizes, the structures were placed in a dodecahedron simulation box and solvated using TIP3P[176] water molecules, applying a minimal distance of 0.8 nm for PT35 and 1.2 nm for ED30 variants from the edge of the box to the protein. Thus, the respective box volume was 632.93 nm^3 or 638.95 nm^3 with an enzyme concentration of 2.62 mM or 2.60 mM. The system was subsequently neutralized by randomly replacing solvent molecules with sodium ions.

The single and 5 alternative triple mutants were simulated in water only, while the wild type as well as the triple and multi-mutants were additionally simulated in 80 % ACN. The wild type simulations in water were provided by A. Jäckering as well as the model of ACN and the AMBER99SB*-ILDNP force field with implemented parameters of ACN,[139] which were originally taken from a previous study.[177] Based on the box volume (V), the molecular weight (M) and the densitiy (ρ) of ACN, the number of molecules required to achieve a final concentration of 80 % ACN were calculated according to equation 2.8.

$$n(ACN) = \frac{\rho(ACN) \cdot V(box) \cdot 0.8}{M(ACN)} \tag{2.8}$$

The calculated number of ACN molecules was inserted by replacing water molecules. Each simulation system consisted of around 50,000 – 70,000 atoms in total. The setup conditions for each enzyme variant are provided in table 2.3.

Energy minimization. Energy minimization was performed using the steepest descent algorithm[178] to reduce the force of the system to below the threshold of 1,000 kJ mol^{-1} nm^{-1}.

Equilibration. The first equilibration step involved a short 100 ps MD simulation. The canonical (NVT) ensemble was used to relax the solvent around the position restricted protein and to bring the temperature close to 303 K (30 °C). In the second equilibration step, a 1 ns simulation using the isothermal-isobaric (NpT) ensemble was performed to adjust the pressure at 1 bar.

In both steps, the velocity-rescaling thermostat[179] and the Berendsen barostat[180] were

Table 2.3: System sizes of the MD simulations. The enzyme variants, wild type (wt), triple mutant (tmut) and multi-mutant (mmut), were simulated in water-only and in 80 % ACN. For each system the total number of atoms, water molecules (H_2O), sodium cations (Na^+) and ACN molecules are stated. The PT35 variants are shaded in gray, ED30 variants in white.

		Atoms	**H_2O**	**Na^+**	**ACN**
water-only	wt	71,262	21,031	48	
		55,944	15,929	6	
	tmut	63,036	18,281	42	
		63,795	18,554	12	
	mmut	63,091	18,280	31	
		63,692	18,542	24	
80 % ACN	wt	49,021	2,140	48	5,838
		49,425	2,178	6	5,894
	tmut	48,684	2,021	42	5,838
		49,156	2,095	12	5,894
	mmut	49,029	2,118	31	5,838
		49,580	2,256	24	5,894

applied to adjust the temperature and pressure. The LINCS algorithm[174] was used for all bond constraints and the PME method[181,182] was employed to calculate the electrostatic interactions in conjuntion with PBC.

Production run. The production simulation time was chosen to be 100 ns for the single and 5 alternative triple mutants. The wild types, the triple mutants and the multi-mutants were simulated for 500 ns. The temperature was set at 303 K and the pressure at 1 bar using the Nosé-Hoover thermostat[183,184] and the Parinello-Rahman barostat.[185–187] A cutoff of 12 Å was applied for van der Waals (vdW) and short-range Coulomb interactions using the PME method. In water-only simulations, VS were used to enable a time step of 4 fs. When simulating in 80 % ACN, the time step was set at 2 fs and no VS were used. As before, the bonds were contrained using the LINCS algorithm.

The simulations of the wild type, triple mutants and multi-mutants were performed in triplicate and thus the total simulation time per variant is 1.5 µs (3 x 500 ns). The single mutants were simulated only once for 100 ns.

2.1.3 Computational Analysis

Unless otherwise stated, GROMACS 2018.6[146,147] was applied for the analyses of the MD simulations.

2.1.3.1 Structural alignment

Structural alignment of PT35 and ED30 was done using PyMOL 2.3[143]. The alignment was used to identify potential substitutions by screening the neighborhood of a mutation candidate within the other structure for suitable residues. The neighborhood of a mutation candidate is defined as the surrounding area within a radius of 5 Å.

2.1.3.2 Determination of conserved residues

To determine conserved residues, the ConSurf server[188,189] was used. A total of 67 and 86 homologous sequences of ED30 and PT35, respectively, were collected from the Clean_Uniprot database (a modified version of the UniProt database aimed to gather the more reliable sequences) using the HMMER homology search algorithm. The search was limited to 40-60 % sequence identity in order to exclude sequences with too high or too low similarities. Conservation scores (CS) were calculated using the Bayesian method. Each residue was assigned a CS between 1 (variable) and 9 (conserved). Residues exhibiting a CS of 8 or 9 were excluded from the substitution selection process.

2.1.3.3 Solvent accessible surface area

The solvent-accessible surface area (SASA) was computed for individual residues and the overall protein according to the double cubic lattice method (DCLM)[190]. For calculating the SASA, the vdW atomic radii of Bondi[191] were used. The SASA was initially determined for the wild type simulated in water-only in order to identify residues that belong to the protein surface.

As defined in equation 2.9, the relative solvent-accessible surface area (rSASA) of each residue was calculated with respect to the maximum solvent-accessible surface area (maxSASA) for that residue type. Tien *et al.* (2013)[192] recently determined maxSASA for all 20 amino acids using one empirical and one theoretical strategy. Here, the theoretical maxSASA values resulting from computationally constructed Gly-X-Gly tripeptides, were used for normalization, as recommended by the authors.

$$rSASA = \frac{SASA}{maxSASA} \tag{2.9}$$

Residues with an rSASA of at least 30 % were considered *exposed* and are counted as being part of the protein surface.[193,194] The residues with an rSASA < 30 % were defined as *buried*.

2.1.3.4 Electrostatic potential surface

The EPS of the wild types and mutants was calculated using the PyMOL plugin adaptive Poisson-Boltzmann solver (APBS)[142,195] in order to visualize the charge differences between PT35 and ED30 and further determine the change of surface charge upon mutation. The default method *pdb2pqr* was selected to prepare the molecule by assigning partial charges.

2.1.3.5 Structure stability

ΔΔG. The free energy (ΔG) of folding was determined for each mutant and wild type using the YASARA Plugin FoldX 4.0[115] in YASARA version 20.10.4[145]. The energy of the mutant and wild type structures were minimized using the *Repair object* command to optimize the residues by removing van der Waals clashes and bad contacts followed by determining $\Delta\Delta G$ using the *Stability of object* command[144]. The relative change in free energy ($\Delta\Delta G$) was then calculated according to equation 2.10.

$$\Delta\Delta G = \Delta G(mutant) - \Delta G(wild\,type) \tag{2.10}$$

According to the CompassR rule,[196,197] substitutions were considered *potentially active* ($\Delta\Delta G \leq +0.36$ kcal mol^{-1}), *unpredictible* ($+0.36$ kcal mol$^{-1} < \Delta\Delta G < +7.52$ kcal mol^{-1}) and *potentially deactivating* ($\Delta\Delta G > +7.52$ kcal mol^{-1}).

RMSD. The root mean square deviation (RMSD) reports the average distance between the C_α atoms of each structure of a trajectory and the initial structure. It is calculated according to equation 2.11 where $M = \sum_i m_i$, $\mathbf{r}_i(t)$ is the position of atom i at time t and $\mathbf{r}_i^{ref}$ is the position of atom i in the reference structure, which in this work is always taken as the initial structure of the corresponding MD simulation. The RMSD was determined in order to observe conformational changes during the simulation, while a deviation of over 3 Å from the reference structure is considered critical.

$$RMSD(t) = \sqrt{\frac{1}{M}\sum_{i=1}^{N} m_i(|\mathbf{r}_i(t) - \mathbf{r}_i^{ref}|)^2} \tag{2.11}$$

The deviation of the averaged RMSD between single mutant and wild type structure was calculated according to equation 2.12. A ΔRMSD > 2 Å and a ΔRMSD > 3 Å are considered *semi critical* and *critical*, respectively. Larger deviations from the wild type

could provide information about an increased or decreased stability of the mutant.

$$\Delta RMSD = \overline{RMSD}(mutant) - \overline{RMSD}(wild\,type) \tag{2.12}$$

RMSF. The root mean square fluctuation (RMSF) describes the time averaged standard deviation of atomic positions in a trajectory from a reference position. The RMSF was calculated using equation 2.13 with the time period T, the position of C_α atom $\mathbf{r}_i(t_j)$ of residue i at the time t_j and the reference position $\mathbf{r}_i^{ref}$, which is the time-averaged position of particle i.

The RMSF provides information about the flexibility of each residue and therefore more or less flexible areas of the protein become visible. A residue is considered flexible, if its RMSF is larger than 2 Å.

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t_i=1}^{T}(|\mathbf{r}_i(t_j) - \mathbf{r}_i^{ref}|)^2} \tag{2.13}$$

Comparing the RMSF of the mutant with the wild type can indicate whether or to what extent a substitution affects the flexibility of its environment.

Both RMSD and RMSF were calculated without considering overall translation and rotation by prior least square fitting of the trajectory frames to the corresponding initial structure.

Rgyr. The radius of gyration (Rgyr) is a measure of the compactness of a protein. It is described by the root means square distance of the structure to its center of mass. The calculation was done with C_α atoms according to equation 2.14 with the mass m_i of atom i and its position $\mathbf{r}_i$ at time t.

$$Rgyr(t) = \sqrt{\frac{\sum_i |\mathbf{r}_i(t)|^2 m_i}{\sum_i m_i}} \tag{2.14}$$

The Rgyr was determined for the enzyme mutants in order to detect a potential increase or loss of compactness compared to the wild type, whereby the latter can be associated with unfolding.

2.1.3.6 Hydration shell

RDF. The radial distribution function (RDF) with the symbol $g(\mathbf{r})$ gives the probability of finding a particle at a distance r from a reference. The $g_{AB}(\mathbf{r})$ between particles of type A and B is defined by equation 2.15 with $\langle\rho_B(\mathbf{r})\rangle$ as the particle density of type B

at a distance r around particles A and $\langle \rho_B \rangle_{local}$ being the particle density of B averaged over all spheres around particles A with radius r_{max}. The distance $\mathbf{r}$ is checked between all particles of type A (N_A) and type B (N_B). If $\mathbf{r}_{ij}$, i.e., the distance between particle i of type A and particle j of type B matches $\mathbf{r}$, then $\delta(\mathbf{r}_{ij} - \mathbf{r})$ is 1, otherwise it is 0.

$$\begin{aligned} g_{AB}(\mathbf{r}) &= \frac{\langle \rho_B(\mathbf{r}) \rangle}{\langle \rho_B \rangle_{local}} \\ &= \frac{1}{\langle \rho_B \rangle_{local} N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(\mathbf{r}_{ij} - \mathbf{r})}{4\pi \mathbf{r}^2} \end{aligned} \tag{2.15}$$

To determine whether the arrangement of solvent molecules in the vicinity of the protein is altered upon substitution, the RDF of water or ACN was determined for the nearest position of a substitution or the total protein. In addition, the cumulative radial distribution function (cumRDF) of ACN and water was determined within 2.0 Å and 3.0 Å, respectively.

H-bonds. Hydrogen bonds (H-bonds) were computed based on cutoffs for the hydrogen–donor–acceptor angle ($\alpha \leq 30°$) and the donor–acceptor distance ($r \leq 3.5$ Å). Hydroxy and amine groups are regarded as donors, oxygen and nitrogen atoms are acceptors. The number of H-bonds between substitutions and water was determined in order to detect differences between the hydration shell of wild type and mutant.

2.2 Experimental Section

2.2.1 Software

The software that was applied in the experimental part of this study is listed in table 2.4.

Table 2.4: Software and its application in the experimental part of this study.

SOFTWARE	APPLICATION
CloneManager Professionell Version 9 (Sci Ed Software)	Design of mutagenic primers and generation of multi-mutant genes
CurveExpert Professional (Hyams Development)	Curve fitting
PR.ThermControl (NanoTemper Technologies)	Inflection point calculation of thermal melting curves

2.2.2 Technical devices

Technical devices used in this study and their maufacturers are shown in table 2.5.

Table 2.5: Devices used in this study and their manufacturers.

DEVICE	MANUFACTURER
Biometra TOne	Analytik Jena GmbH
Centrifuge 5420	Eppendorf AG
Advanced Fluorescence and ECL imager	INTAS Science Imaging Instruments GmbH
Gel IX imager	INTAS Science Imaging Instruments GmbH
Mini-PROTEAN Tetra electrophoresis system	Bio-Rad Laboratories Inc.
Multitron incubator shaker	Infors AG
NanoDrop 2000c Spectrophotometer	ThermoFisher Scientific Inc.
Prometheus NT.Plex nanoDSF	NanoTemper Technologies GmbH
Rotina 35 R centrifuge	Andreas Hettich GmbH & Co. KG
Sonopuls UW 2070 Ultrasonic Homogenizer	BANDELIN electronic GmbH & Co. KG
Sorvall RC 6+ Centrifuge	ThermoFisher Scientific Inc.
SpectraMax 250 Microplate Reader	Molecular Devices LLC
GENESYS 6 Spectrophotometer	ThermoFisher Scientific Inc.
ThermoMixer C	Eppendorf AG
Wide Mini-Sub Cell GT electrophoresis system	Bio-Rad Laboratories Inc.

2.2.3 Consumables

Common consumables, as reaction tubes and pipetting tips, were purchased from Th. Geyer GmbH & Co. KG, VWR International LLC. and RATIOLAB GMBH. Special consumables including purification kits are listed in table 2.6.

Table 2.6: Special consumables used in this study and their manufacturers.

CONSUMABLES	MANUFACTURER
innuPREP Plasmid Mini Kit 2.0	Analytik Jena GmbH
Microtitration plates, 96-well	BRAND GmbH & Co. KG

NucleoSpin Gel and PCR Clean-up	MACHEREY-NAGEL GmbH & Co. KG
PCR plate, 96-well	BRAND GmbH & Co. KG
PD-10 columns	GE Healthcare GmbH
PVDF membrane	Bio-Rad Laboratories Inc.
Sealing film, microtest plates, aluminium	Carl Roth GmbH
Standard 24-Capillary Chips	NanoTemper Technologies GmbH
Syringe filters, sterile, 0.2 µmμmμm	VWR International LLC.
Vivaspin 20 Centrifugal Concentrator 10.000 MWCO	Sartorius AG

2.2.4 Chemicals and biomolecular products

All chemicals used in this study were purchased from Carl Roth GmbH & Co. KG, Bio-Rad Laboratories Inc., ThermoFisher Scientific Inc., Merck KGaA, NIPPON Genetics Europe and Th. Geyer GmbH & Co. KG.

Biomolecular products including restriction enzymes, ligases, DNA polymerases, antibodies and DNA or protein size standards were purchased from ThermoFisher Scientific Inc. and New England Biolabs.

2.2.5 Buffers and solutions

Buffers and solutions were prepared with deionized water (table 2.7).

Table 2.7: Composition of buffers and solutions.

BUFFER	COMPONENTS	
blotting buffer	10 mM	$NaHCO_3$
	3 mM	Na_2CO_3
	20 %	methanol
Coomassie staining solution	0.1 %	Coomassie Brilliant Blue R250
	10 % (w/v)	ammonium sulfate
	1.2 % (v/v)	phosphoric acid (85 %)
elution buffer (pH 7.4)	20 mM	Na_2HPO_4
	500 mM	NaCl
	500 mM	imidazol

FB buffer (pH 6.4)	100 mM	KCl
	67.5 mM	$CaCl_2$
	10 mM	potassium acetate
	10 % (w/v)	glycerol
lysis buffer (pH 7.4)	20 mM	Na_2HPO_4
	500 mM	NaCl
	10 mM	imidazol
100 mM KPi buffer (pH 7.2)	61.6 mM	K_2PO_4
	38.4 mM	KH_2PO_4
	100 mM	NaCl
10x SDS running buffer	25 mM	TRIS-HCl pH 8.8
	192 mM	glycine
	0.1 % (w/v)	SDS
2x SDS sample buffer	65.8 mM	TRIS-HCl pH 6.8
	0.01 %	bromphenol blue
	26.3 % (v/v)	glycerol
	2.1 % (w/v)	SDS
	355 mM	2-mercaptoethanol
12 % SDS-polyacrylamide separating gel (2x gels)	4 ml	acrylamide/bisacrylamide
	2.5 ml	1.5 M TRIS-HCl pH 8.8
	3.4 ml	H_2O
	0.1 ml	10 % (w/v) SDS
	0.1 ml	10 % (w/v) APS
	0.01 ml	TEMED
5 % SDS-polyacrylamide stacking gel (2x gels)	0.83 ml	acrylamide/bisacrylamide
	1.3 ml	0.5 M TRIS-HCl pH 6.8
	2.8 ml	H_2O
	0.05 ml	10 % (w/v) SDS
	0.05 ml	10 % (w/v) APS
	0.01 ml	TEMED
TBS-T buffer	25 mM	TRIS
	150 mM	NaCl
	0.1 %	Tween-20

20 mM TRIS buffer (pH 8.0)	8.7 mM	TRIS base
	11.3 mM	TRIS-HCl
	100 mM	NaCl

2.2.6 Cultivation media

For the cultivation of *Escherichia coli* (*E. coli*), lysogeny broth (LB) medium, LB agar and auto induction (AI) medium were prepared as described in table 2.8 and filled to the desired volume with destilled water. The media were autoclaved to ensure sterile conditions during cell culturing. The components marked by an asterisk (*) were separately sterile-filtered and added to the AI medium before use.

Table 2.8: Composition of cultivation media.

MEDIA	COMPONENTS	
AI medium	20 g/l	tryptone from casein
	5 g/l	NaCl
	5 g/l	yeast extract
	6 g/l	Na_2HPO_4
	3 g/l	KH_2PO_4
	0.6 % (v/v)	glycerol*
	0.2 % (w/v)	lactose*
	0.05 % (w/v)	glucose*
LB agar	25 g/l	LB medium powder
	15 g/l	agar-agar
LB medium	25 g/l	LB medium powder

2.2.7 Oligonucleotides

Mutagenic primer pairs for site directed mutagenesis were designed using Clone Manager (Sci Ed Software). The sequences of each primer pair comprising the substitutions for single and triple mutants are listed in table 2.9.

Table 2.9: Oligonucleotides used for site directed mutagenesis. The complementary primer sequences with substitutions for ED30 are listed at the top, for PT35 at the bottom. The forward and reverse primers are shaded in gray and white, respectively.

SUBSTITUTION	SEQUENCE (5' → 3')
K124D	CTCGGGCGATAACTTGGCGACGGTCGGCGATGTC
	CCAAGTTATCGCCCGAGTAGAGCCCGATGCTGC
R211E	CAAAAGCGGAAGGGCTGTTCCACAAAGCGATCTC
	GAACAGCCCTTCCGCTTTTGGGCTTGCGAGCAAAC
K337E	CACGATGGACGAAGACGAACTTGCGCGCTGGGCCG
	CGGCCCAGCGCGCAAGTTCGTCTTCGTCCATCGTG
E199R	GTCTCGCCCCGGGCTCGGGAGGCAGATCTGTTC
	GAACAGATCTGCCTCCCGAGCCCGGGGCGAGAC
E328K	CGTGAGCGCGTTGCCAAATTCTTCCAGCCCTAC
	GTAGGGCTGGAAGAATTTGGCAACGCGCTCACG
D439K	CAGCGCTATACCGGCAAACCGGATGACCTGAAC
	GTTCAGGTCATCCGGTTTGCCGGTATAGCGCTG

The oligonucleotides used for sequence- and ligation-independent cloning (SLIC) were provided by R. Molitor (table 2.10).

Table 2.10: Oligonucleotides used for SLIC. The forward primer is shaded in gray, the reverse primer in white.

SEQUENCE (5' → 3')
GGCGATGGCCATGGATTGCCTGTCAGGTGGCGGCGGTGG
AGTGGTGGTGGTGGTGGTGCTCGAGGGGCAATACCAGCGGCGGGTCA

2.2.8 Strains and plasmids

Chemically competent *E. coli* DH5α cells[198] and *E. coli* BL21(DE3) cells[199] were used for plasmid replication and enzyme expression, repectively.

Previous work by A. Bollinger involved the cloning of the ED30 and PT35 genes into the pET22b(+) vector (Merck KGaA), with the C-terminus flanking the vector-encoded hexa histidine tag, yielding the recombinant plasmids pET22b_ED30$_{c6H}$ and pET22b_PT35$_{c6H}$. An additional plasmid pET22b_PelB-PT35$_{c6H}$ was obtained by replacing the native signal peptide with the N-terminal pelB signal sequence for periplasmic

localization, provided by the pET22b(+) vector.

To generate single and triple mutants, the plasmids pET22b_ED30$_{c6H}$ and pET22b_PelB-PT35$_{c6H}$ were used for site-directed mutagenesis. The genes of ED30 and PT35 multi-mutants were synthesized by BioCat GmbH (https://www.biocat.com/gene-synthesis) and provided in pUC57 vectors. For enzyme production, the multi-mutant genes were cloned into the pET22b(+) vector beforehand.

2.2.9 Transformation of chemically competent cells

To prepare chemically competent cells of *E. coli* BL21(DE3) and *E. coli* DH5α, a pre-culture was prepared using LB medium and bacterial cells from a cryo stock, and cultivated overnight at 37 °C with shaking (130 rpm). In 5 L Erlenmeyer flasks, 500 ml LB medium was inoculated with the pre-culture to an optical density (OD) of 0.05 (λ = 580 nm) and incubated (37 °C, 130 rpm) until an OD of around 0.5 was reached. After incubation for 10 min on ice, the cells were centrifuged for 10 min at 4 °C and 2,800 rpm. During the following washing steps the cells were constantly chilled on ice. After discarding the supernatant, the pellet was resuspended in 35 ml FB buffer. After 20 min incubation, the cells were centrifuged again (10 min, 4 °C, 2,800 rpm) and the supernatant was discarded. The cells were resuspended in 8.5 ml FB buffer and stored in 100 µl aliquots at –80 °C.

Transformation of *E. coli* BL21(DE3) and *E. coli* DH5α was performed according to the heat shock method.[200] About 100 ng plasmid DNA were usually added to an aliquot of chemically competent cells. After incubating the cells on ice for 30 min, the heat shock was performed for 60 sec at 42 °C. 700 µl LB medium supplemented with 0.5 % glucose was added and the cells were incubated at 37 °C for 1 h according to the recovery time recommended for ampicillin resistance. The cells were then either plated on LB agar or cultivated in LB medium, both supplemented with 100 μg/ml ampicillin (Amp100), and incubated overnight at 37 °C.

2.2.10 DNA manipulation

Site directed mutagenesis

Site directed mutagenesis was performed using QuikChange (QC) polymerase chain reaction (PCR) as described before by Edelheit *et. al.*[201] in order to introduce single and triple substitutions to ED30 and PT35. The QC PCR was performed using about 100 ng template DNA (pET22b_ED30$_{c6H}$ or pET22b_PelB-PT35$_{c6H}$), 0.5 µM each of the mutagenic primers listed in table 2.9 and the High-Fidelity DNA Polymerase (ThermoFisher

Scientific) according to the supplier's specifications. Thermal cycling was carried out with the Biometra TOne cycler (Analytic Jena). To avoid annealing of complementary primer pairs, the first 5 cycles were performed in two separate reactions containing either forward or reverse primer, followed by additional 18 cycles with both reactions combined.

To remove non-modified template DNA from the QC PCR product, a *DpnI* (ThermoFisher Scientific) digestion was carried out at 37 °C for 2 h, followed by inactivation at 80 °C for 20 min. Chemically competent *E. coli* DH5α cells were subsequently transformed with 10 µl of the QC PCR product for plasmid DNA amplification and verification (section 2.2.10). To generate triple mutants, the QC PCR was performed in a three-step procedure, using the modified plasmid DNA from the previous step as template in the following step.

Cloning of the multi-mutant plasmids

For the cloning of the ED30-mmut gene into the pET22b(+) vector, 500 ng pUC57_ED30-mmut vector and 500 ng pET22b(+) were cut using the restriction enzymes *NdeI* (ThermoFisher Scientific) and *XhoI* (ThermoFisher Scientific). The 20 µl mixtures were incubated at 37 °C for 1 h and the enzymes were then inactivated at 80 °C for 20 min. The resulting DNA fragments were separated by agarose gel electrophoresis (section 2.2.10). The pET22b(+) backbone and ED30-mmut insert were purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). The backbone and insert were ligated at a 1:5 ratio using the T4 DNA Ligase (ThermoFisher Scientific) in a total volume of 20 µl. The ligation reaction was incubated for 1 h at 22 °C and inactivation was performed for 10 min at 65 °C. Subsequently, chemically competent *E. coli* DH5α cells were transformed with 10 µl ligation product for plasmid DNA amplification and verification (section 2.2.10).

The cloning of pET22b_PT35-mmut was carried out using the SLIC method.[202] At first, a PCR[203] was performed using about 100 ng pUC57_PT35-mmut as template DNA, 0.5 µM each of the SLIC-primers (table 2.10) and the Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific) according to the supplier's specifications.At the same time, 500 ng of the pET22b(+) vector was cut with the restriction enzymes *NcoI* (ThermoFisher Scientific) and *XhoI* (ThermoFisher Scientific) in a total volume of 20 µl . The enzymes were incubated at 37 °C for 1 h and subsequently inactivated at 80 °C for 20 min. The pET22b(+) backbone and the PT35-mmut insert were separated by agarose gel electrophoresis (section 2.2.10) and subsequently purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). The SLIC reaction contained the purified pET22b(+) backbone and the ED30-mmut insert at a 1:2 ratio, 0.5 µl T4-Polymerase (New England Biolabs), 1 µl 10x BSA and 1 µl NEB buffer 2 in a total volume of 10 µl. The T4-Polymerase was treated at room temperature for 2.5 min to generate 3' overhangs.

Chemically competent *E. coli* DH5α cells were immediately transformed with 5 µl for plasmid DNA amplification and verification (section 2.2.10), after incubating the SLIC reaction at 10 °C for 10 min.

Agarose gel electrophoresis

DNA fragments were separated by agarose gel electrophoresis, using a 1 % agarose gel supplemented with 10 µl/l midori green for DNA visualization. The agarose gel was placed into the Wide Mini-Sub cell GT electrophoresis chamber (Bio-Rad) filled with 0.5x TBE buffer and the samples were loaded onto the gel after mixing them with 6x TrickTrack DNA loading dye (ThermoFisher Scientific). The 1 kb GeneRuler (ThermoFisher Scientific) was used as DNA standard. After separation for 30 min at 120 V, gel documentation was done using the Gel IX imager (INTAS).

Plasmid DNA amplification and verification

After transformation of chemically competent *E. coli* DH5α (section 2.2.9), the cells were cultivated on Amp100 containing LB agar plates overnight at 37 °C. Single bacterial clones were used to inoculate LB medium supplemented with Amp100. After overnight incubation at 37 °C, the plasmid DNA was isolated using the innuPREP Plasmid Mini Kit 2.0 (Analytic Jena) according to the manufacturer's specifications. The DNA concentration was determined using the NanoDrop Spectrophotometer (ThermoFisher Scientific). Site-directed mutations and the cloned pET22b_ED30-mmut and pET22b_PT35-mmut plasmids were verified by Sanger sequencing (LGC genomics GmbH). Prior to that, the ligation of pET22b_ED30-mmut was additionally verified by a test restriction with *NdeI* and *XhoI* and subsequent visualization by agarose gel electrophoresis (2.2.10).

2.2.11 Enzyme production and purification

Recombinant enzyme production

Recombinant proteins were produced as previously described by Studier *et al.*[204] using AI. At first, pre-cultures were prepared in Erlenmeyer flasks filled to 1/10 with LB medium, additionally supplemented with 0.5 % glucose, and inoculated with prior transformed *E. coli* BL21(DE3) cells. Afterwards, 500 mL AI medium supplemented with Amp100 was inoculated with the pre-culture to an OD of 0.05 (λ = 580 nm). The 5 L Erlenmeyer flasks filled to 1/10 with media for sufficent oxygen supply were shaken (160 rpm) at 25 °C for 24 h using the Multitron incubator shaker (Infors). The cells were harvested by centrifugation for 30 min (4 °C, 6000 × g). After discarding the supernatant, the cell pellets were stored at –20 °C if not used immediately.

Enzyme purification

The immobilized metal ion affinity chromatography (IMAC) was performed to purify the enzymes using nickel nitrilotriacetic acid (Ni-NTA) agarose. At first, the cell pellets were resuspended in lysis buffer resulting in a 10 – 15 % (w/v) suspension followed by cell disruption using the Sonopuls Ultrasonic sonotrode (BANDELIN). Ultrasonic waves were emitted twice for 5 min with 30 % power and a 50 % cycle. Cell debris and insoluble aggregates were removed by centrifugation (30 min, 4 °C, 36,000 × g). After washing and equilibrating the Ni-NTA agarose (ThermoFisher Scientific) with water and lysis buffer, respectively, 6 ml Ni-NTA matrix per liter of culture were mixed with the supernatant comprising the soluble proteins and incubated for 30 min at 5 °C. The matrix was transferred to a gravity flow column and washed with at least 10 column volumes (CV) of lysis buffer followed by protein elution with 3 CV of elution buffer. The elution fraction was concentrated to about 2.5 ml by centrifugal ultrafiltration (4 °C, 3,600 rpm) using Vivaspin 20 filters (Sartorius). PD-10 desalting columns (GE Healthcare) were used following the manufacturer's instructions to desalt the proteins with 20 mM TRIS buffer (pH 8.0) or 100 mM potassium phosphate (KPi) buffer (pH 7.2). The purified proteins were stored at –20 °C.

Determination of enzyme concentration

Protein concentrations were determined using the NanoDrop Spectrophotometer (ThermoFisher Scientific). The ProtParam web service[205] was used to calculate the molecular weight ($ED30_{c6H}$: 55,030 Da, $PT35_{c6H}$: 56,937 Da) and the extinction coefficient ($ED30_{c6H}$: 86,315 M^{-1} cm^{-1}, $PT35_{c6H}$: 69,705 M^{-1} cm^{-1}) which are required to calculate the protein concentration with the Protein A_{280} method.

Evaluation of enzyme purification

To estimate the purity of the prepared enzymes, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli[206]. Protein samples from different purification steps were mixed with SDS sample buffer and boiled for 10 min at 98 °C. The PageRuler PrestainedProtein Ladder (ThermoFisher Scientific) and the prepared protein samples were loaded on a polyacrylamide gel composed of a 12 % separating gel and a 5 % stacking gel. The gel was placed into the Mini-PROTEAN electrophoresis chamber (Bio-Rad) and SDS running buffer was added. After separation for 15 min at 100 V followed by 40 min at 200 V, the gel was stained using the Coomassie staining solution. Gel documentation was done using the Advanced Fluorescence and ECL imager (INTAS).

Immunological verification of PT35 and ED30

The proteins, previously separated by SDS-PAGE, were transferred to a PVDF membrane (Bio-Rad) for immunological detection. The membrane was first equilibrated in methanol for 5 min and in water for 2 min. The protein transfer was performed for 15 min at 150 mA and 30 min at 300 mA on ice using the Mini-PROTEAN electrophoresis chamber (Bio-Rad) filled with blotting buffer. For blocking non-specific binding sites, the membrane was subsequently incubated in a 3 % skim milk powder solution overnight at 4 °C. After washing the membrane with TBS-T buffer for 2 min, the membrane was incubated with the Novex Anti-His (C-Term)-HRP Antibody (ThermoFisher Scientific) diluted 1:50,000 in TBS-T buffer for 1 h. The membrane was then again washed three times for 15 min in TBS-T buffer. For antibody detection via chemiluminescence, the Pierce ECL Western Blotting Substrate (ThermoFisher Scientific) was applied according to the manufacturer's instructions and the Advanced Fluorescence and ECL imager (INTAS) was used for visualization.

2.2.12 Determination of enzyme stability

Organic solvent stability assay

The effect of ACN on the esterase activity was determined by mixing purified enzyme in 20 mM TRIS buffer (pH 8.0) or 100 mM KPi buffer (pH 7.2) with ACN as described in table 2.11 followed by incubation at 30 °C with shaking (750 rpm) using the ThermoMixer C (Eppendorf). 10 µl samples were taken after 0, 1, 2, 3 and 24 h and then daily until the activity was reduced to about 50 %. The esterase activity was measured directly after sampling as described below.

Table 2.11: Preparation of the esterase activity assay. Mixtures containing 0, 5, 10 and 20 % ACN for ED30 and 0, 20, 50 and 80 % ACN for PT35 variants were prepared using 20 mM TRIS buffer (pH 8.0) or 100 mM KPi buffer (pH 7.2).

VARIANT	ACN (%)	ENZYME (µl)	BUFFER (µl)	ACN (µl)
ED30	**0**	10	990	0
	5	10	940	50
	10	10	890	100
	20	10	790	200
PT35	**0**	10	990	0
	20	10	790	200
	50	10	490	500
	80	10	190	800

Determination of the half-inactivation temperature

The thermostability was determined for the wild type ED30 and PT35 after incubation of the enzymes at 30 to 50 °C and 40 to 80 °C, respectively. 50 µl enzyme solution per well were added to a 96-well PCR plate (BRAND). The plate was covered with an aluminium coated sealing film (Carl Roth) and incubated for 1 h using the gradient function of the Biometra TOne (Analytik Jena). The esterase activity was subsequently measured as described below.

Esterase activity assay

Colorimetric determination of the esterase activity was carried out by the hydrolysis of the substrate 4-nitrophenyl butyrate (*p*NPB). The colorless substrate is thereby converted into the yellow 4-nitrophenol which can be detected at a wavelength of $\lambda = 410$ nm. The substrate solution was prepared right before measurement by mixing 20 mM *p*NPB in ACN with 20 mM TRIS buffer (pH 8.0) or 100 mM KPi buffer (pH 7.2) at a 1:20 ratio. The measurements were done by combining 10 µl enzyme solution with 190 µl substrate solution in a flat bottom 96-well microtiter plate. The reaction was followed in the SpectraMax microplate reader (Molecular Devices) at 30 °C with $\lambda = 410$ nm for 10 min. The obtained reaction velocities were corrected by a control reaction without enzyme.

Determination of the protein thermal melting point

The effect of ACN on the protein thermal melting point was determined by nano differential scanning fluorimetry (nanoDSF) using the Prometheus NT.Plex device (NanoTemper). NanoDSF is based on measuring the tryptophan and tyrosin fluorescence to monitor protein unfolding. While tryptohan shows a fluorescence maximum at 350 nm upon contact with the polar solvent, the fluorescence intensity of the tyrosin at 330 nm is independent of its close surroundings. Therefore, the ratio of the fluorescence intensities at 350 nm and 330 nm is suitable to detect structural changes, e.g. due to protein unfolding.

Shortly before measurement, purified enzyme was combined with ACN in 20 mM TRIS buffer (pH 8.0) or 100 mM KPi buffer (pH 7.2) to a final concentration of 0, 5, 20 or 50 % ACN. The capillary chips were loaded and placed into the Prometheus device to scan the protein melting from 20 to 95 °C at a heating rate of 1 °C per minute. The PR.ThermControl software (NanoTemper) was used to analyze the melting scan.

3 Results and discussion

The hydration of a protein is suggested to play a key role in preventing organic solvents from entering and disrupting its hydrophobic core, which would cause unfolding.[111,122] Negatively charged residues are known for their strong water-binding properties.[129,130] Therefore, the highly organic solvent tolerant esterase PT35, which possesses an excess of negatively charged residues, represents the optimal candidate to test this hypothesis in a mutagenesis study.[136] The mutagenesis strategy was based on the comparison between PT35 and a related but much less stable esterase ED30, which was used as a counterpart to PT35.[140] A particular feature of PT35, which strongly distinguishes PT35 from ED30, is its extremely negatively charged surface, which was assumed to contribute to its stability in the presence of organic solvents.[139] Therefore, this mutagenesis strategy focused on the hydration shell and surface charge of the enzyme with the aim to generate a more stable ED30 and a less stable PT35.

The mutagenesis approach involved 4 main steps, [1] the selection and combination of substitutions according to rational criteria (section 3.1), [2] the analysis of the generated triple and multi-mutants using comprehensive molecular dynamics (MD) simulations in water and organic solvents (section 3.2), [3] development of an experimental test system for organic solvent stability (section 3.3) and [4] the investigation of the mutants using the experimental test system (section 3.4).

3.1 Selection and combination of substitutions

The first step of the mutagenesis strategy involved screening of residues with substitution potential. Identification of these mutation candidates was based on comparison of the two enzymes, using ED30 as a template for PT35 and vice versa. Since both enzymes have approximately the same sequence length (514 and 507 residues), their total number of charged, polar and nonpolar residues can be directly compared (figure 3.1). Here, H is not considered a positively charged residue, since it is predominantly uncharged at pH $>$ 6.0. PT35 possesses an excess of 68 negatively charged residues (D: 41, E: 27) compared to 20 positively charged residues (R: 14, K: 6). In contrast, ED30 contains a

similar number of 53 negatively charged (D: 25, E: 28) but 47 positively charged (R: 31, K: 16) amino acids.

Surprisingly, the charge difference between PT35 and ED30 is less influenced by the fact that PT35 possesses more negatively charged amino acids than ED30 (+15), but rather contains a considerably lower number of positively charged residues (–27). Moreover, the number of polar residues, in particular N (+11), Q (+12) and Y (+11), are markedly increased in PT35 compared to ED30.

This information was used to establish the following substitution principles: (i) positively charged residues in ED30 were substituted with negatively charged residues or polar residues [ED30: positive (R, K) → negative (D, E) / polar], (ii) polar and negatively charged residues in PT35 were replaced by positively charged residues [PT35: polar/negative → positive] and, in rarer cases, (iii) polar residues were exchanged for negatively charged residues in ED30 and *vice versa* in PT35 [polar ↔ negative]. These principles allowed a charge alteration of at least +1/-1 with each substitution. Among the polar residues, N and Q, which are structurally very similar to E and D, and exceptionally S were used for substitutions. Although its content is higher in PT35 than in ED30, the large aromatic amino acid Y was excluded from the substitution process to keep structural changes due to steric clashes low.

The identification of substitution candidates was based on the EPS of the two enzymes (figure 3.1). The EPS was calculated and visualized using the PyMOL plugin APBS[142,195]. Considering the EPS with a range of −10 to +10 kT/e allowed a more accurate location of the most dominant surface charges of the protein. In both enzymes, the negative charge was pronounced at the active site. A negative potential at the active site was also observed for other lipases and esterases, where it appeared to be correlated with high activity toward triglycerides due to electrostatically induced rapid product release.[207] However, while the surface of ED30 also exhibits positive potentials, the negative surface potential covers almost the entire surface of PT35. The distribution and intensity of potentials were compared between both enzymes by structural alignment. Those areas with the largest surface potential differences between ED30 and PT35, referred to as surface potential *hot spots*, were screened for mutation candidates.

To identify compatible substitutions for the mutation candidates, a structural alignment of ED30 and PT35 was performed. Based on the substitution principles, the aligned structure of the other enzyme was screened for suitable residues in the close proximity (within 5 Å) to the mutation site in the wild type. For example, K99 was considered a suitable mutation candidate of ED30. The structural alignment revealed that PT35 possesses a D93 in close proximity to K99. Therefore, K99D was found a compatible substitution for ED30. Rational criteria were established to determine the suitability of the mutation

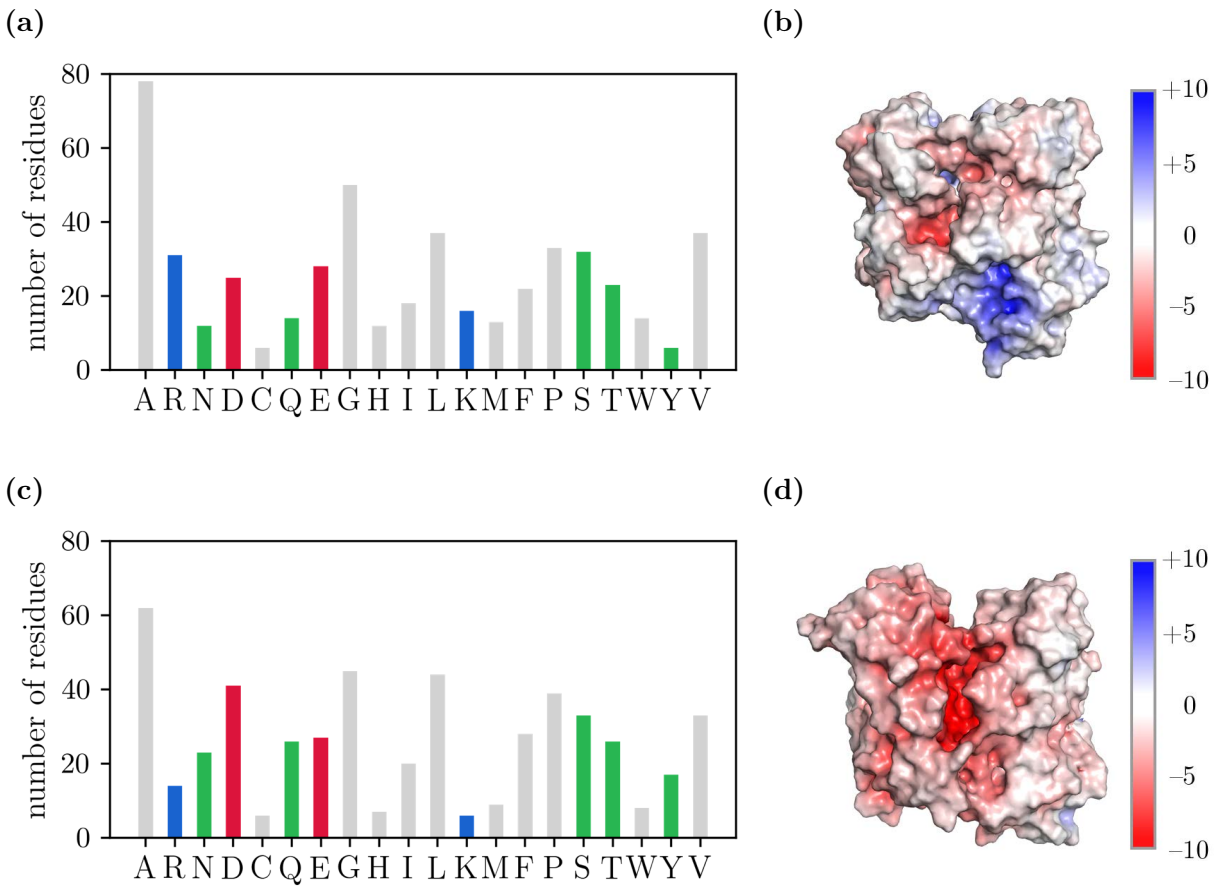

Figure 3.1: The amino acid composition of ED30 and PT35. The number of each type of amino acid contained in (a) ED30 and (c) PT35 is shown as a barplot. Positively charged (blue), negatively charged (red), polar (green) and nonpolar (gray) residues are color-coded. The electrostatic potential surface (EPS) of (b) ED30 and (d) PT35 was calculated using the PyMOL plugin APBS[142,195]. The EPS ranges from negatively charged (red, −10) to uncharged (white, 0) to positively charged (blue, +10) and is given in units of kT/e.

candidates and evaluate the generated single mutants for subsequent combination of the most favorable substitutions.

Residue conservation. Conserved residues are often important to maintain enzyme function and conformation.[60,208] To avoid substitution of conserved residues, the conservation score (CS) was determined for each residue of ED30 and PT35 using the ConSurf server[188,189]. The CS ranges between 1 (variable) and 9 (conserved). Residues with a CS greater than 7 were excluded from the selection process of mutation candidates.

Solvent accessible surface area. The wild-type residues were also evaluated based on their relative solvent-accessible surface area (rSASA), which describes the solvent-accessible surface area (SASA) of an amino acid relative to their theoretical maximum

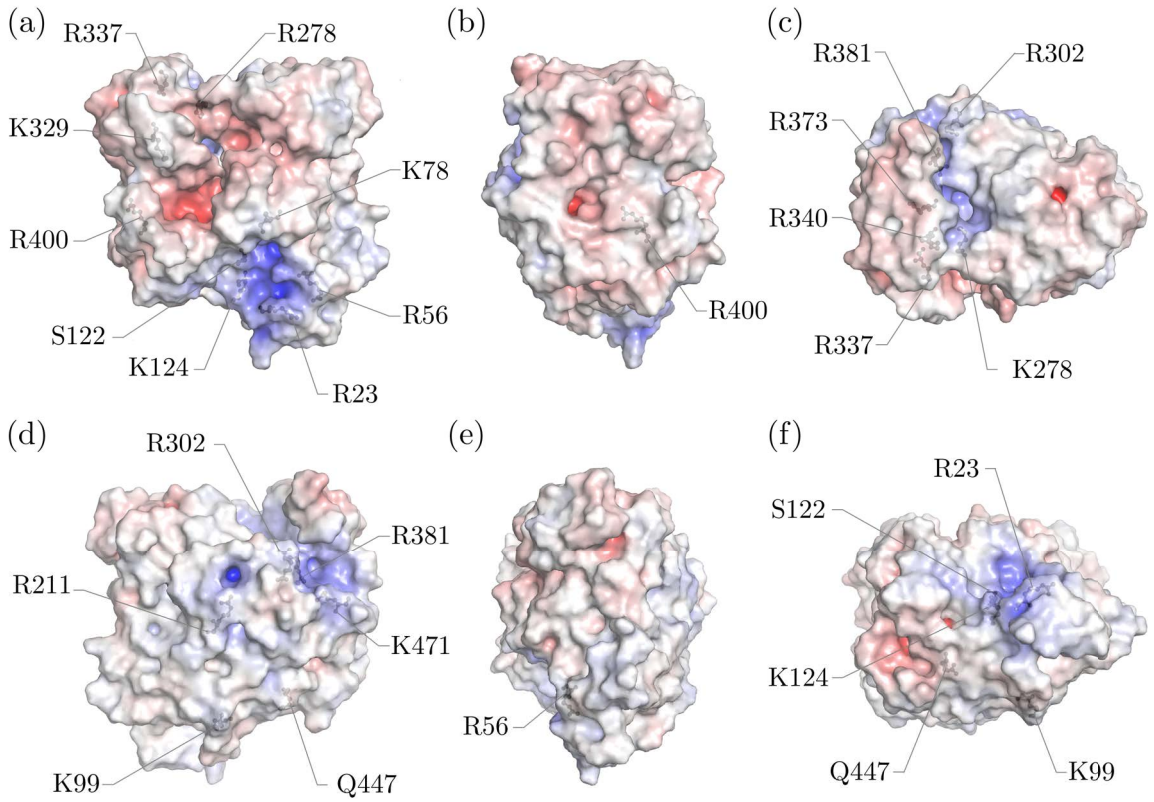

Figure 3.2: Positions of substitutions in the wild-type ED30. The wild-type residues substituted to generate 17 single mutants are shown at the electrostatic potential surface (EPS) of ED30 as black ball and stick representations. The (a) front view, the (d) back, the (b) left and (e) right side as well as the (c) top and (f) bottom view are shown. The front view is defined by the C-and N-Terminus on the upper left and lower right side (C-N), respectively. The EPS ranges from negatively charged (red, -10) to uncharged (white, 0) to positively charged (blue, +10) and is given in units of kT/e (see figure 3.1).

possible SASA. Residues with an rSASA $> 30\,\%$ were counted as being part of the protein surface, hence being *exposed*.[193,194] Since the interaction between surface residues and water molecules is assumed to be critical for organic solvent tolerance, particularly exposed residues are of interest for substitutions.[66,122]

Electrostatic potential surface. One of the most relevant criteria is the estimation of the surface-potential change upon mutation to test the hypothesis whether a negatively charged surface correlates with higher solvent tolerance. After substitution, the EPS of the generated single mutants was computed using the PyMOL plugin APBS[142,195] and compared to the wild type (ΔEPS). The substitutions were assessed based on their effect on the surface potential hot spots in the respective enzyme using values from 1 (low effect) to 3 (high effect). The hot spots in ED30 and PT35 were expected to show a large ΔEPS towards a more negative (ED30) and more positive (PT35) EPS, respectively. Substitu-

tions with the opposite or no detectable effect on the hot spots were excluded from the selection process.

After eliminating conserved residues and substitutions that were not affecting the surface potential hot spots, 17 single mutants of each of ED30 and PT35 were obtained. Two mutation candidates of PT35 were substituted twice (D268R and D268N, Q73K and Q73R). The 17 and 15 positions of the substitutions are shown at the EPS of the wild-type ED30 and PT35 in figure 3.2 and 3.3, respectively. To analyze the hydration shell and the structural stability of the single mutants, 100 ns MD simulations in water only were performed. On the basis of the MD simulations, the radial distribution function (RDF) of water, the number of hydrogen bonds (H-bonds) between substitution and water and the root mean square deviation (RMSD) were computed.

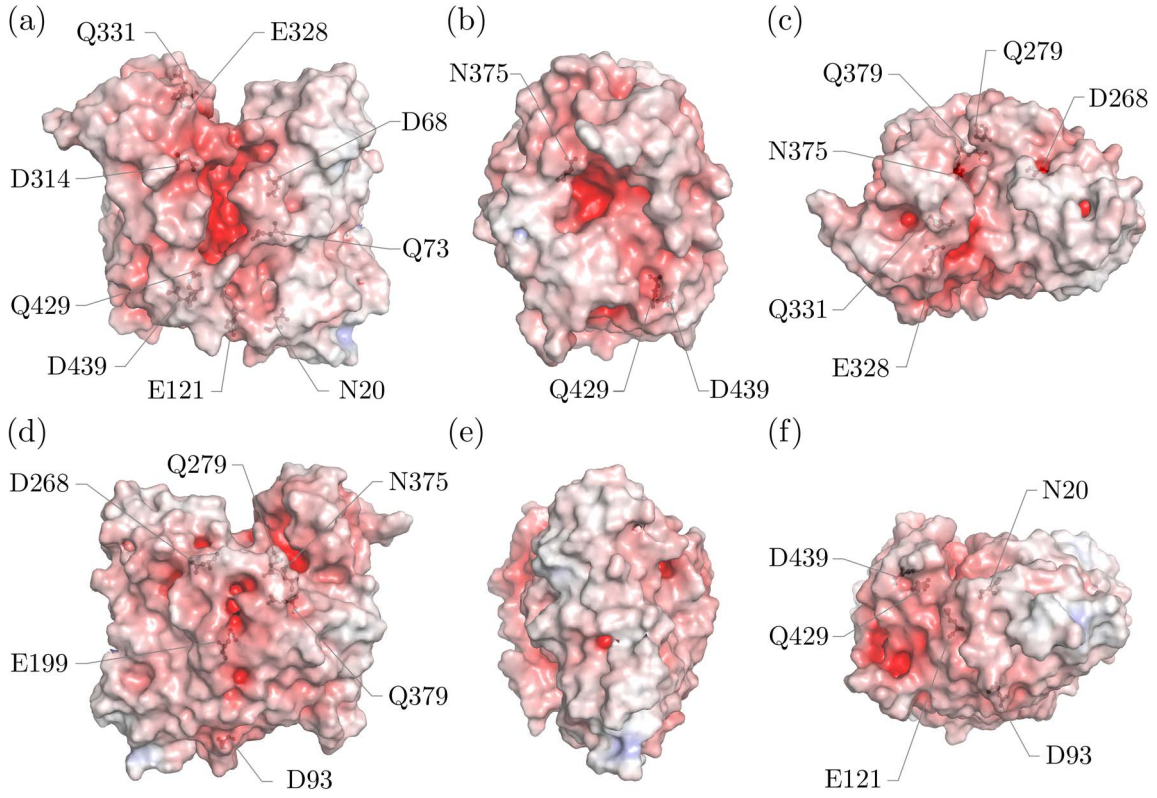

Figure 3.3: Positions of substitutions in the wild-type PT35. The 15 compatible wild-type residues substituted to generate 17 single mutants are shown at the electrostatic potential surface (EPS) of PT35 as black ball and stick representations. The (a) front view, the (d) back, the (b) left and (e) right side as well as the (c) top and (f) bottom view are shown. The front view is defined by the C-and N-Terminus on the upper left and lower right side (C-N), respectively. The EPS ranges from negatively charged (red, −10) to uncharged (white, 0) to positively charged (blue, +10) and is given in units of kT/e (see figure 3.1).

Hydrogen bonds and radial distribution function. The protein hydration shell is assumed to play a key role in the stability of an enzyme in organic solvents. The hydration of a protein or a specific residue can be estimated by the number of H-bonds with water or the radial distribution function (RDF) of water. The RDF of water describes the probability to locate water molecules at a certain distance, while the cumulative RDF (cumRDF) gives the number of water molecules within a certain distance. To assess the change in hydration upon substitution, the difference between wild type and single mutant with respect to the number of hydrogen bonds (ΔH-bonds) between the substitution and water, and the cumRDF (ΔRDF) of water within 2 Å of the substitution was determined. ED30 and PT35 substitutions were considered *favorable*, when the hydration shell was increased (ΔH-bonds > 0, ΔRDF > 0) and decreased (ΔH-bonds < 0, ΔRDF < 0), respectively.

Root mean square deviation. Serious structural changes due to substitution may impair the activity of the enzyme or even lead to full inactivation. Therefore, the RMSD of single mutants and wild types was computed and compared (ΔRMSD). A single mutant with a ΔRMSD > 3 Å was considered structurally different from the wild type, hence unfavorable.

Relative change in free energy. To estimate the change in thermodynamic stability upon mutation, the relative change in free energy ($\Delta\Delta G$) of folding was predicted using FoldX.[115,144] The interpretation of the obtained $\Delta\Delta G$ values was done according to the CompassR rule, postulated by Cui *et al.* through analysis of the correlation between $\Delta\Delta G$ and specific activity.[197] The CompassR rule was evolved as a selection guide for the combination of beneficial substitutions from directed evolution approaches. In this study, it was used as an additional criterion for combination of favorable substitutions. Referring to the CompassR rule, substitutions were categorized into 3 groups: (i) *potentially active* ($\Delta\Delta G \leq +0.36$ kcal mol^{-1}), (ii) *unpredictable* ($+0.36$ kcal mol$^{-1} < \Delta\Delta G < +7.52$ kcal mol^{-1}) and (iii) *potentially deactivating* ($\Delta\Delta G > +7.52$ kcal mol^{-1}). To keep the probability of inactive mutants low, substitutions from group A should be preferred for combination.

The evaluation according to these rational criteria is shown using a color scheme for each of the 17 selected substitutions of ED30 and PT35 (figure 3.4). The substitutions were categorized in groups, (A) positively charged residues $\leftrightarrow$ polar residues, (B) positively charged residues $\leftrightarrow$ negatively charged residues and (C) polar residues $\leftrightarrow$ negatively charged residues.

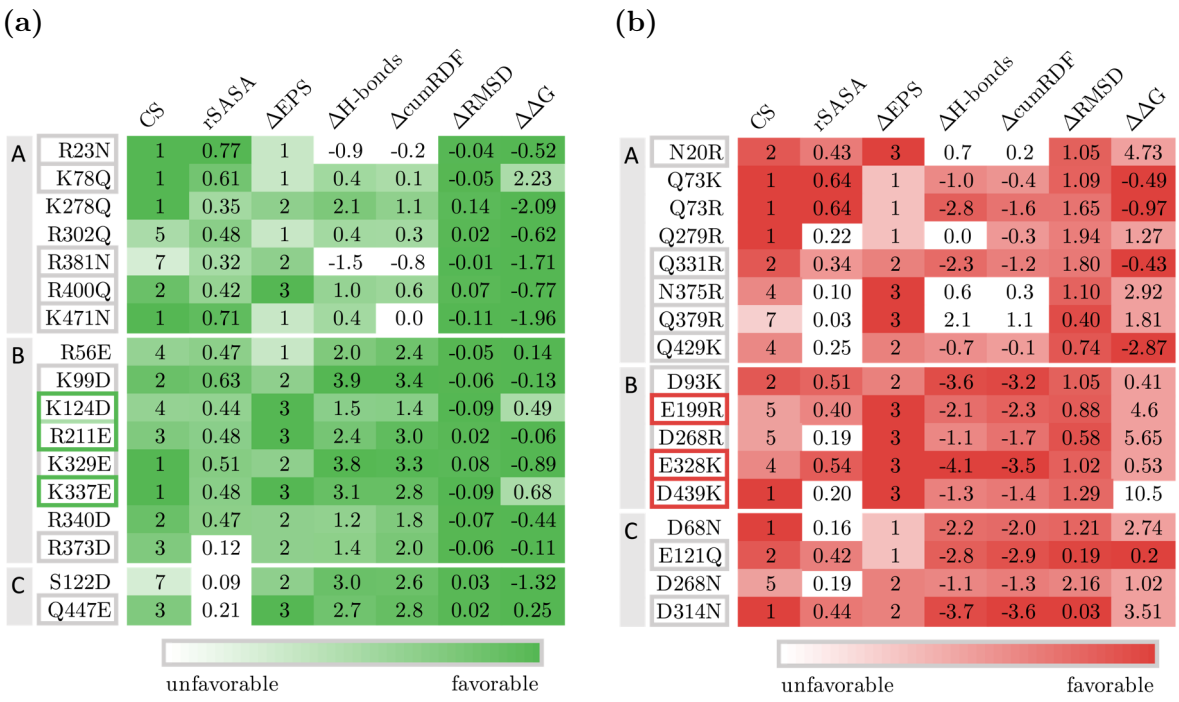

Figure 3.4: Assessment of substitutions by rational criteria. The 17 substitutions of each of (a) ED30 and (b) PT35 were assessed based on the following rational criteria. Wild-type residues were rated by their conservation (conservation score [CS]) and residue exposure (solvent accessible surface area [rSASA]). Substitutions in the single mutants were estimated based on their impact on the surface charge hot spots (electrostatic potential surface [ΔEPS]), prediction of thermodynamic stability (relative change of free energy [$\Delta\Delta G$]), change of hydration shell around the substitution (relative number of hydrogen bonds [ΔH-bonds], relative cumulative RDF at 2.0 Å [ΔRDF]) and structural stability (relative root mean square deviation [ΔRMSD]). The rating ranges from unfavorable (white) to favorable (green or red). Substitutions are categorized in the groups A–C according to their type of substitution. Substitutions combined to triple mutants are framed in green or red, multi-mutant substitutions are framed in gray.

The strongest differences in EPS with respect to surface charge hot spots (ΔEPS) were observed when negatively charged residues were inserted into ED30. In PT35, substitutions of negatively charged residues with positively charged residues accounted for the largest difference in charge, but also some exchanges of polar residues for positively charged residues. No correlation was observed between residue exposure (rSASA) and the resulting level of change in EPS. Instead, a linear correlation was found between the exposure of the wild type residues and the hydration shell of the substitution in the single mutants (figure A1). With respect to ED30, this correlation was positive for group A ($R^2 = 0.54$, but excluding R381N) and negative for group B ($R^2 = 0.89$, but excluding R373D). This indicates that the more exposed a positively charged residue was, the more enhanced was the hydration after substitution with a polar residue. In contrast, the greatest effects on the hydration were observed when less exposed positively charged residues

were substituted with negatively charged residues. In contrast, only negative correlations were observed for all groups of PT35 (A) $R^2 = 0.46$, but excluding N20R, (B) $R^2 = 0.95$, (C) $R^2 = 0.83$). The more exposed a residue was, the more reduced was the hydration shell after substitution with positively charged or polar residues.

A characteristic change of residue hydration was observed by certain substitutions (figure 3.5). The RDF of water ($g(r)$) is representatively shown for one ED30 and one PT35 substitution of each of the groups A–C up to a distance of 5 Å from the residue. This allows the distribution of water of the substitution to be estimated in the close proximity. The RDF of K278Q and Q73R were only slightly different between single mutation (smut) and wild type (wt). In contrast, the RDF of a negatively charged residue clearly differs from that of a positively charged or polar residue. Regardless of the enzyme, negatively charged residues showed two clearly separated RDF peaks at around 1.8 Å and 2.8 Å, which can be interpreted as two layers of the hydration shell. Polar and positively charged residues, instead, exhibited hardly separable, often merging peaks, with only a weak first peak at around 1.9 Å in many cases. In addition, the RDF of negatively charged residues was in general markedly higher at distances below 2.0 Å compared to positively charged and polar residues. This is in particular indicated by the increasing and decreasing RDF within 2.0 Å after introduction (K329E, Q447E) and substitution (D93K, D314N) of negatively charged residues, respectively. The enhanced hydration of negatively charged residues was, moreover, evident from the ΔH-bonds and ΔRDF values of group B and C substitutions, which were generally found to be increased and decreased for ED30 and PT35, respectively (figure 3.4). Therefore, it could be assumed that negatively charged residues facilitate a tighter water bonding and a hydration shell with two organized water layers. This observation agrees well with results from NMR studies that showed that hydration of negatively charged amino acids is higher than that of positively charged or polar residues.[129]

The interpretation of the $\Delta\Delta G$ values according to the CompassR rule revealed that most of the substitutions in ED30 were *potentially active* and only 3 were *unpredictable* (figure 3.4). In contrast, most of the PT35 substitutions were categorized *unpredictable*, 5 were *potentially active* and 1 was predicted *potentially deactivating*. This may indicate that the mutagenesis strategy is indeed suitable to develop a thermodynamically more stable ED30 and a less stable PT35. However, no serious structural changes were evident from the ΔRMSD values from MD simulations. Therefore, the highest priority was given to the effect of the substitution on the potential change and the hydration shell.

In many previous studies, the combination of stabilizing single mutations resulted in mutants with strikingly increased stability. In some cases, the effect of combining single mutations was reported to be additive.[111,126,209,210] Especially in surface charge engineer-

ing, mutants with multiple mutations, such as quadruple or octuplet mutants, showed the greatest improvement of stability.[132,211] Therefore, favorable substitutions were combined to triple and 12-fold (multi) mutants of ED30 and PT35. Since the most favorable substitutions were found in group B, this group was preferred for the generation of the triple mutants by combination. The combination strategy was based on the distribution of substitutions, i.e. a triple mutant with substitutions distributed over the entire surface was preferred to obtain a holistic surface potential change. Thus, one triple mutant of both ED30 (K124D/R211E/K337E) and PT35 (E199R/E328K/D439K) was generated that fully met the rational criteria (figure 3.6). Each substitution exhibited highly favorable ΔEPS, ΔH-bonds and ΔRDF values, while showing an average or low CS. Only D439K was considered *potentially deactivating* based on the CompassR rule. However, this could be compensated by a favorable ΔRMSD.

At the same time, 5 alternative-triple mutants were generated for both ED30 and PT35 based on other combination strategies, e.g. the concentration of substitutions at the front, back or active site of the enzyme. The triple mutants, including the alternative 5 triple mutants per enzyme were first simulated for 100 ns in water only to identify the best variant and combination strategy (listed in section 2.1.1). However, the simulation time and the restriction to water-only solvent were insufficient to discern a difference from the single mutants. Therefore, the triple mutants were considered the most promising variants based on the rational criteria.

To generate the multi-mutant, the substitutions were combined in a manner that the EPS of ED30 matched that of PT35 as closely as possible and *vice versa*. Thus, one multi-mutant of ED30 (R23N/K78Q/K99D/K124D/R211E/K329E/K337E/R373D/R381N/R400Q/Q447E/K471N,) and one of PT35 (N20R/D93K/E121Q/E199R/D268R/D314N/E328K/Q331R/N375R/Q379R/Q429K/D439K) was designed. To investigate the effects of the combined substitutions on the structural stability and hydration shell of the enzymes, the triple and multi-mutants of ED30 and PT35 were simulated for 500 ns in water-only and in 80 % acetonitrile (ACN). As one of the most important water-miscible solvents in industrial processes, ACN has been used for computational and experimental studies of enzyme resistance to organic solvent.

3.2 MD simulations of the triple and multi-mutants

3.2.1 Computational analysis of the triple mutants

The triple mutants were generated by combination of 3 favorable substitutions as described above. The substitutions were distributed across the surface of the enzymes and

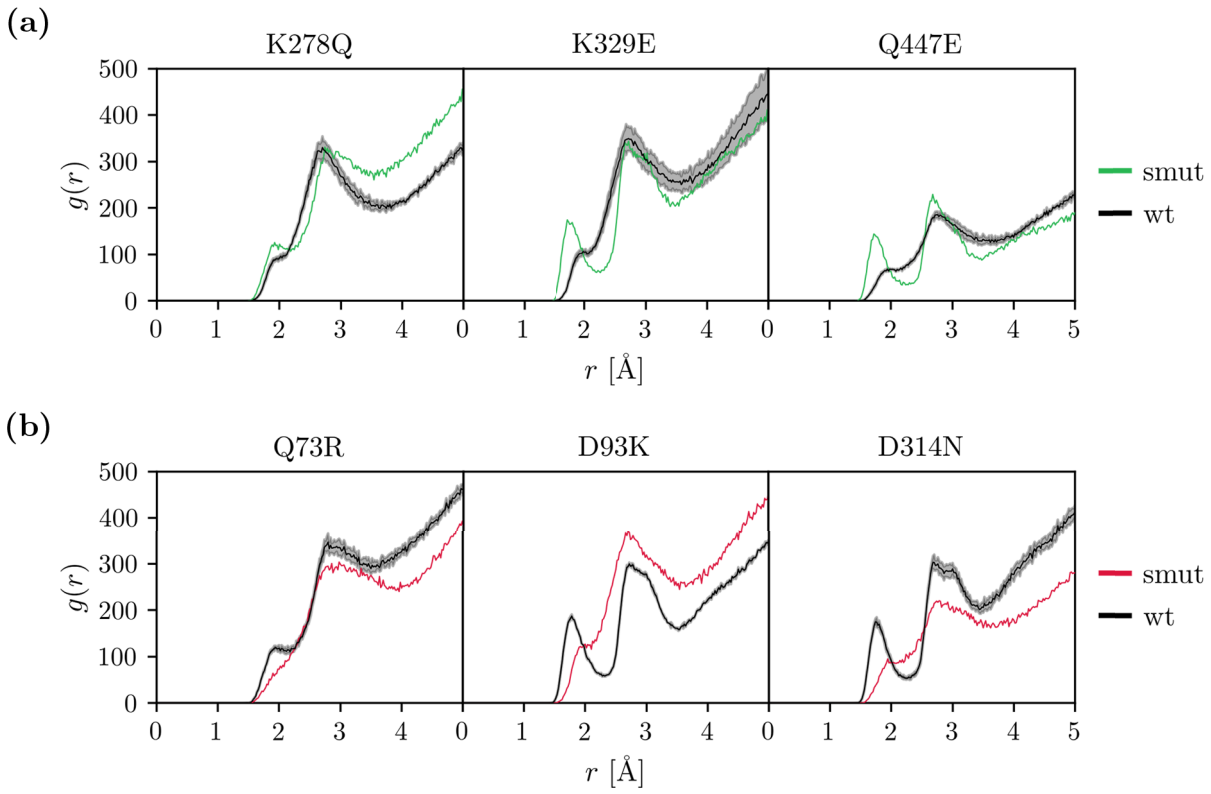

Figure 3.5: RDF of water and substitutions representatively shown by 3 single ED30 and PT35 mutants. The RDF of water within 5 Å, is shown for 3 substitutions of (a) ED30 and (b) PT35. The results of the wild type are presented as the means±SEM from triplicate determinations, while single determination were performed for single mutants.

showed great effect on the surface potential hot spots (figure 3.6). With respect to the ED30 triple mutant, the most pronounced effect was found for K337E, with the EPS indicating a large area of negative potential after substitution. The K124D substitution allowed reduction of the highly positively charged region of the lower front of ED30, which was one of the most prominent hot spots. Using the third substitution, R211E, a negative charge could be inserted to the predominantly positive or neutral surface on the backside of ED30. On the front of PT35, 2 negative surface potential hot spots were affected by E328K and D439K, resulting in a more positively charged surface of the triple mutant. The centrally positioned substitution, E199R, caused a relatively large surface region of the strongly negatively charged back of PT35 to become more positive. Thus, by inserting 3 charged residues at specific positions of ED30 or PT35, an overall more negatively or positively charged surface was obtained that showed some similarities to that of the other wild-type enzyme.

To analyze the changes in dynamics and solvent interaction due to mutation, the triple mutants were simulated in water-only and 80 % ACN for 500 ns and compared to the wild-type enzymes.

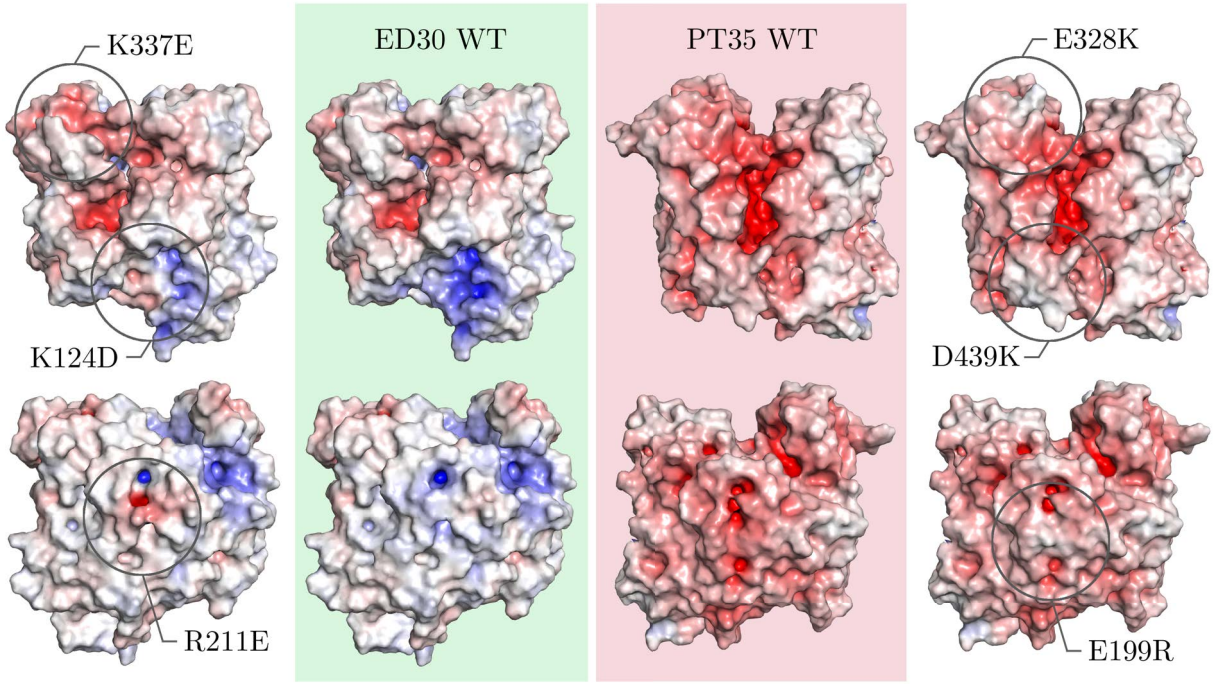

Figure 3.6: EPS of triple mutants compared to wild type. The electrostatic potential surface (EPS) of the triple-mutant ED30 (left) and PT35 (right) was calculated using the PyMOL plugin APBS[142,195] and compared to that of the wild types (center). The front and back views are displayed in the top and bottom row, respectively. The EPS ranges from negatively charged (red, −10) to uncharged (white, 0) to positively charged (blue, +10) and is given in units of kT/e (see figure 3.1)

Stability of the triple mutant. The dynamics of an enzyme can give an indication of how stable the structure of the protein is. Severe changes in the structure can be accompanied by inactivation of the enzyme. Therefore, the effects of the substitutions on the triple-mutant stability were analyzed by comparison with the wild type on the basis of the radius of gyration (Rgyr), the root mean square deviation (RMSD) and the root mean square fluctuation (RMSF) (figure 3.7).

The Rgyr describes the globularity of a protein and can indicate whether unfolding is taking place, which is often characterized by a disruption of the compact form of a protein to an extended conformation. The Rgyr of triple-mutant and wild-type ED30 remained constant in both systems, water-only and 80 % ACN. Only a minor difference between triple mutant and wild type was observed in water with the mutant showing a slightly more compact conformation ($|\Delta\text{Rgyr}| \approx 0.45$ Å). In contrast, the Rgyr of both ED30 variants in 80 % ACN displayed congruent values over the entire simulation time. In both systems, the triple mutant maintained its globular structure and did not show marked differences from the wild type.

In the water-only system, the triple-mutant and wild-type PT35 showed very similar and constant Rgyr values over the entire simulation time. In contrast, higher Rgyr values

were observed at 80 % ACN for both the triple mutant and wild-type PT35 compared to the water-only simulation. This was accompanied by increased variation indicated by a marked standard error of the mean (SEM) of both ED30 variants. In the presence of ACN, the triple mutant showed even further deviation from the wild type ($|\Delta\text{Rgyr}| \approx 0.73$ Å).

The RMSD provides information about the conformational changes from the initial enzyme structure during the MD simulation. Since mutant and wild type had very similar initial structures, their RMSD could be compared. Simulations of both ED30 variants in water showed a constant RMSD of about 4 Å, indicating that a stable conformation was adopted (figure 3.7). As with the Rgyr, a slightly lower RMSD was observed for the triple mutant than for the wild type ($|\Delta\text{RMSD}| \approx 0.53$ Å). In 80 % ACN, on the other hand, both ED30 variants showed a steady but slightly increasing RMSD over the entire simulation time after the rapid increase to about 4 Å within the first 50 ns. Here, no differences were detected between the triple mutant and the wild type.

Similar to the ED30 variants, after the first 50 ns an RMSD of around 4–5 Å was exhibited by both PT35 variants in water. While the RMSD remained constant for the triple mutant, a rise of the RMSD to around 6 Å was observed at around 150 ns for the wild type, resulting in a slightly increased average RMSD compared to the triple mutant ($|\Delta\text{RMSD}| \approx 0.36$ Å). In 80 % ACN, the wild type achieved the stable conformation with an RMSD of 6 Å already after 50 ns, while the RMSD of the triple mutant increased steadily throughout the simulation, yet the difference of the average RMSD between triple mutant and wild type was small ($|\Delta\text{RMSD}| \approx 0.34$ Å). Nonetheless, these results may indicate that the triple mutant was able to find a stable conformation in water, but not in the presence of ACN, which may be correlated with ongoing unfolding or structural rearrangement of the protein.

The RMSF provides information about the flexibility of each residue, allowing the identification of more or less flexible areas within the enzyme. Adaptation of proteins to extreme environments is generally associated with the conservation of functionally important flexible regions.[212,213] To avoid impairment of enzyme function and stability changes unintended by the mutagenesis strategy, the triple mutant was analyzed for flexibility changes in specific regions upon mutation. In previous studies, the most flexible regions of the wild-type ED30 and PT35 were located and linked to specific regions of 5-50 residues on the primary sequence of the enzymes.[139] The regions were denoted as follows e/p(ED30/PT35) + n/c(N-/C-terminal side) + h/s/o(helix/β-strand/others) + numbering(from N- to C-termius). This classification was used to locate flexibility changes in the triple mutants.

To facilitate the comparison of the RMSF between triple mutant and wild type in water only and 80 % ACN, the ΔRMSF_{SOL} (RMSF(ACN)-RMSF(water)) and the ΔRMSF_{MUT}

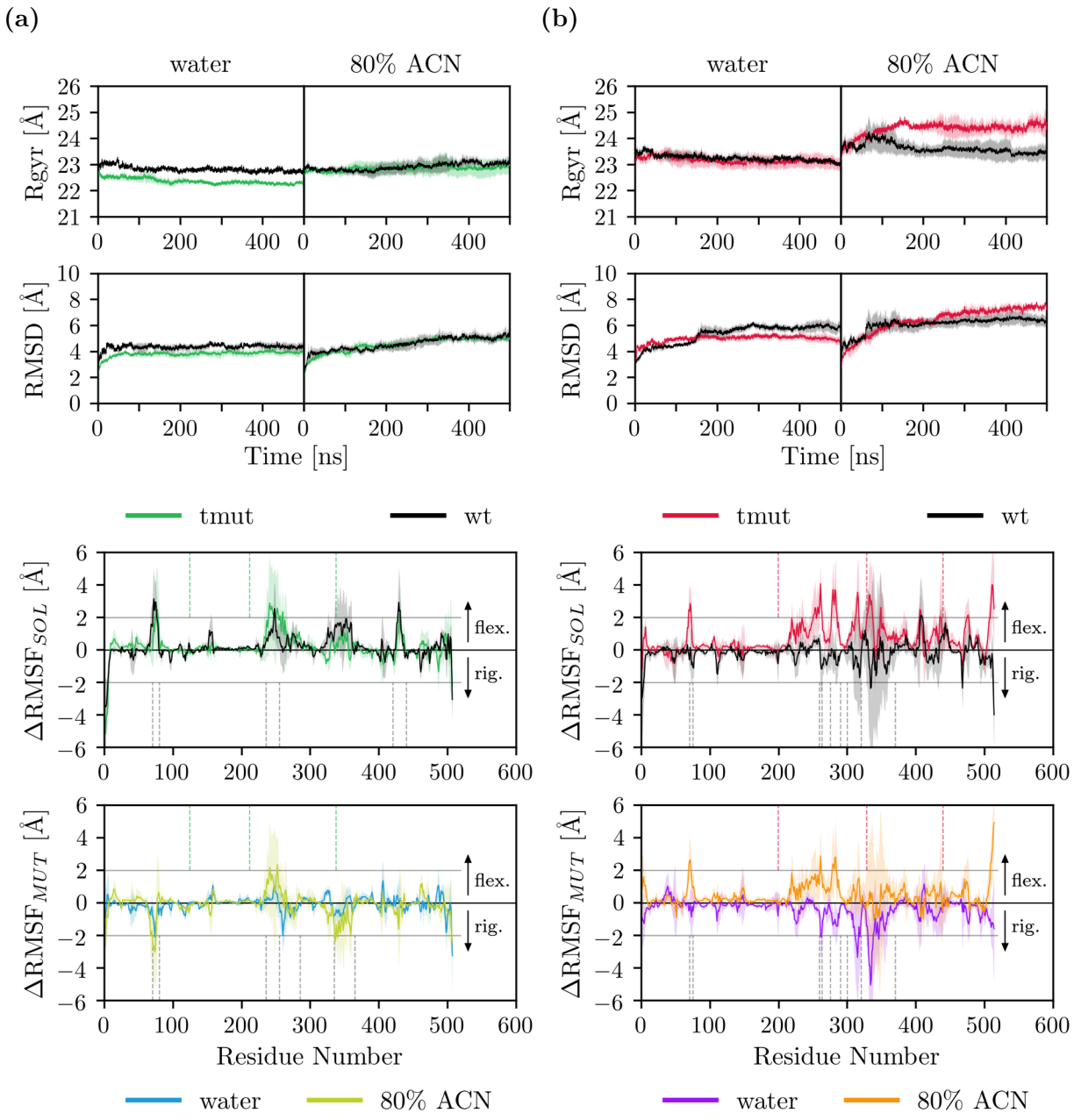

Figure 3.7: Stability of triple mutant and wild-type ED30 and PT35 estimated by Rgyr, RMSD and RMSF. The Rgyr, RMSD and RMSF were determined for the (a) ED30 and (b) PT35 triple mutant (tmut) and wild type (wt). MD simulations were conducted in water and 80 % ACN for 500 ns. The ΔRMSF_{MUT} and ΔRMSF_{SOL} display the change of RMSF due to introduction of mutations (RMSF(tmut)-RMSF(wt)) and organic solvent (RMSF(ACN)-RMSF(water)), respectively. The positions of the triple mutant substitutions of ED30 (K124D,R211E,K337E) and PT35 (E199R,E328K,D439K) are marked by vertical green or red dotted lines and the boundaries of the flexible areas are indicated by vertical gray dotted lines. Residues with a changed flexibility ($|\Delta\text{RMSF}| > 2$ Å) are marked by a horizontal gray line. The arrows at the right-hand side indicate an increase in flexibility or rigidity. The results are presented as the means±SEM from triplicate simulations.

(RMSF(triple mutant)-RMSF(wild type)) were determined (figure 3.7). The $\Delta RMSF_{SOL}$ describes the difference in flexibility between the water-only and the 80 % ACN simulation, while the change in flexibility upon mutation is indicated by $\Delta RMSF_{MUT}$. When $|\Delta RMSF| > 2$ Å, the flexibility is considered changed.

As shown by $\Delta RMSF_{SOL}$, both ED30 variants gained flexibility in the presence of ACN in the same manner with respect to enoh1 (residues 70-80), enoh3 (235-255) and ecoh4 (420-440).

The $\Delta RMSF_{MUT}$ plot revealed that differences in flexibility between the triple-mutant and wild-type ED30 occurred in enoh1, enoh3, enoh4 (255-285) and ecoh2 (335-365). The triple mutant became more rigid in enoh1 and enoh4. In the presence of ACN, the increased rigidity of enoh1 upon mutation was even more pronounced, while no differences between wild type and triple mutant could be observed for enoh4. Moreover, in ACN, the triple mutant gained flexibility with respect to enoh3, while ecoh2 became less flexible.

The $\Delta RMSF_{SOL}$ of PT35 indicated that, due to high SEM values, no certain differences in flexibility between water and 80 % ACN simulation could be detected for the wild-type PT35 (figure 3.7). In contrast, the triple mutant gained in flexibility with respect to pno2 (70-75), pno4 (275-290), pco1 (300-320) and pcoh2 (320-370). Further deviation was observed for the residues 259-263, which can be assigned to a novel flexible region pnoX.

The $\Delta RMSF_{MUT}$ plot displayed that the triple mutant was overall more rigid in water and more flexible in ACN compared to the wild type. While pco1 and pcoh2 showed higher flexibility in water, more rigidity was observed for pno2 and pno4 in ACN. Only pnoX was both, more rigid in water and more flexible in ACN.

The most pronounced increase in rigidity exhibited by pco1 and pcoh2 of the triple-mutant PT35 may have been caused by the substitution E328K. Analysis of the trajectory using PyMOL revealed that E328, which is part of a highly negatively charged region of the wild type, did not interact with any other residue. After substitution, a positively charged area was formed by K328 and R325. This seemed to attract many negatively charged residues, including E319 and E310, which could explain the increased rigidity of the pco1. Moreover, the helix harboring the negatively charged residues D340, D341 and D342 was located closer to the positively charged area of K328 and R325, most probably due to attraction of the opposite charges. These movements pushed the residue Y333 towards F370 and W367. It can be assumed that hydrophobic interactions between these residues caused the increased rigidity in pcoh2.

There are controversial hypotheses regarding the connection of flexibility and organic solvent tolerance. Park *et al.* considered those residues with high flexibility as organic solvent affecting site.[122] Enhanced rigidity of the affecting site due to mutation led to mutants showing higher organic solvent stability. In contrast, Cui *et al.* reported from a

rational-engineering study that increased flexibility near the substitution sites was one of the main factors leading to improved organic solvent-resistant enzymes.[196] Therefore, it is hard to determine what changes in flexibility may affect the stability of the enzyme in organic solvents. In order to prevent inactivation of the mutant enzymes, it is primarily important that the changes in flexibility compared to the wild type are kept rather low. In this respect, the triple-mutant ED30 showed only minor overall changes in flexibility compared to the wild type in both systems, while greater differences in flexibility were observed between the triple-mutant and wild-type PT35. However, besides increased rigidity of specific regions in water and an overall high flexibility in the presence of ACN, differences between the wild type and triple mutant of PT35 could hardly be defined due to the great variance.

Solvent interaction of the triple mutant. Protein surface hydration is assumed to be a relevant factor for organic solvent tolerant enzymes to prevent organic solvents from penetrating and destroying their hydrophobic protein core.[66,122,123] Therefore, the substitutions of the triple mutant were examined for the change in hydration by comparison with the wild type in water and 80 % ACN. The cumRDF is a convenient measure to estimate the number of water molecules within a certain distance. To capture only water molecules tightly bound to the substitution, the cumRDF of water and the substitution was determined for a distance of 2 Å.

First, all residues exhibited a lower cumRDF of water in 80 % ACN than in water, however some substitutions, such as K124D and K337E of ED30, were able to retain a high amount of closely bound water in the presence of ACN. Each substitution of the ED30 triple mutant showed an increased cumRDF compared to the wild type, both in water and in 80 % ACN. The minimum deviation from the wild type was 1.32 in water-only and 1.40 in 80 % ACN. In contrast, a decrease of cumRDF was observed for each substitution of the PT35 triple mutant in both systems. Here, the cumRDF was at least reduced by 2.12 in water-only and by 1.25 in 80 % ACN. Similar results were obtained by calculating the H-bonds between substitution and water (figure A2). These results agree with those of the single mutants simulated in water only, which showed higher hydration of negatively charged residues, while positively charged residues exhibited less tight contact with water (figures 3.4 and 3.5). The negatively charged residues could also bind more water than the positively charged residues.at high ACN concentrations.

To determine the effects of the substitution on the overall enzyme hydration shell, the number of H-bonds between the triple mutant and water was calculated (figure A4). A tendency of the PT35 triple mutant to make less H-bonds than the wild type was observed. However, due to a large SEM, no difference between the number of H-bonds of

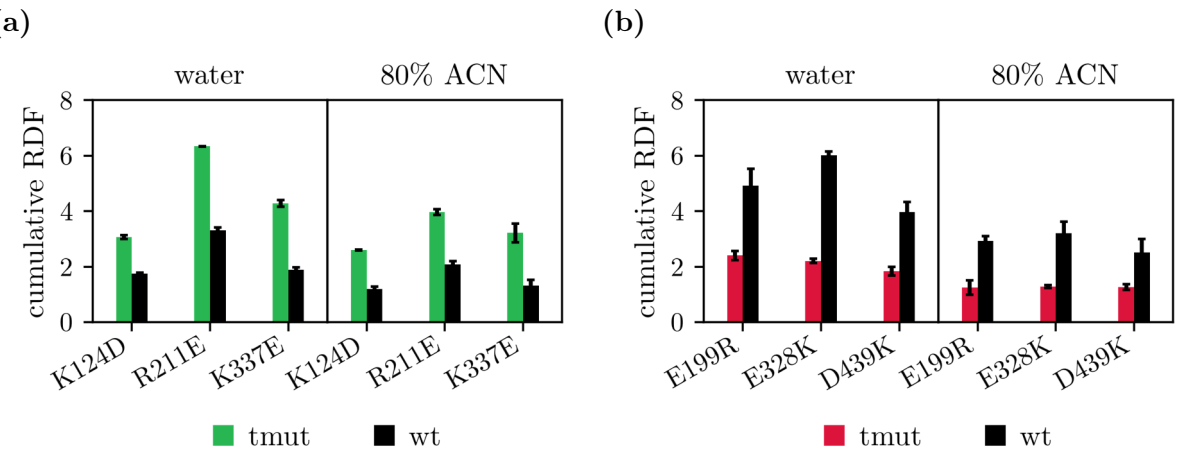

Figure 3.8: Hydration shell of triple-mutant ED30 and PT35 determined by the cumulative RDF of each substitution. The cumRDF of water within 2.0 Å was calculated for each substitution of the (a) ED30 and (b) PT35 triple mutant (tmut) and wild-type (wt) simulated in water and 80 % ACN. The results are presented as the means±SEM from triplicate simulations.

the triple mutants and the wild types could be detected. As observed before with respect to the substitutions, less water contacts were exhibited by both enzymes in the presence of ACN, while, in parallel, the number of intramolecular H-bonds increased (figure A5). Interactions with ACN may be less favored, causing the enzyme to exert more intramolecular H-bonds. This led to the question of how strongly the triple mutant was affected by ACN. The more a residue favors interaction with organic solvents the greater the probability that crucial water molecules will be stripped from the enzyme surface around the residue, allowing the organic solvent molecules to enter the protein core.[64] The cumRDF of ACN within 3.0 Å was determined for the triple-mutant and wild-type ED30 and PT35 (figure 3.9). Here, an opposing effect was shown by the substitutions compared to the cumRDF of water. After substitution, the cumRDF of ACN was reduced in the triple-mutant ED30 by at least 2.48. The interaction with ACN was reduced the most for K337E (ΔcumRDF $\approx$ –4.76). The substitution of the triple-mutant PT35, by contrast, showed increased contacts to ACN. Here, E328K caused the highest deviation (ΔcumRDF $\approx$ 2.75). No difference to the wild type could be defined for D439K due to large SEM values. These results may indicate that positively charged residues exhibit strong interaction with ACN, while negatively charged residues avoid these contacts.

In summary, a triple-mutant ED30 with a more negatively charged surface could be created, where each substitution exhibited improved hydration shells and reduced interactions with ACN. In parallel, a more positively charged triple-mutant PT35 could be obtained with each substitution showing lower contacts to water and more ACN interac-

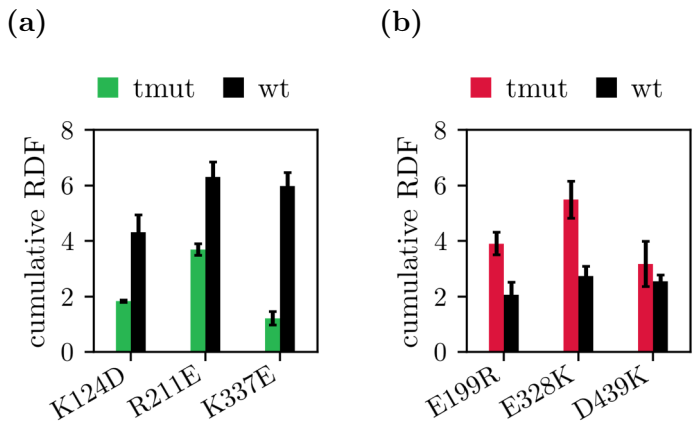

Figure 3.9: Cumulative RDF of ACN and each substitutions of the triple-mutant and wild-type ED30 and PT35. The cumRDF of ACN within 3.0 Åwas calculated for each substitution of the (a) ED30 and (b) PT35 triple mutant (tmut) and wild-type (wt) simulated in 80 % ACN. The results are presented as the means±SEM from triplicate simulations.

tions. Moreover, no significant and systematic overall structural change was observed for the triple mutants of ED30 and PT35 in either solvent, which led to the assumption that the enzyme activity should not be strongly impaired.

3.2.2 Computational analysis of the multi-mutants

With respect to surface charge engineering, it was found that enzyme mutants containing a large number of mutations in particular exhibited extraordinary stability enhancements.[132,211] Therefore, multi-mutants were generated by combining 12 of the 17 selected substitutions (section 3.1). The mutations were combined so that the surface charges of ED30 matched those of PT35 as much as possible and *vice versa* (figure 3.10).

The EPS of the multi-mutant ED30 showed a greatly increased negatively charged surface, especially at the upper front side and the active site (enzyme center) of ED30, which showed great similarity to the EPS of the wild-type PT35. In addition, the highly positively charged hot spot at the lower front side of ED30 could greatly be reduced by replacing positively charged residues with negatively charged and polar residues. The two positively charged hot spots at the upper center and upper right of the back of ED30 were markedly reduced and several additional negative charges were introduced.

The front surface of the multi-mutant PT35 showed a distinctly decreased negative charge around the active site (enzyme center). In addition, the introduction of mainly positively charged residues caused two hot spots, the upper left and the lower right, to become considerably more positive. The backside of the multi-mutant PT35 was found to possess a positive ESP surface in the upper region and diminished negatively charged

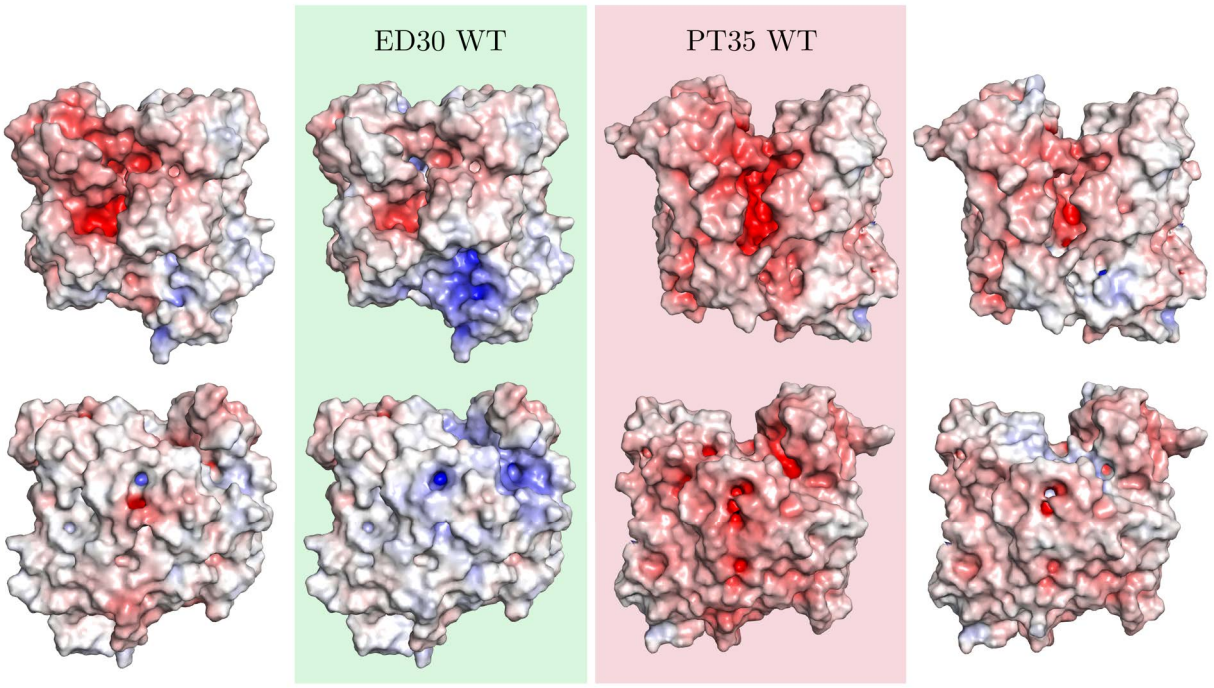

Figure 3.10: Electrostatic potential surface of the multi-mutants compared to the wild types. The electrostatic potential surface (EPS) of the multi-mutant ED30 (leftmost) and PT35 (rightmost) was calculated using the PyMOL plugin APBS[142,195] and compared to that of the wild types (center). The front and back views are displayed in the top and bottom row, respectively. The EPS ranges from negatively charged (red, −10) to uncharged (white, 0) to positively charged (blue, +10) and is given in units of kT/e (see figure 3.1).

hot spots in the central regions.

By inserting 12 substitutions at specific positions of ED30 or PT35, an overall greatly increased negatively or positively charged surface was obtained, which showed certain similarities to the other wild-type enzyme.

The multi-mutants were simulated in water-only and 80 % ACN for 500 ns in order to study the deviations in dynamics and solvent interaction due to mutation.

Stability of the multi-mutant. The effects of the substitutions on the multi-mutant stability were analyzed by determination of Rgyr, RMSD and RMSF, and compared to the wild type (figure 3.11).

The Rgyr and RMSD of the multi-mutant ED30 remained constant in water and barely showed any deviation from the wild type. In 80 % ACN, the multi-mutant exhibited higher Rgyr and RMSD values (ΔRgyr $\approx$ 0.66 Å, ΔRMSD $\approx$ 0.98 Å), whereby the Rgyr converged with that of the wild type towards the end of the simulation, indicating that the mutant has reverted to a more dense conformation. In the same manner as the wild type, the RMSD of the multi-mutant increased steadily in 80 % ACN but remained constant

for the last 100 ns. Higher SEM values of the RMSD indicated increased variance of the multi-mutant compared to the wild type in the presence of ACN, yet the deviations between both variants were low.

The multi-mutant PT35 showed slightly increased Rgyr values compared to the wild type in water (ΔRgyr ≈ 0.39 Å) and from approximately 200 ns on the multi-mutant and the wild type displayed congruent and steady RMSD values. In the presence of ACN, only minor deviations between the Rgyr and RMSD of the multi-mutant and wild type PT35 were observed when considering the relatively high variance of the Rgyr and RMSD as judged by their SEM ranges.

Both multi-mutant enzymes could be considered similar to the respective wild type in terms of their stability during simulation, which means that no obvious mutation-induced stabilization or destabilization effect are to be expected in the presence of ACN.

The ΔRMSF_{SOL} and ΔRMSF_{MUT} were determined for ED30 and PT35 multi-mutants, describing flexibility changes ($|\Delta\text{RMSF}| > 2$ Å) due to the presence of ACN and due to multi-mutation, respectively (figure 3.11).

The ΔRMSF_{SOL} plot of ED30 revealed increased flexibility of the multi-mutant in enoh3 (235-255), enoh4 (255-285) and ecoh2 (335-365) due to ACN. This change of flexibility is more pronounced for the multi-mutant than for the wild type and triple mutant (figure 3.7) Increased rigidity was, on the other hand, observed for eco5 (455-470) in ACN.

According to the ΔRMSF_{MUT} plot, the main difference between the multi-mutant and wild-type flexibility in water were found in the more flexible eco5 (455-470). In the presence of ACN, a more rigid enoh1 (70-80) region and more flexible enoh3 and eco3 (400-410) regions were shown by the multi-mutant compared to the wild type.

The ΔRMSF_{SOL} plot of the multi-mutant PT35 revealed that the pno4 (275-290) and pco1 (300-320) regions of PT35 exhibited increased flexibility due to the presence of ACN, while more rigidity was observed for pco4 (400-415). In the wild type, the regions pno4 and pco1, however, were not affected by ACN, while pco4 showed more flexibility due to presence of ACN.

By comparing the multi-mutant and wild-type PT35 on the basis of ΔRMSF_{MUT}, flexibility changes upon mutation could be detected in pco1 and pcoh2 exhibiting higher rigidity in water, while pco4 indicated enhanced flexibility. With the exception of pco4, similar changes in flexibility were observed earlier for the triple-mutant PT35 in water. The most pronounced difference to the triple mutant was the increased rigidity of the multi-mutant in pcoh2 in the presence of ACN. This could be related to the fact that the pcoh2 region of the multi-mutant exhibited less variance than that of the triple mutant.

Although some flexibility changes induced by the substitutions were detected, no significant differences were found between multi-mutant and wild type ED30 and PT35, when

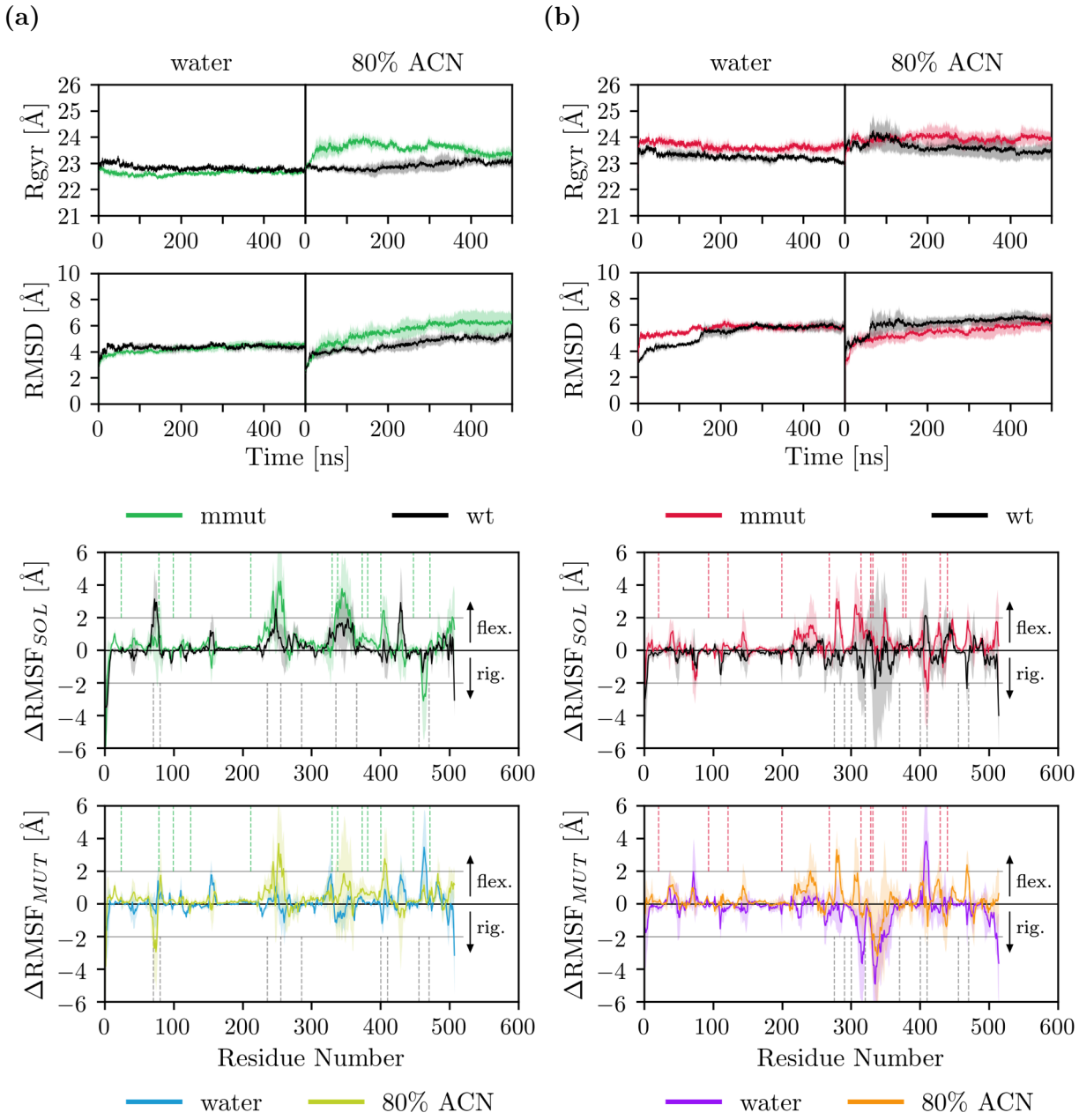

Figure 3.11: Stability of multi-mutant ED30 and PT35 determined by Rgyr, RMSD and RMSF. The Rgyr, RMSD and RMSF was determined for the (a) ED30 and (b) PT35 multi-mutant (mmut) and wild type (wt). MD simulations were conducted in water and 80 % ACN for 500 ns. The ΔRMSF_{MUT} and ΔRMSF_{SOL} display the change of RMSF due to introduction of mutations (RMSF(mmut)-RMSF(wt)) and organic solvent (RMSF(ACN)-RMSF(water)), respectively. The positions of the substitutions are marked by vertical green or red dotted lines and the boundaries of the flexible areas are indicated by vertical gray dotted lines. Residues with a changed flexibility ($|\Delta\text{RMSF}| > 2$ Å) are marked by a horizontal gray line. The arrows at the right-hand side indicate an increase in flexibility or rigidity. The results are presented as the means±SEM from triplicate simulations.

considering the high variance of the flexible regions.

Solvent interaction of the multi-mutant. To estimate the effect of mutation on the hydration of the multi-mutant enzyme, the cumRDF of water within 2.0 Å was determined for each substitution.

Most of the substitutions of the multi-mutant ED30 showed a strongly increased cumRDF of water compared to the wild type in both water only and 80 % ACN (figure 3.12). The greatest increase of hydration was observed for substitutions with negatively charged residues, which agrees with the results from the single mutant simulations in water only. Among them, K329E showed the most pronounced effect in water (ΔcumRDF $\approx$ 3.46).

While the cumRDF indicated slightly increased hydration of R23N and K78Q in the

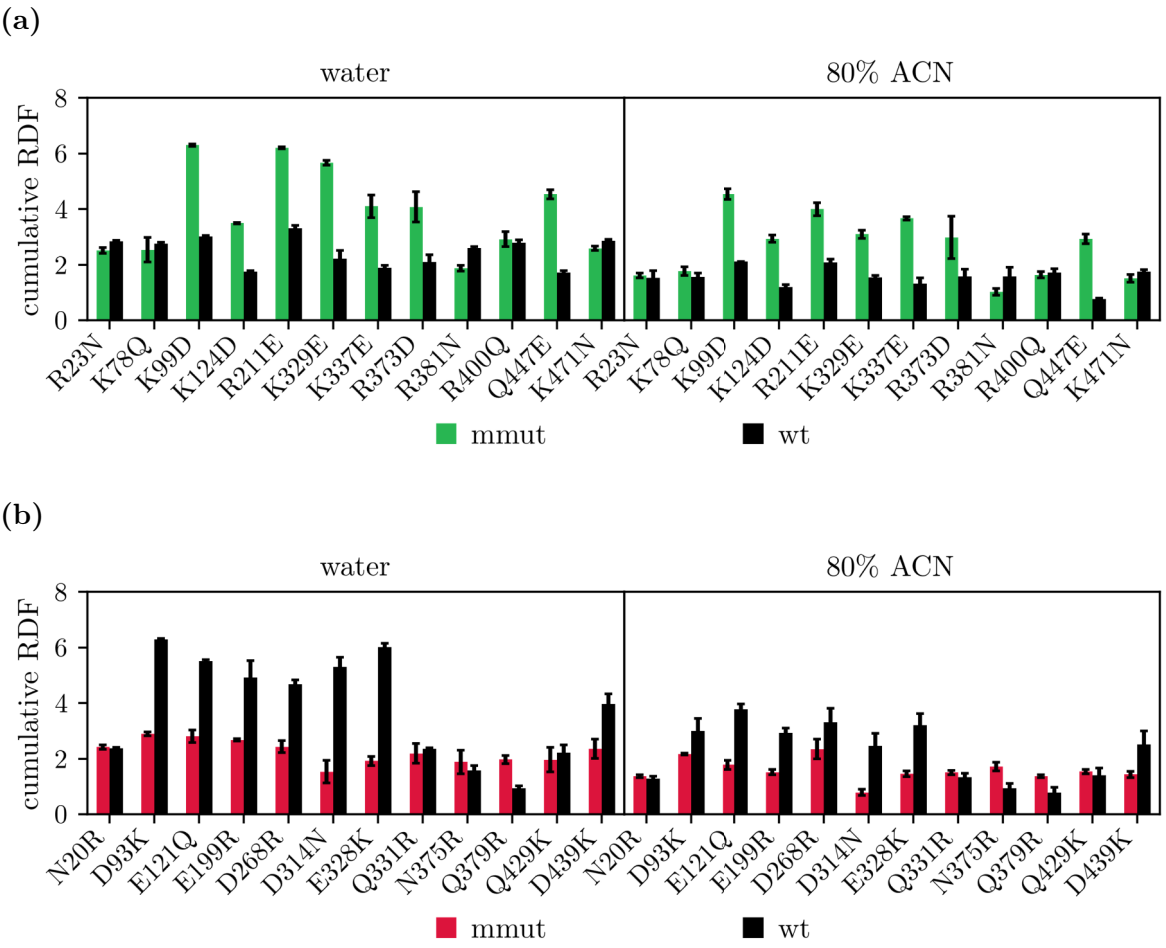

Figure 3.12: Hydration shell of multi-mutant ED30 and PT35 determined by the cumulative RDF of water and each substitution. The cumRDF of water within 2.0 Å was calculated for each substitution of the (a) ED30 and (b) PT35 multi-mutant (mmut) and wild-type (wt) simulated in water and 80 % ACN. The results are presented as the means±SEM from triplicate simulations.

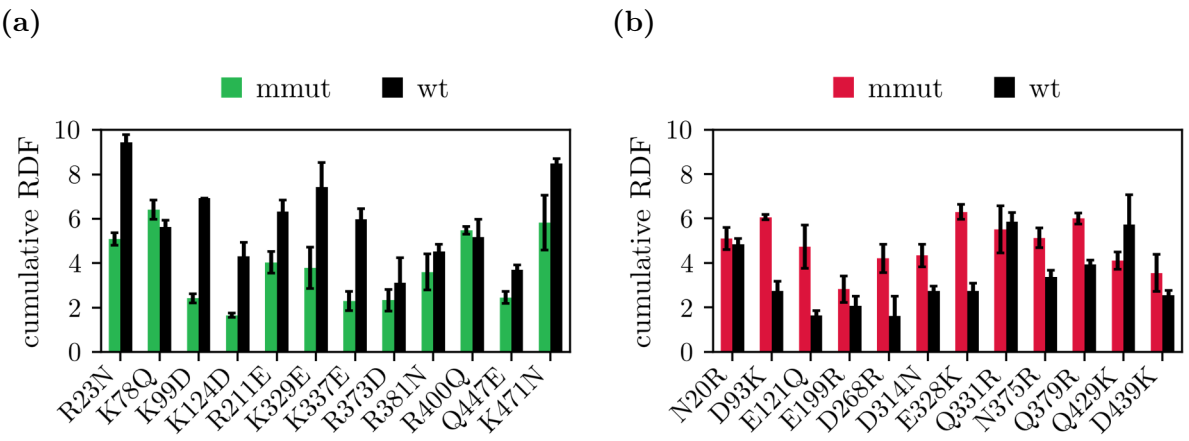

Figure 3.13: RDF of ACN and each substitutions of the multi-mutant ED30 and PT35. The cumRDF of ACN within 3.0 Å was calculated for each substitution of the (a) ED30 and (b) PT35 multi-mutant (mmut) and wild-type (wt) simulated in 80 % ACN. The results are presented as the means±SEM from triplicate simulations.

presence of ACN, slightly reduced hydration was detected in water. Substitutions with decreased cumRDF exhibited only minor deviations with the maximum reduction from the wild type in both solvents shown by R381N (ΔcumRDF $\approx$ –0.33). In contrast to the multi-mutant ED30, a decrease of cumRDF was observed for the most substitutions of the multi-mutant PT35 in both solvents. Few substitutions resulted in increased hydration, with the maximum increase of cumRDF found for Q379R in water (ΔcumRDF $\approx$ 1.04 Å). Similar results were obtained by calculating the H-bonds between substitution and water (figure A3).

In comparison to the triple mutant, a larger difference between multi-mutant and wild-type ED30 was observed with regard to the H-bonds of the overall enzyme and water (figure A4). The ED30 triple mutant showed a tendency to bind more water than the wild type in both solvents. Similar to the triple mutant, the PT35 multi-mutant showed a tendency to bind less water than the wild type in water, while no deviation was detected in ACN. However, the findings exhibited large SEM, which made it difficult to define a difference between the mutants and the wild type.

To examine the interactions between substitution and ACN, the cumRDF of ACN within 3.0 Å was determined for the multi-mutants and wild types in 80 % ACN (figure 3.13). The findings of the ED30 multi-mutant indicated that the interaction with ACN could be reduced by almost all substitutions, among which K99D showed the most pronounced effect (ΔcumRDF $\approx$ –4.50). The substitutions K78Q and R400Q caused an increase of ACN density in close proximity, however, the deviations from wild type were low and within the SEM range.

With the exception of Q331R and Q429K, all substitutions of the multi-mutant PT35

resulted in higher cumRDF of ACN. Here, the greatest effect was found for E328K (ΔcumRDF $\approx$ 3.56). Slightly reduced cumRDF values with deviations within the SEM range were observed for Q331R and Q429K. The results strengthened the hypothesis that positively charged residues favor interaction with ACN, while negatively charged residues avoid these contacts. Instead, negatively charged residues were suggested to facilitate interaction with water molecules.

In summary, a multi-mutant ED30 with a significantly more negatively charged surface could be created whose substitutions predominantly exhibited improved hydration shells, while interactions with ACN were suppressed. Similarly, a more positively charged PT35 could be obtained after introduction of substitutions that exhibited lower contacts to water molecules and mainly favored ACN interactions. In addition, both multi-mutants of ED30 and PT35 showed only minor deviations in stability from the wild type during MD simulation, which may indicate that enzyme activity is not severely affected.

3.3 Experimental investigation of the wild-type enzymes

3.3.1 Production of the wild-type enzymes

The biochemical characterization of ED30 and PT35 was performed in independent previous studies by Tao *et al.*[140] and Bollinger *et al.*, respectively[54,136]. In this study, both wild-type enzymes were directly compared regarding organic solvent, buffer and thermal stability in order to determine optimal conditions for subsequent analyses of the mutated variants.

The wild-type enzymes were, therefore, produced in *E. coli* BL21(DE3) and purified via immobilized metal ion affinity chromatography (IMAC) using an Ni-NTA matrix. Two variants of the wild-type PT35 were prepared, one carrying the native signal peptide (SP) (native PT35) and one with the shorter PelB-SP (PelB-PT35). Each step of the purification was sampled and visualized by SDS-PAGE (figure 3.14). Using immunological detection, ED30 and PT35 could be identified at around 55 kDa and 65 kDa, respectively (figure A6).

The banding pattern of the cell extract (CE) indicates an extensive overexpression of both PT35 variants, however, only a minor amount is found in the soluble fraction (SF). The larger amount of insoluble proteins suggests that inclusion bodies were formed by aggregation of PT35. However, this is more pronounced for the native PT35 than for the PelB variant, for which the soluble fraction contains slightly more target protein. In contrast, a relatively large amount of soluble ED30 enzyme was obtained even though

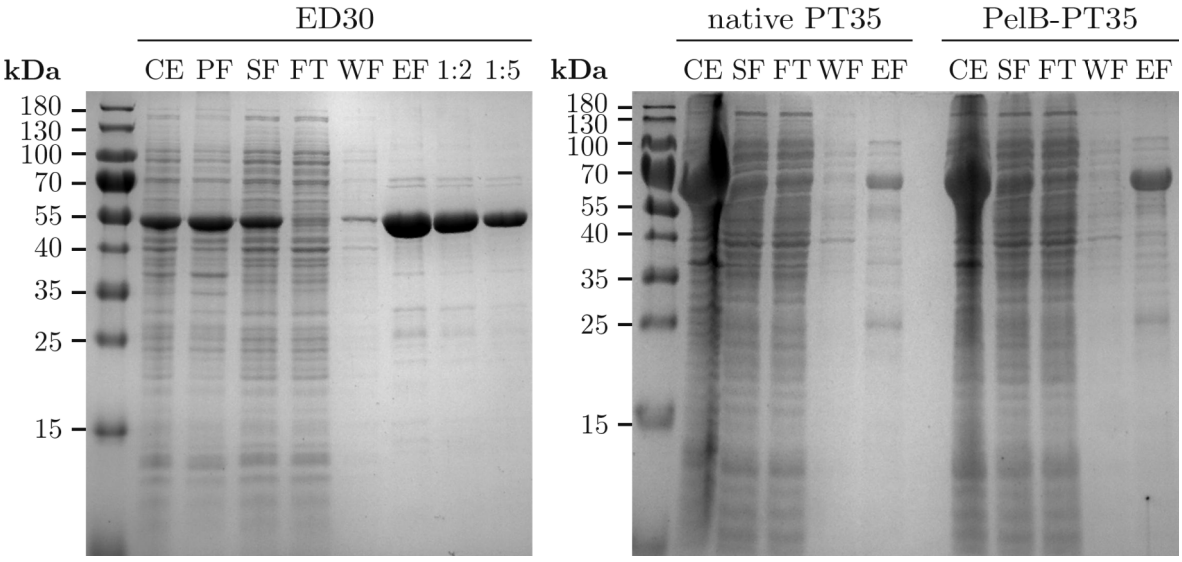

Figure 3.14: Estimation of the wild-type purification by SDS-PAGE. After expression in *E. coli* BL21(DE3) (24 h, 25 °C, 160 rpm), the enzymes were purified by IMAC using Ni-NTA agarose. First row: molecular standard, CE: cell extract, PF: pellet fraction, SF: soluble fraction, FT: flow through, WF: wash fraction, EF: elution fraction (1:2 and 1:5 diluted).

much of insoluble target protein remained in the pellet fraction (PF). Most of the ED30 enzymes bound to the column, while the flow through (FT) of the PT35 purification shows a lot of unbound target protein, indicating a lower binding affinity towards the Ni-NTA matrix.

After washing (WF) the column, the target protein was eluted (EF) from the matrix. The EF fraction of ED30 shows only a few weak bands, which are due to slight contamination by other proteins, indicating a remarkably high purity of greater than 95 % based on the band intensity. The EF of both PT35 variants also show a clear band of the target protein, albeit some other bands are still present. Here, a protein with a molecular weight of 25 kDa contributes most to the contamination. In previous studies, this contamination could be removed by an ion-exchange chromatography following IMAC.[136] However, for the stability assays performed in this study, the purity of the PT35 variants was considered sufficient.

Due to the higher yield and purity of the PelB-PT35 variant and its similar behavior in stability assays compared to native PT35 (figures A7 and A8), PelB-PT35, hereafter only referred to as PT35, was used for comparative studies with ED30 and for mutagenesis to prepare single, triple and multi-mutants.

3.3.2 Organic solvent stability of the wild type

According to the conditions used in the biochemical characterization of ED30 and PT35 in previous studies, the organic solvent stability of both enzymes was investigated in

two buffers, 100 mM potassium phosphate (KPi) buffer (100 mM NaCl, pH 7.2) and 20 mM TRIS buffer (100 mM NaCl, pH 8.0), hereafter only referred to as KPi and TRIS buffer.[136,140] The stability assay was performed at various concentrations of ACN, with the concentration range adjusted to the respective enzyme based on its expected stability. Therefore, PT35 was incubated in 0, 20, 50 and 80 % ACN, while ED30 was examined in lower concentrations of 0, 5, 10 and 20 % ACN. The enzymes were incubated at 30 °C and the activity was measured after different incubation times by the esterase activity assay, which is based on the enzymatic hydrolysis of *p*NPB. The resulting yellow dye can be detected photometrically, allowing a reaction velocity to be determined.

To study the influence of ACN on the enzyme stability, the half-life ($t_{1/2}$) of deactivation, which represents the time required for reducing the original activity by half, was determined by fitting an exponential model $v/v_0 = exp(-K_d t)$ to the residual activity (v/v_0) plot (figure 3.15). v_0 is the esterase activity at the first measurement (0 h) and v the activity after a specific time of incubation in ACN or buffer. K_d represents the deactivation constant of the enzyme and t the time of incubation. Previous studies could show that this model is suitable for other enzyme inactivation processes in water-miscible organic solvents.[111]

The findings showed a steep decline of residual activity for ED30 with increasing ACN concentrations and an especially low stability in TRIS buffer (figure 3.15). Immediate inactivation was observed for ED30 in 20 % ACN and in 10 % ACN in combination with TRIS buffer. In KPi buffer, ED30 showed 5 % residual activity after 1 h and was fully inactivated after 3 h in 10 % ACN. Only in 5 % ACN, a residual activity was measurable for several hours. After 3 h incubation in 5 % ACN, ED30 showed 76 % residual activity in KPi, while only 9 % residual activity was observed in TRIS buffer. Moreover, a relatively rapid loss of activity was shown by ED30 even in the control measurement without ACN (0 % ACN), which was more pronounced in TRIS buffer with 38 % residual activity than in KPi buffer with 67 % residual activity after 24 h incubation. The phenomenon of a rapid loss of activity within a few minutes or hours was observed frequently for ED30, regardless of which buffer was used. The rapidity of the activity decrease could be reduced by prior centrifugation for 30 min of the enzyme solution and discarding of the resultant pellet. Therefore, ED30 was assumed to have a tendency to aggregate, which may have been enhanced by freezing. It is known that protein solubility and stability can be increased by using additives in buffers, such as ionic compounds, salts, detergents or osmolytes. With respect to ED30, the buffer conditions could therefore be further optimized by screening, e.g. different buffer types, additives and salt concentrations.[214]

In previous studies by Tao *et. al*, 22 % and 0.4 % residual activity in 15 % and 30 % ACN after 40 min incubation in 50 mM TRIS buffer (pH 8.0) were reported. This higher

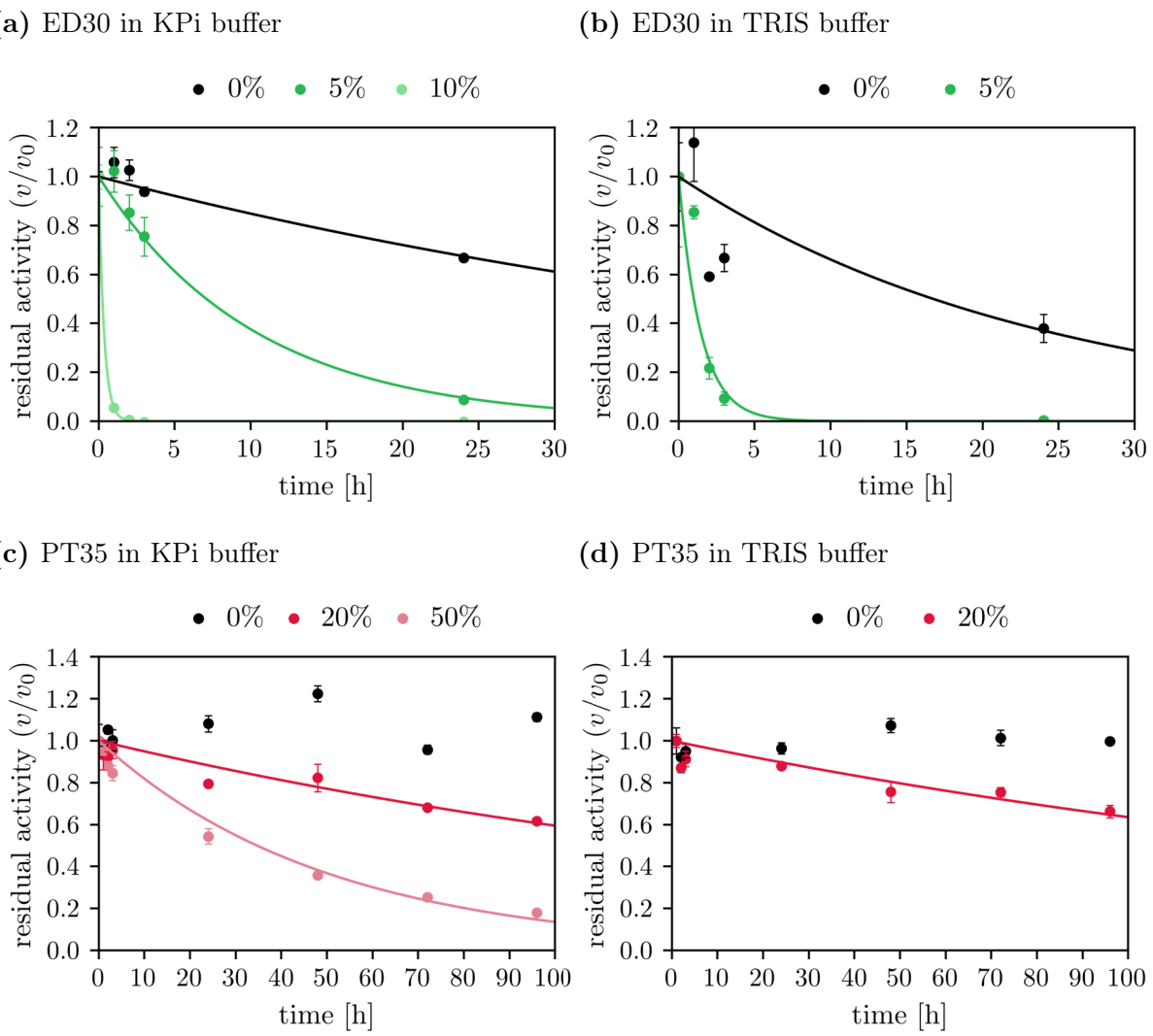

Figure 3.15: Inactivation of the wild type by acetonitrile in different buffers. The residual activity of the wild-type ED30 was detected after different incubation times in the presence of 0, 5, 10 and 20 % ACN in (a) KPi and (b) TRIS buffer. In the same way, the residual activity of wild-type PT35 was determined after different incubation times in the presence of 0, 20, 50, and 80 % ACN in (c) KPi buffer and (d) TRIS buffer. The activity was measured under standard esterase activity assay conditions using 10 µl purified enzyme in 200 µl of KPi or TRIS buffer containing 1 mM *p*NPB. The residual activity (v/v_0) was calculated by comparing the enzyme activity at 0 h (v_0) with the activity after different incubation times (v). The results are presented as the mean±SEM from triplicate measurements.

stability of ED30 may be due to the slightly different conditions, such as a lower incubation temperature of 25 °C. Despite higher concentrations of ACN in comparison to ED30, a slower decrease in residual activity with increasing organic solvent was observed for PT35 (figure 3.15). In both buffers, no PT35 activity could be measured at all in 80 % ACN and in TRIS buffer; PT35 was immediately inactivated in the presence of 50 % ACN. In KPi, by contrast, PT35 maintained 18 % of its activity after the maximal incubation

time of 96 h in 50 % ACN. No differences in PT35 stability between both buffers were observed at 0 and 20 % ACN. After 96 h incubation, PT35 remained fully active in 0 % ACN in both buffers and showed 62 % and 66 % residual activity in 20 % ACN in KPi and TRIS buffer, respectively. In contrast to ED30, there are no indications that PT35 tends to aggregate in aqueous media under the chosen conditions and within the observed time.

In previous studies by Bollinger *et al.*, PT35 was found to be considerably more stable in 80 % ACN, where it maintained 33 % of its activity after 3 h incubation.[54] These results were obtained from stability assays performed with cell lysate, while purified enzyme was used in this study. It is known that impurities often bias the stability of enzymes in the presence of organic solvents, therefore it is recommended to use purified enzymes for stability studies.[15] On the other hand, it also implies that the stability of PT35 could further be enhanced by specific additives present in the cell lysate. In addition, a markedly higher residual activity of around 70 % was reported for PT35 after 96 h incubation at 50 % ACN under similar conditions (100 mM KPi [pH 7.2], 30 °C) but using PT35 with its native signal peptide (SP).[136] When comparing the tolerance of PT35 with native SP and PT35 with PelB-SP to ACN in TRIS buffer, no differences were observed (figure A8). Nevertheless, both PT35 variants should also be compared in KPi buffer to exclude a possible influence of the SP on the PT35 stability differences between the two studies.

Half-life values and inactivation constants of the wild-type ED30 and PT35 were determined for both buffers and different ACN concentrations (table 3.1). Since the correlation coefficient for each model was at least 0.89, high model suitability for the given residual activity plots was assumed. The exponential model could not be fitted to the control measurement of PT35 with 0 % ACN, because no decline of activity was detected during the total incubation time of 96 h. In contrast, ED30 showed relatively rapid declines in absence of ACN with half lives of 42 and 17 h in KPi and TRIS buffer, respectively. These results may indicate either a generally higher stability of PT35 compared to ED30 or that the selected buffer conditions are more favorable for PT35 than for ED30.

Despite a lower tolerance to ACN compared to results from previous studies, PT35 still showed significantly higher stability than ED30 in the presence of ACN, which was particularly evident from the comparison of the half-life values at 20 % ACN. While ED30 was immediately inactivated in both buffers, PT35 exhibited half-lives of 133 h (KPi) and 153 h (TRIS).

The findings indicated a strongly reduced resistance of both enzymes to ACN when incubated in TRIS buffer. The stability of ED30 at 5 % ACN, for instance, was 7-fold lower in TRIS buffer than in KPi buffer and at 10 % and 50 % ACN, even no half-life values could be determined for ED30 and PT35, respectively, due to immediate enzyme deactivation in TRIS buffer. Only at 20 % ACN, a modestly higher stability of 15 % was

Table 3.1: Half-lives and inactivation constants for the wild-type enzymes in different buffers and acetonitrile. The half-lives $t_{1/2}$ and deactivation constants K_d were determined for wild-type ED30 and PT35 at different ACN concentrations, in KPi (gray shaded) or TRIS (no shading) buffer. $t_{1/2}$ was calculated from the exponential model $v/v_0 = exp(-K_d t)$. The correlation coefficient R^2 is also shown for each model. The model could not be fitted to those residual activity plots that did not show a decrease in activity ("no fit"). Measurement series, where no activity at all was detected, are not shown or left blank (–).

	ACN	$t_{1/2}$ [h]	K_d [10^{-2} h^{-1}]	R^2
ED30	0 %	42	1.6	0.95
		17	4.1	0.89
	5 %	7	9.8	0.99
		1	69	0.94
	10 %	14 min	288	1.00
		–	–	–
PT35	0 %		– no fit –	
			– no fit –	
	20 %	133	0.52	0.93
		153	0.45	0.94
	50 %	35	2.0	0.98
		–	–	–

observed for PT35 in TRIS than in KPi. However, because the respective half-lives were beyond the total measurement time of 96 h, they were considered less reliable. The same was true for the determination of the half-life value of ED30 at 10 % ACN and KPi buffer. Since the measurements were performed hourly, the decline in activity with a half-life value of 14 min was too rapid to obtain accurate results. Since the half lives of 7 and 35 h obtained with KPi at 5 % (ED30) and 50 % (PT35) ACN, respectively, were within the overall measurement time, these conditions were considered well suited for subsequent stability approaches with the enzyme mutants, especially since the exponential models show very high correlation coefficients (>0.98).

3.3.3 Thermostability of the wild type

Many previous studies reported a positive correlation between enzyme thermostability and tolerance to organic solvents.[89–93] Therefore, the temperature of half-inactivation T_{50} was determined for the wild-type ED30 and PT35. The enzymes were incubated in KPi or TRIS buffer at different temperatures for 1 h, followed by activity measurements by the *p*NPB assay. The T_{50} values were determined from a polynomial model

$v/v_{max} = a + b \cdot T + c \cdot T^2 + d \cdot T^3 + ...$ fitted to the relative activity (v/v_{max}) data (figure 3.16), where v represents the enzyme activity after 1 h incubation at a given temperature T and v_{max} is the maximum activity measured.

The T_{50} values were determined for ED30 and PT35 after incubation in KPi and TRIS buffer (table 3.2). ED30 stability was studied for a temperature range of 30–50 °C (figure 3.16). After incubation in KPi buffer, a steep drop in ED30 relative activity was observed between 30 and 33°C, followed by a delayed decline, which resulted in a T_{50} of 35.4 °C. In comparison, a slightly lower T_{50} of 33.7 °C was obtained after incubation in TRIS buffer. Complete enzyme deactivation was observed from a temperature of 38 (TRIS) and 42 °C (KPi). These results strikingly disagree with those of Tao *et al.*, where ED30 showed >80 % of the maximum activity after 1 h incubation at 40 °C in TRIS buffer.[140] However, no T_{50} was reported and only 5 % relative activity at 45 °C and complete enzyme deactivation at 50 °C after 1 h incubation was observed. Nevertheless, ED30 showed markedly lower thermostability in this study, as the T_{50} must be lower by more than 6.3 °C compared to the results of Tao *et al.*.

A temperature range of 40–80 °C was selected for the thermostability determination of PT35. A similar manner of decline in relative activity was shown for PT35 in KPi and TRIS buffer as the temperatures increased. However, a steeper drop in relative activity was observed for PT35 when incubated in KPi buffer, resulting in a lower T_{50} of 56.1 °C compared to 60.7 °C for incubation in TRIS buffer. However, the relative activity of PT35 declined very slowly with increasing temperatures in both buffers and even remained present at 80 °C with 16 (KPi) and 29 % (TRIS). This may indicate either moderate stability at very high temperatures or the ability to refold.

Previous work by Bollinger *et al.* revealed a T_{50} of 56 °C, which was determined under similar conditions (100 mM potassium phosphate buffer, pH 7.2) and is consistent with the findings of this study.[54] Another study reported a T_{50} of around 48 °C after 1 h incubation in KPi buffer.[136] On the basis of these results, PT35 was categorized as a mesophilic esterase (optimal temperature ≈ 30–50°C).[215] Despite a higher T_{50} value, the findings of this study support the classification of PT35 as a mesophilic esterase. To examine the correlation between thermostability and tolerance to ACN, the half-life and T_{50} values of ED30 and PT35 were compared (table 3.2). In both buffers, PT35 showed much higher half-lifes (>130 h vs. 0 h in 20 % ACN) and T_{50} values ($\Delta T_{50} > 20$ °C) than ED30. These results support the assumption that higher organic solvent tolerance often comes along with higher thermostability. The comparison of the two buffers used revealed that ED30 showed higher resistance to ACN and temperatures when incubated in KPi buffer. Although, PT35 was also markedly more resistant to 50 % ACN in KPi buffer ($t_{1/2} = 35$ h vs. 0 h), it showed higher thermostability in TRIS buffer ($\Delta T_{50} = 4.6$ °C).

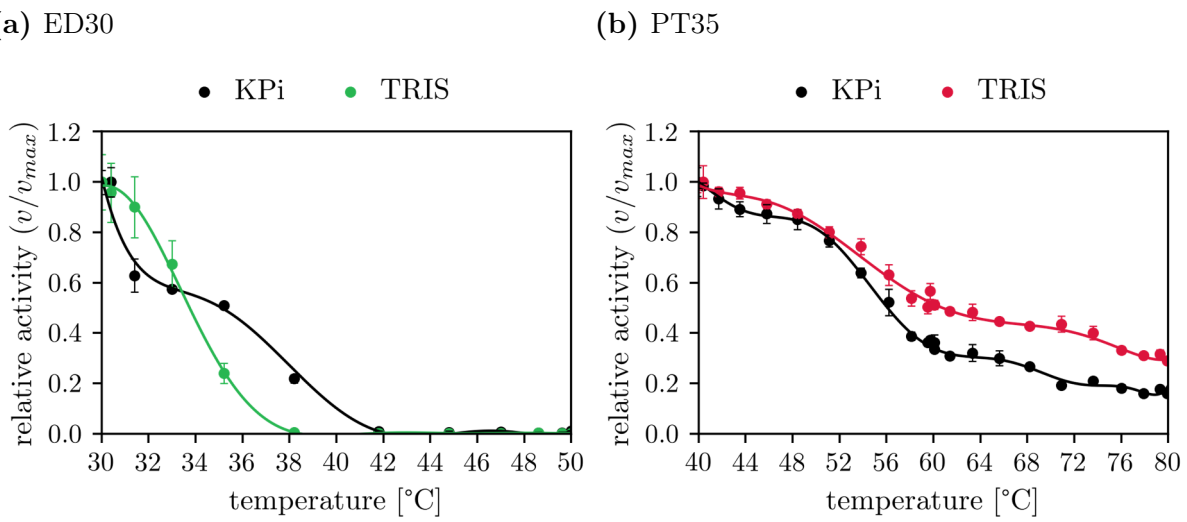

Figure 3.16: Thermally induced inactivation of the wild type in different buffers. The relative activity of the wild-type (a) ED30 and (b) PT35 was detected after incubation at different temperatures for 1 h in KPi buffer and TRIS buffer. The activity was measured under standard esterase activity assay conditions using 10 µl purified enzyme in 200 µl of TRIS or KPi buffer containing 1 mM *p*NPB. The relative activity (v/v_{max}) was calculated by comparing the enzyme activity (v) with the maximal activity (v_{max}) after incubation. The results are presented as the means±SEM from triplicate measurements.

Previous studies could show that PT35 was strongly inhibited in the presence of several divalent cations which indicates a sensitivity of PT35 to positive charges.[136] Since the amino group of Tris(hydroxymethyl)aminomethane (Tris) is partially protonated at pH 8 and thus positively charged, it could be assumed that Tris neutralizes and thus impairs the negatively charged surface of PT35 that is considered to be protective against organic solvents. On the other hand, no differences in PT35 stability between TRIS and KPi buffer were observed at lower concentrations of ACN. Raman spectroscopy of Tris and ACN have shown that ACN formed complexes with deprotonated amino groups of Tris but did not interact with protonated ones, which increased the relative concentration of protonated Tris molecules.[216] This interaction may explain the markedly faster decline in activity with increasing ACN concentrations compared to KPi buffer.

3.3.4 Determination of the melting temperature

A popular method to analyze enzyme stability is to study the thermal tolerance, i.e. the unfolding of an enzyme due to increasing temperature. Thermal tolerance can be distinguished from thermostability by its relation to thermodynamic stability (tendency to unfold and refold) while thermostability refers to kinetic stability (ability to avoid irreversible denaturation). A measure of thermal tolerance is the melting temperature T_m,

at which half of the protein is in the unfolded state.[59] Additives as organic solvents can cause disruption of the hydrophobic protein core and, hence, shift the equilibrium towards the unfolded state. This effect can be investigated by the determination of T_m in the presence of organic solvents or other denaturing compounds.[59,217] Therefore, the thermal unfolding of wild-type ED30 and PT35 was performed in KPi and TRIS buffer as well as in the presence of ACN using nano differential scanning fluorimetry (nanoDSF). ACN concentrations of 5 % for ED30 as well as 20 and 50 % for PT35 were chosen. The protein melting was measured between 20 and 95 °C with a heating rate of 1 °C per minute.

The resulting melting scans are displayed as the fluorescence ratio (350 nm/380 nm) and the first derivative, at which the peaks indicate the melting temperatures T_m (figure 3.17). The melting scans of ED30 and PT35 in absence of ACN differed mostly in their melting onset and melting point. The melting of ED30 started at 40.2 and 36.9 °C with melting points of 45.6 and 43.3 °C in KPi and TRIS buffer, respectively, which indicates an accelerated unfolding in TRIS buffer. In contrast, a slow melting of PT35 was already observed from about 30 °C, while similar melting points were obtained for KPi and TRIS buffer with 64.8 and 65.1 °C, respectively. PT35 showed additional small peaks prior to the melting point, shifted between the scan in KPi (48.3 °C) and in TRIS (44.6 °C) buffer. These findings may indicate either specific structures of PT35 that were more sensitive to lower temperatures and whose unfolding was triggered differently in the respective buffers, or impurities that have lower melting points. The latter can be linked to the results of SDS-PAGE, which showed different bands indicating impurities (figure 3.14). A nanoDSF measurement requires protein samples of high purity to achieve an accurate melting curve and to determine a reliable T_m value.[217] Therefore, destabilizing impurities could also explain the lower stability of PT35 compared to the results from previous studies that reported a T_m of around 70 °C in KPi.[136]

Both enzymes were destabilized by the presence of ACN, as indicated by decreased melting temperatures. The T_m of ED30 at 5 % ACN dropped to 41.6 °C (ΔT_m = 4.0 °C) and 40.8 °C (ΔT_m = 2.5 °C) for KPi and TRIS buffer, respectively. Due to the immediate protein aggregation, apparent by white deposits in the capillaries, no melting scan could be performed for PT35 at 50 % ACN in TRIS buffer. This observation supports the hypothesis that Tris molecules may compensate for the negative surface charge of PT35, which facilitates attack by ACN and finally leads to denaturation and aggregation of the enzyme. In 20 % ACN, PT35 showed melting temperatures of 59.0 °C (ΔT_m = 5.8 °C) and 61.2 °C (ΔT_m = 3.9 °C) for KPi and TRIS buffer, respectively. The thermally induced unfolding is even more accelerated by higher ACN concentrations. 50 % ACN caused the melting temperature of PT35 to drop to 49.2 °C (ΔT_m = 15.6 °C) for KPi buffer. A previous study reported a linear decline in melting temperatures by increasing ACN

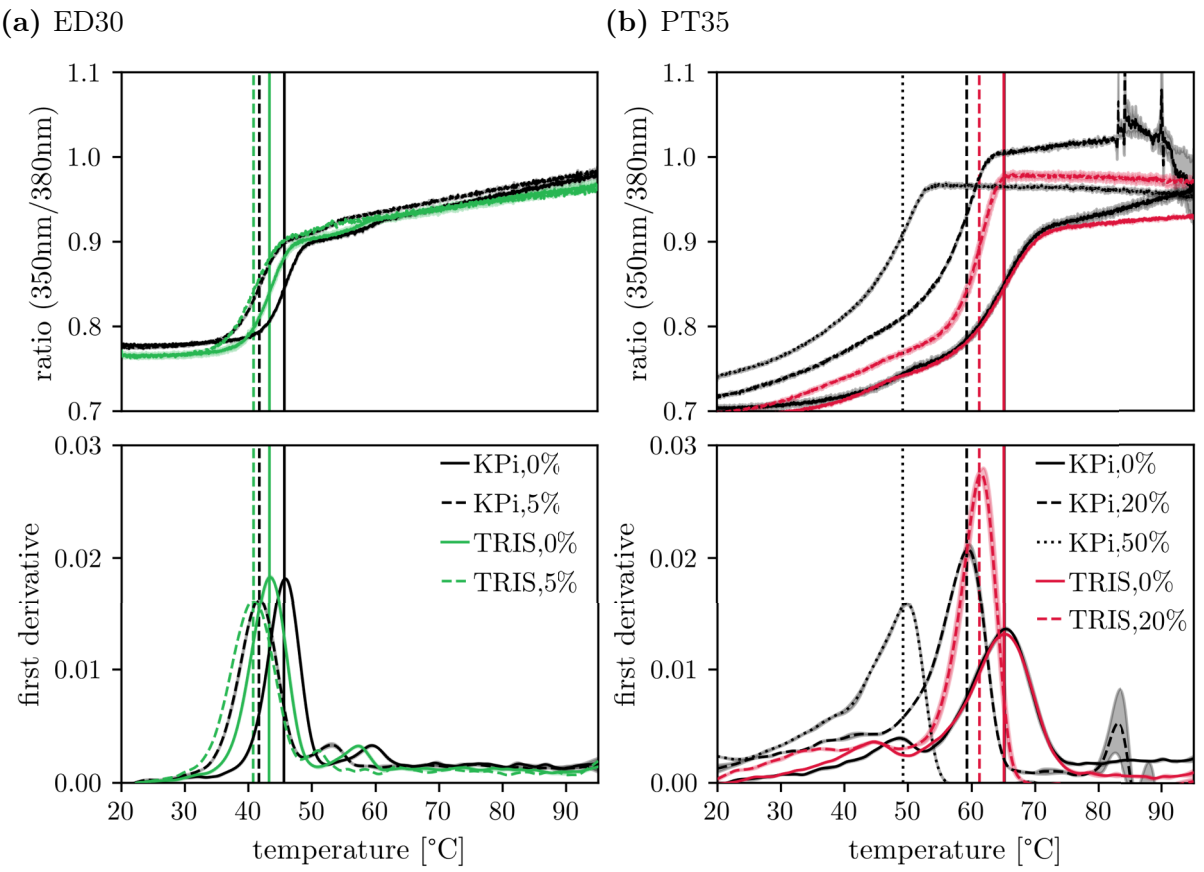

Figure 3.17: Thermally induced unfolding of the wild type enzymes in different buffers and for varying acetonitrile concentrations. The T_m of wild-type ED30 and PT35 was determined by nanoDSF in the presence of 5 % ACN and in 20 and 50 % ACN, respectively, as well as in two different buffers, KPi and TRIS. The protein melting was followed from 20 to 95 °C with a heating rate of 1 °C per minute. Results are presented as mean±SEM from triplicate measurements, with the SEM shown as a shaded area. The calculated T_m values are visualized as vertical lines.

concentrations for different enzymes.[217] A similar manner was observed for the reduced melting temperatures at 20 and 50 % ACN. However, to confirm a linear decrease in melting temperature, one would need to investigate the thermal unfolding of PT35 at additional ACN concentrations.

Comparison of both enzymes with respect to the effects of ACN on thermal stability showed a relatively small ΔT_m of PT35 at 20 % ACN compared to that of ED30 at lower ACN concentration (5 %), indicating higher stability of PT35 in ACN. The obtained melting temperatures T_m were compared to the half-lives $t_{1/2}$ and half-inactivation temperatures T_{50} from organic solvent or thermally induced enzyme deactivation measurements (table 3.2). Both enzymes showed T_m values about 5–10 °C higher than those of T_{50}. The active site may have been affected by the temperatures prior to complete unfolding, which could explain enzyme inactivation at lower temperatures. Conclusive

Table 3.2: Thermodynamic and kinetic stability measures of the wild type enzymes for different ACN concentrations and buffers. The half life ($t_{1/2}$), half-inactivation temperature (T_{50}) and the melting temperature (T_m) were determined by organic solvent stability assay, thermal stability assay and nanoDSF. The stability constants were determined for wild-type ED30 and PT35 at different ACN concentrations, while using KPi (gray shaded) or TRIS (no shading) buffer. No half life and melting temperature could be determined for PT35 in 50 % ACN (–). The T_{50} was determined only in aqueous solution without ACN.

	ACN	$t_{1/2}$ [h]	T_{50} [°C]	T_m [°C]
ED30	0 %	42	35.4	45.6
		17	33.7	43.3
	5 %	7		41.6
		1		40.8
PT35	0 %	no fit	56.1	64.8
		no fit	60.7	65.1
	20 %	133		59.0
		153		61.2
	50 %	35		49.2
		–		–

results were obtained for ED30 when comparing the effects on stability caused by the two buffers used. ED30 was more stable in KPi buffer than in TRIS buffer, regardless of the method used to determine stability in the presence or absence of ACN. In contrast, PT35 showed inconsistent results according to its stability in the different buffers. While inactivation of PT35 in KPi was observed at lower temperatures, no differences between both buffers were found with respect to unfolding. In addition, at 20 % ACN, higher $t_{1/2}$ and T_{50} values were obtained for TRIS buffer. However, PT35 was immediately inactivated and denatured at 50 % ACN in TRIS buffer, while a $t_{1/2}$ of 35 h and a T_m of 49.2 °C was obtained for KPi buffer.

In summary, the hypothesis of PT35 as a stable esterase and ED30 as its unstable counterpart was strongly confirmed in this study. PT35 not only showed higher tolerance to ACN but it was also more thermostable and resistant to thermal unfolding, even in the presence of ACN. Therefore, PT35 and ED30 are appropriate candidates for studying the structural drivers of an esterase organic solvent tolerance using mutagenesis. Based on the wild-type analysis, optimal conditions and measures could be determined for comparing the ACN tolerance of the mutants. These include first the usage of KPi buffer for both enzymes, and second ACN concentrations of 5 % for ED30 as well as 20 and 50 % for PT35 to determine half-lives and melting temperatures in the presence of ACN.

3.4 Investigation of the mutant stability

Previous studies suggested that the highly negatively charged surface of PT35 may be associated with its tolerance to organic solvents.[54,136,139] Due to their strong water-binding properties, the negatively charged residues might strengthen the hydration shell, which protects PT35 against attack by organic solvents.[66,122,123] To test this hypothesis, a mutagenesis strategy was established that aimed to generate a more stable ED30 and a less stable PT35. On the basis of different criteria, selected substitutions were combined to triple and multi-mutants (section 3.1). Special attention was paid to the effects of mutation on the hydration shell and the surface charge.

One triple mutant of both ED30 (K124D/R211E/K337E) and PT35 (E199R/E328K/D439K) was produced in *E. coli* BL21(DE3), purified by IMAC and subsequently tested for organic solvent tolerance. Furthermore, each substitution of the triple mutants was separately produced and tested as a single mutant. The yield and purity of the mutant enzymes were verified by SDS-PAGE (figure A9). The multi-mutants could not yet be produced due to the lack of overexpression.

In order to investigate the effect of substitutions on the organic solvent tolerance of the enzymes, stability assays were performed at ACN concentrations of 5 % (ED30 mutants) and 50 % (PT35 mutants). The enzyme mutants were incubated at 30 °C and the residual activity was measured after different incubation times by the *p*NPB assay. To facilitate estimation of the change between mutant and wild-type stability in the presence of organic solvents, the half-life ($t_{1/2}$) of deactivation was determined for each variant. The half-life values were obtained from the exponential model $v/v_0 = exp(-K_d t)$ fitted to the residual activity (v/v_0) plots (figure 3.18).

The inactivation of each mutant and the respective wild type in presence of ACN was compared (figure 3.18). The ED30 mutants K124D, R211E, K337E, and the triple mutant (tmut) showed similar or enhanced stability compared to the wild type (wt) in 5 % ACN, and they maintained 13, 13, 34, and 24 % of their activity after 24 h incubation, respectively, while the wild type showed only 9 % residual activity under the same conditions. Opposing effects were observed for the PT35 mutants. Here, mostly similar or reduced stability compared to the wild-type (wt) PT35 was observed for the mutants E199R, E328K, D439K and the triple mutant (tmut) in 50 % ACN, which showed residual activities of 15, 18, 6 and 12 %, respectively, after the maximal incubation time (96 h), while the wild type remained 18 % activity.

The half-life and deactivation constants were determined for the wild-type and its mutants (table 3.3). Due to the large correlation coefficients (> 0.94) for the exponential models, a high reliability of the calculated half-lives is assumed. The ED30 single mutant K124D and R211E hardly differ in their half-lives from that of the wild type, indicating no

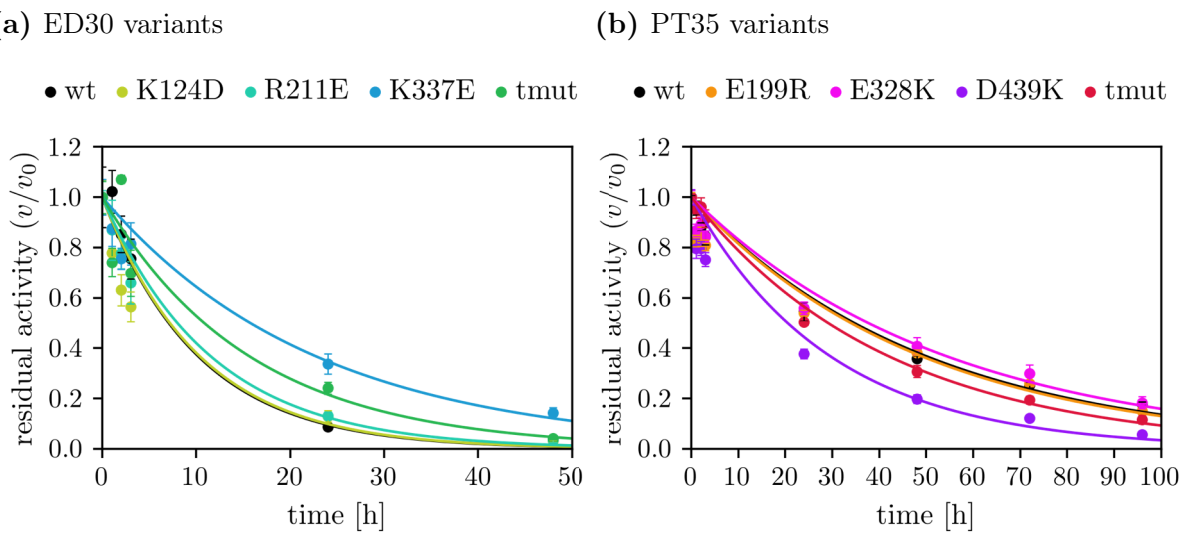

Figure 3.18: Inactivation of the wild types and the mutants by acetonitrile. The residual activity of each (a) ED30 and (b) PT35 wild type (wt), single and triple mutant (tmut) was detected after different incubation times in the presence of 5 and 50 % ACN, respectively. The activity was measured under standard esterase activity assay conditions using 10 µl purified enzyme in 200 µl of KPi buffer containing 1 mM *p*NPB. The residual activity (v/v_0) was calculated by comparing the enzyme activity at 0 h (v_0) with the activity after different incubation times (v). The results are presented as the means±SEM from triplicate measurements.

change in organic solvent tolerance compared to the wild type. The most striking effect was shown for the single mutant K337E which exhibited 2.3-fold improved stability above the wild type. Enhanced tolerance to ACN was also observed for the triple mutant with a 1.6-fold higher half-life value. However, many studies reported stabilizing additive effects by combining single mutations.[111,126,210] This would suggest a similar half-life value of the triple mutant and the K337E single mutant. Therefore, the other substitutions K124D and R211E may have diminished the stabilizing effect of K337E in a way that had no impact on the overall enzyme stability in the respective single mutants.

The PT35 single mutants E199R and E328K expressed similar or slightly improved stability with respect to the wild type. The most notable change was observed for the D439K mutant, which showed a 0.4-fold decreased tolerance to ACN. As with ED30, no additive effect was found for the PT35 triple mutant, whose half-life value is only slightly decreased (<0.2-fold) compared to that of the wild type. This may indicate that the destabilizing effect of D439K was suppressed by one or both of the E199R and E328K substitutions in the triple mutant.

To examine the reproducibility of these findings, nanoDSF was performed with each mutant and wild type. The thermally induced unfolding of the protein in the presence of ACN further enables the differentiation between thermally and organic solvent tolerant enzymes. The protein melting was measured at 5 % (ED30 variants) as well as 20 and

Table 3.3: Half-lives and inactivation constants for the wild types and the mutants in acetonitrile. The half-life value ($t_{1/2}$) was calculated from the exponential model $v/v_0 = exp(-K_d t)$ with the deactivation constant K_d for each ED30 and PT35 wild type (wt), single and triple mutant (tmut) at 5 and 50 % ACN, respectively. The correlation coefficient R^2 is shown in addition for each model.

	VARIANT	$t_{1/2}$ [h]	K_d [10^{-2} h^{-1}]	R^2
ED30 (5 % ACN)	wt	7	9.8	0.99
	K124D	7	9.6	0.95
	R211E	8	8.7	0.99
	K337E	16	4.4	0.97
	tmut	11	6.4	0.96
PT35 (50 % ACN)	wt	35	2.0	0.98
	E199R	34	2.0	0.95
	E328K	38	1.8	0.97
	D439K	21	3.4	0.96
	tmut	30	2.4	1.00

50 % ACN (PT35 variants) between 20 and 95 °C with a heating rate of 1 °C per minute. The melting temperatures (T_m) were determined for the wild types and their mutants in absence and presence of ACN (figure 3.19). T_m represents the temperature at which half of the protein is in the unfolded state. The higher the T_m, the more tolerant the enzyme is to thermally induced unfolding.

The results of ED30 variants at 0 and 5 % ACN showed very minor changes in stability for the single mutants K124D and R211E, while enhanced stability was observed for K337E and the triple mutant. The T_m at 0 and 5 % ACN of the ED30 single mutant K337E was improved by 1.7 and 1.6 °C, respectively, while the triple mutant showed only a slightly enhanced melting temperatures of 0.6 and 0.5 °C under the same conditions. Regardless of the ACN concentration, the PT35 single mutants E199R and E328K exhibited similar stability compared to the wild-type PT35, while the T_m of D439K was reduced by −1.0, −1.1 and −1.5 °C in 0, 20 and 50 % ACN, respectively. However, differences in the stability of the triple-mutant PT35 between the absence and presence of ACN were detected. In absence of ACN, the triple mutant showed similar thermal tolerance as the wild type, while its stability was reduced by −1.5 and −1.4 °C in 20 and 50 % ACN, respectively.

The change in half-life ($\Delta t_{1/2}$) and melting temperature (ΔT_m) upon substitution were compared for each ED30 and PT35 mutant (table 3.4). The results of the nanoDSF approach were found to support those of the stability assay in organic solvents. Single mutants of ED30 and PT35 that exhibited similar half-lives as the wild type also showed minor or no differences in T_m. However, the PT35 single mutant E328K, which showed

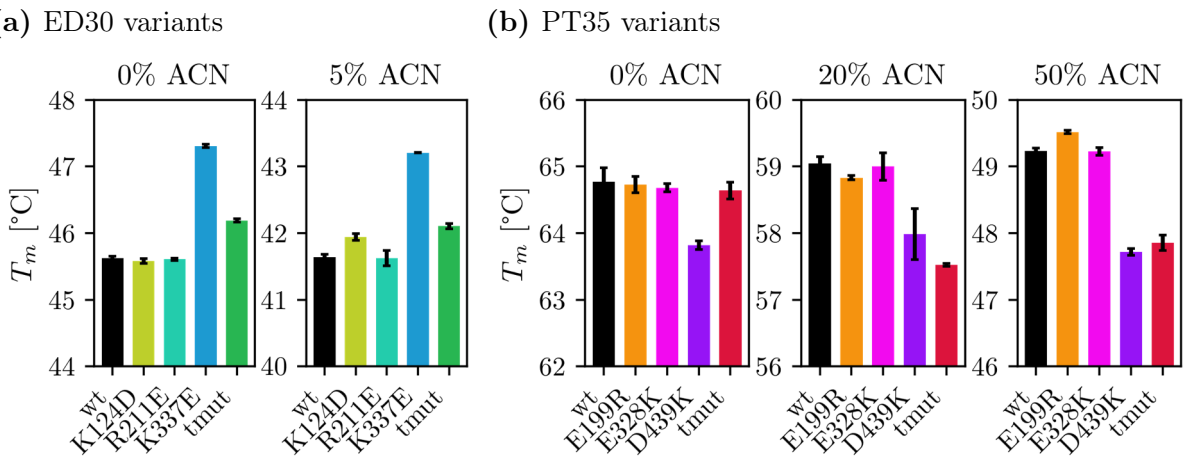

Figure 3.19: Protein melting temperatures of the wild types and mutants in the presence of ACN. The T_m of each (a) ED30 and (b) PT35 wild type (wt), single and triple mutant (tmut) was determined by nanoDSF in the presence of 0 and 5 % and 0, 20 and 50 % ACN, respectively. The protein melting was followed from 20 to 95 °C with a heating rate of 1 °C per minute in KPi buffer.

slightly enhanced resistance to inactivation by ACN, showed no change in melting temperature at all.

The higher stability of K337E compared to the triple mutant was indicated both by higher $\Delta t_{1/2}$ and ΔT_m values. The high ΔT_m value in the absence of ACN led to the assumption that the structural changes due to the K337E substitution resulted not only in increased resistance to ACN but also in improved thermal tolerance.

The PT35 single mutant D439K showed negative ΔT_m values, which was consistent with $\Delta t_{1/2}$ indicating a reduced organic solvent tolerance of D439K. Due to the decreased melting temperature at 0 % ACN, a lower thermal tolerance is assumed in addition to the higher sensitivity to ACN. In contrast, the triple mutant was similar stable as the wild type in the absence of ACN but showed reduced stability in 20 and 50 % ACN. While D439K showed lower half-lives, i.e. higher sensitivity to ACN induced inactivation, than the PT35 triple mutant, similar resistance to thermally induced unfolding was observed for both mutants in the presence of ACN. These findings may indicate that the combination of E199R, E328K, and D439K resulted in a decrease of the triple-mutant stability in organic solvents but not in thermal tolerance.

In summary, the most striking improvement in thermal and ACN tolerance was observed for the ED30 single mutant K337E, while the highest reduction in stability was observed for the PT35 single mutant D439K. The ED30 and PT35 triple mutants, however, were only slightly more stable and unstable in ACN, respectively. While the triple mutant of ED30 showed also higher thermal tolerance, no differences were observed for the triple mutant of PT35 compared to the wild-type resistance to thermally induced

Table 3.4: Change in stability of ED30 and PT35 mutants compared to the wild type. The deviations of the half-lives ($\Delta t_{1/2}$) of the mutants from the wild-type ED30 and PT35 were calculated for the measurements in 5 and 50 % ACN, respectively. The change of the melting temperature (ΔT_m) between mutant and the wild type is shown for 0 and 5 % (ED30 variants) and 0, 20 and 50 % ACN (PT35 variants).

ED30	$\Delta t_{1/2}$ **[h]**	ΔT_m **[°C]**		**PT35**	$\Delta t_{1/2}$ **[h]**	ΔT_m **[°C]**		
		0 %	5 %			0 %	20 %	50 %
K124D	0	0.0	0.3	E199R	-1	0.0	-0.2	0.3
R211E	1	0.0	0.0	E328K	3	-0.1	0.0	0.0
K337E	9	1.7	1.6	D439K	-14	-1.0	-1.1	-1.5
tmut	4	0.6	0.5	tmut	-5	-0.1	-1.5	-1.4

unfolding. To elucidate the molecular basis of the enhanced tolerance and the differences between single and triple mutant stability, the MD trajectories were analyzed to identify possible stabilizing factors.

3.5 MD analysis based on experimental results

Experimental results showed that one single mutant of each ED30 (K337E) and PT35 (D439K) exhibited distinct deviations from the wild type in terms of organic solvent tolerance but also thermal tolerance. It was also found that the effects of substitution on the respective triple-mutant stability were less pronounced. To identify the structural features causing these phenomena, the MD simulations of the respective mutants were analyzed with respect to newly formed or disrupted salt bridges and associated structural changes. Since salt bridges were found to be abundant in thermophilic proteins, a contribution to thermal stability was suggested.[218,219] In addition, engineering of salt bridges has yielded enzymes with increased resistance to organic solvents[105] and improved thermal stability[220]. To elucidate whether the formation or disruption of salt bridges caused the improved or decreased stability of the current mutants, the interactions of the substituted residues with their surroundings were analyzed.

Investigation of the K337E substitution in ED30 mutants The main interactions of the K337E substitution of ED30 were observed between the residues K337-D334 of the wild type, and between E337-R340 of the single and triple mutants (figure 3.20). The minimum distance between the sidechain nitrogen and oxygen atoms of the interacting residues was computed (figure 3.21).

The minimum distance plot of the wild-type ED30 indicated relatively weak interactions between K337-D334 in water and 80 % ACN, since the minimum distance varied

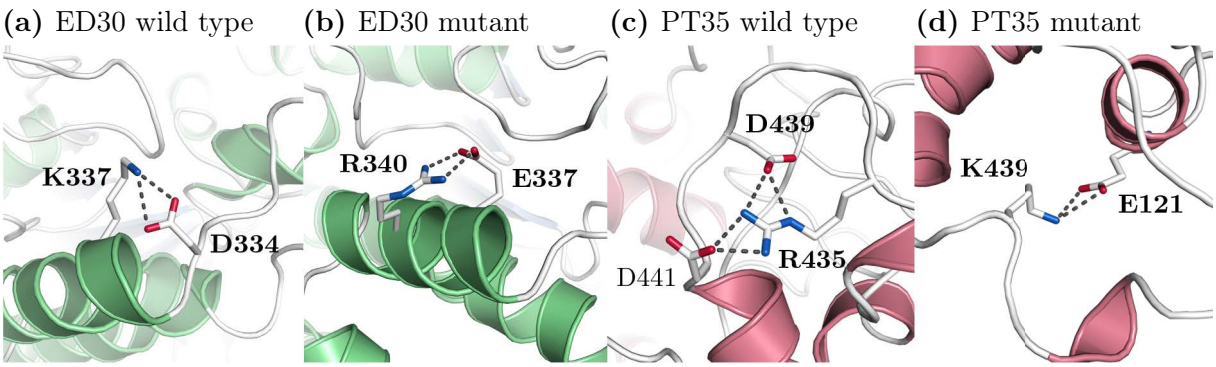

Figure 3.20: Salt bridges formed in the wild type, triple mutant and single mutant of ED30 (K337E) and PT35 (D439K). The residues of the ED30 (a) wild type (K337-D334) and (b) single mutant (E337-R340), representatively shown for single and triple mutant, and the residues of PT35 (c) wild type ([D439-R435]-D441) and (d) single mutant (K439-E121) forming salt bridges are highlighted in different colors as sticks with carbon (white), oxygen (red) and nitrogen (blue). Electrostatic interactions are indicated by dashed lines. The secondary structure elements are colored as followed: helices (light red), β-sheet (light blue), coil (white).

steadily between 2.7 and 10 Å. The residues E337-R340 of the triple and single mutants, by contrast, showed stronger interaction in water with almost constant minimum distances around 2.7 Å. In the presence of ACN, the minimum distance between E337-R340 of the triple mutant increased to about 5 Å starting from 350 ns, while exhibiting a large variance. These results indicated that the substitution of K337E resulted in the formation of a new salt bridge. This salt bridge may have had stabilizing effects on ED30, which were more pronounced in the single mutant than in the triple mutant. However, since no major differences were observed between the two mutants with respect to the minimum distance, other structural changes must have suppressed the stabilizing effect of K337D. It is assumed that one or both of the other substitutions, K124D and R211E, of the triple mutant caused these structural changes.

To analyze overall structural changes of the ED30 enzymes, snapshots of the wild-type, single-mutant and triple-mutant structure were taken at 100 ns and 500 ns (if available) of the simulations in water only (figure A10). The structure of the wild type revealed a denser conformation than that of the triple mutant at both times, 100 ns and 500 ns. The triple mutant was even more expanded after 500 ns. The single mutant showed a slightly more compact structure than the triple mutant. However, the snapshots represent only one of the three trajectories in the case of the wild type and the triple mutant. Moreover, they are inconsistent with the Rgyr values, which were decreased for the triple mutant compared to the wild type.

The ΔRMSF_{T-S} (RMSF(triple mutant)-RMSF(single mutant))was calculated to determine the change in flexibility upon mutation from single to triple mutant in water

(figure 3.21). The flexibility is considered changed when $|\Delta\text{RMSF}_{T-S}| > 2\,\text{Å}$. Minor overall changes in flexibility were observed between triple and single mutant. Only two regions, eco3 (residues 400-410) and ecoh4 (420-440) showed higher flexibility and rigidity in the triple mutant, respectively. It could be assumed that these changes in flexibility were induced by one or both of the other substitutions in the triple mutant. However, the simulation time of 100 ns and the single determination of the single mutant are not sufficient for reliable analyses.

Investigation of the D439K substitution in PT35 mutants The wild type and the D430K single mutant of PT35 were found to exhibit interactions between the residues D439-R435 and K439-E121, respectively (figure 3.20). A further interaction between R435 and D441 was observed and presented. The minimum distance between the sidechain nitrogen and oxygen atoms of the interacting residues was computed (figure 3.21). The results showed a relatively strong interaction between D439-R435 for the wild type in water, which is less pronounced in the presence of ACN. No interaction between K439-E121 was observed in the triple mutant, whereas this contact was favored in the single mutant with increasing time. These findings suggest that the salt bridge D439-R435 in the wild type was disrupted by substitution. After the maximum simulation time of 100 ns and 500 ns in water only, snapshots of the wild-type, triple-mutant and single-mutant structure were taken (figure A11). The findings revealed a shift of K439 and the connected α helix towards the E121 residue in the triple mutant and in the single mutant, where it is even more pronounced. After 500 ns simulation, the coil of the wild type containing D439 interacting with R435 was slightly shifted upward, whereas the position in the triple mutant remained constant. These findings may indicate that the substitution D439K led to a relative huge rearrangement around this residue in the single mutant. To verify whether these structural changes are more prominent in the single mutant than in the triple mutant, more comprehensive MD simulations and analyses of the single mutant and the triple mutant are needed.

The ΔRMSF_{T-S} was determined in order to analyze the differences between triple-mutant and single-mutant flexibility (figure 3.21). Enhanced rigidity was observed for the triple-mutant PT35 in pco1 (300-320), pcoh5 (430-445) and pco6 (465-475) while enhanced flexibility was displayed for pcoh2 (320-370). The increased rigidity of pcoh5 harboring the substitution D439K may indicate that pcoh5 was restrained from shifting toward E121. It could be assumed that the other regions, pco1 and pco6, gained in rigidity due to the other substitutions in the triple mutant (E199R, E328K) and induced the restricted flexibility of pcoh5. To verify this hypothesis, the interactions and motions of the respective regions need to be analyzed in detail, e.g. by dynamic cross-correlation

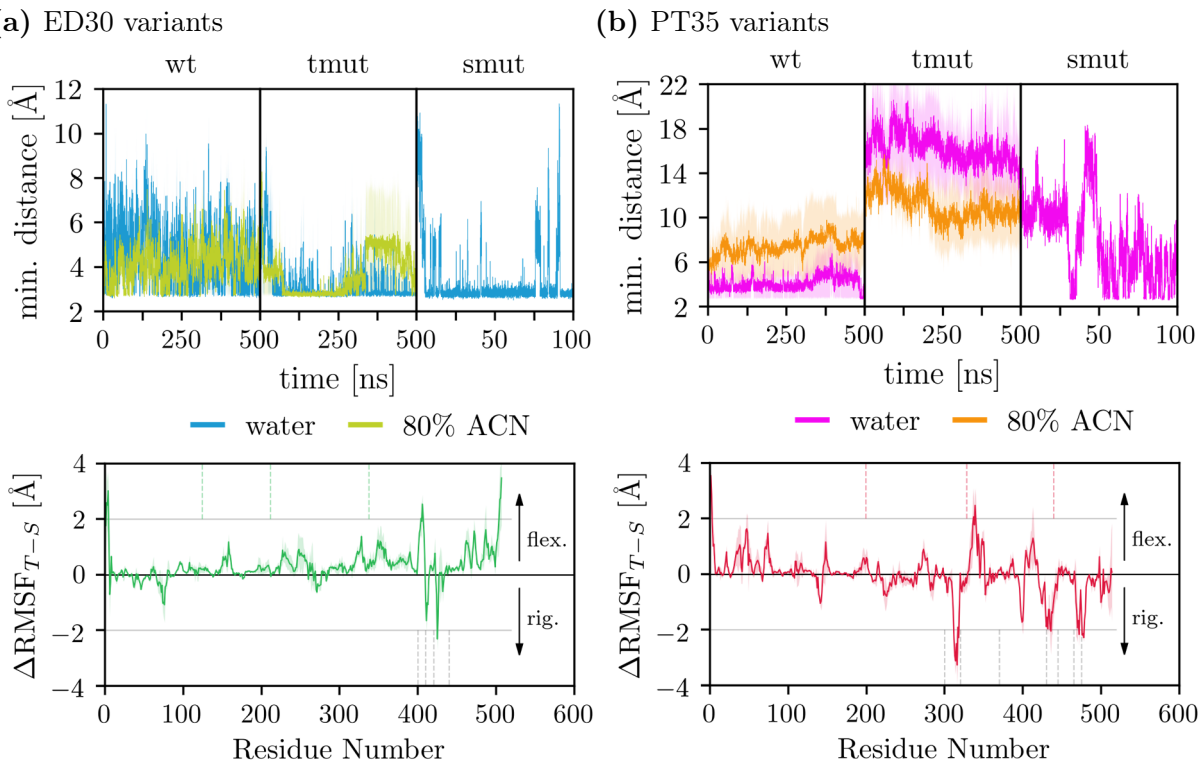

Figure 3.21: Distance and RMSF analyses to unveil substitution-dependent structural changes in the single and triple mutant resulting from K337E in ED30 and D439K in PT35. The minimum distance between sidechain nitrogens and oxygens of (a) ED30 wild type (wt, K337–D334), triple mutant (tmut, E337–R340) and single mutant (smut, E337–R340) and (b) PT35 wt (D439-R435), tmut (K439-E121) and smut (K439-E121) was computed for the simulation in water only and 80 % ACN. The ΔRMSF_{T-S} was determined for the single (100 ns) and triple mutant (500 ns) simulations in water and displays the change in flexibility of the triple mutant and the single mutant. Residues with a changed flexibility ($|\Delta\text{RMSF}_{T-S}| > 2$ Å) are marked by a horizontal gray line. The positions of the substitutions of the triple-mutant ED30 (K124D, R211E, K337E) and PT35 (E199R, E328K, D439K) are marked by vertical green or red dotted lines and the boundaries of the flexible areas are indicated by vertical gray dotted lines. The arrows at the right-hand side indicate an increase in flexibility or rigidity. The results for the triple mutants are presented as the mean±SEM from triplicate simulations, while only a single simulation was performed for the single mutants.

maps (DCCMs), which provide information about correlated and anti-correlated motions in the enzyme.

In summary, the results suggest that salt bridge formation and disruption may have contributed to the increased or decreased stability of the K337E (ED30) and D439K (PT35) single mutants, respectively. With regard to D439K, also structural rearrangement of specific regions associated with this substitution was observed. However, to identify the structural and dynamical differences causing the higher stability of the single mutant compared to the wild type and triple mutant, longer MD simulations of the single mutant also in the presence of ACN are required.

4 Conclusion and outlook

In this study, a mutagenesis strategy was developed and applied to test the hypothesis that the organic-solvent tolerance of an enzyme can be improved by surface charge and hydration shell engineering. The usage of enzymes in organic solvents has many advantages in industrial processes, thus a generally applicable strategy for enzyme stability engineering in organic solvents is of high interest.

The mutagenesis strategy was based on previous studies, where the extremely negatively charged surface of PT35 was associated with its remarkable tolerance to various organic solvents.[136,139] The surface of PT35 was compared with that of the unstable and less negatively charged ED30 to identify mutation candidates for substitutions of charged and polar residues. The most pronounced effect on surface potential change was shown by substitution of opposing charged residues allowing a net charge deviation of $\pm$ 2. Moreover, the introduction and substitution of negatively charged residues were found to allow for the highest increase and decrease of residue hydration in ED30 and PT35, respectively, which is in line with previous studies reporting a stronger binding of water by negatively charged residues than by positively charged or polar residues.[129,130] This suggests that negative surface charge is correlated with enhanced hydration of the enzyme. A strong hydration shell is thought to protect the enzyme from organic solvents entering and destroying the hydrophobic protein core resulting in protein denaturation.[63,64] Therefore, introduction of negatively charged residues at the enzyme surface may evolve as a promising general strategy to improve the hydration and hence the stability of enzymes in organic solvents.

In this regard, triple mutants were generated by 3 highly favorable substitutions of positively charged residues with negatively charged residues in ED30 and *vice versa* in PT35. While no significant structural changes were detected in the simulations in water and ACN for both enzymes, enhanced hydration and less interaction with ACN could be revealed for the respective substitutions of ED30 and vice versa for those of PT35. However, the hydration of the overall enzyme was not markedly changed for either ED30 or PT35 triple mutant compared to the corresponding wild type. Although the surface potential was markedly altered at the mutation sites, 3 substitutions may not be sufficient to change the overall charge and hydration of the enzyme to a significant extent. With

a net charge of −42, the PT35 triple mutant is still much more negative than the ED30 wild type (−6) and the ED30 triple mutant (−12). The importance of a large number of mutations with additive mutational effects was reported in particular by many previous surface charge engineering studies, where mutants with up to 15 substitutions were designed.[132,211,221,222]

Stronger effects were, therefore, expected for the ED30 and PT35 multi-mutants with 12 substitutions, which, with respect to the wild type, displayed a considerably more negative and more positive overall surface charge, respectively. Indeed, MD simulations of the multi-mutant ED30 showed its tendency to bind more water than the triple mutant and wild type; however, these findings were not significant. In addition, for both multi-mutant enzymes, larger changes in flexibility to the wild type were observed compared to the triple mutant. The generally conserved flexible and rigid regions of an enzyme from extreme environments are crucial for its activity and stability.[212,213,223,224] Thus, larger deviations due to substitutions may lead to inactivation. In addition, the introduction of a large number of substitutions increases the probability of misfolding due to disruption of the proper structure.[225] Since both multimutants have not yet been overexpressed, it could be assumed that misfolding have occurred and that the enzymes were degraded by the proteases of the *E. coli* quality control system.[226] However, the present results are not sufficient to clearly identify misfolding as the cause of failed expression. It is therefore necessary to test different strategies to optimize the expression conditions, e.g. by using different host strains, temperatures and media[227,228].

An experimental test system was established to study the organic solvent tolerance of the triple mutants and the corresponding single mutants of ED30 and PT35. The mutants and the wild types were analyzed by inactivation and thermally induced unfolding of the enzymes in presence of ACN, allowing the half-life and the melting temperature to be determined. The findings revealed that the ED30 and PT35 single mutants K337E and D439K showed 2.3-fold improved and 1.6-fold decreased stability to inactivation in ACN, respectively. In addition, higher resistance to thermally induced unfolding in the presence of ACN was observed with a change of the melting temperatures by +1.6 °C of K337E in 5 % ACN as well as −1.1 and −1.5 °C of D439K in 20 and 50 % ACN, respectively. The ED30 and PT35 single mutants showed also in absence of ACN increased (+1.7 °C) and decreased (−1.0 °C) stability. These results support the assumption of a positive correlation between tolerance to organic solvents and thermal stability, which has been reported frequently before.[89–93]

The triple mutants were found to exhibit lower deviations from the wild type compared to the single mutants K337E and D439K with regard to inactivation by ACN. The ED30 triple mutant exhibited a 0.4-fold improved stability, while a slight decrease of < 0.2-fold

was observed for the PT35 triple mutant. The ED30 triple mutant showed also minor increased thermal stability in the absence (+0.6 °C) and presence (+0.5 °C) of ACN, which supports the finding of the change in the half-life of inactivation. However, non-additive mutations are rather uncommon, but are suspected to occur when, e.g. two side chains, each containing a substitution, are in close contact with each other.[229] This, however, is not the case in the triple mutants investigated here. Interestingly, the PT35 triple mutant showed a similar thermal stability as the wild type in the absence of ACN, but a markedly decreased melting temperatures of −1.5 and −1.4 °C in 20 and 50 % ACN. Thus, the triple mutant exhibits similar thermal stability as the wild type, but shows a reduced resistance to thermally induced unfolding in ACN to the same extent as the D439K single mutant. This implies that negatively charged residues may play a critical role in the resistance of enzymes to organic solvents, which supports the current mutagenesis strategy. Since the most pronounced effects are shown by single mutants, it is suggested that other structural influences rather than the surface charge or hydration altered the enzyme stability. Thermal stability is often associated with stabilizing intramolecular interactions, including disulfide bonds, salt bridges and H-bonds.[97,105,218,219] The analysis of the trajectories revealed that salt bridge formation and disruption may have occurred in the K337E and D439K single mutant, respectively. Structural rearrangements were further observed for D439K, which are less pronounced in the PT35 triple mutant. However, because single mutants were only simulated for 100 ns in water, longer simulations, including simulations in ACN, are needed to analyze the causes of the altered stability of the single mutants compared to the wild types and triple mutants.

To validate the hypothesis that the negatively charged surface of PT35 is the main factor for tolerance to organic solvents, further analyses of the wild type are equally required. It is known that disulfide bonds can have highly stabilizing effects.[96,97,99,100] Since PT35 is assumed to possess 3 disulfide bonds, while all 6 Cys were predicted to be reduced in ED30, stability assays in the presence of the reducing agent DTT and ACN may provide information about the influence of disulfide bonds to the organic solvent tolerance of PT35.

It was further found that both ED30 and PT35 wild types showed reduced resistance to inactivation by ACN in TRIS buffer compared to KPi buffer. However, ED30 showed lower stability in all concentrations of ACN in TRIS, while PT35 exhibited similar stability in 0 and 20 % ACN for both buffers, but was considerably more resistant in KPi ($t_{1/2}$ = 35 h) than in TRIS buffer, where it was immediately inactivated in 50 % ACN. Previous studies using Raman spectroscopy indicated that interactions between ACN and deprotonated Tris molecules could result in an increased relative concentration of protonated Tris.[216] Since PT35 showed strong inhibition in the presence of divalent cations,

it could be assumed that it is sensitive to positive charges, which would explain the considerably lower stability at higher ACN concentrations in TRIS.[136] MD simulations of PT35 in TRIS buffer might elucidate certain interactions between TRIS and the negatively charged surface of PT35, which facilitate the deconstruction of the protein by ACN. This would support the hypothesis that the negatively charged surface of PT35 is crucial for its stability in ACN.

To summarize, two esterases, the stable PT35 and the unstable ED30, were modified by substitution of polar or charged residues in order to yield a more ACN tolerant ED30 and a more ACN sensitive PT35. A total of 17 mutation candidates located at electrostatic surface potential hot spots were substituted and analyzed using MD simulations in water. One triple mutant and one multi-mutant of each PT35 and ED30 was generated by combining substitutions based on different criteria, particularly the change of surface potential and residue hydration. The mutants were first simulated in water and 80 % ACN to analyze changes in structure and interactions to solvents, and then tested for stability in organic solvents using an experimental test system adapted to PT35 and ED30.Two single mutants, K337E of ED30 and D439K of PT35, were found to exhibit significantly higher and lower stability in organic solvents, respectively. MD analysis revealed that salt bridge formation and disruption might have caused the stability changes. It could further be shown that the introduction or substitution of negatively charged residues led to the most pronounced increase or decrease of the residues hydration shell. The introduction of negatively charged residues at the enzyme surface could be a promising mutagenesis strategy to enhance the hydration shell of a protein and thereby improve its tolerance to organic solvents.

5 Appendix

Table A1: Parameters used for MD simulations in water and in 80 % ACN. The following parameters were set for energy minimization, NVT and NpT equilibration and the production run of each wild type, triple mutant and multi mutant in water and in 80 % ACN. Single mutants and alternative-triple mutants were simulated for 100 ns in water only.

Parameters	Energy minimization	Equilibration	Production run
define	-DFLEXIBLE	-DPOSRES	-
integrator	steep	md	md
dt [ps]	0.002	0.002	0.002; 0.004 (H_2O)
nsteps	5000 (10 ps)	50000 (NVT, 100 ps); 500000 (NpT:, 1 ns)	50000000 (100 ns); 250000000 (500 ns)
comm-mode	-	Linear	Linear
nstcomm	-	100 (NVT); 1 (NpT)	100
comm-grps	-	Protein Non-Protein	Protein Non-Protein
emtol	1000	-	-
emstep	0.01	-	-
nstcgsteep	10	-	-
nstxout	-	1000 (NVT); 5000 (NpT)	0
nstvout	-	1000 (NVT); 5000 (NpT)	0
nstxtcout	-	-	10000
nstlog	-	1000 (NVT); 5000 (NpT)	500
nstenergy	-	1000 (NVT); 5000 (NpT)	500
cutoff-scheme	Verlet	Verlet	Verlet

nstlist	30	20	20
ns_type	grid	grid	grid
pbc	-	xyz	xyz
rlist	1.0	1.2	1.2
coulombtype	PME	PME	PME
rcoulomb	1.0	1.2	1.2
rvdw	1.0	1.2	1.2
DispCorr	-	EnerPres	EnerPres
fourierspacing	0.15	0.16	0.16
pme_order	6	4	4
ewald_rtol $[kJmol^{-1}Å^{-2}]$	1e-5	-	-
tcoupl	-	V-rescale	Nosé-Hoover
nsttcouple	-	-	10
nhchainlength	-	-	1
tc-grps	-	Protein Non-Protein	Protein Non-Protein
tau_t [ps]	-	0.1 0.1	0.8 0.8
ref_t [K]	-	303 303	303 303
pcoupl	-	no (NVT); Berendsen (NpT)	Parrinello-Rahman
pcoupltype	-	isotropic (NpT)	isotropic
nstpcouple	-	10 (NpT)	10
tau_p [ps]	-	1.0 (NpT)	8.0
compressibility $[bar^{-1}]$	-	4.5e-5 (NpT)	4.5e-5
ref_p [K]	-	1.0 (NpT)	1.0
refcoord_scaling	-	all (NpT)	all
gen_vel	-	yes (NVT); no (NpT)	no
gen-temp [K]	-	303 (NVT)	-
gen-seed	-	-1 (NVT)	-
constraints	none	all-bonds	all-bonds
constraint_algorithm	-	LINCS	LINCS
continuation	-	no (NVT); yes (NpT)	yes
lincs_order	-	6	6

lincs_iter	-	1	1

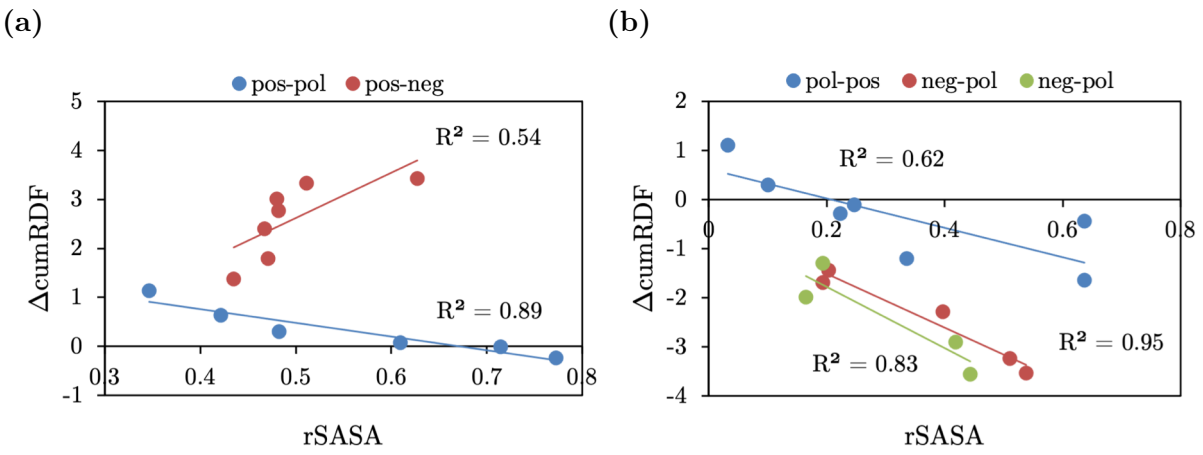

Figure A1: Correlation between rSASA and ΔcumRDF for different types of substitutions. The rSASA and ΔcumRDF values of each mutation candidate and respective substitution (except R373D, R381N and N20R) was plotted for (a) ED30 and (b) PT35. The linear regression was done separately for each substitution group: positive↔polar (pos-pol/pol-pos), positive↔negative (pos-neg/neg-pol) and negative→polar (neg-pol). The correlation coefficient (R^2) is shown for each linear fit.

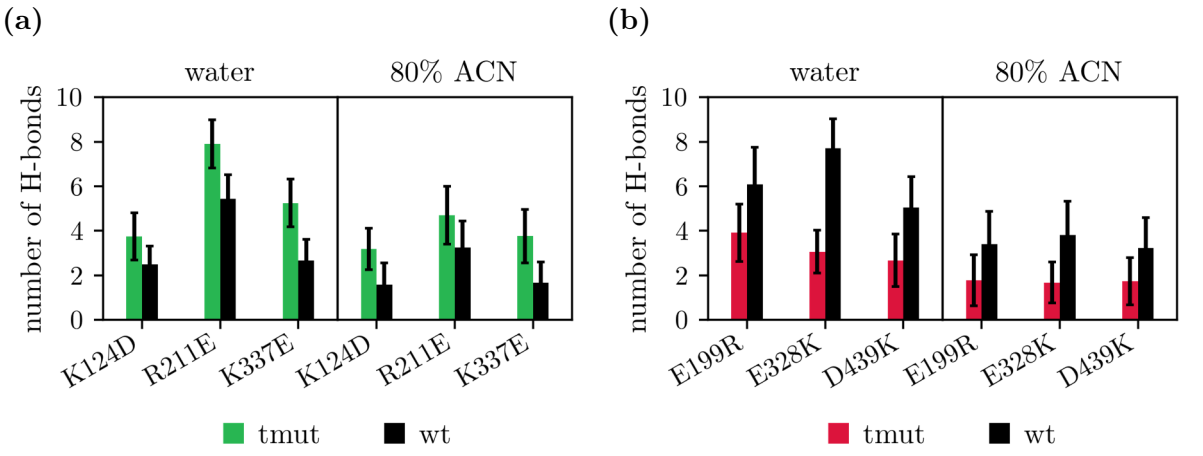

Figure A2: Hydration shell of triple-mutant ED30 and PT35 determined by the number of H-bonds between water and each substitution. The number of H-bonds between water and substitution was calculated for the (a) ED30 and (b) PT35 triple mutant (tmut) and wild-type (wt) simulated in water and 80 % ACN. The results are presented as the means±SEM from triplicate determinations.

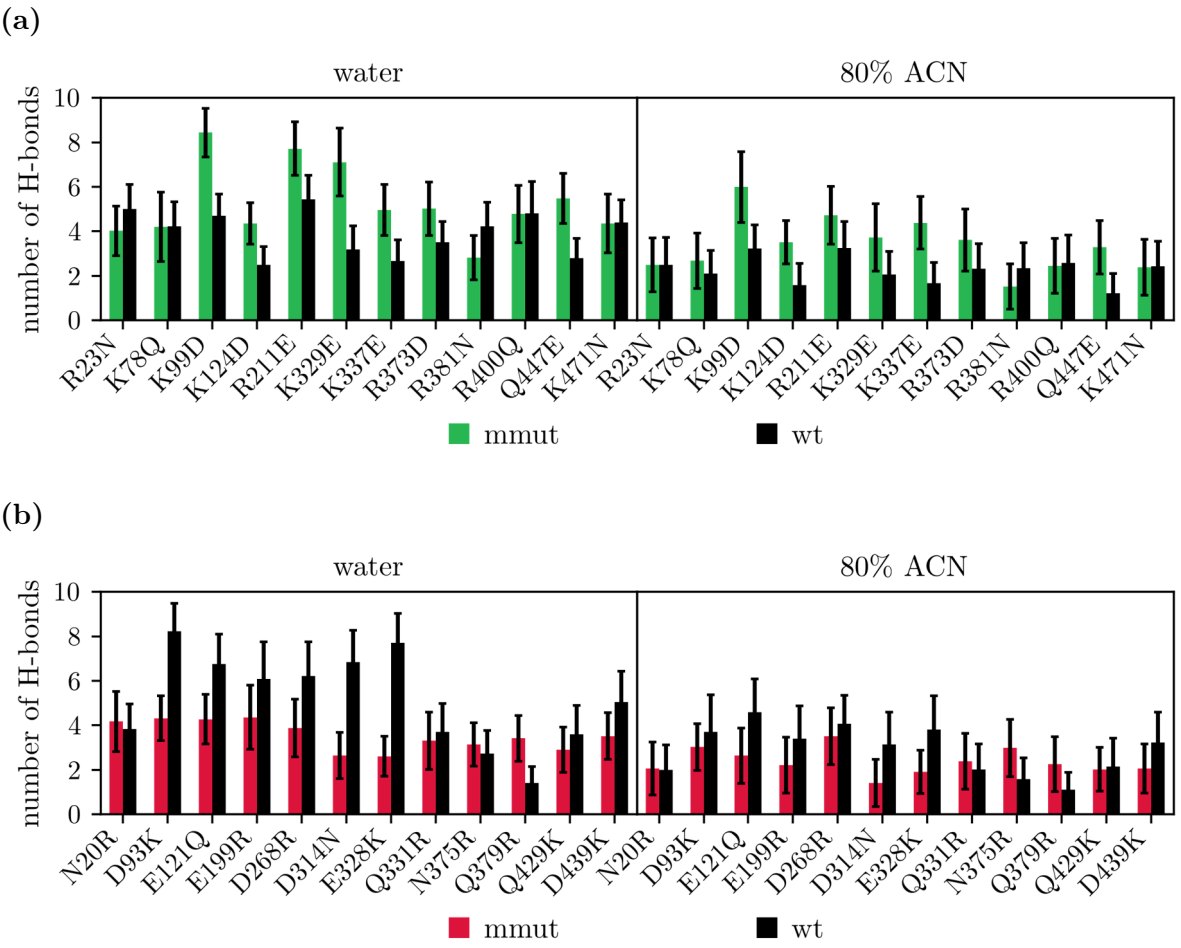

Figure A3: Hydration shell of multi-mutant ED30 and PT35 determined by the number of H-bonds between water and each substitution. The number of H-bonds between water and substitution was calculated for the (a) ED30 and (b) PT35 multi mutant (mmut) and wild-type (wt) simulated in water and 80 % ACN. The results are presented as the means±SEM from triplicate determinations.

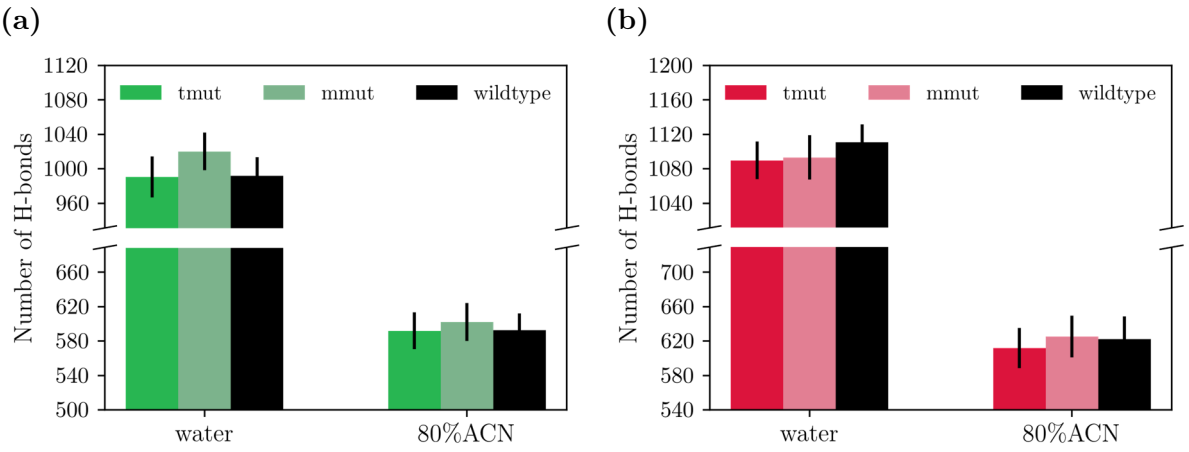

Figure A4: H-bonds between water and the overall enzyme of the wild-type, triple-mutant and multi-mutant ED30 and PT35. The number of H-bonds between water and enzyme was calculated for the (a) ED30 and (b) PT35 triple mutant (tmut), multi mutant (mmut) and wild type simulated in water and 80 % ACN. The results are presented as the means±SEM from triplicate determinations.

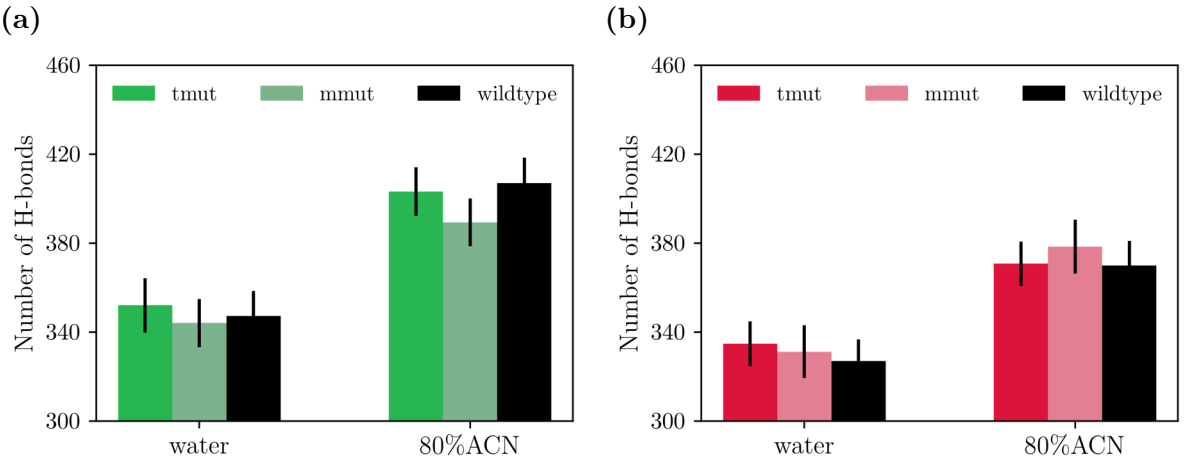

Figure A5: Intramolecular H-bonds of the wild-type, triple-mutant and multi-mutant ED30 and PT35. The number of H-bonds within the enzyme was calculated for the (a) ED30 and (b) PT35 triple mutant (tmut), multi mutant (mmut) and wild type simulated in water and 80 % ACN. The results are presented as the means±SEM from triplicate determinations.

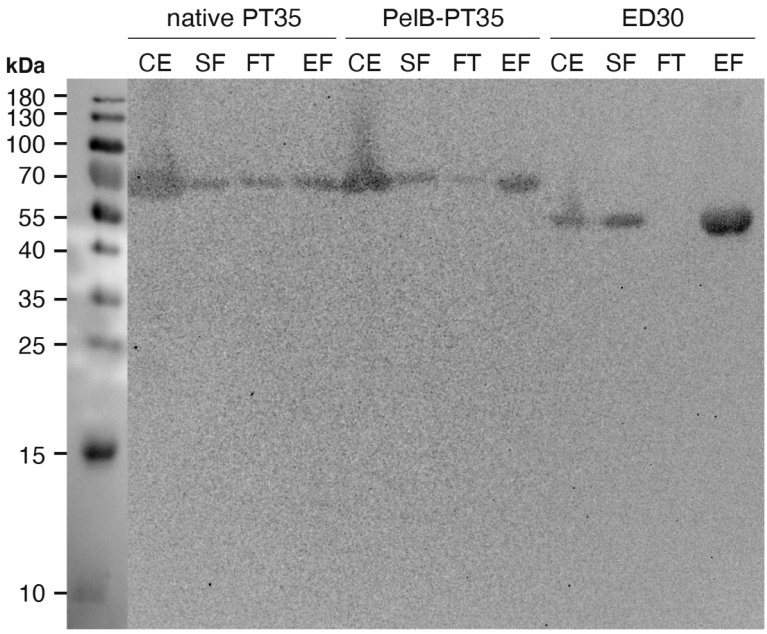

Figure A6: Verification of the wild-type ED30 and PT35 by Western blot analysis. After expression in *E. coli* BL21(DE3) (24 h, 25 °C, 160 rpm), the enzymes were purified by IMAC using Ni-NTA agarose. First row: molecular standard, CE: cell extract, SF: soluble fraction, FT: flow through, EF: elution fraction 2 µl of each sample containing 50 % 2x sample buffer were applied to the gel. The proteins were transferred to a PVDF membrane and immunological detection was performed using the Anti-His (C-Term)-HRP Antibody.

(a) native PT35 **(b)** PT35-PelB

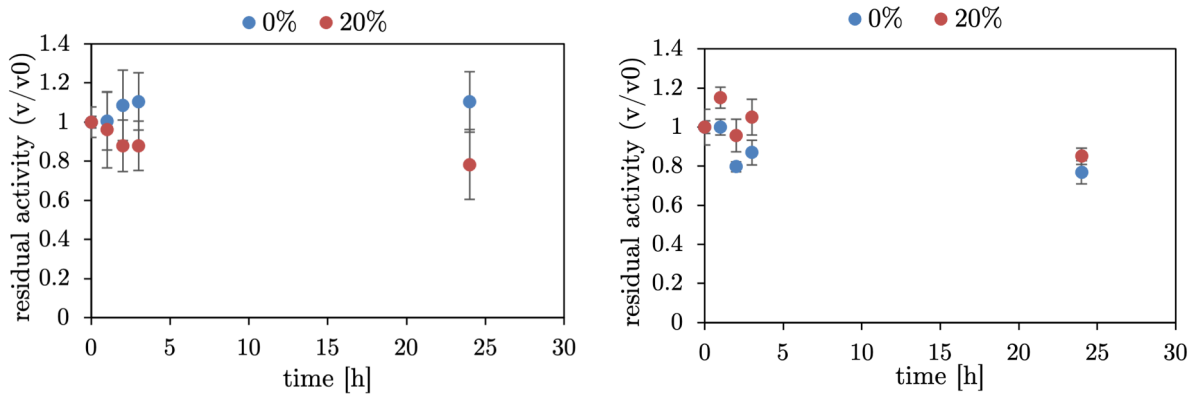

Figure A7: Stability of the native PT35 and PT35-PelB in TRIS buffer and acetonitrile. The residual activity of the native PT35 and PT35-PelB was detected after different incubation times in presence of 0, 20, 50, and 80 % ACN in TRIS buffer. No activity was detected in 50 and 80 %. The activity was measured under standard esterase activity assay conditions using 10 µl purified enzyme in 200 µl of TRIS buffer containing 1 mM *p*NPB. The residual activity (v/v_0) was calculated by comparing the enzyme activity at 0 h (v_0) with the activity after different incubation times (v). The results are presented as the means±SEM from 6 replicates.

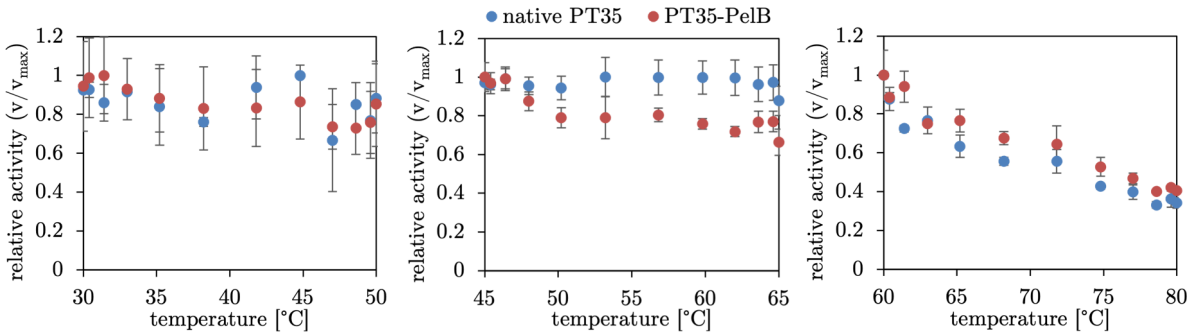

Figure A8: Thermally induced inactivation of native PT35 and PT35-PelB in TRIS buffer. The relative activity of the native PT35 and PT35-PelB was detected after incubation at different temperatures for 1 h in TRIS buffer. The activity was measured under standard esterase activity assay conditions using 10 µl purified enzyme in 200 µl of TRIS buffer containing 1 mM *p*NPB. The relative activity (v/v_{max}) was calculated by comparing the enzyme activity (v) with the maximal activity (v_{max}) after incubation. The results are presented as the means±SEM from duplicate (30-50 °C) and triplicate determinations.

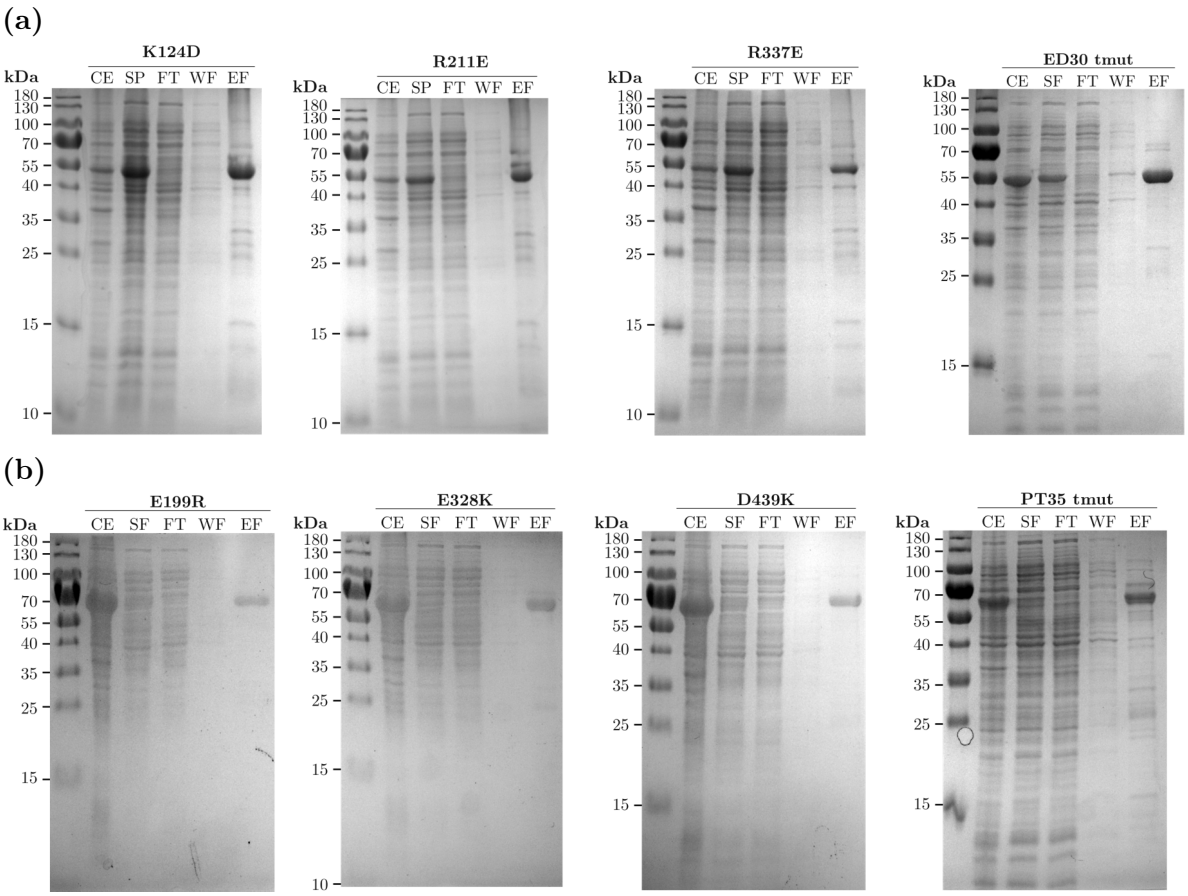

Figure A9: Estimation of the purification of the mutants by SDS-PAGE. After expression in *E. coli* BL21(DE3) (24 h, 25 °C, 160 rpm), the (a) ED30 and (b) PT35 mutants were purified by IMAC using Ni-NTA agarose. First row: molecular standard, CE: cell extract, SF: soluble fraction, FT: flow through, WF: wash fraction, EF: elution fraction. 2 µl of each sample (or 1 µl CE and 3 µl WF) containing 50 % 2x sample buffer were applied to the gel.

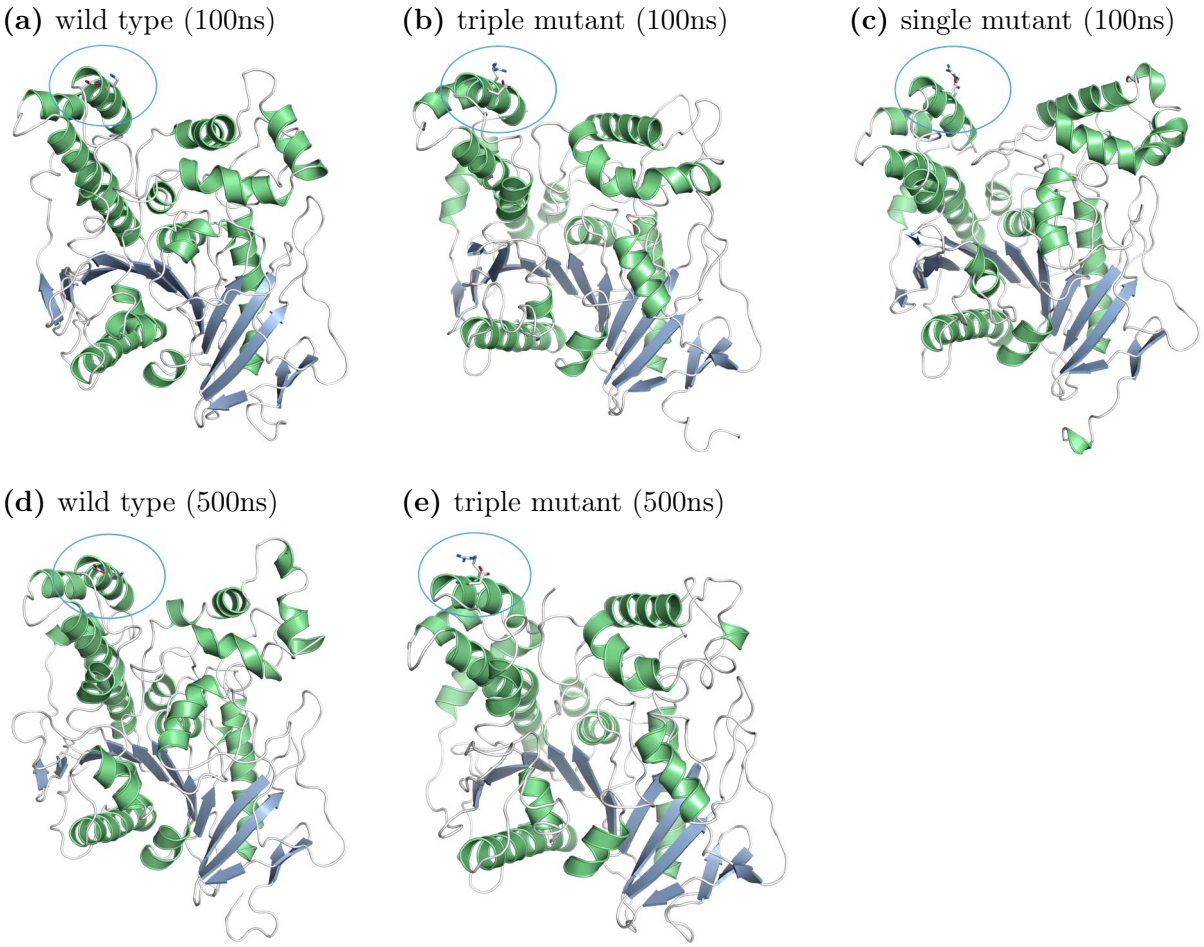

Figure A10: Snapshots of the wild type, triple mutant and K337D single mutant of ED30 after the maximum simulation time. Snapshots were taken from the ED30 (a)/(d) wild type, (b)/(e) triple mutant and (c) K337D single mutant structures after (a-c) 100 ns and (d-e) 500 ns of MD simulation in water only. The salt bridges forming residues K337-D334 (wild type) and E337-R340 (mutants) are shown as sticks with carbon (white), oxygen (red) and nitrogen (blue) and are additionally highlighted by a blue ring. The secondary structure elements are colored as followed: helices (light green), β-sheet (light blue), coil (white).

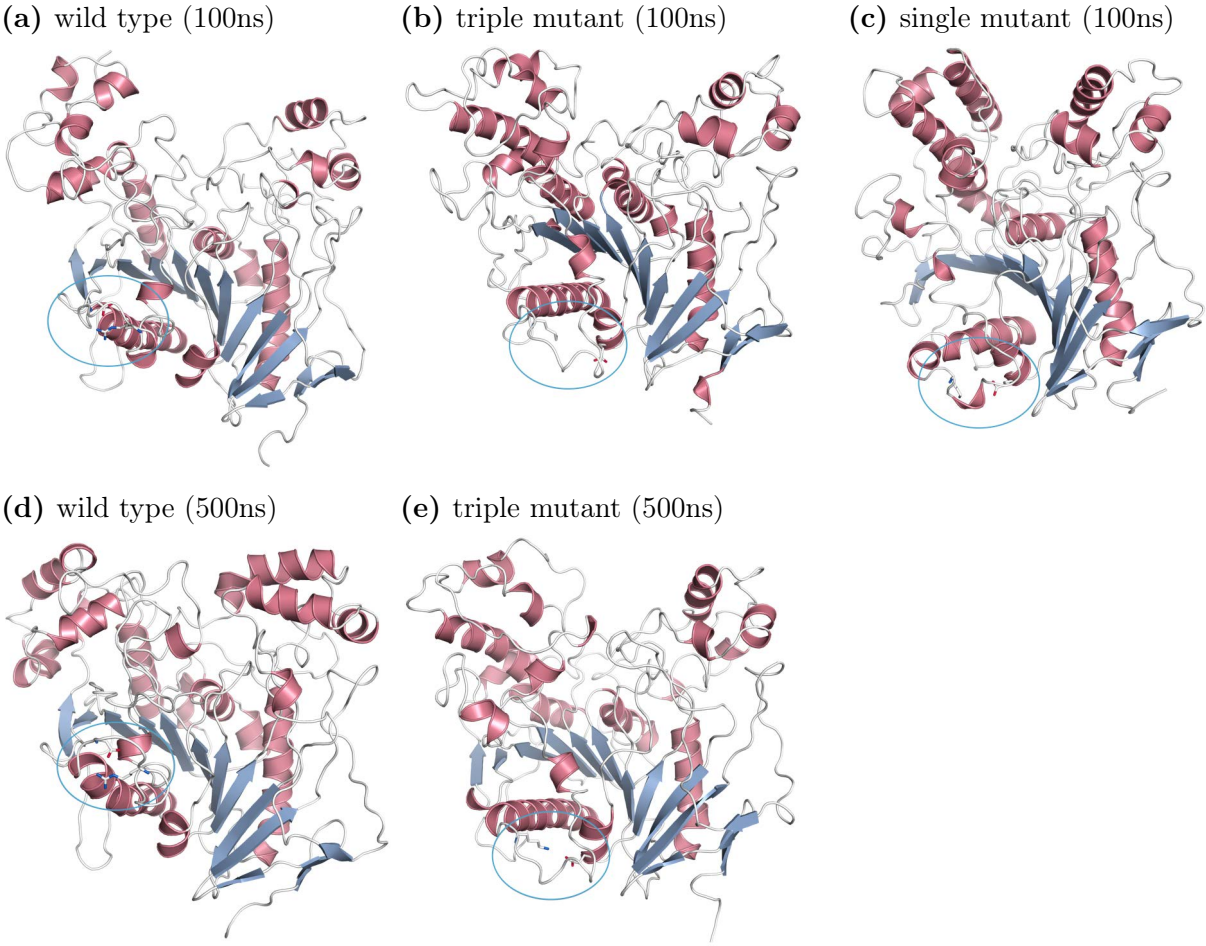

Figure A11: Snapshots of the wild type, triple mutant and D439K single mutant of PT35 after the maximum simulation time. Snapshots were taken from the PT35 (a)/(d) wild type, (b)/(e) triple mutant and (c) D439K single mutant structures after (a-c) 100 ns and (d-e) 500 ns of MD simulation in water only. The salt bridges forming residues D439-R435 (wild type) and K439-E121 (mutants) are shown as sticks with carbon (white), oxygen (red) and nitrogen (blue) and are additionally highlighted by a blue ring. The secondary structure elements are colored as followed: helices (light red), β-sheet (light blue), coil (white).

Bibliography

[1] F. Hasan, A. A. Shah, and A. Hameed. **2006**. Industrial applications of microbial lipases. *Enzyme and Microbial technology*, 39(2):235–251.

[2] J. L. Arpigny and K.-E. Jaeger. **1999**. Bacterial lipolytic enzymes: classification and properties. *Biochemical journal*, 343(1):177–183.

[3] K.-E. Jaeger and F. Kovacic. **2014**. Determination of lipolytic enzyme activities. In *Pseudomonas Methods and Protocols*, pages 111–134. Springer.

[4] K.-E. Jaeger, S. Ransac, B. W. Dijkstra, C. Colson, M. van Heuvel, and O. Misset. **1994**. Bacterial lipases. *FEMS microbiology reviews*, 15(1):29–63.

[5] D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, et al. **1992**. The α/β hydrolase fold. *Protein Engineering, Design and Selection*, 5(3):197–211.

[6] M. Nardini and B. W. Dijkstra. **1999**. α/β Hydrolase fold enzymes: the family keeps growing. *Current opinion in structural biology*, 9(6):732–737.

[7] C. Tölzer, S. Pal, H. Watzlawick, J. Altenbuchner, and K. Niefind. **2016**. A novel esterase subfamily with α/β-hydrolase fold suggested by structures of two bacterial enzymes homologous to l-homoserine O-acetyl transferases. *FEBS letters*, 590(1): 174–184.

[8] F. Kovacic, N. Babic, U. Krauss, and K. Jaeger. **2019**. Classification of lipolytic enzymes from bacteria. *Aerobic Utilization of Hydrocarbons, Oils, and Lipids*, 24: 255–289.

[9] U. T. Bornscheuer. **2002**. Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS microbiology reviews*, 26(1):73–81.

[10] O. Kirk, T. V. Borchert, and C. C. Fuglsang. **2002**. Industrial enzyme applications. *Current opinion in biotechnology*, 13(4):345–351.

[11] R. A. Sheldon and P. C. Pereira. **2017**. Biocatalysis engineering: the big picture. *Chemical Society Reviews*, 46(10):2678–2691.

[12] W. A. Loughlin. **2000**. Biotransformations in organic synthesis. *Bioresource Technology*, 74(1):49–62.

[13] L. Ramnath, B. Sithole, and R. Govinden. **2017**. Classification of lipolytic enzymes and their biotechnological applications in the pulping industry. *Canadian journal of microbiology*, 63(3):179–192.

[14] R. Muralidhar, R. Chirumamilla, R. Marchant, V. Ramachandran, O. Ward, and P. Nigam. **2002**. Understanding lipase stereoselectivity. *World Journal of Microbiology and Biotechnology*, 18(2):81–97.

[15] N. Doukyu and H. Ogino. **2010**. Organic solvent-tolerant enzymes. *Biochemical engineering journal*, 48(3):270–282.

[16] N. Sarmah, D. Revathi, G. Sheelu, K. Yamuna Rani, S. Sridhar, V. Mehtab, and C. Sumana. **2018**. Recent advances on sources and industrial applications of lipases. *Biotechnology progress*, 34(1):5–28.

[17] A. Bollinger, S. Thies, E. Knieps-Grünhagen, C. Gertzen, S. Kobus, A. Höppner, M. Ferrer, H. Gohlke, S. H. Smits, and K.-E. Jaeger. **2020**. A novel polyester hydrolase from the marine bacterium *Pseudomonas aestusnigri*–Structural and functional insights. *Frontiers in microbiology*, 11:114.

[18] M. Lu, A. Dukunde, and R. Daniel. **2019**. Biochemical profiles of two thermostable and organic solvent–tolerant esterases derived from a compost metagenome. *Applied microbiology and biotechnology*, 103(8):3421–3437.

[19] M. Ferrer, R. Bargiela, M. Martínez-Martínez, J. Mir, R. Koch, O. V. Golyshina, and P. N. Golyshin. **2015**. Biodiversity for biocatalysis: a review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatalysis and Biotransformation*, 33(5-6):235–249.

[20] P. Priyanka, Y. Tan, G. K. Kinsella, G. T. Henehan, and B. J. Ryan. **2019**. Solvent stable microbial lipases: current understanding and biotechnological applications. *Biotechnology letters*, 41(2):203–220.

[21] L. Polgár. **2005**. The catalytic triad of serine peptidases. *Cellular and molecular life sciences CMLS*, 62(19):2161–2172.

[22] T. Imai. **2006**. Human carboxylesterase isozymes: catalytic properties and rational drug design. *Drug metabolism and pharmacokinetics*, 21(3):173–185.

[23] A. Kumar, K. Dhar, S. S. Kanwar, and P. K. Arora. **2016**. Lipase catalysis in organic solvents: advantages and applications. *Biological Procedures Online*, 18(1): 1–11.

[24] V. Stepankova, S. Bidmanova, T. Koudelakova, Z. Prokop, R. Chaloupkova, and J. Damborsky. **2013**. Strategies for stabilization of enzymes in organic solvents. *Acs Catalysis*, 3(12):2823–2836.

[25] A. J. Auman, J. L. Breezee, J. J. Gosink, P. Kämpfer, and J. T. Staley. **2006**. *Psychromonas ingrahamii* sp. nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice. *International journal of systematic and evolutionary microbiology*, 56(5):1001–1007.

[26] K. Takai, K. Nakamura, T. Toki, U. Tsunogai, M. Miyazaki, J. Miyazaki, H. Hirayama, S. Nakagawa, T. Nunoura, and K. Horikoshi. **2008**. Cell proliferation at 122 C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proceedings of the National Academy of Sciences*, 105(31):10949–10954.

[27] C. Kato, L. Li, Y. Nogi, Y. Nakamura, J. Tamaoka, and K. Horikoshi. **1998**. Extremely barophilic bacteria isolated from the Mariana Trench, Challenger Deep, at a depth of 11,000 meters. *Applied and environmental microbiology*, 64(4):1510–1513.

[28] S. DasSarma and P. DasSarma. **2017**. *Halophiles*, pages 1–13. American Cancer Society. ISBN 9780470015902.

[29] C. Baker-Austin and M. Dopson. **2007**. Life in acid: pH homeostasis in acidophiles. *Trends in microbiology*, 15(4):165–171.

[30] D. G. Wernick, S. P. Pontrelli, A. W. Pollock, and J. C. Liao. **2016**. Sustainable biorefining in wastewater by engineered extreme alkaliphile *Bacillus marmarensis*. *Scientific reports*, 6(1):1–10.

[31] J. A. Coker. **2016**. Extremophiles and biotechnology: current uses and prospects. *F1000Research*, 5.

[32] D. W. Hough and M. J. Danson. **1999**. Extremozymes. *Current opinion in chemical biology*, 3(1):39–46.

[33] G. Wang, Q. Wang, X. Lin, T. B. Ng, R. Yan, J. Lin, and X. Ye. **2016**. A novel cold-adapted and highly salt-tolerant esterase from *Alkalibacterium* sp. SL3 from the sediment of a soda lake. *Scientific reports*, 6(1):1–10.

[34] P. Fuciños, R. González, E. Atanes, A. B. F. Sestelo, N. Pérez-Guerra, L. Pastrana, and M. L. Rúa. **2012**. Lipases and esterases from extremophiles: overview and case example of the production and purification of an esterase from *Thermus thermophilus* HB27. *Lipases and Phospholipases*, pages 239–266.

[35] F. Kovacic, A. Mandrysch, C. Poojari, B. Strodel, and K.-E. Jaeger. **2016**. Structural features determining thermal adaptation of esterases. *Protein Engineering, Design and Selection*, 29(2):65–76.

[36] P. D. Schloss and J. Handelsman. **2005**. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome biology*, 6(8):1–4.

[37] P. Hugenholtz and G. W. Tyson. **2008**. Metagenomics. *Nature*, 455(7212):481–483.

[38] P. Lorenz and J. Eck. **2005**. Metagenomics and industrial applications. *Nature Reviews Microbiology*, 3(6):510–516.

[39] J. Kennedy, J. R. Marchesi, and A. D. Dobson. **2008**. Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microbial cell factories*, 7(1):1–8.

[40] M. Khan and K. Jithesh. **2012**. Expression and purification of organic solvent stable lipase from soil metagenomic library. *World Journal of Microbiology and Biotechnology*, 28(6):2417–2424.

[41] P. Jin, X. Pei, P. Du, X. Yin, X. Xiong, H. Wu, X. Zhou, and Q. Wang. **2012**. Overexpression and characterization of a new organic solvent-tolerant esterase derived from soil metagenomic DNA. *Bioresource technology*, 116:234–240.

[42] Q. Mo, A. Liu, H. Guo, Y. Zhang, and M. Li. **2016**. A novel thermostable and organic solvent-tolerant lipase from *Xanthomonas oryzae* pv. oryzae YB103: screening, purification and characterization. *Extremophiles*, 20(2):157–165.

[43] J.-M. Park, C.-H. Kang, S.-M. Won, K.-H. Oh, and J.-H. Yoon. **2020**. Characterization of a novel moderately thermophilic solvent-tolerant esterase isolated from a compost metagenome library. *Frontiers in microbiology*, 10:3069.

[44] X. Li and H.-Y. Yu. **2014**. Characterization of an organic solvent-tolerant lipase from *Haloarcula* sp. G41 and its application for biodiesel production. *Folia microbiologica*, 59(6):455–463.

[45] L. Xin and Y. Hui-Ying. **2013**. Purification and characterization of an extracellular esterase with organic solvent tolerance from a halotolerant isolate, *Salimicrobium* sp. LY19. *BMC biotechnology*, 13(1):1–8.

[46] C. Roh and R. D. Schmid. **2013**. Isolation of an organic solvent-tolerant lipolytic enzyme from uncultivated microorganism. *Applied biochemistry and biotechnology*, 171(7):1750–1758.

[47] X.-L. Zhong, Y.-Z. Tian, M.-L. Jia, Y.-D. Liu, D. Cheng, and G. Li. **2020**. Characterization and purification via nucleic acid aptamers of a novel esterase from the metagenome of paper mill wastewater sediments. *International journal of biological macromolecules*, 153:441–450.

[48] S. H. Park, S.-j. Kim, S. Park, and H. K. Kim. **2019**. Characterization of organic solvent-tolerant lipolytic enzyme from *Marinobacter lipolyticus* isolated from the Antarctic Ocean. *Applied biochemistry and biotechnology*, 187(3):1046–1060.

[49] Y. Zhang, F. Ji, J. Wang, Z. Pu, B. Jiang, and Y. Bao. **2018**. Purification and characterization of a novel organic solvent-tolerant and cold-adapted lipase from *Psychrobacter* sp. ZY124. *Extremophiles*, 22(2):287–300.

[50] A. K. Maharana and S. M. Singh. **2018**. A cold and organic solvent tolerant lipase produced by Antarctic strain *Rhodotorula* sp. Y-23. *Journal of basic microbiology*, 58(4):331–342.

[51] J. Li and X. Liu. **2017**. Identification and characterization of a novel thermophilic, organic solvent stable lipase of *Bacillus* from a hot spring. *Lipids*, 52(7):619–627.

[52] A. Ghati and G. Paul. **2015**. Purification and characterization of a thermo-halophilic, alkali-stable and extremely benzene tolerant esterase from a thermo-halo tolerant *Bacillus cereus* strain AGP-03, isolated from 'Bakreshwar' hot spring, India. *Process Biochemistry*, 50(5):771–781.

[53] R. K. Sahoo, M. Kumar, L. B. Sukla, and E. Subudhi. **2017**. Bioprospecting hot spring metagenome: lipase for the production of biodiesel. *Environmental Science and Pollution Research*, 24(4):3802–3809.

[54] A. Bollinger, R. Molitor, S. Thies, R. Koch, C. Coscolín, M. Ferrer, and K.-E. Jaeger. **2020**. Organic-solvent-tolerant carboxylic ester hydrolases for organic synthesis. *Applied and environmental microbiology*, 86(9).

[55] S. Malekabadi, A. Badoei-Dalfard, and Z. Karami. **2018**. Biochemical characterization of a novel cold-active, halophilic and organic solvent-tolerant lipase from *B. licheniformis* KM12 with potential application for biodiesel production. *International journal of biological macromolecules*, 109:389–398.

[56] M. Li, L.-R. Yang, G. Xu, and J.-P. Wu. **2013**. Screening, purification and characterization of a novel cold-active and organic solvent-tolerant lipase from *Stenotrophomonas maltophilia* CGMCC 4254. *Bioresource technology*, 148:114–120.

[57] G. A. Sellek and J. B. Chaudhuri. **1999**. Biocatalysis in organic media using enzymes from extremophiles. *Enzyme and Microbial Technology*, 25(6):471–482.

[58] M. de Lourdes Moreno, D. Pérez, M. T. García, and E. Mellado. **2013**. Halophilic bacteria as a source of novel hydrolytic enzymes. *Life*, 3(1):38–51.

[59] K. M. Polizzi, A. S. Bommarius, J. M. Broering, and J. F. Chaparro-Riggers. **2007**. Stability of biocatalysts. *Current opinion in chemical biology*, 11(2):220–225.

[60] M. Musil, H. Konegger, J. Hon, D. Bednar, and J. Damborsky. **2018**. Computational design of stable and soluble biocatalysts. *Acs Catalysis*, 9(2):1033–1054.

[61] M. Manning and W. Colón. **2004**. Structural basis of protein kinetic stability: resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward β-sheet structure. *Biochemistry*, 43(35):11248–11254.

[62] C. N. Pace, H. Fu, K. L. Fryar, J. Landua, S. R. Trevino, B. A. Shirley, M. M. Hendricks, S. Iimura, K. Gajiwala, J. M. Scholtz, et al. **2011**. Contribution of hydrophobic interactions to protein stability. *Journal of molecular biology*, 408(3): 514–528.

[63] H. Ogino and H. Ishikawa. **2001**. Enzymes which are stable in the presence of organic solvents. *Journal of bioscience and bioengineering*, 91(2):109–116.

[64] M. Lotti, J. Pleiss, F. Valero, and P. Ferrer. **2018**. Enzymatic production of biodiesel: strategies to overcome methanol inactivation. *Biotechnology journal*, 13 (5):1700155.

[65] A. M. Klibanov. **1997**. Why are enzymes less active in organic solvents than in water? *Trends in biotechnology*, 15(3):97–101.

[66] Y. L. Khmelnitsky, A. B. Belova, A. V. Levashov, and V. V. Mozhaev. **1991**. Relationship between surface hydrophilicity of a protein and its stability against denaturation by organic solvents. *FEBS letters*, 284(2):267–269.

[67] T. Kawata and H. Ogino. **2009**. Enhancement of the organic solvent-stability of the LST-03 lipase by directed evolution. *Biotechnology progress*, 25(6):1605–1611.

[68] F. Zhu, B. He, F. Gu, H. Deng, C. Chen, W. Wang, and N. Chen. **2020**. Improvement in organic solvent resistance and activity of metalloprotease by directed evolution. *Journal of biotechnology*, 309:68–74.

[69] J. K. Song and J. S. Rhee. **2001**. Enhancement of stability and activity of phospholipase A1 in organic solvents by directed evolution. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1547(2):370–378.

[70] R. Yamada, T. Higo, C. Yoshikawa, H. China, and H. Ogino. **2014**. Improvement of the stability and activity of the BPO-A1 haloperoxidase from *Streptomyces aureofaciens* by directed evolution. *Journal of biotechnology*, 192:248–254.

[71] T. P. Korman, B. Sahachartsiri, D. M. Charbonneau, G. L. Huang, M. Beauregard, and J. U. Bowie. **2013**. Dieselzymes: development of a stable and methanol tolerant lipase for biodiesel production by directed evolution. *Biotechnology for biofuels*, 6 (1):1–13.

[72] K. Tian, K. Tai, B. J. W. Chua, and Z. Li. **2017**. Directed evolution of Thermomyces lanuginosus lipase to enhance methanol tolerance for efficient production of biodiesel from waste grease. *Bioresource technology*, 245:1491–1497.

[73] U. T. Bornscheuer, B. Hauer, K. E. Jaeger, and U. Schwaneberg. **2019**. Directed evolution empowered redesign of natural proteins for the sustainable production of chemicals and pharmaceuticals. *Angewandte Chemie International Edition*, 58(1): 36–40.

[74] J. L. Porter, R. A. Rusli, and D. L. Ollis. **2016**. Directed evolution of enzymes for industrial biocatalysis. *ChemBioChem*, 17(3):197–203.

[75] L. Zheng, U. Baumann, and J.-L. Reymond. **2004**. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic acids research*, 32(14): e115–e115.

[76] R. C. Cadwell and G. F. Joyce. **1992**. Randomization of genes by PCR mutagenesis. *Genome research*, 2(1):28–33.

[77] G. Williams, A. Nelson, and A. Berry. **2004**. Directed evolution of enzymes for biocatalysis and the life sciences. *Cellular and Molecular Life Sciences CMLS*, 61 (24):3034–3046.

[78] M. T. Reetz, S. Wilensek, D. Zha, and K.-E. Jaeger. **2001**. Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angewandte Chemie International Edition*, 40(19):3589–3591.

[79] H. Leemhuis, R. M. Kelly, and L. Dijkhuizen. **2009**. Directed evolution of enzymes: library screening strategies. *IUBMB life*, 61(3):222–228.

[80] R. A. Chica, N. Doucet, and J. N. Pelletier. **2005**. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current opinion in biotechnology*, 16(4):378–384.

[81] G. Li, X. Fang, F. Su, Y. Chen, L. Xu, and Y. Yan. **2018**. Enhancing the thermostability of *Rhizomucor miehei* lipase with a limited screening library by rational-design point mutations and disulfide bonds. *Applied and environmental microbiology*, 84 (2).

[82] K. Zorn, I. Oroz-Guinea, H. Brundiek, M. Dörr, and U. T. Bornscheuer. **2018**. Alteration of Chain Length Selectivity of *Candida antarctica* Lipase A by Semi-Rational Design for the Enrichment of Erucic and Gondoic Fatty Acids. *Advanced synthesis & catalysis*, 360(21):4115–4131.

[83] J. A. Dietrich, A. E. McKee, and J. D. Keasling. **2010**. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annual review of biochemistry*, 79:563–590.

[84] M. D. Lane and B. Seelig. **2014**. Advances in the directed evolution of proteins. *Current opinion in chemical biology*, 22:129–136.

[85] U. Markel, K. D. Essani, V. Besirlioglu, J. Schiffels, W. R. Streit, and U. Schwaneberg. **2020**. Advances in ultrahigh-throughput screening for directed enzyme evolution. *Chemical Society Reviews*, 49(1):233–262.

[86] K. K. Yang, Z. Wu, and F. H. Arnold. **2019**. Machine-learning-guided directed evolution for protein engineering. *Nature methods*, 16(8):687–694.

[87] M. T. Reetz, D. Kahakeaw, and R. Lohmer. **2008**. Addressing the numbers problem in directed evolution. *ChemBioChem*, 9(11):1797–1804.

[88] U. T. Bornscheuer and M. Pohl. **2001**. Improved biocatalysts by directed evolution and rational protein design. *Current opinion in chemical biology*, 5(2):137–143.

[89] E. Vazquez-Figueroa, V. Yeh, J. Broering, J. Chaparro-Riggers, and A. Bommarius. **2008**. Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media. *Protein engineering, design & selection*, 21(11):673–680.

[90] M. T. Reetz, P. Soni, L. Fernández, Y. Gumulya, and J. D. Carballeira. **2010**. Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chemical communications*, 46(45):8657–8658.

[91] Z.-l. Han, S.-y. Han, S.-p. Zheng, and Y. Lin. **2009**. Enhancing thermostability of a Rhizomucor miehei lipase by engineering a disulfide bond and displaying on the yeast cell surface. *Applied microbiology and biotechnology*, 85(1):117–126.

[92] H. Gao, C. Li, R. Bandikari, Z. Liu, N. Hu, and Q. Yong. **2018**. A novel cold-adapted esterase from *Enterobacter cloacae*: Characterization and improvement of its activity and thermostability via the site of Tyr193Cys. *Microbial cell factories*, 17(1):1–11.

[93] N. Akbulut, M. T. Öztürk, T. Pijning, S. İ. Öztürk, and F. Gümüşel. **2013**. Improved activity and thermostability of *Bacillus pumilus* lipase by directed evolution. *Journal of biotechnology*, 164(1):123–129.

[94] L. J. Perry and R. Wetzel. **1984**. Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science*, 226(4674):555–557.

[95] J. Mansfeld, G. Vriend, B. W. Dijkstra, O. R. Veltman, B. Van den Burg, G. Venema, R. Ulbrich-Hofmann, and V. G. Eijsink. **1997**. Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *Journal of Biological Chemistry*, 272(17):11152–11156.

[96] M.-Y. Jeong, S. Kim, C.-W. Yun, Y.-J. Choi, and S.-G. Cho. **2007**. Engineering a de novo internal disulfide bridge to improve the thermal stability of xylanase from *Bacillus stearothermophilus* No. 236. *Journal of biotechnology*, 127(2):300–309.

[97] Y. Wang, Z. Fu, H. Huang, H. Zhang, B. Yao, H. Xiong, and O. Turunen. **2012**. Improved thermal performance of *Thermomyces lanuginosus* GH11 xylanase by engineering of an N-terminal disulfide bridge. *Bioresource technology*, 112:275–279.

[98] H. Takagi, K. Hirai, Y. Maeda, H. Matsuzawa, and S. Nakamori. **2000**. Engineering subtilisin E for enhanced stability and activity in polar organic solvents. *The Journal of Biochemistry*, 127(4):617–625.

[99] H. Ogino, T. Uchiho, J. Yokoo, R. Kobayashi, R. Ichise, and H. Ishikawa. **2001**. Role of intermolecular disulfide bonds of the organic solvent-stable PST-01 protease in its organic solvent stability. *Applied and Environmental Microbiology*, 67(2): 942–947.

[100] O. R. Siadat, A. Lougarre, L. Lamouroux, C. Ladurantie, and D. Fournier. **2006**. The effect of engineered disulfide bonds on the stability of *Drosophila melanogaster* acetylcholinesterase. *BMC biochemistry*, 7(1):1–7.

[101] C. N. Pace, H. Fu, K. Lee Fryar, J. Landua, S. R. Trevino, D. Schell, R. L. Thurlkill, S. Imura, J. M. Scholtz, K. Gajiwala, et al. **2014**. Contribution of hydrogen bonds to protein stability. *Protein Science*, 23(5):652–661.

[102] G. Vogt and P. Argos. **1997**. Protein thermal stability: hydrogen bonds or internal packing? *Folding and Design*, 2:S40–S46.

[103] A. Horovitz, L. Serrano, B. Avron, M. Bycroft, and A. R. Fersht. **1990**. Strength and co-operativity of contributions of surface salt bridges to protein stability. *Journal of molecular biology*, 216(4):1031–1044.

[104] D. M. Charbonneau and M. Beauregard. **2013**. Role of key salt bridges in thermostability of *G. thermodenitrificans* EstGtA2: distinctive patterns within the new bacterial lipolytic enzyme family XV. *PLoS One*, 8(10):e76675.

[105] T. Kawata and H. Ogino. **2010**. Amino acid residues involved in organic solvent-stability of the LST-03 lipase. *Biochemical and biophysical research communications*, 400(3):384–388.

[106] K. Takano, K. Tsuchimori, Y. Yamagata, and K. Yutani. **2000**. Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry*, 39(40):12375–12381.

[107] G. I. Makhatadze, V. V. Loladze, D. N. Ermolenko, X. Chen, and S. T. Thomas. **2003**. Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *Journal of molecular biology*, 327(5):1135–1148.

[108] R. Luo, L. David, H. Hung, J. Devaney, and M. K. Gilson. **1999**. Strength of solvent-exposed salt-bridges. *The Journal of Physical Chemistry B*, 103(4):727–736.

[109] B. T. Porebski and A. M. Buckle. **2016**. Consensus protein design. *Protein Engineering, Design and Selection*, 29(7):245–251.

[110] A. S. Bommarius, J. K. Blum, and M. J. Abrahamson. **2011**. Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Current opinion in chemical biology*, 15(2):194–200.

[111] A. Dror, E. Shemesh, N. Dayan, and A. Fishman. **2014**. Protein engineering by random mutagenesis and structure-guided consensus of *Geobacillus stearothermophilus* lipase T6 for enhanced stability in methanol. *Applied and environmental microbiology*, 80(4):1515–1527.

[112] E. R. Lindahl. **2008**. *Molecular Modeling of Proteins*.

[113] D. Seeliger and B. L. De Groot. **2010**. Protein thermostability calculations using alchemical free energy simulations. *Biophysical journal*, 98(10):2309–2316.

[114] J. Mendes, R. Guerois, and L. Serrano. **2002**. Energy estimation in protein design. *Current opinion in structural biology*, 12(4):441–446.

[115] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano. **2005**. The FoldX web server: an online force field. *Nucleic Acids Research*, 33(2):W382–W388. ISSN 0305-1048.

[116] Z. Zhang, L. Wang, Y. Gao, J. Zhang, M. Zhenirovskyy, and E. Alexov. **2012**. Predicting folding free energy changes upon single point mutations. *Bioinformatics*, 28(5):664–671.

[117] E. Capriotti, P. Fariselli, and R. Casadio. **2005**. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*, 33(suppl_2):W306–W310.

[118] L. Folkman, B. Stantic, A. Sattar, and Y. Zhou. **2016**. EASE-MM: sequence-based prediction of mutation-induced stability changes with feature-based multiple models. *Journal of molecular biology*, 428(6):1394–1405.

[119] S. Teng, A. K. Srivastava, and L. Wang. **2010**. Sequence feature-based prediction of protein stability changes upon amino acid substitutions. *BMC genomics*, 11(2): 1–8.

[120] L.-T. Huang, M. M. Gromiha, and S.-Y. Ho. **2007**. iPTREE-STAB: interpretable decision tree based method for predicting protein stability changes upon mutations. *Bioinformatics*, 23(10):1292–1293.

[121] Y. Dehouck, J. M. Kwasigroch, D. Gilis, and M. Rooman. **2011**. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC bioinformatics*, 12(1):1–12.

[122] H. J. Park, J. C. Joo, K. Park, Y. H. Kim, and Y. J. Yoo. **2013**. Prediction of the solvent affecting site and the computational design of stable *Candida antarctica* lipase B in a hydrophilic organic solvent. *Journal of biotechnology*, 163(3):346–352.

[123] H. Cui, L. Eltoukhy, L. Zhang, U. Markel, K.-E. Jaeger, M. D. Davari, and U. Schwaneberg. **2021**. Less unfavorable salt-bridges on the enzyme surface results in more organic cosolvent resistance. *Angewandte Chemie*.

[124] Y.-J. Zheng and R. L. Ornstein. **1996**. What happens to salt-bridges in nonaqueous environments: insights from quantum mechanics calculations. *Journal of the American Chemical Society*, 118(45):11237–11243.

[125] J. N. Sarakatsannis and Y. Duan. **2005**. Statistical characterization of salt bridges in proteins. *Proteins: Structure, Function, and Bioinformatics*, 60(4):732–739.

[126] K. Chen, A. C. Robinson, M. E. Van Dam, P. Martinez, C. Economou, and F. H. Arnold. **1991**. Enzyme engineering for nonaqueous solvents. II. Additive effects of mutations on the stability and activity of subtilisin E in polar organic media. *Biotechnology progress*, 7(2):125–129.

[127] T. Takahashi, K. K.-S. Ng, H. Oyama, and K. Oda. **2005**. Molecular cloning of the gene encoding Vibrio metalloproteinase vimelysin and isolation of a mutant with high stability in organic solvents. *Journal of biochemistry*, 138(6):701–710.

[128] H. J. Park, J. C. Joo, K. Park, and Y. J. Yoo. **2012**. Stabilization of *Candida antarctica* lipase B in hydrophilic organic solvent by rational design of hydrogen bond. *Biotechnology and Bioprocess Engineering*, 17(4):722–728.

[129] I. D. Kuntz Jr. **1971**. Hydration of macromolecules. III. Hydration of polypeptides. *Journal of the American Chemical Society*, 93(2):514–516.

[130] R. M. Kramer, V. R. Shende, N. Motl, C. N. Pace, and J. M. Scholtz. **2012**. Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophysical journal*, 102(8):1907–1915.

[131] J. N. Pedersen, Y. Zhou, Z. Guo, and B. Pérez. **2019**. Genetic and chemical approaches for surface charge engineering of enzymes and their applicability in biocatalysis: A review. *Biotechnology and bioengineering*, 116(7):1795–1812.

[132] Y. Zhou, B. Pérez, W. Hao, J. Lv, R. Gao, and Z. Guo. **2019**. The additive mutational effects from surface charge engineering: a compromise between enzyme activity, thermostability and ionic liquid tolerance. *Biochemical Engineering Journal*, 148:195–204.

[133] D. Sánchez, M. Mulet, A. C. Rodríguez, Z. David, J. Lalucat, and E. García-Valdés. **2014**. *Pseudomonas aestusnigri* sp. nov., isolated from crude oil-contaminated intertidal sand samples after the Prestige oil spill. *Systematic and applied microbiology*, 37(2):89–94.

[134] R. J. Brooijmans, M. I. Pastink, and R. J. Siezen. **2009**. Hydrocarbon-degrading bacteria: the oil-spill clean-up crew. *Microbial biotechnology*, 2(6):587–594.

[135] M. Gomila, M. Mulet, J. Lalucat, and E. García-Valdés. **2017**. Draft genome sequence of the marine bacterium *Pseudomonas aestusnigri* VGXO14T. *Genome announcements*, 5(32).

[136] M. Bulka. **2019**. Reinigung und Charakterisierung einer Esterase aus *Pseudomonas aestusnigri* für biotechnologische Anwendungen. Master's thesis.

[137] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. **1990**. Basic local alignment search tool. *J. Mol. Biol.*, 215(3):403 – 410.

[138] D. J. States and W. Gish. **1994**. QGB: Combined Use of Sequence Similarity and Codon Bias for Coding Region Identification. *J. Comput. Biol.*, 1(1):39–50.

[139] A. Jaeckering. **2020**. Computational analyses towards the mechanism behind resistance to organic solvent and sodium dodecyl sulfate of a novel esterase. Master's thesis.

[140] N. Tao, Y. Gao, and Y. Liu. **2011**. Isolation and characterization of a *Pichia anomala* strain: a promising candidate for bioethanol production. *Brazilian Journal of Microbiology*, 42(2):668–675.

[141] D. E. Otzen. **2002**. Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature. *Biophysical journal*, 83(4):2219–2230.

[142] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. **2001**. Electrostatics of nanosystems: Application to microtubules and the ribosome. *P.e Natl. A. Sci.*, 98(18):10037–10041. ISSN 0027-8424.

[143] L. Schrodinger. **2010**. The PyMOL molecular graphics system. *Version*, 1(5):0.

[144] J. Van Durme, J. Delgado, F. Stricher, L. Serrano, J. Schymkowitz, and F. Rousseau. **2011**. A graphical interface for the FoldX forcefield. *Bioinformatics*, 27(12):1711–1712. ISSN 1367-4803.

[145] E. Krieger, G. Koraimann, and G. Vriend. **2002**. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins: Structure, Function, and Bioinformatics*, 47(3):393–402.

[146] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. **2015**. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19 – 25. ISSN 2352-7110.

[147] C. Kutzner, S. Páll, M. Fechner, A. Esztermann, B. L. de Groot, and H. Grubmüller. **2019**. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *Journal of computational chemistry*, 40(27):2418–2431.

[148] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. **2009**. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191. ISSN 1367-4803.

[149] D. Krause and P. Thörnig. **2018**. Jureca: Modular supercomputer at jülich supercomputing centre. *Journal of large-scale research facilities JLSRF*, 4:132.

[150] A. Roy, A. Kucukural, and Y. Zhang. **2010**. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, 5(4):725–738.

[151] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang. **2015**. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*, 12(1):7–8.

[152] J. Yang and Y. Zhang. **2015**. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.*, 43(W1):W174–181.

[153] M. P. Allen et al. **2004**. Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins*, 23(1):1–28.

[154] A. Hospital, J. R. Goni, M. Orozco, and J. L. Gelpi. **2015**. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.*, 8:37–47.

[155] M. González. **2011**. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, 12:169–200.

[156] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. **1995**. A smooth particle mesh Ewald method. *The Journal of chemical physics*, 103(19):8577–8593.

[157] B. Leimkuhler and C. Matthews. **2016**. *Molecular Dynamics.* Springer.

[158] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. **1995**. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197.

[159] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al. **1998**. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18):3586–3616.

[160] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren. **2004**. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry*, 25 (13):1656–1676.

[161] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. **1996**. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236.

[162] R. B. Best and G. Hummer. **2009**. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B*, 113(26): 9004–9015.

[163] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. **2010**. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–1958.

[164] A. E. Aliev, M. Kulke, H. S. Khaneja, V. Chudasama, T. D. Sheppard, and R. M. Lanigan. **2014**. Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics. *Proteins*, 82(2): 195–215.

[165] P. K. Weiner and P. A. Kollman. **1981**. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry*, 2(3):287–303.

[166] J. Wang, P. Cieplak, and P. A. Kollman. **2000**. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of computational chemistry*, 21(12): 1049–1074.

[167] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. **2006**. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725.

[168] L. Verlet. **1967**. Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review*, 159(1):98.

[169] M. P. Allen and D. J. Tildesley. **2017**. *Computer simulation of liquids*. Oxford university press.

[170] M. A. Cuendet and W. F. van Gunsteren. **2007**. On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *The Journal of chemical physics*, 127(18):184102.

[171] D. Frenkel and B. Smit. **2001**. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier.

[172] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. **1982**. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of chemical physics*, 76(1):637–649.

[173] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. **1977**. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341.

[174] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. **1997**. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12): 1463–1472.

[175] K. A. Feenstra, B. Hess, and H. J. Berendsen. **1999**. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798.

[176] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. **1983**. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935.

[177] D. van der Spoel, P. J. van Maaren, and C. Caleman. **2012**. GROMACS molecule & liquid database. *Bioinformatics*, 28(5):752–753.

[178] M. A. Cauchy. **1847**. Méthode généerale pour la résolution des systèmes d' éequations simultanéees. *CR Hebd. Acad. Sci.*, 25:536–538.

[179] G. Bussi, D. Donadio, and M. Parrinello. **2007**. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101.

[180] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. **1984**. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690.

[181] T. Darden, D. York, and L. Pedersen. **1993**. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092.

[182] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. **1995**. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103(19):8577–8593.

[183] S. Nosé. **1984**. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519.

[184] W. Hoover. **1985**. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev.*, 31:1695–1697.

[185] M. Parrinello and A. Rahman. **1980**. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.*, 45:1196–1199.

[186] M. Parrinello and A. Rahman. **1981**. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190.

[187] M. Parrinello and A. Rahman. **1982**. Strain fluctuations and elastic constants. *J. Chem. Phys.*, 76(5):2662–2666.

[188] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, and N. Ben-Tal. **2016**. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, 44(W1):W344–W350.

[189] M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, and N. Ben-Tal. **2005**. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic acids research*, 33(suppl_2):W299–W302.

[190] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf. **1995**. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.*, 16(3):273–284.

[191] A. Bondi. **1964**. van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, 68(3):441–451.

[192] M. Z. Tien, A. G. Meyer, D. K. Sydykova, S. J. Spielman, and C. O. Wilke. **2013**. Maximum allowed solvent accessibilites of residues in proteins. *PloS one*, 8(11): e80635.

[193] P. R. Burney, E. M. Nordwald, K. Hickman, J. L. Kaar, and J. Pfaendtner. **2015**. Molecular dynamics investigation of the ionic liquid/enzyme interface: Application to engineering enzyme surface charge. *Proteins: Structure, Function, and Bioinformatics*, 83(4):670–680.

[194] A. D. White, A. K. Nowinski, W. Huang, A. J. Keefe, F. Sun, and S. Jiang. **2012**. Decoding nonspecific interactions from nature. *Chemical Science*, 3(12):3488–3494.

[195] M. Holst, N. Baker, and F. Wang. **2000**. Adaptive multilevel finite element solution of the Poisson–Boltzmann equation I. Algorithms and examples. *J. Comput. Chem.*, 21(15):1319–1342.

[196] H. Cui, K.-E. Jaeger, M. D. Davari, and U. Schwaneberg. **2021**. CompassR Yields Highly Organic-Solvent-Tolerant Enzymes through Recombination of Compatible Substitutions. *Chemistry–A European Journal*, 27(8):2789–2797.

[197] H. Cui, H. Cao, H. Cai, K.-E. Jaeger, M. D. Davari, and U. Schwaneberg. **2020**. Computer-assisted Recombination (CompassR) teaches us how to recombine beneficial substitutions from directed evolution campaigns. *Chemistry (Weinheim an der Bergstrasse, Germany)*, 26(3):643.

[198] D. Woodcock, P. Crowther, J. Doherty, S. Jefferson, E. DeCruz, M. Noyer-Weidner, S. Smith, M. Michael, and M. Graham. **1989**. Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants . *Nucleic Acids Research*, 17(9):3469–3478. ISSN 0305-1048.

[199] F. W. Studier and B. A. Moffatt. **1986**. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology*, 189(1):113–130.

[200] D. Hanahan. **1983**. Studies on transformation of Escherichia coli with plasmids. *Journal of molecular biology*, 166(4):557–580.

[201] O. Edelheit, A. Hanukoglu, and I. Hanukoglu. **2009**. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC biotechnology*, 9(1):1–8.

[202] J.-Y. Jeong, H.-S. Yim, J.-Y. Ryu, H. S. Lee, J.-H. Lee, D.-S. Seen, and S. G. Kang. **2012**. One-step sequence-and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Applied and environmental microbiology*, 78(15):5440–5443.

[203] R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. **1988**. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839):487–491.

[204] F. W. Studier. **2005**. Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification*, 41(1):207–234.

[205] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch, et al. **2005**. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, pages 571–607. Springer.

[206] U. K. Laemmli. **1970**. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature*, 227(5259):680–685.

[207] M. T. N. Petersen, P. Fojan, and S. B. Petersen. **2001**. How do lipases and esterases work: the electrostatic contribution. *Journal of biotechnology*, 85(2):115–147.

[208] G. M. Cooper and C. D. Brown. **2008**. Qualifying the relationship between sequence conservation and molecular function. *Genome research*, 18(2):201–205.

[209] S. Zhang, G. Wu, S. Feng, and Z. Liu. **2014**. Improved thermostability of esterase from *Aspergillus fumigatus* by site-directed mutagenesis. *Enzyme and microbial technology*, 64:11–16.

[210] M. Pezzullo, P. Del Vecchio, L. Mandrich, R. Nucci, M. Rossi, and G. Manco. **2013**. Comprehensive analysis of surface charged residues involved in thermal stability in *Alicyclobacillus acidocaldarius* esterase 2. *Protein Engineering, Design & Selection*, 26(1):47–58.

[211] E. M. Nordwald, G. S. Armstrong, and J. L. Kaar. **2014**. NMR-guided rational engineering of an ionic-liquid-tolerant lipase. *ACS Catalysis*, 4(11):4057–4064.

[212] A. Karshikoff, L. Nilsson, and R. Ladenstein. **2015**. Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *The FEBS journal*, 282(20): 3899–3917.

[213] R. Jaenicke and P. Závodszky. **1990**. Proteins under extreme physical conditions. *FEBS letters*, 268(2):344–349.

[214] A. P. Golovanov, G. M. Hautbergue, S. A. Wilson, and L.-Y. Lian. **2004**. A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society*, 126(29):8933–8939.

[215] L. Giver, A. Gershenson, P.-O. Freskgard, and F. H. Arnold. **1998**. Directed evolution of a thermostable esterase. *Proceedings of the National Academy of Sciences*, 95(22):12809–12813.

[216] N. N. Brandt, A. Y. Chikishev, and I. K. Sakodinskaya. **2003**. Raman spectroscopy of tris-(hydroxymethyl) aminomethane as a model system for the studies of α-chymotrypsin activation by crown ether in organic solvents. *Journal of Molecular Structure*, 648(3):177–182.

[217] A. O. Magnusson, A. Szekrenyi, H.-J. Joosten, J. Finnigan, S. Charnock, and W.-D. Fessner. **2019**. nanoDSF as screening tool for enzyme libraries and biotechnology development. *The FEBS journal*, 286(1):184–204.

[218] R. Geney, M. Layten, R. Gomperts, V. Hornak, and C. Simmerling. **2006**. Investigation of salt bridge stability in a generalized born solvent model. *Journal of Chemical Theory and Computation*, 2(1):115–127.

[219] H.-X. Zhou. **2002**. Toward the physical basis of thermophilic proteins: linking of enriched polar interactions and reduced heat capacity of unfolding. *Biophysical journal*, 83(6):3126–3133.

[220] H. Leemhuis, H. J. Rozeboom, B. W. Dijkstra, and L. Dijkhuizen. **2004**. Improved thermostability of Bacillus circulans cyclodextrin glycosyltransferase by the introduction of a salt bridge. *Proteins: Structure, Function, and Bioinformatics*, 54(1): 128–134.

[221] T. A. Whitehead, C. K. Bandi, M. Berger, J. Park, and S. P. Chundawat. **2017**. Negatively supercharging cellulases render them lignin-resistant. *ACS Sustainable Chemistry & Engineering*, 5(7):6247–6252.

[222] Y. Liu, L. Huang, L. Jia, S. Gui, Y. Fu, D. Zheng, W. Guo, and F. Lu. **2017**. Improvement of the acid stability of *Bacillus licheniformis* alpha amylase by site-directed mutagenesis. *Process Biochemistry*, 58:174–180.

[223] R. Chiuri, G. Maiorano, A. Rizzello, L. L. Del Mercato, R. Cingolani, R. Rinaldi, M. Maffia, and P. P. Pompa. **2009**. Exploring local flexibility/rigidity in psychrophilic and mesophilic carbonic anhydrases. *Biophysical journal*, 96(4):1586–1596.

[224] P. A. Fields. **2001**. Protein function at thermal extremes: balancing stability and flexibility. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 129(2-3):417–431.

[225] G. Raghunathan, S. Sokalingam, N. Soundrarajan, B. Madan, G. Munussami, and S.-G. Lee. **2013**. Modulation of protein stability and aggregation properties by surface charge engineering. *Molecular BioSystems*, 9(9):2379–2389.

[226] A. Villaverde and M. M. Carrió. **2003**. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnology letters*, 25(17):1385–1395.

[227] D. M. Francis and R. Page. **2010**. Strategies to optimize protein expression in E. coli. *Current protocols in protein science*, 61(1):5–24.

[228] C. Castro-Martínez, S. Luna-Suárez, and O. Paredes-López. **2012**. Overexpression of a modified protein from amaranth seed in *Escherichia coli* and effect of environmental conditions on the protein expression. *Journal of biotechnology*, 158(1-2):59–67.

[229] M. T. Reetz. **2013**. The importance of additive and non-additive mutational effects in protein engineering. *Angewandte Chemie International Edition*, 52(10):2658–2666.

Danksagung

An erster Stelle geht mein besonderer Dank an Frau Prof. Dr. Birgit Strodel für die Bereitstellung dieses interessanten Themas und Unterstützung durch regelmäßige Gespräche und hilfreiche Ratschläge. Ich freue mich sehr darauf, meinen PhD in ihrer Arbeitsgruppe machen zu dürfen.

Herrn Prof. Dr. Karl-Erich Jaeger danke ich sehr dafür, dass ich den experimentellen Teil der Arbeit im IMET durchführen durfte und für das Erstellen des Zweitgutachtens.

Ich möchte mich außerdem bei Dr. Stephan Thies für die Betreuung während meiner Zeit im IMET und für viele hilfreiche Diskussionen bedanken. Bei Dr. Alexander Bollinger möchte ich mich besonders für die großartige Betreuung im Labor, für ausgiebige Diskussionen und für die Unterstützung während des schriftlichen Teils der Arbeit bedanken. Rebecka Molitor bin ich für ihre Hilfe und ihre aufgeschlosse und fröhliche Art im Labor dankbar.

Ein großer Dank geht an Anna Jäckering, welche sich viel Zeit für Fragen, Diskussionen, Ratschläge, Probleme, Lösungen, Korrekturen und motivierende Gespräche genommen hat. Außerdem möchte ich allen Mitgliedern der Arbeitsgruppe Strodel für ihre Hilfe bei verschiedensten Fragen zu Rechenclustern oder Mac Problemen danken.

Ich bin dankbar für die Rechenunterstützung und Infrastruktur, welche durch das 'Zentrum für Informations- und Medientechnologie' (ZIM) der Universität Düsseldorf, dem 'Jülich Supercomputing Centre' (JSC) des Forschungszentrum Jülich und dem 'Regionalen Rechenzentrum' (RRZK) der Universität zu Köln bereitgestellt wurde.

Abschließend möchte ich meiner Familie und besonders meinen Eltern für ihre liebevolle Unterstützung während der gesamten Zeit des Studiums danken. Meinen Freunden bin ich für ihre Geduld und für viele schöne aufmunternde Gepräche dankbar.

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Köln, April 20, 2021

Lara Scharbert