INAUGURAL-DISSERTATION

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich-Mathematischen Gesamtfakultät

der Ruprecht-Karls-Universität

Heidelberg

vorgelegt von

Christopher Kling M.Sc.

geboren in Schwalmstadt, Deutschland

Heidelberg 2023

Tag der mündlichen Prüfung:

A Molecular Flip:

How Polyols Affect the Stability and Activity of an Industrial Relevant Subtilisin Protease in Liquid Formulations

Der Einfluss von Polyolen auf die Stabilität und Aktivität einer Industrie-relevanten Subtilisin-Protease in Flüssigformulierungen

Gutachter: Prof. Dr. Gert Fricker

Prof. Dr. Walter Mier

This study is the result of the work carried out at BASF SE in Ludwigshafen

2020-2023

Eidesstattliche Versicherung gemäß § 8 der Promotionsordnung für die Naturwissenschaftlich-Mathematische Gesamtfakultät der Universität Heidelberg

1. Bei der eingereichten Dissertation zu dem Thema:

“A Molecular Flip: How Polyols Affect the Stability and Activity of an Industrial Relevant Subtilisin Protease in Liquid Formulations”

handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich ~~wie folgt~~/bislang nicht an einer Hochschule des In oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

.....
Ort und Datum Unterschrift

Acknowledgments

An erster Stelle möchte ich mich bei Herrn Prof. Dr. Fricker für die Möglichkeit bedanken, meine Dissertation extern in der Industrie anfertigen zu dürfen. Mein Dank gilt seiner Betreuung sowie seiner stets uneingeschränkten unproblematischen Hilfe, Ideen, Ratschläge und Anregungen. In diesem Zusammenhang bedanke ich mich auch bei Herrn Prof. Dr. Preuß und Herrn Prof. Dr. Sauer der Hochschule Mannheim die den Kontakt zur Universität Heidelberg hergestellt haben und diese Arbeit unterstützt haben. Prof Dr. Mier danke ich für seine Bereitschaft als Zweitgutachter zu fungieren.

Ganz besonders möchte ich mich bei Frau Dr. Grit Baier für Betreuung, Unterstützung und das mir entgegengebrachte Vertrauen bedanken. Danke, dass du mir die Türen geöffnet hast und diese Arbeit in dieser Form ermöglicht hast.

Des Weiteren möchte ich mich bei den aktuellen und ehemaligen Laborleitern des Team Biological Actives, Dr. Sonja Kübelbeck, Dr. Anna Hoschek, Dr. Rute André und Dr. Katrin-Stephanie Tücking und PostDoc Dr. Christina Leichner für ein stets offenes Ohr und ihre Hilfe bedanken. Den Laboranten Anja Wieske, Olga Pinneker, Kornelia Slawinski, Miriam Roth und Bettina Drescher danke ich für Ihre Unterstützung und eine oftmals willkommene Ablenkung im Labor. Ein Dank geht auch an das restliche Team, das mich herzlich aufgenommen hat und mich stet motivierend begleitet hat.

Ebenfalls möchte ich mich bei den „Modellern“, Dr. Eduard Schreiner, Dr. Dmitry Suplatov, Dr. Priya Anand und Dr. Stephan Köhler bedanken, die mir die zuvor unbekannte Welt der „Molekular-Dynamik“-Simulationen nähergebracht haben und mit ihrem Wissen, ihrer Unterstützung und Hartnäckigkeit diese Dissertation signifikant bereichert haben.

In gleichen Maßen bedanke ich mich bei Dr. Rafael Teixeira Freire, der mich an seinem unglaublichen Wissen im Bereich der Chemometrie hat teilhaben lassen und die Möglichkeiten der multivariaten Datenanalyse nähergebracht hat. Danke für deine uneingeschränkte Erreichbarkeit und Hilfsbereitschaft. Unsere regelmäßig überzogenen Treffen werden mir definitiv in Erinnerung bleiben.

Ich bedanke mich auch bei Dr. Stefan Jenewein, der stets ein offenes Ohr für meine Enzym-bezogenen Fragen hatte und der mit seinem unglaublichen Wissen und Ideen einen wesentlichen Beitrag zum Erfolg dieser Arbeit beigetragen hat. In diesem Zusammenhang

danke ich auch Dr. Jeremy Sloan und Dr. Konstantin Schneider, die ihr Wissen im Bereich der Enzymaufreinigung und -analytik mit mir geteilt haben.

Des Weiteren bedanke ich mich bei meinem ehemaligen Gruppenleiter Dr. Frank Runge sowie Abteilungsleiter Dr. Jürgen Barwich und später Dr. Carla Seidel für die Möglichkeit der Anfertigung meiner Doktorarbeit in der BASF.

Abschließend möchte ich mich bei meiner Freundin Sonja und meiner Familie bedanken, die mich während meines Studiums und vor allem während meiner Zeit als Doktorand unterstützt, aufgebaut und motiviert haben. Ihr seid einfach die Besten!

„Fantasie ist wichtiger als Wissen, denn Wissen ist begrenzt“

-Albert Einstein-

Table of Contents

Zusammenfassung

Enzyme weisen aufgrund ihres natürlichen Ursprungs eine hohe Anpassung an die physiologischen Bedingungen in lebenden Systemen auf. Folglich steht ihrer Anwendung in industriellen Prozessen häufig eine zu geringe Stabilität entgegen. Neben der gezielten Modifikation der Enzymstruktur, Immobilisation an Oberflächen, oder Einkapselung, kann die gezielte Anpassung des umgebenden Mediums zu einer erhöhten Enzymstabilität beitragen. Dementsprechend werden Flüssigformulierungen nach der Fermentation, im *Downstream Processing* oder vor der Lagerung, Additive wie Polyole zugesetzt. Dadurch soll eine möglichst hohe Stabilität und langanhaltende Aktivität sichergestellt werden.

Die Lagerstabilität wird durch die Bestimmung des aktiven Enzymgehaltes zu mehreren Zeitpunkten überprüft. Dadurch lässt sich der Effekt verschiedener Additive untereinander vergleichen. Als Alternative zu herkömmlichen photometrischen Assays und analytischen Methoden galt es in diesem Zusammenhang das MIRA-Infrarot-Spektrometer zur Bestimmung der Enzym-/Lösungsmittelkonzentration zu validieren. MIRA vereint klassische IR-Spektroskopie und KI-gestützte Ansätze zur reproduzierbaren Quantifizierung von Analyten in wässrigen Umgebungen. Durch die Verwendung multivariater Datenanalyse und des entwickelten Arbeitsablaufs zur Reduzierung des Einflusses des Lösungsmittelsignals durch die Verwendung von Differenzspektren und *External Parameter Orthogonalization* (EPO), konnten auf Basis verschiedener Subtilisin-Protease/Endocellulase-Formulierungen mit 1,2-Propandiol, PCR/PLS-Modelle mit einer bislang unerreichten Genauigkeit generiert werden. Bei der Bestimmung der Enzym-/Lösungsmittelkonzentration in Flüssigformulierungen mittels IR-Spektroskopie handelt es sich um einen vollständig neuen Ansatz, der im Vergleich zu herkömmlichen Methoden mehrere Vorteile aufweist. Dazu gehören eine fehlende Notwendigkeit der Probenverarbeitung, die Schnelligkeit der Methode und der hohe Informationsgehalt einer Messung, der in Kombination mit moderner Datenverarbeitung weitere Messmethoden überflüssig macht.

Da Unterschiede in Lagertests, je nach Enzym und Formulierung, erst nach Tagen, Wochen oder Monaten ersichtlich werden, hat sich die Bestimmung der thermischen Stabilität mittels dynamischer Differenzkalorimetrie (DSC) als Alternative etabliert. Allerdings ist die thermische Stabilität in Formulierungen, insbesondere bei Proteasen, nur bedingt mit Lagertestergebnissen in Einklang zu bringen. Die Ursache für die Diskrepanz galt es anhand von DSC-Messungen, Molekulardynamik (MD)-Simulationen und

Langzeitlagertest-Studien näher zu untersuchen. In diesem Zusammenhang wurde der Einfluss von Polyolen (u.a. Glycerin, Sorbitol und 1,2-Propandiol) und deren Konzentration (10-50 wt%) auf die thermische sowie Langzeitstabilität einer Subtilisin Protease bestimmt. Die ermittelten T_m-Werte der DSC-Messungen zeigten eine starke lineare Korrelation mit molekularen Deskriptoren der MD-Simulationen. Zwischen DSC- und Lagertest-Ergebnissen konnte hingegen kein Zusammenhang ermittelt werden.

Während sich ein möglicher Einfluss der Polyole auf die Anzahl der gebundenen Calcium-Ionen und damit auf Enzymstabilität/-aktivität nicht festgestellt werden konnte, konnte über Enzymkinetiken eine inhibitorische Wirkung von kleineren/hydrophoberen Polyolen auf K_m und V_{max} in Gegenwart des künstlichen Protease-Substrats *Suc-AAPF-pNA* nachgewiesen werden. SDS-PAGE-Analysen ermöglichten es zudem den Anteil der Autoproteolyse am Gesamtverlust aktiven Enzyms in den Lagertestproben zu bestimmen und damit die Wechselwirkung zwischen thermischer Entfaltung und Autoproteolyse nachzuweisen. Anhand der MD-Trajektorien konnte ein Umklappen der katalytischen Histidin-Seitenkette und eine damit verbundene Vergrößerung des Abstandes zwischen dem katalytischen Histidin-$N^{\varepsilon 2}$ und Serin-O^{γ} in Abhängigkeit der Größe und Hydrophilie/Phobie der Polyole als mögliche Ursache bestimmt werden. Folglich konnte erstmals der Nachweis erbracht werden, dass Polyole - neben ihrer teilweise stabilisierenden Wirkung auf die Enzymstruktur - einen Einfluss auf die katalytische Aktivität bzw. Autolyse von Proteasen haben können. Da dieser Effekt in DSC-Messungen nicht ersichtlich wird, liefern die erzielten Ergebnisse neue Erkenntnisse bezüglich der Diskrepanz zwischen DSC- und Lagertestergebnissen. Darauf aufbauend ergeben sich neue Ansätze im Bereich der Enzymstabilisierung und darüber hinaus neue Ideen für die gezielte Anpassung der Enzymstruktur an verschiedene Formulierungsumgebungen.

Abstract

Due to their natural origin, enzymes are highly adapted to the physiological conditions in living systems. Thus, their application in industrial processes is often hindered by insufficient stability. In addition to the targeted modification of the enzyme structure, immobilization on surfaces, or encapsulation, targeted adjustment of the surrounding medium can contribute to increased enzyme stability. Consequently, osmolytes such as polyols are added to liquid formulations after fermentation, in downstream processing or before storage. This is to ensure the highest level of stability and prolonged activity.

Storage stability is monitored by determining the active enzyme content at several points in time. This allows the effect of different additives to be compared with each other. In this context, the MIRA infrared spectrometer was validated as an alternative to conventional photometric assays and analytical methods for determining the enzyme/solvent concentration. MIRA combines classical IR spectroscopy and AI-based approaches for the reproducible quantification of analytes in aqueous environments. By using multivariate data analysis and the developed workflow to reduce the influence of the solvent signal by using difference spectra and external parameter orthogonalization (EPO), PCR/PLS models with a so far unmatched accuracy could be generated based on different subtilisin protease/endocellulase formulations with 1,2-propanediol. The determination of enzyme/solvent concentration in liquid formulations by means of IR spectroscopy is a completely new approach that has several advantages over conventional methods. These include a missing need for sample processing, the speed of the method and the high information content of a measurement, which in combination with modern data processing makes additional analytical methods obsolete.

Since differences in storage tests, dependent on the enzyme and formulation, only become apparent after days, weeks or months, the determination of thermal stability by differential scanning calorimetry (DSC) has established itself as an alternative. However, the thermal stability in formulations, especially for proteases, can only be partially matched with storage test results. The reason for this discrepancy was to be investigated further using DSC measurements, molecular dynamic (MD) simulations and long-term storage test studies. In this context, the influence of polyols (including glycerol, sorbitol and 1,2-propanediol) and their concentration (10-50 wt%) on the thermal and long-term stability of a subtilisin protease was evaluated. The determined T_m values of the DSC measurements

showed a strong linear correlation with molecular descriptors of the MD simulations. In contrast, no correlation could be determined between DSC and storage test results.

While a possible influence of the polyols on the number of bound calcium ions and thus on enzyme stability/activity was not confirmed, an inhibitory effect of smaller/hydrophobic polyols on K_m and V_{max} in the presence of the artificial protease substrate *Suc-AAPF-pNA* could be demonstrated by the characterization of enzyme kinetics. SDS-PAGE analyses further allowed to determine the contribution of autoproteolysis to the total loss of active enzyme in the storage test samples and thus to demonstrate the interaction between thermal unfolding and autoproteolysis. Based on the MD trajectories, a flipping of the catalytic histidine side chain and an associated increase in the distance between the catalytic histidine $N^{\varepsilon 2}$ and serine O^{γ} depending on the size and hydrophilicity/hydrophobicity of the polyols could be determined as a possible cause. Consequently, it could be demonstrated, for the first time, that polyols, in addition to their partial stabilizing effect on the enzyme structure, can have an influence on the catalytic activity or autolysis of proteases. Since this effect is not evident in DSC measurements, the results obtained provide new insights into the discrepancy between DSC and storage test results. Based on this, new approaches in the field of enzyme stabilization and furthermore new ideas for the targeted adaptation of the enzyme structure to different formulation environments arise.

Introduction

Current global crises such as the COVID-19 pandemic and the war in Ukraine, have once again made us painfully aware of how depend modern society is on chemicals and associated products. From the food we eat, to household products or home and personal care, chemicals accompany us wherever we go. However, the production of such, consumes a large number of raw materials and energy. In addition, a large amount of waste is generated, which not only pollutes the environment but also affects our quality of life. In the future, the situation is likely to be exacerbated by a steadily increasing global population and the hunger for resources of emerging markets. As a result, industries around the world are looking for possible alternatives to conventional approaches, that will allow them to produce more products, while using fewer resources and minimizing the influence on the environment [1].

Industrial biotechnology is considered to play a key role in this context, due to the use of bio-based materials and natural production processes [2, 3]. These could be used to either replace or supplement conventional technologies, towards cleaner production processes and sustainable products. In this regard, the use of enzymes in industrial processes and products is seen as one of the promising and sustainable alternatives to conventional approaches [4, 5].

1.1 Enzymes

Albeit unknowingly, the first use of enzymes by humankind dates to the earliest time of civilization. Numerous activities in the first primitive communities such as the production of certain types of foods and beverages, and the tanning of hides and skins to produce leather for garments, were based on the use of enzymes [6, 7]. But only in the late 19th

century enzyme directed research and, subsequently, understanding of enzymes started to develop.

At the beginning of the 19th century the digestion of meat by stomach secretes and the conversion of starch to sugars by plant extracts and saliva were already known, but the mechanism had not been identified [8, 9]. In 1833 the first enzyme, diastase, was discovered by Anselme Payen [10]. In the years after, additional enzymes as pepsin and invertase followed [8, 11]. At the same time (1836) Berzelius first postulated the concept of catalysts which are released unchanged from reactions [10]. A few decades later, Louis Pasteur concluded that the alcoholic fermentation of glucose to ethanol was caused by a vital force within the yeast cells, which he called “ferments” [12]. These, however, were thought to function only in living organisms, which lead to the term enzyme (Greek: ἔνζυμον “in yeast”) being introduced by Wilhelm Kühne in 1877 [13]. This, however, was contradicted by Hans and Eduard Buchner in 1897, who discovered that glucose was fermented by yeast extracts even though living cells were absent in the mixture. Following, the term enzyme was used to refer to nonliving substances [14].

As knowledge about enzymes, expanded in the 19th century, the first industrial applications emerged. In the 1870’s the Danish chemist Christian Hansen isolated pure rennet from calves’ stomachs [7, 15]. The use of this, in the production of cheese, lead to considerable improvements in product quantity and quality. This paved the way for the industrial production of rennet and thus the first enzyme-producing industry ever. The industrial use of enzymes as we know it today started in 1913 when Otto Röhm discovered the efficacy of pancreatic trypsin for the removal of proteinaceous stains from clothing [16].

The development of fermentation processes in the 20th century by James B. Sumner and Kaj Lindstrøm-Lang further boosted this development [17, 18]. Together with the introduction of suitable purification and analysis methods, enzymes could be expressed and purified on a larger scale. Consequently, enzymes advanced into industrial products and processes in detergent, textile, and starch industry in the 1960s [19, 20]. The use of recombinant gene technology in the eighties further improved the manufacturing processes and commercialization of new enzymes. Furthermore, breakthroughs within modern biotechnology, as protein engineering and directed evolution, have further accelerated the development of industrial relevant enzymes [21, 22]. These advances, led to the development of enzymes with improved adaptation to novel process environments and previously non-existent substrate affinity and activity. As a result, enzymatic processes and

products are not only expanding into new areas of application in the detergent, food, animal nutrition, cosmetics, chemical and pharmaceutical industries, but are also advancing into other areas of our everyday lives.

1.1.1 Protein/Enzyme Structure

The ability of enzymes to catalyze chemical reactions is primarily a result of their molecular structure. As proteins, this is characterized by several levels of organization, with an amino acid sequence as their basis.

Primary Structure

The number and type of the amino acids is determined by the sequence of bases, which code for the protein in the desoxyribonucleic acid (DNA). Consequently, each protein is defined by an individual sequence of amino acids, referred to as primary structure. Based on the 22 proteinaceous amino acids a huge variety of different proteins can be expressed [23]. The individual amino acids are linked by peptide bonds, which consist of acid amide bonds. In this regard, an α-carboxyl group of the preceding amino acid is linked with the α-amino group of the following amino acid. As part of this process, a water molecule is released. The resulting sequence of nitrogen atoms, from the amino groups, C_α atoms, from which the side chains branch off, and C atoms, from the carbonyl groups, forms the protein backbone, which has a high potential for hydrogen bond formation [23, 24].

Secondary Structure

The potential of hydrogen bond formation is driven by the interaction of amino groups and carbonyl groups. Structures that emerge from such interactions are described as secondary structures. Among these α-helices and β-sheets were the first ones to be discovered by Linus Pauling and Robert Corey in 1951 [25, 26]. Later, additional structures as the β-turn and the Ω-loop were identified [27-29].

α-Helices are composed of a tightly coiled backbone, which results in a rod like shaped inner structure. The side chains face outward in a helical arrangement. Hydrogen bonds between the amino and carbonyl groups of the main chain stabilize the helix. Each carbonyl group of one amino acid forms a hydrogen bond with the amino group of the

amino acid four residues away in the primary structure [30, 31]. Consequently, every amino and carbonyl group of the main chain, except the one at the beginning and end, are involved in the formation of H-bonds. In this formation, each residue is shifted against the next by 0.15 nm and rotated by 100° along the helix axis. Therefore, each helical turn is composed of 3.6 amino acid residues [32]. Due to this arrangement, amino acids that are three or four residues apart in the primary sequence are in close proximity within the α-helix. The direction of rotation of an α-helix can be clockwise or counterclockwise. However, clockwise helices are energetically more favorable, because they tend to produce fewer steric collisions between the side chains and the backbone [33].

The β-sheet structure is composed of β-strands. In contrast to the tightly coiled polypeptide chain in an α-helix they are almost completely stretched out. Consequently, the distance between neighboring amino acids is 0.35 nm compared with 0.15 nm in α-helices [34]. The orientation of amino acid side chains away from each other in β-strands, leads to the formation of β-sheets, which result from the formation of H-bonds between one or more β-strands. Such β-sheets can have purely the same direction (parallel) or can run in purely opposite directions (antiparallel), or they can be mixed [31]. In the case of the antiparallel orientation amino and carbonyl group of the amino acid in one strand are equally linked to the amino and carbonyl group of a partner in the opposing strand. However, in a parallel orientation the amino group of the amino acid on one strand is linked to a carbonyl group of an amino acid two residues away on the neighboring strand, whereas the carbonyl group is linked to an amino group [35]. In schematic drawings β-strands are commonly depicted as broad arrows. These are orientated towards the carboxyl terminus, to depict the orientation of the β-sheet. In general, ß-sheets have a bigger structural diversity than α-helices and can be relatively flat or they can have a slightly twisted orientation [36].

Since many proteins form a globular structure, turns and loops are required as part of the secondary structure. β-turns for example are formed by the interaction of an amino acid side chain carbonyl group with an amino acid side chain amino group two residues away. In other cases, more sophisticated structures, like Ω-loops are formed [29]. In contrast to α-helices and β-sheets, these are characterized by non-repeating periodic structures. Nevertheless, in many cases they are rigid and well defined. Loops and turns, are not only important secondary structure elements. Far more, they are often involved in interactions between proteins and other molecules. Consequently, they are commonly located on the protein surface.

Tertiary Structure

The tertiary structure of a protein refers to its three-dimensional structure. It may be composed of one or several domains. Respective secondary structure elements, as α-helices and β-sheets, are folded into a compact globular structure [31]. The formation of the tertiary structure is driven by non-specific hydrophobic interactions, in order to bury hydrophobic residues from water [35]. In consequence, polar residues are orientated towards the protein surface. This distinct distribution of polar and nonpolar residues highlights the fundamental aspect of protein architecture [37, 38]. In an aqueous environment, protein folding is driven by the tendency of hydrophobic residues to escape water. The reason for this is that a system is thermodynamically more stable, when the hydrophobic side groups are in close proximity and not exposed to the aqueous environment. Therefore, the formation of the tertiary structure is a spontaneous reaction to hide the hydrophobic side chains in the core of the protein and present polar groups on the surface. The tertiary structure is stabilized by tertiary interactions. These comprise salt bridges, H-bonds, tight packing of side chains, and disulfide bridges. Some polypeptides fold into two or more compact domains, connected by a flexible polypeptide segment. Such separated compact globular units are referred to as domains [35].

Quaternary Structure

For proteins consisting of more than one polypeptide chain, a fourth and highest structural level is considered, the quaternary structure [31, 35]. The quaternary structure is the three-dimensional structure consisting of the aggregation of two or more individual polypeptide chains that operate as a single functional unit. The individual polypeptide chains inside this multimer are often referred to as subunit. The quaternary structure describes the spatial arrangement of these subunits and the nature of their interactions. The simplest case of a quaternary structure is a dimer consisting of two identical subunits. However, a more complex quaternary structure with a different and variable numbers of subunits is possible. The resulting multimer is stabilized by similar non-covalent interactions and disulfide bonds as the tertiary structure. In contrast to the first three structural levels, however, not all proteins have a quaternary structure since some proteins function as a single unit.

1.1.2 Thermodynamics of Enzyme Catalysis

Many chemical reactions can occur spontaneously, others need to be catalyzed to proceed at a significant rate. This is especially important in the case of reactions, which are relevant to sustain live. In this context, the catalyst has to significantly increase the reactivity and velocity of the substance turnover without changing the reaction equilibrium [39]. Enzymes are such proteins, whose catalytic action, referred to as enzyme activity, lead to an increase of the reaction rate by a factor of up to 10^{17} [40]. It is characteristic to such processes, that the catalyst is not consumed or modified. In principle, it can therefore be used without limit. However, the lifetime of a catalyst is limited by its stability under the given reaction conditions.

To ensure a successful reaction, it is mandatory for participating molecules to meet in a certain spatial orientation. Following, the precursor substance is transferred into an activated reactive transition state before it is converted into the reaction product. The energy barrier that has to be overcome in this regard, is referred to as free activation enthalpy (ΔG^{+}), respectively activation energy. The amount of ΔG^{+} required, influences the reaction rate, while the free enthalpy (ΔG) of the reaction, determines whether the reaction takes place spontaneously or not [41]. Enzymes increase the reaction rate, by lowering the necessary activation energy to form the transition state. The binding of the enzyme to its substrate, results in the formation and stabilization of a transition state, so that less energy is required compared to the uncatalyzed reaction. This leads to the product formation upon dissociation of the enzyme-product complex. The energy necessary to decrease the activation energy, derives from the interactions between the substrate and enzyme, which results in free enthalpy, also referred to as binding energy [35, 39].

1.1.3 Enzyme-Substrate Binding

Enzymes are highly specific with regard to the catalyzed reaction, but also in choosing their substrate. Therefore, only a single reaction or a group of related reactions is catalyzed by a specific enzyme. The high specificity of enzymes is mediated by the active site [35, 42]. A region of the enzyme which binds the substrate and catalyzes the formation and release of the product by the formation and dissolution of bonds. Residues in this region are referred to as catalytic group and consist of amino acids, which are often positioned far away from each other in the primary structure of the enzyme. However, upon protein

folding they reach spatial proximity. In general, active sites are often located in cleft-like cavities or cavity-like folds on the surface of enzymes [43].

The binding of the substrate in the active site is induced by the formation of non-covalent interactions (H-bonds, ionic interactions, hydrophobic interactions, and van der Waals interactions). The specificity of the enzyme-substrate interaction is mediated by directed hydrogen bridges and the shape of the active site [44]. In this way, molecules are excluded that are not sufficiently complementary in structure. According to a model developed by Emil Fischer (1894), enzymes have a structure complementary to that of their substrate (lock-and-key principle) [45]. However, some enzymes have shown to undergo a substrate induced conformation change to achieve an optimal positioning of substrate and enzyme. This mechanism was described by Daniel Koshland as “induced fit” and describes the dynamic adaptation of enzyme and substrate based on the conformational flexibility of the enzyme [46]. The interaction between the active site and the substrate facilitates the formation of a transition state in a dynamic process, which is preferentially bound by the enzyme. During this process, water is excluded from the active site and the substrate is bound. The binding is accompanied by conformational changes in the active site, resulting in catalysis.

Substrate specificity ensures that even small differences between substrates are detected, and the right substrate is bound. This can either relate to the substrate as a whole molecule or to specific structural elements of the substrate [41]. Substrates of low molecular weight can be recognized and bound by an enzyme as such. Large molecular substrates as proteins, polysaccharides, and nucleic acids in contrast have defined substructures, which interact with the enzyme. Besides, enzymes are reaction specific, which ensures that only one of many thermodynamically possible reaction types is used for a substrate [47, 48]. If needed, however, enzymes can generate alternating products from one substrate. Further, enzymes are stereo specific, due to which differentiation between mirror-image isomeric substrates is possible [49-51]. At last enzymes can be highly regulated, which enables precise dosing of the enzyme activity in a variety of processes [52-54].

1.1.4 Cofactors

Apart from the catalytic activity of the active site, some enzymes depend on the presence of small molecules, the so-called cofactors. An enzyme without a cofactor is called

apoenzyme [55-57]. The fully activated, catalytic enzyme is referred to as holoenzyme. Cofactors are relevant for chemical reactions, which cannot be catalyzed by the amino acid residues in the active site alone. These include inorganic ions, but also low molecular weight non-proteinaceous organic molecules, which are referred to as coenzymes [58, 59]. Cosubstrates in turn are coenzymes [60], which bind to the enzyme and are released in an altered form after catalysis. The modified cosubstrates are consequently returned to their original state and can be reused in the next catalytic reaction. In many cases, the binding of a cofactor is considered as an inevitable step to obtain an active enzyme or to enlarge its catalytic spectrum. Besides the majority of the enzymes using vitamins as cosubstrates [61-63], two thirds rely on metal ions as cofactors [41, 64, 65]. Metalloenzymes contain metal ions, which are bound to the apoenzyme in a stoichiometric ratio. In contrast, metal activated enzymes bind metal ions specifically, but reversibly. Common metal ion cofactors primarily derive from the group of alkali and alkaline earth metals (Na^+, K^+, Mg^{2+}, Ca^{2+}).

1.1.5 Enzyme Kinetics

Enzyme activity describes the ability of enzymes to increase the velocity of a chemical reaction. All units associated with the enzyme activity are therefore related to the reaction velocity [41]. In general, enzyme activity is determined by measuring the decreasing concentration of the substrate or the increasing product concentration. This is most commonly done by using spectroscopic methods which are based on the Lambert-Beer' law [39, 66, 67]. This means, that the decrease of substrate, respectively the increase in product, is followed by a proportional change in the absorption of monochromatic light (extinction) in the solute (*eq. 1*). Consequently, a prerequisite for this is application is the absorption of monochromatic light by substrate or product. The enzyme activity can then be calculated based on the measured absorbance change per unit time.

(1)

$$E_\lambda = log_{10}\left(\frac{I_0}{I}\right) = c * \varepsilon_\lambda * d$$

In this regard, the extinction E_λ is a dimensionless unit. I_0 and I represent the intensity of the light entering (I_0) the measuring cell respectively leaving it (I) on the other side. ε_λ is the extinction coefficient of the light absorbing substance at the respective wavelength and

d is the path length of the light travelling through the analyte. The variable *c* refers to the concentration of the light absorbing analyte.

The reaction velocity (*V*) (*eq. 2*) is defined to be the change of the substrate respectively product concentration over a defined period of time.

(2)

$$V = \frac{d[P]}{dt} = -\frac{d[S]}{dt}$$

In enzymatic reactions, the reaction velocity often shows a hyperbola dependence on substrate concentration. To describe this behavior Michaelis and Maud Leonora Menten (1912) developed a simplified mathematical model to describe the hyperbola kinetic of enzymes, the Michaelis-Menten model [68]. In a first reversible step, described as pre-steady state, the enzyme (E) and substrate (S) form an enzyme substrate complex (ES) (*eq. 3*). The concentration of such remains almost constant in the following steady state. The concentration of ES depends on the substrate concentration and corresponds to the total enzyme concentration at high levels of substrate.

(3)

$$E + S \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} ES \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} E + P$$

The enzyme substrate complex finally decomposes releasing the product. This leads to a linear increase of the product concentration in the steady state. The release of the product is accompanied by the regeneration of the enzyme, whereby it can repeatedly take part in catalysis.

Kinetics of enzyme catalyzed reactions are most often monitored under so called initial conditions. Under this condition substrate conversion and product release are very small, whereby the velocity of the enzyme substrate complex formation out of enzyme and product is neglectable. Thereby, the reaction velocity (*V*) is proportional to the enzyme-substrate complex ([ES]) concentration (*eq. 4*).

(4)

$$V = k_{f2} * [ES]$$

However, the concentration of the enzyme-substrate complex is not directly measurable. Therefore, the equation can be converted, under consideration of steady state conditions (no major changes in enzyme-substrate concentration), to rely on the enzyme and substrate concentration (*eq. 5*).

(5)

$$\frac{d[ES]}{dt} = \left(k_{f1} * [E] * [S] - \left(k_{r1} + k_{f2}\right) * [ES]\right) = 0$$

In this cause, the free substrate concentration is described by [S], whereas [E] is related to the free enzyme concentration. The total enzyme concentration [E_T] is thereby given by the sum of [E] and [ES] (*eq. 6*).

(6)

$$[E_T] = [E] + [ES]$$

If the total enzyme concentration is considerably smaller than the total substrate concentration, the free substrate concentration is more or less equal to the total substrate concentration. Consequently, equation *5* and *6* can be combined to determine the enzyme-substrate complex concentration under steady state conditions (*eq. 7*).

(7)

$$[ES] = [E_T] * \frac{[S]}{\left(\frac{k_{r1} + k_{f2}}{k_{f1}} + [S]\right)}$$

Further, by taking equation *4* into account, the Michaelis-Menten equation is obtained. This describes the dependency between the reaction velocity and the substrate concentration of an enzyme catalyzed reaction under initial reaction conditions (*eq. 8*).

(8)

$$V = k_{+2} * [E_T] * \frac{[S]}{\left(\frac{k_{r1} + k_{f2}}{k_{f1}} + [S]\right)} = V_{max} * \frac{[S]}{(K_M + [S])}$$

The parameter V_{max} describes the maximum velocity possible and is dependent on the type of enzyme and substrate. With increasing substrate concentrations [S], the reaction velocity (V) gets close to the maximum velocity (V_{max}). Therefore, based on the dependency between reaction velocity and substrate concentration a hyperbolic dependency respectively functional relationship can be observed. In general, enzyme activity is described by the unit U, which is defined to be the turnover (µmol) of substrate to product in one minute [69]. In some cases, however, the unit katal (kat) is used, which is related to the enzyme activity, necessary to turnover one mol of substrate in one second [70]. If the enzyme activity is related to the sample volume, one speaks of the catalytic activity concentration or volume activity. In this regard Unit per milliliter (U/mL) or katal per liter (kat/L) are commonly used [67]. The use of catalytic activity concentrations, is however, highly impracticable when characterizing or comparing different enzymes, since it is not directly associated with the respective enzyme. Therefore, the quotient of the catalytic activity concentration and the protein concentration of the enzyme in solution is used, which is referred to as specific activity. The unit of the specific activity is Unit per milligram (U/mg) respectively katal per kilogram (kat/kg) [67].

The Michaelis-Menten constant K_m refers to the substrate concentration at which the reaction velocity is half V_{max}. In contrast to V_{max}, however, the obtained K_m value is independent of the enzyme concentration. Besides, K_m can be specified using the dissociation constant (K_D) of the enzyme substrate complex (*eq. 9*).

(9)

$$K_m = \frac{(k_{r1} + k_{f2})}{k_{f1}} = K_D + \frac{k_{f2}}{k_{f1}}$$

If the dissociation of the enzyme substrate complex is faster than the release of the product, the Michaelis constant is similar to the dissociation constant. In this case, the K_m value can be regarded as a measure of the affinity of the enzyme for its substrate [41]. Accordingly, an enzyme with high substrate affinity is characterized by a low Michaelis-Menten constant respectively an enzyme with low affinity with a high value.

1.1.6 Enzyme Inhibition

Enzyme activity can be negatively influenced by small molecules or ions. Inhibition of enzymatic processes can play an import role in biologic systems or industrial products and is often utilized in pharmacological applications. In this regard two major inhibition types, reversible and irreversible inhibition, can be distinguished.

An irreversible inhibitor is characterized by a strong binding, either covalent or non-covalent, to the enzyme [71, 72]. Therefore, dissociation of the inhibitor from the enzyme is very slow and the binding of the actual substrate is prevented. Subsequently, binding of an irreversible inhibitor can lead to partial or permanent loss of enzyme activity.

Reversible inhibitors, in contrast, are characterized by the fast dissociation of the enzyme inhibitor complex [73]. In this regard, several different mechanisms can be distinguished. Competitive inhibition *e.g.*, is mediated by inhibitors, which are structurally very similar to the enzyme substrate. Consequently, inhibitor and substrate compete for the active site of the enzyme. Depending on the inhibitor concentration, this leads to different levels of enzyme inhibitor and enzyme substrate complexes. This results in a diminished catalytic rate, since not all enzymes can take part in the reaction. However, by significantly increasing the substrate concentration, inhibitor molecules can be displaced, and the turnover is again increased. Under these conditions V_{max} remains unchanged, while the Michaelis constant K_m is increased [39].

Uncompetitive inhibitors in turn, are characterized by the binding of the reversible inhibitor to the enzyme substrate complex, rather than the lone enzyme [73, 74]. This is mediated by the formation of an inhibitor binding pocket upon substrate binding. In contrast to competitive inhibition, this state cannot be counteracted by increasing the substrate concentration. As before, this leads to a diminished catalytic rate characterized by a decrease of V_{max} and K_m [39].

In the case of non-competitive inhibition, substrate and inhibitor can simultaneously bind to distinct positions of the enzyme [75]. Therefore, non-competitive inhibitors are free to bind the enzyme or the enzyme substrate complex. The inhibitory effect is mediated by the inactivation of the enzyme, rather than preventing the substrate binding or enzyme substrate formation. This achieved by a conformational change of the active site upon inhibitor binding. Since the amount of catalytic enzyme is changed, V_{max} is reduced, while K_m remains unchanged [39]. As in the case of the uncompetitive inhibition the effect

cannot be reverted by an increased substrate concentration since the inhibitor binding site is not influenced by such.

1.1.7 Enzyme Nomenclature

Enzymes can be classified by two main criteria. Either the amino acid sequence similarity and thus their evolutionary relationship or their enzymatic activity. In biochemistry, a hierarchical nomenclature and classification system by the International Union of Biochemistry and Molecular Biology (IUBMB) is applied, based on the reaction catalyzed by the enzyme [76]. The systematic name of an enzyme consists of two parts. The first part of the name specifies the catalyzed substrate, the second specifies the type of reaction catalyzed and ends in "-ase" [77]. Due to this, enzymes that cleave peptide bonds are referred to as proteases or peptidases.

Besides the systematic naming of the enzymes, they are further assigned an Enzyme Commission (EC) -number, which consists of four digits [77]. The first number classifies the enzyme according to seven different main classes. The following numbers refer to the catalyzed reaction and are used for sequential numbering. In the case of the alkaline subtilisin protease, classified as serine protease, this leads to the EC-number 3.4.21.62:

- EC 3: Hydrolases, enzymes that use water to break peptide bonds
- EC 3.4: Hydrolases that act on peptide bonds
- EC 3.4.21: Serine endopeptidases
- EC 3.4.21.62: Subtilisin

1.2 Proteases

Proteases are ubiquitous in biological systems, having multiple functions in the biochemical, physiological, and regulatory aspects of cells and organisms. This is due to their ability to break down proteins by hydrolyzing peptide bonds of the polypeptide chain [78]. Furthermore, proteases represent the largest segment of the industrial enzyme market, as they are used in detergents, food, leather, and fabric processing, as catalysts in organic synthesis, and therapeutics [79-82]. With annual sales of US$1.5-1.8 billion, proteases account for 60 % of the total enzyme market [83].

According to their mode of action, proteases are classified by the IUBMB in the third group (EC3) of Hydrolases [76]. Based on the type of bond that is hydrolyzed they are further subdivided into subgroups from 1 to 13. These subgroups, include hydrolases which *e.g.*, act on ester bonds (3.1), ether bonds (3.3), or among others, peptide bonds (3.4). Based on the site of protease action, they can be further subdivided into exopeptidases or endopeptidases. Exopeptidases catalyze the hydrolysis of peptide bonds near the N- and C-terminal tails of the substrate [84, 85]. In contrast, endopeptidases break the peptide bonds by hydrolysis within the polypeptide chain. Based on the IUBMB nomenclature, exopeptidases can be further subclassified into aminopeptidases, dipeptidases, dipeptidyl-peptidases, tripeptidyl-peptidases, carboxypeptidases, and omega peptidases [86].

Besides the IUBMB enzyme classification, Rawlings *et al*. have introduced a classification scheme, which is based on statistical similarities in sequence and structure of all known proteolytic enzymes [87]. The database behind this classification system has been termed MEROPS. It separates proteases into clans, based on the catalytic mechanism, and families, based on common ancestry. A clan comprises all proteases that share a single evolutionary origin. Therefore, clans can include more than one family of peptidases. Each clan has been assigned a unique identifier consisting of two letters. The first letter represents the catalytic type of the family members, such as aspartic acid (A), cysteine (C), glutamine (G), asparagine (N), serine (S), threonine (T), metal (M), or mixed residues (P). Further, families are defined to group proteases with homologies on the amino acid sequence level. The different families are termed by a single letter, which refers to the catalytic mechanism in the clan identification. In addition, the letter is followed by a unique number.

1.2.1 Serine/Subtilisin Proteases

Out of many proteases, serine proteases, a group of endopeptidases, are one of the well-established types of proteases. More than a third of all known proteolytic enzymes are serine proteases [88]. According to MEROPS they can be classified into 13 clans and 40 families [87]. The classification as serine protease derives from the nucleophilic serine in the enzyme active site, which attacks the carbonyl moiety of the substrate peptide bond to form an acyl-enzyme intermediate [89]. Serine proteases are widely distributed in nature and are found in all kingdoms of cellular life as well as many viral genomes [88]. Among

others the serine protease family comprises proteases such as chymotrypsin, kexin, carboxypeptidase, and subtilisin [90].

Due to their characteristics, proteases of the subtilisin family, have evolved to one of the most well-known classes of proteases, due to their wide range of thermostability and pH compatibility. The name subtilisin is, originally derived from the species, *Bacillus subtilis*, from which the protease was first isolated [91]. The first enzyme of the subtilisin family, however, *subtilisin Carlsberg* (*subtilisin A*), from *B. licheniformes*, was discovered by Linderstrøm-Lang and Ottensen [92], while they were studying the conversion of ovalbumin to plakalbumin. Till today, several species variants have been isolated, including those from *B. amyloliquefaciens* (*subtilisin BPN'*) [93], *B. lentus* (*subtilisin 147* or *esperase*) [94], *B. alcalophilus* (*maxacal*, *savinase* (N85S)) [94], and *B. subtilis* (subtilisin, *subtilisin E*) [95]. Furthermore, subtilisin proteases have also been identified in eukaryotic and other prokaryotic genii [96, 97].

The first sequence of a subtilisin protease was determined in 1966 when Smith published the full amino acid sequence of *subtilisin BPN'* and *Carlsberg* [98]. Followed by X-ray crystallographic structure elucidation by Wright *et al.* of *subtilisin BPN'* in 1969 [93]. Based on this, subtilisins were classified as separate family of proteases. Before, it had been suggested that subtilisin's were homologous to the pancreatic serine protease, chymotrypsin [99]. However, based on sequential and structural information, it could be shown that subtilisin's are serine proteases with a catalytic triad composed off aspartic acid, histidine, and serine residues rather than the chymotrypsin, respectively trypsin characteristic order, histidine, aspartic acid, and serine.

When expressed, subtilisin proteases are secreted in a precursor state, which is referred to as preprosubtilisin [100]. This is characterized by a significantly larger size (up to 381 amino acids) than the active enzyme. The preprosubtilisin is composed of the 29 residue signal peptide, relevant for secretion, followed by the 77 residue propeptide, and the mature subtilisin of varying size. The propeptide has shown to play a vital role for the activity and the regulation of subtilisin [54]. By mutating the aspartic acid residue at position 32 to asparagine, it could be shown that the mutated prosubtilisin was not further processed and therefore remained inactive [101]. This suggests, that the propeptide acts as an intramolecular regulator for the activation of subtilisin. Furthermore, the propeptide was shown to act as intramolecular chaperone [102], mediating the folding of the enzyme in its active state.

Signal peptide and propeptide are both not part of the active enzyme. Subtilisin proteases are in general monomeric proteins, which are composed of 268-275 amino acid residues [103]. Consequently, masses range from 26.8 to 27.5 kDa. The resulting structural differences lead to a large variation of the pI from 7.8 (*subtilisin BPN'*) to ≈11 (*B. lentus*). Still, structural differences between the members of the subtilisin family are small and their structures are highly superimposable.

The secondary structure of subtilisin consists of a central seven-stranded parallel β-sheet surrounded by nine α-helices, with two additional strands of antiparallel β-sheet near the C-terminus (structure and numbering refers to *subtilisin BPN'*) [103]. The catalytic triad is located near the protein surface with Ser221 and His64 at the end of two nearby α-helices, and Asp32 at the end of a strand of the central β-sheet [89, 104]. The oxyanion hole, which is essential for the stabilization of the negative charge of the tetrahedral intermediate state is comprised of the Asn155 side chain amide group, the backbone amide of Ser221, and the Thr220 side chain hydroxyl group [105]. Substrate binding occurs up on the interaction of a nine residue stretch of the substrate backbone with the subsites S6 to S3' [103]. The subsite S1 is a large open cleft that has residues 125-127 on one side, 153-156 on the other, and the 165-170 loop at the base. Residues 125-127 play an important role during substrate binding as the N-terminal part (P4-P1) of the substrate backbone is bound in between residues 125-127 and 100 and 102, and thereby kept in place.

Further, the enzyme has one low affinity and one high affinity calcium binding site. The high affinity binding site consists of a surface loop opposite the active site. It is formed by three backbone carbonyls (residue 75, 79, and 81), and the side-chain oxygen of Asn77, and is coordinated on either side by side-chain oxygens from Gln2 and Asp41 [106, 107]. The low affinity binding site is formed by backbone carbonyls from Gly169, Tyr171, Val174, the side chain of Glu195, and two water molecules [108].

1.2.2 Mechanisms of Subtilisin Protease Catalysis

Enzyme catalysis is mediated by the interactions between reactive side groups of amino acids in the active site with the substrate. In some cases, additional cofactors are involved. During this process, H-bonds, ionic, hydrophobic, van der Waals interactions and temporary covalent bonds are formed. The variety of interactions can significantly influence the possible number of catalytic mechanisms. In general, three principal

mechanisms can be distinguished [41]: metal ion catalysis, acid base catalysis, and covalent catalysis.

In metal ion catalysis, metal ions are utilized to cover up negative charges, activate water molecules, reversibly take up electrons during redox reactions, and/or they can induce an improved conformation of the substrate towards the enzyme [41]. In acid base catalysis, the side chains of amino acids in the active site function as Brønsted acid or base. Consequently, they can reversibly take up or release protons [41]. Common amino acids in this regard, are histidine, cysteine, tyrosine and lysine. For covalent catalysis, temporary covalent bonds between the functional group of the enzyme and the substrate are formed [41]. The structural basis of covalent catalysis are nucleophilic groups, as serine residues, in the active site of the enzyme. This type of mechanism is especially common in the case of serine proteases.

The protease EC 3.4.21.62 is an extracellular alkaline serine endopeptidase/subtilisin protease, which is expressed by various Bacillus strains and especially *Bacillus subtilis* on an industrial scale. The catalytic mechanism of this serine protease combines acid base catalysis and covalent catalysis [35, 109, 110]. The active site forms a catalytic triad, which is composed of serine (residue 215), histidine (residue 62), and asparagine (residue 32). Upon substrate binding the catalyzed reaction is initiated by the nucleophilic attack of the Ser32 residue towards the targeted peptide bond. Thereby, a tetrahedral intermediate state is formed at the C-atom of the peptide bond. This intrinsically unstable tetrahedral intermediate contains a negative charge on the oxygen atom, originating from the carbonyl group. The charge is stabilized by interactions with the NH-groups of an enzyme region, referred to as oxyanion pocket. This results in the release of the peptide fragment with the amino group of the cleaved peptide bond. This process is mediated by the transfer of a proton from the positively charged His62 imidazole ring onto the respective amino group. At the same time the C-terminal substrate fragment is covalently bound to the Ser32 residue, which leads to the formation of an acyl enzyme intermediate. The function of the β-carboxylate group of Asp215 is to stabilize the imidazolium group of His62 by a hydrogen bond. In the next step, a water molecule is attracted by the His62 residue, whereby a proton is extracted from the water molecule. The generated OH^- ion in turn attacks the carbonyl carbon atom of the acyl group, forming a second tetrahedral intermediate, which is stabilized by the oxyanion pocket. However, this structure disintegrates, which leads to the release of the carboxylic acid ending peptide. As a result, the enzyme is regenerated.

1.3 Enzyme Stabilization

Enzymes are highly adapted to the physiological conditions they originate from. These are, however, considerably different from the conditions used for enzymes in industrial applications, as *e.g.* in detergents, food processing, animal nutrition, cosmetics, pharmaceuticals, and the production of chemicals [111, 112]. The use of enzymes as efficient catalysts is therefore only feasible and productive if the catalysts are stable under the present temperature, extreme pH, or other detrimental influences. Therefore, the need to stabilize enzymes is inevitable. Subsequently, different measures are applied to stabilize enzymes for their intended use. These measures are supposed to counteract the denaturing factors and prevent the loss of enzyme during production, storage, and application.

Denaturation refers to the loss of the tertiary structure of the enzyme, whereby the ordered structure of the protein gives way to an arbitrary structure [113]. Consequently, functional important residues are no longer in proximity, which leads to a loss of activity and/or stability. In some cases, this effect is reversible, if the denaturating influence is removed. In others, it leads to an irreversible loss of active enzyme.

1.3.1 Enzyme Replacement

One simple way to obtain a stable and active enzyme for an industrial process, is to replace a mesophilic enzyme with an extremophile alternative [114, 115]. Extremophiles are organisms that have evolved under extreme conditions and thereby adapted themselves and their metabolisms to the harsh conditions they originate from. Due to this, they produce unique biocatalysts that remain active under conditions that their mesophilic counterpart could not endure. Among other, extremophiles can endure at high and low temperatures (-40 to >130 °C), extreme pH (0-13), high salt concentrations (rainwater to 5 M NaCl), and high pressure (0.3-1200 atm) [116]. Consequently, the use of extremophile biocatalysts have enabled the development of processes in food, paper, detergent, pharmaceutical, and chemical industries, which had been limited or impossible before.

1.3.2 Enzyme Engineering

Instead of exchanging the whole enzyme, another option is to alter parts of the primary structure, to change and optimize its characteristic towards its intended use. In this regard

two general strategies are used for protein engineering: rational protein design and directed evolution.

Rational protein design utilizes detailed knowledge of the enzyme structure and the underlying function, to make targeted changes [117, 118]. Together with sophisticated methods, as molecular modelling, this enables to predict the outcome of changes to the protein structure, and thereby to obtain the desired result. This has the advantage of being relatively inexpensive and technically easy, since site directed mutagenesis methods are well established. However, a major drawback of this approach is the necessity for detailed structural knowledge, which is not always available especially at an early development stage.

Directed evolution in turn, uses mutant libraries which are generated by introducing random changes in the protein structure [119, 120]. This is achieved by random mutagenesis, *e.g.*, by error-prone polymerase chain reaction (EP-PCR) or sequence saturation mutagenesis. Afterwards, mutants with desired properties are selected and exposed to further rounds of evolution. Thereby, natural evolution is mimicked, which can lead to superior mutants compared to rational protein design. With additional methods, as DNA shuffling, the obtained results can be even further improved by combining different parts of superior mutants. Compared to rational protein design, directed evolution needs only very limited amounts of information about the structural identity of an enzyme, nor is it necessary to predict the influence of the different mutations on the enzyme function beforehand. A clear disadvantage, however, is the need for high throughput methods to achieve the desired goal in a reasonable amount of time, which due to high costs for equipment and testing is not feasible for all applications.

1.3.3 Enzyme Functionalization

Besides changing parts of the enzyme to achieve the wanted result, it has been shown that adding covalent chemical modifications to the protein can be beneficial for activity and stability [121, 122]. A large number of available chemical moieties, in combination with different amino acid side chains, provide a multiplicity of possible modifications that can be introduced into the enzyme structure. In this regard, amino acids have been phosphorylated, hydrophilic/-phobic groups have been introduced, or enzymes have been crosslinked with glutaraldehyde to increase stability. However, the chemical modification

of an enzyme comes with the risk of destroying, inactivating, or creating inhomogeneous enzyme populations, due to poor discrimination or insufficient chemistry.

1.3.4 Enzyme Immobilization

Enzyme Immobilization in turn comprises methods that chemically or physically confine enzymes onto or within a support or matrix. By doing so, their full or at least most of their activity is retained. In such a way, product contaminations and the necessity to recover and reuse the enzyme can be neglected. The binding mechanism can either be physical, involving weak interactions, or chemical with covalent bonds [123, 124]. Available approaches, include different groups of methodologies like entrapment with nanofibrous polymers, nanoparticles, cross-linked enzyme aggregates, crystals, covalent binding to carrier surfaces or adsorption [125].

Although adsorption to carrier materials is one of the simplest and gentlest methods for enzyme immobilization it is not suitable for all enzymes and applications, since the weak interactions as Van der Waals forces, ionic interactions, and hydrogen bonds are prone to enzyme leakage [123, 126]. The major advantage of such an approach is, that neither additional coupling reagents nor modifications of the protein or carrier is required. Nitrocellulose membranes [127] or polylysine coated slides [128] are *e.g.*, widely used in this regard. However, besides leakage, it occurs that the absorption of proteins to surfaces results in conformational changes and denaturation, which can lead to activity and protein loss.

For more stable attachment, the formation of covalent bonds is mandatory [129, 130]. In most cases, these are formed by the reaction with functional groups on the protein surface. In this regard, additional modifications are not necessary since the naturally present functional groups, as lysine side groups, are readily available and reactions partners as N-hydroxysuccinimide (NHS), aldehyde groups, cysteine residues or epoxide-functionalized materials can be used.

If the binding of a protein to a carrier material is associated with an adverse effect on the enzyme structure, encapsulation offers a structure preserving alternative. In addition, the shell acts as a barrier that shields the enzyme from other detrimental influences. Well established techniques to encapsulate biological species, such as enzymes, antibodies, and other proteins, are the use of sol-gels respectively silica particles [131], liposomes [132]

and biodegradable polymer nano systems as poly lactic acid (PLA) and poly(-lactic-co-glycolic acid) (PLGA) [133, 134].

1.3.5 Liquide Enzyme Formulations

Beside the active alteration or binding of protective carriers to/or surrounding the enzyme, modifying its environment in the liquid state can be an approach to minimize the rate of activity loss. Especially industrially relevant enzymes are often sold with the promise of being long lived and thus active over a long period of time. Even though manufacturers will recommend optimal storage conditions in many cases, some enzymes will retain their activity for weeks while others diminish in days. Therefore, certain additives are needed following fermentation, downs stream processing and storage. Among others, the addition of inorganic salts, polyols, and sugars to liquid enzyme formulations has shown to be beneficial.

Particularly in liquid formulations, the interaction of enzymes with their environment plays a crucial role, as structural stability is particularly dependent on the interaction between the protein and the solvent molecules. Changes in the environment can not only facilitate changes in the activity and stability, but also selectivity can be altered. Especially the water content can significantly influence the stability of the enzyme [135], by influencing the chemical equilibrium or acting as a reaction substrate. Further does water increase the solvation of polar residues of the enzyme, whereby conformational changes can be facilitated, with a positive or negative effect on activity and stability [136].

The effect of polar and unipolar solvents can be quite diverse with respect to enzyme activity and stability. Polar solvents tend to displace water molecules from the protein surface and take their place, whereby a stiffening of the protein structure is induced. Based on the enzyme, this can have diverse overall effects. A more rigid structure *e.g.*, can lead to an increased structural stability, however, it can also negatively affect the turnover, due to the inhibition of the movement of catalytic residues. Furthermore, polar solvents can interfere with the ionic interactions of the protein, which promote its unfolding [137].

It has been shown that polyols and sugars have a positive effect on protein stability in liquid formulations [138-141]. This effect is mediated by multiple factors, like an increase of interactions, a change in the microenvironment of the protein, or a reduction of the active water content. Especially polyols have shown to increase the interactions between

non-polar amino acid residues [142-144]. This in turn, leads to an increased thermostability, due to a more rigid structure. Furthermore, sugars are known to replace water molecules from the hydration shell of the protein, whereby a protective layer is formed, which acts as a stabilizing shield [145, 146].

In addition, salts can positively, but also negatively, influence the stability of proteins by increasing the ionic strength of the environment [147, 148]. Salts tend to bind to charged groups and dipoles. Consequently, water molecules are pulled away from the enzyme, causing hydrophobic moieties to interact with each other, increasing thermal resistance. However, the loss of local water patterns and the interaction of charged ions with the surface, can also lead to a disruption of the protein structure. The nature of the interactions of different ions with proteins is defined by the Hofmeister series, which describes the adverse effects of salts on the solubility of proteins [149]. Some ions, such as sulfate, phosphate, and ammonium, promote the folding of proteins, intermolecular association and even aggregation, while others, as iodide and guanidinium, promote dissociation and unfolding of the protein chain.

1.4 Aims and Objectives

Enzymes are very efficient catalysts of biological respectively chemical processes and in many cases, they are superior to artificial catalysts. Consequently, it seems attractive to transfer them to industrial applications. However, due to their adaption to their original environment and its conditions, enzymes are very often not well suitable for industrial processes or product applications. This can be counteracted by specific modifications in the enzyme structure or by creating a formulation that provides an enzyme friendly environment. Depending on the application, the enzyme can be stabilized for the desired approach by various methods, such as the development of liquid formulations, encapsulation, or immobilization.

Development of liquid enzyme formulations can be thought of as adding one or more additives to an enzyme. Depending on the number of additives and their concentrations, this will result in *n* different formulations. Subsequently, these are stored for a defined period, at a defined temperature for later use. To check the extent of enzyme stabilization, the residual enzyme activity is used as an evaluation criterion. Therefore, samples are withdrawn from the storage test at different times and the remaining enzyme activity after *x* days is determined relative to the initial value.

In the past various methods have been proposed to accelerate this process by determining the enzyme stability using a single measurement. In this context, differential scanning calorimetry (DSC) has shown to significantly speed up the development process by determining the thermal stability of an enzyme. However, thermal stability measures, depending on the enzyme and especially for proteases, are rarely transferable to results of long-term storage tests.

Therefore, the aim of this work was, to increase the understanding about the missing correlation in between thermal stability and long-term storage test results. For this, the thermal and long-term stability of a subtilisin protease in different formulation environments in the presence of different polyols and concentrations should be investigated. This should broaden the understanding of the influence of polyols on the enzyme structure beyond thermal stability.

To do so, the thermal stability of a subtilisin protease in different polyol formulations, of varying type and concentration, was to be determined by DSC and compared to long-term storage results of the same formulation stored at different temperatures. Furthermore, the

stability of the subtilisin protease was to be assessed on an atomistic scale using MD simulations. Thereby, possible correlations between computational stability studies and laboratory results were to be identified. In the event of a missing correlation in between DSC/MD simulation data, the reasons and mechanisms leading to this observation were to be identified, to improve the future development of liquid formulations in the context of enzyme stabilization.

Besides, another objective of this work was to evaluate the use of the MIRA mid infrared spectrometer in the context of enzyme stability testing. In the majority of the cases the extent of enzyme stabilization is determined based on the content of active enzyme, by photometric or turnover based assays. However, besides requiring a high degree of sample preparation and thus being time consuming, they are also based on a variety of buffers and substrates, which can be costly. Subsequently, in the context of the conducted subtilisin protease stability studies, the use of the MIRA mid infrared spectrometer for the prediction of the active enzyme and solvent concentration was to be evaluated. The twin solution comprised by MIRA, combining classical IR spectroscopy with AI driven approaches to correct the spectral data, was expected to be a promising approach to determine enzyme-related quantities in the context of liquid formulations.

Materials & Methods

2.1 Materials

2.1.1 Chemicals

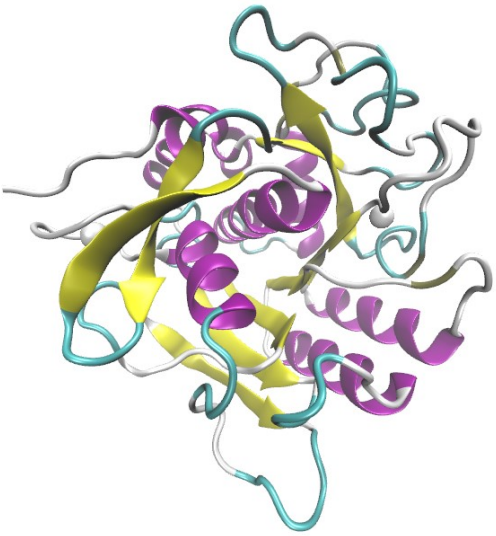
Chemicals	Supplier	Other
1,2-Propanediol	BASF SE	≥99 %; MPG
3-(*N*-morpholino)propanesulfonic acid	Sigma-Aldrich	MOPS
4-Nitroaniline	TCI	>98 %
4-Nitrophenol	Sigma-Aldrich	pNP
4-ntrophenyl-β-D-lactopyranoisde	Goldbio	pNPL
Acetic acid	Bernd Kraft	100 %
Calcium chloride dihydrate	Honeywell	
Coomassie brilliant blue	Sigma-Aldrich	
Dimethyl sulfoxide	Sigma-Aldrich	DMSO
Disodium ethylenediaminetetraacetate dihydrate	Bernd Kraft	EDTA
Erythritol	Borcher	
Ethylene glycol	BASF SE	≥99 %
Ethylenediamine-*N*,*N*′-disuccinic acid	BASF SE	EDDS
Glycerol	Honeywell	
Hydrochloric acid	Carl Roth	
Methanol	Sigma	
Methylglycinediacetic acid trisodium salt	BASF SE	MGDA

Polyoxyethylene(23)lauryl ether	Sigma-Aldrich	Brij® L23
SeeBlue™ Plus2 Prestained Standard	Invitrogen	
Sodium carbonate	Honeywell	
Sodium hydroxide	Sigma-Aldrich	Concentrated
Sorbitol	Fisher Scientific	≥ 99.5%
Suc-Ala-Ala-Pro-Phe-*para*-nitroaniline	Bachem	Suc-AAPF-pNA
Sulfuric acid	Bernd Kraft	Concentrated
Tris(hydroxymethyl) aminomethane	Sigma-Aldrich	≥ 99.5%; TRIS
Xylitol	Sigma-Aldrich	≥99 %

2.1.2 Enzymes

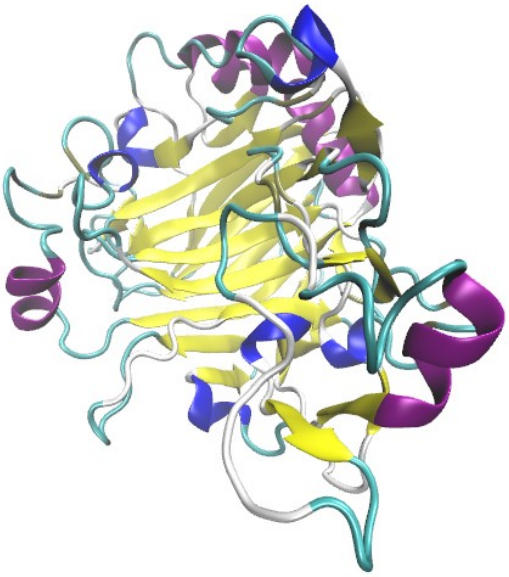
Subtilisin Protease

Genus	Serine endopeptidase
Family	Subtilisin
Origin	*Bacillus lentus*
Amino acid count	269
Molecular weight	26.80 kDa
Catalytic triad	Aspartate (Asp32) Histidine (His64) Serine (Ser221)
Metal Ion Binding Sites	2 x Ca^{2+}
EC-Number	3.4.21.62
MEROPS-Identifier	S08.009
PDB-Structure	1ST3



Endocellulase

Genus	Hydrolase
Family	Endocellulase
Origin	*Humicola insolens*
Amino acid count	398
Molecular weight	44.99 kDa
Active Site	Alanine (Met198) Glutamic acid (Glu202)
Metal Ion Binding Sites	-
EC-Number	3.2.1.4
PDB-Structure	1DYM



2.1.3 Buffers and Solutions

Protease Assay

Buffer / Solution	Ingredients
10x Stabi buffer (TRIS buffer)	In deionized water: 1 M TRIS 1 % Brij® L23 pH adjusted to 8.6 at 30 °C with HCl
Substrate solution	60 mg/mL Suc-AAPF-pNA in DMSO
Starter solution	Gallery: 9.86 mL 1x Stabi buffer 139.1 µL substrate solution (Final Suc-AAPF-pNA concentration 0.8 mg/mL)

Cellulase Assay

Buffer / Solution	Ingredients
MOPS buffer (assay buffer)	In deionized water: 50 mM MOPS pH adjusted to 7.1 at 30 °C with diluted NaOH
Substrate solution	1 mg/mL pNPL in assay buffer

Stop solution	In deionized water: 500 mM Sodium carbonate

SDS-PAGE

Buffer / Solution	Ingredients
NuPAGE MES SDS Running Buffer (pH 7.3) Invitrogen (20x)	1 M MES 1 M Tris Base 2% SDS 20 mM EDTA
Coomassie Staining Solution (0.5 L)	250 mg Coomassie blue (final 0.5 g/L) 250 mL Methanol (final 50 %) 50 mL Acetic acid (final 10 %) 200 mL Deionized water
Distaining Solution (0.5 L)	25 mL Methanol 35 mL Acetic acid 440 mL deionized water

2.1.4 Instruments

Instrument	Manufacturer
Äkta pure™ Chromatography System	Cytiva
ChemiDoc™ MP Imaging System	Bio-Rad
CLARIOstar Plus (Plate Reader)	BMG Labtech
Degassing Station	TA Instruments
Fraction Collector	Cytiva
Gallery Analyzer (Automated Photometric Analyzer)	ThermoFisher Scientific
MIRA mid infrared spectrometer	CLADE
nanoDSC	TA Instruments
SH-642, Bench-top Type, Temperature & Humidity Chamber	ESPEC
ThermoMixer/Shaker® C	Eppendorf

XCell Sure Lock™ Gel Chamber	Invitrogen

2.1.5 Consumable Materials

Consumables	Manufacturer
Stericup®, Millipore Express® PLUS, 0.22 µm	Merck KGaA
NuPAGE™ 4-12% Bis-Tris Gel (1.0 mm x 10 well)	Invitrogen
Steriflip®, Millipore Express® PLUS, 0.22 µm	Merck KGaA
50 mL Tube	Sarstedt
5 mL Eppendorf Tube®	Eppendorf
1.5 mL Safe-Lock Tubes	Eppendorf
Spectra/Por 3 RC molecularporous membrance tubing / Dialysis Membrane	Spectrum™
HiLoad® 16/600 Superdex® 75 pg	Cytiva
96-Well Deep Well Plate	Brand
UV-STAR® Microplate 96 Well, F-Bottom	Greiner Bio-One
2 mL Glass Crimp Top HPLC Vials	Thermo Fisher Scientific
Syringe Filters Millex® 0.22 µm	Merck KgaA

2.1.6 Software

Software / Programming Language	Company/Publisher
nanoAnalyze software suite (Version 3.10.0)	TA Instruments
Python 3.9.12 • LMFIT 1.1.0 • Pandas 1.5.2 • Matplotlib 3.6.2 • Scipy 1.9.3 • Sklearn 1.2.0 • Numpy 1.23.5	-
ImageLab 6.1	BioRad

Packmol 20.0.0	University of Campinas/São Paulo [150]
Propka 3.4.0	Søndergaard & Olsson [151, 152]
Visual Molecular Dynamics (VMD) • Autoionize Plugin (Balabin et al.)	University of Illinois at Urbana-Champaign / Theoretical and Computational Biophysics Group [153]
Nanoscale Molecular Dynamics (NAMD) v2.11	University of Illinois at Urbana-Champaign / Theoretical and Computational Biophysics Group [154]

2.2 Methods

2.2.1 Formulation Preparation, Storage, and Sampling

Formulations were prepared, if not stated differently, by adding appropriate amounts of solvent and $CaCl_2$ to a 50 mL reaction tube. Subsequently, the enzyme in water was added and if necessary, the pH was adjusted using hydrochloric acid (HCl) or sodium hydroxide (NaOH). Before use, samples were sterile filtrated (Steriflip®, Millipore Express® PLUS Membrane (0.22 µm)).

For storage, samples were split into 5 mL Eppendorf reaction tubes under sterile conditions and transferred into temperature and humidity chambers with the desired temperature (4-45 °C) and ambient humidity.

At certain times during storage, samples were withdrawn (200 µL) under sterile conditions and transferred into a 1.5 mL Eppendorf reaction tube. Until further use samples were stored at -80 °C

2.2.2 Enzyme Activity Assays

Background

The enzyme activity respectively the active enzyme concentration of different enzymes was determined using photometric activity assays. In this regard, ultraviolet/visible (UV/Vis) spectroscopy was used to determine the activity of an endocellulase and subtilisin protease. Both approaches utilize the presence of a light absorbing moiety, which is released during the turnover of an artificial substrate in the presence of the enzyme.

The amount of the released moiety is determined by using electromagnetic waves in the range of ultraviolet and visible light. These are guided through the samples by mirrors and

the remaining radiation is analyzed by a detector. As the waves pass through, energy is transferred to the sample. The effect is mediated by the band structure of the samples, where photons of certain energy excite atoms or molecules that have quantum transitions with the same energy difference in the electron shell or their molecular vibrations (*e.g.* infrared light). As a result, the intensity of the UV or visible light after the sample is decreased, when compared to the primary beam. Following, the intensity of the light before (I_0) and after (I) the sample respectively transmittance (T) can be used to determine the absorbance (A) (*eq. 10*) [155, 156].

(10)

$$A = log_{10}\frac{I_0}{I} = -log_{10}T$$

The determined absorbance, in turn, correlates with the amount of released moieties, which is further dependent on the amount of enzyme present in the samples.

Protease Assay

The protease activity/active enzyme concentration was determined with an assay based on the cleavage of an artificial peptide substrate from DelMar *et al.* [157]. The artificial substrate is composed of a short peptide sequence (*Ala-Ala-Pro-Phe*), which is linked to a succinyl residue (Suc) at the N-terminus and a *para*-nitroaniline molecule (pNA) at the C-terminus. Upon cleavage of *Suc-AAPF-pNA* by a protease pNA is released and absorption can be measured at 405 nm. Consequently, absorption increases with the activity/active enzyme concentration of the protease present in the sample.

Prior to the measurement, 10 x Stabi-buffer was prepared according to *chapter 2.1.3*. The used ingredients were dissolved in deionized water and the pH was adjusted to 8.6 at 30°C, using concentrated hydrochloric acid (HCl). Thereafter, the buffer was sterile filtrated (Stericup®, Millipore Express® PLUS, 0.22 µm PES) and diluted to onefold buffer with deionized water.

Using onefold reaction buffer, the standard, with a known enzyme concentration (50 g/L), and samples were diluted. In this regard 200 µL of the standard were filled up to 25 g with 1x reaction buffer. Following, the standard was diluted 1:20 once and samples twice in a ratio of 1:10. The first dilution was done gravimetrically and the second volumetrically.

Further, assay substrate was prepared according to *chapter 2.1.3*. Afterwards, standard, samples, buffer, and starter solution were transferred into the Gallery™ automated photometer system and the analysis was started according to *Table 1*.

Each sample was measured in three different dilutions in between 1:1-1:15. During the analysis, each sample was analyzed four times, which returned the absorption change per minute (Abs/min). The linear range of the AAPF assay is between 0.04 and 0.25 Abs/min, which correlates with approximately 0.10-0.67 AAPF-Units/g. Values above or below the linear range were discarded. The aberration of the calculated activity of the standard and its index value of 6000 AAPF-Units/s were determined, and sample activities were corrected by the relative aberration. Following, the enzyme concentration (g/L) of the samples was calculated by determining the AAPF-Units/g, by considering the dilution factor of the samples, and by dividing the obtained value by the specific activity of the used protease.

Table 1: AAPF-assay protocol for the Gallery™ automated photometer.

Step	Description	Time [s]	Volumne [µL]	Temperatue [°C]
1.	Collect assay substrate	-	184	-
2.	Incubate assay substrate	300	-	30
3.	Dilute sample (1:1 – 1:15) and add to assay substrate	-	8	-
4.	Kinetic Measurement (405 nm, 4 measurements)	60	-	30

Cellulase Assay

The residual endocellulase concentration in the conducted experiments was determined using a 4-nitrophenyl-β-D-lactopyranoside (pNPL) based assay [158]. Cellulases hydrolyze pNPL to p-nitrophenol (pNP). The quantity of pNP liberated per unit of time is measured based on the mediated absorbance at 405 nm. The absorbance is used to calculate the activity based on a pNP calibration curve.

Prior to the measurement MOPS buffer was prepared according to *chapter 2.1.3*. The used ingredients were dissolved in deionized water and the pH was adjusted to 7.1 at RT, using diluted sodium hydroxide (NaOH). Thereafter, the buffer was sterile filtrated (Stericup®, Millipore Express® PLUS, 0.22 µm PES) and stored at room temperature until further use.

Using the MOPS buffer, the standard with a concentration of 3.94 mg/mL active endocellulase was diluted, by adding 0.1 mL of it to a 15 mL reaction tube and filling the recorded weight to 10 g with assay buffer. In addition, a calibration solution was prepared by adding 0.2 g of pNP in a 15 mL vessel. The weight was recorded and filled up to 5 g using MOPS buffer. Further, samples were prepared by performing gravimetric dilutions to an approximate concentration of 0.04 mg/mL.

Prior to the measurement, assay substrate and stop solution were prepared according to *chapter 2.1.3*. Subsequently, calibration, standard, samples, buffer, substrate, and stop solution were transferred into the Gallery™ automated photometer system and the analysis was started according to *Table 2*.

Table 2: pNPL-assay protocol for the Gallery™ automated photometer.

Step	Description	Time [s]	Volumne [µL]	Temperatue [°C]
1.	Collect assay (MOPS) buffer	-	40	-
2.	Collect pNPL substrate and mix with 1.	-	20	-
3.	Incubate mixture	180	-	RT
4.	Dilute and add sample (1:1 – 1:24) to mixture	-	20	-
5.	Incubate mixture	270	-	-
6.	Add stop solution	60	-	-
7.	Measure Absorption (405 nm)	-	-	RT

Before analyzing the samples, a dilution series of the calibrator and the standard were conducted, to check whether absorption values were inside the linear range of the assay. The linear range of the pNPL assay is between 0.05 and 0.25 Abs, which corresponds to approximately 0.005-0.04 U/g. Values above or below the linear range were discarded. Subsequently, samples were measured five times and if required further diluted using the Gallery™ device. The Gallery™ directly calculates the enzyme units based on the conducted pNP calibration. The reference material can be used as a positive control to check, whether the correct enzyme unit number is determined. Subsequently, the active enzyme concentration can be calculated by considering the dilution factor and dividing the enzyme units by the specific activity of the endocellulase.

2.2.3 Differential Scanning Calorimetry (DSC)

Background

With differential scanning calorimetry (DSC) changes of the physical properties of a biopolymer can be determined by monitoring the behavior along a temperature gradient. In this regard, the specific heat of a system is measured as a function of temperature at a given scan rate relative to a reference solution. Changes in the protein or enzyme structure throughout this process, lead to the emission or absorption of heat by the sample, when a certain temperature is reached. In general, two types of DSC instruments are used: heat flux and power compensated DSCs [159-161].

Heat-flux DSCs, use a thermoelectric disc as a heat flow sensor. As part of a measurement, sample and reference are placed symmetrically on this disk, which is surrounded by a furnace. When the furnace is heated in a linear fashion, heat is transmitted from the disk onto the samples. Subsequently, temperature in sample and reference increase in the same manner. However, if this steady-state equilibrium is altered by a conformational change of the sample, the difference in temperature between sample and reference is detected as electrical potential [159, 160].

In the case of power-compensated DSC two separate furnaces are used, for sample and reference. During the measurement, both are kept at the same temperature, while running a temperature gradient. In the event of sample transition, the differential power required to maintain the temperature in the sample is detected. Subsequently, the thermogram (*fig. 1*) can be plotted as the required power versus time. From a thermodynamic point, energy

values can be converted to obtain the time to temperature or molar heat capacity [159, 160].

The transition midpoint (T_m) corresponds to the temperature where half of the protein is in its native state, while the other is denaturated. With regard to protein stability testing, the T_m value can be used to compare different enzymes with regard to thermal stability. In general, a higher T_m value corresponds with a higher stability of the enzyme. In addition, in different environments respectively formulations, the influence of these measures on the enzyme stability can be investigated and used for comparison. Further, the molar heat capacity (C_p) can be determined based on the observed peak height. The calorimetric enthalpy (ΔH), in turn, can be obtained by integration of the area under the peak [160].

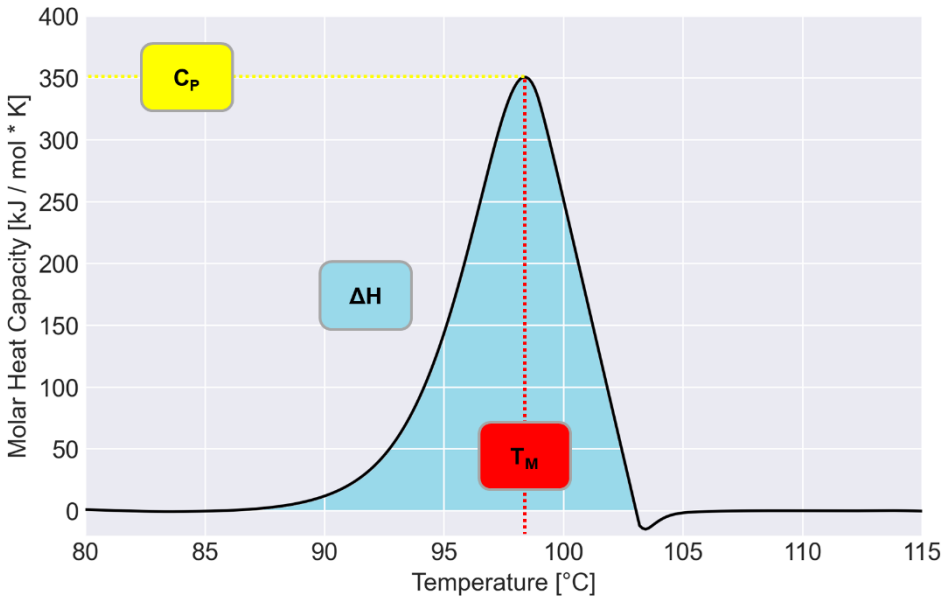


Figure 1: DSC Thermogram with relevant measures as the transition midpoint (T_m), molar heat capacity (C_p), and calorimetric enthalpy (ΔH).

Methodology

For the conducted experiments a power compensated DSC (nanoDSC) from TA Instruments was used. The samples and references were prepared and processed according to *chapter 2.2.1*. Reference (formulation environment without enzyme) and liquid samples (with enzyme) were degassed for 12 min prior to experiment before 300 µL of sample were manually added to the reference/sample cell. If not stated differently, an enzyme concentration of ≈50 g/L was used. Sample and reference were measured with the same

temperature profile up to 105 °C, with a heating rate of 1.5 K/min. The obtained results were analyzed with the nanoAnalyze software suite (Version 3.10.0, TA Instruments). In this regard, thermograms were baseline corrected and the thermal transition temperature (T_m) and the unfolding enthalpy ΔH was determined by fitting gaussian models to the observed peaks.

2.2.4 Molecular Dynamics (MD) Simulations

Background

MD simulations are a powerful computational tool used to study the behavior of materials and bio-/molecules on an atomistic scale. In MD simulations, the position and velocities of individual atoms are calculated based on Newton's equations of motion, which are integrated numerically over time to simulate the motion of the underlying system. By varying the initial conditions and parameters, it is possible to study the behavior of the system under varying conditions and thereby gain insights about the underlying physical processes [162].

MD simulations are commonly used in chemistry and biology to study a wide range of phenomena, such as protein folding/unfolding and drug binding. The simulations can be used to predict the thermodynamic and kinetic properties of a system, such as energy, temperature, pressure, and diffusion coefficients. In addition, MD simulations can provide detailed information about the structure and dynamics of a system, such as bond length, angles, and dihedral angles, as well as the movement of individual atoms over time. The major advantage of MD simulations is the possibility to study the behavior of complex systems, which are difficult to study experimentally. However, these kinds of simulations require significant computational resources, and the accuracy of the simulations depends on the accuracy of the underlying physical model and force fields used to describe the system [163].

MD simulations involve several steps, that are repeated iteratively over a defined period of time. In the first step, the studied system has to be defined, including atomic coordinates, the molecular topology (*e.g.*, bond angles, and dihedral angles), and the initial velocities of the atoms. Next, a potential energy function, which describes the underlying interactions between atoms in the system, is selected. This function can either be empirical or based on quantum mechanics and can be selected from existing force fields or specifically developed for the system being studied. In the initialization step, the simulation parameters

are defined, including temperature, pressure, and box size. In addition, the simulation time step is defined, which determines how frequently the positions and velocities of the atoms are updated. At each time step, the position and velocities of the atoms are used to calculate the forces acting on each atom. The forces are calculated based on the selected potential energy function. Subsequently, the equation of motion for each atom is integrated numerically using a numerical algorithm, such as the Verlet algorithm. The motion of atoms is calculated based on classical mechanics, which assumes that atoms follow laws of motion defined by Newton. The forces acting on each atom are calculated based on the selected potential energy function, which describes the interactions between atoms due to bond stretching, angle bending, and non-bonded interactions, such as van der Waals forces and electrostatic interactions. These forces are used to update the positions and velocities of the atoms. Subsequently, this process is repeated for each time step to simulate the motion of the system over time [162, 164].

System Setup

Simulated systems were setup using the PACKMOL v20.0.0 software package [150], which is used in computational chemistry to generate initial configurations for MD simulations. PACKMOL works by taking as input the geometries of the molecules to be packed, along with their desired number and concentration, and generates an initial configuration that places the molecules randomly within the specified volume. By specifying the volume and number of molecules of the system, the density of polyol concentrations with 50 wt% solvent were considered and reproduced (*fig. 2*). The structure of the subtilisin protease was obtained from the protein database (https://www.rcsb.org/; ID code 1ST3) and solvent/metal ion structures from the Charmm Small Molecule Library (https://www.charmm-gui.org/?doc=archive&lib=csml). The box size for the simulations was set to ≈262.000 Ångström^3 (Å^3) resulting in a box of roughly 64 Å x 64 Å x 64 Å. Further, the protonation state of each enzyme residue at pH 6.0 was predicted with Propka v3.4.0 software tool [151, 152] and adapted in the enzyme structure using the Visual Molecular Dynamics (VMD) v1.9.3 software package [153]. At last, all systems were neutralized by adding sodium and chloride ions (conc. 0.15 mol/L), using the Autoionize v1.5 plugin in VMD.

Ingredient	Concentration [weight %]
Subtilisin Protease	5
Glycerol	50
Water	44.7
$CaCl_2$	0.3

Density
≈1.13 g/mL

Ingredient	Number of Molecules
Subtilisin Protease	1
Glycerol	944
Water	3784
$CaCl_2$	8

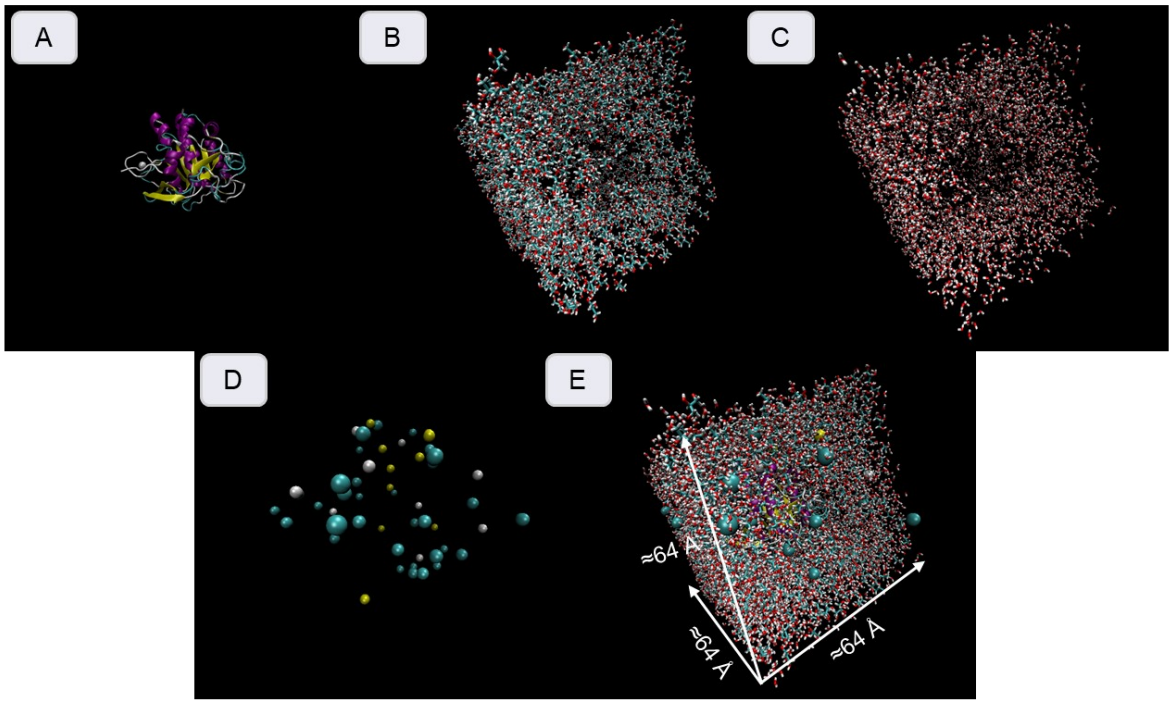


Figure 2: System setup of a subtilisin protease formulation with 50 wt% glycerol. *Shown is the transfer of a laboratory formulation in a system used for MD simulations. (A) Subtilisin protease, (B) glycerol, (C) water, (D) Ca^{2+}, Cl^-, and Na^+, (E) final system composed of molecules A-D*

Simulation Run

MD simulations were run with the Nanoscale Molecular Dynamics (NAMD) v2.11 software package [154], which is suitable for simulations of large biological systems such as proteins, nucleic acids, and membranes. The main integration algorithm used by NAMD is the reversible reference system propagator algorithm (RESPA), which is a multiple step (MTS) integration method. In the case of the RESPA algorithm, forces on the atoms are split into two or more components with different time scales, and the atoms are moved using different time steps for each component. Typically, a shorter time step is used for the fast, high-frequency forces such as bond-stretching and angle bending, while a longer time step is used for the slower, low-frequency forces such as non-bonded interactions.

Forces acting on the atoms were calculated based on the CHARMM36 force field files (Chemistry at Harvard Macromolecular Mechanics). Besides the particle mesh Ewald

(PME) method was used to compute long-range electrostatic interactions under periodic boundary conditions. For temperature control Langevin dynamics were applied to maintain a constant temperature of 373 K throughout the simulations and Nosé-Hoover Langevin piston pressure control to retain a constant pressure of 1.01325 bar. Simulations were run over a period of 200 ns with a timestep 1 fs.

To prevent the two calcium ions from leaving the subtilisin protease structure throughout the simulations, extra bonds were applied from atom 20, 546, 565, 1007, 1030, 1065 and 1088 to the first calcium ion (3752) and 2212, 2257, 2291, and 2574 to the second calcium ion (3752).

Data Evaluation

The obtained trajectories were analyzed by determining a variety of different molecular descriptors, as the solvent accessible surface (SASA), the root mean square deviation (RMSD), the root mean square fluctuation (RMSF), the radius of gyration (RG), and the Q-value. The evaluation was done using the .tcl scripting language in VMD.

SASA is a molecular descriptor that quantifies the surface area of a biomolecule, that is accessible to solvent molecules. The most commonly used method for calculating the SASA is the Lee-Richards algorithm, which was introduced in 1971 [165]. The Lee-Richards algorithm uses a probe sphere to represent a solvent molecule and rolls the probe sphere over the surface of the molecule. The surface area of the molecule, that is in contact with the probe sphere is then calculated and added to the total SASA. This process is repeated for multiple probe sphere sizes to obtain a more accurate estimate of the SASA.

The RMSD (*eq. 11*) is commonly used to measure the similarity in between two structures of a molecule. It quantifies the average distance between the corresponding atoms in two structures, after the structures have been aligned with each other. For trajectory analysis the changed structures throughout the simulation are most often compared to the starting conformation of the enzyme. A low RMSD indicates that the two structures are highly similar, while a high RMSD indicates significant changes of the enzyme structure. Following, the RMSD can be used to determine the magnitude of conformation changes of a molecule, upon changes of the protein folding or upon ligand binding, or to identify regions of a molecule that are rigid or flexible [166].

(11)

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - r_i^{ref})^2}$$

$$r_i = x_i + y_i + z_i$$

$$r_i^{ref} = x_i^{ref} + y_i^{ref} + z_i^{ref}$$

n= number of atoms; r_i^{ref} = coordinates of the i-th atom/residue in the reference structure; r_i= coordinates of the same atom at later timepoint

With regard to the protein or enzyme structure, the RMSF provides information about the flexibility or rigidity of individual residues. It is calculated (*eq. 12*) as the standard deviation of the residue fluctuations, measured as the deviation of the residue position from its average position over a MD simulation trajectory. The RMSF can be used to identify flexible regions in a molecule, such as loops, termini, or active sites, or to compare the flexibility of different regions in a molecule, such as protein domains or ligand binding sites. It can also be used to compare the flexibility of different conformations or mutants of a molecule. In general, a high RMSF indicates that the atomic or residue positions fluctuate significantly over the simulation or experimental ensemble, suggesting that the region is flexible, while a low RMSF indicates that the atomic or residue positions are relatively stable, suggesting that the region is rigid.

(12)

$$RMSF = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(r_i - r_i^{ref})^2}$$

$$r_i = x_i + y_i + z_i$$

$$r_i^{ref} = x_i^{ref} + y_i^{ref} + z_i^{ref}$$

T= number of frames; r_i^{ref} = position of the i-th atom/residue in the reference structure; r_i= position of the i-th atom/residue at later timepoint

The RG can be used to compare the size and shape of different conformations of a biomolecule. It is calculated (*eq. 13*) as the root-mean-square distance between the atoms or residues in a molecule and their center of mass, which represents the overall spread of the molecule's mass around its center. Generally, a large RG indicates that the molecule is more spread out, while a small RG indicates that the molecule is more compact.

(13)

$$r_{gyr}^2 = \frac{(\sum_{i=1}^{n} w_i(r_i - \bar{r})^2)}{\sum_{i=1}^{n} w_i}$$

r_i= position of the i-th atom/residue; $\bar{r}$= is the weighted center; w= weight

The Q-value provides information about the similarity between two structures or conformations of biomolecules. It is calculated (*eq. 14*) as the fraction of native contacts, which are defined as the contacts between pairs of residues or atoms that are in contact in the native, or reference structure of the *e.g.,* enzyme. The Q-value ranges from 0 to 1, where a Q-value of 1 indicates that the two structures are identical in terms of their native contacts, while a Q-value of 0 indicates the two structures share no native contacts. The Q-value is commonly used in the context of protein folding simulations or studies on conformational changes of proteins.

(14)

$$Q(d, d_0) = \frac{1}{1 + e^{\beta(d - \lambda d_0)}}$$

d= contact distance at time t; d_0= contact distances at time t= 0; β= softness of the switching function (5 Å); λ= reference distance tolerance (1.8)

2.2.5 Complexation and Dialysis

In the cause of different experiments, the calcium content in the subtilisin protease structure and related formulations was reduced by adding chelators in a concentration of 5 wt%. To prevent the subtilisin protease from denaturating upon addition of ethylenediamine-*N*,*N*′-disuccinic acid (EDDS), ethylenediaminetetraacetic acid (EDTA), and methylglycine-diacetic acid trisodium salt (MGDA) the pH level of the chelators, solubilized in deionized water, was adjusted to pH 6.0.

In some cases, the chelator was removed after complexation, to study the stability of the subtilisin protease in its absence respectively after the removal of calcium in and around the protein structure. For this, 100 mL of treated sample were dialyzed, by placing the subtilisin-protease containing solution into a dialysis membrane with a molecular weight cut off of 3.5 kDa. Subsequently, the filled dialysis membrane was transferred into 10 L of deionized water at a temperature of 4 °C. Based on the resulting concentration gradient, chelators and calcium were to be driven out of the protein-containing solution. The dialysis was conducted over a period of 36 h. During this time, deionized water was exchanged four times. To reduce the loss of active subtilisin protease by autolysis, dialysis was performed in a cooling space at 4 °C and exchange water was precooled before use.

2.2.6 Size Exclusion Chromatography (SEC)

Background

Size exclusion chromatography (SEC), also known as gel permeation chromatography (GPC), is a type of liquid chromatography used for the separation and purification of biological molecules based on their size. SEC is a gentle and non-destructive method that can be used under a variety of conditions and in both analytical and preparative modes. Subsequently, it is commonly used in the production of biopharmaceuticals, as well as in research laboratories for protein purification and characterization. In principle, a sample is injected into a column containing a stationary phase made up of porous beads with defined pore size. Smaller molecules tend to enter the pores, while larger molecules are excluded and pass around the beads. As a result, smaller molecules are retained in the column for a longer period of time, while larger molecules elute faster [167, 168].

Methodology

In the conducted experiments a HiLoad® 16/600 Superdex® column in an Äkta pure chromatography device was used for the separation of different fractions in subtilisin protease samples. To do so 500 µL of a sample with ≈30 g/L active enzyme was applied to the column and eluted (H_2O) with a flow rate of 1 mL/min. The released fractions were analyzed by absorbance readings at a wavelength of 280 nm. The SEC was conducted at 4°C and obtained fractions were transferred into a 96 deep well plate by an automated fraction collector. Until further use samples were stored -80 °C.

2.2.7 Michalis-Menten (MM) Kinetics

Background

The Michalis-Menten (MM) kinetics is a widely used mathematical model for describing the rate of enzymatic reactions. The model is based on the theory of quasi-steady-state enzyme kinetics, which include the assumption that the rate-limiting step of the reaction is the formation and breakdown of the enzyme-substrate complex. Furthermore, the enzyme concentration is assumed to be significantly lower when compared to the substrate concentration. For a more in-depth description see *chapter 1.1.5* [68].

Methodology

To experimentally determine the kinetic parameters of the used subtilisin protease in different buffer/polyol-environments, varying AAPF substrate concentrations were added to a defined concentration of the subtilisin protease.

MM-Assay

In this regard, prior to the experiment different buffer/polyol-environments were prepared to be tested. Therefore, 10x Stabi buffer was diluted to 1x with the respective amount of glycerol, sorbitol and 1,2-propanediol (MPG), as to achieve concentrations of 15, 45 and 75 wt% polyol, and filling the rest with deionized water.

Consequently, a dilution series (n= 11) of an enzyme substrate solution in a concentration range from 0.06-3.38 mM AAPF substrate in the different buffer/polyol-environments and a subtilisin protease standard with a concentration of 1 µg/mL, in 1x Stabi buffer was prepared.

Following, the different substrate dilutions were added (100 µL) to a 96-well plate in triplicates. Besides, a blank was added as triplicate, consisting of the respective buffer/polyol-environment without the AAPF substrate. 50 µL of the protease standard were added to all wells, resulting in a total volume of 150 µL per well and a final concentration of polyol of 10, 30 and 50 wt%.

Right after, plates were thoroughly mixed on a thermoshaker and transferred into a plate reading device. The plates were incubated for additional 3 min at 30 °C, shaken and the extinction was determined over a period of 10 min, every minute, at 405 nm.

Extinction Coefficient and Specific Activity

Based on the measured extinction increase per minute, the specific activity of the subtilisin protease in the respective buffer/polyol-environment and in presence of a defined substrate concentration was calculated according to (*eq. 15*).

(15)

$$Specific\ Actvitiy = \frac{\Delta E}{\Delta t} * \frac{V}{\varepsilon_\lambda * d * v_E * c_E} * 10^6 \quad \left[\frac{U}{mg}\right]$$

In this case, V (150 µL) and v_E (50 µL) represent the total volume respectively the volume of the enzyme stock solution added to each well. Further, d (0.43 cm) represents the optical path length and c_E (1 µg/mL) the enzyme concentration of the enzyme stock solution. The extinction coefficient ε_λ was determined prior to the experiment for the different buffer/polyol-environments.

Therefore, different concentrations (n= 10) of the *para*-nitroaniline (pNA) moiety (0.0014-0.18 mM) were dissolved in 10, 30 and 50 wt% of the different TRIS-buffer/polyol environments. Following, the extinction (ε_λ) of the different pNA concentrations was determined by calculating the slope, resulting from the linear correlation between the pNA concentration and the extinction values at 30 °C. The extinction-coefficients for the different environments were determined to be as follows:

Table 3: Extinction coefficients (ε_λ) of different TRIS-buffer/polyol environments used to determine the specific activity of the subtilisin protease at 30 °C.

Solvent	Concentration [weight%]	ε (n= 3)
TRIS	-	8931 ± 34
Glycerol	10	9381 ± 19
	30	10336 ± 16
	50	11318 ± 20
Sorbitol	10	9475 ± 39
	30	10617 ± 19
	50	11640 ± 66
1,2-Propanediol (MPG)	10	9545 ± 32
	30	10579 ± 45
	50	11300 ± 52

Data Evaluation and Kinetic Parameter Determination

The experimental data was plotted as a graph of the substrate concentration versus the reaction rate. Following, the datapoints were fitted with the Michalis-Menten equation using the "LMFIT: Non-Linear Least-Squares" package in python, which utilizes the Levenberg-Marquardt algorithm. The resulting curve is typically hyperbolic, with the reaction rate increasing with substrate concentration until it reaches a maximum value (V_{max}). The substrate at which the reaction rate is half of V_{max} is known as Michalis constant (K_m).

2.2.8 SDS-PAGE to Determine Influence of Autolysis

Background

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) is a technique applied to separate proteins based on their size and charge. The principle behind SDS-PAGE is to denature and linearize the protein molecules and then utilize an electric field to move them through a polyacrylamide gel matrix. In this regard, SDS is added to the samples to coat the protein with negative charges and to denature them into linear polypeptide chains. The applied electric field then causes the negatively charged protein molecule to migrate towards the positively charged electrode. While moving through the gel matrix the protein fragments encounter resistance from the cross-linked acrylamide molecules, which slow their migration and separate them based on size and charge. Smaller protein fragments encounter less resistance than larger and therefore begin to separate. The separation results in distinct protein bands on the gel, which can be visualized by staining [169, 170].

Methodology

To determine the influence of autolysis on the long-term stability of the subtilisin protease, SDS-PAGE experiments were conducted. For this, the intensity of the subtilisin protease band at ≈27 kDa was set as a criterium. The intensity of the full-length band is expected to decay upon the proteolytic degradation of the protease by itself.

Sample preparation

Samples were prepared by adding 50 µL (0-50 g/L active enzyme) of aliquots, collected throughout the long-term storage tests, to 19.95 mL of deionized water. Afterwards, 100

µL of the diluted sample were mixed with 100 µL of 0.05 M sulfuric acid (H_2SO_4). The acidification of the alkaline subtilisin protease has shown to be very effective for greatly inhibiting protease activity throughout sample preparation. In the last step, the 200 µL of the sample were mixed with 200 µL of 2x SDS sample buffer. All steps were performed gravimetrically to increase the reproducibility and accuracy of the results by applying corrections based on the gravimetrically determined dilution factor.

SDS-PAGE

SDS-PAGE was performed by adding 5 µL of standard and 10 µL of sample to an SDS-Gel (NuPAGE™ 4-12 % Bis-Tris Gel). Therefore, 1x MES SDS running buffer was prepared by mixing 40 mL 20x buffer concentrate with 760 mL of deionized water, which was added to an XCell Sure Lock™ gel chamber. After the addition of the gel and standard respectively samples, SDS-PAGE analysis was started at 200 V. After ≈45 min the gel separation was stopped, and the gel was transferred into the Coomassie color solution for 15 min. Thereafter, the gel was decolorized with a distaining solution until bands were visible and the gel turned lucid. An image (*fig. 3*) of the gel was acquired with the BioRad ChemiDoc™ MP Imaging System, and the relative band intensity was determined with BioRad ImageLab 6.1 software. The obtained intensity values were corrected based on the gravimetric dilution factor and determined relative to the intensity of the sample after 0 days of the respective storage test.

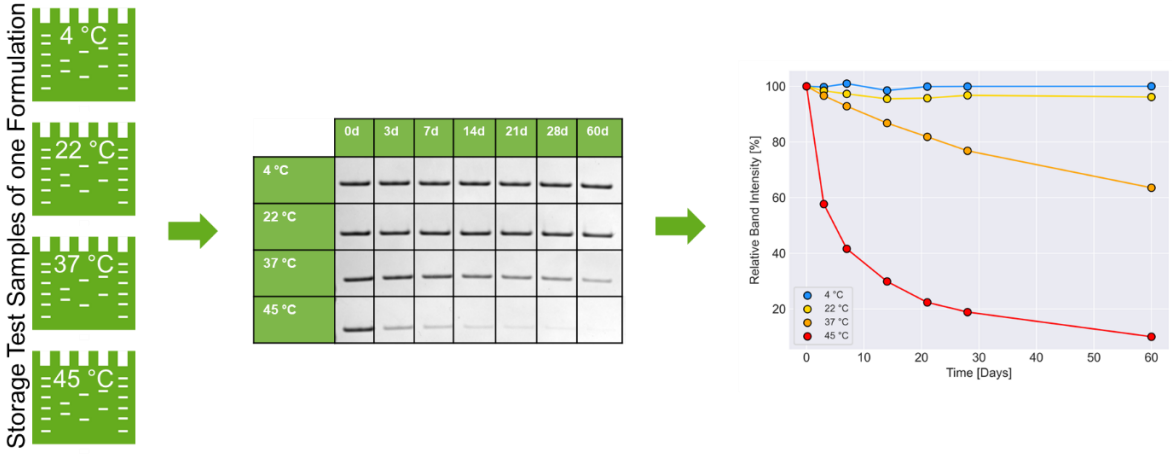


Figure 3: SDS-PAGE approach to determine the effect of autolysis on the storage test results of a subtilisin protease in different polyol formulations.

2.2.9 MIRA Infrared (IR) Spectroscopy

Background

Infrared spectroscopy is an analytical technique to identify and analyze the chemical composition of a sample by measuring its interaction with infrared light. Infrared radiation is an electromagnetic radiation located above the visible light spectrum and below microwave radiation in the range of 800 nm to 1 mm. From a spectroscopic point of view an additional distinction is made between the near infrared (800-2,500 nm), the mid infrared (2,500-25,000 nm) and far infrared (25,000-1,000,000 nm). Radiation in this range can be absorbed by molecules and converted into vibrations. In the mid- and far infrared spectral regions, this is the case when the frequencies of light and vibration match and when the molecular dipole moment changes during vibration. The position of an infrared absorption band is defined by the vibrating masses and the type of bond, which is influenced by electron withdrawing or donating effects of the intra- and intermolecular environment and by coupling with other vibrations [171].

Modern infrared spectrometers are usually Fourier transform infrared (FTIR) spectrometers. The core of such, is composed of an interferometer, which is composed of a fixed and a moveable mirror. Further, most interferometers employ a beam splitter, which takes the incoming infrared beam from a source and divides it into two optical beams. One beam is reflected from the mirror fixed in place. The other is reflected from the moving second mirror. This mirror constantly moves back and forth and depending on its position the second beam travels a longer or shorter distance. When the two beams recombine, at the beam splitter, they exhibit an interference pattern reflecting the constructive and destructive interference of the recombination. The interference pattern is guided onto the sample, and the transmitted proportion of the resulting interferogram is sent to the detector. The interferogram is mathematically transformed using a Fourier transform algorithm to obtain the infrared spectrum of a sample. The spectrum is a plot of the intensity of the infrared radiation absorbed or transmitted by the sample as a function of wavelength. It has become common practice to indicate the wavenumber (cm^{-1}) rather than the wavelength on the abscissa, which is the reciprocal value of the wavelength [172].

Methodology

The MIRA mid infrared spectrometer was developed for the analysis of samples in an aqueous environment. In this context, it combines advanced approaches of an ultra-thin

flow cell with 7-10 µm and AI assisted processing of the measured spectra. This is to reduce the influence of water on the measurement and thus ensure a high accuracy and reproducibility of the results. Measurements in aqueous solution can be performed in the range of 930-3050 cm^{-1}.

Analyzed samples were prepared according to *chapter 2.2.1*. Before transferring samples (1 mL) into HPCL vails, samples were sterile filtered using syringe filters with a pore size of 0.22 µm. Subsequently, sample veils were sealed and transferred into the MIRA device. If not directly measured, samples were stored at 4 °C until further use.

Spectra were processed using the python programming language. In this context, a variety of packages were used for the multivariate data analysis:

Table 4: Python packages used for multivariate data analysis.

Analysis	**Package**
Preprocessing:	
• Savitzky Golay Filter	Scipy / savgol_filter
• Autoscale and Mean Center	Sklearn / StandardScaler
• External Parameter Orthogonalization	"Scripted"
Multivariate analysis:	
• Dendrogram/Cluster Analysis	Scipy / Dendrogram + Linkage
• Principle Component Analysis (PCA)	Sklearn / PCA
• Principle Component Regression (PCR)	Sklearn / PCA
• Partial Least Squares (PLS)	Sklearn / PLSRegression

2.2.10 Data Analysis and Visualization

Data organization and analysis was conducted in Python using the *pandas* and *numpy* package for data analysis and manipulation. The shown plots were generated with the *matplotlib* package for data visualization.

Results & Discussion

3.1 Subtilisin Protease Characterization

In a first step, the used subtilisin protease was to be characterized according to its general stability and robustness in a non-stabilized environment. Therefore, the enzyme was stored in water at different temperatures to determine the long-term stability over period of 60 days. Subsequently, the stability of the protease in water at different pH levels should be investigated to determine an optimal pH for maximum stability and reproducibility.

3.1.1 Long-Term Stability of a Subtilisin Protease in Water

At first, the general stability of the used subtilisin protease was to be determined in water at four different temperatures. In this context, the enzyme concentration was adjusted to ≈50 g/L using deionized water. Besides, 0.3 percent by mass (wt%) calcium chloride was added to improve solubility as well as the general stability of the enzyme. Subsequently, samples were stored at 4, 22, 37 and 45 °C for 60 days, whereby samples were withdrawn after 0, 3, 7, 14, 21, 28, and 60 days. The residual active enzyme concentration was determined with the AAPF assay on the Gallery™ system and the results were normalized [%] with respect to the starting concentration (0 days).

The obtained results show a strong correlation between elevated temperatures and the decrease of protease stability. At the lowest temperature, of 4 °C (*fig. 4; blue*), an activity loss of ≈12 % could be determined after 60 days. The loss increases to ≈55 % at 22 °C and 88 % at 37 °C (*fig. 4; yellow & orange)*. The most significant loss, however, could be determined for the samples stored at 45 °C, which show a total loss of activity already after 21 d *(fig. 4; red)*.

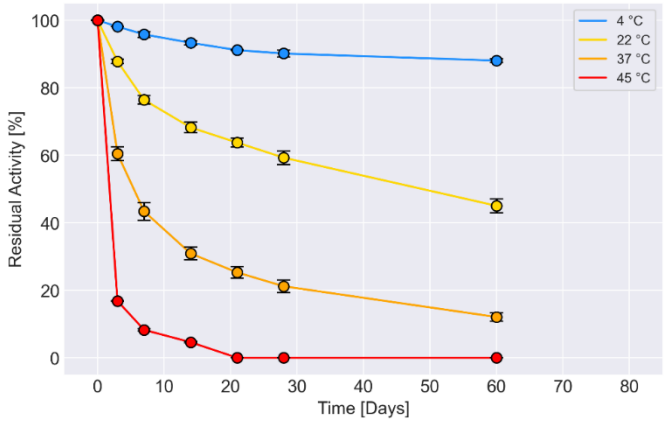


Figure 4: Residual enzyme activity [%] of the subtilisin protease in deionized water at 4, 22, 37 and 45 °C (n=2) over 60 days. *Samples were withdrawn after 0, 3, 7, 14, 21, 28, and 60 days. The determined active enzyme concentration is illustrated relative to the starting concentration of ≈50 g/L.*

The obtained results provide a general stability profile of the used subtilisin protease in an unstabilized state, in water. Further, a strong negative correlation between temperature and the subtilisin protease stability could be observed. While the enzyme was generally found to be stable at 4 °C, the residual enzyme activity was already largely lost after 3 days at 45 °C.

3.1.2. Long-Term Stability of a Subtilisin Protease at Different pH Values

In order to specify a pH value for following experiments, the stability of the subtilisin protease was determined at different pH levels. This should ensure maximum stability of the enzyme as well as reproducibility and comparability of the results obtained. Therefore, the enzyme concentration was adjusted to ≈50 g/L with deionized water and 0.3 wt% $CaCl_2$. Different samples with pH values of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 were generated based on the enzyme solution, using diluted hydrochloric acid (HCl) respectively sodium hydroxide (NaOH). Subsequently, samples were stored at 4 °C, 22 °C, and 37 °C for 60 days, whereby samples were withdrawn after 0, 14, 28 and 60 days. The pH-level and residual enzyme activity of each sample was determined using a pH-meter and the AAPF-assay on the Gallery™ system.

The results demonstrate a high stability of the used subtilisin protease (*fig. 5A*), especially in the range of pH 4-11, across the different temperatures. In general, as already described

in *chapter 3.1.1*, protease stability significantly decreased at elevated temperatures. Consequently, all samples, independent of their pH level, had lower residual enzyme activity in presence of higher temperatures. Besides pH levels below 4 and above 11 showed to have a detrimental effect on the enzyme activity. At pH 2 and 12 almost no residual enzyme activity could be determined after 60 days of storage. Similarly, the protease stability appeared to be significantly reduced at pH 3. In the range of pH 4-11 the enzyme remained stable, however, at different levels. The highest content of active enzyme after storage, could be determined in the samples stored at pH 4 and 5.

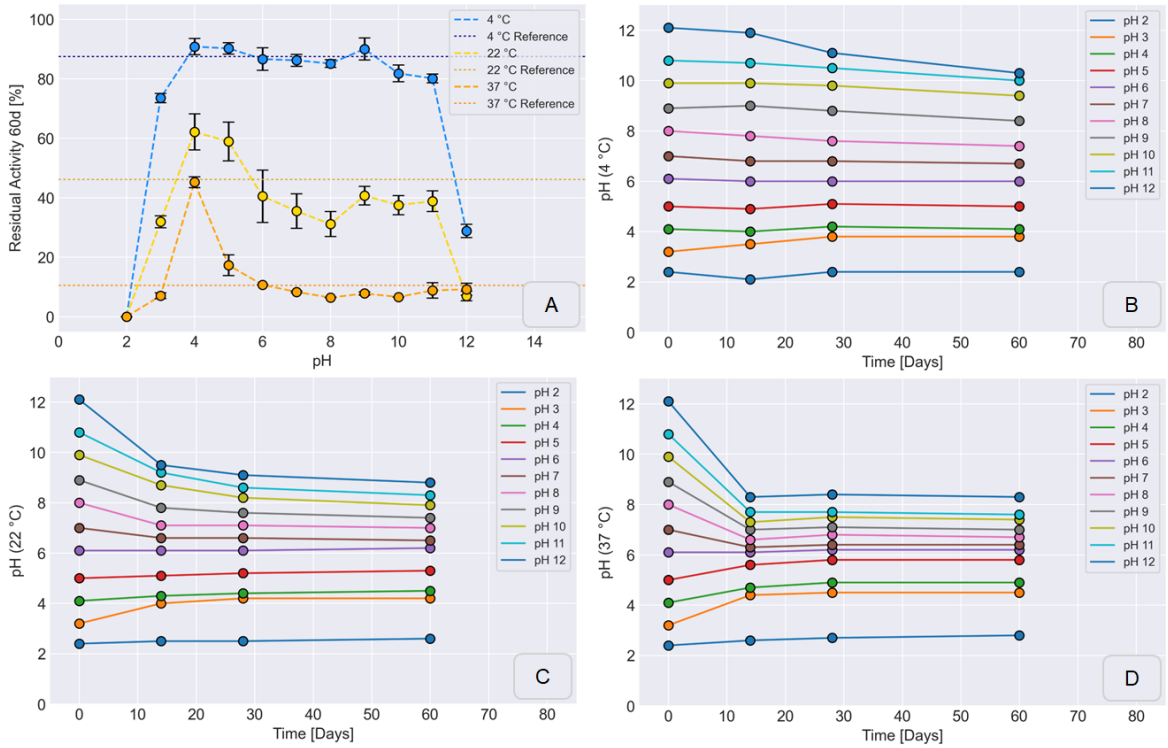


***Figure 5**: **Residual subtilisin protease activity [%] and formulation pH after storage.** (A) Active enzyme content after 60 days at pH 2-12 at 4, 22, and 37 °C. (B-D) pH drift of the formulations over 60 days at 4, 22 and 37 °C.*

In addition, monitoring the pH levels at different temperatures (*fig. 5B-D)*, showed that pH levels fluctuate during storage depending on the underlying temperature. Especially, pH maxima and minima exhibited significant changes of the pH levels, within the first 14-28 days of storage. The pH plots at different temperatures further illustrate a constant drift of the pH-values towards pH 6. Consequently, formulations with pH levels with/close to pH 6, remained constant throughout the 60 days of storage.

The experiment conducted shows, that in addition to temperature, pH values can significantly influence the stability of an enzyme in an unstabilized environment. Overall, the subtilisin protease appears to be stable in a broad pH spectrum from 4-11. Furthermore, the pH level has shown to vary throughout storage, with a tendency towards pH 6. Thereby, it can be assumed that the pH value is changed by dynamic processes in the aqueous environment.

3.1.3 Discussion

The stability of an enzyme is defined by its ability to retain its active structural conformation despite of disruptive forces, as temperature and pH [173]. High thermal stability is considered an essential criterion for many industrial relevant enzymes [174, 175]. Besides, they have to be stable in a pH range, relevant for their intendent use [176]. In this regard the used subtilisin protease is stable in a broad range of different temperatures and pH levels.

In general, elevated temperatures are known to be very effective at denaturating proteins, as they can alter the protein structure. These changes are mediated by an increase of the available kinetic energy, which is reflected by an increased vibration and fluctuation of the protein structure. In the end, extensive vibrations can lead to a disruption of H-bonds, hydrophobic interactions, and van der Waals forces, which are essential to maintain the secondary and tertiary structure motifs [35, 177]. Subsequently, the enzyme can be inactivated or irreversibly denaturated.

Similarly, the pH level of the enzyme's environment can significantly influence its stability. Enzymes are composed of a great number of charged amino acid residues, on and inside their three-dimensional structure. In some cases, the interactions mediated by these charged groups can be essential for the enzyme to maintain its structure. Varying pH levels can affect the charge of these residues, disrupting stabilizing interactions and thereby altering the protein structure [178-180].

In addition, temperature and pH can have a significant influence on enzyme activity [67, 181, 182]. With regard to subtilisin proteases this has a significant influence on the long-term stability of the enzyme. As proteases are capable of hydrolyzing peptide bonds, they do not differentiate between their own and foreign proteins. Subsequently, the rate of autolysis is directly affected by temperature and pH.

Similar to the turnover of a classical substrate, the process of autolysis is strongly temperature dependent. As for chemical reactions, enzyme activity constantly increases with temperature. According to the van't Hoff rule or the Arrhenius equation there is basically no limit to this increase [183, 184]. However, above a certain temperature, the reaction rate approaches zero, as the effect is counteracted by the irreversible thermal denaturation of the enzyme. The optimal temperature for enzymes is subsequently not the one with the highest activity, rather than the one, which retains the optimal ratio between activity and stability.

Besides, higher temperatures increase the number of collisions between molecules in solution [185]. This is due to an increase in velocity and kinetic energy that is mediated by elevated temperatures. Thereby, the contact time between subtilisin proteases and their substrate is further decreased. Consequently, elevated temperatures not only result in more enzymes reaching the activation energy to hydrolyze peptide bonds, but also the number of contacts with substrate or other proteases in solution is increased.

The enzyme activity is further dependent on the pH of the environment. Especially, if charged residues are involved in the catalytic process. The catalytic rate of serine proteases and consecutively subtilisin proteases are therefore significantly influenced by the protonation state of the residues in the catalytic triad. This comprises, the hydroxy group of serine, the imidazole group of histidine, and the carboxy group of aspartate, which are essential for the transfer of a proton from the serine hydroxy group to the imidazole nitrogen [31, 35]. In the past, it was shown that especially alkaline proteases as the subtilisin protease used, can be efficiently inhibited at lower pH [176, 186]. Consequently, pH levels above the detrimental level of 2 and above the transition region of 3 seem to minimize the proteolytic activity and therefore result in an increased protease stability at pH values of 4 to 5 after 60 days.

The proteolytic activity appears to further affect the pH level of the different samples. The effect is assumed to be caused by the continuous release of peptide fragments and protons into the environment mediated by the hydrolysis reaction in the catalytic triad [35]. The catalytic activity of the subtilisin protease appears to significantly alter the pH throughout storage, especially at alkaline pH levels, where a high activity of the subtilisin protease is expected. The observed trend of most samples towards pH 6 is assumed to be mediated by the increased availability of N/C-terminal amino acid residues and side chains.

To minimize the effects of pH fluctuations and to maintain a constant pH throughout storage, the pH was set at 6.0 for the following experiments. This should maintain and improve reproducibility and comparability between different formulation environments.

3.2. Subtilisin Protease Stabilization with Polyols

In addition to the general characterization, the effect of different polyols on the subtilisin protease stability was determined. In this context, differential scanning calorimetry (DSC) was used to determine the thermal stability of the protease in presence of glycerol, sorbitol and 1,2-propandiol (MPG). Subsequently, the stability of the subtilisin protease was to be evaluated on an atomistic scale, in different alcohol/polyol environments, by using molecular dynamic (MD) simulations. In this context, a correlation with DSC results was to be assessed. In addition, the long-term stability of the enzyme was determined in different subtilisin protease/polyol formulations by conducting long-term storage tests. Following, results from DSC and activity measurements were to be compared, in order to evaluate the predictive power of DSC measurements in the context of long-term stability testing.

3.2.1 Thermal Stability of a Subtilisin Protease in Presence of Polyols determined by DSC

To determine the thermal stability of the used subtilisin protease in different polyol formulations, DSC measurements were conducted. In this way, the influence of glycerol, sorbitol and MPG on the thermal stability should be determined. DSC allows the simultaneous acquisition of the unfolding temperature (T_m) and unfolding enthalpy (ΔH), in a single experiment. These variables can be used to characterize an analyte in different molecular environments. With regard to enzyme formulation development, DSC measurements have shown to significantly accelerate the development process [187, 188]. Subsequently, DSC has established itself as a powerful and versatile tool for determining optimal conditions for stabilizing proteins in liquid and solid formulations.

The thermal stability of the subtilisin protease was determined in the presence of three different solvents and five different solvent concentrations. Therefore, formulations with an enzyme concentration of ≈50 g/L and an inherent solvent concentration of 10-50 wt% (n=5; interval= 10 wt%) glycerol, sorbitol or MPG were prepared. Further, 0.3 wt% $CaCl_2$ were added to each formulation to increase the general stability and solubility of the enzyme. The rest of the formulation was filled with deionized water. When not directly analyzed samples were stored at -80 °C until further use.

The thermograms of the DSC measurements illustrate a positive effect of glycerol and sorbitol on the thermal stability of the enzyme (*fig. 5A+B*). Already, in the presence of 10 wt% glycerol respectively sorbitol, thermograms are shifted towards higher temperatures, when compared to the enzyme in water. This trend increases with the amount of glycerol and sorbitol. For MPG, however, a negative trend between increasing concentrations and thermal stability can be observed (*fig. 6C*). Although, formulations with 10 wt% MPG exhibit a similar thermal stability than the enzyme in water, increasing MPG concentrations shifted the thermogram towards lower temperatures.

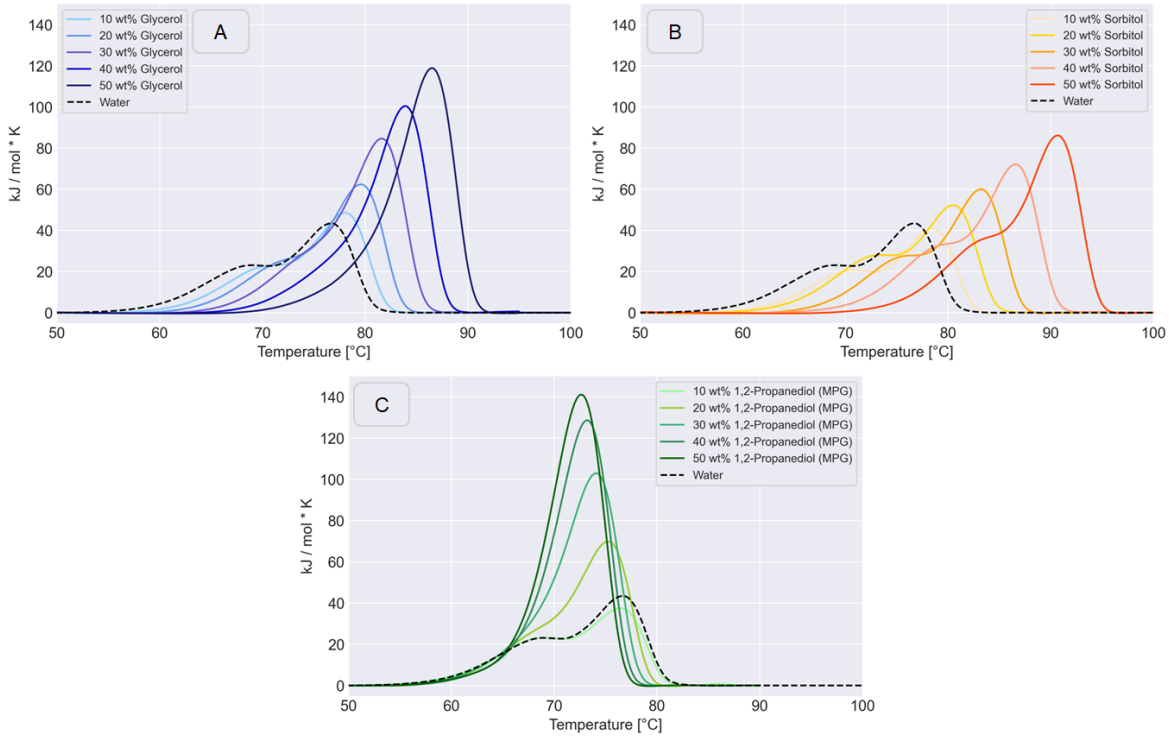


Figure 6: DSC results for subtilisin protease/polyol formulations with varying polyol concentrations (10-50 wt%). *(A) Glycerol, (B) Sorbitol and (C) 1,2-Propanediol (MPG).*

Besides, samples with low solvent concentrations, show to exhibit two distinct peaks. This is particularly evident for sorbitol formulations, where this observation is visible up to 50 wt% polyol content. The observation is less frequent for glycerol formulations and even less in the case of MPG.

Since the subtilisin protease/polyol formulations have two distinct peaks, two T_m values per formulation could be determined (*tab. 5*). The T_m values for glycerol and sorbitol show to increase with the solvent concentration (*fig. 7 blue & orange*). For MPG, increasing solvent concentrations show to exhibit decreasing T_m values (*fig. 7; green*). The T_m values

in presence of 10 and 20 wt% glycerol and sorbitol are similar for both peaks. However, at higher polyol concentrations a clear separation between both solvents is visible. Subsequently, samples with a similar concentration achieve higher T_m values in the presence of sorbitol, when compared to glycerol. In contrast, T_m values of formulations with MPG are considerably lower. The T_m values for the first peak remain constant (*fig. 7A*), while the values for the second (*fig. 7B*) appear to decrease with increasing MPG concentration.

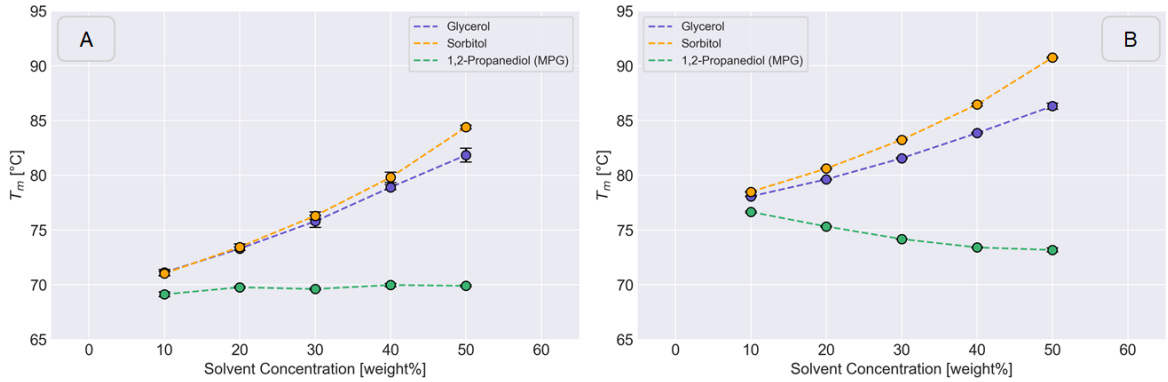


***Figure 7: T_m* values of two distinct peaks observed for different subtilisin protease/polyol formulations.** Shown are the observed T_m values of (A) Peak 1 and (B) Peak 2 with respect to the solvent concentration of the different formulations.

***Table 5: T_m* values of two distinct peaks observed for different subtilisin protease/polyol formulations.**

A / Peak1

Solvent	Concentration [weight%]	Mean T_m (n=2) [°C]	SD (n=2) [°C]
Glycerol	10	71.13	0.26
	20	73.30	0.15
	30	75.80	0.55
	40	78.91	0.24
	50	81.85	0.64
Sorbitol	10	71.03	0.21
	20	73.45	0.29
	30	76.30	0.35
	40	79.81	0.47
	50	84.39	0.17
1,2-Propanediol (MPG)	10	69.13	0.22
	20	69.78	0.10
	30	69.61	0.09
	40	69.98	0.16
	50	69.90	0.08

B / Peak2

Solvent	Concentration [weight%]	Mean T_m (n=2) [°C]	SD (n=2) [°C]
Glycerol	10	78.07	0.03
	20	79.63	0.05
	30	81.56	0.07
	40	83.86	0.05
	50	86.32	0.28
Sorbitol	10	78.49	0.05
	20	80.62	0.06
	30	83.24	0.04
	40	86.47	0.14
	50	90.75	0.05
1,2-Propanediol (MPG)	10	76.66	0.09
	20	75.33	0.05
	30	74.18	0.03
	40	73.41	0.05
	50	72.18	0.18

The unfolding enthalpy (ΔH) for each measurement is obtained by determining the area under each thermogram. Due to the bi-peaked shape, two gaussian peaks were fitted to the thermogram and evaluated separately. Both peaks show an increase of the enthalpy of deconvolution, with respect to the solvent and increasing solvent concentrations. However,

the unfolding enthalpy for the first peak appears to be comparable (*fig. 8A*) in between similar solvent concentrations. Further, only moderate changes, when compared to the subtilisin protease in water, can be observed.

In contrast, significant differences, between the different polyols and concentrations are visible for the second peak (*fig. 8B*). In addition, formulations with solvent concentrations above 10 wt% polyol exhibit a significantly higher unfolding enthalpy than the enzyme in water. In the presence of 10 wt% solvent, the highest unfolding enthalpy can be observed for glycerol, followed by sorbitol and MPG. When further increasing the polyol concentration, however, the order is shifted and remains the same for following concentrations. In this regard, MPG exhibits the highest unfolding enthalpy, followed by glycerol and sorbitol.

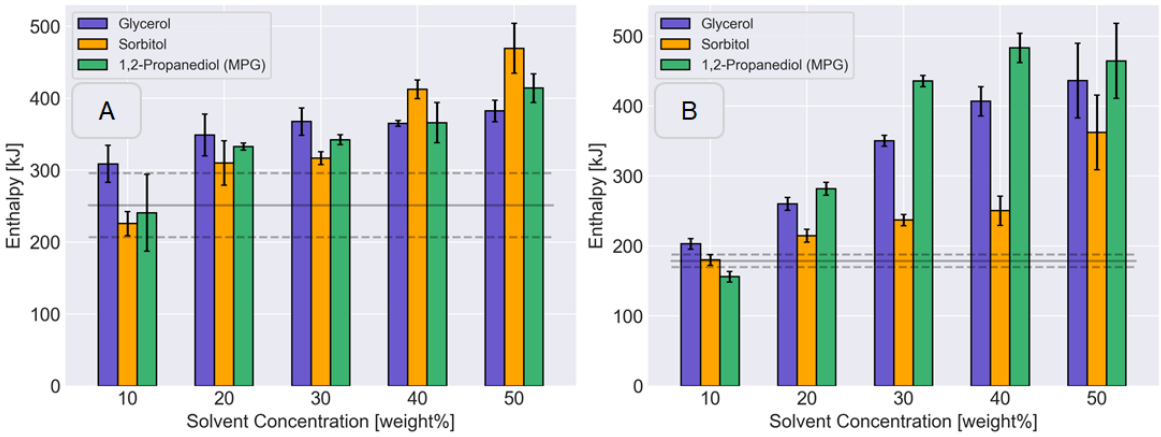


Figure 8: Unfolding enthalpies determined for different subtilisin protease/polyol formulations. *Illustrated is the unfolding enthalpy for two distinct peaks: (A) Peak 1, (B) Peak 2. The average enthalpy and standard deviation of the enzyme in water is depicted by the solid grey line respectively dashed lines.*

Glycerol and sorbitol, both increase the thermal stability of the subtilisin protease, when compared to the unstabilized sample in water. In contrast, the enzyme exhibits a decreased thermal stability in the presence of MPG. The differences were illustrated by the obtained thermograms as well as based on the determined T_m values. However, the order of the different solvents according to their unfolding enthalpy is different. Especially, MPG appears to exhibit considerably higher values, when compared to glycerol and sorbitol. The unfolding enthalpy is, however, strongly dependent on the enzyme concentration in the sample [161]. Subsequently, changes in the enzyme concentration during the measurement have to be considered. The autolytic activity of the protease is expected to increase with

temperature, until the enzyme is denaturated [189, 190]. Consequently, this effect is prolonged in solvents in which the enzyme exhibits a higher thermal stability. Following, the active enzyme concentration in sorbitol is expected to be lower upon unfolding, when compared to glycerol or MPG. Subsequently, in the case of the subtilisin protease, the unfolding enthalpy, is negligible with regard to the evaluation of thermal stability. Consequently, based on the T_m values, the used polyols can be classified in terms of their ability to increase the thermal stability of the subtilisin protease as follows:

Sorbitol > Glycerol > Water > MPG

Besides, the appearance of two distinct peaks in the thermogram was the subject of an in-depth analysis in *chapter 3.3*.

3.2.2 Thermal Stability of a Subtilisin Protease in Presence of Polyols determined by MD Simulations

The results from the DSC measurements demonstrate the effect of different polyols on the stability of the subtilisin protease. To gain additional insights on the effect of the different polyols on the enzyme structure, MD simulations were conducted. Thereby, polyol induced changes of the enzyme structure, were to be studied on an atomistic scale. In addition, a possible correlation between DSC and MD results was to be assessed. In this context, molecular descriptors, and experimental T_m values/predicted logP values, of different subtilisin protease/alcohols/polyols formulations were to be determined. The studied molecular descriptors comprised the solvent accessible surface area (SASA), the root mean square deviation (RMSD), the radius of gyration (RG), and the Q-value. Molecular descriptors were calculated based on the returned simulation trajectories, using the tcl-scripting language in VMD. The T_m values were obtained by DSC-analysis and logP values were predicted based on the molecular structure of the solvents, by molinspiration online prediction tool (https://www.molinspiration.com) [191].

The molecular descriptors were determined, by conducting MD simulations in atomistic environments resembling the laboratory subtilisin protease/polyol formulations. For this, systems were generated, which resembled laboratory formulations with approximately 50 wt% alcohol/polyol and 0.3 wt% $CaCl_2$. Following, MD simulations were conducted over a period of 200 ns at a constant temperature of 373.15 K (100 °C). Besides, the polyols used in *chapter 3.2.1*, further alcohols/polyols, as xylitol, erythritol, ethylene

glycol, and methanol were included in the analysis. Subsequently, six repeats for each subtilisin protease/alcohol/polyol system were simulated. Following, the molecular descriptors were calculated based on the simulation trajectories. Besides, the mean fluctuation of each amino acid residue was determined, by calculating the root mean square fluctuation (RMSF).

For the better visualization of the results, the values were averaged over the six independent repeats, obtained for each solvent. The correlation analysis was conducted by determining the average value of each descriptor over the last 10 ns of the trajectory, which was plotted against the T_m value of the formulation respectively the estimated logP-value of the solvent. Erythritol and methanol were excluded from the DSC analysis due to solubility and denaturation issues under laboratory conditions, at a concentration of 50 wt%.

Calcium-Bound Structure

MD simulations were conducted based on the native structure of the subtilisin protease. In its active state this is expected to comprise two occupied calcium binding sites [192]. Thereby, the stability of the active state of the subtilisin protease in different alcohol/polyol formulations was to be monitored. In order to prevent one or both calcium ions from leaving the protein structure, artificial bonds were applied to retain both calcium ions in their binding pockets throughout the simulations.

The average values of the molecular descriptors illustrate a dense picture for the different subtilisin protease formulations (*fig. 9*). Especially, the results for the RG (*fig. 9C*) exhibit only a minor separation in between the different formulations. This state appears to improve for the SASA and the Q-value (*fig. 9A+D*), however, classification of the solvents is not possible. The RMSD (*fig. 9B*) returns the best separation for the underlying dataset, but still, for the majority of formulations a visual separation is not possible. In general, the results shown, do not allow a clear classification of the different solvents according to their ability to stabilize the subtilisin protease. Even though, some solvents show to be isolated exceptions, as erythritol and sorbitol, the obtained results illustrate an arbitrary behavior in the corresponding correlation analysis (*fig. 10*).

For the correlation analysis, the average value of the last 10 ns of each system and its six repeats is plotted against the experimental values obtained from DSC measurements and

logP predictions. The plots lack any evident correlation between the results of the molecular descriptors and the experimental/predicted T_m and logP values. Especially, the correlation of the logP value with the different descriptors shows to fluctuate significantly in between the different solvents, independent from their hydrophobicity/hydrophilicity. Subsequently, some solvents that have shown to increase the thermal stability of the enzyme, exhibit similar values for molecular descriptors as solvents, which have shown to be destabilizing. In this context, glycerol has shown to considerably increase the thermal stability of the subtilisin protease during DSC measurements, in contrast to MPG. However, based on some molecular descriptors glycerol appears to perform similar, or even worse, than MPG. Fare more, results show to significantly fluctuate in between the different subtilisin protease/alcohol/polyol systems.

Therefore, based on the obtained results, no correlation between the experimental T_m values respectively predicted logP-values of different alcohol/polyol formulations/solvents and molecular descriptors from MD simulations can be determined. Far more, it can be assumed that the results are rather arbitrarily distributed for the different alcohol/polyol formulations. The missing differentiation between the molecular descriptors, is expected to be the result of normal structural fluctuations, rather than structural changes associated with thermal unfolding. Consequently, observed differences between the formulations, may be regarded as “noise” rather than representing actual structural changes caused by the environment. Subsequently, it can be assessed that the general stability of the subtilisin protease in its native state is too high, to monitor significant structural changes in the time frame of 200 ns.

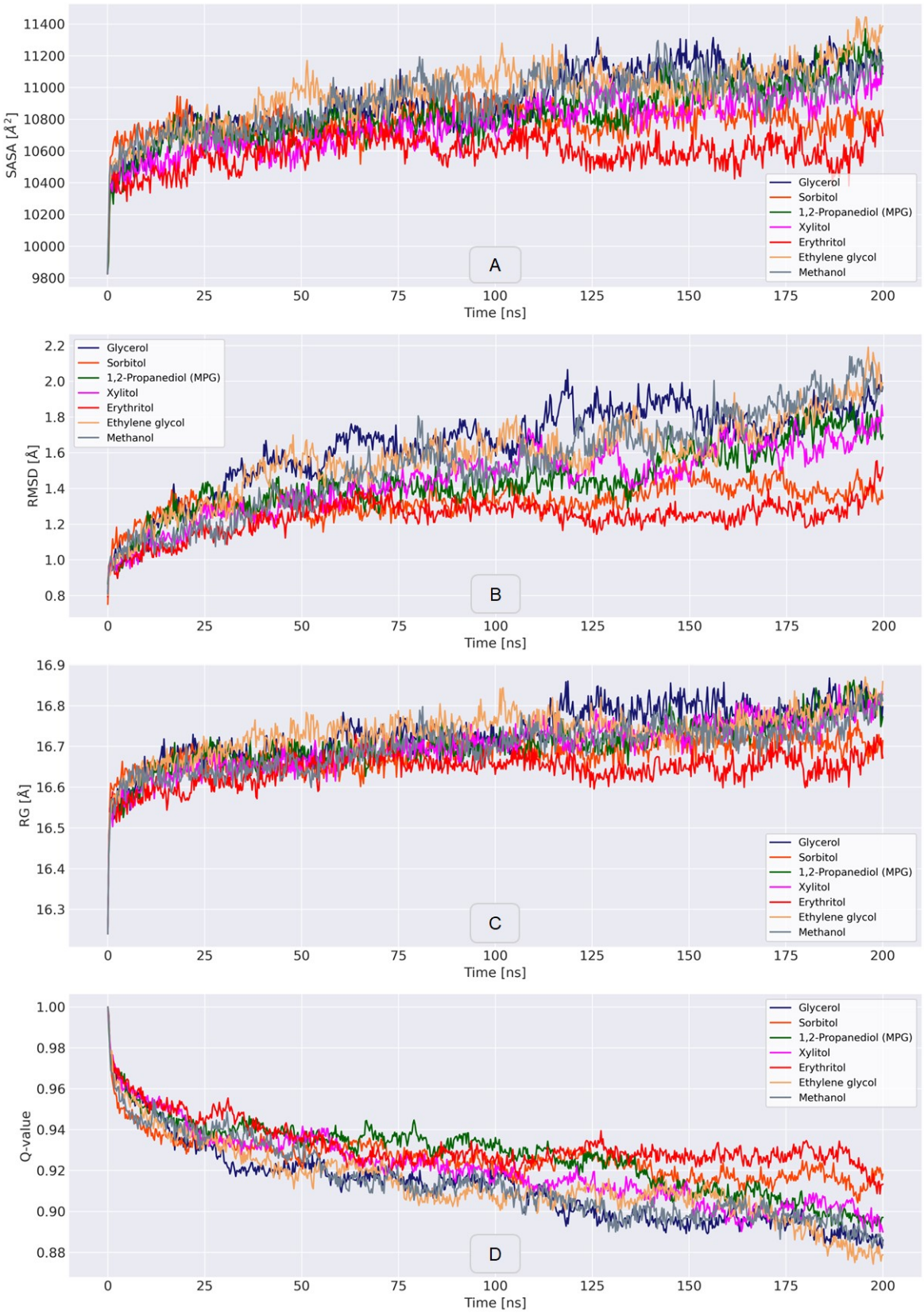


Figure 9: Molecular descriptors determined for the different subtilisin protease/alcohol/polyol systems with Ca^{2+}. *(A) SASA, (B) RMSD, (C) RG, and (D) Q-value. The illustrated values were averaged over six independent MD simulation repeats for each subtilisin protease/alcohol/polyol system.*

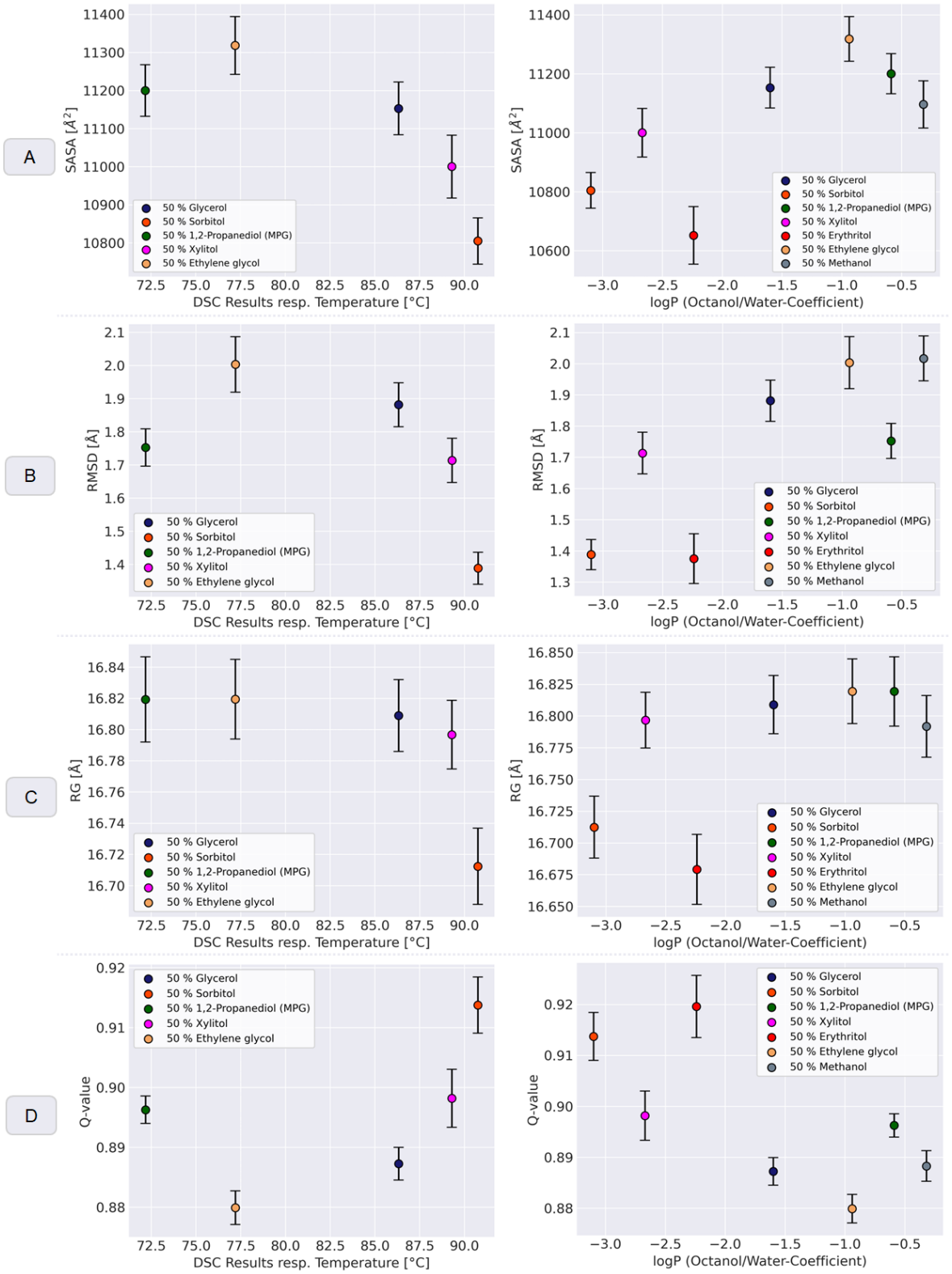


Figure 10: Correlations between T_m values/predicted logP-values and the average value of the last 10 ns of different molecular descriptors in presence of Ca^{2+}. *(A) SASA, (B) RMSD, (C) RG, and (D) Q-value.*

Calcium-Free Structure

The MD simulations with the native structure of the subtilisin protease have illustrated a high stability of the subtilisin protease in the presence of the different alcohols/polyols. Subsequently, differentiation between the subtilisin protease/alcohol/polyol formulations is not possible and the molecular descriptors lack a correlation with experimental T_m and predicted logP values. Therefore, additional MD simulations were performed, to determine the molecular descriptors in the absence of bound and free calcium in the subtilisin protease structure and formulation. By removing calcium, a decreased stability of the enzyme structure was expected. Thereby, it was hoped, to shift stability related unfolding processes in the simulated time frame of 200 ns.

For this, molecular systems were prepared, which resembled laboratory subtilisin protease formulations with 50 wt% alcohol/polyol. Besides, removing the calcium from the enzyme structure, no $CaCl_2$ was added to the molecular ensembles. This was supposed to prevent the re-binding of calcium throughout the simulation. The enzyme-solvent systems were simulated for 200 ns, whereafter the molecular descriptors (SASA, RMSD, RG, and Q-value) were determined based on the simulation trajectories. For a better illustration of the results, the average value of each descriptor was calculated based on the six independent repeats.

The obtained results show a high similarity in between the different molecular descriptors (*fig. 11*). The spread in between the different enzyme-solvent systems is, however, larger than in the presence of calcium. Similar to before, the resolution is the best for the RMSD, followed by SASA and Q-value (*fig. 11A-B & D*). RG results (*fig. 11C*) appear to be more condensed. Overall, the increased separation enables a better interpretation and classification of the solvents according to the different molecular descriptors.

Subsequently, the average value of the last 10 ns of each descriptor, was compared with experimentally determined T_m and predicted logP-values. In contrast to before, all plots illustrate a strong correlation of the MD results with the experimental and predicted values (*fig. 12*). While the correlation with the T_m values appears to be linear, correlations with the logP show to be exponential.

The observed correlation between the values shows to be the highest for the SASA (*fig. 12A*). In this regard, correlation analysis for T_m and SASA-values illustrate a high negative linear correlation between both data sets. Overall, Pearson, Spearman, and Kendall tau correlation analysis exhibit values of -0.999.

Figure 11: Molecular descriptors determined for to the different subtilisin protease/alcohol/polyol systems without Ca^{2+}. *(A) SASA, (B) RMSD, (C) RG, and (D) Q-value. The illustrated values were averaged over six independent MD simulation repeats.*

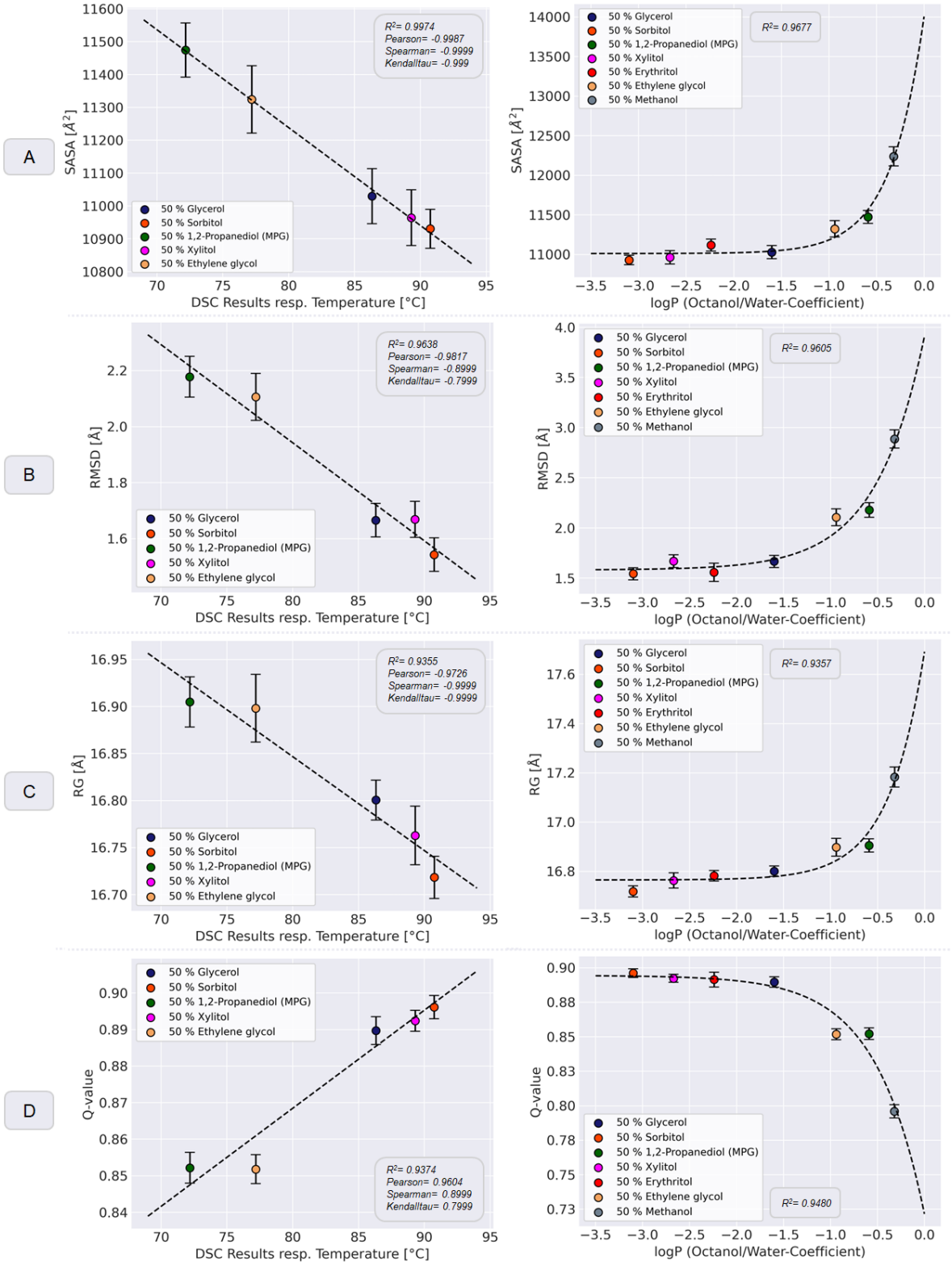


Figure 12: Correlations between T_m values/predicted logP-values and the average value of the last 10 ns of different molecular descriptors in absence of Ca^{2+}. *(A) SASA, (B) RMSD, (C) RG, and (D) Q-value.*

In addition, an R^2-value (0.997) close to 1 indicates a high similarity in between the linear regression model and the MD/experimental values. Similarly, the predicted logP-values for the solvents and the SASA-values show an exponential correlation, as was illustrated by fitting an exponential equation to the datapoints. The fitted exponential equation exhibits an R^2-value of 0.967, which illustrates the high-level conformity between the datapoints and fitted model.

Experimental and predicted values show a similar high correlation with regard to the RMSD (*fig. 12B*). T_m values and RMSD exhibit a strong negative correlation (Pearson= 0.982; Spearman= 0.900; Kendall tau= 0.800). Further, the R^2-value of 0.964 illustrates a high accordance with the linear regression model. In addition, the correlation between the logP-values and the RMSDs exhibit the already experienced exponential behavior. The fitted exponential model returns an R^2-value of 0.961.

RG and Q-value (*fig. 12C+D*) show to be similarly correlated with regard to the experimental T_m and predicted logP-values. For both descriptors, the linear regression model appears to be in good accordance with the determined datapoints. In this regard, an R^2-value of 0.936 can be determined for the RG and 0.937 for the Q-value. The calculated linear correlation coefficients for the RG (Pearson= -0.973; Spearman= -0.999; Kendall tau= -0.999) illustrate a strong negative correlation, while the Q-value exhibits a strong positive correlation (Pearson= -0.960; Spearman= -0.900; Kendall tau= -0.800). Further, both descriptors exhibit an exponential correlation with the logP-values of the solvents. In this context, an R^2-value of 0.936 can be determined for the RG and 0.948 for the Q-value.

The conducted MD simulations, in absence of calcium, in and around the protein, illustrate the influence of the alcohols/polyols on the enzyme structure on an atomistic scale. While differentiation between the subtilisin protease/alcohol/polyol formulations was not possible in the presence of calcium, the enzyme shows to be significantly less stably in the absence of calcium in the monitored 200 ns time frame. Subsequently, results from molecular descriptors appear to be in good accordance with DSC results. In general, the subtilisin protease exhibits a smaller surface area and diameter in the presence of larger/hydrophilic polyols and shows to better retain native contacts inside the protein structure. Furthermore, fluctuations of the amino acid residues show to be decreased, when compared to formulations with smaller/hydrophobic alcohols/polyols.

Root Mean Square Fluctuations (RMSF)

When considering the root mean square fluctuations (RMSF) of the single amino acid residues, differences between the calcium occupied and -free (*fig. 13*) subtilisin protease become evident. Especially amino acids from loop regions show to fluctuate significantly throughout the simulations. While fluctuations of these regions appear to be arbitrary in the presence of different alcohols/polyols for the calcium carrying protease (*fig. 13A*), fluctuations for the calcium-free phenotype (*fig. 13B*) appear to be related to the molecular weight respectively hydrophilicity/-phobicity of the solvent. In this context, the RMSF illustrates a similar picture as observed for the different molecular descriptors. The observation further adds to the hypothesis, that the observed changes in the calcium occupied subtilisin protease structure are rather natural fluctuations than the onset of thermal unfolding.

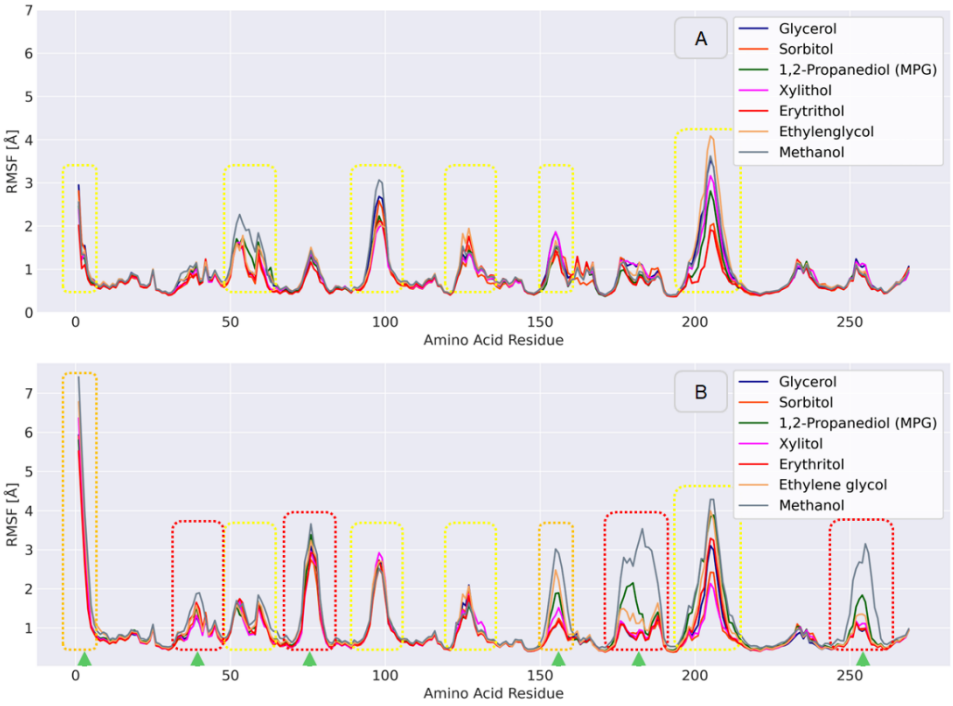


Figure 13: RMSF of subtilisin protease/polyol systems in the presence/absence of Ca^{2+} in and around the enzyme structure. *(A) Both calcium binding sites occupied, (B) calcium-free.*

Besides, several enzyme regions exhibit increased fluctuations in the absence of the two calcium ions (*fig. 14*). In this regard, three regions that have already shown increased fluctuations in the presence of calcium (*fig. 14; yellow)*) exhibit even stronger fluctuations in the absence of calcium (*fig. 14; orange*). Further, additional four regions (*fig. 14; red*) illustrate significantly increased fluctuations, in the calcium-free protease structure. These

four regions, show to be loops in close proximity of the calcium binding sites. Among others, these regions comprise the N- and C-terminal region of the subtilisin protease.

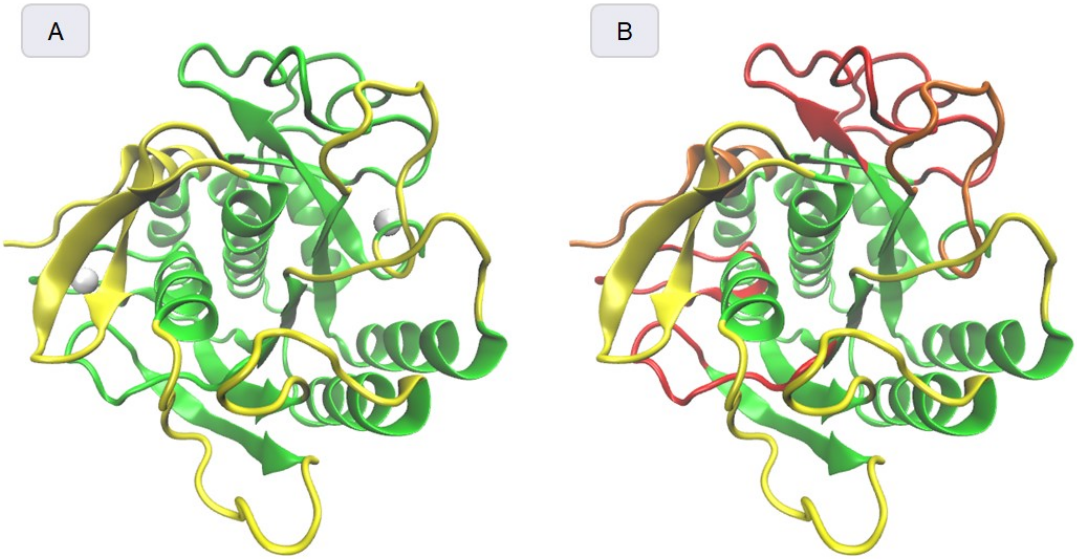


Figure 14: Regions of the subtilisin protease with increased RMSF in the presence/absence of Ca^{2+}. *(A) Ca^{2+}-bound enzyme structure and (B) Ca^{2+}-free enzyme structure. Yellow regions have shown to significantly fluctuate in the Ca^{2+}-bound structure respectively have shown similar fluctuations in the Ca^{2+}-free state. Orange regions have shown to fluctuate in the Ca^{2+}-bound structure and have shown even stronger fluctuations in the absence of Ca^{2+}. Red regions exhibited minor fluctuations in the Ca^{2+}-bound structure but illustrate significant fluctuations in the Ca^{2+} free state.*

The obtained results, illustrate the influence of the two calcium ions on the subtilisin protease structure. While several regions show to fluctuate in the presence of calcium, fluctuations show to significantly increase in the absence of both calcium ions. Subsequently, it can be assessed that the presence/absence of both calcium ions significantly influences the flexibility respectively rigidity of the subtilisin protease structure. The increased fluctuations in the proximity of the calcium binding sites, illustrate the importance of both calcium ions as interactions partners to maintain the enzyme structure and stability. The overall influence of both calcium ions on the protein stability and activity is further evaluated and discussed in *chapter 3.3*.

3.2.3. Long-Term Stability of a Subtilisin Protease in Presence of Polyols

To determine the long-term stability and thereby, the extent to which the different polyols stabilize the subtilisin protease over time, storage tests were performed over a period of 60 days. In this regard, formulations were prepared with 10, 20, 30, 40 and 50 wt% glycerol (G), sorbitol (S), and MPG (M). The enzyme concentration was set to ≈50 g/L and a consistent amount of 0.3 wt% $CaCl_2$ was added to each formulation. Following, samples were stored at four different temperatures, 4, 22, 37 and 45 °C, and samples were withdrawn after 0, 3, 7, 14, 21, 28, 60 days. After storage, the residual active enzyme concentration, was determined with the AAPF assay. Following, the residual activity of the samples was illustrated relative [%] to the starting concentration (0 d).

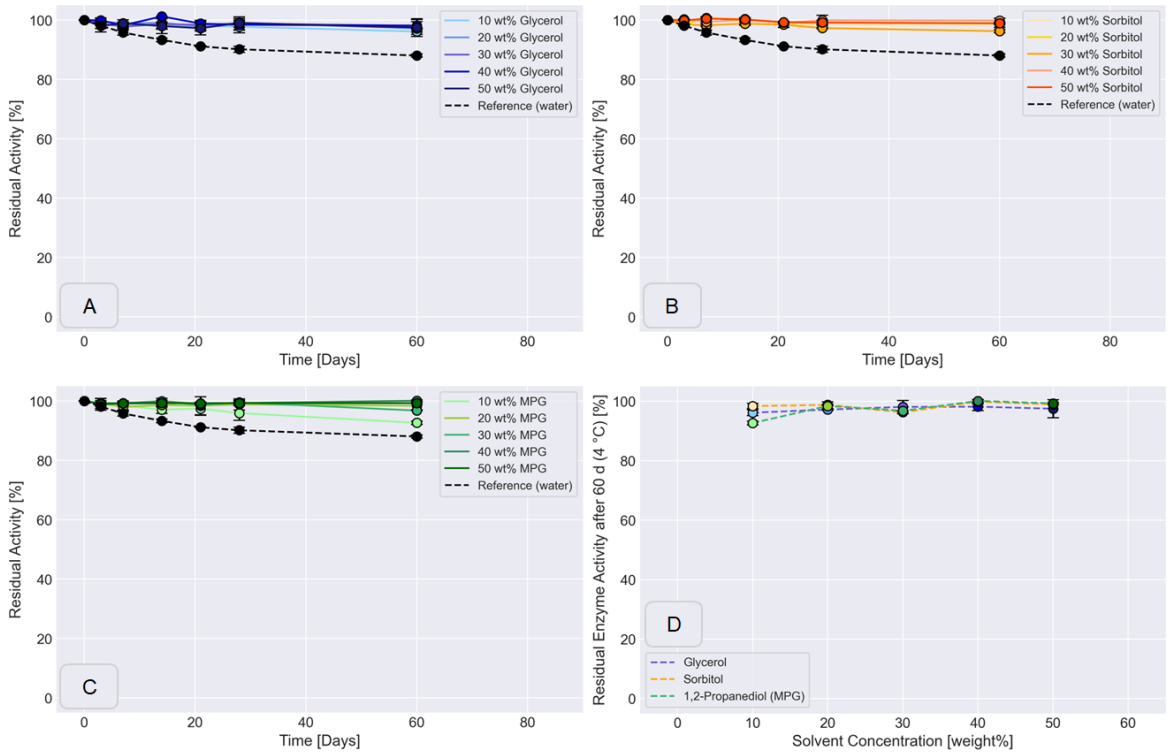


Figure 15: Residual enzyme activity [%] of the subtilisin protease in different polyol formulations (10-50 wt%) after storage at 4 °C over 60 days. *Samples were withdrawn after 0, 3, 7, 14, 21, 28 and 60 days. (A) Glycerol, (B) Sorbitol, (C) 1,2-Propanediol (MPG), and (D) Residual enzyme activity [%] after 60 days of storage.*

At 4 °C (*fig. 15*), only minor changes of the active enzyme content in between the different formulations can be determined. Compared to the reference in water, however, all formulations illustrate a higher residual enzyme activity after 60 days. Overall, differences

in between the different solvents and concentrations are only scarcely visible. Even after 60 days, the data points of similar solvent concentrations are still overlapping. Subsequently, no significant differences in between the solvents and concentrations can be detected. However, in the presence of 10 wt% solvent a separation shows to develop.

At 22 °C (*fig. 16*), the separation between the samples and similar concentrations of different solvents becomes more evident. Especially, the samples formulated with 10 wt% MPG (*fig. 16C*) show a significant decline in activity (-41 % (60 d)), when compared to samples with similar amounts of glycerol (-23 % (60 d)) and sorbitol (-24 % (60 d)) (*fig. 16A+B*). At higher solvent concentrations, however, the clear separation between solvents and concentrations fades (*fig. 16D*). In comparison, glycerol (G20/60 d (91 %), G30/60 d (97 %)) appears to maintain the active enzyme content on the highest level, while sorbitol (S20/60 d (85 %), S30/60 d (91 %)) and MPG (M20/60 d (85 %), M30/60 d (93 %)) exhibit similar results at solvent concentrations of 20 and 30 wt%. In the presence of 40 and 50 wt% solvent, a clear separation based on the residual active enzyme content is not possible.

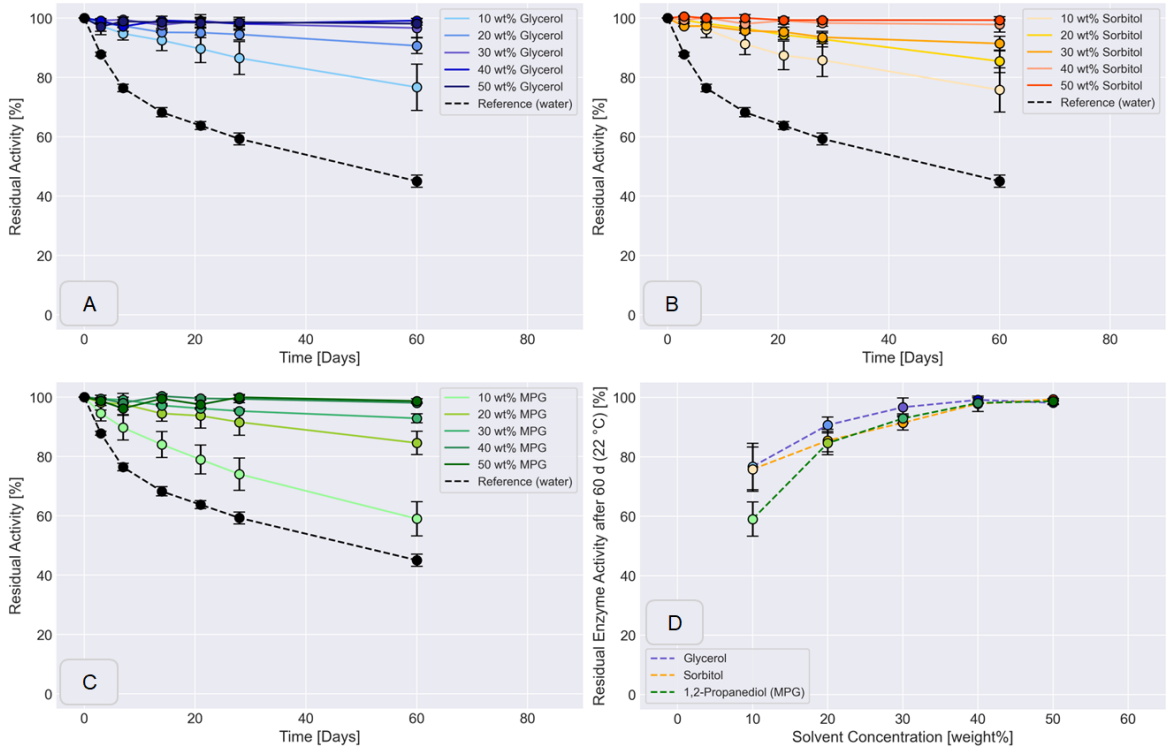


Figure 16: Residual enzyme activity [%] of the subtilisin protease in different polyol formulations (10-50 wt%) after storage at 22 °C over 60 days. *Samples were withdrawn after 0, 3, 7, 14, 21, 28 and 60 days. (A) Glycerol, (B) Sorbitol, (C) 1,2-Propanediol (MPG), and (D) Residual enzyme activity [%] after 60 days of storage.*

When incubating the samples at 37 °C (*fig. 17*), a clear separation between the different solvents and concentrations becomes evident. In the case of the samples with 10 wt% MPG (60 d; 16 %) (*fig. 17C*) only a minor improvement of stability with regard to the reference in water (60 d; 12 %) can be determined. Similar formulations with glycerol and sorbitol in turn (*fig. 17A+B*), exhibit an active enzyme content of 27 and 25 % after 60 days. As solvent concentrations increase, the active enzyme concentration of sorbitol and MPG begin to converge (*fig. 17D*), starting in the presence of 20 wt% solvent (S20 (60 d; 37 %), M20 (60 d; 35 %)). In the presence of higher concentrations, MPG (M30 (60 d; 58 %), M40 (60 d; 79 %), M50 (60 d; 91 %)) exhibits slightly higher active enzyme concentrations, when compared to sorbitol (S30 (60 d; 55 %), S40 (60 d; 76 %), S50 (60 d; 91 %)). Formulations with glycerol in turn, illustrate considerably higher stabilities in the presence of 20 (60 d; 45 %) and 30 wt% (60 d; 72 %) solvent. This level is maintained at 40 wt% (60 d; 89 %) glycerol, however, begins to converge towards 50 wt% (60d; 95 %).

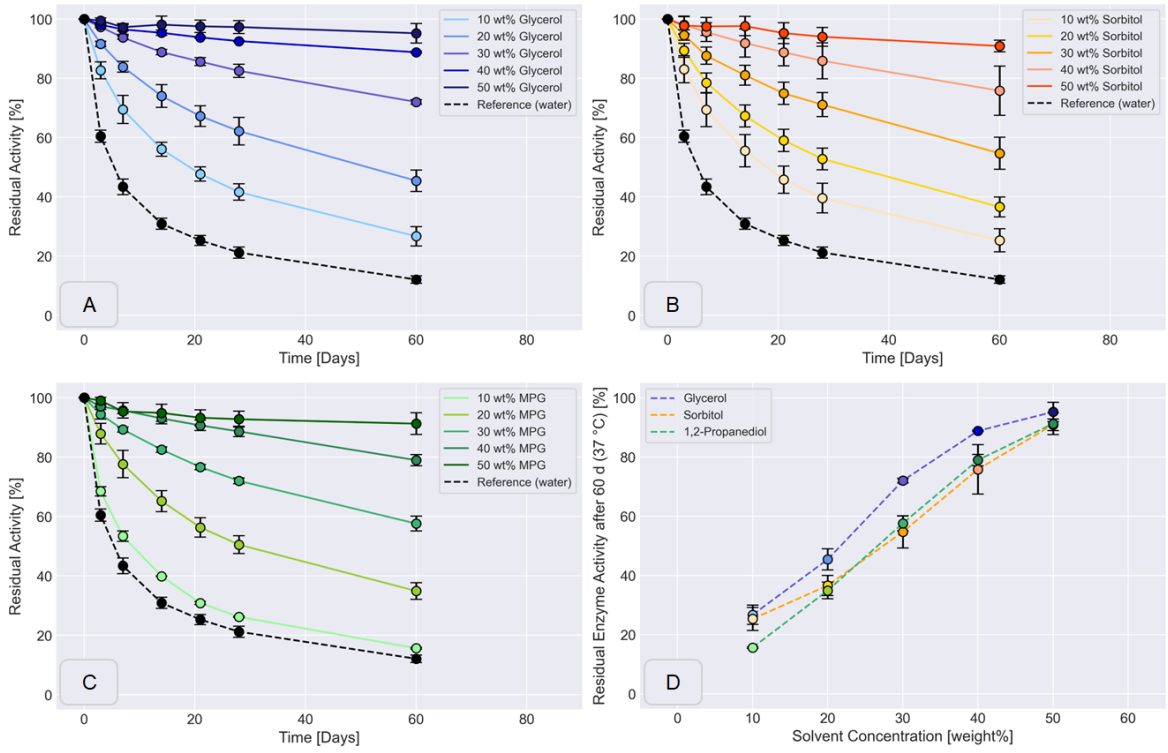


Figure 17: Residual enzyme activity [%] of the subtilisin protease in different polyol formulations (10-50 wt%) after storage at 37 °C over 60 days. *Samples were withdrawn after 0, 3, 7, 14, 21, 28 and 60 days. (A) Glycerol, (B) Sorbitol, (C) 1,2-Propanediol (MPG), and (D) Residual enzyme activity [%] after 60 days of storage.*

The most significant change of the active enzyme concentration can be determined for the samples stored at 45 °C (*fig. 18*). Especially, samples with a solvent concentration of 10-20

wt% illustrate a drastic decline of activity within the first sampling timepoints. For samples with 10 wt% solvent, no enzyme activity was measurable after 60 days. Samples with 20 wt% polyol exhibit only minor differences close to zero (G20 (60 d; 4 %), S20 (60 d; 0 %), M20 (60 d; 3%). At higher solvent, concentrations, however, a clear trend for the different solvents begins to develop. Similar to before, values of sorbitol (60 d; 10 %) and MPG (60 d; 8 %) show to converge in the presence of 30 wt% solvent. Above, however, the residual enzyme activity of the samples with sorbitol (S40 (60 d; 21 %, S50 (60 d; 48 %)) show to significantly increase, when compared to MPG (M40 (60 d; 13 %), M50 (60 d; 20 %). Formulations with Glycerol in contrast, illustrate a considerably higher residual enzyme activity and therefore stabilizing effect at all concentrations above 10 wt%. This effect is first visible at 20 wt% and then maintained across higher concentrations ((G30 (60 d; 16 %), (G40 (60 d; 33 %)) and culminates in a residual active enzyme content of 60 % after 60 days at 45 °C.

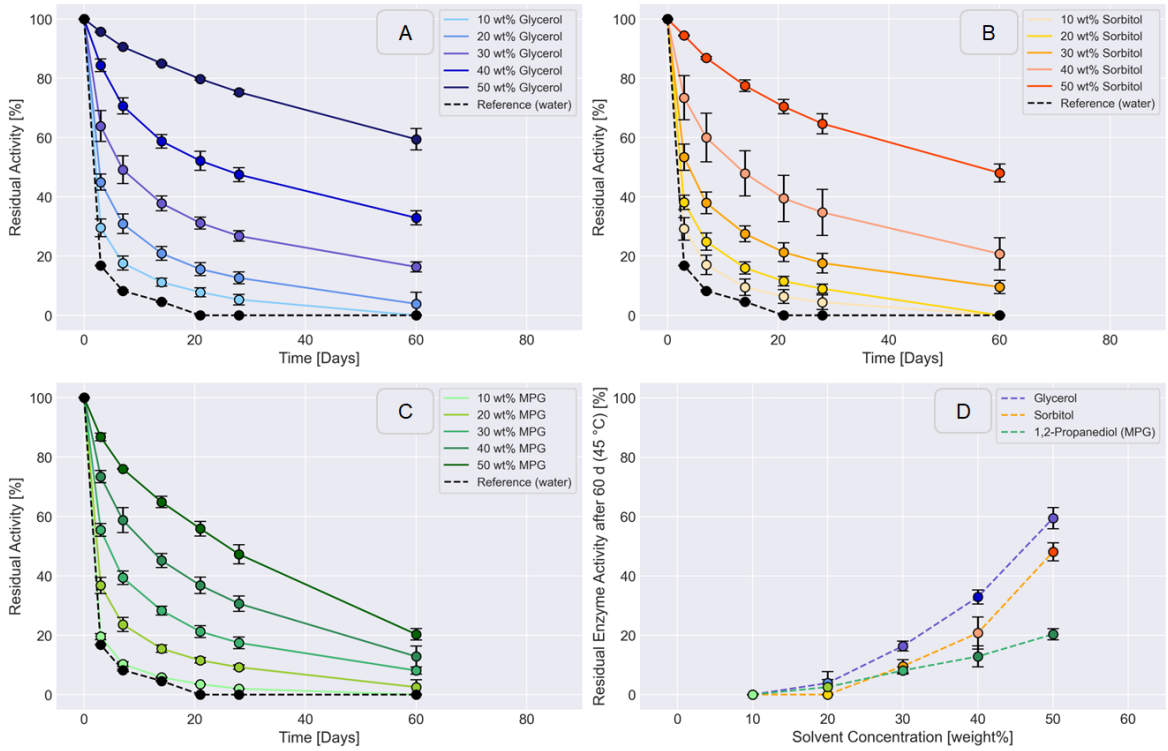


Figure 18: Residual enzyme activity [%] of the subtilisin protease in different polyol formulations (10-50 wt%) after storage at 45 °C over 60 days. *Samples were withdrawn after 0, 3, 7, 14, 21, 28 and 60 days. (A) Glycerol, (B) Sorbitol, (C) 1,2-Propanediol (MPG), and (D) Residual enzyme activity [%] after 60 days of storage.*

The conducted long-term storage tests illustrate the stabilizing effect of polyols on the subtilisin protease, when compared to the enzyme in water. Besides, higher solvent

concentrations have shown to increase this effect. Based on the obtained results formulations with glycerol exhibit the highest residual active enzyme content. Formulations with Sorbitol and MPG in turn, illustrate varying results dependent on concentration and temperature.

3.2.4 Correlation Analysis between Thermal Stability and Long-Term Stability Results of a Subtilisin Protease in Presence of Polyols

In the past, it was demonstrated that results of DSC measurements, especially *Tm* values, show a good correlation with results from storage stability studies. Among others, it was shown to be applicable for subtilisin proteases in liquid detergents [187]. Consequently, a possible correlation, between the T_m values determined in *chapter 3.2.1* and the residual active enzyme content after 60 days at 37 °C from *chapter 3.2.3* was investigated.

For better comparability of the results, T_m and the residual activity values (60 d) were normalized on the basis of the results obtained with the respective sorbitol concentration. Following, the obtained values were plotted based on the solvent present in the formulations. Since two distinct peaks were visible in the DSC measurements both T_m values were used, in separate approaches.

The obtained results show to be similar (*fig. 19*), independent of the peak used for the analysis. During DSC measurements, sorbitol has shown to increase the thermal stability of the subtilisin protease the most, when compared to the water reference. Subsequently, glycerol and MPG formulations exhibit values below 1 after normalization. In contrast, several formulations with glycerol and MPG have shown to exhibit higher residual enzyme activities, when compared to sorbitol. With regard to the normalization, this leads to values above 1. Especially, glycerol shows to retain a significantly higher active enzyme content, after 60 days, when compared to sorbitol. For MPG, T_m values have shown to be significantly lower than for similar sorbitol concentrations. Besides, increasing MPG concentrations appear to destabilize the subtilisin protease, which becomes evident by reduced T_m values. However, when looking at the active enzyme content after 60 days of storage, several MPG concentrations illustrate a higher residual activity, then the respective sorbitol formulations.

Subsequently, it could be illustrated that for the subtilisin protease, T_m values from DSC measurements are not suitable for accessing the residual activity after several days of

storage. In contrast to earlier reports, results of both datasets were missing any correlation. Thereby, it is not possible to predict the effect of different polyols on the long-term stability of the used subtilisin protease based on DSC measurements.

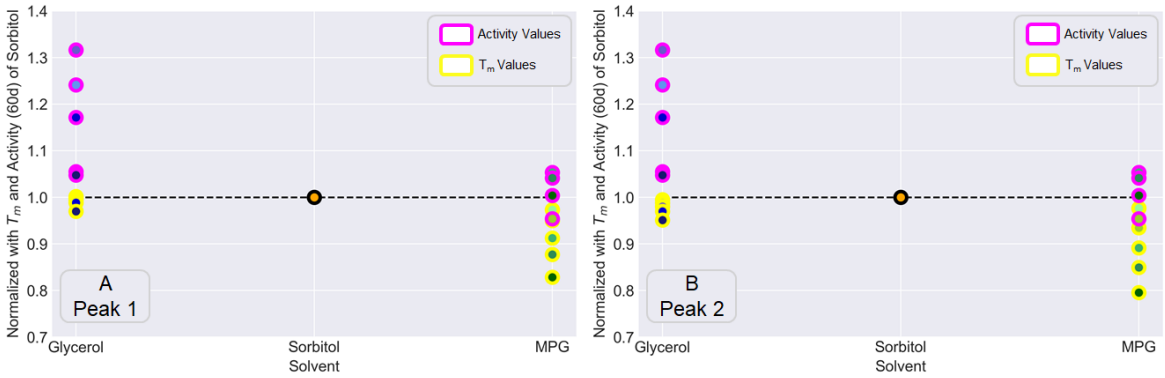


Figure 19: Correlation between T_m values and residual enzyme activities of a subtilisin protease in different polyol formulations after storage at 37 °C over 60 days. *Obtained values for glycerol and MPG were normalized with regard to the corresponding sorbitol results. Since two distinct peaks were visible in DSC, T_m values of both peaks were evaluated.*

3.2.5 Discussion

A popular strategy to increase protein respectively enzyme stability in liquid formulations is the use of cosolvents. These so called osmolytes, are in most cases, small molecules, as sugars, polyols, and neutral amino acids [140, 142, 193], which stabilize the folded state of a protein. In the past, various studies have shown their ability to increase thermal and long-term stability of globular proteins. This effect appears to be mediated by the physicochemical characteristics of these molecules, such as structure, charge, size, and their ability to interact with other solvent molecules in aqueous solutions [194].

Numerous mechanisms have been proposed to explain the molecular basis of polyol-induced protein stabilization. Among others, Wyman linkage function [146], excluded volume effect [195], transfer of protein energy to chemical groups [196], modification of surface tension [197], and preferential exclusion [198] have been considered. Besides, the type and the amount of cosolvent seem to play a critical role in enhancing conformational stability.

Preferential exclusion is mediated by the ability of polyols to repel non-polar substances. In an aqueous environment several protein residues are exposed to the outside. When

nonaqueous solvents are added, the solvent molecules displace water, whereby the environment of these residues is changed. Subsequently, the composition of the water lattice is changed. This effected is mediated by the ability of polyols to form hydrogen bonds with water molecules, which reinforce water interactions [199]. This, however, makes contacts between non-polar residues of the protein and the polyol solution even more unfavorable, than the contact with water [200]. Non-polar groups on the proteins surface, would normally react to these changes by orientating themselves towards the interior of the protein, however, there ability to do so is hindered by the protein structure. Consequently, the solvent structure surrounding the protein is changed, as to keep the chemical potential respectively activity of the solvent components constant. Subsequently, polyol molecules tend to migrate away from the protein surface. The resulting space is occupied by water molecules, which form a protective layer around the protein. Since unfolding of the protein structure would expose additional hydrophobic residues to the solvent, this would result in a thermodynamically unfavorable situation, which would require more free energy than in water. Subsequently, polyols tend to preserve the proteins native fold [143-145, 201].

In the conducted DSC and MD experiments, especially larger/hydrophilic polyols show to significantly increase the thermal stability of the subtilisin protease. Therefore, the obtained results, show to be in good accordance with the preferential exclusion mechanism. The observed effect appears to be mediated by a more compact structure of the subtilisin protease, which is illustrated by a reduction of the solvent accessible surface area, radius gyration and an increased number of preserved native contacts. Besides, the enzyme structure appears to be more rigid, as the RMSD and RMSF are reduced in the presence of larger/hydrophilic polyols.

Since larger/hydrophilic polyols tend to mediate a local accumulation of water molecules around hydrophobic protein residues, an increase of the hydrophobic surface area would not be favorable. The extend of this effect in the presence of different polyols/-concentrations is expected to be the result of indirect interactions, mediated by the ability of polyols to reduce the entropy around the protein, based on their molarity and volume [202]. A reduced entropy in the first hydration shell around the protein, is expected to shift the equilibrium towards the folded state, enabling higher thermal stability. It was found that this effect is positively correlated with both the molecular volume and the fractional polar surface area of polyols.

Subsequently, smaller/hydrophobic alcohols/polyols are expected to reduce the need of hydrophobic residues to evade the surrounding environment. Far more, smaller/hydrophobic alcohols and polyols tend to displace water from the protein surface, whereby the increased entropy in the protein's proximity, is expected to make the unfolded state thermodynamically more favorable. Consequently, the subtilisin protease stability appears to be significantly reduced in the presence of smaller/hydrophobic alcohols/polyols respectively different MPG concentrations, when compared to the water reference. This is further illustrated by an increased surface area and diameter of the subtilisin protease in the presence of smaller/hydrophobic alcohols/polyols, which is assumed to be the beginning of thermal unfolding. Subsequently, the fraction of native contacts appears to decrease with the molecular volume and fractional polar surface area of the alcohols and polyols. Besides, the fluctuations of the protein structure and single amino acids shows to significantly increase with size and hydrophobicity of the underlying solvent.

The observed effect of the different alcohols/polyols on the T_m values of DSC measurements, together with the structural changes determined by different molecular descriptors based on MD simulation trajectories, have illustrated the high conformity between the experimental and computational results. Subsequently, T_m values and averaged values from molecular descriptors exhibit a high correlation in the conducted correlation analysis. Both methods are extensively used to study the stability of proteins on a molecular level. In this regard, both methods utilize temperature as an evaluation criterion. While DSC measures the heat absorbed or released by an analyte, MD simulations mimic temperature by adapting the velocity of the atoms in the molecular ensembles. In the past, single molecular descriptors have shown to correlate well with experimental T_m values, however, not to the extend and rarely with the here reported accuracy [203-205]. The applicability of the conducted approach, however, has yet to be proven for other enzyme/solvent formulations. Nonetheless, the current approach has illustrated the possibility of predicting the thermal stability of an enzyme in the presence of different alcohols/polyols based on MD simulation derived from molecular descriptors.

In this context, simulation time and temperature are expected to influence the prediction performance [203]. Temperature was chosen to be slightly above the maximum T_m value of the most stabilizing subtilisin protease/polyol formulation (50 wt% sorbitol). Subsequently, the optimal temperature could appear to be enzyme respectively formulation dependent. Furthermore, the missing separation and correlation between the molecular descriptors, in the presence of both calcium ions in the subtilisin protease structure, has demonstrated the

significant influence of simulation time on the observed result. The enzyme appears to be too stable, in the presence of calcium, to observe unfolding related changes in the monitored time frame. Therefore, either simulation time has to be increased or the stability of the enzyme structure has to be decreased. However, the effects of a deliberate destabilization of the protein structure are difficult to assess, which makes the transferability to experimental results questionable. In the underlying case, the possibility has to be assessed, whether the loss of one or both calcium ions from the subtilisin protease structure is part of the thermal unfolding mechanism. If true, thermal stability studies would have to be assessed with the calcium free enzyme structure, and activity related mechanisms by utilizing the calcium occupied enzyme structure.

Even though, DSC and MD simulation results have illustrated a high level of correlation, residual enzyme activities have shown to significantly deviate from these results. In contrast to former reports [187], a correlation between thermal stability and residual enzyme activities, could not be shown with regard to the subtilisin protease. Subsequently, based on the obtained results, additional factors and processes affecting long-term stability of the subtilisin protease in different alcohol/polyol formulations, had to be considered. Although, DSC and MD simulations can provide a larger understanding of protein behavior in different environments, both methods try to assess the protein stability from a thermodynamic point of view. However, especially with regard to proteases, additional mechanisms for activity loss have to be taken into consideration.

Subsequently, the influence of polyols on the calcium binding state, and catalytic activity were considered in following chapters. MD simulation results have demonstrated the significant influence of calcium on the subtilisin protease stability and DSC results have shown to exhibit two distinct peaks. In this context, the presence of two distinct populations of subtilisin protease, with regard to the calcium binding state, was to be assessed (*chapter 3.3*). Furthermore, it could be assumed that autolysis contributes to a considerable extent to the loss of active enzyme. Consequently, a possible influence of the polyol formulation environment on the proteolytic activity, was investigated (*chapter 3.4*).

3.3 Influence of Calcium on the Subtilisin Protease Stability

The conducted MD simulations in *chapter 3.2.2* had illustrated the influence of calcium on the subtilisin protease stability. The subtilisin protease structure was observed to be considerably less stable in the absence of calcium in and around the enzyme structure. Considering, the two distinct peaks in the DSC measurements and the missing correlation between DSC and long-term storage test results, a possible influence of the calcium-binding state on the long-term stability was to be subject of further investigations. In this context, the possible existence of different populations of the subtilisin protease was to be investigated. Resulting differences in thermal stability and catalytic activity could account for the missing correlation in between DSC results and residual enzyme activities. Furthermore, possible changes in the ratio of the two peaks could be an explanation for the observed differences between the solvents/-concentrations. In this regard, DSC experiments were conducted to determine the influence of calcium on the subtilisin protease thermal stability, as well as on the ratio in between the two hypothetical states represented by the two distinct peaks.

3.3.1 Influence of Added Calcium on the Subtilisin Protease Thermal Stability

Prior to this chapter, 0.3 wt% calcium had been added to formulations to increase the general stability and solubility of the subtilisin protease. To investigate a possible influence of calcium on the subtilisin protease thermal stability and on the ratio between the two peaks, the DSC measurements conducted in *chapter 3.2.1* were repeated without the addition of calcium. For this, analogous to before, subtilisin protease/polyol formulations with ≈50 g/L active enzyme and 10-50 wt% (n= 5; interval= 10 wt%) glycerol, sorbitol, and MPG were prepared. Following, formulations were analyzed with DSC and the change of the T_m values in the absence of additional calcium was determined. Besides, the relative enthalpy [%] of both peaks was determined, by fitting the peaks with gaussian curves relative to the total enthalpy (ΔH). The T_m values and the relative area were afterwards compared to the results obtained in *chapter 3.2.1*.

The absence of additional calcium in the formulations results in significant differences (*fig. 20*) in between the observed T_m values and the results from *chapter 3.2.1*. In the absence of calcium, T_m values show to be decreased for the majority of the formulations. In numbers, formulations with glycerol (*fig. 20A*) illustrate an average decrease by 1.14 (± 0.18) °C for

the first peak and 1.03 (± 0.09) °C for the second peak. In a similar manner the T_m values for sorbitol (*fig. 20B*) are decreased by 1.45 (± 0.36) °C for the first and 1.83 (± 0.29) °C for the second peak. With MPG (*fig. 20C*), differences are only visible at lower polyol concentrations, while at higher differences diminish. The T_m values appear to be reduced by 0.98 °C (peak 1)/1.16 °C (peak 2) in the presence of 10 wt% MPG, whereas formulations with 50 wt% MPG exhibited differences of 0.03 °C for the first and 0.20 °C for the second peak.

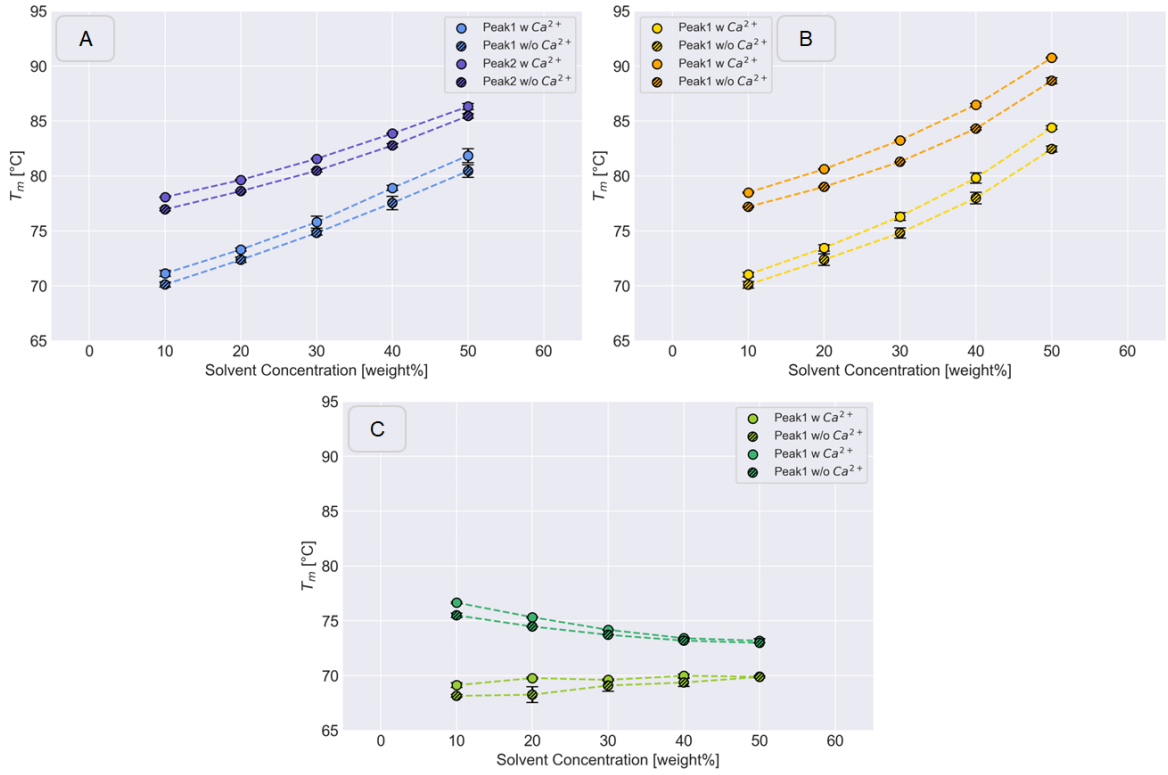


Figure 20: Comparison of the T_m values of the subtilisin protease in presence/absence of additional 0.3 wt% Ca^{2+} in different polyol formulations (10-50 wt%). *(A) Glycerol, (B) Sorbitol, and (C) MPG.*

Subsequently, the relative enthalpy [%] of both peaks was determined in the absence and presence of additional calcium. In this regard (*fig. 21*), results appear to be unchanged in between the subtilisin protease/polyol formulations. Differences, however, are visible in between the different solvent concentrations of glycerol and MPG (*fig. 21A+C*). While in both cases the first peak shows to comprise ≈60 % of the total enthalpy in the presence of 10 wt% solvent, the proportion decreases to ≈50 % at higher solvent concentrations. Subsequently, the area comprised by the second peak shows to increase by the same proportion. The formulations with sorbitol (*fig. 21D*), in contrast, maintain the ratio

between peak 1 and 2 in presence of the different concentrations. In this regard ≈60 % of the area is represented by the first and ≈40 % by the second peak.

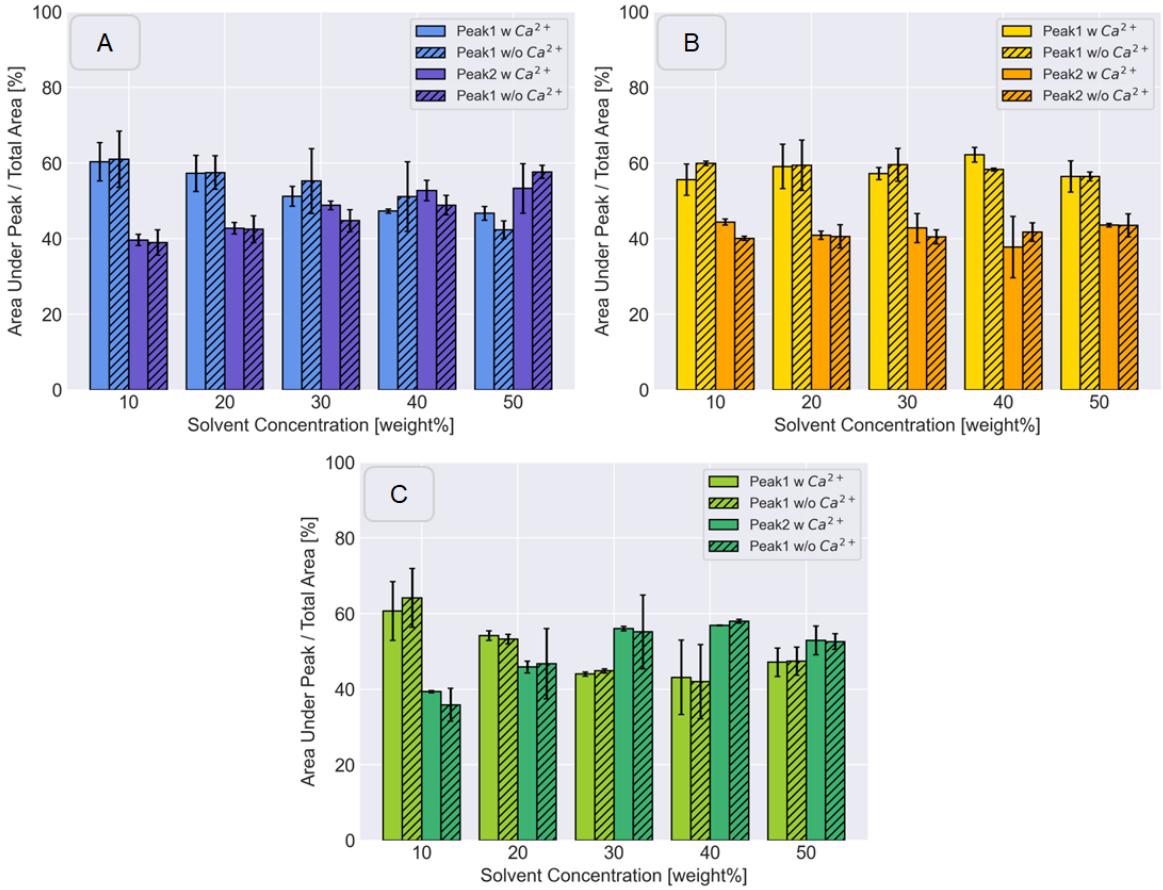


Figure 21: Enthalpy of the first and second Peak relative [%] to the total enthalpy (ΔH) in presence/absence of additional 0.3 wt% Ca^{2+}, determined for different subtilisin protease/polyol formulations. *(A) Glycerol, (B) Sorbitol, and (C) 1,2-Propanediol (MPG).*

The comparison of the subtilisin protease stability, in absence and presence of additional calcium, illustrates the stabilizing effect of calcium, when added to formulations. In general, it can be assessed that the presence of calcium leads to an increased thermal stability of the subtilisin protease with regard to the measured T_m values. However, an influence on the ratio of the two hypothetical subtilisin protease species, with regard to the calcium binding state, could not be observed. However, what can be observed, is a solvent concentration dependent change of the ratio in between both peaks in the presence of glycerol and MPG.

3.3.2 The Effect of Chelators on the Observation of two Distinct Peaks

The results obtained in *chapter 3.3.1* had demonstrated the stabilizing effect of calcium on the subtilisin protease structure. However, the presence respectively absence of additional calcium had shown to not affect the ratio between the observed peaks. Since the exact amount of calcium in the used enzyme concentrate is unknown, chelating agents were used to further reduce the free calcium content in selected subtilisin protease/polyol formulations. In this context, two random solvent concentrations were selected, which had shown to exhibit two distinct peaks. Subsequently, 5 wt% (pH 6.0) ethylenediamine-*N*,*N*′-disuccinic acid (EDDS), ethylenediaminetetraacetic acid (EDTA), and methylglycinediacetic acid trisodium salt (MGDA) were added to the formulations in three separate approaches. For this, two formulations with ≈50 g/L subtilisin protease and 30 wt% glycerol respectively 10 wt% MPG were prepared. Following, the thermal stability of the subtilisin protease was to be assessed by conducting DSC measurements.

In the presence of the chelating agents, the morphology of the thermogram (*fig. 22*), with two distinct peaks, is largely lost or diminished. In the case of EDDS, peaks appear more gaussian shaped and overlap with the second peak of the former thermogram. For MGDA similar peaks are visible, however, slightly shifted towards lower temperatures. This shift is observed to further increase in the presence of EDTA. In this case, both peaks show to be considerably shifted to lower temperatures and appear to be broader when compared to the other formulations.

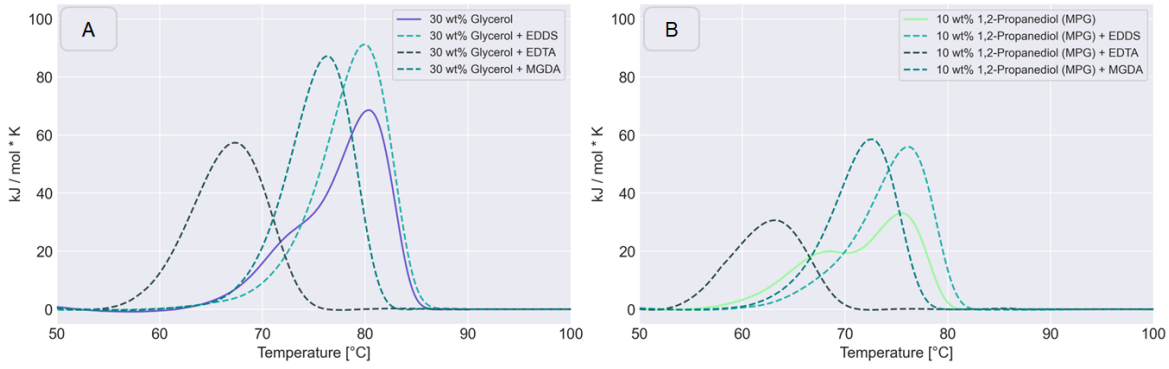


Figure 22: DSC results of formulations with the subtilisin protease (≈50 g/L) and glycerol (30 wt%) respectively 1,2-Propanediol (MPG; 10 wt%) after the addition of EDDS, EDTA and MGDA (5 wt%; pH 6.0). *(A) glycerol and (B) MPG.*

The use of chelating agents had shown to shift the bi-peaked morphology of the thermogram towards a single gaussian peak. Subsequently, the use of chelators has shown to change the ratio between both peaks. However, upon addition of the chelators, the formulations environments is expected to be considerably altered. Consequently, the observed effect, could either be the result of different populations or an artifact caused by the presence of the chelating agents in the subsequent analysis. In order to be able to better compare the resulting subtilisin protease populations, before and after complexation, the chelator was to be removed after complexation. Thereby, the effect of the chelator on the experimental results was to be diminished, resulting in a better comparability and significance of the results.

3.3.3 Removal of Chelating Agents after Calcium Complexation

In order to diminish the influence of the chelating agents on the conducted DSC measurements, the chelator was to be removed after complexation by dialysis. Thereby, the thermostability of the treated subtilisin protease was studied in the same environment as the native enzyme. In this regard, different protease samples were generated, with the goal to reduce the calcium content in the enzyme structure and the surrounding environment. For this, the subtilisin protease, with a concentration of ≈100 g/L, was dialyzed in deionized water, in its native state (WT/dia) and after the treatment with 5 wt% (pH 6.0) MGDA (WT/dia+MGDA). Subsequently, the thermostability of WT/dia and WT/dia+MGDA was compared to the native enzyme (WT) and to WT/dia+MGDA after the repeated addition of calcium (WT/dia+MGDA^{+}). Since, dialysis had significantly reduced the active enzyme content, the concentration of all samples was adjusted to ≈30 g/L. The thermal stability of the different approaches was afterwards determined by DSC.

With regard to the unprocessed sample in water, the thermogram (*fig. 23; black*) shows a high similarity with the results obtained in *chapter 3.2.1*. The thermogram of WT/dia (*fig. 23; blue*), in turn, shows to be shifted towards lower temperatures. The treatment with MGDA and the following dialysis (*fig. 23; grey*) seems to further amplify this effect, shifting the thermogram to even lower temperatures. When repeatedly adding calcium to WT/dia+MGDA (*fig. 23; yellow*) the effect appears to be reverted and the thermogram returns to the position of the subtilisin protease in its native state, however, with a reduced amplitude.

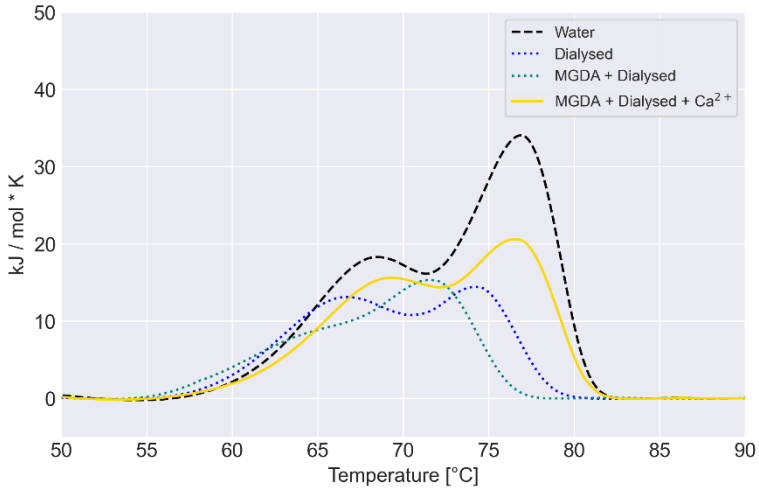


Figure 23: DSC results of the subtilisin protease in water after different treatments to reduce the calcium in and round the enzyme structure.

The analysis of the T_m values of the two distinct peaks, reveals additional differences. For WT/dia a similar shift of the T_m values of both peaks (*fig. 24A+B*), when compared to the unprocessed sample can be observed. In this context, the first peak (*fig. 24A*) appears to be shifted by 2.1 °C (68.98 → 66.87 °C), and the second (*fig. 24B*) by 2.3 °C (76.81 → 74.50 °C). The sample incubated with MGDA prior to dialysis (*fig. 24C+D*) illustrates additional changes, besides the general shift of the thermogram towards lower temperatures.

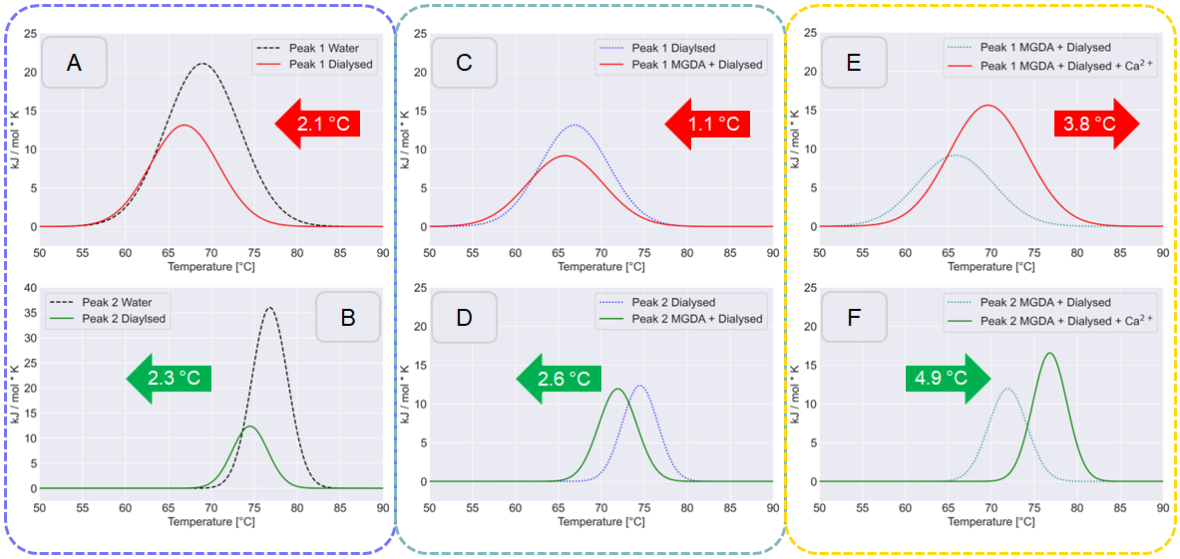


Figure 24: Obtained peaks after the deconvolution of the thermograms and comparison of the temperature shift from sample to sample. (Blue A + B) *From unprocessed to dialyzed sample, Peak 1 and Peak 2.* ***(Teal C + D)*** *From dialyzed to MGDA treatment and dialyzed, Peak 1 and Peak 2.* ***(Yellow E + F)*** *From MGDA treatment and dialyzed to MGDA treatment, dialyzed and calcium addition.*

In this regard, the first peak (*fig. 24C*) appears to be shifted by 1.1 °C (66.87 → 65.80 °C), while the second (*fig. 24D*) appears to be shifted by 2.6 °C (74.50 → 71.91 °C). Subsequently, the peaks appear to move closer together. The trend is reverted (*fig. 24E+F*) upon the addition of 0.3 wt% calcium, whereby both peaks are shifted to higher temperatures and the temperature gap between both peaks is again enlarged. In this context, the first peak (65.80 → 69.60 °C) (*fig. 24E*) is shifted by 3.8 °C, while the second (71.91 → 76.81 °C) (*fig. 24F*) is shifted by 4.9 °C. Subsequently, the effect of MGDA and dialysis appears to be reverted and the thermograms of WT and WT/dia+MGDA^{+} show to overlap.

The determined T_m values have shown, that besides a general shift of the thermograms, the different peaks appear to be shifted to one another. Therefore, the difference in between the T_m values of both peaks was set as an evaluation criterion (*fig. 25*). The distance for the unprocessed sample is determined to be 7.83 °C (*fig. 25; gray*). Upon dialysis in deionized water (*fig. 25; blue*) only a moderate change to 7.63 °C can be observed. However, as was already visible based on the T_m values, the temperature gap significantly decreases upon incubation with MGDA (*fig. 25; dark gray*), which returns a distance of 6.11 °C in between both peaks. When calcium is added (*fig. 25; yellow*) to this sample, the distance is again increased, which results in a difference of 7.20 °C.

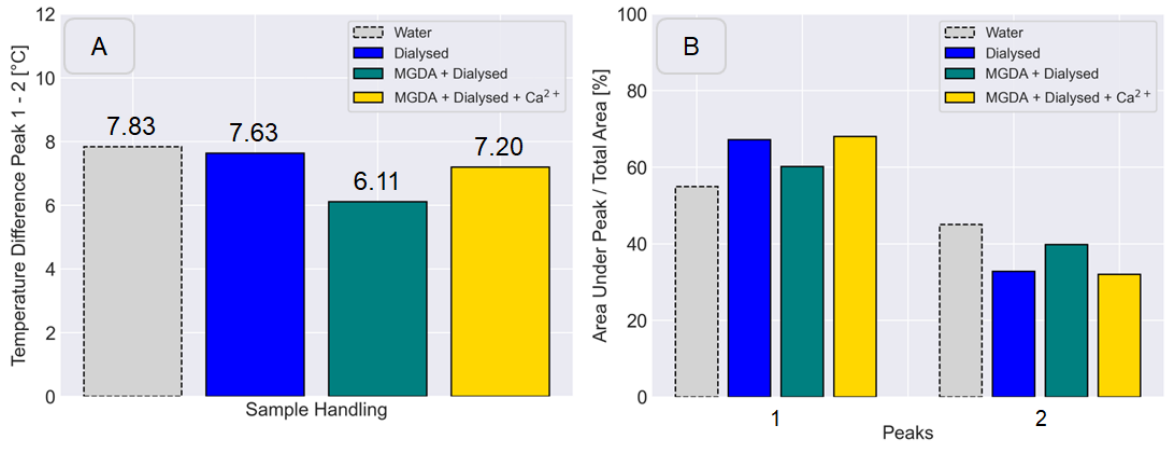


Figure 25: Thermal gap between the first and second peak (A) of the thermogram and relative enthalpy [%] of the peaks with regard to the total enthalpy (ΔH) (B).

Besides monitoring the changes of the T_m values and the resulting temperature gap in between both peaks, the area under the two distinct peaks was determined relative [%] to the total enthalpy (ΔH) (*fig. 25B*). With respect to the unprocessed sample, the analysis

illustrates the first peak to comprise 55 % of the total enthalpy and the second 45 %. After dialysis, the relative area of the first peak shows to increase to 67 %, when compared to the unprocessed sample, while the area under second peak decreases to 33 %. When adding MGDA to the subtilisin protease prior to dialysis, the area of the first peak is again decreased to 60 %, whereas the second peak increases to 40 %. Upon the addition of 0.3 wt% calcium to this sample, the relative area of the first peak is again increased to 68 % and the second decreases to 32 %.

In contrast to the results in *chapter 3.3.2*, the thermogram of WT/dia+MGDA did not exhibit the trend towards a single gaussian shaped peak. However, the treatment with the chelator had shown to greatly affect the thermostability and the ratio of the two peaks. Besides, WT/dia had illustrated the already destabilizing effect of dialysis on the thermostability of the subtilisin protease. Furthermore, the chelator showed to not only further reduce the thermostability but also affect the distance and ratio between both peaks.

3.3.4 Discussion

Metal ions play a major role in biological systems, especially for the structure, regulation, and enzymatic activity of proteins. More than one quarter of all proteins in the Protein Data Bank, have shown to contain bound metal ions [206]. The greater part of these proteins contains bound zinc, iron, calcium, and magnesium.

Subtilisin proteases are classified to the S8 family and assigned to the SB clan of serine proteases [207]. They are characterized by a catalytic triad of aspartate, histidine, and serine, which is common for all subtilisins of Bacillus species [89]. Subtilisins, further contain two or more cation binding sites, which are occupied by calcium or sodium ions [106, 208, 209]. Members from family A, subtilisins, generally have two Ca^{2+} sites, a high and low affinity binding site, although a calcium independent phenotype has been reported. Both sites are shared by thermitase (family B), which contain a third binding site of intermediate binding strength [210]. Furthermore, proteinase K, from family C, has two binding sites, which are however different to the ones found in subtilisins and thermitase [211]. Lastly, the Ak. 1 protease has shown to comprise four cation binding sites, three occupied by Ca^{2+} and the fourth by Na^{+} [106].

The Ca-1 binding site has shown to be conserved in various members of the subtilisin superfamily and is often referred to as high-affinity binding site. It is shared by subtilisins, thermitase, and Ak.1 protease. In the past, protein engineering, thermodynamic and kinetic

studies had demonstrated the unique role of this binding site for the stabilization of subtilisins [212, 213]. Ca-1 appears to be formed only, after removal of the 77 amino acid residue propeptide, which mediates the correct folding of the enzyme. Subsequently, once calcium is bound, there is large enthalpic barrier for unfolding, as the Ca^{2+} ion is buried inside the protein structure [214]. Therefore, the stabilizing effect of Ca-1 on the folded enzyme structure is likely to be very high in all enzymes. This effect is expected to be mediate by interactions with the backbone carbonyl oxygen atoms of Leu75, Ile79, and Val81 and the side-chain oxygen of Asn77 (*subtilisin BPN* numbering) [215]. Structurally the Ca-1 binding site is primarily composed by the loop (residue 75-83) that protrudes from the α3-helix, which provides three oxygen ligands from the side chains of Glu2 and Asp41, and thereby coordinates the cation from either side of the loop.

The low affinity binding site of subtilisins is located in a narrow crevice more than 30 Å away from the high affinity binding site [215]. Ca-2 is primarily formed by a single loop that is part of the ß2-α3 connection, which consists of the peptide carbonyl oxygen atoms of Gly169, Tyr171, Val 174, and Glu195, and two water molecules. Therefore, it is concluded that the calcium binding strength of Ca-2 is significantly smaller than for Ca-1. Pantoliano *et al.* have reported, that the calcium binding strength ($K(M^{-1})= 10^2$) in subtilisin BPN is indeed considerably lower than for the strong calcium binding site Ca-2 ($K(M^{-1})= 10^8$) [216].

In the past, metal ions have shown to play a significant role in stabilizing protein structures against denaturation [208] and proteolysis [217]. This ability was assessed to be mediated by their ability to reduce the flexibility of the polypeptide chain and thereby prevent local unfolding. This is particularly true for calcium, as it tends to bind carboxylate and oxygen ligands, which are the binding groups most commonly found in external loops [106]. In enzymes of the subtilisin family, calcium binding has shown to increase thermal stability and reduce the loss of enzyme due to autolysis. For several other subtilisins it could be, further shown to contribute to full enzymatic activity, by stabilizing the active conformation [218-220]. In the past, the loss of calcium from Ca-1 binding site has shown to result in destabilization and loss of activity, while deletion of the whole binding site retained full activity [221]. Similarly, it has been reported that the binding of a cation in Ca-2 dramatically improves the stability of subtilisin [108, 216].

The stabilizing effect of both calcium ions on the subtilisin protease structure was demonstrated by conducting MD simulations (*chapter 3.2.2*) of the calcium occupied and -

free enzyme structure in different formulation environments. Fluctuations of the subtilisin protease structure and single amino acids have shown to considerably increase, in the absence both calcium ions. Especially, the loops involved in calcium binding and loops in close proximity appear to be considerably more flexible in the absence of calcium. Subsequently, the subtilisin structure is expected to be less stable and therefore more susceptible to thermal unfolding and autolysis.

The absence of additional Ca^{2+}, similar to the dialysis of the native enzyme, has further shown to reduce the thermal stability of the subtilisin protease in the conducted experiments. The addition of low salt concentrations is known to increase the protein solubility in a process described as “salting in” [222]. When salt ions are added to enzyme solutions, they compete with enzymes for water and bind to charged groups or dipoles. Subsequently, interactions, between hydrophobic residues are increased, which results in a more compact protein structure, increasing its resistance to thermal unfolding reactions [223]. The absence of additional Ca^{2+} and a reduced Ca^{2+} concentration, thereby showed to reduce the thermal stability of the subtilisin protease in the conducted DSC measurements.

In contrast, the addition of chelators has shown to affect the two visible peaks in different ways. Based on the reduced distance between both peaks respectively the stronger shift of the second peak after the incubation with MGDA a direct effect on the bound calcium in the subtilisin structure is to be assumed. Since the calcium binding strength of Ca-2 was determined to be low ($K(M^{-1})= 10^2$) by Pantoliano *et al.* [216], MGDA is expected to withdraw the Ca^{2+} from the subtilisin protease structure. The binding strength of Ca-1 in turn, is assumed to be too large ($K(M^{-1})= 10^8$) to be affected by MGDA ($K(M^{-1})= 10^7$), thus the chelator is less likely to compete with the enzyme for the bound Ca^{2+}. Consequently, the change of the distance in between both peaks, is to be mediated by the loss of the Ca^{2+} from the Ca-2 binding site. Therefore, the second peak, is most likely to represent a subtilisin protease structure, with only one Ca^{2+} bound to Ca-1.

Since the first peak remained rather unaffected in the presence of MGDA, the shift appears to be mediated by the further reduction calcium in the environment, rather than being directly affected by the chelator. Consequently, the first peak is assumed to represent a calcium-free state of the subtilisin protease. What appears to contradict this assumption, however, is the fact that the relative area of the first peak shows to decrease upon the addition of MGDA rather an increase. Subsequently, the exact composition of the first peak remained unclear and was further elaborated in *chapter 3.4*.

The addition of calcium to WT/dia+MGDA showed to revert the effect of MGDA and dialysis, and partly restored a similar stability than observed for the native state of the subtilisin protease. This and the increased distance between both peaks appears to be mediated by the repeated binding of Ca^{2+} to the Ca-2 binding site. In the past, it was shown that the stability of subtilisin BPN' is highly dependent on the Ca^{2+} concentration [216]. In this regard, the addition of Ca^{2+} in the range of 0.1-100 mM had decreased the rate of thermal inactivation of the enzyme by 100-fold at 65 °C. A sequence of refined X-ray crystal structures had further shown that the binding of Ca^{2+} to Ca-2 is affected by the calcium concentration in the surrounding environment. Subsequently, the second peak in the native state of the subtilisin protease is assumed to resemble the active state of the subtilisin protease, with two bound calcium ions.

The conducted experiments added further insights on the origin of the two distinct peaks observed for different subtilisin protease/polyol formulations during DSC. In this context, the second peak appeared to represent the native state of the subtilisin protease with two occupied calcium binding sites. The exact composition of the first peak, however, remained largely unclear. Even though the missing reaction in presence of MGDA led to the assumption it could be a population of calcium-free subtilisin protease, the ratio in between both peaks appears to be unrelated with conditions favoring the calcium-free state. Furthermore, the repeated addition of Ca^{2+} to the same sample shows to increase the proportion of the first peak rather than decreasing it. Therefore, either the calcium-free state is not revertible, or the first peak is composed of something different. Subsequently, as part of *chapter 3.4* the composition of the first peak was to be further elaborated.

3.4 The Effect of Polyols on the Proteo-/Autolytic Activity of a Subtilisin Protease

The experiments in *chapter 3.3* had demonstrated the significant influence of Ca^{2+}, bound to the enzyme structure and in the surrounding environment, on the subtilisin protease stability. However, the exact composition of the two distinct peaks observed in DSC, and thereby the reason for the missing correlation between DSC and storage-test results remained largely unclear. Although the results indicated that the second peak was caused by the unfolding of the native state of the subtilisin protease and the first appeared to be calcium-free, it remained doubtful whether the first peak solemnly consisted of the calcium-free enzyme structure. Subsequently, the influence of autolysis on the observed DSC results and thereby on the missing correlation with the residual enzyme activities, was to be assessed. In this context, the possibility was investigated whether the observed first peak could originate from the thermal unfolding of released peptide fragments as part of the proteolytic activity of the subtilisin protease. Besides, the influence of the different alcohols/polyols on the catalytic activity of the subtilisin protease was determined. In this regard, enzyme kinetic parameters of the subtilisin protease in the presence of different polyols was assessed, based on the artificial AAPF substrate. Furthermore, the approximate autolysis rate in the different formulations, was to be determined by SDS-PAGE. At last, an influence of the alcohols/polyols on the catalytic mechanism of the subtilisin protease was to be investigated on an atomistic scale using MD simulation trajectories from *chapter 3.2.2*.

3.4.1 The Effect of Peptide Fragments on DSC Results

Subtilisin proteases, belong to the family of serine proteases, an enzyme class known for their ability to hydrolyze peptide bonds [89]. In the absence of other proteins, the main target of this degrading process are the proteases themselves. In a process, referred to as autolysis, the active enzyme content is reduced due to the proteolytic activity. Since subtilisin proteases are rather unspecific, peptide fragments of varying length and structure are released [224]. Subsequently, it was investigated whether the first peak observed during DSC measurement, could be the result of the fragmentation of the enzyme before and during the analysis. In this regard, size exclusion chromatography (SEC) was performed to separate shorter peptide fragments present in the formulations from the enzyme of interest. For this, a sample with ≈30 g/L subtilisin protease in deionized water

was used. Following, the composition of the different fractions obtained from SEC was analyzed by SDS-PAGE. In an additional step, several fractions were selected and their ability to exhibit a signal during DSC measurement was investigated.

Interestingly, the elution profile (*fig. 26*) obtained from SEC by UV detection, similarly to the DSC thermograms, exhibits two distinct peaks. The first after an elution volume of ≈70-80 mL and the second after ≈90-100 mL. To further elaborate on the composition of both peaks, several fractions of both peaks were analyzed by SDS-PAGE. Therefore, the still active protease, was first fractionated, to prevent autolysis from progressing and thereby influencing the result of the SDS-PAGE.

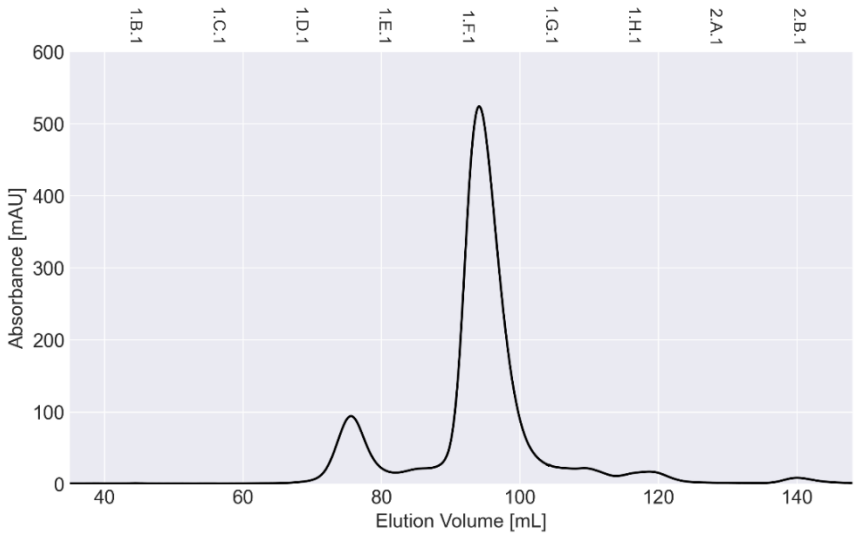


Figure 26: Size exclusion chromatography (SEC) elution profile for the subtilisin protease. *The line corresponds to the absorbance readings at 280 nm. Besides the elution volume, the top x-axis represents the different fractions, which were collected during SEC. The gap between these fractions is further distributed in 12 subfractions.*

Subsequently, different fractions of peak 1 and peak 2 were loaded to and analyzed by SDS-PAGE (*fig. 27A*). As was to be expected, the fractions of the two peaks exhibited significant differences, with regard to the observed peptide fragment pattern and distribution. Besides illustrating the expected band for the subtilisin protease at ≈27 kDa several additional peptide fragments, especially of smaller size, are visible. The quantity respectively intensity of the smaller fragments appears to be considerably higher, in the first peak (*fig. 27A; D8-11*), when compared to the second. However, also a small amount of the subtilisin protease appears to be present. In the case of the second peak (*fig. 27A; F2-F6*), the intensity of the subtilisin protease band significantly increases, due to the

considerably higher concentration. In contrast, the intensity/quantity of the smaller peptide fragments shows to be considerably lower, when compared to the first peak. Besides, fractions of both peaks show to exhibit a band with ≈54 kDa, which is assumed to be due to the formation of dimers in between subtilisin protease peptide chains.

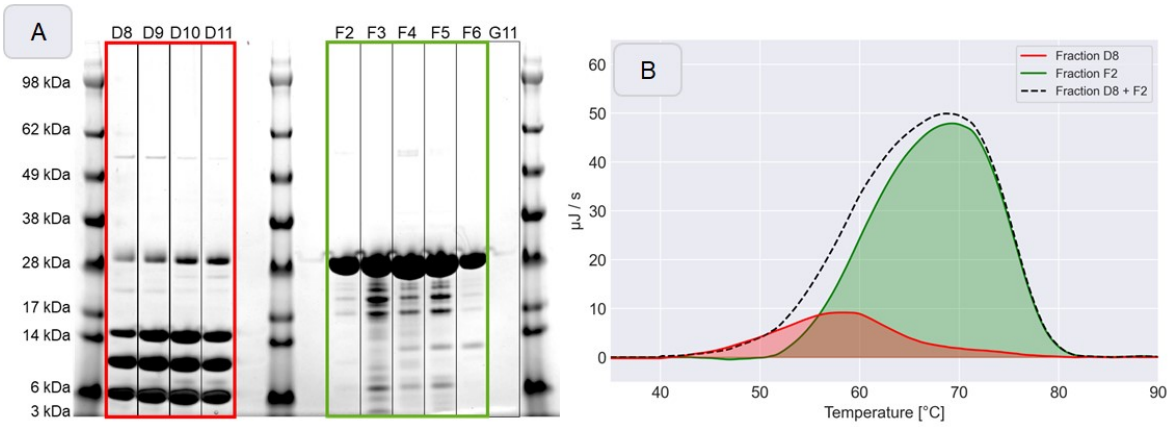


Figure 27: Analysis of the SEC fractions from peak 1 and 2. *(A) SDS-PAGE of the first peak (fractions D8-11) and second peak (fractions F2-6). The subtilisin protease was detect at ≈27 kDa. (B) DSC results of fractions D8 and F2.*

To determine the capability of the smaller peptide fragments to exhibit a measurable enthalpy in DSC, various fractions of both peaks were analyzed. The illustrated results from fractions D8 and F2 (*fig. 27B*), represent a general picture, observed during the analysis. In general, results show to considerably vary from the thermograms obtained in *chapter 3.2.1* and *3.3.3*. However, it has to be considered that the SEC had significantly reduced the peptide fragment/protein concentration in the different fractions. Subsequently, T_m value and enthalpy are assumed to be considerably lower than for the dialyzed enzyme in water. Nonetheless, the resulting peaks from fraction D8 and F2, and others, appear to overlap, similar to the two-peaks determined before. Furthermore, the SDS-PAGE has revealed small fractions of the native enzyme and smaller peptides in both peaks, which appear to become visible in the overlapping region of both peaks.

The SEC elution profile has revealed two distinct peaks similar to result obtained by DSC. Furthermore, fractions of the first peak, which are mainly composed of smaller peptide fragments, exhibit a signal in the conducted DSC measurements. This shows to overlap with the peak observed for fractions of the second peak, which have shown to be composed mainly of the full-length subtilisin peptide chain. Therefore, it is to be concluded, that the peptide fragments released as a result of the proteolytic activity of the subtilisin protease,

are expected to be the reason for the two distinct peaks observed in the DSC thermograms of different subtilisin protease formulations. Subsequently, the difference between the ratio of both peaks is assumed to be related to the autolytic rate of the subtilisin protease in the respective formulation environment. Therefore, the influence of the different alcohols/polyols on the autolysis rate was to be assessed in following chapters.

3.4.2 Michaelis Menten Kinetics of the Subtilisin Protease in the Presence of Different Polyols

The result obtained from *chapter 3.4.3* had illustrated the ability of peptides, which are released as result of the catalytic activity of the subtilisin protease, to exert a measurable enthalpy during DSC. Since these results contradicted the theory of different protease populations, with regard to the calcium binding state, to cause the observed differences between thermal stability and residual activity results, autolysis moved into focus. Therefore, the influence of different polyols, namely glycerol, sorbitol and 1,2-propanediol (MPG), on the subtilisin protease activity was investigated. For this Michalis-Menten kinetic parameters were determined, in order to assess the influence of the solvents on the Michaelis constant (K_m) and the maximum velocity (V_{max}) of the subtilisin protease, with regard to the used AAPF substrate. Subsequently, the AAPF assay protocol for microtiter plates was adapted to be performed in the respective buffer/polyol environment, in the presence of varying AAPF substrate concentrations.

Since the addition of polyols to the TRIS buffer affects the extinction coefficient of the reaction environment, it was to be redetermined in the presence of 10, 30, and 50 wt% glycerol, sorbitol, and MPG. The extinction coefficients were assessed, based on the slope of a dilution series performed, with the para-nitroaniline moiety, in the respective reaction environment at 30 °C. The exact values can be obtained from *chapter 2.2.8*. Subsequently, MM-kinetic parameters were determined by measuring the turnover of a fixed concentration of the subtilisin protease in presence of varying AAPF concentrations. For this, an enzyme stock solution with a concentration of 1 µg/mL and different starter solutions with a substrate concentration in between 0.06 to 3.38 mM were generated. Based on the average absorption change measured for the different solutions, in the presence of the subtilisin protease, the specific activity (U/mg) was determined. Following, K_m and V_{max} were obtained, by fitting the hyperbolic curve of the MM-equation to the datapoints. Further, the reciprocal values of the substrate concentration and specific

activity were determined, in order to visualize the proposed inhibition mechanism based on the Lineweaver-Burk plot.

Enzyme Kinetics

Based on the extinction coefficients for the different TRIS buffer/polyol environments, the specific activity of the subtilisin protease in the presence of varying AAPF concentrations was determined (*fig. 28*). For the subtilisin protease in TRIS buffer a K_m value of 3.9 mM and a V_{max} of 172.4 U/mg could be determined.

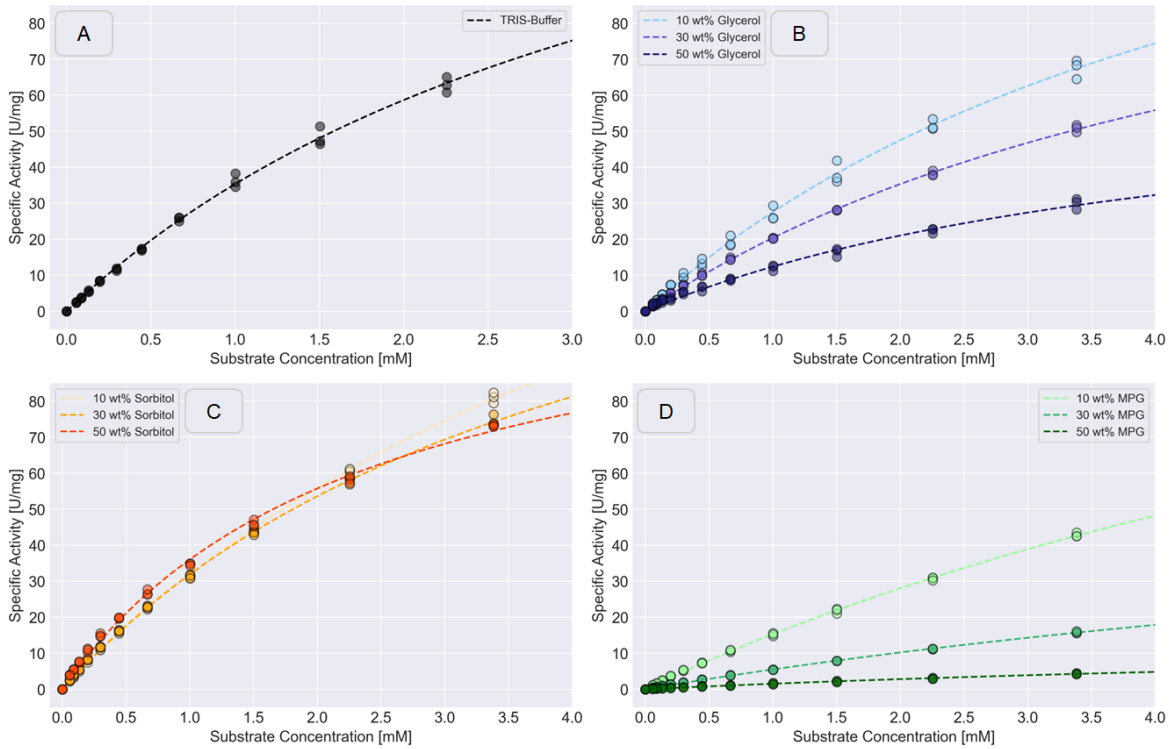


Figure 28: Michalis-Menten kinetics of the subtilisin protease in the presence of different polyol concentrations (10, 30, 50 wt%). *(A) TRIS-buffer, (B) Glycerol, (C) Sorbitol, and (D) MPG.*

The MM-constants of the other TRIS buffer/polyol systems show to considerably vary based on the present solvent/-concentration (*tab. 6*). In the presence of 10 wt% glycerol a minor increase of the K_m value to 5.2 mM can be observed, while at elevated glycerol concentrations the Michalis-constant appears to be mostly stable. Therefore, at a concentration of 30 wt% glycerol a K_m of 5.5 mM and at 50 wt% a K_m of 4.6 mM can be determined. In the case of sorbitol, the increase of the specificity appears to be even more related to the solvent concentration. Subsequently, in the presence of 10 wt% sorbitol a K_m

value of 6.1 mM, followed by 4.3 mM at 30 wt% and lastly 2.4 mM at 50 wt% can be observed. Considerably higher K_m values and thereby a significantly reduced specificity of the enzyme can be determined in the presence of MPG. In this case, K_m values show to drastically increase already in the presence of 10 wt% MPG. Showing a K_m value of 10.3 mM. In the presence of 30 and 50 wt% the specificity remained almost constant exhibiting values of 11.9 and 10.0 mM.

Table 6: Michaelis-Menten kinetic parameters determined for the subtilisin protease in different TRIS buffer/polyol environments.

Solvent	Concentration [weight%]	K_M (n= 3)	V_{max} (n=3)
TRIS	-	3.9 ± 0.2	172.4 ± 5.4
Glycerol	10	5.2 ± 0.4	171.7 ± 9.6
	30	5.5 ± 0.2	133.2 ± 3.3
	50	4.6 ± 0.5	69.5 ± 5.2
Sorbitol	10	6.1 ± 0.3	225.0 ± 6.7
	30	4.3 ± 0.1	167.8 ± 3.5
	50	2.4 ± 0.1	122.9 ± 3.2
1,2-Propanediol (MPG)	10	10.3 ± 0.6	171.9 ± 7.6
	30	11.9 ± 0.8	71.2 ± 4.1
	50	10.0 ± 1.9	16.8 ± 2.6

In the case of the maximum reaction velocity (*tab. 6*), changes mediated by the different solvents/-concentrations, appear to be even larger. However, in the presence of 10 wt% solvent, the V_{max} values of glycerol and MPG appear to be similar in size, when compared to the TRIS buffer environment. In this regard, the enzyme has shown to exhibit a V_{max} of 172.4 U/mg. In a similar range, reactions environments with 10 wt% glycerol and MPG have shown to illustrate a maximum reaction velocity of 171.7 respectively 171.9 U/mg. With 10 wt% sorbitol, surprisingly, a larger V_{max} of 225 U/mg can be determined. However, increasing the solvent concentration, shows to reduce the maximum velocity. This is particularly evident in the case of 30 and 50 wt% MPG, as V_{max} shows to be reduced to 71.2 U/mg respectively 16.8 U/mg. Similarly, reaction environments with 30 and 50 wt% glycerol illustrate a reduction of the maximum velocity, when compared to the original TRIS buffer. At first the V_{max} was reduced to 133.2 U/mg and then to 69.5 U/mg. In comparison, reaction environments with sorbitol appear to be only mildly affected by

increasing concentrations. After exceeding the maximum velocity in TRIS buffer, in the presence of 10 wt% sorbitol, values show to further decrease to 167.8 U/mg (30 wt%) and 122.9 U/mg (50 wt%).

The conducted experiments illustrate the effect of different polyols on the subtilisin protease substrate specificity and catalytic rate. In this context, MPG has shown to considerably reduce the specificity of the subtilisin protease towards the AAPF substrate. Furthermore, MPG has illustrated to significantly reduce the maximum reaction velocity. Thereby, it can be assumed that in the presence of high MPG concentrations, the subtilisin protease proteolytic activity is greatly diminished. A similar effect is observed to be mediated by glycerol, however not to the same extend. Still, glycerol is expected to considerably decrease the autolytic rate of the subtilisin protease at high concentrations. Sorbitol, however, is expected to be a weaker inhibitor, as the catalytic rate and maximum reaction velocity remains greatly uninfluenced by the surrounding solvent environment. Subsequently, different polyols have been shown to affect the proteolytic activity of the subtilisin protease at varying rates. Therefore, polyols could not only be shown to affect the thermal stability of the subtilisin protease, but also the activity.

Lineweaver Burk Plot

The Lineweaver-Burk plot is a graphical representation of the relationship between the catalytic rate of a reaction and the corresponding substrate concentrations. Historically these plots have been used to determine the Michalis-Menten kinetic parameters of enzymes, which are the maximum reaction velocity (V_{max}) and the substrate specificity (K_m). In a Lineweaver-Burk plot, the reciprocal of the reaction rate ($1/V$) is plotted against the reciprocal of the substrate concentration (1/[S]). The result is a linear plot, whose x-intercept represents the reciprocal K_m ($1/K_m$) and the y-intercept the reciprocal V_{max}. The advantage of using a Lineweaver-Burk plot besides the simple way of determining the kinetic parameters, is the visualization of different enzyme reaction conditions. This allows to distinguish between different inhibitory mechanisms affecting enzyme activity.

While the Lineweaver-Burk plot has historically been used for the evaluation of kinetic parameters, all linearized forms lost importance as computational approaches, being significantly more accurate, became accessible. Nonetheless, this approach is a convenient way visualize different inhibitory effects, comprised by different kinetic parameters. Subsequently, the Lineweaver-Burk plot was used, with regard to the obtained MM-kinetic

parameters, to visualize the effect of the different polyols on the subtilisin protease specificity and catalytic rate.

For glycerol K_m values have shown to be slightly increased (*fig. 29A*), when compared to TRIS buffer, though remain relatively constant in between the different concentrations. The V_{max} in turn, appears to be unchanged in the presence of 10 wt% glycerol. Thereby, the TRIS buffer and TRIS buffer/glycerol result share the same y-intercept. Thereupon, it could be inferred that at low concentrations glycerol appears to act as a competitive inhibitor. In this context, the V_{max} remains unchanged while the K_m is increased. A similar result could be determined in the presence of 10 wt% MPG (*fig. 29C*), whereby the same competitive inhibition is assumed to affect subtilisin protease activity.

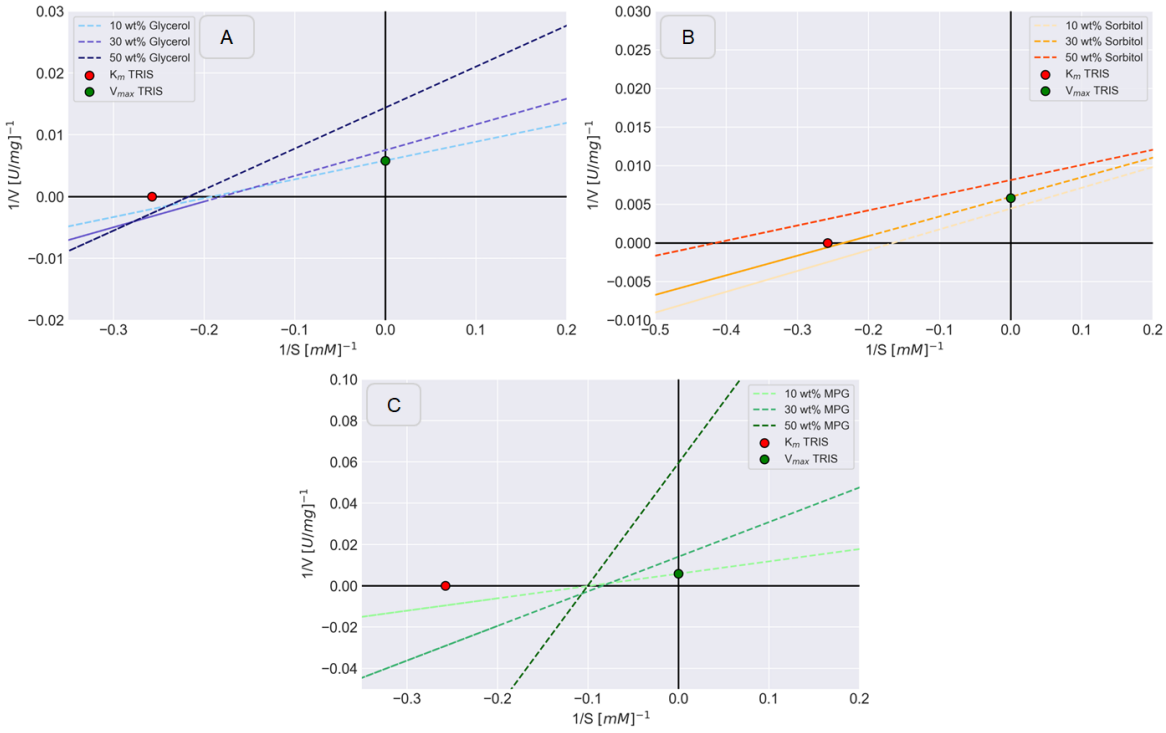


Figure 29: Lineweaver-Burk plot of the results obtained from MM-kinetic experiments. *(A) Glycerol, (B) Sorbitol, and (C) 1,2-Propanediol (MPG).*

An exception, however, can be determined for sorbitol (*fig. 29B*), as the decreased specificity would tend towards a competitive inhibition mechanism, which however, appears to be inconsistent with an increased reaction velocity. Furthermore, the inhibition type of the solvents appears to change with increasing concentrations. In the case of glycerol and MPG, higher concentrations show a noncompetitive (mixed) inhibition, which is illustrated by a largely unaffected K_m value and a considerably reduced V_{max}. In the

presence 30 wt% sorbitol K_m and V_{max} values appear to resemble the result obtained in TRIS buffer. Therefore, lower sorbitol concentrations appear to not affect the protease activity at all. In contrast, 50 wt% sorbitol shows to further reduce the K_m and V_{max}, which corresponds to an uncompetitive inhibition type.

The Lineweaver-Burk plot has provided an overview on the inhibition mechanisms, by which different polyols tend to reduce the proteolytic activity of the subtilisin protease. In this context, glycerol and MPG appear to act as a competitive inhibitor at lower concentrations while exhibiting the mechanism of an uncompetitive (mixed) inhibitor at higher concentrations. In contrast, sorbitol appears to hardly affect the subtilisin protease activity at 10 and 30 wt%. At higher concentration however, sorbitol appears to act as an uncompetitive inhibitor, increasing the substrate specificity, while decreasing the maximum reaction velocity.

3.4.3 The Effect of Polyols on Autolytic Activity of a Subtilisin Protease determined by SDS-PAGE

The results in *chapter 3.4.2* had demonstrated the effect of polyols on the substrate specificity and maximum reaction velocity of the subtilisin protease. In order, to further investigate the extent, to which this effect is relevant for the missing correlation in between DSC and storage-test results, the influence of autolysis was to be determined. In this context, the intensity of the ≈27 kDa subtilisin protease band, observed by SDS-PAGE, was selected as an evaluation criterion to monitor progression of autolysis throughout storage. Thereby, the influence of the polyols on the proteolytic activity was assessed, based on the degradation of the subtilisin protease itself. In this context, the decrease in the intensity of the full-length band was expected to correspond to the amount of enzyme/full length peptide lost due to autolysis. In contrast to earlier reports [225], the results were, however, not expected to correspond to the actual autolytic rate of the subtilisin protease, as the proportion of unfolded/inactivated enzyme comprised by the full-length band remains unknown throughout storage. Nonetheless, the results were regarded to provide an accurate overview about the extent to which long-term stability is influenced by autolysis in different subtilisin protease/polyol formulations.

For this, samples from a long-term storage test, conducted over 60 d at 4, 22, 37 and 45 °C in the presence of 10, 20, 30, 40 and 50 wt% glycerol, sorbitol and 1,2-propanediol (MPG), were collected (0, 3, 7, 14, 21, 28, 60 days) and prepared according to *chapter 2.2.8*.

Afterwards, samples from the same formulation and temperature were analyzed by SDS-PAGE. The obtained intensity values of the full-length band were normalized with regard to the intensity at timepoint 0 days. Furthermore, the active enzyme content of all samples was determined with the AAPF assay on the Gallery™ system and the residual enzyme activity was determined relative to the concentration at timepoint 0 days.

The determined residual enzyme activities show a high accordance with the results obtained in *chapter 3.2.3*. Consequently, samples at low temperature and in the presence of high solvent concentrations show to retain their stability over the sampled time period. Thus, no differences between residual enzyme activities and full-length band intensity could be determined. In contrast, samples with low solvent concentrations, as with 10 wt% polyol at 37 °C (*fig. 30*), illustrate a considerable decrease of the residual enzyme activity

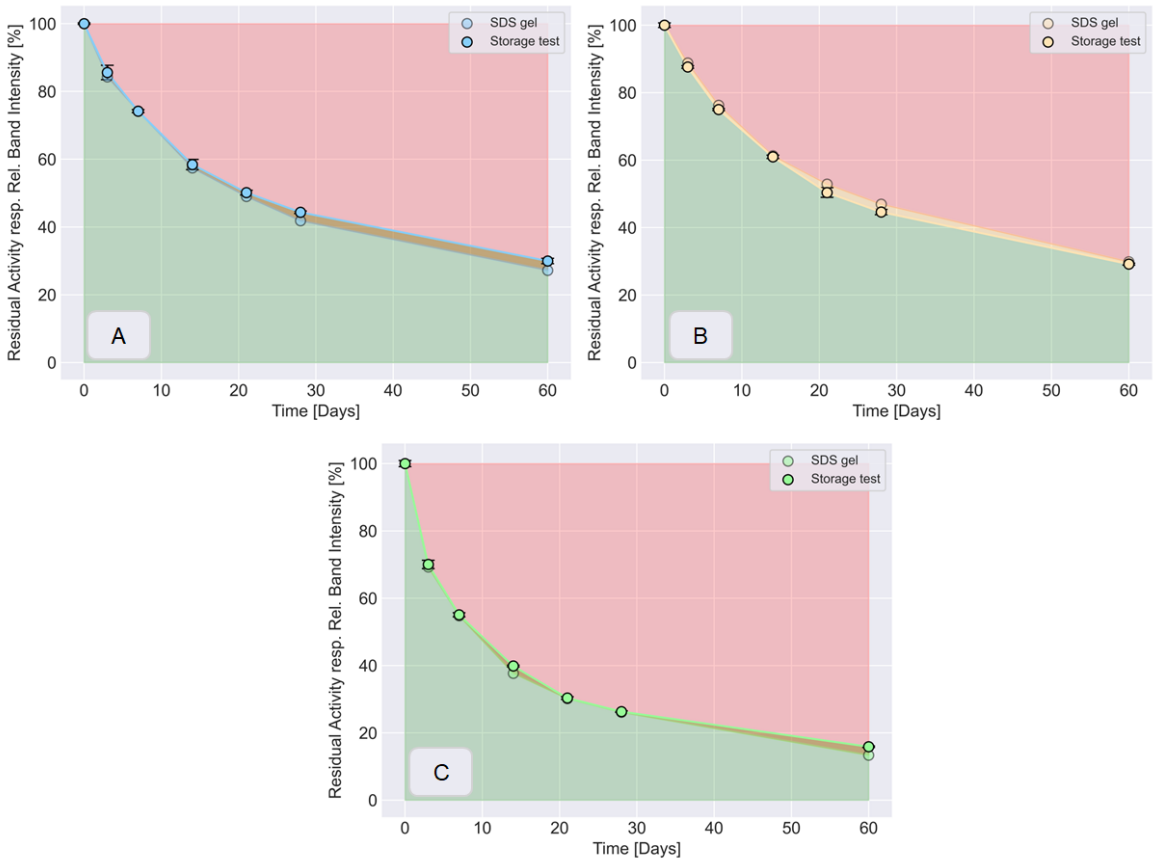


Figure 30: Residual enzyme activity [%] and relative full-length band (≈27 kDa) intensity [%] of subtilisin protease/10 wt% polyol formulations, stored at 37 °C over 60 days. *(A) Glycerol, (B) sorbitol, and (C) 1,2-Propanediol.*

and full-length intensity. Furthermore, the relative intensity loss shows to match the results from the residual activity measurements. Subsequently, the driving force for the enzyme

loss in these formulations, is assessed to be autolysis. However, as explained before, the result is not expected to resemble the actual ratio between thermal unfolding and autolysis. Rather, the thermal unfolding rate appears to be masked by a considerably higher autolysis rate, leading to a rapid degradation of already inactivated/unfolded subtilisin proteases.

Besides these results, some show significant differences in between the relative activity and intensity in presence of different polyol concentrations. Especially, formulations with 50 wt% polyol at 45 °C and 30 wt% at 37 °C provide additional insights on the ratio between activity loss and autolysis.

In general, formulations with 50 wt% polyol (*fig. 31*) have shown to exhibit the highest residual enzyme activity after 60 days, when compared to other concentrations. However, the relative full-length intensity showed to considerably differ from the relative activity results.

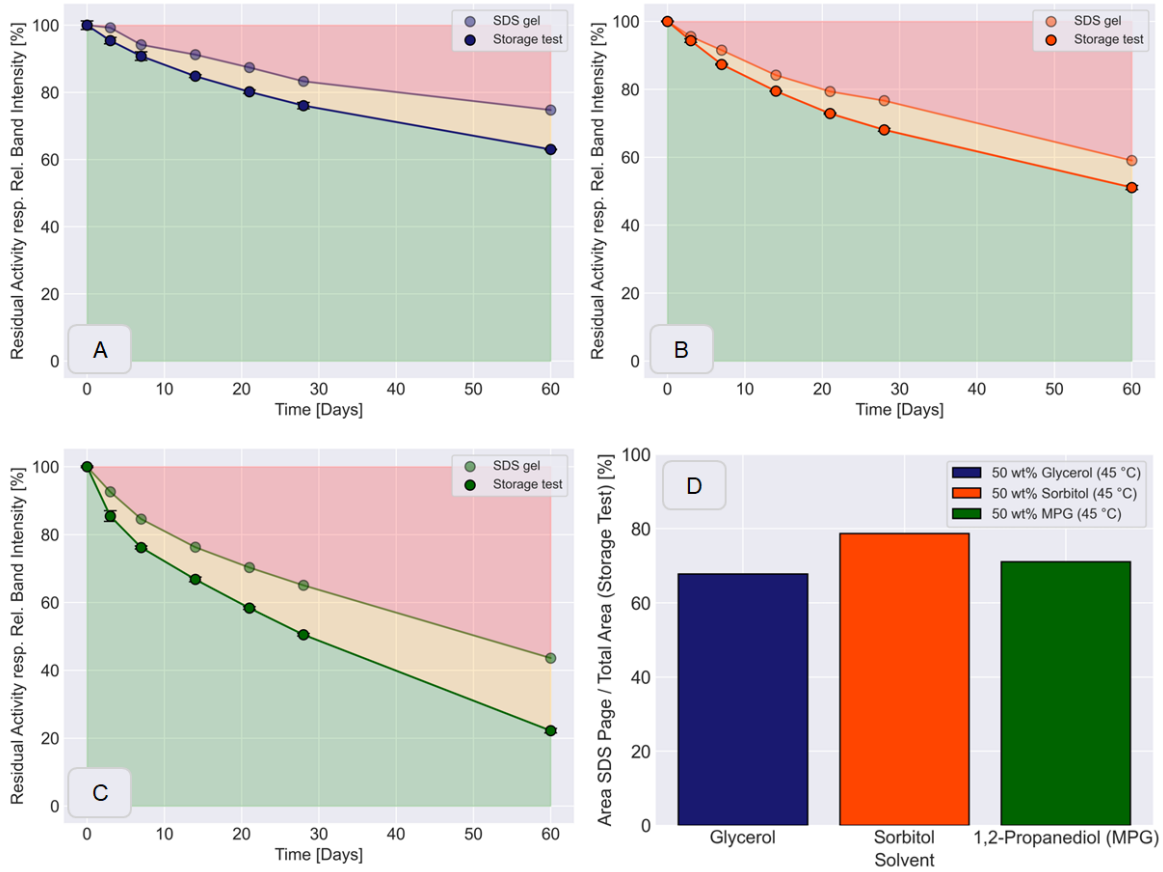


Figure 31: Residual enzyme activity [%] and relative full-length band (≈27 kDa) intensity [%] of subtilisin protease/50 wt% polyol formulations, stored at 45 °C over 60 days. *(A) Glycerol, (B) sorbitol, and (C) 1,2-Propanediol.*

Therefore, only a minor part of the overall activity loss, appears to be the result of autolysis. Subsequently, it can be assessed that the autolysis rate is either smaller or equal to the activity loss mediated by thermal unfolding. Furthermore, the proportion of the area described by autolysis appears to vary in between the different polyols. While in glycerol (*fig. 31D*), SDS-PAGE results appear to describe 67.8 % of the overall loss, values show to increase to 78.6 % for sorbitol and 71.1 % for MPG. Thereby, under these conditions', glycerol appears to be the strongest inhibitor of autolysis, followed by sorbitol and MPG. The high proportion of autolysis observed in the formulation with MPG, however, partly contradict the results obtained in *chapter 3.4.2*, where MPG has shown to be the by far strongest inhibitor of the subtilisin protease specificity and maximum reaction velocity. The destabilizing effect of high concentrations of MPG, however, could counteract the reduced activity by increasing the number of partially unfolded and thereby susceptible protease binding sites. Subsequently, the rate of autolysis appears to be increased, as a result of the partial destabilization of the subtilisin protease structure in the presence of high concentrations of MPG.

Results obtained in the presence of 30 wt% solvent at 37 °C (*fig. 32*), provide additional insights on the high similarity in between the storage test results obtained with sorbitol and MPG. As before glycerol (*fig. 32A*) has illustrated to retain the active enzyme content on the highest level. Sorbitol and MPG (*fig. 32B+C*) in turn, appear to be on the same level after 60 days, even though thermal stability results have shown to be considerably higher for sorbitol, when compared to MPG. Based, on the determined full-length intensity values, differences in between glycerol, sorbitol, and MPG become visible. While for sorbitol (*fig. 32B*) residual enzyme activities and full-length intensities illustrate a high similarity, glycerol and MPG show (*fig. 32A+C*) considerable differences in between residual enzyme activity and full-length intensity. Therefore, in the case of sorbitol 90.2 % (*fig. 32D*) of the overall loss appears to be explained by the relative intensity result, whereas for glycerol and MPG it shows to cover 71.6 % respectively 50.8 %. Subsequently, the inhibitory effect of MPG on the autolysis rate appears to be stronger, when compared to glycerol or sorbitol. At the same time however, MPG shows to be considerably more affected by thermal unfolding, as glycerol and sorbitol. Subsequently, glycerol which had shown to neither exhibit the highest inhibitory effect nor the highest thermal stability, illustrates the highest residual enzyme activity when compared to MPG or sorbitol, which are either strongly affected by thermal unfolding or their missing ability to reduce autolysis.

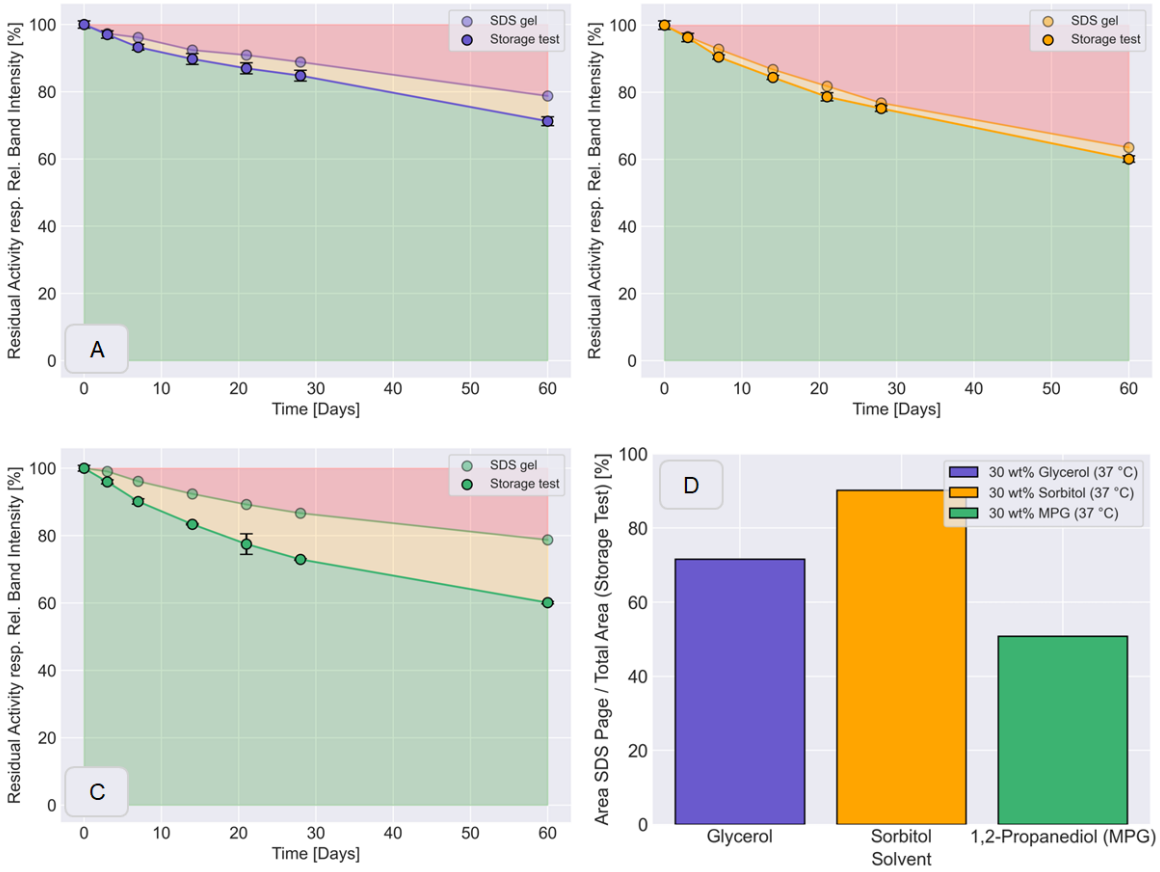


Figure 32: Residual enzyme activity [%] and relative full-length band (≈27 kDa) intensity [%] of subtilisin protease/30 wt% polyol formulations, stored at 37 °C over 60 days. *(A) Glycerol, (B) sorbitol, and (C) 1,2-Propanediol.*

Subsequently, the performed experiments have once again demonstrated the significant influence of different polyols on the proteolytic activity of the subtilisin protease. Furthermore, it was shown, that the observed long-term storage test results derive from a complex interplay of different detrimental factors. Even though, thermal unfolding and autolysis are expected to be the major cause of inactivation, additional factors are expected to contribute. Overall, the obtained results provided an explanation for the missing correlation in between residual enzyme activities and thermal stability. Furthermore, they clarified the need to consider additional factors, besides thermal stability, when trying to predict the outcome of long-term storage tests, conducted with the subtilisin protease. In summary, polyols have shown to affect the autolysis rate of the subtilisin protease, the mode of action and the exact rate, however, have jet to be determined.

3.4.4 The Effect of Polyols on Proteo-/Autolytic Activity of a Subtilisin Protease determined by MD Simulations

The catalytic activity of serine proteases is highly dependent on the interplay of a nucleophile, a general base, and an acid. In the case of the subtilisin family this catalytic triad is composed of serine, histidine, and an aspartic acid residue. The catalytic mechanism is initiated upon substrate binding. Subsequently, the catalytic serine residue in the enzymes active site undergoes a nucleophilic attack on the carbonyl carbon of the peptide bond, which results in the formation of a tetrahedral intermediate. The histidine residue stabilizes the intermediate by accepting a proton from the serine hydroxyl group, whereby a protonated histidine residue is formed.

In the past, it was shown that the proteolytic activity of an enzyme is influenced by the orientation of the serine residue and by the accessibility of the substrate. Radisky and Koshland demonstrated, that the activity of a subtilisin protease was greatly diminished in the presence of inhibitors, due to a change of the nucleophilic attack angle and by increasing the distance of the serine oxygen to the carbonyl carbon of the peptide bond [226]. Besides lipase b with engineered amidase activity had exhibited reduced activity when mutations affected the nucleophilic attack angle or the distance to the carbonyl carbon [227].

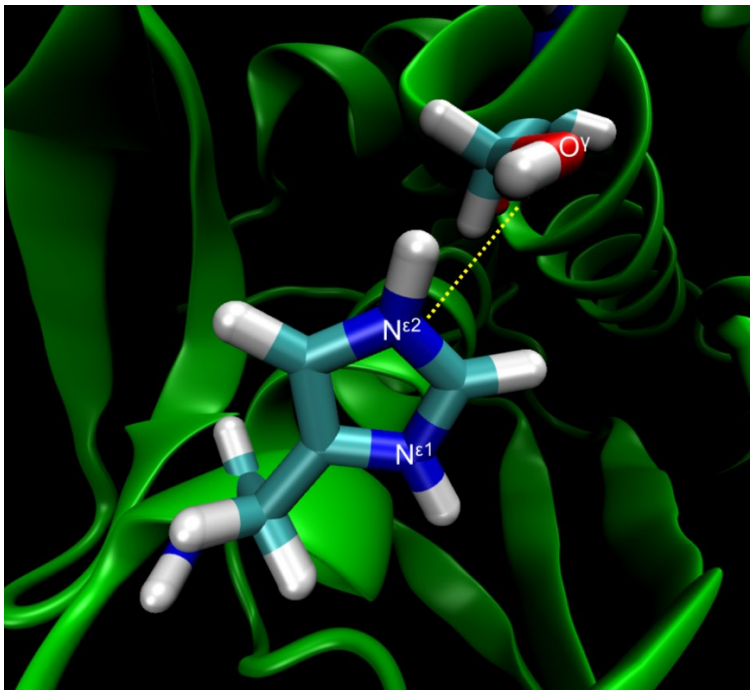


Figure 33: Distance between the catalytic serine oxygen (Ser O^{γ}) and histidine nitrogen (His $N^{\varepsilon 2}$) of the subtilisin protease

Further, the distance between the hydrogen donor and acceptor was shown to be an important factor that determines the catalytic nature of the active site. Katz *et al.* have shown, that in the presence of an inhibitor, the distance between the serine hydrogen attached to the oxygen and the nitrogen of histidine is increased [228]. The change in distance, resulted in a loss of catalytic activity of the enzyme. Subsequently, it can be assessed that the bonding distance between the donor and acceptor atoms of serine and histidine play a crucial role for the catalytic activity [229].

Based on these findings, the distance between the catalytic serine oxygen (Ser O^{γ}) and histidine nitrogen (His $N^{\varepsilon 2}$) (*fig. 33*) was set as an evaluation criterion to monitor the catalytic activity of the subtilisin protease in the presence of different alcohols/polyols in MD simulations. In this regard, trajectories obtained in *chapter 3.2.2* were used for further analysis. The distance between Ser O^{γ} and His $N^{\varepsilon 2}$ was determined using the calcium bound and free enzyme structure. Following, the relative proportion of the frames, exhibiting a distance >3 Å between both residues was determined for the subtilisin protease/alcohol/polyol systems. The results of the six independent repeats per solvent were averaged and used for comparison.

Calcium-Bound Structure

The results obtained for the enzyme structure with two bound calcium ions, show significant differences in between the different enzyme/solvent systems (*fig. 34*). While in the presence of long chained polyols, as sorbitol, xylitol, and erythritol, the distance between the Ser O^{γ} and His $N^{\varepsilon 2}$ appears to mainly fluctuate in between 3-6 Å, smaller, more hydrophobic alcohols/polyols show to increase the rate and amplitude of the observed fluctuations. Subsequently, the distance in some of the trajectories with glycerol, MPG, and ethylene glycol, appears to be increased throughout the simulation time of 200 ns. In this context, distances in between 6-10 Å and above can be observed.

When looking at the number of frames in which the distance between Ser O^{γ} and His $N^{\varepsilon 2}$ are >3 Å, the proportion (*fig. 35*) appears to be highly solvent dependent. While the distance appears to be larger in the presence of sorbitol in 63 % of all frames, the proportion increases to 87 % in the case of MPG. Furthermore, the proportion shows to increase based on the size/hydrophobicity of the polyols. In this context, formulations with xylitol (67 %), erythritol (71 %), glycerol (75 %), and ethylene glycol (81 %) show to exhibit proportions in between.

Methanol shows to be an exception, as the average result illustrates a distance larger than 3 Å in 74 % of the cases. However, the observed frequency and amplitude of the fluctuations appears to be different, when compared to rest. In this context, corresponding trajectories show to fluctuate in the range of 3-14 Å, however, more scarcely when compared to other solvents.

When looking for the reason of the drastic increase in the distance between the Ser O^{γ} and His $N^{\varepsilon 2}$ in subtilisin protease systems with glycerol, MPG, and ethylene glycol, a changed orientation of the catalytic histidine (H64) side chain could be identified (*fig. 36*). While both histidine nitrogen's appear to interact with the serine respectively aspartate in its native state, the histidine side chain appears to be flipped in the cause of some simulations. This change in orientation, shows to result in a loss of the interactions with the catalytic serine (Ser215) and later on with the aspartate (Asp32). Subsequently, the distance between the Ser O^{γ} and His $N^{\varepsilon 2}$ is significantly increased.

In summary, the trajectories of the subtilisin protease with two bound calcium ions could be used to identify a possible reason for the observed decrease of proteolytic respectively autolytic activity in chapter *3.4.2* and *3.4.3* in the presence of smaller/hydrophobic alcohols/polyols. The distance between the Ser O^{γ} and His $N^{\varepsilon 2}$ appears to be affected by alcohols/polyols present in the formulations. Subsequently, smaller/hydrophobic residues have shown to promote a flipping of the histidine side chain, significantly increasing the distance between the catalytic residues. This is expected to interfere with the proteolytic activity of the subtilisin protease and shows the ability of polyols to mediate changes in the subtilisin structure that can affect the enzymes catalytic activity. The mechanism, leading to the observed changes is, however, yet to be discovered.

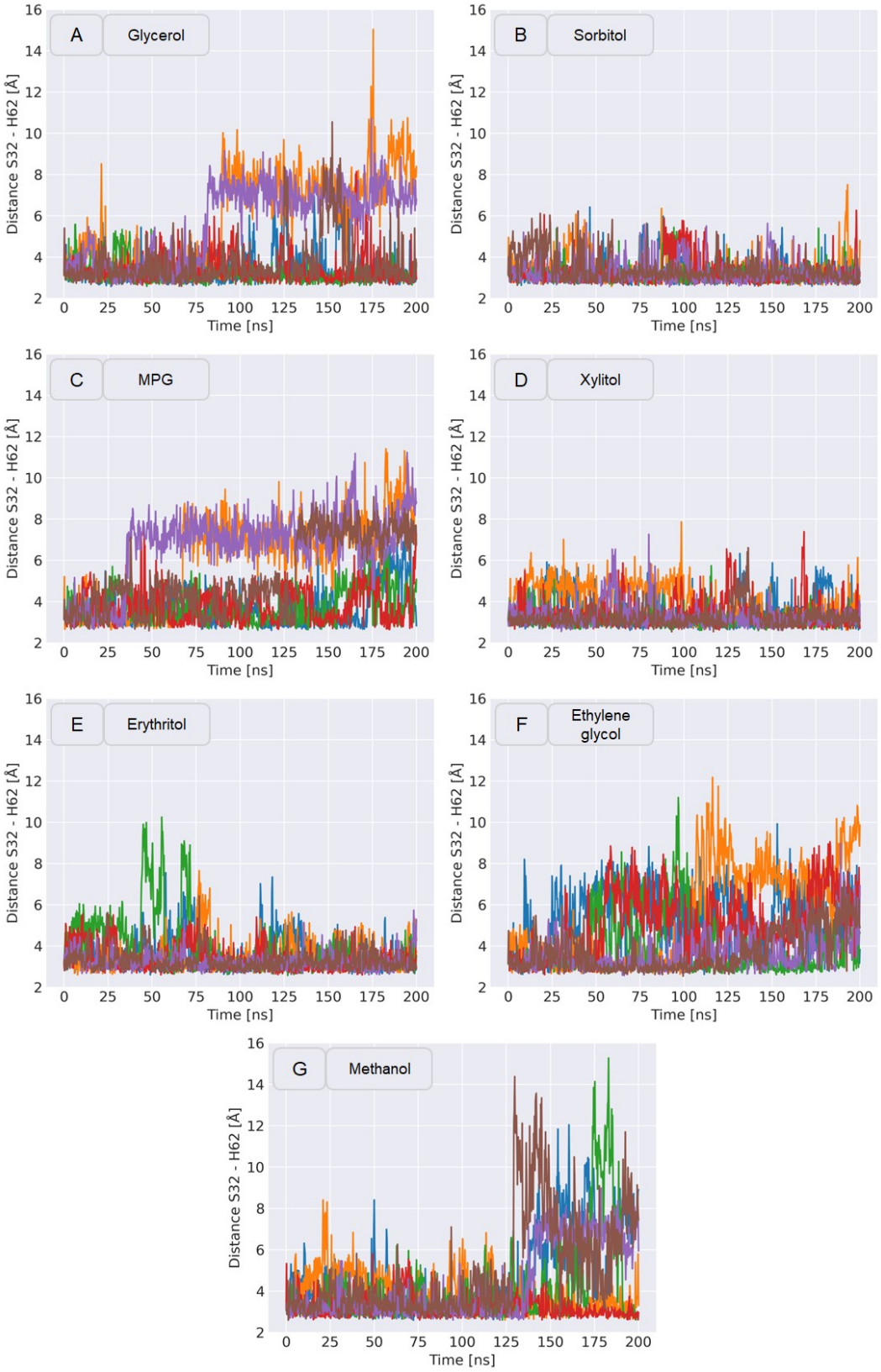


Figure 34: Distance between the Ser O^{γ} and His $N^{\varepsilon 2}$ atom in different subtilisin protease/alcohol/polyol systems with Ca^{2+} in and around the enzyme structure.

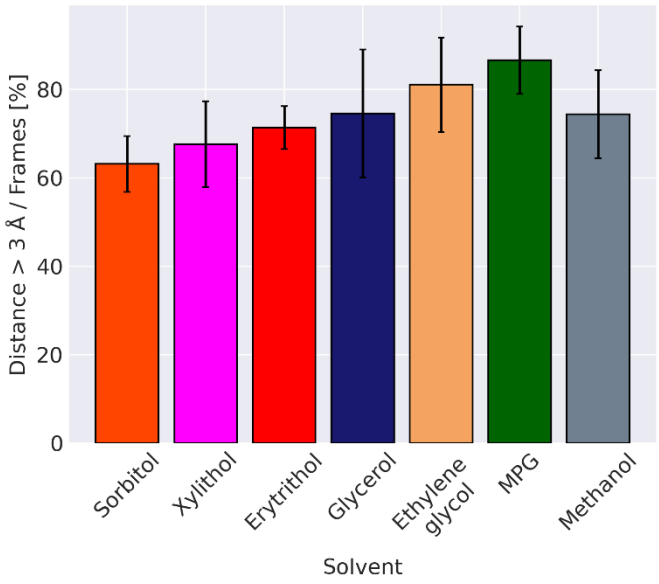


Figure 35: Relative number of frames [%], which exhibit a distance larger than >3 Å between Ser O^{γ} and His $N^{\varepsilon 2}$ in presence of Ca^{2+}in and around the enzyme structure.

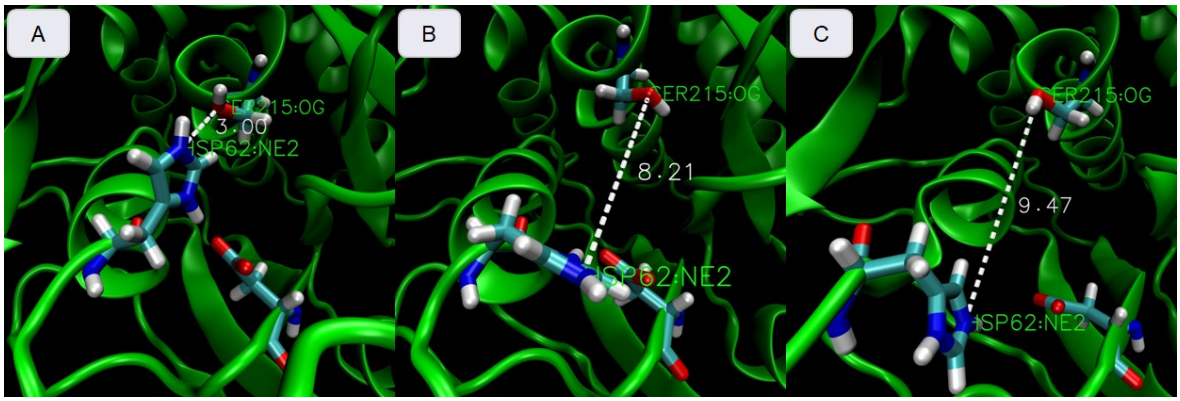


Figure 36: Change in the orientation of the histidine side chain, observed for simulations with glycerol, MPG, and ethylene glycol. *(A) Resembles the native active state, (B + C) illustrate the change in the histidine side chain orientation throughout the simulation.*

Calcium-Free Structure

The calcium-free structure of the subtilisin protease shows a totally different picture in the conducted analysis. In contrast to the results of the calcium occupied structure, the distance between the Ser O^{γ} and the His $N^{\varepsilon 2}$ shows to significantly fluctuate independent of the present solvent (*fig. 37*). Subsequently, in the majority of the cases the distance shows to vary in the range from 3-12 Å. In contrast, to the observed histidine flipping in the subtilisin structure with bound calcium ions, the flipping is observed to be non-permanent.

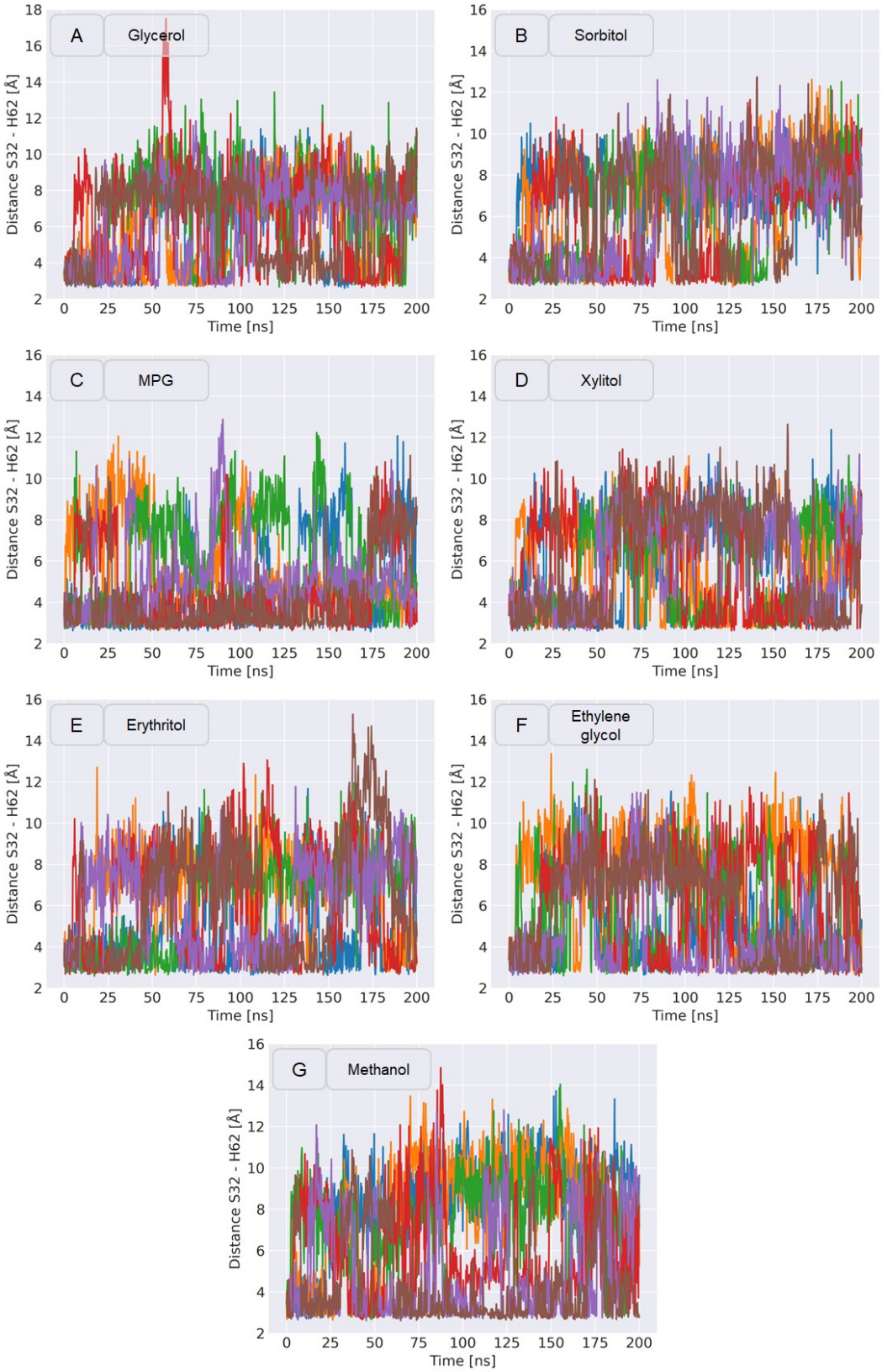


Figure 37: Distance between the Ser O^{γ} and His $N^{\varepsilon 2}$ atom in different subtilisin protease/alcohol/polyol systems in absence of Ca^{2+}.

Much more, the distance appears to fluctuate significantly in between the different states, which is represented by the increased amount of noise in the underlying data. Subsequently, the histidine side chain is observed to move erratically rather than targeted as in the presence of calcium.

As a result, the average relative proportion of frames exhibiting a distance >3 Å shows similar results (*fig. 38*) for the different subtilisin protease/alcohol/systems. Consequently, on average the distance appears to be out of bounds in 87 % of the frames over all systems. This is the same result obtained for MPG with regard to the calcium bound structure.

Subsequently, the results show that the distance between the Ser O^{γ} and His $N^{\varepsilon 2}$ appear to be significantly influenced in the absence of calcium in the subtilisin protease structure. However, the mechanism leading to the observed erratic change of distance between the Ser O^{γ} and His $N^{\varepsilon 2}$ is expected to be different. However, the effect on the proteolytic activity of the subtilisin protease could be the same. Therefore, the loss of both calcium ions is expected to similarly reduce the catalytic rate of the subtilisin protease as the polyols do in the active respectively calcium bound state.

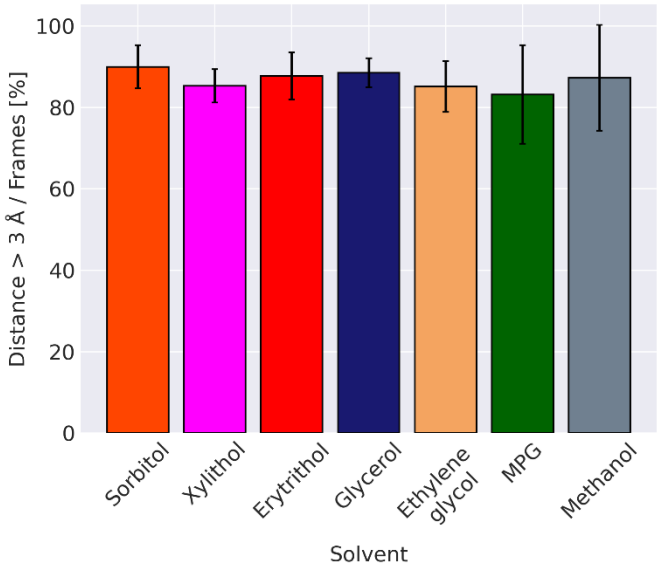


Figure 38: Relative number of frames [%], which exhibit a distance larger than >3 Å between Ser O^{γ} and His $N^{\varepsilon 2}$ in absence of Ca^{2+}in and around the enzyme structure.

3.4.5 Discussion

Subtilisins are one of the largest classes of serine proteases that are encoded in the genomes of all life forms, including viruses [230-234]. Bacterial subtilisins form the largest group and are normally associated with Gram-positive *Bacillus* or *Chlostridium* species [230, 235]. The majority have low substrate specificity and are secreted in the extracellular space to degrade proteins for defense and nutrition. Besides, less common intracellular forms show to be important for the degradation of intracellular proteins [54].

Proteases preferentially recognize and then hydrolyze peptide bonds, of polypeptide substrates, based on the amino acids preceding and/or following the cleavage site [236]. This specificity for the substrate is mediated based on favorable interactions of the substrate amino acid side chains with the residues that make up the binding site. The broad specificity of subtilisins is mediated by a large, hydrophilic binding pocket, with a preference for aromatic or large non-polar substrate residues [237, 238]. However, subtilisin are not specific for a particular hydrophobic amino acid. Subsequently, the subtilisin protease structure itself provides a broad spectrum of potential cleavage sites for the protease itself. Thus, a high proportion of peptide fragments with varying length and quantity could be determined in the SEC and corresponding SDS-PAGE. Furthermore, released peptide chains are expected to partially retain secondary and tertiary structure motifs, which leads to the observed enthalpy changes during DSC.

The process of substrate binding is defined by the active site and several smaller binding pockets (S), that interact with side chains of the substrate protein/peptide residues (P). The target bond for hydrolysis, is located in the center. Substrate residues upstream (N-terminal) are termed P_1-P_n, residues downstream (C-terminal) are referred to as P_1'-P_n'. While aspartate and histidine in the catalytic triad and the S_2, S_3, and S_5 sub-sites are primarily involved in binding the N-terminal domain, serine and S_1, S_4, S_6 are linked to the fixation of the C-terminal domain. S_1 of subtilisin proteases, forms a large hydrophobic pocket accommodating non-polar residues located at P_1, while S_2 is in general a smaller pocket that accommodates smaller residues as alanine and proline. Consequently, the substrate accumulates in between the catalytic residues, bridging the interface and initiating the catalytic reaction [238].

The used Suc-AAPF-pNA has shown to be a good substrate for many proteases in terms of K_m/k_{cat}. In its native state, AAPF is a white solid, which absorbs UV light at a wavelength of 315 nm. When cleaved, the pNA moetiy is released, which leads to increased

absorbance in the range of 405 nm. The low K_m for most proteases is mediated by the comprised peptide chain. In this regard, the phenylalanine with its high hydrophobicity and the subsequent proline have shown to provide a high affinity with the P_1 and P_2 position in the active site of subtilisin proteases. The K_m and V_{max} in the conducted experiments had shown to significantly increase respectively decrease in the presence of different polyols. In this regard, proposed inhibition mechanism ranged from competitive, to uncompetitive, and non-competitive dependent on solvent/-concentration. In general, classical protease inhibitors interact with the enzyme structure in two major ways [239]. The first, is the irreversible trapping reaction, which is mediated by protease inhibitors of the serpine-, α2-macroglobuline-, and balcovirus protein p35 inhibitor families [239]. These inhibitors are cleaved when interacting with the active site of proteases, triggering an irreversible conformational change. Subsequently, the active site is blocked and not further susceptible to substrate binding. The other mechanism observed, are protease inhibitors which act over a tight-binding reaction. These inhibitors have unique properties, mediated by an ultrahigh affinity, fast association to the target enzyme combined with extremely slow dissociation of the from the enzymes active site.

Since polyols are structurally quite diverse to common protease inhibitors and up to today no such mechanism was proposed, structural changes mediated in the presence of polyols appear more likely to affect the proteolytic rate. This is also reflected by the various proposed inhibition mechanisms, which are known to be the result of interactions with protease domains, other than the active site. Furthermore, as discussed in *chapter 3.2.5*, polyols can have a considerable influence on the protein structure. While large/hydrophilic polyols tend to mediate a more compact folding of the enzyme structure, smaller/hydrophobic polyols have shown to promote an increase of the surface area, diameter, and loss of native contacts. Subsequently, flexibility and fluctuations of the enzyme structure are to be increased, which will further comprise defined domains and residues.

The observed change in the distance of the Ser O^{γ} and His $N^{\varepsilon 2}$ in the presence of smaller/hydrophilic polyols, is expected to be the result of such changes. However, the specific reason respectively responsible mechanism for the observed histidine flipping has yet to be determined. Nonetheless, an increased distance between both residues is assumed to have a considerable effect on the proteolytic activity of the subtilisin protease and may thereby affect the enzyme long-term stability in a liquid formulation environment [226-229]. In general, the conformation of the active site and binding pocket is highly conserved

and stable. The bound substrate, therefore, adapts to the given structure, which mediates the broad specificity of subtilisin proteases. When, however, the interface of the active site is destabilized, the flexibility of this region will increase, leading to an adaption of the enzyme to the substrate. Subsequently, the position of certain residues will be changed, leading to reduced or no catalytic activity.

Reduced proteolytic activity, however, is believed to be just one essential component, to achieve an increased long-term stability of the subtilisin protease in liquid formulations. In the past, high long-term stability was shown to result from a complex interplay of thermal stability and proteolysis. In this context, it was shown that many proteins in their native compact conformation appear to be resistant to degradation [240-242]. This ability, is however, rapidly lost upon partial unfolding of the protein structure, leading to rapid degradation. Even though, high concentrations of MPG have illustrated to significantly reduce the proteolytic activity of the subtilisins protease, this effect appears to be counteracted by the observed destabilization of the enzyme structure. Nonetheless, proteases with a broad specificity can still attack proteins in their native conformation, especially if they can contain accessible loop regions that are flexible enough for cleavage. Subsequently, sorbitol appears to be no better than MPG at retaining the active state of the subtilisin protease, due to its lack of reducing the autolysis rate. Consequently, a middle way of both, represented by glycerol, showed to lead to the best result. Greatly retaining the native compact structure of the subtilisin protease, while at the same time reducing the impact of autolysis on the long-term stability.

3.5 Summary & Conclusion

The stability of enzymes in liquid formulations is commonly monitored by determining the residual enzyme concentration after storage. Thereby, the effect of different additives on the long-term stability is to be assessed under conditions similar to the later application. However, based on the degree of stabilization and the native stability of the enzyme, this can take days, weeks, months or even years. Due to the rapid availability of the results, DSC measurements are regarded an alternative for testing the stabilizing effect of various additives on the basis of thermal stability. Unfortunately, depending on the enzyme, and especially for proteases, the results are rarely transferable to the conducted long-term stability studies.

This lack of correlation can also be observed in the course of the experiments carried out with a subtilisin protease. In this context, DSC results appear to significantly deviate from the long-term storage test results. Subsequently, it could be shown that this methodology is only partially suitable for making predictions about the long-term stability of enzymes in liquid formulations. Moreover, results would have led to the false assumption that sorbitol provides a better long-term stability than glycerol or MPG. Despite the fact that glycerol achieved a significantly better result at moderate temperatures.

The results, thereby highlight the need to consider other factors besides thermal stability when predicting the stabilizing effect of different additives. In this context, the calcium binding state was found to be insignificant for the observed differences between thermal stability and residual enzyme activity results. Conversely, a considerable influence of glycerol, sorbitol, and MPG on the proteolytic activity of the subtilisin protease could be demonstrated. In addition, the influence of the polyols on autolysis and subsequent interaction between thermal stability and autolysis could be shown. While the use of sorbitol was associated with a significant increase in thermal stability, MPG proved to be a potent inhibitor of autolysis. However, the lack of effect on the opposite mechanism led to a comparable long-term stability in the presence of both solvents. Consequently, the best long-term stability could be achieved with glycerol, which had shown to increase thermal stability and to limit degradation by autolysis. Therefore, the use of polyol combinations appears to be particularly promising in order to minimize the influence of both factors on the long-term stability as much as possible.

In addition, the use of MD simulations provided further insights, that are a useful complements/substitute to the DSC data. Not only do the results obtained allow prediction of the thermal stability based on different molecular descriptors, but also the cause and correlation of the polyol size/hydrophilicity/-phobicity with the reduced proteolytic activity could be determined. However, the exact mechanism causing the histidine flipping, as well as possible correlations with the proteolytic/autolysis rate, will be part of future investigations. Furthermore, these results could be of interest in the scope of protein engineering to specifically induce the inactivated state of the subtilisin protease in the storage medium. Upon actual use, and the associated dilution of the formulation environment, the effect should be reversed, and the full activity restored. This would open the door to inhibitor free protease formulations, which would significantly reduce costs and increase the long-term stability of proteases in liquid formulations. Nevertheless, the transferability of the results to other proteases/enzymes and additives has yet to be verified.

In summary, the results have highlighted the complex relationship between several factors on the long-term stability of a subtilisin protease. This emphasis the need to take other factors into account, when predicting storage stability of enzyme in liquid formulations. Particularly, in the case of proteases, it is necessary to consider not only the influence of the solvent on thermal stability, but also the influence on activity. Due to the incredible diversity of enzymes and the associated structures and catalytic mechanisms, it remains to be seen whether the applicability is limited to a single, a few, or a group of enzymes.

3.6 Mid Infrared Spectroscopy for Enzyme and Solvent Quantification in Liquid Enzyme Formulations

Photometric assays are widely used for the quantification of enzymes in biochemical research and clinical diagnostics. These assays rely on measuring the absorbance or fluorescence of a product or substrate that is released as part of an enzymatic reaction. The amount of product or substrate consumed or produced is in many cases directly proportional to enzyme activity. This in turn, allows the quantification of the enzyme concentration.

While photometric assays for enzyme quantification are commonly used, they do have several limitations. The presence of interfering substances in the sample, as salts, detergents, or other proteins, can negatively affect the accuracy of the assay by binding to assay reagents or altering the enzyme activity. Further, the range of enzyme activity that can be measured is limited, whereby high enzyme concentrations often require dilutions, to stay within the linear range of the assay. This can result in loss of accuracy at higher concentrations or underestimation of enzyme activity at lower concentrations. Besides, depending on the enzyme being studied and the specific assay used, the cost of substrates and buffers can greatly vary and be a significant expense.

Following, several experiments were performed to evaluate the use of mid-infrared spectroscopy for enzyme and solvent quantification in varying enzyme/polyol formulation environments. In this regard spectra of the formulations were acquired with the mid-infrared analyzer (MIRA) and evaluated with multivariate statistics. For this, different formulations with varying concentrations of 1,2-propanediol (MPG) and a subtilisin protease respectively endocellulase were formulated and analyzed. As a result, different principal component regression (PCR) and partial least squares (PLS) models for the prediction of the enzyme and solvent concentration of different formulations were generated and tested.

3.6.1 Subtilisin Protease/MPG Formulations for Solvent Concentration Prediction

First, the capability of the MIRA technology in combination with multivariate analysis, to quantify the solvent concentration of different enzyme formulations containing the subtilisin protease was determined. Based on a design of experiment (DoE) (*tab. 7*),

Table 7: DoE setup used to create the PCR and PLS model for enzyme and solvent concentration prediction of different subtilisin protease/MPG formulations.

Run Order	Design Order	Sample Name	Solvent [weight%]	Enzyme Conc. [g/L]
1	27	MI-14	50	38.75
2	32	MI-15	30	27.5
3	4	MI-16	10	16.25
4	31	MI-17	30	27.5
5	20	MI-18	30	50
6	22	MI-19	50	5
7	19	MI-20	30	50
8	16	MI-21	30	27.5
9	28	MI-22	50	38.75
10	21	MI-23	50	5
11	17	MI-24	30	38.75
12	26	MI-25	50	27.5
13	33	MI-26	30	27.5
14	3	MI-27	10	16.25
15	1	MI-28	10	5
16	12	MI-29	30	5
17	18	MI-30	30	38.75
18	34	MI-31	30	27.5
19	6	MI-32	10	27.5
20	10	MI-33	10	50
21	25	MI-34	50	27.5
22	7	MI-35	10	38.75
23	11	MI-36	30	5
24	8	MI-37	10	38.75
25	29	MI-38	50	50
26	30	MI-39	50	50
27	5	MI-40	10	27.5
28	14	MI-41	30	16.25
29	23	MI-42	50	16.25
30	24	MI-43	50	16.25
31	2	MI-44	10	5
32	13	MI-45	30	16.25
33	15	MI-46	30	27.5
34	9	MI-47	10	50

different formulations with three MPG and five enzyme concentrations were prepared (n= 34). Solvent concentrations were calculated based on the weight (wt%) of the solvent added to the formulations. Following, spectra of all samples were acquired using MIRA and the obtained spectra were analyzed by multivariate data analysis. MIRA spectra consisted of 1101 variables in a range of 929-3051 cm^{-1}.

The spectra (*fig. 39*) show significant differences based on enzyme and solvent concentration. Especially, in the range from 1500-1750 cm^{-1} and 2500-3000 cm^{-1} an organization of the spectra according to the three underlying solvent concentrations is visible. In general, higher concentrations of MPG seem to be associated with increased absorption, showing the distribution of the solvent bands in all MIRA spectra. Besides, a smaller variance of the three groups, especially in the range of 1500-1750 cm^{-1} can be observed. The visible variance into five-line shapes, in this region, suggests a connection with the enzyme concentration in the samples.

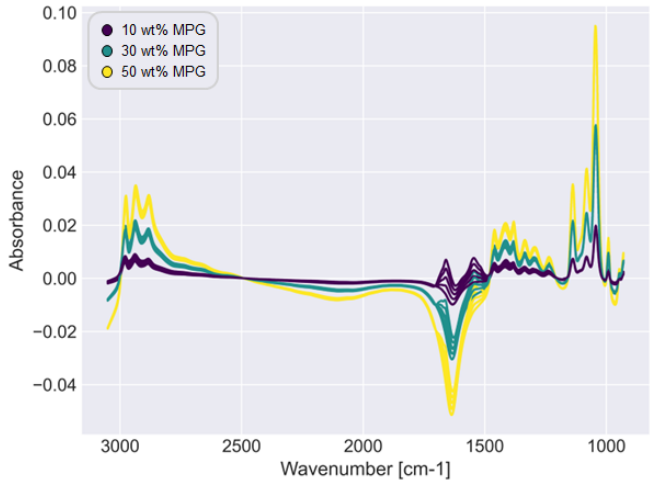


Figure 39: Spectra of different subtilisin protease formulations with 1,2-propanediol (MPG) based on the performed DoE.

Preprocessing

Prior to the statistical analyses and model creation, MIRA-spectra were preprocessed by applying the Savitzky-Golay (SavGol) smoothing and differentiation filter (*fig. 40A*). A SavGol filter is commonly used to reduce the noise and baseline variations in spectra, while preserving the spectral features. This achieved by fitting a polynomial function to a small segment of the signal, and then using the coefficients of the polynomial to estimate

the smoothened value of the signal at the center of the window. The process is repeated for each point in the signal, resulting in a smoothened signal that approximates the underlying trend of the data. In this context, a SavGol-filter using a 5th order polynomial and 2nd order derivative were applied.

Further, spectral data was auto scaled (*fig. 40B*). Thereby, the variability in the spectral data, caused by regions with higher intensity than others, was to be reduced. The purpose of this normalization is to remove any systematic variations in the data and ensure that the data is on the same scale giving the same weight to all variables in the MIRA spectra, making it easier to compare and analyze different features in the dataset. Autoscaling transforms the data by subtracting the mean value of each feature from the original values and then dividing it by the standard deviation of the feature. The mean of each feature is thereby zero, and the standard deviation is one, which mean that the transformed data is centered around zero and has a similar scale for each feature.

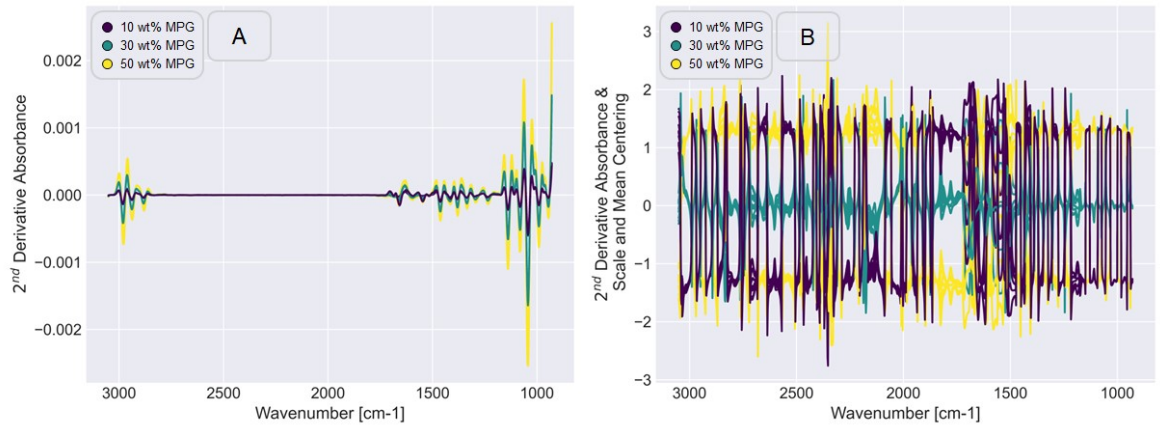


Figure 40: Preprocessing of MIRA-spectra with varying subtilisin protease/MPG formulations. *(A) Savitzky-Golay smoothening and differentiation (5th order polynomial;2nd derivative); (B) Autoscaling.*

Hierarchical Clustering

Hierarchical cluster analysis is a multivariate data analysis technique used to group similar observations of variables into clusters. It is therefore suitable to provide a first overview about the data similarities and differences. In the underlying analysis the mathematical distance between the variables was determined based on the Euclidian distance algorithm. It determines the distance between all points in Euclidian space, with a fixed number of dimensions. The algorithm then calculates the distance between all points by taking the

square root of the sum of the square differences between their corresponding features in each dimension. The Euclidean distance algorithm thereby defines how the clusters are formed, while the linkage method specifies the rule for how the distance between the clusters is calculated.

In the present analysis, the linkage method by Joe H. Ward Jr. [243], was used, which minimizes the variance within each cluster. This is achieved by treating each data point as its own cluster. Subsequently, the algorithm iteratively merges the two closest clusters together, which will result in the smallest increase in total variance within the resulting cluster. The distance between two clusters is defined as the increase in the sum of squared distances between the data points in the merged cluster and their respective cluster centroids. The result is represented visually as a dendrogram.

The dendrogram of the analyzed samples (*fig. 41*) exhibits three major clusters, which can be attributed to the three different solvent concentrations of the DoE samples with 10, 30, and 50 wt% MPG. While the purple cluster represents the samples with 10 wt% MPG, the blue/green cluster is composed of samples with 30 wt% MPG and the yellow cluster of samples with 50 wt% solvent.

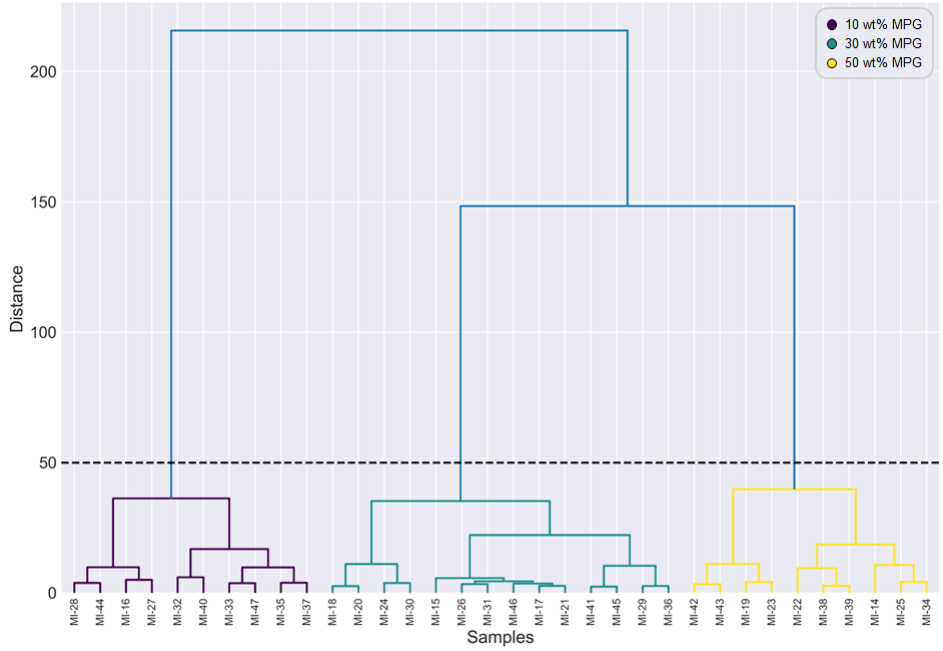


Figure 41: Hierarchical cluster analysis of preprocessed MIRA spectra of different subtilisin protease/MPG formulations. *The three colored clusters are composed of samples with a similar MPG concentration: purple= 10 wt% MPG, blue/green= 30 wt% MPG, and yellow= 50 wt% MPG. Subcluster are composed of samples with a similar enzyme concentration.*

Furthermore, additional subclusters are visible, which represent the different enzyme concentrations. In general, the three clusters with different solvent concentrations are composed of additional subclusters comprising samples with similar enzyme concentration.

Hierarchical clustering is often used prior to more sophisticated multivariate statistical approaches to identify structures and patterns in the underlying data. In this regard, the obtained dendrogram illustrates the information contained within the spectra, retrieved based on the mathematical distance and linkage method used. The dendrogram illustrates the differences between the samples, resulting from different solvent concentrations, but also based on the inherent enzyme concentration. Following approaches, were aimed at extracting this information and using it to predict enzyme and solvent concentration based on the variability of this properties contained in the spectra.

Principal Component Analysis (PCA)

The hierarchical clustering analysis had identified different clusters based on the solvent concentration of the samples. Besides, smaller subclusters were visible which contained samples with similar enzyme concentrations. To reduce the dimensionality of the data and identify the origin of the variability in the spectra, principal component analysis (PCA) was applied to the spectral data. The goal of PCA is to find a new set of orthogonal variables, principal components (PCs), which are linear combinations of the original variables and capture the majority of the variance in the underlying data.

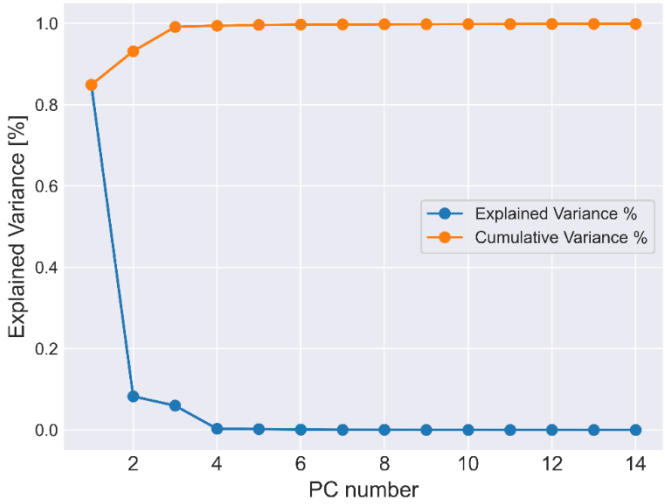


Figure 42: PCA scree plot with the explained and cumulative variance of the underlying dataset

The determined scree plot (*fig. 42*) illustrates the percentage of variance in the original dataset, which is explained by each of the PCs. The explained variance of a PC represents the amount of variability in the original data that can be attributed to that component. For the given data, PC1 shows to explain 84.91 % and PC2 8.25 % of the variance of the original data set. Already with 3 PCs, the PCA model shows to capture more than 99 % of the variance. The high proportion of PC1 on the total variability, suggests that the obtained spectra have a strong underlying structure or pattern assigned to this component.

The PCA scores plot *(fig. 43)*, further adds to this observation by visualizing the relationship of the variability of the samples according to the respective principal components. Each data point represents a sample of the data, which is plotted based on its score on the selected PC. Based on the values of the PC1 scores it can be assessed that the variance captured, relates to the solvent concentration. This is illustrated by the fact, that samples with a similar MPG concentration have the same scores values for PC1. Further, PC2 illustrates a tendency to separate the samples according to the present enzyme concentration. Consequently, datapoints with a similar solvent concentration were subdivided in five groups according to PC2. However, the five groups illustrate a varying spread and shift in between the different concentrations. Thereby, it can be assumed, that the variability of PC2 is not due to the enzyme concentration alone.

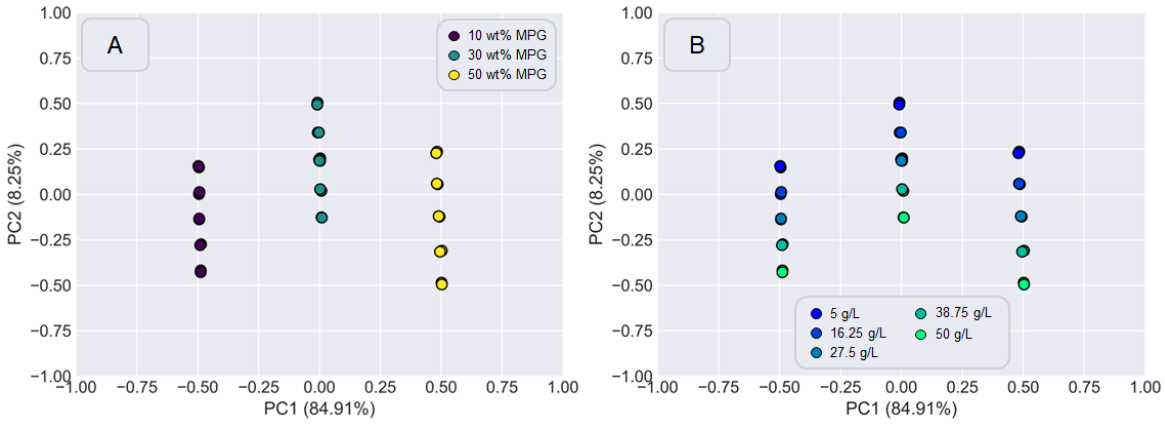


Figure 43: PCA scores plot for the first two PCs colored according to the (A) solvent and (B) enzyme concentration of the subtilisin protease/polyol formulations.

To understand the cause for this separation, the loadings plot (*fig. 44*) can provide additional information. A PCA loadings plot, can be visualized, to illustrate the contribution of different variables to the principal components in a data set. It shows the

correlation between variables and PCs, allowing one to identify which variables influence the components the most. For better visualization, the mean IR spectra was colored according to the loading's values of PC1 and PC2. These plots, illustrate the contribution of the variables to the separation of the samples and their role during PCA model creation. Negative scores values correlate with negative loadings values (*fig. 44; red*) and positive scores values correlate with positive loading values (*fig. 44; blue*). In summary, a negative scores value indicates that the variable is less likely to be associated with the PC, than a variable with a high loadings value.

Since high positive and negative loadings indicate a positive respectively negative correlation between a variable and a principal component, the contribution of the solvent to the overall variability seems to be present in the whole spectrum, as seem by initial assumptions. In addition, the high magnitude of the loadings suggests a strong relationship of these variables with PC1. Nevertheless, some of the variables, especially with a wavelength in the region from 1500-1750 cm^{-1} show no contribution at all to the first PC. For protein samples this region is of special interest, as it comprises the amide I and II bands, which are the major signals of the protein infrared spectrum [244]. In the case of PC2, especially these regions show a strong contribution to the overall variability, showing a high positive and negative contribution.

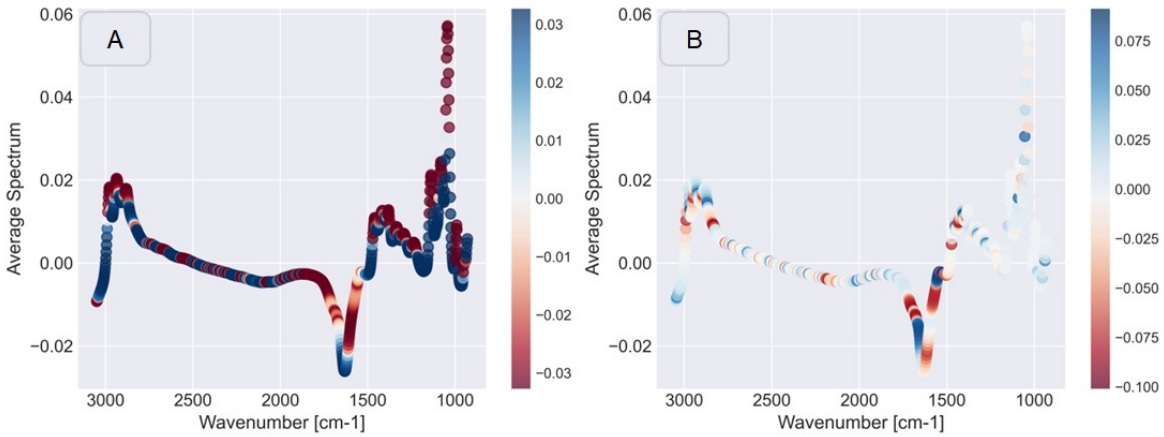


Figure 44: Mean IR spectra colored according to the PCA loadings values. (A) PC1 and (B) PC2.

Even though, this fact could lead to the assumption that the variability encoded by PC2 is due to the different enzyme concentrations, it could be seen that several other variables spread over the spectrum are comprised by PC2. Furthermore, several of these regions are

adjacent or overlapping with regions, which show a high correlation with PC1. Therefore, from a chemical perspective, it could be assumed that PC2 does not only cover the variability caused by different enzyme concentrations, but also by other substances present in the formulations.

Principle Component Regression (PCR) for Solvent Concentration Prediction

Since the solvent concentration shows to be strongly correlated with the overall variability of PC1, a model for the prediction of the solvent concentration in different subtilisin protease formulations was developed using principal component regression (PCR). The goal of PCR is to predict a response variable based on an unsupervised method and a set of predictor variables by reducing the dimensionality of the predictor variables through PCA and then using the principal components as predictors in a regression model. Thereby, multicollinearity and noise, in the underlying data, are to be prevented respectively reduced. The use of PCA transformation, makes the PCR an unsupervised approach, meaning that it does not consider the target variable when determining the PCs.

In order to train the model, the DoE dataset was split in training (n= 23) and test dataset (n= 11). Following a PCR model was to be created for solvent concentration prediction, based on the MIRA-spectra. Since PC1 has shown to describe the variability caused by the three different solvent concentrations, the model was created using PC1 only. The PCR model (*fig. 45*), based on the training dataset, exhibits a high accuracy with regard to the obtained calibration (R^2= 0.999; RMSE= 0.24; bias= $1.0*10^{-15}$) and cross validation (R^2= 0.999; RMSE= 0.30; bias= 0.0002) results. Furthermore, the prediction of the “unknow” solvent concentrations of the test data, returns results with a similarly high accuracy (R^2= 0.998; RMSE= 0.42; bias= 0.11).

The predicted results based on the created PCR model demonstrate a high accuracy for the prediction of the solvent concentration of different subtilisin protease/MPG formulations. Furthermore, it could be demonstrated that the solvent concentration of different samples can be predicted based on a mid-infrared spectrum using an unsupervised PCR model. The high quality of the MIRA spectra shows to be suitable for model creation, returning highly accurate values for validation and cross-validation.

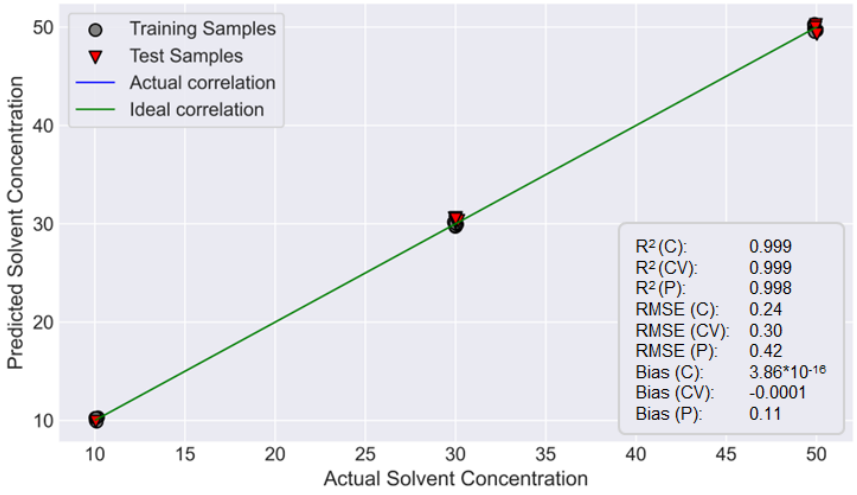


Figure 45: Calibration, cross-validation, and prediction results for the generated PCR model for solvent concentration prediction of different subtilisin protease/MPG formulations.

Partial Least Squares (PLS) Analysis for Solvent Concentration Prediction

Besides using a PCR model to predict the solvent concentration in different subtilisin protease/MPG formulations a partial least squares model (PLS) was generated. In contrast to the unsupervised PCR, PLS is trained on a set of predictor variables (X-variables) and corresponding response variables (Y-variables), which makes it a supervised approach. Following, the PLS model can be used to predict the value of the response variables for unknown samples based on their predictor variable values. The goal of PLS is to find a set of latent variables, so called components, that explain the maximum amount of variance in both the predictor and response variable. These components are found by projecting the data onto a new set of orthogonal axes, which derive from linear combinations of the original predictor and response variable.

In order to determine a reasonable number of PLS components for the prediction of the solvent concentration, the root mean square error (*fig. 46*) of the cross validation is used. For the conducted experiment, one PLS component, shows to be sufficient to obtain accurate results. The cross-validation exhibits an RMSE 0.25 with the first PLS component for the used test dataset.

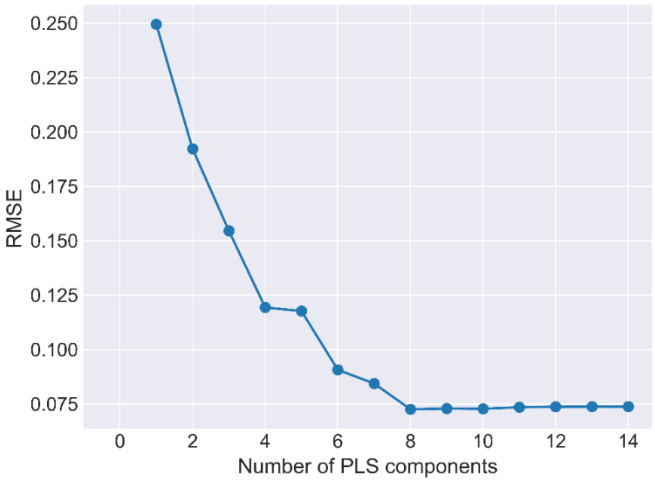


Figure 46: Root mean square error (RMSE) of the cross-validation results with increasing number of PLS components for the solvent concentration prediction of different subtilisin protease/MPG formulations.

The X-scores plot (*fig. 47*) for the first two PLS components illustrates a similar picture as the scores from PCA/PCR. The X-scores for the first PLS component are grouped according to the MPG concentration of the samples. Similar, the second PLS component shows a grouping of the samples according to the five different enzyme concentrations.

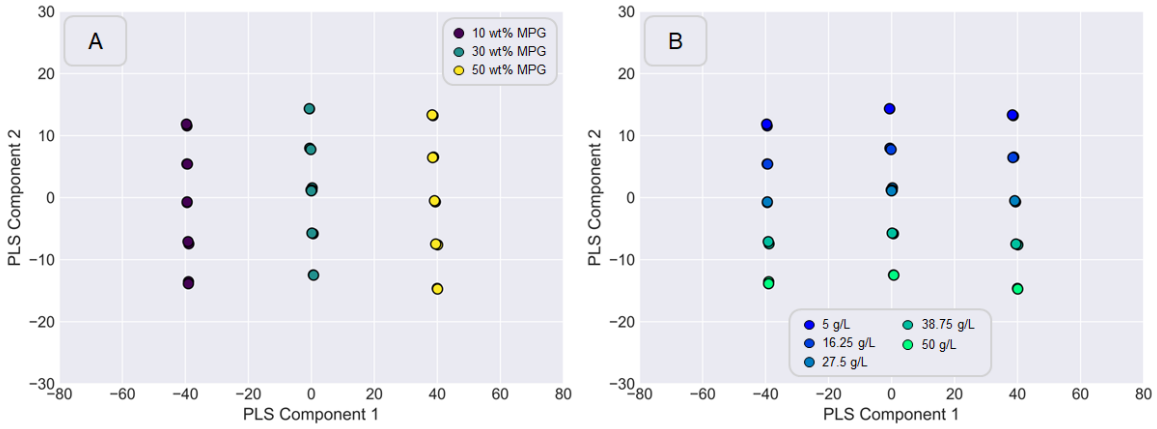


Figure 47: PLS component 1 x PLS component 2 plot of the PLS model for the whole DoE dataset. *Samples are colored according to (A) the MPG and (B) enzyme concentration present in the samples according to the DoE.*

For the determined PLS model, the variable of importance in projection scores (VIPs) were determined to assess the importance of the X-variables in explaining the variation in the Y-variable. The VIP score for each X-variable is calculated by summing the squares of the product of the loadings and the amount of variance explained by each component.

Thereby, the VIPs can be used to rank the X-variables in order of importance, with higher VIP scores indicating greater importance. Subsequently, these variables can be used to further optimize the model.

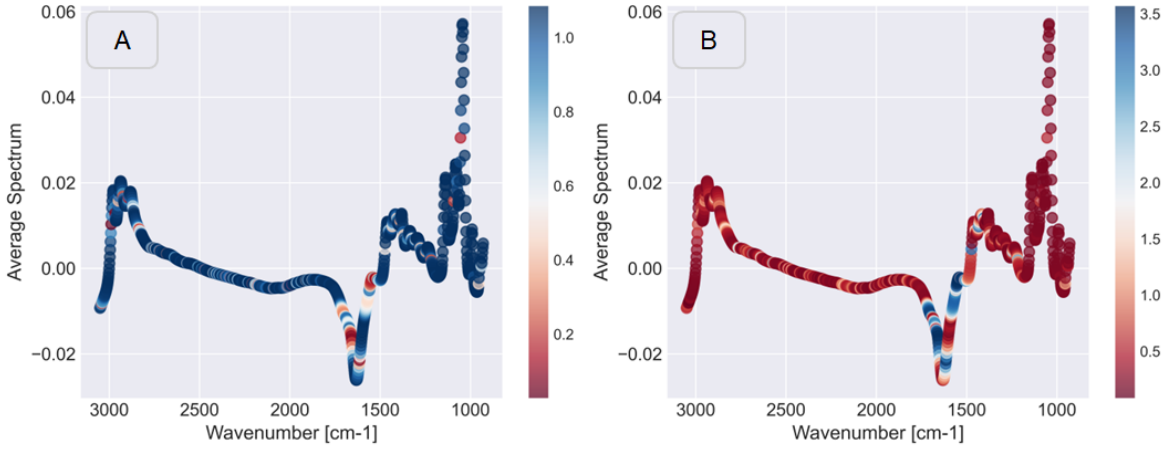


Figure 48: VIPs for the created PLS model to predict the (A) solvent concentration of different subtilisin protease/MPG formulations. *In comparison the VIPs gaining relevance when predicting (B) the enzyme concentration.*

In the shown case (*fig. 48*), VIPs were calculated, and the average IR spectrum of the DoE was colored according to its values. The VIPs for the solvent concentration prediction are distributed over the whole spectrum and show a high level of importance of almost all variables. As before, minor important regions are scarcely scattered along the spectrum and seemed to accumulate in the region from 1500-1750 cm^{-1}. This region is known to comprise MIR relevant signals for proteins as the amide I and II band [171]. This region gains significant importance when exchanging the solvent for the enzyme concentration as predictor variable. In this case, the importance of other variables is significantly decreased and the variables in the region from 1500-1750 cm^{-1} gain importance.

To train and test the PLS model, the DoE data set was again split into a training (n= 23) and test dataset (n= 11). Based on the generated model, the solvent concentration of the different subtilisin protease/polyol formulations in the test dataset was determined. As the first PLS component comprises the variance related to the solvent concentrations of the DoE, the model was created using one PLS component only. Similar to the PCR model the generated PLS model (*fig. 49*) exhibits a high accuracy for the solvent concentration predictions, with regard to the calibration (R^2= 0.999; RMSE= 0.24; bias= $3.01*10^{-15}$) and cross validation (R^2= 0.999; RMSE= 0.30; bias= 0.001) results. Further, the predicted

results (R^2= 0.999; RMSE= 0.33; bias= 0.08) for the solvent concentrations of the test dataset show to exceed the already highly accurate results from the PCR model.

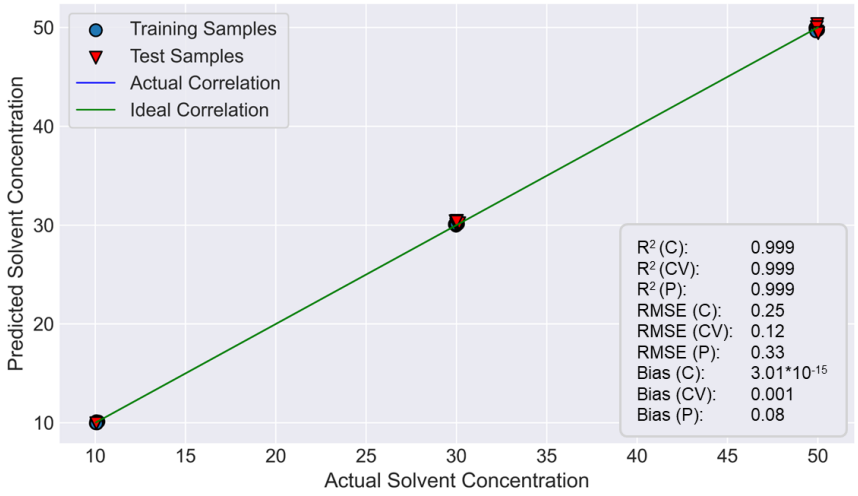


Figure 49: Calibration, cross-validation, and prediction results for the generated PLS model for solvent concentration prediction of different subtilisin protease/MPG formulations.

Therefore, it could be shown that the generated PLS model is able to predict the solvent concentration of different subtilisin protease/polyol formulations with a high accuracy, based on the underlying MIRA-spectra. PCR and PLS model both exhibit a high accuracy for the prediction of the solvent concentration. Consequently, both approaches have shown to be suitable for the intended use.

3.6.2 Subtilisin Protease/MPG Formulations for Enzyme Concentration Prediction

After the solvent concentration prediction had turned out to be consistent and highly accurate, a second model to predict the enzyme concentration of different subtilisin protease/MPG formulations was developed. In this regard, the DoE samples from *chapter 3.5.1* were used. As a reference, the enzyme concentration of the samples was determined with the AAPF assay using the automated photometer Gallery™ system. In parallel, PCR and PLS models were created in order to predict the inherent active enzyme content based on the MIRA-spectra.

Solvent Signal Removal from the MIRA-Spectra

Prior to model creation, however, the influence of the solvent was to be diminished in the obtained spectra, to prevent the results being influenced by the underlying solvent signal. The PCA/PLS analysis in *chapter 3.5.1* has shown, that most of the variability as well as most of the variables of the spectrum have a high correlation with the solvent concentration. Therefore, the idea was, to reduce the influence caused by the solvent, by predicting and subtracting an artificial solvent spectrum from the subtilisin protease/MPG MIRA results. Thereby, a possible cross correlation/multicollinearity in between the enzyme and solvent concentrations, was to be prevented.

To reduce the variability and the influence caused be the solvent on the prediction of the enzyme concentration, first, a PLS model for the prediction of the solvent concentration was generated based on the approach in *chapter 3.5.1*. The created model comprised all samples from the DoE. Following, the model was generated using the first PLS component describing 82.41 % of the total variance of the dataset (*fig. 50*). The overall accuracy shows to be high, for the calibration (R^2= 0.999; RMSE= 0.27; bias= $-9.93*10^{-16}$) and cross validation (R^2= 0.999; RMSE= 0.24; bias= 0.010).

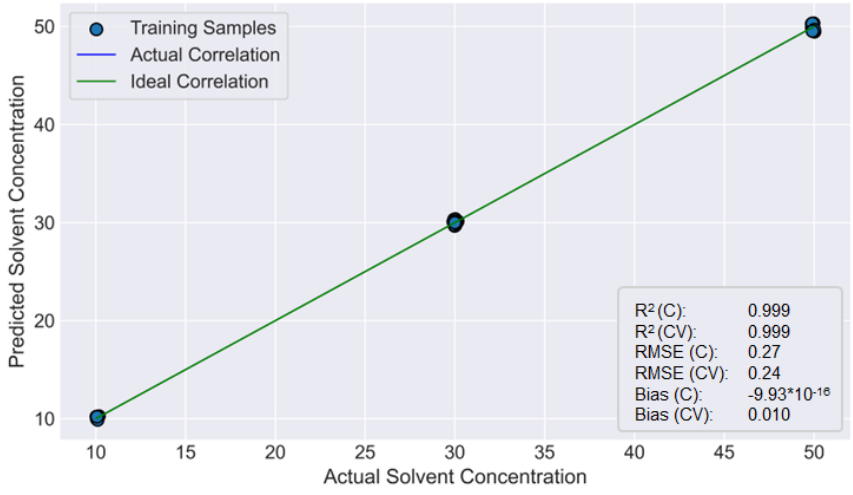


Figure 50: PLS-Model for solvent concentration prediction of subtilisin protease/MPG formulations in order to diminish the influence of the solvent on the overall variability of the spectra.

In order to remove the solvent signal from the MIRA-spectra of the DoE, several enzyme-free solvent spectra of 1,2-propanediol (MPG) were acquired. In this context, samples in a concentration range from 0 wt% to 80 wt% MPG, with an interval of 10 wt%, were

generated in duplicates. The solvent spectra (*fig. 51A*) show a clear separation according to their inherent solvent concentration. Especially, in the region from 1500-1750 cm^{-1} and 2500-3000 cm^{-1} eight distinct groups of spectra can be observed. In the cause of this, absorbance values show to increase in the presence of elevated solvent concentrations.

Following, an average spectrum for each MPG concentration was calculated based on the duplicates. Since the absorbance of neighboring MPG concentrations show a close to linear correlation (*fig. 51B*), spectra of any concentration in between the acquired can be calculated. With larger concentration differences, however, linearity is lost, which is why the difference has to be kept as small as possible. This non-linear behavior of the MIRA spectra is expected to be the result of unknown corrections applied by the device. One plausible explanation could be polynomial baseline corrections, which is however hard to assess since this is inherent to the nature of the MIRA results.

Consequently, based on the predicted concentrations for the DoE samples, obtained from the calibration results (*fig. 50*), an artificial solvent spectrum for each sample was generated and subtracted from their subtilisin protease/MPG counterpart.

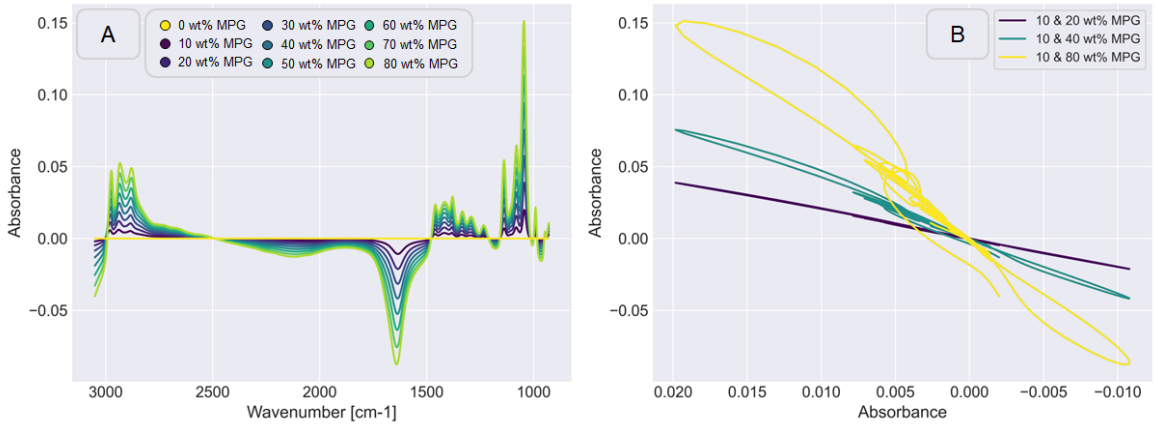


Figure 51: Generation of artificial solvent spectra of different concentrations based on (A) spectra with different MPG concentrations 10-80 wt%. *Due to (B) the linear correlation of the absorbance values of neighboring concentrations the spectrum of any MPG concentration in between can be calculated. When increasing the distance between the concentrations, the linear correlation is lost.*

The resulting spectra (*fig. 52*), show a high similarity with a native protein MIR-spectra, comprising several characteristic protein bands. The amide I is by far the most intense absorption band, which is found in the range between 1600 and 1700 cm^{-1} [171, 245]. The resulting band is formed due to the stretching vibrations of the C=O (70-85 %) and C-N

(10-20 %) groups of the amide bonds. Further, the amide II band can be determined in the expected range in between 1480 to 1580 cm^{-1}. Chemically, the amide II band is more complex than the amide I. The visible band is the result of in-plane N-H bending (40-60 % potential energy), C-N (18-40 %) and the C-C (≈10 %) stretching vibrations. Besides these two dominant bands, several smaller bands are visible, which cannot be clearly assigned to additional amide related molecular vibrations. Signals represented by the amide III and IV bands are in general very complex [171, 246], since the underlying signal is dependent on the force field in the respective environment created by individual side chains and hydrogen bonds of proteins. Therefore, the pattern of the individual bands is specific for different proteins, which makes the clear assignment of additional bands highly complex.

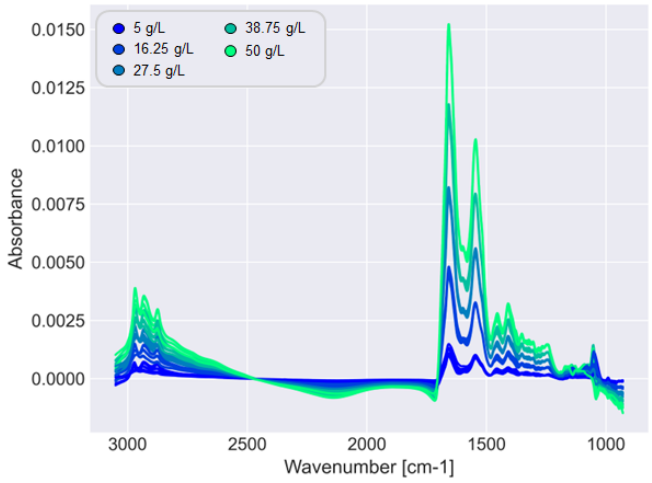


Figure 52: Resulting MIRA-spectra after solvent spectra subtraction. *The resulting spectra illustrate the morphology of a native protein MIR-spectra. Amide I (1600-1700 cm^{-1}) and II band (1480-1580 cm^{-1}), as well as several additional bands resulting from the protein structure can be identified.*

Preprocessing

After the solvent spectra had been subtracted from the subtilisin protease/MPG spectra, the resulting "solvent-free" spectra were used to generate PCR/PLS models for enzyme concentration prediction. To further increase the prediction accuracy and prevent non-protein regions of the spectrum influencing the prediction results, the range of the used spectra was reduced (*fig. 53A*) to the spectral range in between 1461-1700 cm^{-1}. Thereby, the number of variables was reduced from 1101 to 125 datapoints. The resulting spectral range was expected to comprise the enzyme related signals, namely the amide I and II

band. The restricted range exhibits five different groups of spectra, based on the five different enzyme concentrations of the DoE.

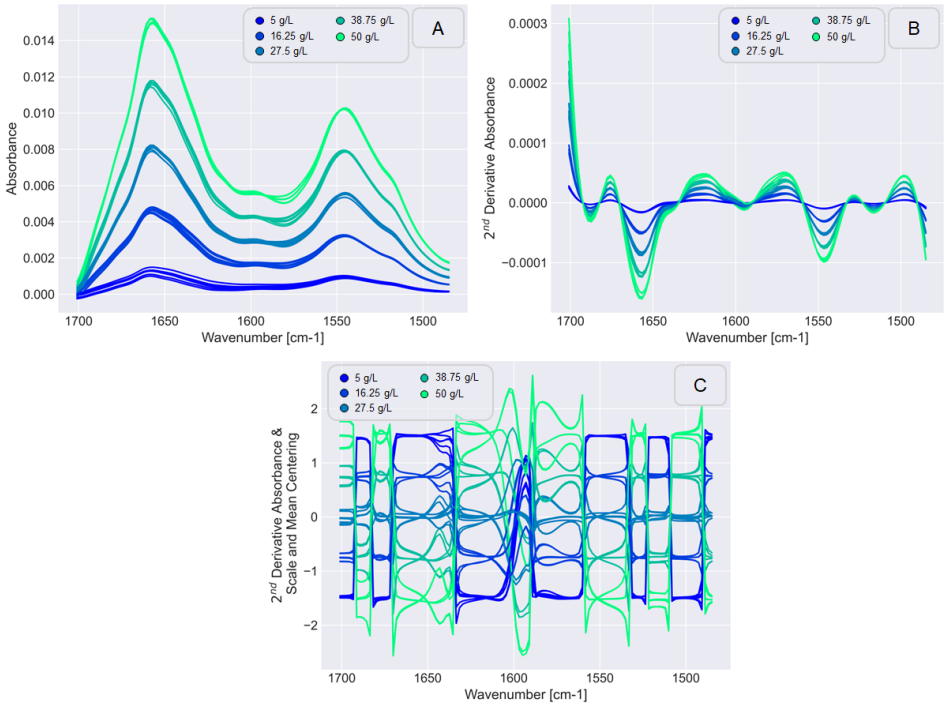


Figure 53: Preprocessing of the (A) reduced spectral range after solvent subtraction. *(B) SavGol-filtering (5th order polynomial; 2nd derivative) and (C) autoscaling was applied.*

Following, the range selection in the spectra was further preprocessed by applying a SavGol-filter (*fig. 53B*) and by autoscaling (*fig. 53C*) the dataset. With regard to the SavGol-filter, a 5th order polynomial was used for smoothening and 2nd order derivative to remove noise and possible offsets in the spectra. Besides, autoscaling, was applied to remove any systematic variations in the data and ensure that the data is on the same scale, making it easier to compare and analyze different features in the dataset.

Hierarchical Clustering

The subsequent hierarchical clustering analysis (*fig. 54*), based on the Euclidean distance algorithm and Ward linkage method, revealed three major clusters, which comprise samples based on the different enzyme concentrations. While the orange cluster contains samples with enzyme concentrations of 38.75 and 50 g/L, the green cluster comprises samples with an enzyme concentration of 27.5 g/L. At last, the red cluster is composed of the two lowest enzyme concentrations of 5 and 16.25 g/L. The three main clusters are further subdivided according to the solvent concentrations of the respective samples.

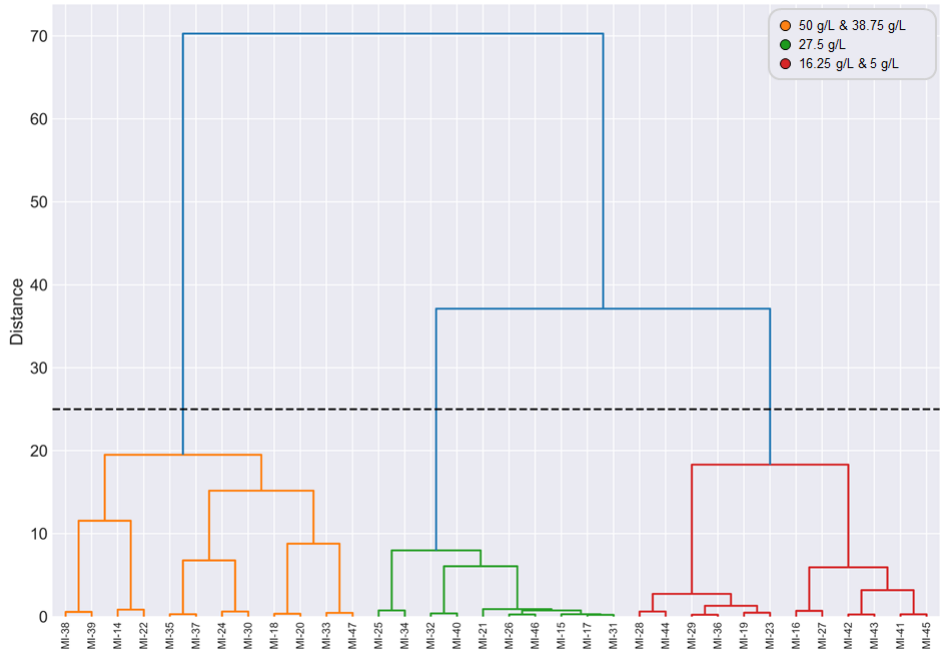


Figure 54: Dendrogram of the hierarchical cluster analysis performed with the resulting subtilisin protease spectra after subtraction. *The orange cluster contains samples with an enzyme concentration of 50 and 38.75 g/L, green 27.5 g/L, and red 5 and 16.25 g/L. In addition, samples inside the different enzyme concentration clusters were assigned according to the solvent concentration present in the sample.*

Subsequently, the hierarchical cluster analysis led to the assumption, that by subtracting the solvent spectra the variability of the solvent had been diminished but not totally removed. Therefore, samples were primarily clustered according to the inherent enzyme concentration and secondly according to the solvent concentration.

Principal Component Analysis (PCA)

To reduce the dimensionality of the underlying dataset, in context of enzyme concentration prediction, a PCA analysis was conducted. The resulting scree plot (*fig. 55*) visualizes the comprised variability by each individual PC. With regard to the spectral data after subtraction, PC1 accounts for 90.37 % of the total variance. Already with PC2 99.58 % (+ 9.04 %) of the total variability is covered. Consequently, following PCs covered only minor proportions of additional variability.

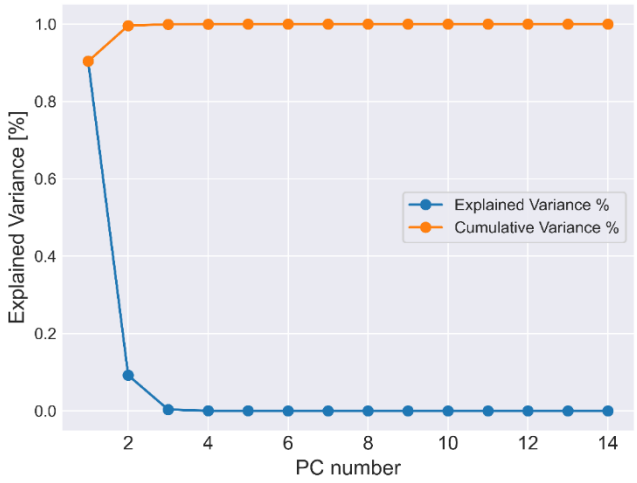


Figure 55: PCA scree plot with the explained and cumulative variance of the underlying dataset after solvent spectra subtraction.

The scores plot (*fig. 56*) led to additional insights on the roots of the variability explained by PC1. In contrast, to the original spectra in *chapter 3.5.1*, where PC1 scores illustrate a grouping according to the solvent concentration, scores values of PC1 after subtraction, show a tendency to correlate with the enzyme concentration values. Subsequently, scores values show five distinct groups based on their PC1 score. From left to right, the five groups can be assigned to the different enzyme concentrations ranging from 5-50 g/L. Further, score values of PC2 show to form three different groups based on the different solvent concentrations of the samples.

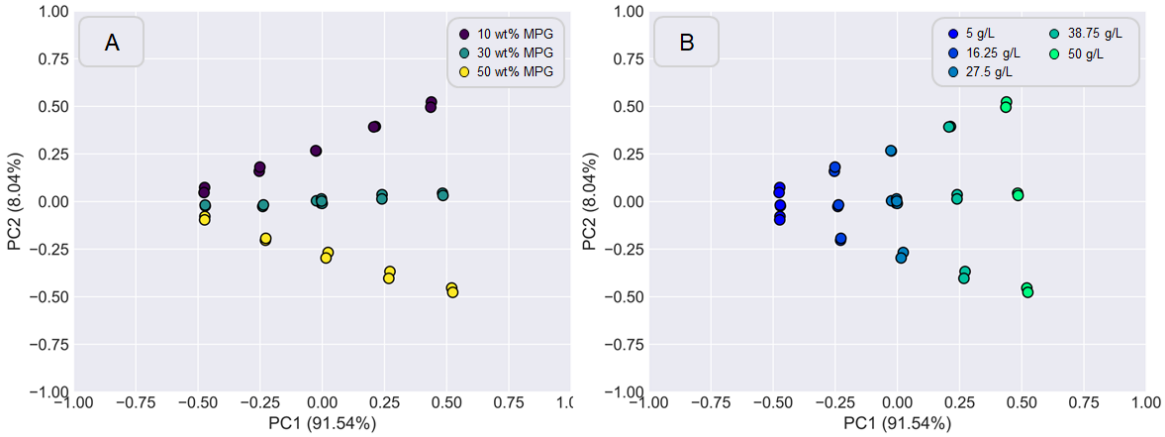


Figure 56: PCA scores plot colored according to the (A) solvent and (B) enzyme concentration of the subtilisin protease/MPG formulations.

This, together with the results from the hierarchical clustering, strengthened the assumption that the subtraction approach had not totally removed the influence of the solvent on the spectrum. Furthermore, the subtraction method seems to have varying degree of success based on the underlying enzyme concentration of the samples. While scores with a high enzyme concentration appear to be extensively distinct, lower enzyme concentrations were only slightly spread with regard to PC2.

Based on the loadings plot (*fig. 57*), variables with a strong correlation to the variability comprised by PC1, can be determined over the full range of the restricted area of the spectrum. With the exception of a small region in between both bands, the majority of the variables show a strong correlation with PC1. The maximum intensity of both bands shows a strong negative correlation, while the body of the bands are strongly positively/negatively correlated. The loadings of PC2 in turn, are in most cases negatively correlated with some few exceptions. However, the region between both bands shows a strong positive correlation with the variability comprised by PC2.

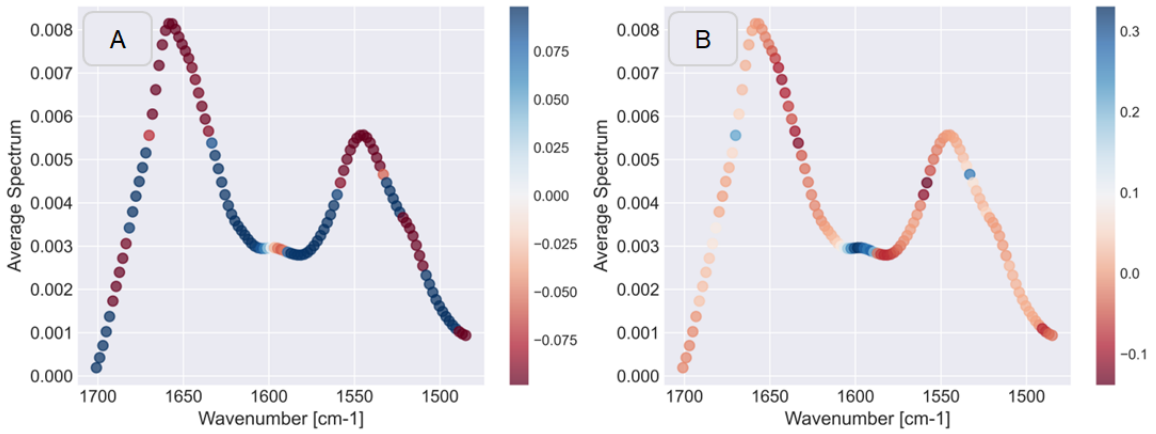


Figure 57: Mean IR of the used spectral region colored according to PCA loadings for (A) PC1 and (B) PC2.

Principle Component Regression (PCR) for Enzyme Concentration Prediction

Since the PCA analysis has shown a tendency for a separation according to the enzyme concentration based on PC1, a principal component regression (PCR) was undergone. Thereby, the enzyme concentration of different subtilisin protease/MPG formulations from the DoE was to be predicted. For this the DoE data was again split in a training dataset (n= 23) and a test dataset (n= 11) and the solvent content in the spectra was diminished based

on the described subtraction approach. Following, a PCR model was created based on the test dataset, using one PC only. To validate the predicted results, the agreement with experimentally determined enzyme concentrations based on the AAPF assay was verified.

A PCR model with one principal component shows to be sufficient to achieve a high accuracy, with regard to the prediction of the enzyme concentration of the test dataset. The obtained results (*fig. 58*) for the calibration (R^2= 0.997; RMSE= 0.74; bias= $1.62*10^{-15}$) and cross validation (R^2= 0.996; RMSE= 0.95; bias= 0.05) showed to overlap with an ideal correlation. In general, calibration and cross-validation yield a high accuracy respectively small offsets for the linear regression coefficient (R^2), root mean square error (RMSE), and bias. Consequently, the PCR model returned predicted enzyme concentrations with a high accuracy and low error for the test dataset (R^2= 0.998; RMSE= 0.68; bias= 0.28).

Besides a PCR model (*fig. 59*) was created to predict the solvent concentration based on the restricted spectrum. Thereby, a possible cross correlation/multicollinearity between enzyme and solvent concentration prediction was to be excluded. The PCR model was created by using PC1 only. As can be observed, the variability comprised by PC1 is not able to describe the inherent solvent concentration of the underlying samples. The generated model shows a significant discrepancy between the ideal and actual correlation. Subsequently, R^2 exhibits values close to zero and the RMSE shows a high root mean square error in between 14.77, for the predicted and 28.39 for the cross-validation results. In addition, the bias of the predicted values shows an offset of -7.83.

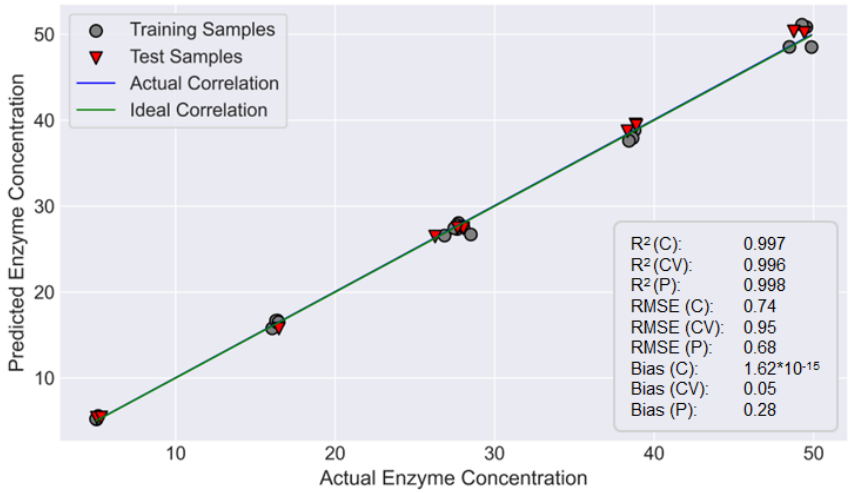


Figure 58: Calibration, cross-validation, and prediction results for the generated PCR model for enzyme concentration prediction based on the magnified spectral range.

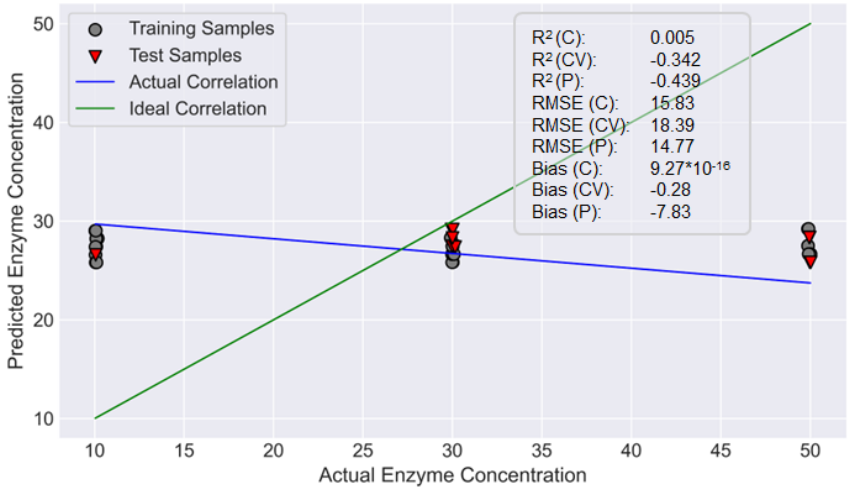


Figure 59: Calibration, cross-validation, and prediction results for the generated PCR model for solvent concentration prediction based on the magnified spectral range.

The established workflow has shown to significantly reduce the effect caused by the solvent in the underlying MIRA-spectra. The combination of solvent subtraction and restriction of the spectral range to protein associated regions, has shown to lead to highly accurate enzyme concentration predictions, with regard to the created PCR model. Overall, the created PCR model has shown to be reliable and consistent way to determine the enzyme concentration of different subtilisin protease/MPG formulations.

Partial Least Squares (PLS) Analysis for Enzyme Concentration Prediction

In addition to the unsupervised PCR model for enzyme concentration prediction, a supervised PLS model was created. Analogous to the PCR model, the DoE data with the solvent signal removed was split in a training (n= 23) and test dataset (n= 11). To be able to compare both results, the PLS model was generated using a single PLS component. In addition, a solvent model was generated to monitor the capability of the first PLS component to predict the solvent concentration. The determined enzyme concentrations, based on the AAPF assay, were used to create the model and to evaluate the predicted concentrations of the test dataset.

To determine the influence of the number of PLS components on the accuracy of the created model (*fig. 60*), the RMSE of the cross-validation results were determined. Already, with a single component a high accuracy (RMSE= 0.87) of the model can be determined. By further increasing the number of PLS component up to three, the RMSE can be reduced to ≈0.51. However, to retain the comparability with the PCR results the

PLS model was created based on the first PLS component only. Furthermore, increasing the number of PLS components to more than three is expected to capture noise and thereby to overfit the model.

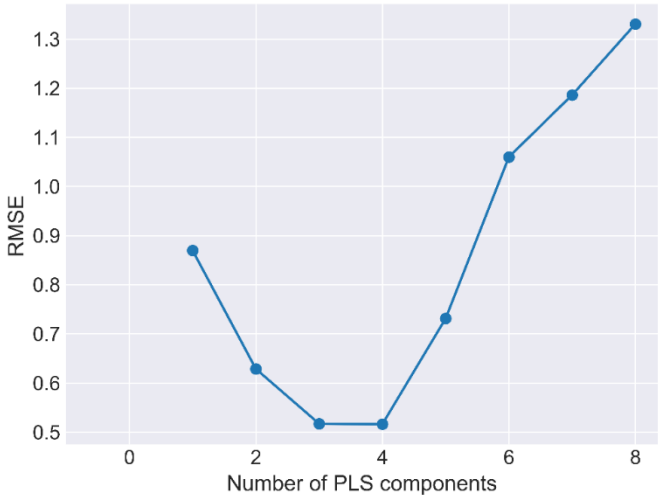


Figure 60: Root mean square error (RMSE) of the cross-validation with increasing number of PLS components.

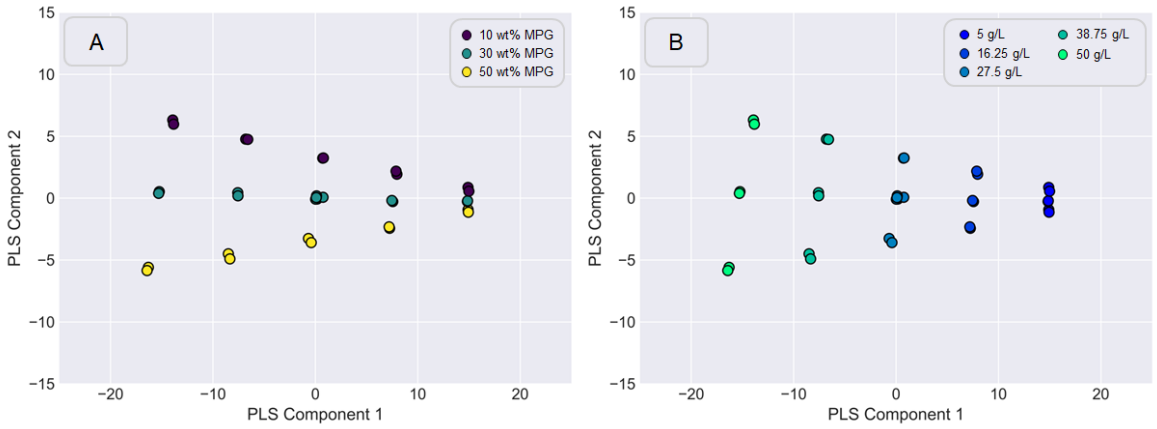


Figure 61: X-scores of the PLS model for the whole DoE dataset. *X-scores are colored according to (A) the MPG and (B) enzyme concentration.*

The X-scores (*fig. 61*) for the samples reveal a similar tendency as observed in the PCA. For the first PLS component the scores appear to be grouped according to their inherent enzyme concentration. Consequently, five different groups for the enzyme concentrations in between 5-50 g/L can be observed. Further, based on the scores for the second PLS component, samples appear to form groups according to the solvent concentration of the

samples. Therefore, based on the coloring according to the solvent concentration three groups could be identified.

When looking at the VIPs (*fig. 62*), of the created PLS model, the majority of the variables, in the range of the used IR spectrum (1461-1700 cm^{-1}), show a high influence (above 1) on the model creation. In general, VIPs are used to measure the importance of each variable in predicting the response variable. Since enzyme concentration is expected to correlate with the intensity of the amide I and II, chemically, a high importance of this region is expected. Subsequently, the majority of the variables show to contribute to the explained variance of the PLS model. Since VIP values for the different variables appear to be similar, most variables in the selected range appear to be equally important.

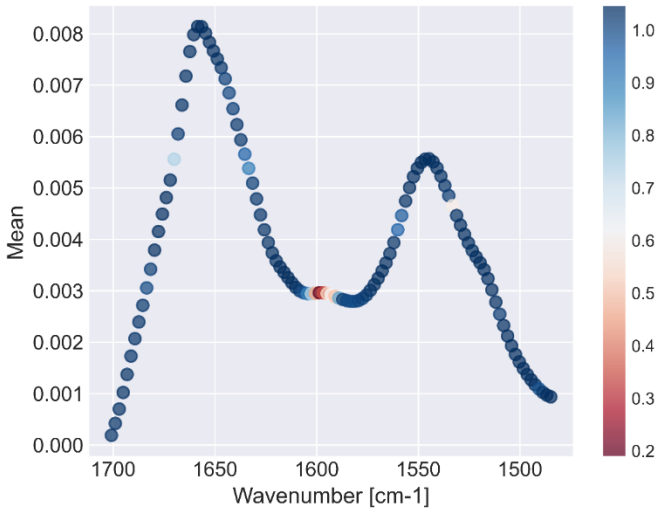


Figure 62: VIPs fort he created PLS model to predict the enzyme concentration.

The created PLS model (*fig. 63*) exhibits a high accuracy for calibration (R^2= 0.997; RMSE= 0.75; bias= $1.78*10^{-15}$), cross-validation (R^2= 0.996; RMSE= 0.87; bias= 0.69) and prediction (R^2= 0.57; RMSE= 0.69; bias= 0.30). In this regard, the PLS shows a similar accuracy as the generated PCR model.

Besides, the capability of the selected region to predict the solvent concentration was evaluated. As could be shown (*fig. 64*) based on the first PLS component, it is still possible to predict the solvent concentration of the different samples, however the accuracy of calibration (R^2= 0.83; RMSE= 6.56; bias= $2.08*10^{-15}$), cross-validation (R^2= 0.57; RMSE= 10.40; bias= 1.37) and predictions (R^2= 0.46; RMSE= 9.06; bias= -2.64) are diminished. Especially, the endpoint samples exhibit a broad spread and thereby a discrepancy in

between actual and calibration results. Subsequently, actual, and ideal correlation show a certain offset at minima and maxima.

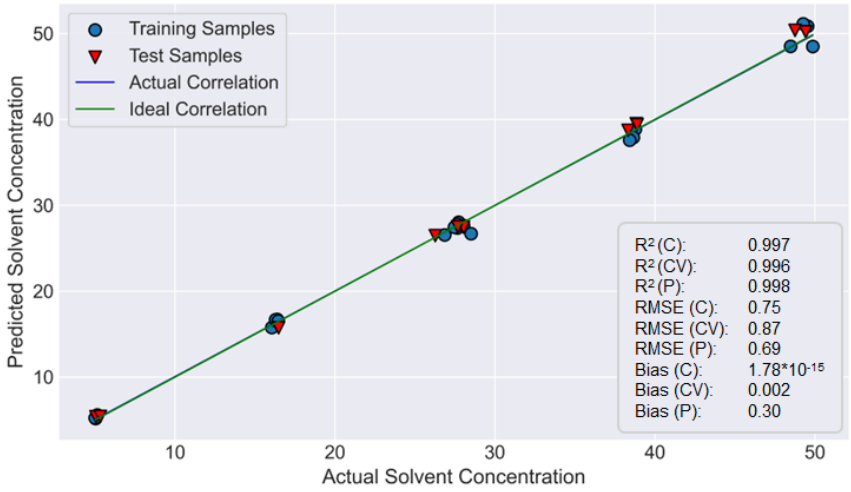


Figure 63: Calibration, cross-validation, and prediction results for the generated PLS model for enzyme concentration prediction of different subtilisin protease/polyol formulations after solvent subtraction. *The generated PLS model was generated based on PLS1 only.*

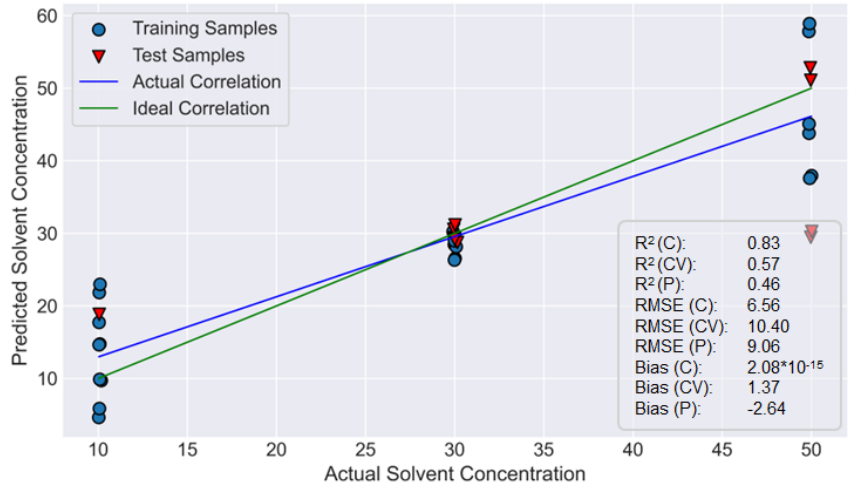


Figure 64: Calibration, cross-validation, and prediction results for the generated PCR model for solvent concentration prediction of different subtilisin protease/polyol formulations after solvent subtraction. *The generated PLS model was generated on PLS1 only.*

Overall, however, the conducted experiment has shown that based on MIRA-spectra the active enzyme content of different subtilisin protease/MPG formulations can be predicted with a high accuracy. The significant reduction of the solvent signal in the spectra, not only greatly diminished the influence of the variability caused by the different solvent

concentrations, but also added further insights on the behavior of the underlying protein signals in the spectrum, which have not been visible before. The conducted workflow has shown to be suitable in this regard, however, was not able to totally remove the influence of the solvent. In this context, further improvements appear to be necessary to further reduce the influence of the solvent on enzyme concentration predictions in subtilisin protease/polyol formulations.

3.6.3 Evaluation of Solvent and Enzyme Concentration Prediction Models

In order to evaluate the ability of the created PLS models to monitor the solvent and enzyme concentration in the course of long-term storage tests, additional samples were generated and tested. Thereby, the applicability of the developed workflow in the context of a practical application was to be evaluated. For this, several subtilisin protease/MPG formulations were stored over a defined period of time. Subsequently, the solvent and enzyme concentration of the samples was predicted, based on the PLS models created in *chapter 3.5.1* and *3.5.2*. Following, prediction results were compared to the calculated solvent concentrations and measured enzyme concentrations determined based on the AAPF assay on the Gallery™ system.

Solvent and Enzyme Concentration of Stored Subtilisin Protease Samples

First, the enzyme and solvent concentration of different subtilisin protease/MPG formulations after storage was to be determined. For this, different formulations with a starting concentration of ≈50 g/L active enzyme were generated, with MPG concentrations in the range of 10-53 wt% (interval 10 wt% (last 13 wt%)). All samples were prepared in duplicates. Following, the samples (n= 20) were stored at 45 °C over a period of 3 days in order to allow enzyme unfolding and degradation to progress. Thereafter, the active enzyme concentration of each sample, at timepoint 0 and after 3 days, was determined using the AAPF assay based on the Gallery™ automated photometer. Simultaneously, a MIRA spectrum was acquired for each sample and evaluated based on the developed PLS models for enzyme (*chapter 3.5.2*) and solvent (*chapter 3.5.1*) concentration prediction. In contrast to prior experiments, the created models were based on the entire DoE dataset (n= 34).

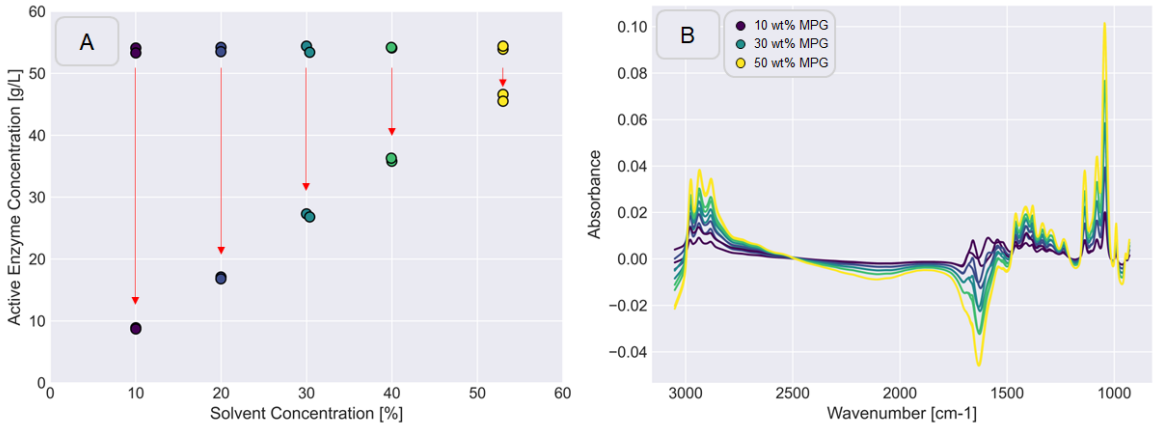


Figure 65: Residual enzyme activities and MIRA spectra obtained from a storage test conducted at 45 °C over a period of 3 days. *(A) Illustrates the active enzyme content of each sample at the beginning and after 3 days based on the present solvent concentration (n= 20). (B) MIRA-spectra of the storage test samples, colored according to the solvent concentration. All timepoints were sampled in duplicates.*

The storage test results (*fig. 65A*), show that the applied temperature has greatly diminished the active enzyme concentration of the different samples based on their inherent solvent concentration. The loss is observed to decrease from 10-50 wt% MPG. Furthermore, structural changes of the MIRA-spectra (*fig. 65B*) due to the reduction of the active enzyme content can be observed. Especially, the region comprising the amide I and II bands, in the range from 1500-1750 cm^{-1}, appear to be altered throughout storage. Besides, the already experienced grouping of the spectra according to the solvent concentrations can be observed. In this regard, five separated groups of spectra, representing the five MPG concentrations, are visible in the range of 1500-1750 cm^{-1} and 2500-3000 cm^{-1}.

Analogous to *chapter 3.5.1* the solvent concentration was determined using the PLS model based on the DoE samples. For this, all spectra (DoE, and storage samples) were preprocessed in a single approach. A SavGol-filter (5^{th} grade polynomial; 2^{nd} order derivative) was applied and the data was autoscaled. Following, the dataset was split, and the DoE samples were used to create the PLS models for solvent concentration prediction.

Similar to *chapter 3.5.1* the DoE solvent model (*fig. 66*) exhibits a high accuracy for calibration (R^2= 0.999; RMSE= 0.24; bias= $-7.84*10^{-16}$) and cross-validation (R^2= 0.999; RMSE= 0.27; bias= 0.01). Furthermore, the prediction of the solvent concentration in the storage test samples returns results with a negligibly lower accuracy. For the storage test samples an R^2-value of 0.997, a RMSE of 0.81 and a bias of 0.67 can be determined. In

addition, samples outside the calibration range, with 53 wt% MPG, return similar accurate values. Subsequently, the created PLS model shows to be accurate beyond the calibration range.

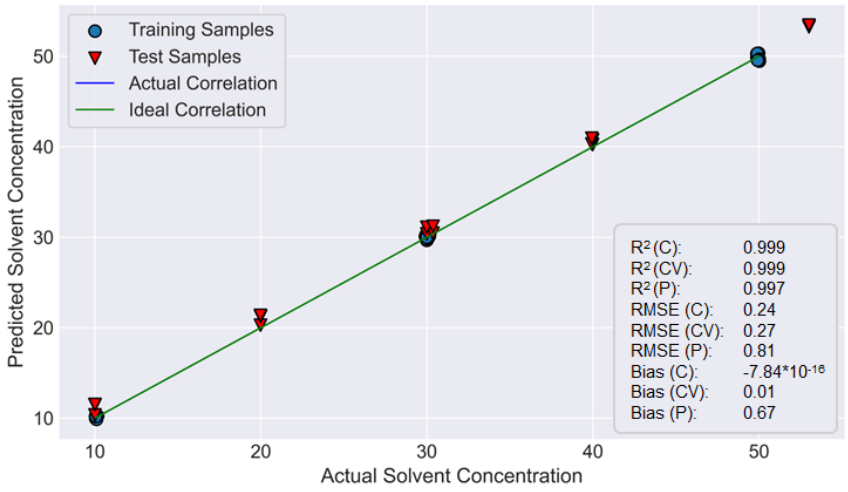


Figure 66: Calibration, cross-validation, and prediction results for the generated PLS model for solvent concentration prediction of different subtilisin protease/MPG storage test samples. *The generated PLS model was generated based on PLS1 only.*

Following, the enzyme concentration of the storage test samples was predicted based on the PLS model generated in *chapter 3.5.2*. Therefore, the solvent concentration predictions were used to generate artificial MPG spectra, which were subtracted from the MIRA-results of the storage test. The resulting spectra (*fig. 67A*) illustrate the already experienced morphology of the amide I and II bands, characteristic for protein samples.

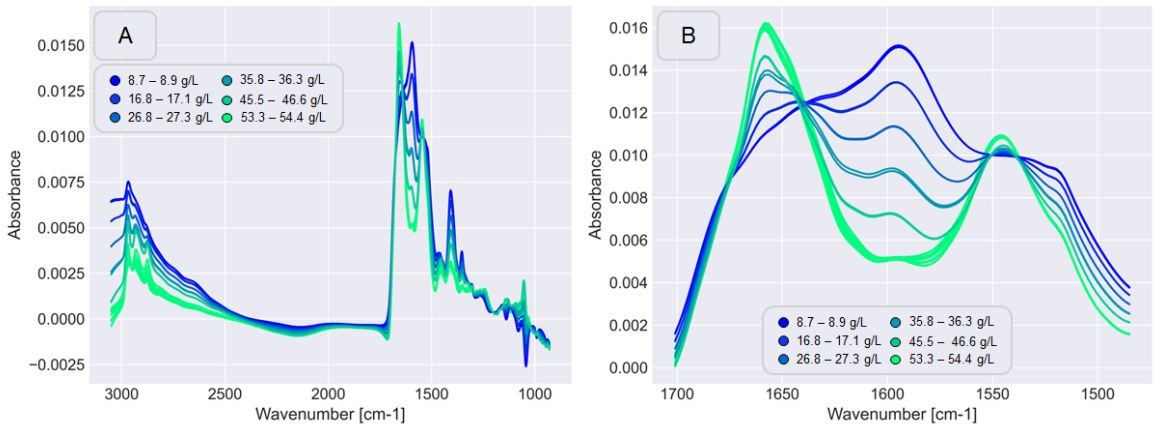


Figure 67: MIRA-spectra of the storage test samples (A) after solvent subtraction and in the region from 1461-1700 cm^{-1}.

Besides several smaller bands are visible.

In contrast, to the original spectra returned by the DoE samples, several structural changes are visible for the amide I and II bands. Not only is the intensity of both bands changed, but also a third band appears to develop in between. This is especially visible when magnifying the range (*fig. 67A*) of the spectrum to the amide I and II band (1461-1700 cm^{-1}). Subsequently, this range was further processed by applying a SavGol-filter (5^{th} order polynomial; 2^{nd} derivative) and autoscaling the dataset.

Based on the preprocessed, restricted spectral range, of the DoE samples, a PLS model (*fig. 68*) for enzyme concentration prediction was generated. Following, the enzyme concentration in the storage test samples was determined. As was to be expected results from the DoE calibration (R^2= 0.996; RMSE= 0.92; bias= $1.65*10^{-15}$) and cross validation (R^2= 0.995; RMSE= 1.00; bias= 0.77) showed similar values as in *chapter 3.5.2*. Furthermore, the predicted results (R^2= 0.994; RMSE= 1.22; bias= 0.44) show to be in good agreement with the enzyme concentrations determined with the AAPF assay. For higher enzyme concentrations, however, a slight offset of the predicted values could be observed, when compared to the actual enzyme concentrations.

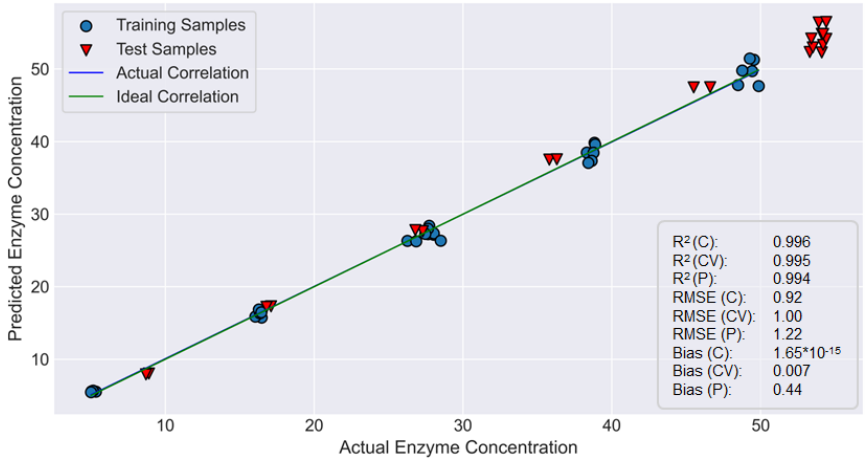


Figure 68: Calibration, cross-validation, and prediction results for the generated PLS model for enzyme concentration prediction of different subtilisin protease/MPG storage test samples. *The generated PLS model was generated based on PLS1 only.*

The conducted experiment has shown the high potential of MIRA to be used in the context of enzyme and solvent concentration prediction of storage test samples. As in the experiments before, the PLS solvent model is capable of predicting the solvent

concentration with a high accuracy. For the enzyme concentration a minor offset of the predictions, when compared to the actual values can be observed. Since the region used for the prediction is altered during sample storage, an effect of the observed changes on the prediction seems likely. However, this problem is expected to be overcome by increasing the sample space for model creation by including stored samples.

In addition, the used workflow has provided additional insights on the influence of storage on the morphology of the amide I and II band. In general, the position of the amide I band is sensitive to the hydrogen bonding pattern and local electrostatic interactions within the peptide backbone, while the amide II band reflects the local conformation of the peptide backbone [171, 244]. By analyzing the positions and relative intensities of these bands, it is possible to determine the proportion of the secondary structure motifs, as wells as the degree of structural disorder or unfolding [247, 248]. Since unfolding and autolysis are expected to gain influence at elevated temperatures, changes and loss of the secondary structure and thereby changes of the amide I and II can be expected.

Evaluating Long Term Storage Test Results of a Subtilisin Protease using MIRA

In addition, to the evaluation of the generated PLS prediction models with aged subtilisin protease/MPG samples, with varying solvent concentrations, another approach focused on the prediction of the long-term stability. Thereby, the capability of the developed approach to be used in the scope of long-term storage tests was evaluated. In this context, a subtilisin protease formulation with ≈46 g/L and 10 wt% MPG was prepared in triplicates. Following, each sample was stored over a period of 28 days at 37 °C, whereby samples were withdrawn after 0, 7, 14, 21 and 28 days. The obtained samples were analyzed with regard to their residual active enzyme content using the AAPF assay on the Gallery™ system. In addition, each sample was analyzed with MIRA to obtain the spectral data necessary to perform solvent and enzyme concentration predictions.

The residual enzyme concentrations (*fig. 69A*) show to drastically decline over the 28 days of storage, as can be illustrated based on the AAPF results. The average value for the active enzyme concentration of the three independent repeats show to decrease from 46.2 g/L to 12.6 g/L over the period of 28 days. Similarly qualitative changes in the region of the amide I and II bands can be observed in the MIRA-spectra (*fig. 69B*).

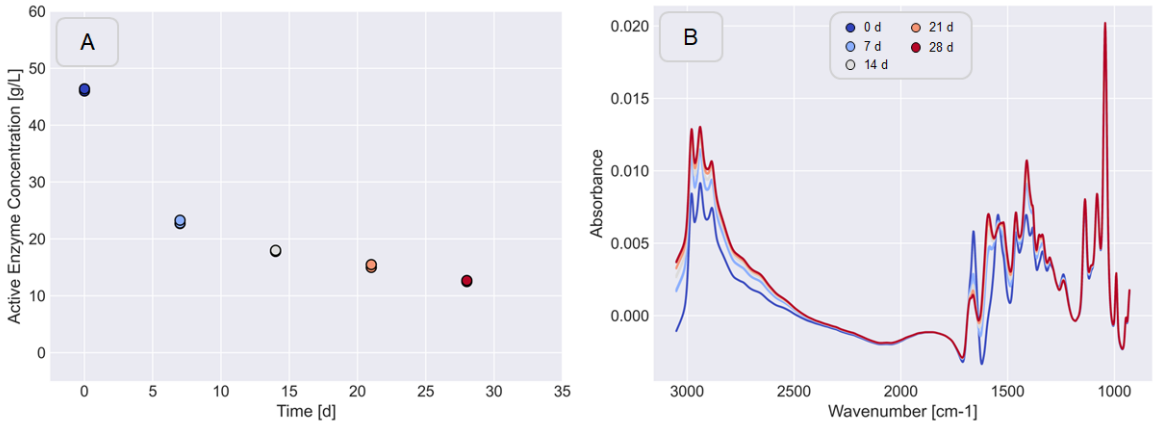


Figure 69: Enzyme activities and MIRA-spectra obtained based on long-term storage test conducted at 37 °C over a period of 28 days. *(A) Illustrates the active enzyme content of the samples drawn after 0, 7, 14, 21, and 28 days. (B) MIRA-spectra of the storage test samples. All timepoints were sampled in triplicates.*

After the combined pre-processing of the DoE and long-term storage test data, the solvent concentration (*fig. 70*) of the underlying samples was determined. As was to be expected based on the former experiment's, calibration (R^2= 0.999; RMSE= 0.24; bias= $-9.93*10^{-16}$) and cross-validation (R^2= 0.999; RMSE= 0.27; bias= 0.01) of the DoE data returned highly accurate values. The predicted solvent concentrations for the long-term storage samples provide results with a minimally lower accuracy. In this regard, a RMSE of 1.01 and a bias of 0.98 could be determined for the data set.

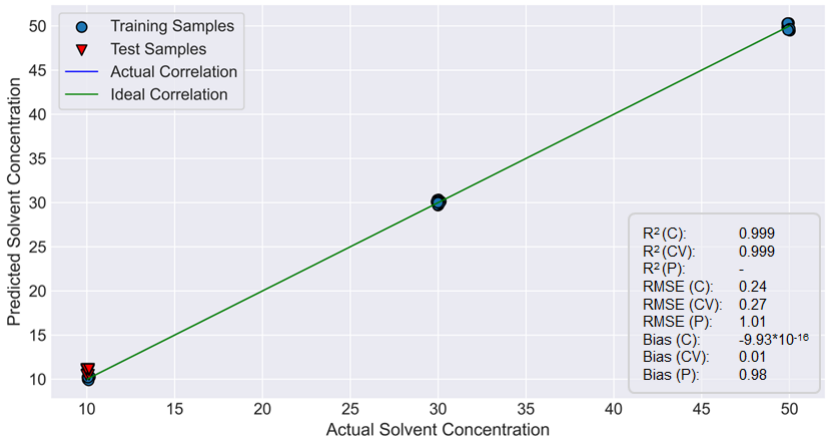


Figure 70 Calibration, cross-validation, and prediction results for the generated PLS model for solvent concentration prediction of long-term storage test samples. *The generated PLS model was generated based on PLS1 only.*

Following, the predicted solvent concentrations were used to create artificial MPG spectra with the respective solvent concentrations. These were subtracted from the original spectra to reduce the influence of the solvent on the subsequent enzyme concentration prediction.

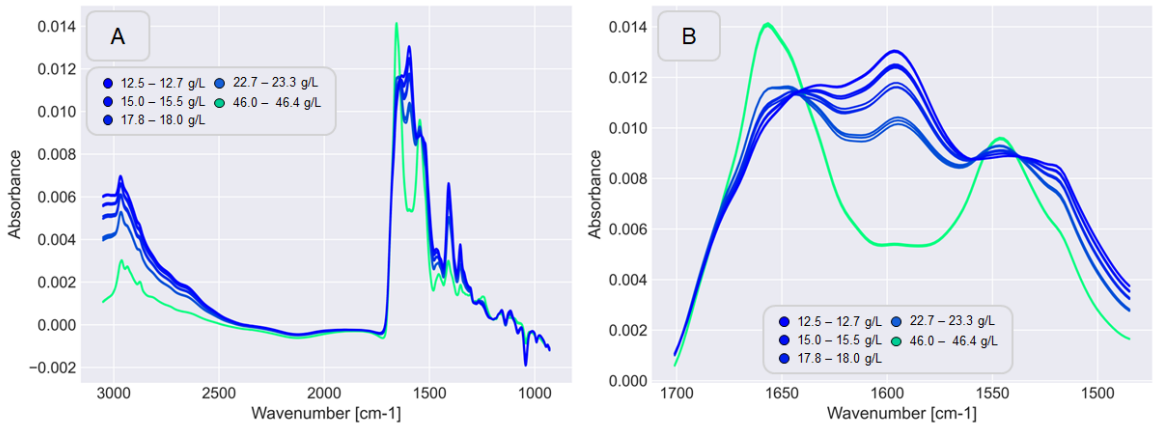


Figure 71: MIRA-spectra of the long-term storage test samples (A) after solvent subtraction and in the region from 1461-1700 cm^{-1}.

The resulting spectra (*fig. 71A*) illustrate the structural changes accompanying the storage of the subtilisin protease over time. As observed before, amide I and II show to be the two most prominent bands. Furthermore, the structure of both bands appears to be considerably altered throughout storage. Since the intensity of amide I and II is directly related to the underlying secondary structure, changes in this regions are to be expected upon unfolding and degradation of the enzyme.

Subsequently, the restricted spectral range (*fig. 71B*) of the DoE was used to generate a PLS model for enzyme concentration prediction of the storage test samples. First, however, DoE and storage test spectra were pre-processed by applying a SavGol-filter (5^{th} polynomial; 2^{nd} derivative) and autoscaling the dataset.

Based on the created PLS model (*fig. 72*) the enzyme concentration of the long-term storage test results was predicted and compared to values obtained from the activity measurements. The predicted storage test results, exhibit an R^2-value of 0.994, a RMSE of 0.93, and a bias of -0.059. These deviated only slightly from the results of calibration (R^2= 0.996; RMSE= 0.92; bias= $1.70*10^{-15}$) and cross validation (R^2= 0.995; RMSE= 1.00; bias= 0.007).

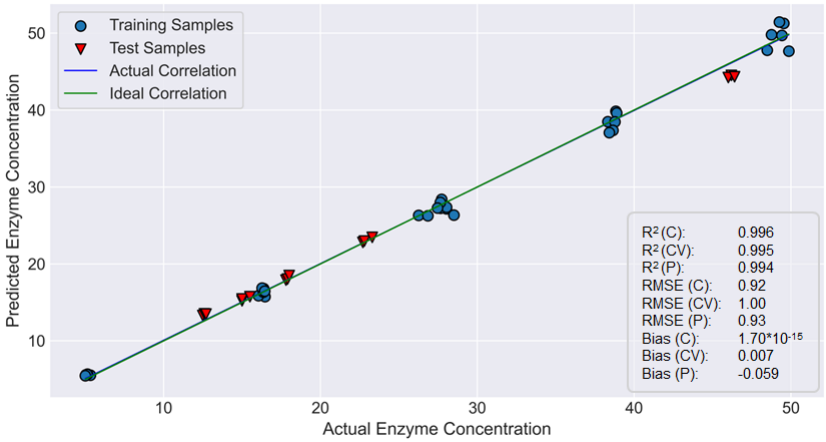


Figure 72: Calibration, cross-validation, and prediction results for the generated PLS model for enzyme concentration prediction of long-term storage test samples. *The generated PLS model was generated based on PLS1 only.*

When the long-term stability of the subtilisin protease is presented as a relative loss of activity over time, only minor differences between the predicted and measured content of active enzyme becomes apparent. To do so, the results of both approaches were normalized [%] with respect to the initial respectively predicted enzyme concentration (0 d). The resulting plot (*fig. 73*) describes the stability of the enzyme over time. While the assay-based enzyme activity was 27 % after 60 days, a residual activity of 30 % could be determined on the basis on the predicted values. The offset of 3 % is below the average error of 5 % in generally experienced with the AAPF assay.

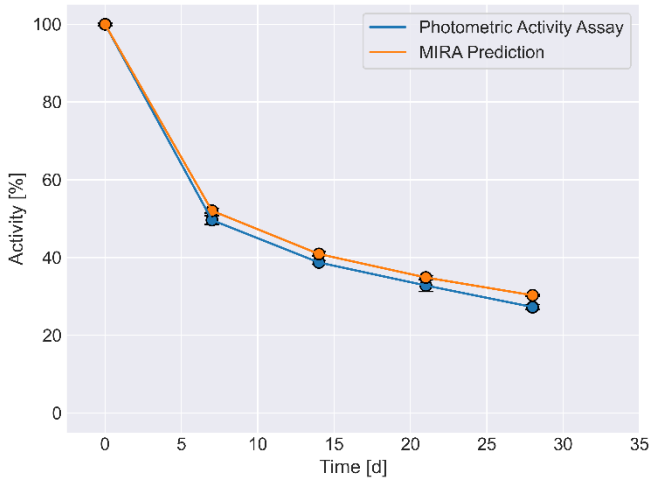


Figure 73: Comparison of the residual enzyme concentration results of the subtilisin protease concentrations determined with the photometric AAPF assay and the predictions based on MIRA spectra. *The determined enzyme concentrations were normalized with respect to the initial enzyme concentration present in the samples at timepoint 0 d.*

The obtained results highlight the capability of the described approach to determine the enzyme concentration of different subtilisin/protease formulations based on MIRA-spectra. The negligible difference between the storage test results obtained from the photometric assay and the PLS-model provide additional support that the developed methodology can be used to analyze long-term storage test samples. The use of undiluted/-processed samples and the missing dependency on different substrates and buffers makes the presented approach convenient and easy to apply to different samples with varying formulation matrices. Besides the enzyme concentration, additional information about the formulation composition can be obtained and integrated into the analysis, which provides additional insights on the behavior of the enzyme in the respective environment.

3.6.4 Endocellulase/MPG Formulations for Solvent Concentration Prediction

The conducted experiments based on subtilisin protease formulations in *chapter 3.5.1-3* have demonstrated the high accuracy of the enzyme and solvent concentration predictions based on MIRA-spectra. To further evaluate the capability of this methodology to be used in the scope of enzyme formulations, additional experiments were conducted with an endocellulase. In contrast, to the subtilisin protease the missing enzyme loss due to autolysis was expected to cause minor changes in the MIRA-spectra. Following, it was evaluated whether changes of the MIR-spectrum due to unfolding are sufficient to determine the enzyme concentration of different endocellulase/MPG formulations. First, however, the solvent concentration of the different formulations was to be predicted.

In this context, a long-term storage test with different endocellulase/MPG formulations was conducted. The formulations contained ≈30 g/L of active enzyme and 10-50 wt% MPG (intervals of 10 wt%). All formulations were prepared in duplicates and subsequently stored over a period 69 days at a temperature 45 °C. Throughout, samples were withdrawn after 0, 21, 28, 35 and 69 days. The concentration of the solvent in the different formulations was calculated based on the weighed amount (wt%) of solvent used to prepare the samples. In addition, it was assumed that the solvent concentration is not changed throughout the experiment. The MIRA-spectra were acquired in the range of 929-3051 cm^{-1} and were composed of 1101 datapoints.

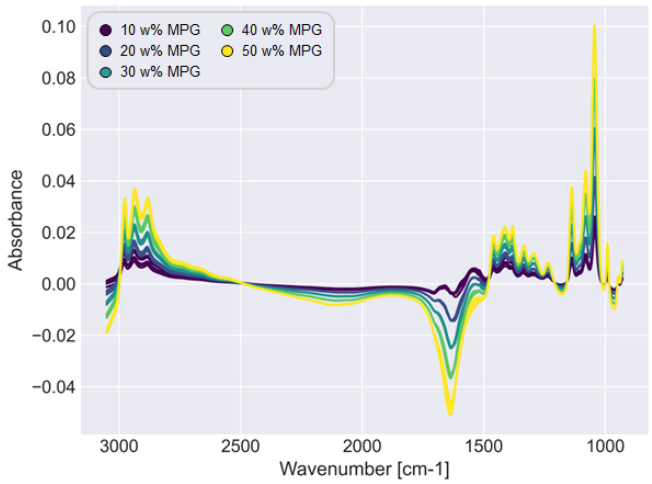


Figure 74: Spectra of different endocellulase/MPG formulations in the range of 10-50 wt% MPG.

Already by visual inspection (*fig. 74*), significant differences in between the samples can be observed. With regard to the solvent concentration of the samples several regions show a clear intensity difference in between the spectra. Especially in the range from 1500-1750 cm^{-1} and 2500-3000 cm^{-1} a separation in five separated groups of spectra based on the solvent concentrations, can be observed. In contrast to the subtilisin samples, however, only minor changes in the region of 1500-1750 cm^{-1} can be observed, with regard to the altered enzyme activity throughout storage.

Preprocessing

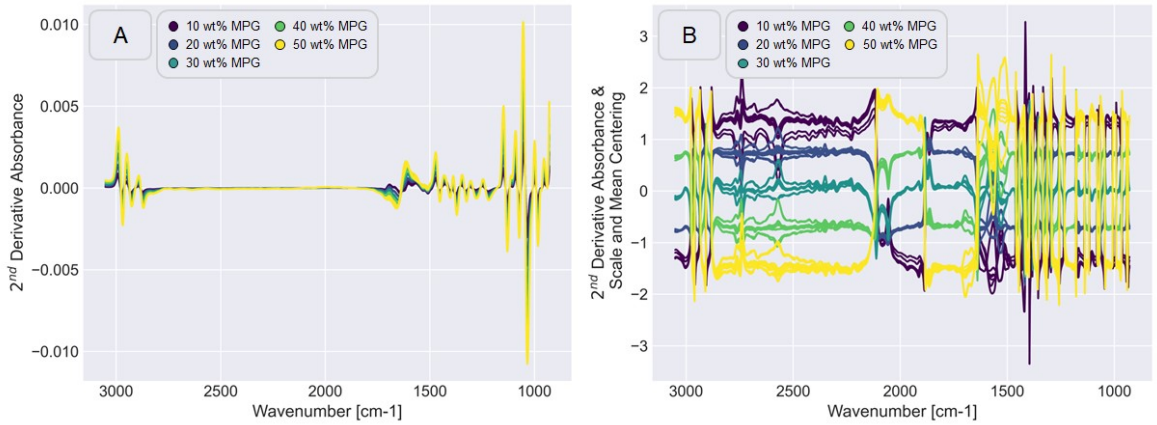


Figure 75: Preprocessing of MIRA-spectra with different endocellulase/MPG formulations. *(A) Savitzky-Golay smoothening and differentiation (5^{th} order polynomial; 1^{st} derivative) and (B) autoscaling.*

The MIRA-spectra were preprocessed by applying different filters and mathematical transformations to reduce the influence of interfering signals and unwanted offsets. In this regard (*fig. 75A*), a SavGol smoothening (5th grade polynomial) and differentiation (1st derivative) filter was applied. In addition, the data was autoscaled (*fig. 75B*), to reduce the influence of spectral regions with high intensity and to remove baseline variations.

Hierarchical Clustering

The hierarchical clustering analysis (*fig. 76*) based on the Euclidian distance algorithm and Ward linkage method provides a first overview about similarity-based differences in between the spectra. The resulting dendrogram illustrates five main clusters corresponding to five MPG concentrations present in the formulations. The five solvent clusters are further subdivided according to the five different sampling timepoints throughout storage. Since samples from the same sampling time presumably have a similar active enzyme content, classification on the basis of enzyme concentration also seems feasible.

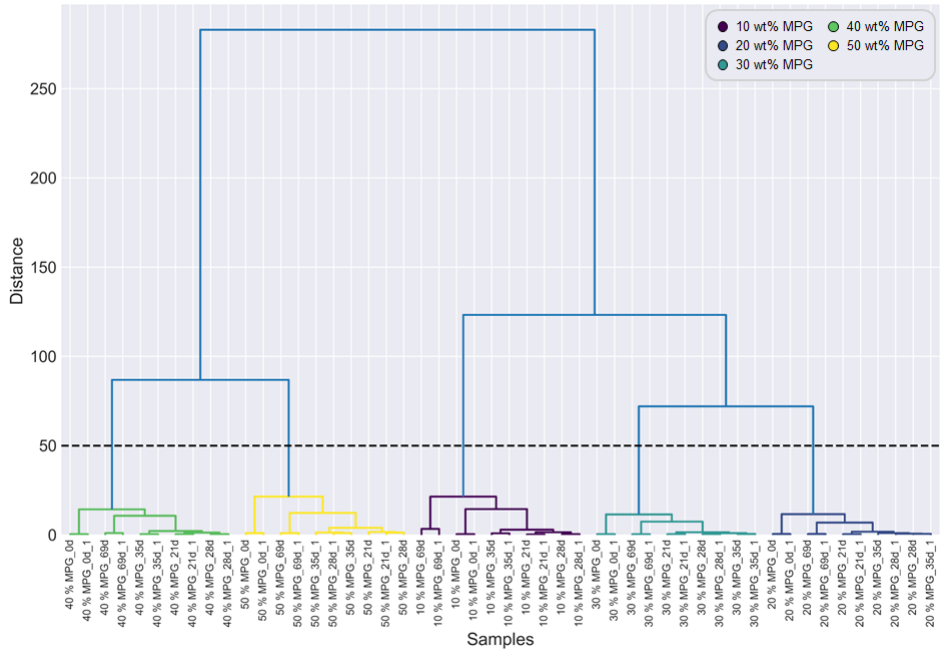


Figure 76: Hierarchical Cluster analysis of preprocessed MIRA-spectra of different endocellulase/MPG formulations. *The five colored clusters are composed of samples with a similar MPG concentration: dark purple= 10 wt% MPG, purple= 20 wt% MPG, blue/green= 30 wt% MPG, green= 40 wt% MPG, yellow= 50 wt% MPG. Subclusters are composed of samples from similar sampling time points and repeats.*

Subsequently, duplicates of each sample are represented by the smallest level of organization in the dendrogram.

The hierarchical cluster analysis provides an overview of the information stored in the spectral data. Based on the formed clusters, differences due to the solvent concentrations appear to be the dominant factor. In addition, information about the sampling timepoints respectively enzyme concentration appears to be present. Accordingly, further statistical analyses were conducted to extract the information of interest from the spectral data.

Principal Component Analysis (PCA)

In order to reduce the dimensionality of the underlying spectral data and to identify the origin of the variability comprised by the spectra a principal component analysis (PCA) was conducted. PCA searches for a set of new variables, the PCs, which capture the majority of the variance in the analyzed data set. The amount of variability described by each PC is illustrated by the scree plot (*fig. 77*). The first principal component (PC1) shows to describe 94.36 % of the total variance present in the spectra. Upon adding the variance of PC2 (+4.15 %) already 98.51 % are comprised by the first two PCs. Consequently, following PCs showed to describe only minor amounts of the total variability comprised by the spectra.

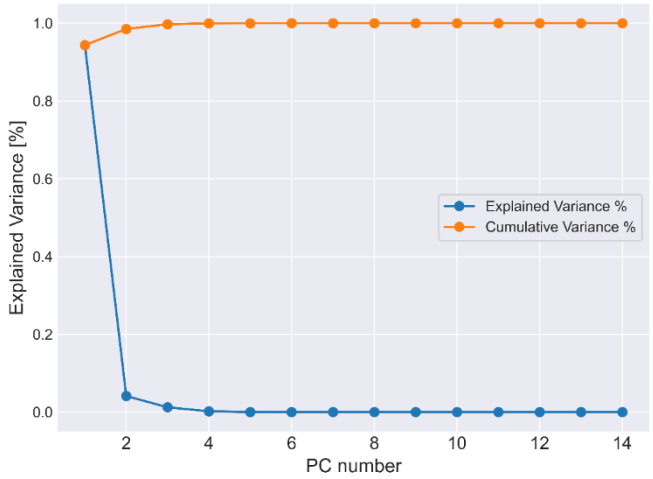


Figure 77: PCA scree plot with the explained and cumulative variance of the underlying dataset.

The scores (*fig. 78*) of the PCA further visualize the relationship of the variability of the samples according to the respective PCs. As can be observed, the resulting scores assigned

for PC1 result in the formation of five distinct groups. The coloring according to the solvent concentration present in these samples leads to the assumption, that similar to the results in *chapter 3.5.1*, the majority of the variance is related to MPG respectively the different solvent concentrations. The classification of the scores according to PC2 show differences in the spread of the datapoints, which leads to the assumption that PC2 could, among other things, comprise the variability associated with changes of the enzyme concentration. However, the shifted positions of the groups to one another and the coloring according to the present enzyme concentration contradict this assumption.

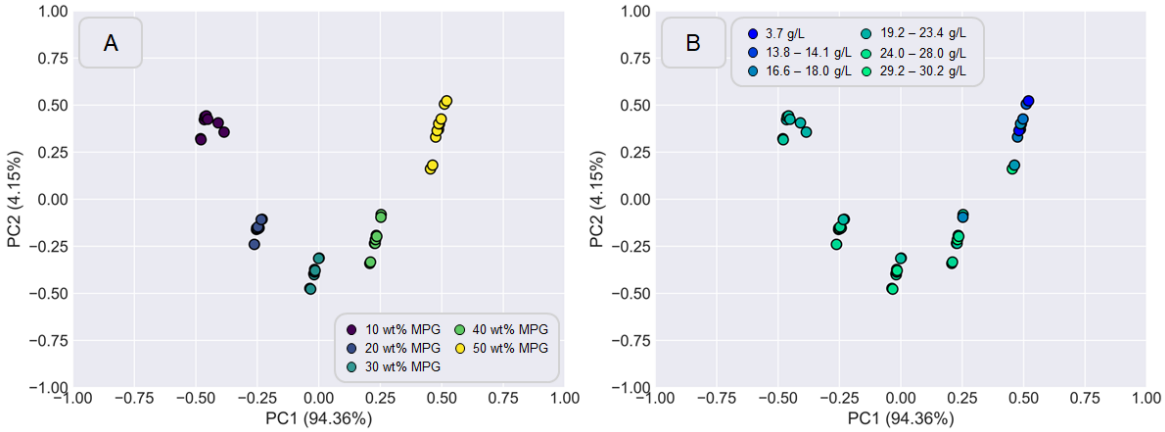


Figure 78: PCA scores plot for the first two PCs colored according to the (A) solvent and (B) enzyme concentration of the subtilisin endocellulase/MPG formulations.

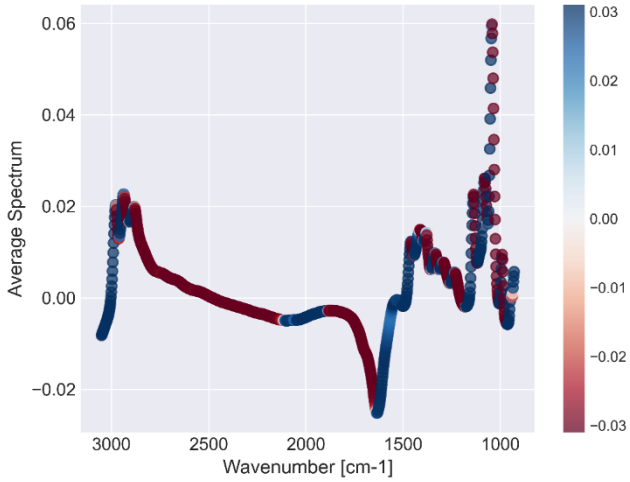


Figure 79: PCA loadings of PC1. *Shown is the average spectrum of the underlying MIRA-spectra, which were color coded according to the magnitude of the present correlations.*

To understand the origin of the observed grouping of the scores according to the solvent concentration for PC1, PCA loadings (*fig. 79*) were determined. The loadings are visualized by coloring the mean spectrum of all samples with their respective values. The colors represent the contribution of the variables to the observed separation of the samples. While the red color represents negative loadings, with a low association to PC1, blue color illustrates positive loadings with a high contribution. The loadings appear to be either highly positively or negatively correlated with PC1. Subsequently, the contribution of the solvent appears to be present in the whole spectrum.

Principle Component Regression (PCR) for Solvent Concentration Prediction

The strong correlation of PC1 with the solvent concentration, was assumed to be sufficient, to predict the solvent concentration of different endocellulase/MPG formulations with a high accuracy. Therefore, a PCR model to predict the solvent concentration of different endocellulase/MPG formulations was created with the variance comprised by PC1 only. The principle of PCR is to predict a response variable based on a set of predictor variables. However, before doing so, the dimensionality of the dataset is reduced by using PCA. Following, the obtained PCs are then used as predictor variables for a regression model. In contrast, to PLS, however, the use of PCA makes PCR an unsupervised method.

Prior to creating the PCR model, the endocellulase dataset was split in a training and test dataset. For this, datapoints obtained after 0, 28 and 69 days were transferred into the training dataset (n= 30) and datapoints obtained after 21 and 28 days into the test dataset (n= 20). The generated results based on this two datasets returned a PCR model (*fig. 80*) with a high accuracy for the solvent concentration prediction. Not only does the calibration (R^2= 0.995; RMSE= 1.11; bias= $5.21*10^{-15}$) and cross-validation (R^2= 0.994; RMSE= 1.25; bias= 0.06) show a high accordance with an ideal correlation, but also the solvent concentrations of the test dataset are predicted (R^2= 0.993; RMSE= 0.82; bias= -0.64) with a high accuracy.

The PCA analysis of the endocellulase/MPG spectra has again highlighted the significant influence of the underlying solvent on the variability comprised in the spectra. In comparison to the same analysis in *chapter 3.5.2,* with the subtilisin protease, the variance caused by the solvent appears to be even larger in the case of the endocellulase formulations than observed before. Subsequently, the created PCR model for solvent concentration prediction exhibits a high accuracy of the results with PC1 only.

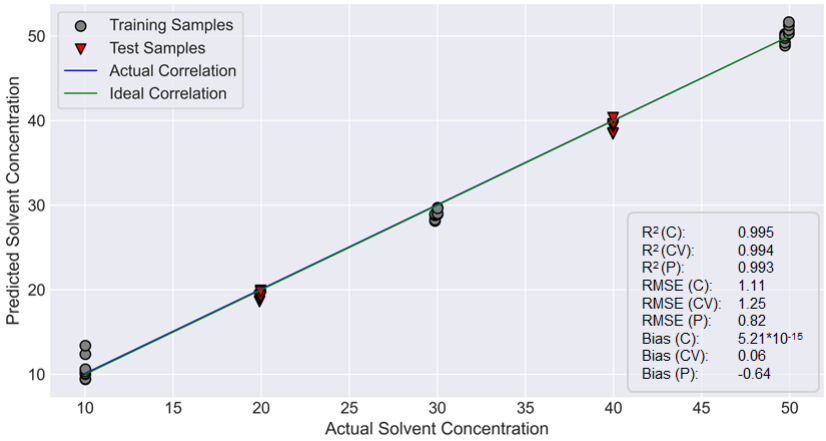


Figure 80: Calibration, cross-validation, and prediction results for the generated PCR model for solvent concentration prediction of different endocellulase/MPG formulations.

Partial Least Squares (PLS) Analysis for Solvent Concentration Prediction

In addition to the generated PCR model for solvent concentration prediction a similar model was to be created based on the supervised partial least square (PLS) analysis. A PLS model is trained based on a set of predictor variables (X-variables) and corresponding response variables (Y-variables). In the ideal case, the resulting PLS model can predict the response variable based on the predictor variable values. Similar to PCA, PLS is optimized to find a set of latent variables, the PLS components, that explain the majority of the variability found, by orthogonalizing the original dataset.

The calculated RMSE (*fig. 81*) of the cross-validation results, for PLS-models with varying numbers of PLS components, returned a high accuracy of the solvent prediction model already with one PLS component (RMSE= 1.22). By further increasing the number of PLS components the RMSE shows to be reduced even further reaching local minima with three (RMSE= 0.58) and eight (RMSE= 0.43) PLS components. However, to prevent overfitting the model and to retain the comparability with the PCR model the first PLS component was selected to be sufficient.

The X-scores (*fig. 82*) for the created PLS solvent model, illustrate a similar grouping of the datapoints, when compared to results of the PCA scores. According to the first PLS component the X-scores show to be clustered in five distinct groups. Further, coloring the datapoints according to the known solvent concentration of the samples, revealed a

grouping according to their inherent MPG concentration. The second PLS component in turn showed no visible correlation with the solvent or enzyme concentration.

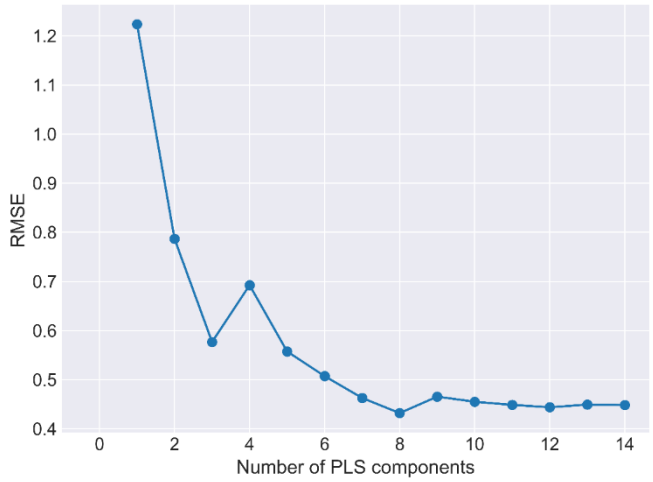


Figure 81: Root mean square error (RMSE) of the cross-validation results with increasing number of PLS components.

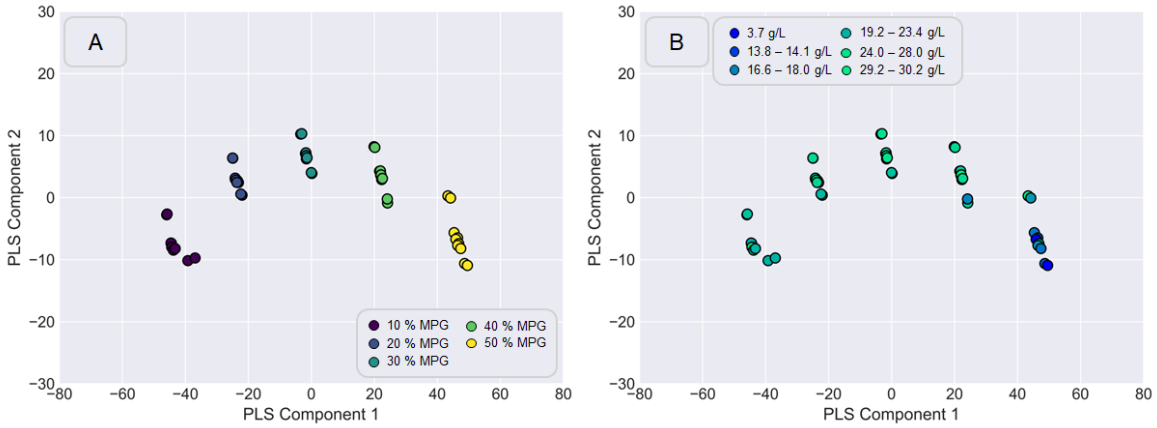


Figure 82: X-scores of the PLS model for the endocellulase/MPG formulations. *The X-scores are colored according to (A) the MPG and (B) enzyme concentration present in the samples according to the DoE.*

The returned VIPs (*fig. 83*) of the PLS model, for the solvent concentration prediction illustrate the X-variables of importance for the prediction of the Y-variables/solvent concentration. By coloring the average spectra of the samples according to the VIPs, regions of high importance for the prediction can be identified. The illustrated results show a high importance of the majority of the variables for the solvent concentration prediction. Basically, all variables of the MIRA-spectrum appear to contribute.

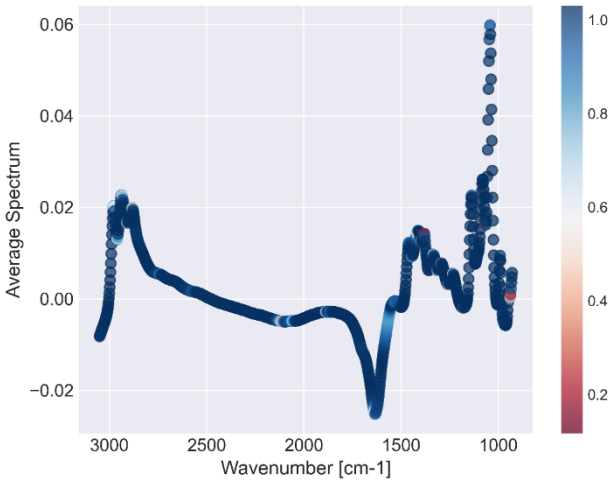


Figure 83: VIPs fort he created PLS model to predict the (A) solvent concentration of different endocellulase/MPG formulations.

Subsequently, the created PLS model (*fig. 84*) for the solvent concentration prediction, returned similar values than observed for the PCR model. For this, the overall data set had been split in a training (n= 30) and a test dataset (n= 20), composed of the samples after 0, 28, 69 days, and 21 and 35 days of storage. The PLS model returned almost the same values for the calibration (R^2= 0.996; RMSE= 1.08; bias= $3.91*10^{-15}$), cross-validation (R^2= 0.994; RMSE= 1.22; bias= 0.5) and predictions (R^2= 0.993; RMSE= 0.79; bias= -0.61) as the unsupervised PCR model.

The obtained results illustrate the high accuracy of the solvent concentration predictions based on the PCR and PLS models. Similar to the results obtained with different subtilisin protease/MPG formulations, the presence of a different enzyme has shown to not affect the quality of predictions. In contrast, to the protease samples, however, the solvent has gained significant influence on the overall variability encoded by the spectrum. This could negatively affect enzyme concentrations predictions, since the loadings and variables of importance are basically spread over the whole range of the spectrum. Similarly, the region relevant for protein quantification, in the range from 1500-1750 cm^{-1}, appears to be affected. Subsequently, cross-correlation and a possible multicollinearity of the enzyme concentration predictions with the solvent concentration had to be considered in following experiments.

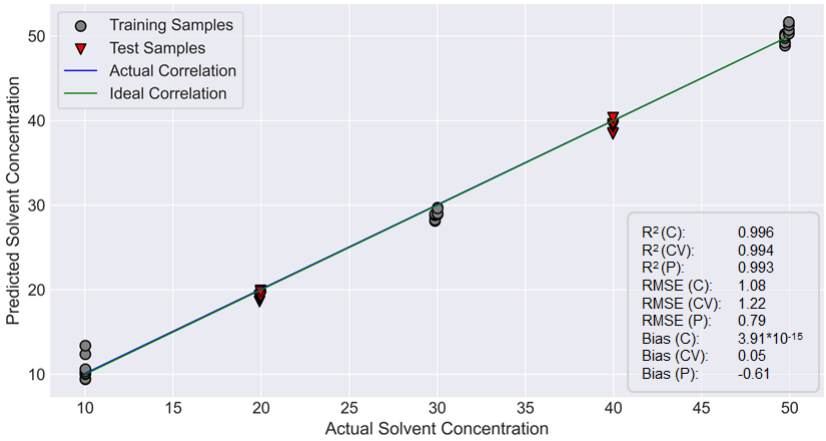


Figure 84: Calibration, cross-validation, and prediction results for the generated PLS model for solvent concentration prediction of different endocellulase/MPG formulations.

3.6.5 Endocellulase/MPG Formulations for Enzyme Concentration Prediction

The results from *chapter 3.5.5* had demonstrated the significant amount influence of the underlying solvent concentration on the observed variability in the MIRA-spectra of different endocellulase/MPG formulations. The solvent subtraction approach from *chapter 3.5.2* had shown to greatly diminish the variability caused by the solvent but could not totally remove its influence on the enzyme concentration predictions. In this regard, prior to the PCR and PLS model creation, for endocellulase concentration prediction, an external parameter orthogonalization method (EPO) was to be applied. Thereby, the variability caused by the solvent was to be removed from the MIRA-spectra and the accuracy of the predictions was to be increased.

For the analysis, the same endocellulase/MPG formulations were used, as for the solvent concentration predictions in *chapter 3.5.4*. The experimental setup consisted of different samples with ≈30 g/L endocellulase and 10-50 wt% MPG (10 wt% interval). The samples were stored at 45 °C, over period 69 days, whereby samples were withdrawn after 0, 21, 28, 35 and 69 days. Besides the obtained MIRA spectra, the residual active enzyme concentration was determined using the pNPL-assay on the Gallery™ system. Following, PCR and PLS models to predict the endocellulase concentration in different formulations were generated based on the acquired MIRA-spectra and the determined active enzyme concentration.

The result of the long-term storage test (*fig. 85A*) returned a varying influence of the MPG concentration on the stability of the studied endocellulase. In addition, the absence of autolysis showed to lead to minor losses of enzyme activity throughout storage. The endocellulase exhibited a good overall stability in the presence of 10 wt% MPG (69 d; 21.8 g/L). The stability shows to increase with concentrations of 20 (69 d; 22.6 g/L) and 30 wt% MPG (69 d; 23.4 g/L). However, in the presence of 40 and 50 wt% MPG the active enzyme content drastically declines and results in an average concentration of 17.3 respectively 3.7 g/L after 69 days.

In contrast, only minor changes in the determined MIRA-spectra (*fig. 85B*) were visible over the sampled time period, with regard to the enzyme concentration. Especially, in the range of 1500-1750 cm^{-1} spectra appeared to be hardly altered. In the absence of autolysis, unfolding of the enzyme is expected to be major reason for activity loss. In contrast to autolysis, unfolding of the protein structure is not expected to be accompanied by the loss of peptide bonds. Therefore, visible changes of the protein spectrum are expected to be mainly due to structural changes of the secondary and tertiary structure of the endocellulase, which are, however, expected to be less significant in the MIRA spectrum than the loss of peptide bonds.

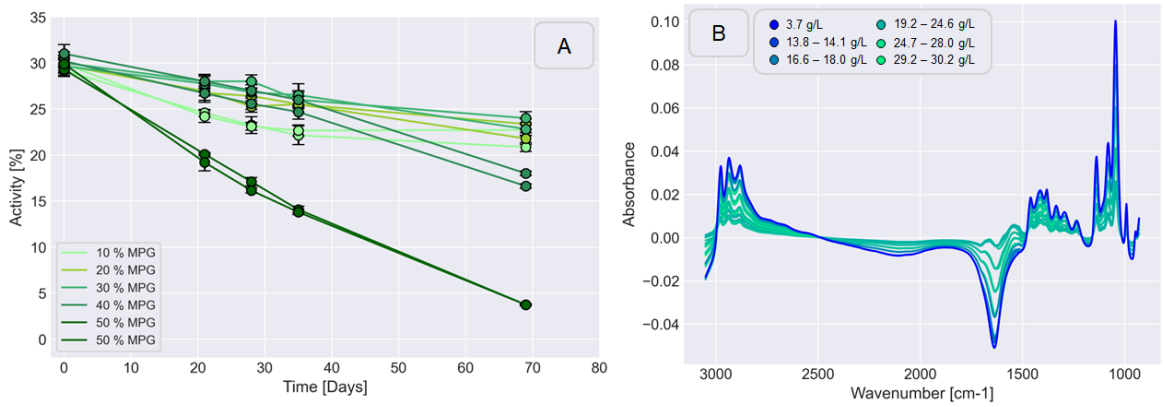


Figure 85: (A) Residual active enzyme concentrations determined based on the pNPL-assay and (B) and MIRA-spectra for the endocellulase/MPG formulations. *Samples were drawn after 0, 21, 28, 35, and 69 days, after storage at 45 °C. The formulations contained 10-50 wt% MPG.*

Preprocessing

In order to reduce the influence of MPG on the underlying spectra (*fig. 86A*), the dataset was transformed using EPO. EPO can be used as preprocessing method to exclude the

influence of one or more PCs from the analysis prior to model building process. To do so, EPO performs a PCA on the predictor variables, to identify PCs to exclude from the analysis. Following, respective columns in the matrix of eigenvectors are removed and the remaining predictor variables are transformed into orthogonal components using the modified matrix of eigenvectors. This has the effect, that the influence of certain predictor variables is removed from the dataset, which leads to improved accuracy and interpretability of the results by reducing the effects of multicollinearity and confounding.

Based on the PCA analysis conducted in *chapter 3.5.5* the variability of the first three principal components was removed from the spectral data. During this process the clutter mean was to be ignored and the mean of the spectra was to be maintained.

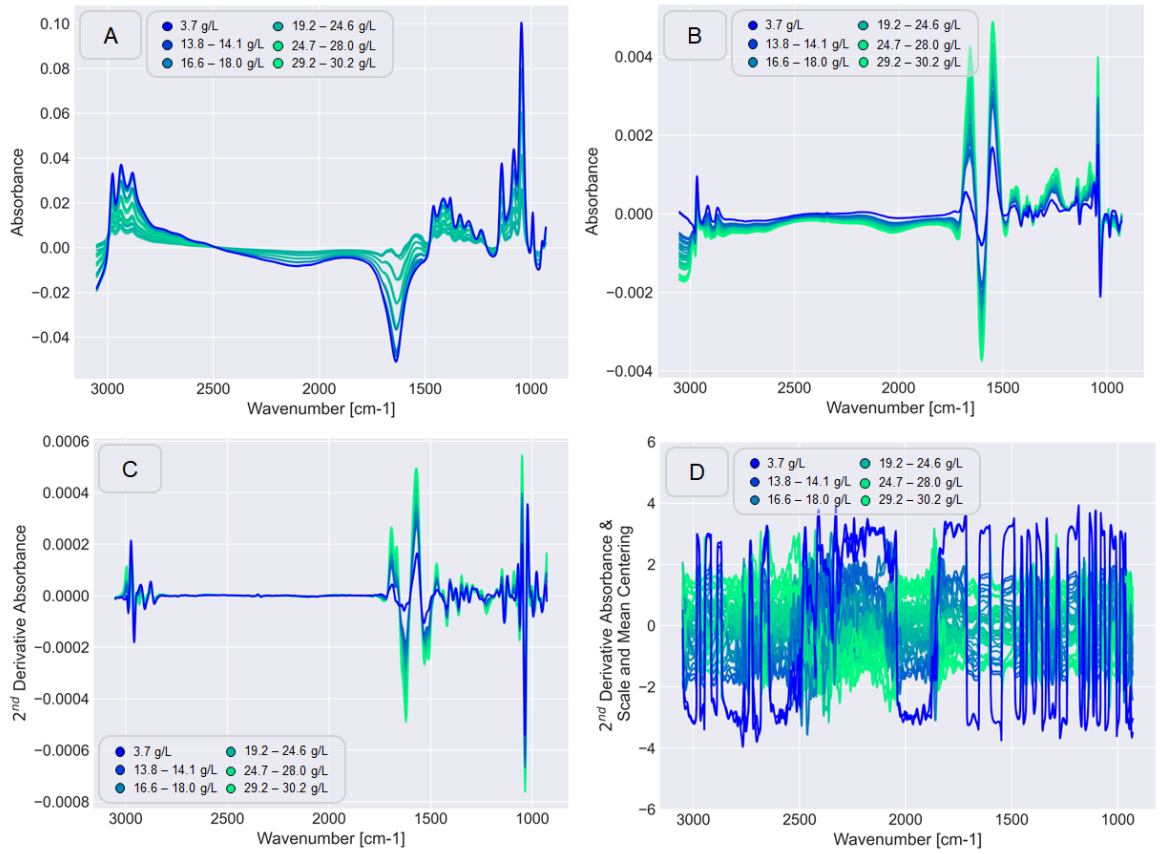


Figure 86: Endocellulase/MPG formulation MIRA-spectra before and after the different preprocessing steps. *(A) Unprocessed MIRA-spectra obtained for the individual samples. (B) Pseudo-spectrum obtained after applying EPO. (C) Resulting pseudo-spectrum after applying SavGol smoothing (5th grad polynomial) and differentiation (1st derivative). (D) Pseudo-spectrum after autoscaling the spectral data.*

The returned result of the EPO (*fig. 86B*) is a pseudo spectrum which was used for further analysis. Interestingly, the reduced spectral data illustrated two dominant bands in the region of 1500-1750 cm^{-1}, which resembled that amide I and II bands after solvent subtraction in *chapter 3.5.2*. Furthermore, the intensity of the absorbance values had significantly decreased, when compared to the original spectrum.

The reduced dataset was further processed by applying SavGol smoothening (*fig. 86C*) (5^{th} polynomial) and differentiation (*fig. 86D*) (1^{st} derivative). In addition, the data was autoscaled.

Hierarchical Clustering

The conducted hierarchical cluster analysis (*fig. 87*), based on the Euclidian distance algorithm and Ward linkage method, reveals several smaller and bigger clusters, which are composed of different samples, of varying composition and age. Even though, some of the clusters appear to be composed of formulations with identical solvent concentrations, others comprise arbitrary samples, which could not be explained by solvent or enzyme concentration. Consequently, the hierarchical cluster analysis did not provide a classification according to enzyme or solvent concentration.

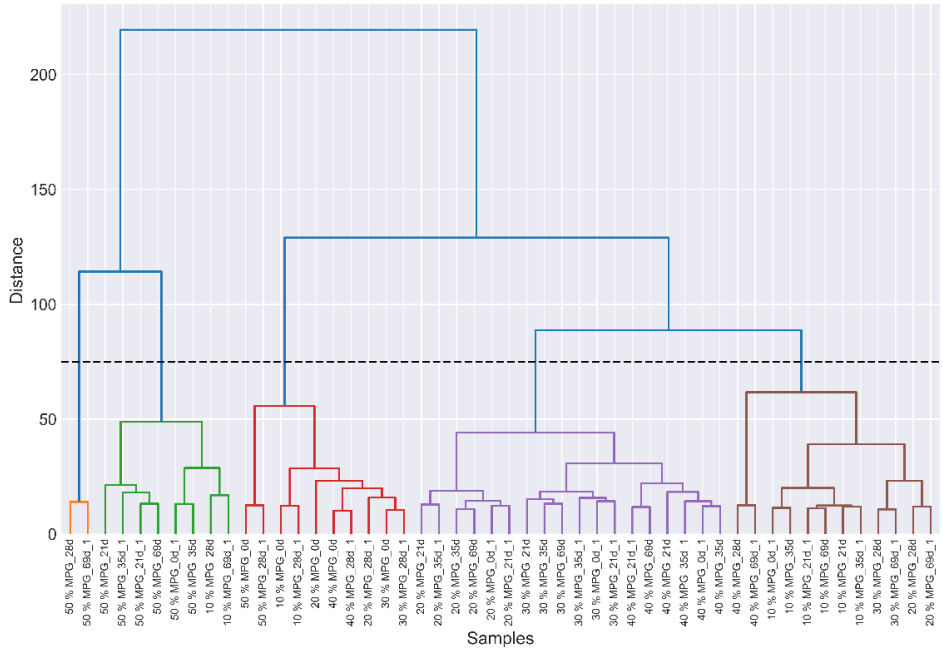


Figure 87: Dendrogram of the hierarchical cluster analysis performed with the preprocessed pseudo-spectra of the endocellulase/MPG formulations at different timepoints throughout storge.

Principle Component Analysis (PCA)

The consecutive PCA revealed that the majority of the variability (*fig. 88*) in the pseudo spectra, is described by the first PC, which covers 75.34 % of the total variance. In contrast, to former analyses the inclusion of additional parameters only moderately increases the cumulative variance explained. For the further analysis, the first three PCs were considered, which account for 88.53 % (PC2: +9.83 %; PC3: +3.35 %) of the total variance.

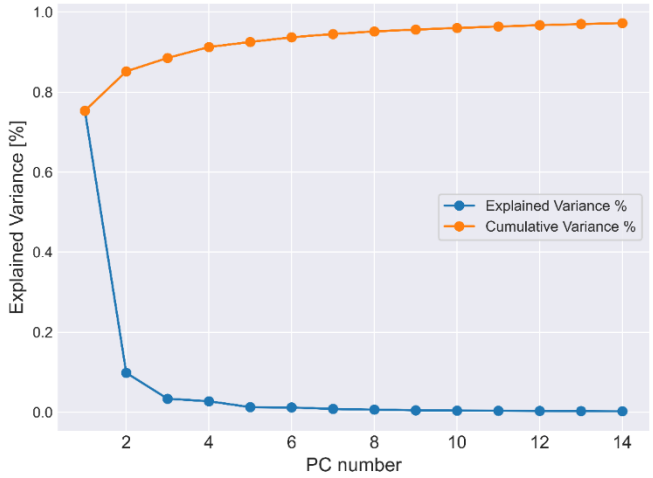


Figure 88: PCA scree plot of the endocellulase/MPG pseudo-spectra after preprocessing.

Unlike the hierarchical cluster analysis, the PCA scores (*fig. 89*) gave additional insights on the variability and differences described by the pseudo spectra. The scores plots of PC1-PC3 illustrate that the variability described by PC1 results in a grouping according to the active enzyme concentration present in the different samples. This is illustrated by the coloring of the sample scores according to their inherent enzyme concentration. The resulting trend shows that negative score values for PC1 appear to describe high enzyme concentrations, shown in green, while high scores values seemed match to samples with low enzyme concentrations, shown in blue. In contrast to that PC2 and PC3 show no obvious or further correlation with the parameters of interest, even though PC2 and PC3 appear to increase the spread in between samples of certain groups.

The loadings (*fig. 90*) of PC1-3 illustrate the contribution of each pseudo spectra variable to the variability of the different PCs. As was to be expected, based on the amount of explained variance, PC1 shows to include the largest proportion of the variables in the spectrum.

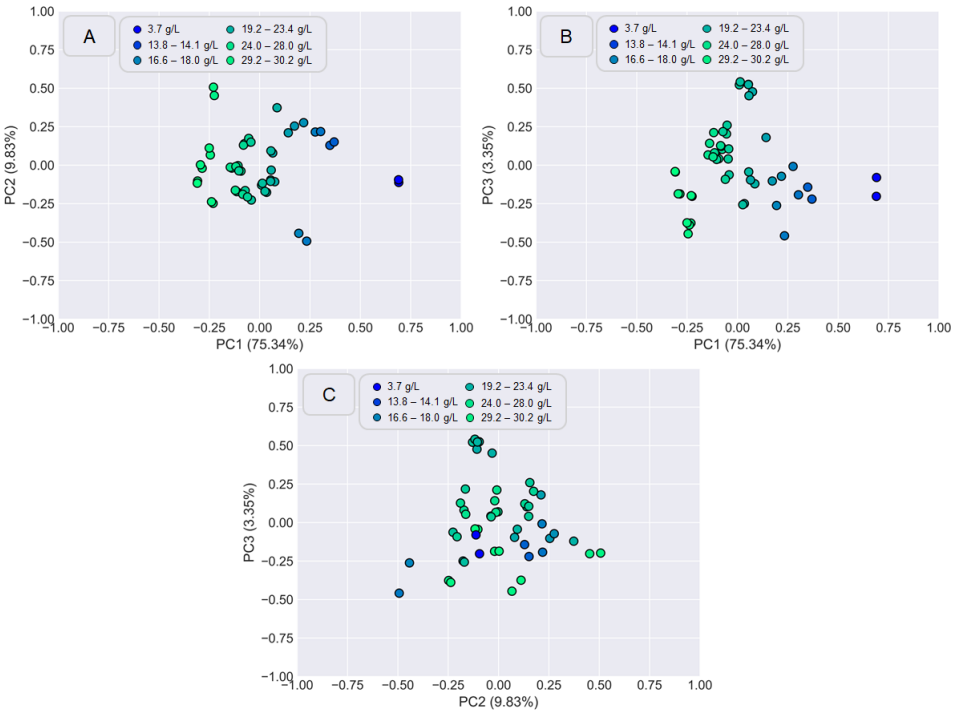


Figure 89: PCA scores plot for PC1-3 colored according to the endocellulase concentration in the samples. *The single plots illustrate the relationship between (A) PC1 and 2, (B) PC1 and 3, and (C) P2 and 3.*

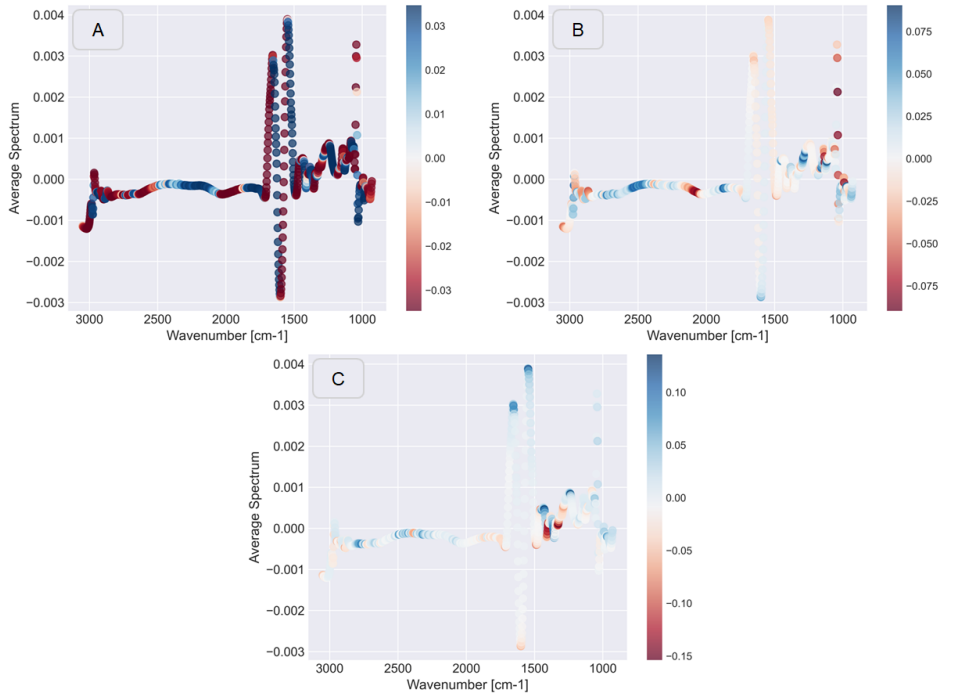


Figure 90: PCA loadings for (A) PC1, (B) PC2, and (C) PC3. *Shown is the average pseudo spectrum after EPO colored according to magnitude of correlation*

In this context, the majority of the datapoints appear to have a strong negative or positive correlation with the variability described by PC1. Subsequently, as PC2 and PC3 describe minor proportions of the variability, only few variables show a low positive/negative correlation with the variability described by these PCs.

Principle Component Regression (PCR)

Before creating the PCR model, the data comprising the pseudo-spectra of all samples was split in a training (n= 30) and a test dataset (n= 20). The training set comprised the datapoints acquired after 0, 28 and 69 days and the test dataset contained the datapoints after 21 and 28 days. The PCR model (*fig. 91*) based on the first three PCs returned an accurate model with a small deviation from the ideal correlation for low enzyme concentrations. In general, however, calibration (R^2= 0.988; RMSE= 0.74; bias= $2.09*10^{-15}$) as well as cross-validation (R^2= 0.984; RMSE= 0.85; bias= 0.027) return precise values for the training data set. In addition, the prediction (R^2= 0.960; RMSE= 0.72; bias= -0.16) of the enzyme content in the test dataset, returns similarly accurate values.

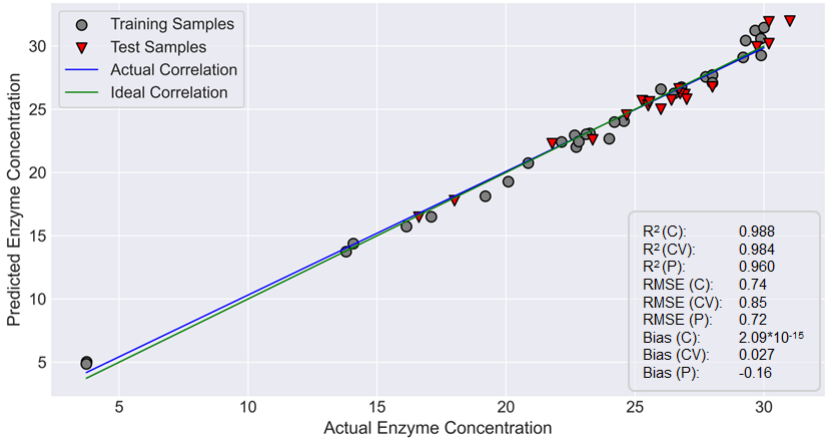


Figure 91: Calibration, cross-validation, and prediction results for the generated PCR model for enzyme concentration prediction of different endocellulase/polyol formulations.

To exclude the possibility of cross correlations and multicollinearity between the enzyme and solvent concentrations predictions a similar PCR model was created based on the solvent data. As could be illustrated (*fig. 92*), however, the model returned, was not able to accurately predict the solvent concentration of the different samples. Subsequently, it was demonstrated that the created PCR model for enzyme concentration prediction is not

influenced, or the result of cross-correlations or multicollinearity caused by the solvent variability in the pseudo spectra. This further demonstrates that the applied EPO has successfully removed and thereby reduced the variability attributed to the solvent.

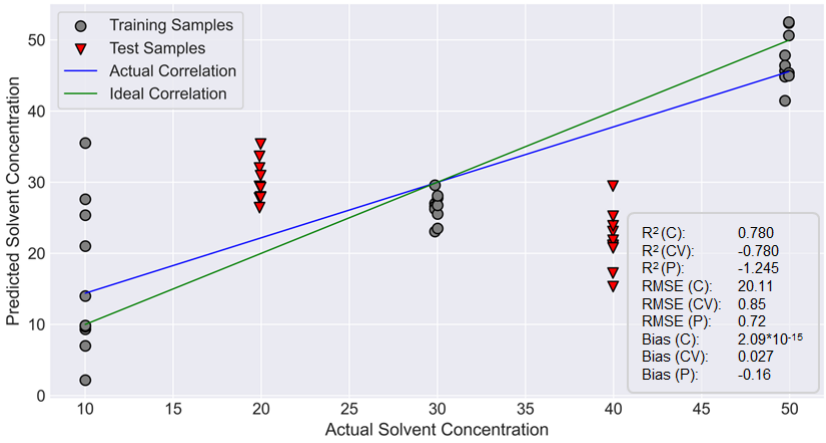


Figure 92: Calibration, cross-validation, and prediction results for the generated PCR model for solvent concentration prediction of different endocellulase/polyol formulations.

Partial Least Squares (PLS) Analysis

In addition to the PCR model for enzyme concentration prediction, a PLS model was to be created. The RMSE (*fig. 93*) of the cross validation returned a small error for the first PLS component (RMSE= 0.96). In accordance with the created PCR model, three PLS components were selected for model creation, which further decreased the RMSE to 0.61.

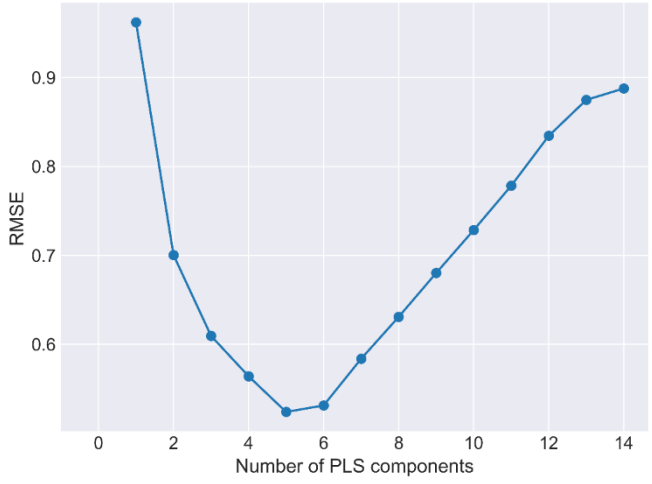


Figure 93: Root mean square error (RMSE) of the cross-validation with increasing number of PLS components.

The X-scores (*fig. 94*) for the whole dataset exhibit a similar result as illustrated by the PCA scores. According to the first PLS component, the datapoints show a grouping according to the inherent enzyme concentration of the samples. The correlation between the enzyme concentration and the first PLS component can be illustrated by coloring the datapoints according to the solvent concentration of the samples. For the second and third PLS component, however, no visible correlation between the datapoints and enzyme or solvent concentration can be observed.

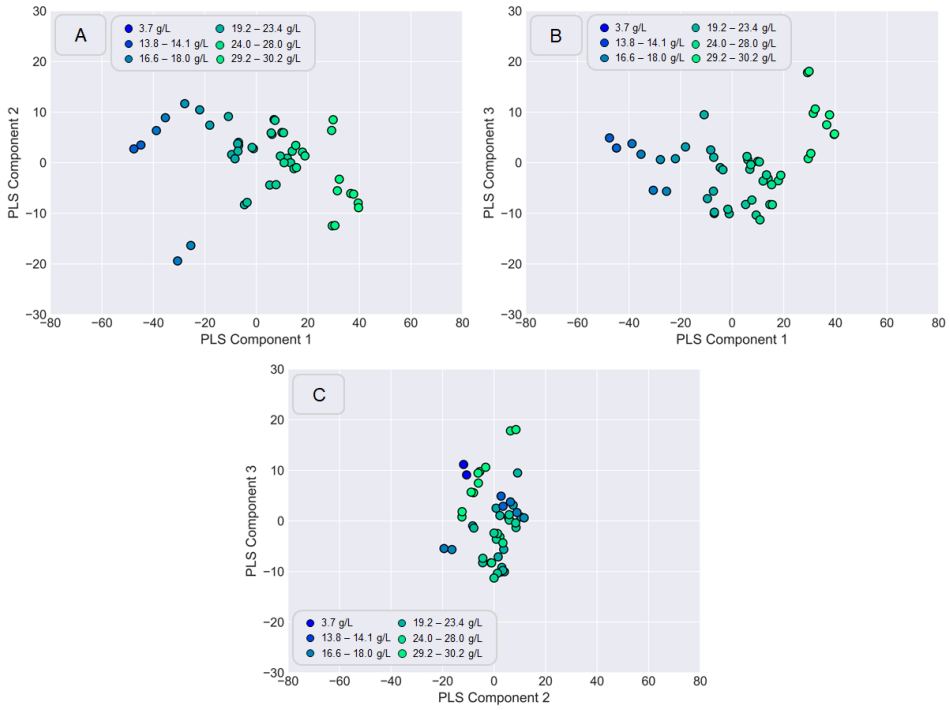


Figure 94: X-scores of the PLS model for the endocellulase/MPG formulations. *The X-scores are colored according to the enzyme concentration present in the samples according to the DoE.*

The VIPs (*fig. 95*) for the created PLS model, show a great importance of the majority of regions in the pseudo spectra for the prediction of the enzyme concentration. Not only does the region associated with amide I and II show a high importance of the variables, but also other variables with higher or lower wavenumber appear to contribute. A minor exception, however, was the region from 2000-2750 cm^{-1}. In this area, several variables exhibit only a minor importance for the prediction of the enzyme concentration. These region, however, shows only minor variations in between the different pseudo spectra. Following, it can be assessed that due to a low information content and a chemical perspective this region is not excessively important for the prediction of the enzyme concentration.

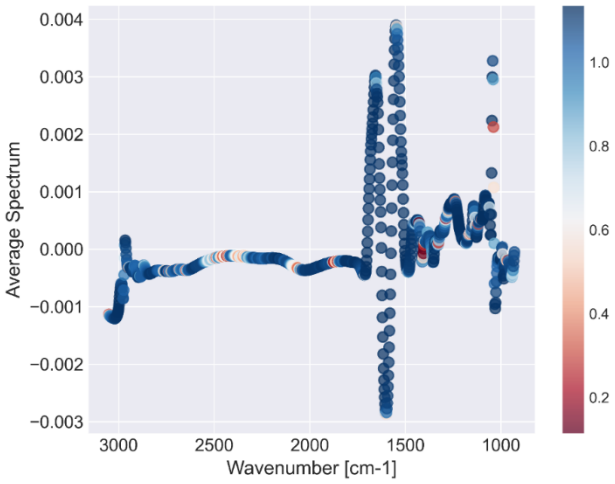


Figure 95: VIPs fort he created PLS model to predict the enzyme concentration of different endocellulase/MPG formulations.

The created PLS-model (*fig. 96*) for enzyme concentration prediction, illustrates an improved prediction accuracy, when compared to the PCR model. Especially, at low enzyme concentrations the offset between the actual and ideal correlations is not visible. Subsequently, calibration (R^2= 0.996; RMSE= 0.45; bias= $1.5*10^{-15}$) and cross-validation (R^2= 0.991; RMSE= 0.61; bias= 0.008) exhibit a high accuracy for the samples in the training dataset. In addition, predictions (R^2= 0.949; RMSE= 0.81; bias= 0.03) of the enzyme concentration of samples in the test dataset return results with a similar high accuracy.

Similar as for the PCR model, the ability of the created PLS model (*fig. 97*) to predict solvent concentrations based on the pseudo spectra was to be determined. As before, the solvent prediction model illustrates a poor ability to do so. Therefore, it can be concluded that by applying the EPO the variability caused by the solvent could greatly be diminished by removing the first three PCs from the dataset. The resulting PCR and PLS models exhibit a high accuracy for the prediction of the enzyme concentration. The developed workflow for the prediction of the endocellulase concentration based on EPO has proven itself to be a convenient method to reduce the variability caused by the solvent. In contrast to the subtraction of the solvent spectrum, which was used in *chapter 3.5.2-3*, the use of EPO is not reliant on pure solvent spectra, which makes it easy to apply and universally useable for different enzyme formulations.

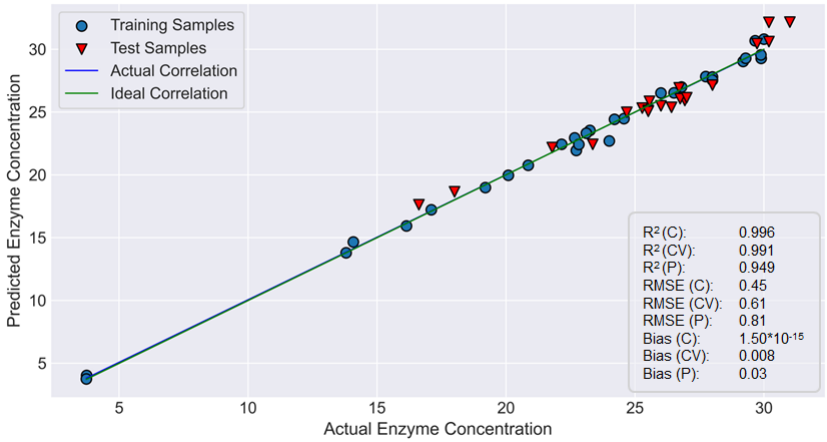


Figure 96: Calibration, cross-validation, and prediction results for the generated PLS model for enzyme concentration prediction of different endocellulase/MPG formulations.

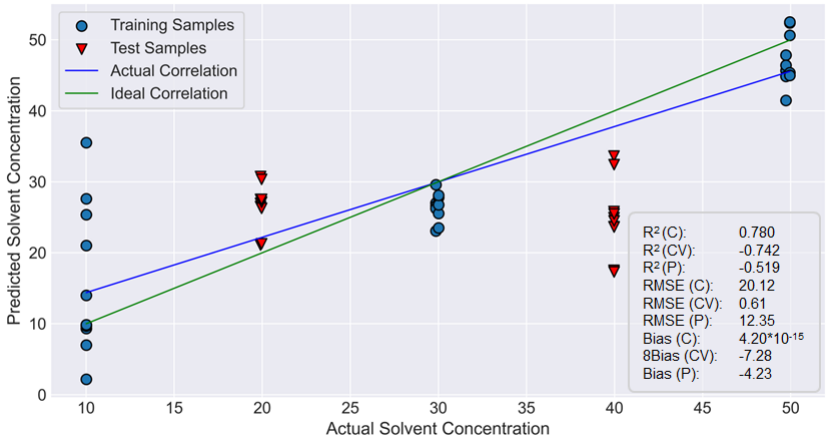


Figure 97: Calibration, cross-validation, and prediction results for the generated PLS model for solvent concentration prediction of different endocellulase/MPG formulations.

3.6.6 Summary & Conclusion

Infrared light (IR) can be absorbed by molecular vibrations when the frequencies of light and vibrations match. The frequency of the vibrations and the probability of the absorption depend on the strength and polarity of the vibrating bonds and are thus influenced by intra and intermolecular effects [171, 244]. Thus, manifold of information about the structure and environment of a sample, molecule or solute can be obtained from spectral parameters, as the band position, bandwidth, and absorption coefficient. Subsequently, IR spectroscopy is a commonly used technique for the structural characterization of small molecules. This

is promoted by its sensitivity, large application range, high time resolution and low sample volume required for the analysis.

This unique combination of properties has promoted the transfer of IR to biological fields of application. Besides, its common use for protein secondary structure elucidation [244, 249], it is used for the identification of molecular mechanisms [250, 251] and to study protein folding, unfolding and misfolding [252-254]. Secondary structure analysis is conducted based on the amide I band. However, in the past also amide II and III, as well as the near infrared region have been used. The amide I vibration, absorbing near 1650 cm^{-1}, arise from the C=O stretching vibration with minor contributions from CN stretching vibrations, the CCN deformation and the NH in-plane bending [171]. The conformation of the amide I is highly dependent on the secondary structure of the backbone and is therefore commonly used for secondary structure analysis. Subsequently, changes of the protein structure due to the loss of peptide bonds by proteolysis or thermal unfolding and thereby the loss of active enzyme are expected to be visible in mid infrared (MIR) spectroscopy.

The analysis of aqueous samples by MIR spectroscopy represents a certain challenge in this context, as the strong absorbance of water in the mid-infrared spectral region (near 1645 cm^{-1}), overlap with the amide I band of proteins and some side chain bands [255, 256]. These limitations are to be overcome with the mid infrared analyzer (MIRA) which uses a combination of an ultra-thin flow cell (7-10 µm) with AI driven approaches to correct the obtained spectral data and provide a digital twin solution. In combination with chemometrics, this method is a valuable methodology for the identification and quantification of known and unknown substances in solution.

The developed workflow thereby illustrates the possible application of this technique for the quantification of the active enzyme content in different liquid formulations. In the past, IR spectroscopy was already extensively used to determined enzyme activity in biotechnological processes. These approaches utilize the ability of IR measurement to provide an “on-line” monitor of enzymatic reactions, as educt and product concentrations can be monitored. In this context, the developed workflow for a subtilisin protease and endocellulase provides an addition to quantify the active enzyme content directly in solution. The proposed approach is independent of artificial substrates and needs no extensive sample dilution prior to the analysis. Thereby, analysis time and costs are significantly reduced, and the accuracy of the measured results is expected to be increased. In combination with multivariate analysis techniques additional formulation ingredients

can be quantified, whereby the information content of a single measurement is further increased.

To further extend the use of the MIRA technology in combination with chemometrics, future approaches will focus on the applicability of the developed workflow in the context of a more diverse formulation matrix. Besides, increasing the complexity in the presence of more than one solvent, the portability of the developed models to other enzyme classes must be assessed. In this context, the presented approach may provide a serious and universally applicable alternative to common photometric assays, which are in many cases limited to certain groups of enzymes.

Literature

1. Jegannathan, K.R. and P.H. Nielsen, *Environmental assessment of enzyme use in industrial production – a literature review.* Journal of Cleaner Production, 2013. **42**: p. 228-240.
2. Kirk, O., T.V. Borchert, and C.C. Fuglsang, *Industrial enzyme applications.* Curr Opin Biotechnol, 2002. **13**(4): p. 345-51.
3. Soetaert, W. and E.J. Vandamme, *Industrial Biotechnology: Sustainable Growth and Economic Success*. 2010: Wiley.
4. Development, O.f.E.C.-O.a., *Biotechnology for Clean Industrial Products and Processes.* OECD Publications, 1998.
5. Press, N.A., *Industrialization of Biology: A Roadmap to Accelerate the Advanced Manufacturing of Chemicals*. 2015: National Acadmies Press (US).
6. Heckmann, C.M. and F. Paradisi, *Looking Back: A Short History of the Discovery of Enzymes and How They Became Powerful Chemical Tools.* ChemCatChem, 2020. **12**(24): p. 6082-6102.
7. Polaina, J. and A.P. MacCabe, *Industrial Enzymes: Structure, Function and Applications*. 2007: Springer Netherlands.
8. Fruton, J.S., *A history of pepsin and related enzymes.* Q Rev Biol, 2002. **77**(2): p. 127-47.
9. de Réaumur, R.A.F., *Observations sur la digestion des oiseaux.* Histoire de l'Académie Royale des Sciences, 1752. **266**: p. 461.
10. Wisniak, J., *The History of Catalysis. From the Beginning to Nobel Prizes.* Educación Química, 2010. **21**(1): p. 60-69.
11. Manoochehri, H., et al., *A review on invertase: Its potentials and applications.* Biocatalysis and Agricultural Biotechnology, 2020. **25**.
12. Maicas, S., *The Role of Yeasts in Fermentation Processes.* Microorganisms, 2020. **8**(8).
13. Robinson, P.K., *Enzymes: principles and biotechnological applications.* Essays Biochem, 2015. **59**: p. 1-41.
14. Buchner, E., *Alkoholische Gährung ohne Hefezellen.* Berichte der deutschen chemischen Gesellschaft, 2006. **30**(1): p. 117-124.
15. Garg, S.K. and B.N. Johri, *Rennet: Current trends and future research.* Food Reviews International, 1994. **10**(3): p. 313-355.
16. Sarmiento, F., R. Peralta, and J.M. Blamey, *Cold and Hot Extremozymes: Industrial Relevance and Current Trends.* Front Bioeng Biotechnol, 2015. **3**: p. 148.
17. Simoni, R.D., R.L. Hill, and M. Vaughan, *Urease, the First Crystalline Enzyme and the Proof That Enzymes Are Proteins: the Work of James B. Sumner.* Journal of Biological Chemistry, 2002. **277**(35): p. e1-e2.
18. Schellman, J.A. and C.G. Schellman, *Kaj Ulrik Linderstrom-Lang (1896-1959).* Protein Sci, 1997. **6**(5): p. 1092-100.
19. Arbige, M., *Industrial enzymology: a look towards the future.* Trends in Biotechnology, 1989. **7**(12): p. 330-335.
20. Olsen, H.S. and P. Falholt, *The role of enzymes in modern detergency.* Journal of Surfactants and Detergents, 1998. **1**(4): p. 555-567.
21. Cherry, J.R. and A.L. Fidantsef, *Directed evolution of industrial enzymes: an update.* Curr Opin Biotechnol, 2003. **14**(4): p. 438-43.
22. Demain, A.L. and S. Sánchez, *Enzymes of industrial interest.* Mexican journal of biotechnology, 2017. **2**(2): p. 74-97.
23. Sanger, F., *The arrangement of amino acids in proteins.* Adv Protein Chem, 1952. **7**: p. 1-67.
24. Fruton, J.S., *Early theories of protein structure.* Ann N Y Acad Sci, 1979. **325**: p. xiv, 1-18.
25. Pauling, L. and R.B. Corey, *The pleated sheet, a new layer configuration of polypeptide chains.* Proc Natl Acad Sci U S A, 1951. **37**(5): p. 251-6.
26. Pauling, L., R.B. Corey, and H.R. Branson, *The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain.* Proc Natl Acad Sci U S A, 1951. **37**(4): p. 205-11.
27. Sun, P.D., C.E. Foster, and J.C. Boyington, *Overview of protein structural and functional folds.* Curr Protoc Protein Sci, 2004. **Chapter 17**(1): p. Unit 17 1.
28. Wilmot, C.M. and J.M. Thornton, *Analysis and prediction of the different types of beta-turn in proteins.* J Mol Biol, 1988. **203**(1): p. 221-32.

29. Fetrow, J.S., *Omega loops; nonregular secondary structures significant in protein function and stability.* The FASEB Journal, 1995. **9**(9): p. 708-717.
30. Eisenberg, D., *The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins.* Proc Natl Acad Sci U S A, 2003. **100**(20): p. 11207-10.
31. Branden, C.I., & Tooze, J. , *Introduction to Protein Structure (2nd ed.)*. Garland Science. 1998.
32. Dunitz, J.D., *Pauling's Left-Handed alpha-Helix.* Angew Chem Int Ed Engl, 2001. **40**(22): p. 4167-4173.
33. Shepherd, N.E., et al., *Left- and right-handed alpha-helical turns in homo- and hetero-chiral helical scaffolds.* J Am Chem Soc, 2009. **131**(43): p. 15877-86.
34. Vojisavljevic, V., et al., *Hybrid approach to analysis of β-sheet structures based on signal processing and statistical consideration.* Proceedings: Mathematical, Physical and Engineering Sciences, 2011. **467**(2128): p. 1052-1072.
35. Jeremy M. Berg , J.L.T., Gregory J. Gatto jr. , Lubert Stryer, *Stryer Biochemie*. Vol. 8. 2017: Springer-Verlag GmbH
36. Cheng, P.N., J.D. Pham, and J.S. Nowick, *The supramolecular chemistry of beta-sheets.* J Am Chem Soc, 2013. **135**(15): p. 5477-92.
37. Anfinsen, C.B., *Principles that govern the folding of protein chains.* Science, 1973. **181**(4096): p. 223-30.
38. Baker, D. and D.A. Agard, *Kinetics versus thermodynamics in protein folding.* Biochemistry, 1994. **33**(24): p. 7505-9.
39. Bisswanger, H., *Enzyme Kinetics: Principles and Methods, Third, enlarged and improved Edition*. 2017: Wiley-VCH Verlag GmbH & Co. KGaA
40. Kiss, G., et al., *Computational enzyme design.* Angew Chem Int Ed Engl, 2013. **52**(22): p. 5700-25.
41. Heinrich, P.C., M. Müller, and L. Graeve, *Löffler/Petrides Biochemie und Pathobiochemie*. 2014: Springer Berlin Heidelberg.
42. Palmer, T. and P.L. Bonner, *Enzymes: Biochemistry, Biotechnology, Clinical Chemistry*. 2007: Elsevier Science.
43. Liang, J., H. Edelsbrunner, and C. Woodward, *Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design.* Protein Sci, 1998. **7**(9): p. 1884-97.
44. Copeland, R.A., *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*. 2023: Wiley.
45. Cramer, F., *Biochemical correctness: Emil Fischer's lock and key hypothesis, a hundred years after — an essay.* Pharmaceutica Acta Helvetiae, 1995. **69**(4): p. 193-203.
46. Koshland, D.E., Jr., *Correlation of Structure and Function in Enzyme Action.* Science, 1963. **142**(3599): p. 1533-41.
47. Khersonsky, O. and D.S. Tawfik, *Enzyme promiscuity: a mechanistic and evolutionary perspective.* Annu Rev Biochem, 2010. **79**: p. 471-505.
48. Nam, H., et al., *Network context and selection in the evolution to enzyme specificity.* Science, 2012. **337**(6098): p. 1101-4.
49. Skarydova, L., et al., *Enzyme stereospecificity as a powerful tool in searching for new enzymes.* Curr Drug Metab, 2010. **11**(6): p. 547-59.
50. Rose, I.A., *Enzyme reaction stereospecificity: a critical review.* CRC Crit Rev Biochem, 1972. **1**(1): p. 33-57.
51. Hanson, K.R., *Enzyme Symmetry and Enzyme Stereospecificity.* Annual Review of Plant Physiology, 1972. **23**(1): p. 335-366.
52. Boyer, P.D., *The Enzymes*. 1971: Elsevier Science.
53. Russell, D.W., *The enzymes, regulation, and genetics of bile acid synthesis.* Annu Rev Biochem, 2003. **72**: p. 137-74.
54. Gamble, M., et al., *Regulation of an intracellular subtilisin protease activity by a short propeptide sequence through an original combined dual mechanism.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3536-41.
55. Fruk, L., et al., *Apoenzyme reconstitution as a chemical tool for structural enzymology and biotechnology.* Angew Chem Int Ed Engl, 2009. **48**(9): p. 1550-74.
56. Akamine, P., et al., *Dynamic features of cAMP-dependent protein kinase revealed by apoenzyme crystal structure.* J Mol Biol, 2003. **327**(1): p. 159-71.
57. Murakami, Y., et al., *Artificial Enzymes.* Chem Rev, 1996. **96**(2): p. 721-758.
58. Kirschning, A., *Coenzymes and Their Role in the Evolution of Life.* Angew Chem Int Ed Engl, 2021. **60**(12): p. 6242-6269.
59. Bugg, T.D.H., *Introduction to Enzyme and Coenzyme Chemistry*. 2012: Wiley.

60. Nelson, D.L. and M. Cox, *Lehninger Principles of Biochemistry: International Edition*. 2021: Macmillan Learning.
61. Ball, G.F.M., *Vitamins: Their Role in the Human Body*. 2008: Wiley.
62. Abeles, R.H. and D. Dolphin, *The vitamin B12 coenzyme.* Accounts of Chemical Research, 2002. **9**(3): p. 114-120.
63. Saini, R., *Coenzyme Q10: The essential nutrient.* Journal of Pharmacy and Bioallied Sciences, 2011. **3**(3): p. 466-467.
64. Andreini, C., et al., *Metal ions in biological catalysis: from enzyme databases to general principles.* J Biol Inorg Chem, 2008. **13**(8): p. 1205-18.
65. Broderick, J., *Coenzymes and Cofactors*. 2001.
66. Reymond, J.L., *Enzyme Assays: High-throughput Screening, Genetic Selection and Fingerprinting*. 2006: Wiley.
67. Bisswanger, H., *Enzyme assays.* Perspectives in Science, 2014. **1**(1-6): p. 41-55.
68. Michaelis, L., et al., *The original Michaelis constant: translation of the 1913 Michaelis-Menten paper.* Biochemistry, 2011. **50**(39): p. 8264-9.
69. *Units of Enzyme Activity. Recommendations 1978.* European Journal of Biochemistry, 1979. **97**(2): p. 319-320.
70. Tipton, K.F., et al., *Standards for Reporting Enzyme Data: The STRENDA Consortium: What it aims to do and why it should be helpful.* Perspectives in Science, 2014. **1**(1-6): p. 131-137.
71. Ramsay, R.R. and K.F. Tipton, *Assessment of Enzyme Inhibition: A Review with Examples from the Development of Monoamine Oxidase and Cholinesterase Inhibitory Drugs.* Molecules, 2017. **22**(7).
72. Gonzalez-Bello, C., *Designing Irreversible Inhibitors--Worth the Effort?* ChemMedChem, 2016. **11**(1): p. 22-30.
73. Patadiya, N., N. Panchal, and V. Vaghela, *A Review on Enzyme Inhibitors.* International Research Journal Of Pharmacy, 2021. **12**(6): p. 60-66.
74. Ring, B., S.A. Wrighton, and M. Mohutsky, *Reversible mechanisms of enzyme inhibition and resulting clinical significance.* Methods Mol Biol, 2014. **1113**: p. 37-56.
75. Blat, Y., *Non-competitive inhibition by active site binders.* Chem Biol Drug Des, 2010. **75**(6): p. 535-40.
76. Cornish-Bowden, A., *Current IUBMB recommendations on enzyme nomenclature and kinetics.* Perspectives in Science, 2014. **1**(1-6): p. 74-87.
77. McDonald, A.G. and K.F. Tipton, *Enzyme nomenclature and classification: the state of the art.* FEBS J, 2021.
78. Lopez-Otin, C. and J.S. Bond, *Proteases: multifunctional enzymes in life and disease.* J Biol Chem, 2008. **283**(45): p. 30433-7.
79. Maurer, K.-H., *Detergent proteases.* Current Opinion in Biotechnology, 2004. **15**(4): p. 330-334.
80. Tavano, O.L., *Protein hydrolysis using proteases: An important tool for food biotechnology.* Journal of Molecular Catalysis B: Enzymatic, 2013. **90**: p. 1-11.
81. Ward, O.P., *Proteases.* Comprehensive Biotechnology, 2011: p. 14.
82. Craik, C.S., M.J. Page, and E.L. Madison, *Proteases as therapeutics.* Biochem J, 2011. **435**(1): p. 1-16.
83. Moo-Young, M., *Comprehensive Biotechnology*. 2019: Elsevier Science.
84. Naqvi, N., et al., *Molecular basis of exopeptidase activity in the C-terminal domain of human angiotensin I-converting enzyme: insights into the origins of its exopeptidase activity.* J Biol Chem, 2005. **280**(8): p. 6669-75.
85. Motyan, J.A., F. Toth, and J. Tozser, *Research applications of proteolytic enzymes in molecular biology.* Biomolecules, 2013. **3**(4): p. 923-42.
86. Sakamoto, Y., et al., *S46 peptidases are the first exopeptidases to be members of clan PA.* Sci Rep, 2014. **4**: p. 4977.
87. Rawlings, N.D., F.R. Morton, and A.J. Barrett, *MEROPS: the peptidase database.* Nucleic Acids Res, 2006. **34**(Database issue): p. D270-2.
88. Di Cera, E., *Serine proteases.* IUBMB Life, 2009. **61**(5): p. 510-5.
89. Carter, P. and J.A. Wells, *Dissecting the catalytic triad of a serine protease.* Nature, 1988. **332**(6164): p. 564-8.
90. Krem, M.M. and E. Di Cera, *Molecular markers of serine protease evolution.* EMBO J, 2001. **20**(12): p. 3036-45.
91. Harwood, C.R. and Y. Kikuchi, *The ins and outs of Bacillus proteases: activities, functions and commercial significance.* FEMS Microbiol Rev, 2022. **46**(1).

92. Linderstrom-Land, K. and M. Ottesen, *A new protein from ovalbumin.* Nature, 1947. **159**(4050): p. 807.
93. Wright, C.S., R.A. Alden, and J. Kraut, *Structure of subtilisin BPN' at 2.5 angstrom resolution.* Nature, 1969. **221**(5177): p. 235-42.
94. Betzel, C., et al., *Crystal structures of the alkaline proteases savinase and esperase from Bacillus lentus.* Adv Exp Med Biol, 1996. **379**: p. 49-61.
95. Zhao, H. and F.H. Arnold, *Directed evolution converts subtilisin E into a functional equivalent of thermitase.* Protein Eng, 1999. **12**(1): p. 47-53.
96. Hallenberger, S., et al., *The role of eukaryotic subtilisin-like endoproteases for the activation of human immunodeficiency virus glycoproteins in natural host cells.* J Virol, 1997. **71**(2): p. 1036-45.
97. Tripathi, L.P. and R. Sowdhamini, *Genome-wide survey of prokaryotic serine proteases: analysis of distribution and domain architectures of five serine protease families in prokaryotes.* BMC Genomics, 2008. **9**: p. 549.
98. Smith, E.L., et al., *The complete amino acid sequence of two types of subtilisin, BPN' and Carlsberg.* J Biol Chem, 1966. **241**(24): p. 5974-6.
99. Azrin, N.A.M., et al., *Versatility of subtilisin: A review on structure, characteristics, and applications.* Biotechnol Appl Biochem, 2022.
100. Ikemura, H., H. Takagi, and M. Inouye, *Requirement of pro-sequence for the production of active subtilisin E in Escherichia coli.* Journal of Biological Chemistry, 1987. **262**(16): p. 7859-7864.
101. Zhu, X.L., et al., *Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process.* Nature, 1989. **339**(6224): p. 483-4.
102. Li, Y., et al., *Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding. Refolding and inhibitory abilities of propeptide mutants.* J Biol Chem, 1995. **270**(42): p. 25127-32.
103. Barrett, A.J., N.D. Rawlings, and J.F. Woessner, *Handbook of Proteolytic Enzymes*. 2012: Elsevier Science.
104. Abdizadeh, H., et al., *Perturbation response scanning specifies key regions in subtilisin serine protease for both function and stability.* J Enzyme Inhib Med Chem, 2015. **30**(6): p. 867-73.
105. Bryan, P., et al., *Site-directed mutagenesis and the role of the oxyanion hole in subtilisin.* Proc Natl Acad Sci U S A, 1986. **83**(11): p. 3743-5.
106. Smith, C.A., et al., *Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of Bacillus Ak.1 protease at 1.8 A resolution.* J Mol Biol, 1999. **294**(4): p. 1027-40.
107. Uehara, R., et al., *Formation of the high-affinity calcium binding site in pro-subtilisin E with the insertion sequence IS1 of pro-Tk-subtilisin.* Biochemistry, 2013. **52**(50): p. 9080-8.
108. Kidd, R.D., et al., *A Weak Calcium Binding Site in Subtilisin BPN' Has a Dramatic Effect on Protein Stability.* Journal of the American Chemical Society, 1996. **118**(7): p. 1645-1650.
109. Kraut, J., *Serine proteases: structure and mechanism of catalysis.* Annu Rev Biochem, 1977. **46**: p. 331-58.
110. Hedstrom, L., *Serine protease mechanism and specificity.* Chem Rev, 2002. **102**(12): p. 4501-24.
111. Vogel, A. and O. May, *Industrial Enzyme Applications*. 2019: Wiley.
112. Aehle, W., *Enzymes in Industry: Production and Applications*. 2008: Wiley.
113. Burtis, C.A. and T.D. Geary, *Glossary of bioanalytical nomenclature - Part 1: General terminology, body fluids, enzymology, immunology (IUPAC Recommendations 1994).* Pure and Applied Chemistry, 1994. **66**(12): p. 2587-2604.
114. Karan, R., M.D. Capes, and S. Dassarma, *Function and biotechnology of extremophilic enzymes in low water activity.* Aquat Biosyst, 2012. **8**(1): p. 4.
115. Haki, G.D. and S.K. Rakshit, *Developments in industrially important thermostable enzymes: a review.* Bioresour Technol, 2003. **89**(1): p. 17-34.
116. Fongaro, G., et al., *Extremophile Microbial Communities and Enzymes for Bioenergetic Application Based on Multi-Omics Tools.* Curr Genomics, 2020. **21**(4): p. 240-252.
117. Bornscheuer, U.T. and M. Pohl, *Improved biocatalysts by directed evolution and rational protein design.* Curr Opin Chem Biol, 2001. **5**(2): p. 137-43.
118. Korendovych, I.V., *Rational and Semirational Protein Design.* Methods Mol Biol, 2018. **1685**: p. 15-23.
119. Wang, Y., et al., *Directed Evolution: Methodologies and Applications.* Chem Rev, 2021. **121**(20): p. 12384-12444.

120. Packer, M.S. and D.R. Liu, *Methods for the directed evolution of proteins.* Nat Rev Genet, 2015. **16**(7): p. 379-94.
121. Davis, B.G., *Chemical modification of biocatalysts.* Curr Opin Biotechnol, 2003. **14**(4): p. 379-86.
122. Giri, P., et al., *Chemical modification of enzymes to improve biocatalytic performance.* Biotechnol Adv, 2021. **53**: p. 107868.
123. Datta, S., L.R. Christena, and Y.R. Rajaram, *Enzyme immobilization: an overview on techniques and support materials.* 3 Biotech, 2013. **3**(1): p. 1-9.
124. Khan, M.R., *Immobilized enzymes: a comprehensive review.* Bulletin of the National Research Centre, 2021. **45**(1).
125. Asgher, M., et al., *Recent trends and valorization of immobilization strategies and ligninolytic enzymes by industrial biotechnology.* Journal of Molecular Catalysis B: Enzymatic, 2014. **101**: p. 56-66.
126. Jesionowski, T., J. Zdarta, and B. Krajewska, *Enzyme immobilization by adsorption: a review.* Adsorption, 2014. **20**(5-6): p. 801-821.
127. Kumar, A. and S.S. Kanwar, *Catalytic potential of a nitrocellulose membrane-immobilized lipase in aqueous and organic media.* Journal of Applied Polymer Science, 2012. **124**(S1): p. E37-E44.
128. Subramanian, A., et al., *Comparison of techniques for enzyme immobilization on silicon supports.* Enzyme and Microbial Technology, 1999. **24**(1-2): p. 26-34.
129. Sheldon, R.A., *Enzyme Immobilization: The Quest for Optimum Performance.* Advanced Synthesis & Catalysis, 2007. **349**(8-9): p. 1289-1307.
130. Zucca, P. and E. Sanjust, *Inorganic materials as supports for covalent enzyme immobilization: methods and mechanisms.* Molecules, 2014. **19**(9): p. 14139-94.
131. Wang, Y. and F. Caruso, *Mesoporous Silica Spheres as Supports for Enzyme Immobilization and Encapsulation.* Chemistry of Materials, 2005. **17**(5): p. 953-961.
132. Yoshimoto, M., *Stabilization of Enzymes Through Encapsulation in Liposomes.* Methods Mol Biol, 2017. **1504**: p. 9-18.
133. Brito, A.E.M., et al., *Poly (lactic-co-glycolic acid) nanospheres allow for high l-asparaginase encapsulation yield and activity.* Mater Sci Eng C Mater Biol Appl, 2019. **98**: p. 524-534.
134. Othman, A., A. Karimi, and S. Andreescu, *Functional nanostructures for enzyme based biosensors: properties, fabrication and applications.* J Mater Chem B, 2016. **4**(45): p. 7178-7203.
135. Rezaei, K., E. Jenab, and F. Temelli, *Effects of water on enzyme performance with an emphasis on the reactions in supercritical fluids.* Crit Rev Biotechnol, 2007. **27**(4): p. 183-95.
136. Stepankova, V., et al., *Strategies for Stabilization of Enzymes in Organic Solvents.* ACS Catalysis, 2013. **3**(12): p. 2823-2836.
137. Hudson, E.P., R.K. Eppler, and D.S. Clark, *Biocatalysis in semi-aqueous and nearly anhydrous conditions.* Curr Opin Biotechnol, 2005. **16**(6): p. 637-43.
138. Balcão, V.M. and M.M.D.C. Vila, *Structural and functional stabilization of protein entities: state-of-the-art.* Advanced Drug Delivery Reviews, 2015. **93**: p. 25-41.
139. Back, J.F., D. Oakenfull, and M.B. Smith, *Increased thermal stability of proteins in the presence of sugars and polyols.* Biochemistry, 1979. **18**(23): p. 5191-6.
140. Sola-Penna, M. and J.R. Meyer-Fernandes, *Stabilization against thermal inactivation promoted by sugars on enzyme structure and function: why is trehalose more effective than other sugars?* Arch Biochem Biophys, 1998. **360**(1): p. 10-4.
141. Singh, S. and J. Singh, *Effect of polyols on the conformational stability and biological activity of a model protein lysozyme.* AAPS PharmSciTech, 2003. **4**(3): p. E42.
142. Arakawa, T. and S.N. Timasheff, *The stabilization of proteins by osmolytes.* Biophys J, 1985. **47**(3): p. 411-4.
143. Gekko, K. and S.N. Timasheff, *Thermodynamic and kinetic examination of protein stabilization by glycerol.* Biochemistry, 1981. **20**(16): p. 4677-86.
144. Gekko, K. and S.N. Timasheff, *Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures.* Biochemistry, 1981. **20**(16): p. 4667-76.
145. Lee, J.C. and S.N. Timasheff, *The stabilization of proteins by sucrose.* Journal of Biological Chemistry, 1981. **256**(14): p. 7193-7201.
146. Kaushik, J.K. and R. Bhat, *Why is trehalose an exceptional protein stabilizer? An analysis of the thermal stability of proteins in the presence of the compatible osmolyte trehalose.* J Biol Chem, 2003. **278**(29): p. 26458-65.

147. Braham, S.A., et al., *Effect of Concentrated Salts Solutions on the Stability of Immobilized Enzymes: Influence of Inactivation Conditions and Immobilization Protocol.* Molecules, 2021. **26**(4).
148. Hani, F.M., A.E. Cole, and E. Altman, *The ability of salts to stabilize proteins in vivo or intracellularly correlates with the Hofmeister series of ions.* Int J Biochem Mol Biol, 2019. **10**(3): p. 23-31.
149. Zhao, H., *Protein Stabilization and Enzyme Activation in Ionic Liquids: Specific Ion Effects.* J Chem Technol Biotechnol, 2016. **91**(1): p. 25-50.
150. Martinez, L., et al., *PACKMOL: a package for building initial configurations for molecular dynamics simulations.* J Comput Chem, 2009. **30**(13): p. 2157-64.
151. Olsson, M.H., et al., *PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.* J Chem Theory Comput, 2011. **7**(2): p. 525-37.
152. Sondergaard, C.R., et al., *Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values.* J Chem Theory Comput, 2011. **7**(7): p. 2284-95.
153. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
154. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.
155. Wilson, K. and J. Walker, *Principles and Techniques of Biochemistry and Molecular Biology*. 2005: Cambridge University Press.
156. Bisswanger, H., *Practical Enzymology*. 2019: Wiley.
157. DelMar, E.G., et al., *A sensitive new substrate for chymotrypsin.* Anal Biochem, 1979. **99**(2): p. 316-20.
158. Deshpande, M.V., K.E. Eriksson, and L.G. Pettersson, *An assay for selective determination of exo-1,4,-beta-glucanases in a mixture of cellulolytic enzymes.* Anal Biochem, 1984. **138**(2): p. 481-7.
159. Höhne, G., W.F. Hemminger, and H.J. Flammersheim, *Differential Scanning Calorimetry*. 2013: Springer Berlin Heidelberg.
160. Kübelbeck, S., *Stabilization of Industry-Relevant Enzyme Formulations*. 2020, Technische Universität: tuprints:11666.
161. Johnson, C.M., *Differential scanning calorimetry as a tool for protein folding and stability.* Arch Biochem Biophys, 2013. **531**(1-2): p. 100-9.
162. Frenkel, D. and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*. 2001: Elsevier Science.
163. Hansson, T., C. Oostenbrink, and W. van Gunsteren, *Molecular dynamics simulations.* Curr Opin Struct Biol, 2002. **12**(2): p. 190-6.
164. Leach, A.R., *Molecular Modelling: Principles and Applications*. 2001: Prentice Hall.
165. Lee, B. and F.M. Richards, *The interpretation of protein structures: estimation of static accessibility.* J Mol Biol, 1971. **55**(3): p. 379-400.
166. Knapp, B., et al., *Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible?* J Comput Biol, 2011. **18**(8): p. 997-1005.
167. Hong, P., S. Koza, and E.S. Bouvier, *Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates.* J Liq Chromatogr Relat Technol, 2012. **35**(20): p. 2923-2950.
168. Fekete, S., et al., *Theory and practice of size exclusion chromatography for the analysis of protein aggregates.* J Pharm Biomed Anal, 2014. **101**: p. 161-73.
169. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
170. Smith, B.J., *SDS Polyacrylamide Gel Electrophoresis of Proteins.* Methods Mol Biol, 1984. **1**: p. 41-55.
171. Barth, A., *Infrared spectroscopy of proteins.* Biochim Biophys Acta, 2007. **1767**(9): p. 1073-101.
172. Günzler, H. and H.U. Gremlich, *IR-Spektroskopie: Eine Einführung*. 2012: Wiley.
173. Yeoman, C.J., et al., *Thermostable enzymes as biocatalysts in the biofuel industry.* Adv Appl Microbiol, 2010. **70**: p. 1-55.
174. Rigoldi, F., et al., *Review: Engineering of thermostable enzymes for industrial applications.* APL Bioeng, 2018. **2**(1): p. 011501.
175. Ward, O.P. and M. Moo-Young, *Thermostable enzymes.* Biotechnol Adv, 1988. **6**(1): p. 39-69.

176. Nigam, P.S., *Microbial enzymes with special characteristics for biotechnological applications.* Biomolecules, 2013. **3**(3): p. 597-611.
177. Seelig, J. and H.J. Schonfeld, *Thermal protein unfolding by differential scanning calorimetry and circular dichroism spectroscopy Two-state model versus sequential unfolding.* Q Rev Biophys, 2016. **49**: p. e9.
178. Yang, A.S. and B. Honig, *On the pH dependence of protein stability.* J Mol Biol, 1993. **231**(2): p. 459-74.
179. Konermann, L., *Protein Unfolding and Denaturants*. 2012: John Wiley & Sons, Ltd.
180. Kishore, D., S. Kundu, and A.M. Kayastha, *Thermal, chemical and pH induced denaturation of a multimeric beta-galactosidase reveals multiple unfolding pathways.* PLoS One, 2012. **7**(11): p. e50380.
181. Daniel, R.M., et al., *The effect of temperature on enzyme activity: new insights and their implications.* Extremophiles, 2008. **12**(1): p. 51-9.
182. Peterson, M.E., et al., *The dependence of enzyme activity on temperature: determination and validation of parameters.* Biochem J, 2007. **402**(2): p. 331-7.
183. Yang, Z. and K.N. Houk, *The Dynamics of Chemical Reactions: Atomistic Visualizations of Organic Reactions, and Homage to van 't Hoff.* Chemistry, 2018. **24**(16): p. 3916-3924.
184. Peleg, M., M.D. Normand, and M.G. Corradini, *The Arrhenius equation revisited.* Crit Rev Food Sci Nutr, 2012. **52**(9): p. 830-51.
185. Van Ness, H.C., *Understanding Thermodynamics*. 2012: Dover Publications.
186. Fujita, T., et al., *Simple method for analyzing the purity of protease-containing samples by acid-treatment SDS-PAGE.* J Biosci Bioeng, 2019. **128**(5): p. 630-635.
187. Lund, H., et al., *Correlation Between Enzyme Activity and Stability of a Protease, an Alpha-Amylase and a Lipase in a Simplified Liquid Laundry Detergent System, Determined by Differential Scanning Calorimetry.* Journal of Surfactants and Detergents, 2011. **15**(1): p. 9-21.
188. Chiu, M.H. and E.J. Prenner, *Differential scanning calorimetry: An invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions.* J Pharm Bioallied Sci, 2011. **3**(1): p. 39-59.
189. Zhang, S. and J. Lv, *Purification and properties of heat-stable extracellular protease from Pseudomonads fluorescens BJ-10.* J Food Sci Technol, 2014. **51**(6): p. 1185-90.
190. Tanaka, T., H. Matsuzawa, and T. Ohta, *Stability of Thermostable Enzyme, Aqualysin I ; a Subtilisin-type Serine Protease from Thermus aquaticus YT-1.* Biosci Biotechnol Biochem, 1998. **62**(9): p. 1806-8.
191. Grob, S. *Molinspiration Cheminformatics free web services*. [cited 2023; Available from: https://www.molinspiration.com.
192. Goddette, D.W., et al., *The crystal structure of the Bacillus lentus alkaline protease, subtilisin BL, at 1.4 A resolution.* J Mol Biol, 1992. **228**(2): p. 580-95.
193. Lozano, P., D. Combes, and J.L. Iborra, *Effect of polyols on alpha-chymotrypsin thermostability: a mechanistic analysis of the enzyme stabilization.* J Biotechnol, 1994. **35**(1): p. 9-18.
194. Zaroog, M.S., H. Abdul Kadir, and S. Tayyab, *Stabilizing effect of various polyols on the native and the denatured states of glucoamylase.* ScientificWorldJournal, 2013. **2013**: p. 570859.
195. Kuznetsova, I.M., et al., *Beyond the excluded volume effects: mechanistic complexity of the crowded milieu.* Molecules, 2015. **20**(1): p. 1377-409.
196. Liu, Y. and D.W. Bolen, *The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes.* Biochemistry, 1995. **34**(39): p. 12884-91.
197. Romero, C.M. and A. Albis, *Influence of Polyols and Glucose on the Surface Tension of Bovine α-Lactalbumin in Aqueous Solution.* Journal of Solution Chemistry, 2010. **39**(12): p. 1865-1876.
198. Kumar, A., P. Attri, and P. Venkatesu, *Effect of polyols on the native structure of α-chymotrypsin: A comparable study.* Thermochimica Acta, 2012. **536**: p. 55-62.
199. Politi, R., L. Sapir, and D. Harries, *The impact of polyols on water structure in solution: a computational study.* J Phys Chem A, 2009. **113**(26): p. 7548-55.
200. Timasheff, S.N., *The Control of Protein Stability and Association by Weak Interactions with Water: How Do Solvents Affect These Processes?* Annual Review of Biophysics and Biomolecular Structure, 1993. **22**(1): p. 67-97.
201. Xie, G. and S.N. Timasheff, *Mechanism of the stabilization of ribonuclease A by sorbitol: preferential hydration is greater for the denatured then for the native protein.* Protein Sci, 1997. **6**(1): p. 211-21.

202. Liu, F.F., et al., *Molecular basis for polyol-induced protein stability revealed by molecular dynamics simulations.* J Chem Phys, 2010. **132**(22): p. 225103.
203. Chen, Q., et al., *Semi-rational design and molecular dynamics simulations study of the thermostability enhancement of cellobiose 2-epimerases.* Int J Biol Macromol, 2020. **154**: p. 1356-1365.
204. Zeiske, T., K.A. Stafford, and A.G. Palmer, 3rd, *Thermostability of Enzymes from Molecular Dynamics Simulations.* J Chem Theory Comput, 2016. **12**(6): p. 2489-92.
205. Bekker, G.J., B. Ma, and N. Kamiya, *Thermal stability of single-domain antibodies estimated by molecular dynamics simulations.* Protein Sci, 2019. **28**(2): p. 429-438.
206. Permyakov, E.A., *Metal Binding Proteins.* Encyclopedia, 2021. **1**(1): p. 261-292.
207. Rawlings, N.D., et al., *MEROPS: the database of proteolytic enzymes, their substrates and inhibitors.* Nucleic Acids Res, 2014. **42**(Database issue): p. D503-9.
208. Pantoliano, M.W., et al., *Large increases in general stability for subtilisin BPN' through incremental changes in the free energy of unfolding.* Biochemistry, 1989. **28**(18): p. 7205-13.
209. Frommel, C. and W.E. Hohne, *Influence of calcium binding on the thermal stability of 'thermitase', a serine protease from Thermoactinomyces vulgaris.* Biochim Biophys Acta, 1981. **670**(1): p. 25-31.
210. Gros, P., K.H. Kalk, and W.G. Hol, *Calcium binding to thermitase. Crystallographic studies of thermitase at 0, 5, and 100 mM calcium.* J Biol Chem, 1991. **266**(5): p. 2953-61.
211. Betzel, C., G.P. Pal, and W. Saenger, *Three-dimensional structure of proteinase K at 0.15-nm resolution.* Eur J Biochem, 1988. **178**(1): p. 155-71.
212. Bryan, P., et al., *Energetics of folding subtilisin BPN'.* Biochemistry, 1992. **31**(21): p. 4937-45.
213. Strausberg, S.L., et al., *Directed evolution of a subtilisin with calcium-independent stability.* Biotechnology (N Y), 1995. **13**(7): p. 669-73.
214. Strausberg, S., et al., *Catalysis of a protein folding reaction: thermodynamic and kinetic analysis of subtilisin BPN' interactions with its propeptide fragment.* Biochemistry, 1993. **32**(32): p. 8112-9.
215. Scott, R.A., *Encyclopedia of Inorganic Chemistry, 5 Volume Set*. 2012: Wiley.
216. Pantoliano, M.W., et al., *The engineering of binding affinity at metal ion binding sites for the stabilization of proteins: subtilisin as a test case.* Biochemistry, 1988. **27**(22): p. 8311-8317.
217. Briedigkeit, L. and C. Frömmel, *Calcium ion binding by thermitase.* FEBS Letters, 1989. **253**(1-2): p. 83-87.
218. Bajorath, J., W. Hinrichs, and W. Saenger, *The enzymatic activity of proteinase K is controlled by calcium.* Eur J Biochem, 1988. **176**(2): p. 441-7.
219. Gamble, M., et al., *The role of substrate specificity and metal binding in defining the activity and structure of an intracellular subtilisin.* FEBS Open Bio, 2012. **2**: p. 209-15.
220. Ghorbel, B., A. Sellami-Kamoun, and M. Nasri, *Stability studies of protease from Bacillus cereus BG1.* Enzyme and Microbial Technology, 2003. **32**(5): p. 513-518.
221. Almog, O., et al., *Structural basis of thermostability. Analysis of stabilizing mutations in subtilisin BPN'.* J Biol Chem, 2002. **277**(30): p. 27553-8.
222. Okur, H.I., et al., *Beyond the Hofmeister Series: Ion-Specific Effects on Proteins and Their Biological Functions.* J Phys Chem B, 2017. **121**(9): p. 1997-2014.
223. Silva, C., et al., *Practical insights on enzyme stabilization.* Crit Rev Biotechnol, 2018. **38**(3): p. 335-350.
224. Coutte, L., et al., *Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway.* EMBO J, 2001. **20**(18): p. 5040-8.
225. Stoner, M.R., et al., *Protease autolysis in heavy-duty liquid detergent formulations: effects of thermodynamic stabilizers and protease inhibitors.* Enzyme and Microbial Technology, 2004. **34**(2): p. 114-125.
226. Radisky, E.S. and D.E. Koshland, Jr., *A clogged gutter mechanism for protease inhibitors.* Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10316-21.
227. Suplatov, D.A., et al., *Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities.* Protein Eng Des Sel, 2012. **25**(11): p. 689-97.
228. Katz, B.A., et al., *Elaborate manifold of short hydrogen bond arrays mediating binding of active site-directed serine protease inhibitors.* J Mol Biol, 2003. **329**(1): p. 93-120.
229. Gupta, V., et al., *Recognition of active and inactive catalytic triads: A template based approach.* Int J Biol Macromol, 2010. **46**(3): p. 317-23.
230. Rozanov, A.S., et al., *Production of subtilisin proteases in bacteria and yeast.* Vavilovskii Zhurnal Genet Selektsii, 2021. **25**(1): p. 125-134.

231. Gao, B.J., et al., *Subtilisin-like Pr1 proteases marking the evolution of pathogenicity in a wide-spectrum insect-pathogenic fungus.* Virulence, 2020. **11**(1): p. 365-380.
232. Chen, P., et al., *Structure of the human cytomegalovirus protease catalytic domain reveals a novel serine protease fold and catalytic triad.* Cell, 1996. **86**(5): p. 835-43.
233. Figueiredo, J., M. Sousa Silva, and A. Figueiredo, *Subtilisin-like proteases in plant defence: the past, the present and beyond.* Mol Plant Pathol, 2018. **19**(4): p. 1017-1028.
234. Kiefer, M.C., et al., *Identification of a second human subtilisin-like protease gene in the fes/fps region of chromosome 15.* DNA Cell Biol, 1991. **10**(10): p. 757-69.
235. Adams, C.M., et al., *Structural and functional analysis of the CspB protease required for Clostridium spore germination.* PLoS Pathog, 2013. **9**(2): p. e1003165.
236. Song, J., et al., *PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites.* PLoS One, 2012. **7**(11): p. e50300.
237. Bode, W., et al., *Refined 1.2 A crystal structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin c. Molecular structure of eglin and its detailed interaction with subtilisin.* EMBO J, 1986. **5**(4): p. 813-8.
238. McPhalen, C.A. and M.N.G. James, *Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin-C-subtilisin Carlsberg and CI-2-subtilisin Novo.* Biochemistry, 2002. **27**(17): p. 6582-6598.
239. Rawlings, N.D., D.P. Tolle, and A.J. Barrett, *Evolutionary families of peptidase inhibitors.* Biochem J, 2004. **378**(Pt 3): p. 705-16.
240. Akasako, A., et al., *High resistance of Escherichia coli ribonuclease HI variant with quintuple thermostabilizing mutations to thermal denaturation, acid denaturation, and proteolytic degradation.* Biochemistry, 1995. **34**(25): p. 8115-22.
241. Daniel, R.M., et al., *A correlation between protein thermostability and resistance to proteolysis.* Biochem J, 1982. **207**(3): p. 641-4.
242. Parsell, D.A. and R.T. Sauer, *The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli.* J Biol Chem, 1989. **264**(13): p. 7590-5.
243. Everitt, B.S., S. Landau, and M. Leese, *Cluster Analysis*. 2001: Wiley.
244. Jackson, M. and H.H. Mantsch, *The use and misuse of FTIR spectroscopy in the determination of protein structure.* Crit Rev Biochem Mol Biol, 1995. **30**(2): p. 95-120.
245. Surewicz, W.K., H.H. Mantsch, and D. Chapman, *Determination of protein secondary structure by Fourier transform infrared spectroscopy: a critical assessment.* Biochemistry, 1993. **32**(2): p. 389-94.
246. Cai, S. and B.R. Singh, *A distinct utility of the amide III infrared band for secondary structure estimation of aqueous protein solutions using partial least squares methods.* Biochemistry, 2004. **43**(9): p. 2541-9.
247. Guler, G., et al., *Real time observation of proteolysis with Fourier transform infrared (FT-IR) and UV-circular dichroism spectroscopy: watching a protease eat a protein.* Spectrochim Acta A Mol Biomol Spectrosc, 2011. **79**(1): p. 104-11.
248. Kristoffersen, K.A., et al., *Fourier-transform infrared spectroscopy for monitoring proteolytic reactions using dry-films treated with trifluoroacetic acid.* Sci Rep, 2020. **10**(1): p. 7844.
249. Yang, H., et al., *Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy.* Nat Protoc, 2015. **10**(3): p. 382-96.
250. Barth, A. and C. Zscherp, *Substrate binding and enzyme function investigated by infrared spectroscopy.* FEBS Lett, 2000. **477**(3): p. 151-6.
251. Gerwert, K., *Molecular reaction mechanisms of proteins monitored by time-resolved FTIR-spectroscopy.* Biol Chem, 1999. **380**(7-8): p. 931-5.
252. Fabian, H. and D. Naumann, *Methods to study protein folding by stopped-flow FT-IR.* Methods, 2004. **34**(1): p. 28-40.
253. Sassi, P., et al., *Unfolding and aggregation of lysozyme: a thermodynamic and kinetic study by FTIR spectroscopy.* Biophys Chem, 2011. **158**(1): p. 46-53.
254. Sarroukh, R., et al., *ATR-FTIR: a "rejuvenated" tool to investigate amyloid proteins.* Biochim Biophys Acta, 2013. **1828**(10): p. 2328-38.
255. Venyaminov, S. and F.G. Prendergast, *Water (H2O and D2O) molar absorptivity in the 1000-4000 cm-1 range and quantitative infrared spectroscopy of aqueous solutions.* Anal Biochem, 1997. **248**(2): p. 234-45.
256. Doherty, J., et al., *Increased optical pathlength through aqueous media for the infrared microanalysis of live cells.* Anal Bioanal Chem, 2018. **410**(23): p. 5779-5789.

List of Abbreviations

$\bar{r}$	Weighted center
$[E_T]$	Total enzyme concentration
°C	Degrees Celsius (Unit)
µ	Mikro
A	Alanine
A	Absorption
Abs	Absorption
AI	Artificial intelligence
Asn	Asparagine
Asp	Aspartic acid
Atm	Standard atmosphere (Unit)
Bis	Bis(2-hydroxyethyl)amino
c	Concentration
C	Cysteine
Ca^{2+}	Calcium ion
C_p	Heat Capacity
d	Layer thickness
d	Contact distance at time t
d_0	Contact distances at time t= 0
Dia	Dialysed
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
e.g.	For example
E/[E]	Enzyme/-Concentration
EC	Enzyme Commission
EDDS	Ethylenediamine-N,N'-disuccinic acid
EDTA	Ethylenediaminetetraacetic acid
EPO	External Parameter Orthogonalization
eq.	Equation
et al.	“and others”
Eλ	Extinction
FT-IR	Fourier Transform Infrared
g	Gram (Unit)
G	Glycine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine

GPC	Gel permeation chromatography
H-bonds	Hydrogen bonds
HCl	Hydrochloric acid
I	Intensity after analyte
I_0	Intensity before analyte
ID	Identification
IR	Infrared
IUBMB	International Union of Biochemistry and Molecular Biology
K	Association/Dissociation constant
K	Kelvin (Unit)
K^+	Potassium ion
kat	Katal (Unit)/[E] to turnover mol [S] in 1s
kD	Kilodalton (Unit)
K_D	Dissociation coefficient
kg	Kilogram (Unit)
K_m	Substrate specificity
L	Liter (Unit)
LMFIT	Non-linear least-squares minimization and curve fitting for python
M	Metal
MD	Molecular Dynamics
MEROPS	Proteolytic enzyme database
MES	2-(*N*-morpholino)ethanesulfonic acid
Mg^{2+}	Magnesium ion
MGDA	Ethylenediaminetetraacetic acid
MIRA	Mid infrared analyzer
MM	Michaelis-Menten
mol	Mol (Unit)
MOPS	3-(*N*-morpholino)propanesulfonic acid
MPG	1,2-Propanediol
MTS	Multiple step integration method
N	Asparagine
n	Number of atoms
Na^+	Sodium ion
NAMD	Nanoscale Molecular Dynamics (Software)
NaOH	Sodium hydroxide
NHS	N-hydroxysuccinimide
nm	Nanometer (Unit)
P	Product
P	Mixed residues (MEROPS classification)
PACKMOL	Initial configurations for Molecular Dynamics (Software)
PCA	Principle Component Regression

PCR	Polymerase chain reaction/Principle component regression
PES	Polyethersulfone
pH	Potential of hydrogen
pI	Isoelectric point
PLS	Partial Least Squares
PME	Particle mesh Ewald
pNA	4-Nitroaniline
pNP	4-Nitrophenol
pNPL	4-ntrophenyl-ß-D-lactopyranoisde
Q-value	Relative number of native contacts
RESPA	Reversible reference system propagation algorithm
RG	Radius of gyration
r_i	Coordinates of the same atom at later timepoint
r_i^{ref}	Coordinates of the i-th atom/residue in the reference structure
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RT	Room temperature (22 °C)
S	Serine
S/[S]	Substrate/-Concentration
SASA	Solvent accessible surface area
SDS-PAGE	Sodium Dodecyl Sulfate – PolyAcrylamid Gel Electrophoresis
SEC	Seize exclusion chromatography
Suc-AAPF-pNA	Protease Substrate/Succinyl-alanine-alanine-proline-phenylalanine
t	Time
T	Threonine/Number of frames/Transmittance
T_m	Transitions midpoint
TRIS	Tris(hydroxymethyl)aminomethane
Trp	Tryptophane
U	Turnover [µmol/min]
UV	Ultraviolet
V	Volume
V	Volt (Unit)
Val	Valine
V_{max}	Maximum reaction velocity
VMD	Visual Molecular Dynamics (Software)
W	Weight
WT	Wildtype
wt%	Weight%
Zn^{2+}	Zinc ion
α	Ancient Greek: ἄλφα/Alpha
β	Ancient Greek: βῆτα/Beta / Softness of the switching function

Δ	Ancient Greek: δέλτα/Delta
ΔG	Free enthalpy
ΔG^{+}	Reaction enthalpy/Activation Energy
ΔH	Enthalpy
ε_{λ}	Extinction coefficient
Λ	Ancient Greek: λάβδα/Lambda / reference distance tolerance (1.8)
Ω	Ancient Greek: ὦ μέγα/Omega

Table of Figures

Table of Tables