Modelling of Enzymatic Protein Hydrolysis

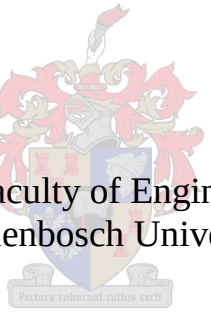
by

Raphahle Nthabiseng Seleme

Thesis presented in partial fulfilment
of the requirements for the Degree

of

MASTER OF ENGINEERING
(CHEMICAL ENGINEERING)



in the Faculty of Engineering
at Stellenbosch University

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Supervisor
Dr. TM Louw

Co-*Supervisor*
Dr. N Goosen

April 2019

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: April 2019

Plagiarism declaration

1. Plagiarism is the use of ideas, material and other intellectual property of another's work and to present is as my own.
2. I agree that plagiarism is a punishable offence because it constitutes theft.
3. I also understand that direct translations are plagiarism.
4. Accordingly all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.
5. I declare that the work contained in this assignment, except where otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

Initials and surname: RN Seleme

Date: April 2019

Abstract

Enzymatic hydrolysis is a method which is used to produce nutritional supplements, aquaculture feed, plant fertilisers and food ingredients. Mathematical modelling of enzymatic protein hydrolysis is a valuable tool for predicting complex protein hydrolysis reactions, but is highly complex.

Previous models are not adequate due to their simplified description of the reaction. Protein hydrolysis involves different substrates and reactions. The products can also act as new substrates resulting in continuous hydrolysis. This study was aimed at developing and validating a theoretical model for protein hydrolysis.

To develop a mathematical model to describe enzymatic protein hydrolysis, the current study proposes a population balance approach along with a modified Michaelis-Menten equation. Population balance modelling was used to predict the population of polymer chains while the modified Michaelis-Menten equation describes the hydrolysis of proteins to form shorter chains. Three kinds of catalytic actions were considered: Endopeptidase, exopeptidase, and endopeptidase-exopeptidase.

In order to validate the proposed models for endopeptidase and endopeptidase-exopeptidase, hydrolysis of whey protein using commercial enzymes, papain and bromelain, were studied at different operating conditions. The optimum conditions of the enzymes were determined experimentally using a central composite experimental design. The effect of enzyme concentration and buffer type on protein hydrolysis were studied using a multilevel factorial design. The degree of hydrolysis and total heat flow during hydrolysis was measured using o-phthaldialdehyde and isothermal microcalorimetry methods. The heat flow and degree of hydrolysis data were used to calculate the model parameters.

The optimum conditions under the studied experimental conditions: Temperature, pH, substrate concentration, enzyme-to-substrate ratio, were 65 ºC, 5.15, 6 % (w/v), 3% (w/w) respectively for papain, and 50 ºC, 6, 10 % (w/v), 3 % (w/w) for bromelain. The heat flow results showed that the heat measured using isothermal microcalorimetry is the apparent heat rather than a heat of the reaction. The comparison between the two techniques for measuring hydrolysis indicated that isothermal microcalorimetry is more accurate and easy to use than spectroscopy method. The parameters of the models were estimated using nonlinear regression

analysis to obtain the model predictions. The model predictions from the proposed models and a model by Marquez-Moreno and Fernandez-Cuadrado (1993) were compared to the experimental data. The proposed model for endopeptidase was on average 60.5 % and 37.4 % better than the model found in the literature for heat flow and degree of hydrolysis experimental data. The model for endopeptidase-exopeptidase was on average 55.4 % and 46.5 % better than the model found in literature. This implied that the proposed models presented a promising approach for modelling a protein-peptidase system and proved to predict the experimental data better than the empirical model in literature. Sensitivity analysis was performed to determine the parameters that have maximum impact on protein hydrolysis. The results showed that two parameters had a great influence on hydrolysis.

In conclusion, the proposed models can be used to predict complex enzymatic protein hydrolysis reactions with reasonable certainty. Knowledge of the mathematical model of protein hydrolysis is important for process optimization and process control.

Keywords: Enzymatic hydrolysis, Isothermal calorimetry, Population balance model, Enzyme kinetics, Mathematical modelling.

Abstrak

Ensimatiese hidroliese is ’n metode wat gebruik word om voedingsaanvullings, awkakultuurvoer, plant bemesting en voedselbestanddele te produseer. Wiskundige modellering van ensimatiese proteïen hidrolise is ’n waardevolle instrument om komplekse proteïen hidrolise reaksies te voorspel, maar is hoogs kompleks.

Vorige modelle is nie voldoende nie as gevolg van hul vereenvoudigde beskrywings van die reaksie. Proteïen hidrolise behels verskillende substrate en reaksies. Die produkte kan ook optree as nuwe substrate wat kontinue hidrolise tot gevolg het. Hierdie studie is gerig op die ontwikkeling en validering van ’n teoretiese model vir proteïen hidrolise.

Om ’n wiskundige model te onwikkel wat ensimatiese proteïen hidrolise beskryf, stel die huidige studie ’n populasie balans benadering saam met ’n gewysigde Michaelis-Menten vergelyking voor. Populasie balans modellering is gebruik om die populasie polimeerkettings te voorspel terwyl die gewysigde Michaelis-Menten vergelyking die hidrolise van proteïne wat korter kettings vorm, beskryf. Drie soorte katalitiese aksies is oorweeg: endopeptidase, eksopeptidase, en endopeptidase-eksopeptidase.

Om die voorgestelde modelle vir endopeptidase en endopeptidase-eksopeptidase te valideer, is hidrolise van weiproteïen deur kommersiële ensieme, papaïen en bromelien, bestudeer by verskillende bedryfstoestande. Die optimale toestande van die ensieme is eksperimenteel vasgestel deur ’n sentrale samestelling eksperimentele ontwerp te gebruik. Die effek van ensiemkonsentrasie en buffer tipe op proteïen hidrolise is bestudeer deur ’n multivlak faktoriaalontwerp te gebruik. Die grade van hidrolise en totale hittevloei is gemeet deur o-tartaardialdehied en isotermiese mikrokalorimetriemetodes te gebruik. Die hittevloei en grade van hidrolise data is gebruik om die model parameters te bereken.

Die optimale toestande in die bestudeerde eksperimentele toestande was temperatuur, pH, substraatkonsentrasie, ensiem-tot-substraatverhouding van 65 °C, 5.15, 6% (w/v), 3% (w/w) onderskeidelik vir papaïen, en 50 °C, 6, 10% (w/v), 3% (w/w) vir bromelien. Die hittevloei resultate het gewys dat die hitte wat deur isotermiese mikrokalorimetrie gemeet is, is die oënskynlike hitte eerder as die hitte van die reaksie. Die vergelyking tussen die twee tegnieke om hidrolise te meet het aangedui dat isotermiese mikrokalorimetrie die metode is wat meer akkuraat en makliker om te gebruik, is. Die parameters van die modelle is beraam deur nie-

liniêre regressie analise te gebruik om die model voorspellings te verkry. Die model voorspellings van die voorgestelde modelle en ’n model deur Marquez-Moreno en Fernandez-Cuadrado (1993) is vergelyk met die eksperimentele data. Die voorgestelde model vir endopeptidase het die kleinste fout van 18.04% gehad en 14.67% minder as dié van die model uit literatuur vir hittevloei en grade van hidrolise korrelasie met eksperimentele data. Die model vir endopeptidase-eksopeptidase het die kleinste fout van 16.16% gehad en 23.11% minder as dié van die literatuurmodel. Dit het geïmpliseer dat die voorgestelde modelle ’n belowende benadering lewer vir modellering van ’n proteïen-peptidase-stelsel en het bewys dat dit die eksperimentele data beter voorspel as die empiriese model in literatuur. Sensitiwiteitsanalise is uitgevoer om die parameters te bepaal wat maksimum impak op proteïen hidrolise het. Die resultate het gewys dat twee parameters ’n beduidende invloed op hidrolise gehad het.

Ten slotte, die voorgestelde modelle kan gebruik word om komplekse ensimatiese proteïen hidrolise reaksies te voorspel. Kennis van die wiskundige model van proteïen hidrolise is belangrik vir proses optimalisering en prosesbeheer.

Sleutelwoorde: Ensimatiese hidrolise, Isotermiese kalorimetrie, Populasie balans model, Ensimatiese kinetika, Wiskundige modellering.

Acknowledgements

To God, I owe all the glory and thanks. I would like to thank my Lord Jesus Christ for giving me the strength to complete this thesis. I would also like to thank and express my appreciation to the following people:

- My supervisors, Dr Tobi Louw and Dr Neill Goosen, for your patience, support, guidance and encouragement throughout the project. I truly appreciate the effort and understanding.
- University of Cape Town, Department of Chemical Engineering, Centre for Bioprocess Engineering for allowing me to use the microcalorimeter. Dr Marijke Fagan-Endres for training me to use the instrument, the support and valuable inputs. Lab manager, Tich.
- Department of Process of Engineering (University of Stellenbosch), National Research Funding (NRF) and Thermaspray (Pty) Ltd for project and personal funding
- Central Analytical Facility, Stellenbosch university
- My family for all the support and accommodating me while I was away from home. I am always thankful for your love. My friends, Cleopatra and Lunga for moral support and listening to all the things regarding this work.
- Bioprocess research group
- Department of process engineering, fellow students and support staff, Thobeka Mkwananzi , Sefakor Dogbe, Mr Alvin Petersen and Ms Francis Layman.

Dedication

I dedicate this work to my grandmother, Raphahle, my parents: Regina and Morris, and aunt, Lily, thank you so much for your encouragements, support and love. To my mother, you are the best. I am very grateful to have all of you in my life.

Table of contents

List of figures

List of tables

1 Introduction

1.1 Background and motivation

Proteins have significant applications in the food and feed industry as nutrition (Le Maux, et al., 2016). Proteins are used to produce nutritional supplements, aquaculture feed, plant fertilisers and food ingredients (Wisuthiphaet, et al., 2015). These applications have sparked interest in enzyme hydrolysis for protein recovery from wastes such as seafood and whey (Bhaskar, et al., 2008) (Hathwar, et al., 2011) (Kristinsson & Rasco, 2000) (Valencia, et al., 2014). In addition, enzymatic protein hydrolysis is used for protein modification to improve the chemical, physical as well as the functional properties of the different proteins sources, such as: whey, casein, sesame seed, corn gluten, for nutritional purpose (Demirhan, et al., 2011) (Apar & Ozbek, 2009) (Marquez-Moreno & Fernandez-Cuadrado, 1993) (Sinha, et al., 2007).

Enzymatic protein hydrolysis is used mostly in the food industry to convert food by-products into useful compounds that display nutritional properties in food systems like foaming and emulsification ability (Kristinsson & Rasco, 2000) (Pelegrine & Gasparetto, 2005). Products of enzymatic protein hydrolysis have mainly been studied due to their functional properties. These functional properties have been associated with peptides and amino acids produced during protein hydrolysis (Kristinsson & Rasco, 2000) (Sinha, et al., 2007). Usually, protein hydrolysis is performed using inorganic chemicals instead of enzymes. However, chemical hydrolysis produces products with low nutritional quality and reduced functionality (Adler-Nissen, 1986) (Sinha, et al., 2007) (Kristinsson & Rasco, 2000). The use of enzymes for protein hydrolysis maintains the nutritional quality of the product since it is carried out in milder biological conditions than the chemical methods and it preserves the original enantiomeric form of the product (Apar & Ozbek, 2009).

There are two ways in which enzymes catalyse protein hydrolysis. It can break the internal bonds or external bonds. Thus, enzymatic protein hydrolysis is known as a highly complex reaction due to the following factors:

1. The undefined nature of the substrate resulting from the diversity of the amino acids in the protein (Valencia, et al., 2015).

2. The multiplicity of the reactions, that is several peptide bonds are broken in parallel and in series simultaneously (Martínez-Araiza, et al., 2012), (Valencia, et al., 2015), (Qi & He, 2006) and (Marquez & Vazquez, 1999)
3. The possibility of substrate inhibition, product inhibition and enzyme inactivation during protein hydrolysis (Martínez-Araiza, et al., 2012), (Qi & He, 2006) and (Valencia, et al., 2015)
4. Multiple operating conditions including temperature, pH, ionic strength and substrate concentration on the reaction rate (Martínez-Araiza, et al., 2012) and (Qi & He, 2006).

All these factors make it difficult to develop a theoretical or fundamental mathematical model for enzymatic protein hydrolysis. Even so, several authors have developed kinetic models for enzymatic protein hydrolysis using empirical and Michaelis-Menten equations.

Recently, Valencia et al. (2015) proposed a novel modelling methodology for the enzymatic hydrolysis of proteins using an empirical equation obtained by Marquez-Moreno and Fernandez-Cuadrado (1993). This model successfully predicted enzymatic protein hydrolysis and was used on multiple sources of proteins and enzymes. Generally, empirical models do not give a fundamental understanding of enzymatic protein hydrolysis and the kinetic constants do not have physical meaning nor can they be extended beyond the range studied. On the other hand, the Michaelis-Menten mechanism is also used to model protein hydrolysis.

The Michaelis-Menten mechanism is a simple approach deduced from fundamentals of enzyme kinetics and it is used to model enzymatic hydrolysis of proteins (Valencia, et al., 2015). The well-known Michaelis-Menten equation for single substrate and single enzyme based on different assumptions has been used to model enzymatic hydrolysis of different proteins using various enzymes (Apar & Ozbek, 2009) (Qi & He, 2006) (Valencia, et al., 2014) (Demirhan, et al., 2011) (Marquez & Vazquez, 1999). However, during protein hydrolysis, the original substrate after being hydrolysed results in two products each of which can also be a new substrate susceptible to hydrolysis. This means that enzymatic protein hydrolysis involves multiple substrates (polypeptides) rather than a single substrate. Furthermore, it involves several peptide bonds broken in parallel and in series. Therefore, population balance and multiple substrates Michaelis-Menten approaches should be considered as an alternative for modelling complex enzymatic protein hydrolysis reactions.

There is a profound lack of population balance models for protein hydrolysis. A population balance approach has been implemented for enzymatic cellulose hydrolysis by Lebaz et al. (2015) as well as Hosseini and Shah (2011). The main difference between hydrolysing cellulose and protein is cellulose involves one type of monomer while proteins consist of different types of monomers (amino acids) which involves specificity. No studies were found in the literature using the population balance method for enzymatic protein hydrolysis. To study and validate protein hydrolysis reactions using experimental data, different laboratory methods are used to measure enzymatic protein hydrolysis.

Reaction kinetic models for enzymatic protein hydrolysis provide insight into the reaction rates as well as a measure of products (peptides) formed during the process. The methods for measuring enzyme hydrolysis include the degree of hydrolysis obtained using spectrophotometry and isothermal calorimetry techniques. The degree of hydrolysis is the percentage of the total number of peptide bonds in a protein, which have been cleaved during hydrolysis (Adler-Nissen, 1986). Isothermal calorimetry is a technique for real-time monitoring of chemical, physical and biological processes which is used to measure reaction rates (Bianconi, 2007) (Todd & Gomez, 2001). There are two types of isothermal calorimetry, namely Isothermal Titration Calorimeter (ITC) and isothermal microcalorimetry (IMC). The difference between IMC and ITC is that IMC is a batch technique while ITC is a closed system which consists of an injection syringe. Although a few authors studied enzymatic protein hydrolysis using ITC, no single study was found that uses isothermal microcalorimetry IMC (Maximova & Trylska, 2015).

In light of the above-mentioned considerations, the problem of the study is how to predict complex enzymatic protein hydrolysis reactions? In this study, it is proposed to model enzymatic protein hydrolysis using a population balance approach along with modified Michaelis-Menten for multiple substrates. In addition, a measure of the heat produced by enzymatic protein hydrolysis may be quantified using IMC.

1.2 Aim and objectives

This study aims to propose a theoretical model that predicts the degree of hydrolysis over time for enzymatic protein hydrolysis. The objectives of the study are as follows:

1. To experimentally determine the optimal hydrolysis conditions for two enzyme-substrate combinations;
2. To investigate the effect of an enzyme to substrate ratio and buffer type on the total heat released during enzymatic protein hydrolysis for two enzyme-substrate combinations;
3. To compare the degree of hydrolysis data obtained from spectrometry and total released data attained from calorimetry;
4. To propose mathematical models for enzymatic protein hydrolysis using population balance approach;
5. To validate the proposed model for two enzymes using both degree of hydrolysis and total heat released data.

1.3 Scope of the study

To model complex enzymatic protein hydrolysis reactions, whey protein concentrates and two commercial enzymes, namely papain and bromelain, were used as the model system, considering the following outline.

1. Whey protein has been well studied and the composition has been identified.
2. Papain and bromelain are two different types of enzymes, exopeptidase and endopeptidase respectively, possesses high specificity and are commercial enzymes.

To achieve the first objective, the hydrolysis of whey protein by papain and bromelain were studied using a central composite design at different operating conditions (protein concentration, enzyme concentration, temperature and pH) to determine optimum conditions defined by the desirability plots. The degree of hydrolysis was measured using OPA (o-phthaldialdehyde) method for experimental studies during optimization.

To achieve the next two objectives, the experiments were performed using a factorial design at three enzyme concentrations with two different buffer types. The degree of hydrolysis, as well as heat flow, were measured as the reaction progress. The heat flow data were obtained upon adding of the substrate and enzyme at the optimum condition in an isothermal microcalorimeter. The degree of hydrolysis and heat flow data was compared using the error bars.

To achieve the fourth objective, a mathematical model was proposed to predict enzymatic protein hydrolysis. The model was based on the population balance approach as well as

multiple substrate Michaelis-Menten equations. The model had the total number of peptide bonds or primary amino acids and a total number of specific or preferred sites as inputs. The model predicted the concentrations of chain length, the degree of hydrolysis and heat flow as a function of time.

To achieve the last objective, the predicted degree of hydrolysis and heat flow curves were compared to the experimental data to determine the accuracy of the model.

1.4 Outline of the thesis

The thesis consists of five chapters. The current chapter introduces the study. A review of the literature is presented in chapter two. First, the description of proteins as well as enzymes is provided. The different methods for protein hydrolysis are reviewed. Furthermore, the process kinetics used to model enzymatic protein hydrolysis are also discussed. Finally, the two types of methods used to measure enzymatic protein hydrolysis are presented.

Chapter three focus on the methodology, which consists of two sections. The first section aims to generate data to validate the proposed model. The section describes the planning of the experiments, methods, materials as well as equipment used to generate the data. In section two, a model for enzymatic protein hydrolysis is proposed and explained. The contents provide a theoretical background on the process, a mathematical explanation of enzymatic protein hydrolysis mechanism, a description of the equations as well as methods used in modelling protein hydrolysis and the implementation of the model.

In chapter four, the results obtained in chapter three are presented and discussed. The results include the effects of operating conditions on enzymatic protein hydrolysis, a comparison between the two methods used to measure protein hydrolysis reaction and the results from the proposed model.

Finally, the last chapter provides a summary of the conclusions of the study, limitations of the model as well as the recommendations for future research work.

2 Literature review

2.1 Proteins

Proteins are biopolymers that are essential for living organisms and represent a vast group of complex nitrogenous molecules widely present in plants and animals (Ahluwalia, et al., 2006). They are defined as macromolecules consisting of one or more polypeptides. Polypeptides are made of chains of amino acids connected together in a specific sequence by peptide bonds, with each protein characterised by a unique amino acid sequence. Amino acids are simple organic compounds made up of a carboxyl (-COOH), amino ($-NH_2$) and side chain (R) groups.

2.1.1 Protein structure

Proteins consist of amino acids linked through a peptide bond to form a long chain (Walsh, 2002). Their structural and functional diversity is caused by different combinations of the amino acids. There are twenty basic amino acids which make up proteins in nature (Walsh, 2002). The sequence of the amino acids and bonding of the side groups determine the structure of proteins. Figure 2-1 shows the common groups in the amino acids. All the amino acids are similar except in the R- group. Figure 2-2 shows the generic structure of a peptide or polypeptide.

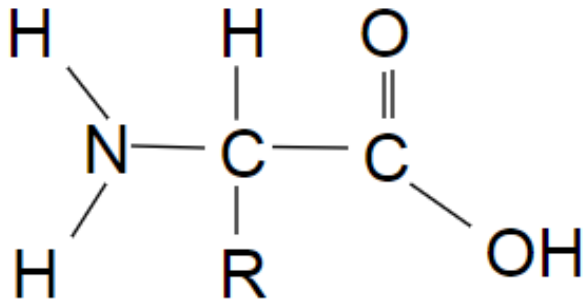


Figure 2-1: Common groups in amino acids

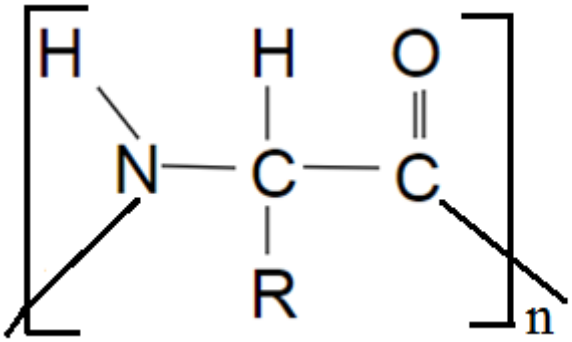


Figure 2-2: Generic structure of a peptide

Proteins have four levels of structures. These include primary, secondary, tertiary and quaternary. A primary structure describes a linear sequence of different amino acids linked together by peptide bonds. Secondary structure refers to the folding of a polypeptide chain due to mainly hydrogen bonds and backbone interactions. Tertiary structure refers to the three-dimensional shape of the protein. Quaternary structure denotes a protein macromolecule formed by interactions between numerous polypeptide chains. In this study, a protein is described as a linear sequence of different amino acids linked together by peptide bonds.

2.1.2 Protein properties

Proteins perform important and different roles in the industry. Proteins show many functional properties governed by their physiochemical activities. Among other functional properties, protein solubility is important because it has a major influence on other functional properties such as emulsification and foaming (Kristinsson & Rasco, 2000) (Pelegrine & Gasparetto, 2005).

Protein solubility is the maximum concentration of proteins in the solvent in a two-phase system (liquid and solid). Solubility relates to the hydrophobic and ionic interactions. Hydrophobic interactions support protein-protein interactions which decrease solubility by forming solids. Ionic interactions promote the protein-water interactions increasing the solubility. The solubility of a protein depends mainly on pH and temperature. The pH of the solution affects the distribution and nature of the charge on the protein (Ahluwalia, et al., 2006). At extreme acidic or alkaline conditions, the proteins unfold and exposing more hydrophobic molecules. The process of unfolding is called denaturing.

Protein denaturation is the alteration of the secondary and tertiary, quaternary structures by breaking down or rupturing the noncovalent interactions between the side chains of the amino acids residues (Ahluwalia, et al., 2006). However, the process of denaturation does not affect the primary structure i.e. the amino acid sequence remains the same (Ahluwalia, et al., 2006). Proteins are denatured by the effect of temperature and pH as explained above. According to Pelegrine and Gasparetto (2005), proteins denature at high temperatures and protein solubility decreases at high temperatures. This is because denaturation leads to aggregation (Pelegrine & Gasparetto, 2005). Therefore, it is important to carefully choose the process conditions for protein hydrolysis.

2.1.3 Whey protein

Whey protein is a by-product that is obtained from milk during cheesemaking. Whey protein concentrate (WPC) and whey protein isolate (WPI) are products derived from whey. WPC and WPI differ in protein content which ranges from 30 to 90 % (the rest made up of fat, ash, lactose). Hence, whey protein consists of a large number of amino acids and most of the essential amino acids (Engelen, et al., 2012).

One of the properties of whey is good solubility over wide pH ranges which is important to perform hydrolysis. However, it is known that whey protein does not maintain good solubility during high heat treatment. Pelegrine and Gasparetto (2005) studied the effect of temperature (40-60 ºC) and pH (3.5-7.8) on the solubility of whey. The authors found out that at the isoelectric point of whey protein, 4.5, the solubility is low and it decreases with increasing temperature.

2.2 Protein hydrolysis

Protein hydrolysis is the breaking down of the peptide bonds in a protein macromolecule into peptides of different sizes and free amino acids. The peptides and amino acids are produced for a wide variety of uses, including as fertilizers, aquaculture, and nutritional supplements in the food industry (Kristinsson & Rasco, 2000).

Most protein hydrolysis processes are carried out in batch processes in the industry (Kristinsson & Rasco, 2000). Furthermore, a batch process is simply controlled, and its behaviour can be

easily studied under laboratory experiments. Other protein hydrolysis processes that are considered are the use of immobilised enzyme (Sousa, et al., 2004).

Mostly, proteins are hydrolysed by chemical or biological methods. However, products obtained from the two methods are different in quality and quantity (Wisuthiphaet, et al., 2015). The chemical methods are acid and alkaline hydrolysis. The biological method is enzymatic hydrolysis. These methods are explained in detail in the next section.

2.2.1 Protein hydrolysis methods

2.2.1.1 Acidic and alkaline hydrolysis

Acid hydrolysis is a conventional method for generating protein hydrolysates by using hydrochloric or sulfuric acid (Wisuthiphaet, et al., 2015). It is preferred over alkaline hydrolysis (Kristinsson & Rasco, 2000) as the acid can be used both as liquid as well as gas phases (Fountoulakis & Lahm, 1998). Acid hydrolysis method is inexpensive and requires less time. However, this process is severe and difficult to control (Ovissipour, et al., 2009). Also, it destroys the essential amino acids in the protein hydrolysates such as tryptophan, tyrosine and cysteine thus yielding products with reduced nutritional quality (Kristinsson & Rasco, 2000).

Even though this method has several disadvantages, the United State of America uses this process to hydrolyse vegetable proteins for flavour and taste enhancements (Kristinsson & Rasco, 2000). Acid hydrolysis to produce hydrolysed vegetable protein is well known and their use as functional ingredients in the food industry is nearly a century old (Adler-Nissen, 1986). In addition, fish protein hydrolysates formed from this process is used as an additive in animal feed, culture media and plant fertilisers.

The alkaline method uses basic solutions to produce hydrolysates. The use of alkali chemicals, such as sodium hydroxide, to break down a protein often results in poor functionality and more importantly can affect the nutritive value of the products (Kristinsson & Rasco, 2000). In short, acid hydrolysis and alkaline methods are not ideal for producing protein hydrolysates due to its harsh conditions.

2.2.1.2 Enzymatic hydrolysis

Enzymatic protein hydrolysis is an attractive technology and an important biological process used to produce and improve physical, chemical and functional properties of natural as well as synthetic proteins (Martinez-Araiza, et al., 2012). Enzymatic protein hydrolysis is an effective method for breaking the peptide bonds without destroying the essential products and residue (Fountoulakis & Lahm, 1998). This is because it is carried out in milder biological conditions, which maintains the nutritional quality of the amino acids (Apar & Ozbek, 2009) (Demirhan, et al., 2011). Enzymatic protein hydrolysis is also an alternative way of making food proteins using natural products. However, its disadvantage is the need for inactivation of the enzyme using pH or temperature treatment after obtaining the required degree of hydrolysis (Kristinsson & Rasco, 2000). This step tends to denature the protein.

Recently, enzymatic protein hydrolysis has become a commonly used method to obtain plant proteins and commercial protein hydrolysates with improved functional properties (Ahmadifard, et al., 2016). Plant protein hydrolysates are used commonly as protein ingredients or supplements in food and beverages (Ahmadifard, et al., 2016). Enzymatic protein hydrolysis has also been applied to fish as an alternative to reuse or utilise the fish processing waste (Kristinsson & Rasco, 2000). Some of the protein hydrolysates application obtained from the fish waste is used primarily as feed for enhancing nutrition in aquaculture (Wisuthiphaet, et al., 2015). The general process for producing enzymatic protein hydrolysate is shown in Figure 2-3. Figure 2-4 shows the role of water during the breaking of a peptide bond in a protein.

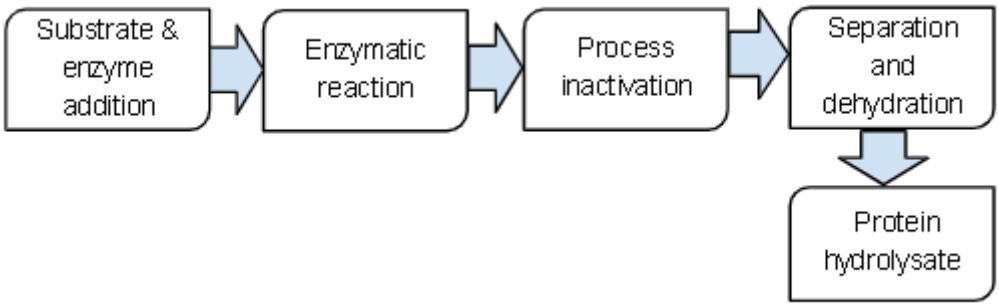


Figure 2-3: The general process of enzymatic hydrolysis of proteins (Adapted from Kristinsson & Rasco (2000))

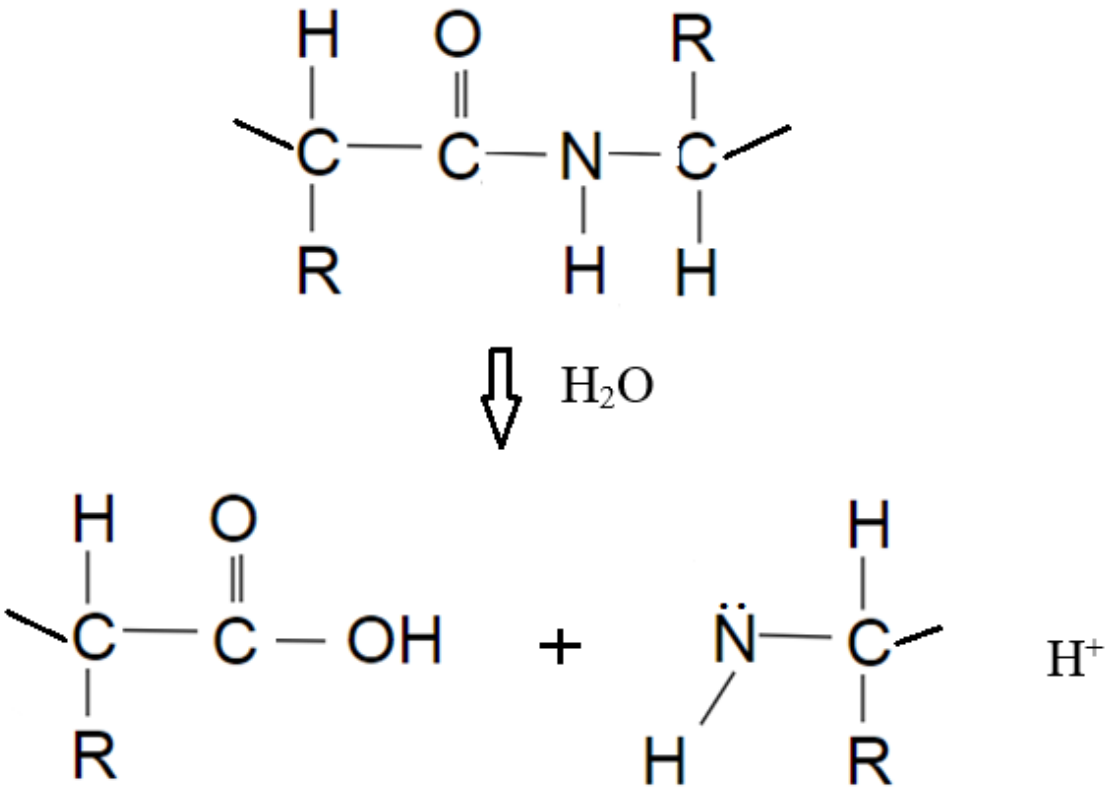


Figure 2-4: Enzymatic protein hydrolysis mechanism

2.2.2 Factors affecting hydrolysis and enzyme activity

Several authors proved that the enzymatic hydrolysis of proteins is characterised initially by fast reaction rate where most of the peptide bonds are cleaved. Therefore, the reaction rate decreases and reaches a steady state. The formation of products as time progress or substrate depletion is initially linear then the rate decline at longer times (Eisenthal & Danson, 2002). Several authors reported that substrate depletion, as well as enzyme concentration, could possibly be the reason for this departure from linearity and temperature and pH affects the reaction rate (Marquez & Vazquez, 1999) (Apar & Ozbek, 2009) (Demirhan, et al., 2011) (Butré, et al., 2014) (Valencia, et al., 2014). The effects of these factors are discussed below.

2.2.2.1 Substrate concentration

It is mostly known that the reaction rate or products formation slows down during a reaction because of substrate depletion. This might explain why the reaction rate slows down as protein hydrolysis progress since as the substrate concentration decreases as shown in Figure 2-5, the enzyme molecules get less engaged with the substrate thus decreasing the reaction speed. If the substrate slows down the reaction rate, the addition of more substrate or the use of excess initial substrate concentration should delay the fall out of the reaction rate. However, a high concentration of the substrate might result in inhibition thus slowing down the enzyme activity and reaction rate instead of increasing (Eisenthal & Danson, 2002) (Zapata-Montoya, et al., 2018) (Butré, et al., 2012) (Butré, et al., 2014).

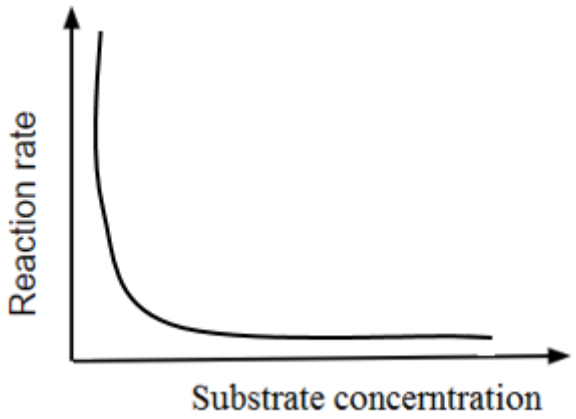


Figure 2-5: The effect of substrate concentration on the enzyme activity at a constant Enzyme concentration

2.2.2.2 Enzyme concentration

As enzymes are catalysts, the reaction rate is expected to be directly proportional to the concentration of the enzyme as shown in Figure 2-6 (Valencia, et al., 2014) (Valencia, et al., 2015). However, there are cases where the enzyme concentration is not proportional to the initial reaction velocity. In such cases where the linearity is not valid, this may suggest changes in pH or ionic strength of the mixture as increasing amounts of the enzyme solution is added (Eisenthal & Danson, 2002).

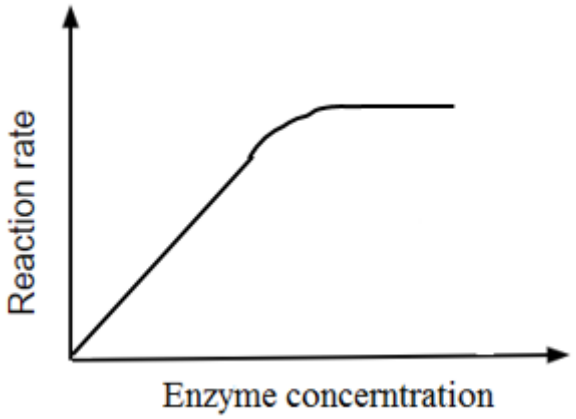


Figure 2-6: The effect of enzyme concentration on the enzyme activity at a constant enzyme concentration.

2.2.2.3 Temperature

According to Ovissipour et al. (2009), the reaction rate will increase with increasing temperature. This is because the activity of enzymes is dependent on temperature. However, all the enzymes deactivate if they are heated above its optimal temperature range and the structure of the enzyme is changed with the loss of catalytic activity (Cornish-Bowden, 2012), subsequently decreasing the reaction rate. This implies that there exists an optimum or maximum temperature, as shown in Figure 2-7, where the reaction rate is high during enzymatic protein hydrolysis. Marquez and Vazquez (1999) reported that the influence of temperature on enzymatic protein hydrolysis follows the Arrhenius equation only up to a certain point.

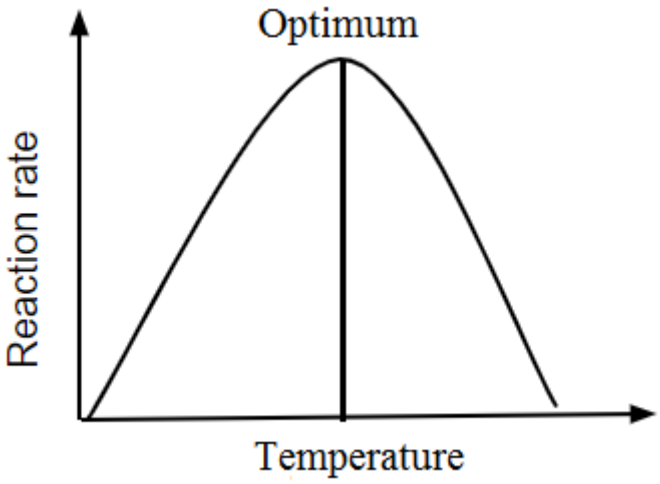


Figure 2-7: The effect of temperature on the enzyme activity.

2.2.2.4 pH

A change in pH affects both the substrate and the enzyme by changing the charge distribution and structure of the molecules (Adler-Nissen, 1986). This means that an increase or decrease in pH can inactivate or deactivate as well as affect the normal function of the enzyme. The effect of pH on the reaction rate is presented in Figure 2-8. The curve shows that the activity of the enzyme is maximum at the optimum pH. The enzyme deactivates above and in certain cases below the optimum pH range. It is known that enzymatic protein hydrolysis is influenced by the changes in pH. A number of authors studied the hydrolysis of proteins under constant pH conditions. No study was found that mathematically related the pH to the reaction rate or the kinetic constants.

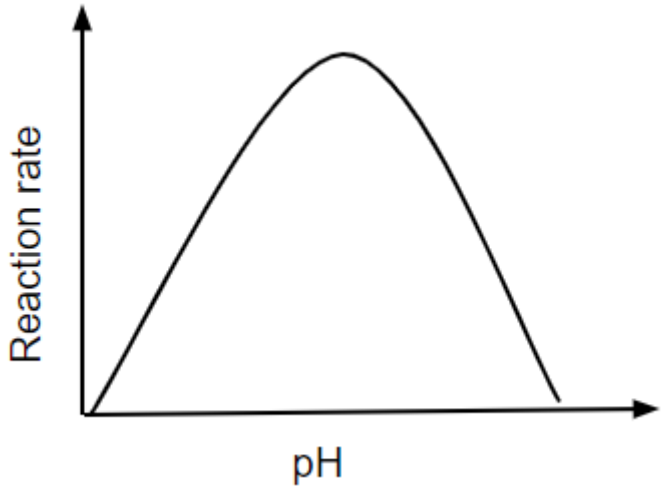


Figure 2-8: The effect of pH on the enzyme activity.

2.3 Enzymes

Enzymes are used currently in a different and growing number of industries - from dairy products to detergents; from animal feed to alcohol production; from meat tenderization to starch processing (Adler-Nissen, 1986) (Walsh, 2002). They are present in all animals, organisms, and plants, to help break down various foods into components that are readily utilized. Enzymes are derived from animals, plants and microbial sources.

Enzymes are proteins which act as biological catalysts. They are able to change the rates and the extent of biochemical reactions. In the absence of enzymes, most biochemical reactions in living organisms occur at a slower rate. Enzymes are considered as a true catalyst that remain unchanged at the end of the reaction. This does not imply that the enzymes play an inert role and is unchanged during the reaction, but that any such changes in either the physical or chemical state of the enzyme are reversible within the operational conditions (Tucker & Woods, 1995). It is therefore important to use optimum conditions in order to keep the functional properties and to achieve excellent enzyme activity or high enzyme performance.

Enzymes have defined optimal ranges of operation (e.g. temperature, pH), and these optima are specific for each enzyme-substrate combination. According to Bhaskar et al. (2008) various conditions like pH, time, temperature, and enzyme to substrate percentage influence the enzymatic activity co-operatively and thus offers the possibility to control the hydrolysis

process. Enzymes catalyse specific reactions and in some instances require specific functional groups and/or binding sites in molecules. Therefore, their commercial use is also specific.

2.3.1 Characteristics of enzymes

There are six main classes of enzymes. These include oxidoreductase, transferases, lyases, isomerases, ligases and hydrolases. As the focus of this study is the development of models for enzymatic proteolysis, the hydrolases will be used to initiate the reaction. More specifically, proteases will be used for protein hydrolysis. Proteolytic enzymes break down bonds in a protein or polypeptides to form even smaller peptides or amino acids depending on the catalytic action (Palmer, 1995) (Walsh, 2002).

Proteases differ in their catalytic action namely endopeptidase or exopeptidase and nature of the catalytic site. Endopeptidases hydrolyse non-terminal amino acids of a protein while exopeptidases hydrolyse the terminal peptide bonds. The enzyme choice depends on the desired final characteristics of the hydrolysed protein. However, sometimes the endopeptidases are combined with exopeptidase in order to achieve a complete process in the industry (Adler-Nissen, 1986). Usually, protein hydrolysis processes at an industrial scale use commercial enzymes.

Commercial enzymes are available at high volumes and lower cost than non-industrial enzymes (Adler-Nissen, 1986). They can contain mixtures of different enzymes for efficiency from an application point of view (Adler-Nissen, 1986). This means that commercial enzymes may consist of both endopeptidase and exopeptidase catalytic mechanism. But, commercial enzymes have a lower purity.

In summary, the different commercial enzymes can be used to catalyse proteins from different sources. Enzymes catalyse protein hydrolysis in three ways depending on the type of enzyme used: endopeptidase, exopeptidase and endopeptidase-exopeptidase. These enzymes are specific, which means that they hydrolyse peptide bonds next to one or two certain amino acids. Furthermore, enzymes have different optimum temperature and pH ranges.

2.3.2 Proteolytic enzymes

Proteolytic enzymes are also known as peptidases, proteases or proteinases. Peptidases are mostly classified by the position of the peptide bond it hydrolyses (endopeptidase or

exopeptidase) and molecular mechanism by which hydrolysis is achieved (Palmer, 1995) (Walsh, 2002). Based on the molecular mechanism, peptidases can be classified into four groups: cysteine protease, serine protease, aspartic protease and metalloprotease (Adler-Nissen, 1986) (Walsh, 2002).

Cysteine proteases are characterised by the presence of a cysteine and histidine amino acid at the enzyme's active site. Serine protease involves serine residue at the enzyme's catalytic site. Aspartic protease are acidic proteases that use an aspartic acid at the active site for catalysis. Metalloprotease are characterised by the presence of metal ion.

2.3.3 Enzymes: papain and bromelain

The two enzymes chosen for our study are bromelain and papain. The enzymes were chosen based on their different catalytic action. Bromelain displays endopeptidase activity while papain displays both endopeptidase-exopeptidase activity.

Bromelain is a protease derived from pineapple stem and the fruit itself. Maurer (2001) and Elavarasan and Shamasundar (2016) stated that bromelain preferentially cleaves glycyl, leucyl and alanyl. Adler-Nissen (1986) reported that the typical optimum operating conditions for bromelain are pH in the range 5-8. Papain is a cysteine protease recovered from papaya. According to Elavarasan and Shamasundar (2016), papain is specific to leucine, glycine, arginine, lysine and phenylalanine. The optimal pH operating conditions for papain are between 5-7. The optimum conditions are subject to change depending on the substrate used and the conditions are often recommended by the supplier.The supplier recommends pH between 4.8-6.2 for papain and pH in the range 4-8 for bromelain.

2.4 Reaction kinetic models for enzyme protein hydrolysis

There are reaction kinetics models for enzymatic protein hydrolysis in the literature (Marquez & Vazquez, 1999) (Marquez-Moreno & Fernandez-Cuadrado, 1993) (Valencia, et al., 2015) (Zapata-Montoya, et al., 2018) (Qi & He, 2006). Currently, enzyme kinetics are used to model protein hydrolysis. However, many assumptions are made to simplify the process to mathematically model it and it is important that the assumptions keep the fundamentals of the process. Three types of reaction kinetic models are used to model protein hydrolysis, namely enzyme kinetics, population balance approach and empirical models.

Enzyme kinetics does not necessarily predict enzymatic protein hydrolysis since the equation models single substrates while protein hydrolysis exhibits multiple substrates during the reaction. Since it is a challenge to model protein hydrolysis, Marquez-Moreno and Fernandez-Cuadrado (1993) reported an empirical model for protein hydrolysis. However, empirical models are derived from the experimental data hence they don't possess a theoretical background of the process. In addition, the constants do not have any physical meaning. Recently, Lebaz et al. (2015) and, Hosseini and Shah (2011) proposed a population balance approach for modelling enzymatic hydrolysis for cellulose. The current study will propose a model for enzymatic hydrolysis for proteins using the population balance approach. The following three sections give more information on enzyme kinetics, population balance approach and empirical models.

2.4.1 Enzyme kinetics

The study of enzyme kinetics is most commonly used to model and investigate enzymes. Enzymes are known as efficient catalysts since they are more active than inorganic catalysts; therefore, they are usually present in a lower concentration than the substrate concentration (Hammes, 1982). This means that all the enzyme-substrate complex can be assumed to be in equilibrium after a very short induction period (Hammes, 1982).

The simple steady mechanism of the enzymatic protein hydrolysis process proceeds according to the following sequence of elementary reaction equations as proposed by Michaelis and Menten in 1913 (Blanch & Clark, 1998):

$$E + S \overset{k_1}{\rightarrow} ES \tag{2-1}$$

$$ES \overset{k_2}{\rightarrow} E + S \tag{2-2}$$

$$ES + w \overset{k_{cat}}{\longrightarrow} P + E \tag{2-3}$$

Where S, E, ES, w and P represent substrate, free enzyme, enzyme-substrate complex, water and products respectively. The first step includes the binding of the enzyme to the peptide bond in the substrate forming an enzyme-substrate complex. In the second step, the enzyme-substrate complex can dissociate or reacts with water to form products which can act as new substrate and excess substrate according to an irreversible reaction shown in eq. 2-3. It is assumed that

the ES complex forms fast and the rate of the reverse reaction of the second step is negligible. The above mechanism or reaction is described by the Michaelis-Menten equation.

The Michaelis-Menten equation is a simple model used to model enzymatic hydrolysis of proteins. Numerous studies were found in modelling enzymatic hydrolysis process using the Michaelis-Menten approach (Apar & Ozbek, 2009) (Demirhan et al., 2011) (Marquez & Vazquez,1999) (Valencia et al, 2014). Below is the derivation of the Michaelis-Menten reaction rate equation.

For the mechanism in eqns. 2-1 to 2-3, the rate of product formation is described by eq. 2-4 and the equilibrium constant by eq. 2-5.

$$v = \frac{d[P]}{dt} = k_{cat}[ES] \tag{2-4}$$

$$k_m = \frac{k_2}{k_1} = \frac{[E][S]}{[ES]} \tag{2-5}$$

Where v is the substrate consumption to form ES complex or the rate of product formation and k_{cat} , k_1 and k_2 are the kinetic constants. However, the free enzyme (E) and the enzyme-substrate concentration cannot be measured. Therefore, in the absence of enzyme denaturation, the total enzyme concentration or initial enzyme concentration (E_o) is constant and given by eq. 2-6:

$$[E_o] = [E] + [ES] \tag{2-6}$$

Rearranging eq. 2-6 and substituting in eqns. 2-5 and 2-6 gives eq. 2-7. Finally, substituting eq. 2-7 into eq. 2-4 gives the reaction rate equation described in eq. 2-8.

$$[ES] = \frac{[E_o][S]}{k_m + [S]} \tag{2-7}$$

$$v = \frac{k_{cat}[E_o][S]}{k_m + [S]} \tag{2-8}$$

$$v_{max} = k_{cat}[E_0] \tag{2-9}$$

Where k_m is the dissociation constant of the ES complex, also known as the Michaelis-Menten constant. v_{max} shown in eq. 2-9 is the maximum velocity when the substrate concentration is very large, and the enzyme is saturated by the substrate.

The chemistry of enzymatic hydrolysis is not yet established. Enzymatic hydrolysis of proteins is a very complex process as there are different reaction components (products and reactants) as well as the breaking of peptide bonds in parallel and series. This implies that products can become new substrates making it multiple substrate reactions.

The simple Michaelis-Menten equation for the single enzyme and single substrate does not describe enzymatic protein hydrolysis processes. This study will present a modified Michaelis-Menten equation for multiple substrates and a single enzyme to describe the reaction rate. Multi-substrate Michaelis-Menten kinetics was proposed by Palmer (1995). Qi and He (2006), Zapata-Montoya, et al. (2018) and, Barros and Malcata (2004) studied the double substrate Michaelis-Menten kinetics.

2.4.2 Population balance

Protein hydrolysis produces products that also become new substrates. This means that hydrolysis involves multi-substrate reactions. The idea of using population balance approach for enzymatic cellulose hydrolysis was first proposed by Hosseini and Shah (2011) using multiple equations. Then, Lebaz, et al. (2015) reported a simple and single model for cellulose hydrolysis using the population balance approach.

Population balance modelling describes the behaviour of any systems with population distribution. It is used to predict the population of bubbles, polymers and particles. Enzyme hydrolysis of proteins is the decrease in the polymer chains because of the breaking of peptide bonds. The formation of particle size distribution due to break up and constant growth is described by eq. 2-10 as shown below (Lebaz, et al., 2015).

$$\frac{\partial n(x)}{\partial t} = \frac{\partial [G(x)n(x)]}{\partial x} - n(x)T(x) + \int_{x}^{\infty} \beta(x, x_k)T(x_k)n(x_k)dx_k \qquad (2\text{-}10)$$

Where $n(x)$ represents the number density as a function of chains with length x. G is the growth rate for chains of length x. $T(x)$ is the breakage frequency for a chain of length x. $\beta(x, x_k)$ is

the number of new chains formed in length x due to breakage of chain x_k. The first term on the right side is the accumulation related to the growth of particles, the second term is the loss or death of particles and the last term corresponds to the formation or birth of new particles.

The study will propose a model for the protein hydrolysis system using population balancing by considering concentrations of polymer chains and the two types of enzymes (endopeptidase and exopeptidase). The population balance method is different for endopeptidase and exopeptidase by the term $\beta(x, x_k)$. This is because endopeptidase and exopeptidase catalyse a polymer chain in different ways. In the context of the protein hydrolysis, the growth rate $G(x)$ in eq. 2-10 is equal to zero since protein hydrolysis is a pure breakup process. In addition, eq. 2-10 will be applied in the discrete form shown in eq. 2-11 as developed by Kumar and Ramkrishna (1996).

$$\frac{dN_i}{dt} = -T_i N_i(t) + \sum_{k=i+1}^{M} n_{ik}\, T_k N_k(t) \qquad (2\text{-}11)$$

Where N_i is the total number of polypeptide or amino acid with chain length i, t denotes the reaction time, T_i is the breakage frequency of the polypeptide with chain length i, T_k defines breakage frequency of the polypeptide with chain length k which break to form new chains length i, and the term n_{ik} refers to the probability of breaking a chain with length k to form a new chain of length i.

2.4.3 Empirical models

Empirical models can be used to model enzymatic protein hydrolysis. There are pure empirical and semi-empirical models. Pure empirical models are not derived from theoretical mechanisms and the constants do not have physical meaning. However, they offer a simple description of the processes. Semi-empirical are based on a combination of empirical observations and theoretical fundamentals.

For enzymatic protein hydrolysis, Marquez-Moreno and Fernandez-Cuadrado (1993) proposed a pure empirical kinetic equation. The kinetic model contained an exponential equation for predicting the degree of hydrolysis presented by eqn. 2-12. The analytical solution is as shown in eq. 2-13.

$$\frac{\mathrm{dDH}}{\mathrm{dt}} = \mathrm{ae}^{-\mathrm{bDH}} \tag{2-12}$$

$$DH = \frac{ln(abt)}{b} \tag{2-13}$$

Where DH is the degree of hydrolysis, t is time, a and b are the kinetic constants.

Marquez-Moreno and Fernandez-Cuadrado (1993) further reported that the parameter a is dependent on the enzyme concentration and substrate concentration. The parameter b does not change when the enzyme and substrate concentrations vary. Qi and He (2006) derived different mechanisms to determine the equation for the kinetic parameters a and b. The authors found that parameter a changed with variation in enzyme concentration and parameter b changed with substrate concentration. However, the authors did not describe the kinetic parameters as the function of the experimental or process operating conditions.

Valencia et al. (2015) proposed a methodology for describing the kinetic parameters as a function of the operating conditions. This model showed good predictability of enzymatic protein hydrolysis. The authors used eqn. 2-12 to predict the degree of hydrolysis and the response surface method for describing the kinetic constants. In addition, Valencia et al. (2015) concluded that the parameter a is affected by the substrate concentration, enzyme concentration and temperature while the parameter b is affected by only substrate concentration. This disagrees with what was reported by Marquez-Moreno and Fernandez-Cuadrado (1993) but agrees with Qi and He (2006). Recently, Zapata-Montoya et al. (2018) adjusted the kinetic parameters in eqn. 2-12 using substrate and product inhibition mechanisms proposed by Qi and He (2006) to describe the enzymatic hydrolysis of protein in the red tilapia viscera.

In summary, although the empirical models shows good correlation with the experimental data. The models are not derived from the fundamental of enzymatic protein hydrolysis. In addition, the kinetic parameters do not have any physical meaning.

2.5 Methods for measuring enzymatic protein hydrolysis

Reaction kinetics for enzyme protein hydrolysis is an informative study of enzyme catalysed reactions. These studies predict how the reactions proceed and highlight the factors affecting the process. However, most methods for measuring hydrolysis are indirect, which implies that

to allow fast process control, high cost and long-time delay are involved to measure protein hydrolysis reactions.

The methods for measuring enzymatic protein hydrolysis reaction to investigate enzyme kinetics are spectrophotometry, pH stat and isothermal microcalorimetry (IMC) methods. The spectrophotometry methods measure the concentration of terminal amino groups from which the degree of hydrolysis is deduced. The pH stat technique measures the consumption of the base during the reaction to maintain pH constant and relates it to degree of hydrolysis. Isothermal microcalorimetry measures the heat produced or consumed during a reaction. In this study, an isothermal microcalorimetry (IMC) and spectrophotometry methods will be used instead to measure enzymatic protein hydrolysis.

2.5.1 Degree of hydrolysis

Protein hydrolysis extent is defined as the degree of hydrolysis. The degree of hydrolysis (DH) is the percentage of the total number of peptide bonds in a protein, which has been cleaved during hydrolysis as shown in eq. 2-14 (Adler-Nissen, 1986).

$$DH = \frac{h}{h_{tot}} \times 100 \qquad (2\text{-}14)$$

Where h is the number of the hydrolysed bond at a certain time and h_{tot} is the total number of peptide bonds in the particular protein substrate.

Several methods for determining the degree of hydrolysis have been described in the literature. Spellman, et al. (2003) reported that pH-stat, trinitro-benzene-sulfonic acid (TNBS) and o-phthaldialdehyde (OPA) are the most commonly used methods to determine DH during food protein hydrolysis.

The pH-stat technique monitors the amount of base added during the reaction to keep the pH constant and relates it to DH (Adler-Nissen, 1986). However, the pH-stat method is limited to pH conditions greater than 7 (Adler-Nissen (1986). The TNBS method is based on the primary reaction of the amino acids with TNBS reagent (Adler-Nissen, 1979). But, the TNBS reagent is unstable, toxic and must be handled carefully due to the risk of explosion (Nielsen, et al., 2001). An advantage of the OPA method is it is safe and non-toxic as compared to the TNBS method.

The OPA method is similar to the TNBS method, but o-phthaldialdehyde is used as a reagent instead of trinitro-benzene-sulfonic acid. To measure DH, O-phthaldialdehyde reagent reacts with amino groups formed during hydrolysis to form a compound that will absorb light at 340 nm (Nielsen, et al., 2001). In this study, the OPA method was chosen to measure enzymatic protein hydrolysis.

2.5.2 Microcalorimetry

All chemical, physical and biological processes result in either production or consumption of heat. A microcalorimeter is used to measure the amount of heat produced or consumed during a reaction and it can be able to detect a small amount of heat in nanowatt and microwatt. There are different types of calorimetry namely bomb, dynamic, indirect and direct calorimetry. For this study, a batch isothermal microcalorimetry (IMC) a type of direct calorimetry will be used to investigate enzyme protein hydrolysis. Another kind of direct calorimetry is isothermal titration calorimetry (ITC). The difference between IMC and ITC is that with ITC multiple injections of the enzyme can be introduced into the reaction cell containing protein solution.

2.5.2.1 Isothermal microcalorimeter

Microcalorimetry deals with heat flow measured in nanowatt and microwatt hence the name microcalorimeter. IMC keeps the temperature or the surrounding of the system constant while the heat flow is measured using a thermostat. A thermostat is a liquid-based system using water or oil to dissipate heat and minimise temperature changes in the system of interest.

The main advantage of a microcalorimeter is that it produces a real-time heat flow data. Also using a calorimeter in enzyme-catalysed reaction enables studying a direct assay without the requirement of using modified substrates (Bianconi, 2007). On the other hand, spectroscopic techniques use modified substrates to determine the reaction kinetics. Williams and Toone (1993) and Maximova and Trylska (2015) reported that there is a good correlation between calorimetric and spectrophotometric data for reaction kinetics. Maximova and Trylska (2015) reported that ITC can be a comparable and a faster way to determine the kinetic constants than standard spectrophotometric.

2.5.2.2 Application of isothermal microcalorimetry in enzyme kinetics

Isothermal microcalorimetry is used to determine thermodynamic and kinetic parameters for biological processes (Falconer, 2016) (Mazzei, et al., 2016). It is a direct approach for studying complex reactions such as enzyme-catalysed reactions, as a microcalorimeter directly measures the heat generated or absorbed during a reaction. Because all the reactions are associated with heat produced or consumed as the reaction progress, it is possible to obtain kinetic using the calorimetry technique (Bianconi, 2007) (Eftink, et al., 1981). In this section, the application of isothermal microcalorimetry on the study of enzyme-catalysed reactions is discussed.

There are several studies related to the application of calorimetry to the study of enzyme-catalysed reaction. Todd and Gomez (2001) determined enzyme kinetics constants using an isothermal titration calorimetry (ITC). Stodeman and Schwarz (2002) studied the hydrolysis of N-acetyl-L-methionine, N-acetyl glycine, N-acetyl-L-phenylalanine, and N-acetyl-L-alanine at 298.35K by porcine kidney acylase in a microcalorimeter. The authors determined the reaction enthalpies of N-acetyl-L-phenylalanine and N-acetyl-L-methionine. Recently, Maximova and Trylska (2015) used ITC to determine the hydrolysis kinetics of an insoluble macromolecule substrate (Casein) and a small substrate (Na-benzoyl-DL-arginine b-naphthylamide) by trypsin. Maximova and Trylska (2015) determined a heat of reaction for hydrolysis of casein using trypsin to be 103.8 kJ/ mol. These studies prove that isothermal calorimetry can be used to study enzymatic protein hydrolysis and detect protein hydrolysis heat flow.

The application of enzyme kinetics is described here according (Todd & Gomez, 2001) (Bianconi, 2007) (Mazzei, et al., 2016). A microcalorimeter is used to study enzyme catalysed reactions given that the amount of heat evolved (Q) is directly proportional to the molar enthalpy of the reaction as defined in eq. 2-15:

$$Q = n\Delta H_{app} = [P]_{Total}.V.\Delta H_{app} \qquad (2\text{-}15)$$

Where n is the total number of moles of product generated in mol, ΔH_{app} is the experimentally determined molar enthalpy for the hydrolysis reaction in J mol^{-1}, P is the total concentration of products generated in mol mL^{-1} and V is the solution volume in mL. Thus, the heat flow or thermal power is proportional to the rate of formation of products as described in eq. 2-16:

$$\frac{dQ}{dt} = \frac{d[P]}{dt}.V.H_{app} \tag{2-16}$$

Rearranging eq. 2-16 gives eq. 2-17:

$$Rate = \frac{d[P]}{dt} = \frac{1}{V.\Delta H_{app}}.\frac{dQ}{dt} \tag{2-17}$$

The heat flow or heat measured using a microcalorimeter is the total of all events taking place during the reaction. Enzyme protein hydrolysis contains proton exchange step where the released protons are taken by the buffer-conjugate base or neutralised by the addition of a base, as shown in eqns. 2-18 and 2-19, to maintain the pH constant, therefore, producing an additional heat inside the microcalorimeter reaction cell. This means that the heat flow measured is not the true heat released during the protein hydrolysis reaction, but includes heat associated with the protonation of the buffer as presented in eq. 2-20 (Mazzei, et al., 2016) (Freyer & Lewis, 2008).

$$S \rightarrow P + nH^{+}; \quad \Delta H_{S\rightarrow P} \tag{2-18}$$

$$A^{-} + H^{+} \rightarrow AH; \quad \Delta H_{1} \tag{2-19}$$

$$H_{app} = \Delta H_{S\rightarrow P} + n\Delta H_{1} \tag{2-20}$$

Where S is the protein, P is the products, $A^{-}$is the conjugate base, ΔH_1 is the heat of protonation, $\Delta H_{S\rightarrow P}$ is the heat of reaction and n is the number of protons released during protein hydrolysis. In short, isothermal microcalorimetry can be used to detect the enzyme catalysed reaction and its measurements are fast, direct, accurate and reliable.

2.6 Summary on modelling enzyme protein hydrolysis

Models provide an understanding of the factors affecting the system, which are used to determine restrictions and possible optimization strategies that can be implemented. Models can also be used in process design, simulations and performance predictions.

There are studies in the literature about modelling protein hydrolysis. The studies used the batch reactor model as well as the Michaelis-Menten equation to study enzymatic protein hydrolysis. However, this approach does not completely describe protein hydrolysis since the

system is modelled as a single substrate process, even though protein hydrolysis consists of multiple substrates during the reaction.

Marquez-Moreno and Fernandez-Cuadrado (1993) reported an empirical model for the protein hydrolysis. Again, Valencia et al (2016) proposed a combination of empirical and theoretical mathematical model using the batch reactor approach. Empirical models are derived from experimental data, therefore, their constants lack physical meaning. In addition, these models might not hold under different conditions which the process was setup. This calls for protein hydrolysis model using population balance approach.

The idea of modelling enzymatic hydrolysis using population balance approach was first reported by Hosseini and Shah (2011), and Lebaz et al. (2015) for enzymatic cellulose hydrolysis. No research was found about modelling enzymatic protein hydrolysis using population balance approach.

In the literature, many studies involve investigating the reaction rate kinetics and factors affecting the reaction rate. The studies prove that the rate at which the substrate depletes follows the Michaelis-Menten equation. The studies also reported that the temperature, pH and substrate concentration are the main factors that influence hydrolysis. This is because failure to control the temperature or pH during hydrolysis can destroy the functional properties of the substrate or enzyme.

Apar and Ozbek (2009), and Demirhan et al. (2011) studied protein hydrolysis by a commercial enzyme using the basic Michaelis-Menten model and the effect of product inhibition, temperature and pH on the rate. Valencia et al (2014) researched the effect of substrate and product on the hydrolysis curve. This was achieved by using the Michaelis-Menten approach along with the OPA method to analyse the degree of hydrolysis. Fernandez and Kelly (2016) studied the effects of controlling and not controlling the pH during the hydrolysis. These studies investigated and described the shape of the hydrolysis curve that results from plotting the degree of hydrolysis against the reaction time. Bianconi (2007) reported that calorimetry is an alternative and advantageous way to study enzyme catalysed reaction and to determine the kinetic parameters.

The use of microcalorimetry is a new technique in enzymatic protein hydrolysis. More specifically, isothermal titration calorimetry (ITC) is mainly used to study enzyme catalysed

reactions. Other than Maximova and Trylska (2015), no study was found using ITC for macromolecules such as casein and whey proteins. This study aims to use batch isothermal microcalorimetry (IMC) to investigate enzymatic protein hydrolysis. No study was found that use IMC for enzymatic protein hydrolysis.

In conclusion, there is a need for a new model for protein hydrolysis. This is because empirical simpler models do not offer a theoretical background. In addition, basic Michaelis-Menten kinetics does not describe enzymatic protein hydrolysis since the reaction involves products that can act as new substrates. An isothermal microcalorimeter can be used to study enzymatic protein hydrolysis. In addition, temperature, pH and substrate concentration play an important role during protein hydrolysis. The study proposed to use population balance model to account for the multiple products that can act as new substrates during enzymatic protein hydrolysis. The basic Michaelis-Menten kinetics was modified to describe the reaction rates due to multiple products (Palmer, 1995) (Barros & Malcata, 2004) (Qi & He, 2006) (Zapata-Montoya, et al., 2018).

3 Methodology

From the introduction (Chapter 1) and literature review (Chapter 2), the study aims to propose a theoretical mathematical model for enzymatic protein hydrolysis. The objectives of the study are as follows:

1. To experimentally determine the optimal hydrolysis conditions for two enzyme-substrate combinations;
2. To investigate the effect of an enzyme to substrate ratio and buffer type on the total heat released during enzymatic protein hydrolysis for two enzyme-substrate combinations;
3. To propose mathematical models for enzymatic protein hydrolysis using population balance approach;
4. To validate the proposed model for two enzymes using both the degree of hydrolysis and total heat released data;
5. To compare the degree of hydrolysis data obtained from spectrometry and total released data attained from calorimetry.

This chapter aims to outline the methods used to achieve the above-stated objectives. The objectives of the study were achieved in two ways, namely developing a model and generating data. In generating data, enzymatic protein hydrolysis was performed using papain and bromelain as enzymes and whey protein as the substrate. Firstly, the optimum enzymatic hydrolysis conditions for each enzyme- substrate concentration was determined from the degree of hydrolysis data using the o-phthaldialdehyde (OPA) method. Secondly, to determine the effects of enzyme concentration and buffer type, the hydrolysis time course data was generated at the optimum conditions of each enzyme-substrate combination using microcalorimetry technique and OPA method. Thirdly, three types of models were proposed by considering the two catalytic actions (endopeptidase and exopeptidase) and a combination of the two actions. Lastly, to validate the model, two models (endopeptidase and endopeptidase-exopeptidase) were taken further to compare with experimental data for bromelain and papain. The heat released and degree of hydrolysis data for the two different enzymes were used to obtain the model constants and compared to the model predictions for endopeptidase as well as endopeptidase- exopeptidase catalytic actions.

1.1 Experimental

3.1.1 Experimental design

In planning experiments, it is essential to consider the objectives of the experiments or research project. As stated above, the project consists of two experimental parts. The main objectives of the experiments in the first part are to obtain the optimum conditions that maximise the rate of the degree of hydrolysis. In the second part, it is desired to see the effect of enzyme concentration, buffer type and comparisons between the degree of hydrolysis and microcalorimeter data. For the planning of experiments, it is also important that there exists a relationship between the independent and the dependent variables.

The factors or independent variables were chosen for the first part as temperature (T), pH, substrate concentration (S), an enzyme to substrate ratio (E/S) and the enzyme type, with the degree of hydrolysis as the dependent or response variable. In the second part, enzyme concentration and buffer type were chosen as the independent variable and the heat released as the response variable for both enzymes. The experimental designs chosen for generating the degree of hydrolysis and heat released data are central composite inscribed and multilevel full factorial.

A central composite inscribed design was chosen because the enzymatic protein hydrolysis process takes place within the enzyme's working/active range. Central composite designs (CCD) are efficient and the response variable can be modelled as a function of the independent variables (Montgomery, 1997). In addition, CCD permits the interactions and effects of a factor to be estimated at numerous levels of the other independent variables, resulting in conclusions that are valid over a range of experimental conditions (Montgomery, 1997). The multilevel full factorial design was chosen because they offer the same advantages as the full factorial design. However, in factorial designs, all the factors have the same number of levels while in multilevel factorial design the levels can differ. Similar to CCD, multilevel factorial can provide a conclusion relating to the effects of variables.

The degree of hydrolysis data was generated for each enzyme type (Table 3-1) by using the central composite inscribed design with four factors or independent variables each at five levels. Table 3-2 and Table 3-3 show the coded levels and actual levels of each independent

variable for both enzymes. A central composite design with twenty-six experimental runs for each enzyme with three replications were performed.

Table 3-1: Groups for the experimental runs

Experimental runs	Enzyme
A	Papain
B	Bromelain

Table 3-2: Coded level of the variables for whey hydrolysis with papain according to central composite rotatable design.

Coded	Independent variable			
Level	Temperature(°C)	pH	S (%)	E:S (%)
2	45.00	4.8	2.00	1
1	50.00	5.15	4.00	1.5
0	55.00	5.5	6.00	2
-1	60.00	5.85	8.00	2.5
2	65.00	6.2	10.00	3

Table 3-3: Coded level of the variables for whey hydrolysis with bromelain according to central composite rotatable design.

Coded	Independent variable			
Level	Temperature(°C)	pH	S (%)	E:S (%)
2	40	4	2.00	1
1	42.5	5	4.00	1.5
0	45	6	6.00	2
-1	47.5	7	8.00	2.5
-2	50	8	10.00	3

In the second step, the heat released data were generated for each enzyme type using a multilevel full factorial design for the last two steps. The design consists of two factors namely enzyme to substrate ratio and buffer type. The enzyme to substrate ratio variable contains three levels as indicated in Table 3-4. The buffer type variable consists of two levels as shown in Table 3-5.

Table 3-4: Enzyme to substrate percentage (E:S) levels for whey protein hydrolysis with bromelain and papain.

Level	E:S (%)
-1	2
0	3
1	4

Table 3-5: Buffer type levels for whey protein hydrolysis

Level	Bromelain	Papain
1	Citric Acid – Disodium hydrogen phosphate	Citric Acid – Disodium hydrogen phosphate
2	Disodium hydrogen phosphate- Potassium dihydrogen phosphate	Sodium acetate -acetic acid

3.1.2 Materials

Whey protein concentrate (WPC) powder was used as a hydrolysis reaction substrate. WPC was supplied by Leprino foods and contained 80 % protein. Papain and Bromelain were used for the hydrolysis of WPC. Papain was supplied by Enzybel International SA. Bromelain was supplied by Enzyme Technologies. Potassium hydroxide was added as the reaction progressed to maintain the pH constant during hydrolysis. Disodiumtetraborate decahydrate and 200 mg Na-dodecyl-sulphate (SDS), o-phthaldialdehyde and dithiothreitol were used to prepare the OPA reagent. L- serine was used to make the standard solution. All the chemicals were supplied by Sigma-Aldrich and Merck.

3.1.3 Buffer solutions

3.1.3.1 Citric acid- disodium hydrogen phosphate

To prepare 1000 ml citric acid- disodium hydrogen phosphate buffer at pH 5.15 and pH 6 (papain and bromelain optimum), 469.23 ml and 368.50 ml (0.1 M) of citric acid and 530.75 ml and 631.50 ml (0.2 M) of disodium hydrogen phosphate respectively were mixed together using an overhead stirrer. The solution was adjusted to the desired pH by adding either disodium hydrogen phosphate or citric acid.

3.1.3.2 Sodium acetate-acetic acid

To make 1000 ml sodium acetate-acetic acid buffer at pH 5.15, 767.5 ml (0.2 M) of sodium acetate trihydrate and 235.5 ml (0.2 M) of acetic acid were mixed together and the solution was adjusted to the desired pH.

3.1.3.3 Potassium dihydrogen phosphate – disodium hydrogen phosphate

To make 1000 ml potassium dihydrogen phosphate – disodium hydrogen phosphate buffer at pH 6, 438.50 ml (0.2 M) of potassium dihydrogen phosphate, 61.5 ml (0.2 M) of disodium hydrogen phosphate and 500 ml deionized were mixed together and the solution was adjusted to the desired pH.

3.1.4 Equipment

The heat released during enzymatic protein hydrolysis was measured using a Thermal Activity Monitor III (TAMIII) isothermal microcalorimeter (TA instruments, Thermometric AB, Sweden) at the Centre for Bioprocess Engineering, Department of Chemical Engineering, University of Cape Town. The TAMIII consists of isothermal titration calorimeter and multicalorimeter modules. The two modules operate independently but use the same heatsink or thermostat. The thermostat uses oil as the thermal media and operates between 15-150 ºC. It also has a long-term stability of < ± 100 µK/24 h, short-term stability of < ± 10 µK as well as an accuracy of < ± 0.1 ºC.

Multicalorimetry was employed for this study. Multicalorimeter holds six independent microcalorimeters with a heat output with a short-term noise of < ± 100 nW, a baseline drift of < 200 nW/24h, an accuracy of < 5% and a precision of ± 200 nW.

3.1.5 Methods

3.1.5.1 Determining optimum enzymatic hydrolysis conditions

Protein hydrolysis using papain

To obtain the optimum conditions for papain, hydrolysis was carried out in a 600 ml reaction vessel with an overhead stirrer, temperature control and pH control as shown in Figure 3-1. A measured amount of whey protein was added to a reaction vessel containing 200 ml distilled

water to make a substrate concentration ranging from 2 % to 10 % (w/v) as proposed by Butré et al. (2014), Adler-Nissen (1986) and Valencia (2015). Before hydrolysis, the solution was preheated to the reaction temperature as in experimental design Table 3-2 and adjusted the pH using either KOH or o-phthaldialdehyde HCl. The reaction was initiated by adding an enzyme of 1 % to 3 % (w/w) of the substrate concentration as determined in the experimental design. During the process, the temperature was controlled using a water bath and pH was maintained by addition of KOH using an automatic pH controller as shown in Figure 3-1. Before deactivating the enzyme, 200 µl samples were taken at regular time intervals (10 min) and added to 1 ml 3 % (w/v) Na-dodecyl-sulphate (SDS) solution. The reaction was inactivated as soon as the samples were taken by heating the samples to 95 ºC using a heating stove with boiling water for 5 min. Finally, the samples were cooled to room temperature and stored at 20 ºC until determining the degree of hydrolysis. The experiments were carried out for 120 min at temperature and pH range as set out in the experimental plan.

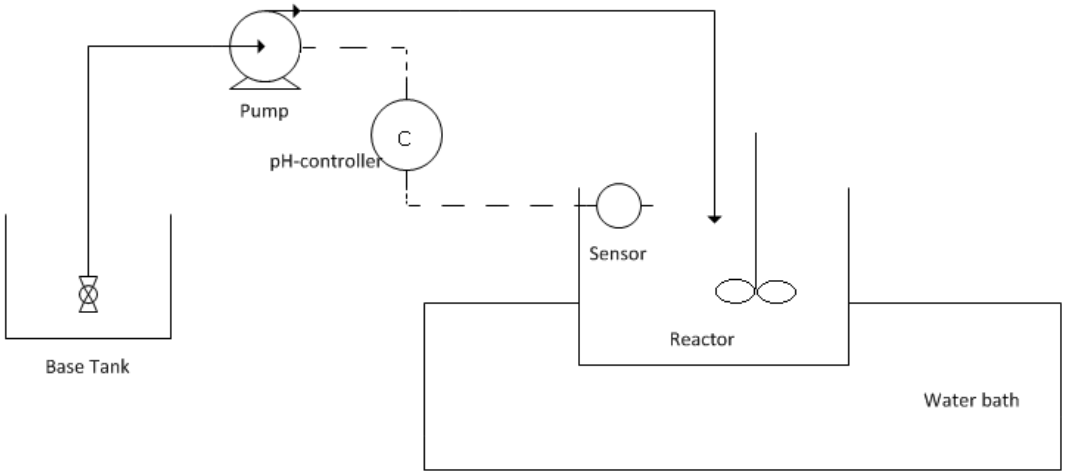


Figure 3-1: Experimental setup for enzymatic protein hydrolysis

Protein hydrolysis using bromelain

To determine optimum conditions for bromelain, a similar procedure as above was followed. The reaction temperature and pH were changed as set out in experimental design Table 3-3.

3.1.5.2 Measuring degree of hydrolysis using OPA method

The degree of hydrolysis of the samples obtained in the above section was determined using the OPA (o-phthaldialdehyde method as previously presented by Nielsen (2001). The method

was also used by Valencia (2014). The OPA reagent was prepared by completely dissolving 7.620 g di-Natetraborate decahydrate and 200 mg Na-dodecyl-sulphate (SDS) in 150 ml of deionized water. Then, 160 mg o-phthaldialdehyde powder was dissolved in 4 ml ethanol and mixed with the above solution. Finally, 176 mg dithiothreitol (DTT) was added then rinse with deionized water to make 200 ml solution. The serine standard was prepared by dissolving 50 mg of serine powder in 500 ml of deionized water (0.9516 meqv/ l).

The following procedure was followed to measure the absorbance of the samples, standard and blank. To prepare the blanks, 400 µl of distilled water was added to 3 ml OPA reagent. The samples and serine standard were prepared by the adding 400 µl of hydrolysates sample to 3ml of OPA reagent. The absorbance values of the solutions were measured at 340 nm using a spectrophotometer. The readings for the reaction of the amino acids with OPA reagent was taken after 2 min to equilibrate and after 10 min and average absorbance values were used to calculate the degree of hydrolysis. Finally, the degree of hydrolysis was calculated using eq.3- 1.

$$DH = \frac{h}{h_{tot}} \times 100\% \tag{3-1}$$

Where h_{tot} represent the total number of peptide bonds per protein molecule and determined as 8.8 meqv/g protein according to Adler-Nissen (1986). h is defined as the number of hydrolysed bond at a certain time and calculated using eqns. 3-2 and 3-3.

$$h = \frac{(Serine - NH_2) - \beta}{\alpha} \tag{3-2}$$

$$Serine - NH_2 = \left(\frac{A_{sample} - A_{blank}}{A_{standard} - A_{blank}}\right)\frac{C_o V_s}{m} \tag{3-3}$$

Where Serine-NH_2 = meqv serine NH_2/g protein, m is the mass of the protein in gram, $\beta = 1$ and $\alpha = 0.4$ are constants for whey protein (Adler-Nissen, 1986), V_s is the sample volume in L (litre), C_o is the standard concentration in meqv/L and A_{sample}, A_{blank} and $A_{standard}$ are the sample absorbance, blank absorbance and standard absorbance respectively.

3.1.5.3 Model validation

Enzyme hydrolysis measured using degree of hydrolysis

For model validation, enzymatic hydrolysis was performed at the optimum conditions as set out in the experimental design. The experiments were performed as above except, a buffer solution was used instead of KOH to control the pH. Also, whey, bromelain and papain powders were dissolved in the buffer solution instead of distilled water. The experiments were carried out for 6 h. Samples for determining the degree of hydrolysis were drawn every 10 min for 1 h, then every 20 min during the 2nd h and after 30 min for the last 3 h. This is because hydrolysis is fast in the beginning and slower after an hour. This was deduced from the results obtained in section 3.1.5.1 and 3.1.5.3.

Enzyme hydrolysis measured using isothermal microcalorimetry

In order to determine the heat flow during the reaction, the heat released or thermal power as time progressed was measured. This was carried out in a single injection experiment using an isothermal batch microcalorimeter.

To measure heat released or thermal power during protein hydrolysis, 2 ml whey protein solution containing 10% or 6% protein as set out in the experimental design was pipetted into 4 ml ampoules. The whey protein solution consisted of whey powder and buffer solution. After 30 min of baseline confirmation, 100 µl enzyme solution was added into the solution to initiate the reaction. Then the ampoule was tightly closed using a hand clamp and slowly placed in the microcalorimeter equilibration position using an ampoule lifter. After 15 min, the samples were lowered into the measuring position to obtain the thermal power or heat released during the reaction. The thermal power was monitored for 5 h and 9 h at 10 s time intervals for hydrolysis using bromelain and papain respectively. For both enzymes, a control or blank experiment of the substrate to buffer was carried out. To stop the reaction, the ampoules were removed and once the heat flow stabilised to the default (moderate) signal stability, the final baseline was recorded for 30 min. TAM Assistant software provided by the manufacturers was used to automatically subtract the average baseline (baseline data before ampoules were inserted and after the ampoules were removed) from the heat flow signal recorded during the reaction.

3.1.6 Statistical analysis

The experimental data obtained from measuring the degree of hydrolysis using OPA method was used to see the factor effects and interaction and find optimum reaction conditions. The heat released, or thermal power data was used to validate the mathematical model. Statistica software was used for statistical testing. The optimum conditions were obtained from the desirability plots. The effect of the factors on the degree of hydrolysis and heat released during the reaction was determined using analysis of variance (ANOVA). The model constants were determined by the least-squares method using non-linear regression in Python 3 (Notebook jupyter) using an optimize.least_squares function. The coefficient of determination (R^2) for experimental data and model predictions was calculated using r2_score function in Python 3 to describe the goodness of fit.

3.2 Mathematical model

A few models and expressions have been developed mathematically to describe enzymatic protein hydrolysis. However, finding a model that fits the experimental data is a challenge. This might be due to the over-simplified assumptions which make it difficult to predict the process. In process modelling, it is important to make assumptions that would simplify the model but keeps the fundamentals of the process. This section aims to describe and explain the choices as well as the assumptions made to develop the model for enzymatic protein hydrolysis.

3.2.1 Process description

Enzymatic hydrolysis is a biological reaction and it can be used in batch as well as continuous processes. In this study, we are modelling a batch enzymatic hydrolysis process, the system is shown in Figure 3-1. The reactor contains water with volume V, a protein with concentration S and enzyme with concentration E. The reactor operates at temperature T and a certain pH. The vessel contents are stirred and heated using a water bath that provides a heating rate Q. The change in solution pH resulting from the produced amino-acids was controlled using an automatic pH controller or a buffer.

Protein hydrolysis involves multiple cleavage sites or peptide bonds that are considered as substrates and each site can be present in different peptides during hydrolysis (Butré, et al., 2014). This implies that products (polypeptides) can act as new substrates resulting in

progressive hydrolysis (Valencia, et al., 2015) (Martínez-Araiza, et al., 2012). Because there are multiple different cleavage sites in the protein or peptides, enzymes are specific or prefer a certain type of amino acids after which it can break the peptide bond. The enzyme breaks either the internal or the external bonds in a protein. Enzymatic hydrolysis of proteins is considered as a pure breakage or chain-end process depending on the type of enzyme used.

As stated in section 2.4.2, the study is interested in modelling populations of polymer chains. Figure 3-2 and Figure 3-3 present how enzymatic protein hydrolysis proceeds as well as the nomenclature used in the proposed model. To model enzymatic protein hydrolysis, the study described a protein before adding an enzyme as a long chain with chainlength M and J number of bonds breakable by the enzyme. M and J were used to define the group of polymer chains into classes described as chainlength. During protein hydrolysis, the original chain can break to form new chains which can also be broken to form other chains.

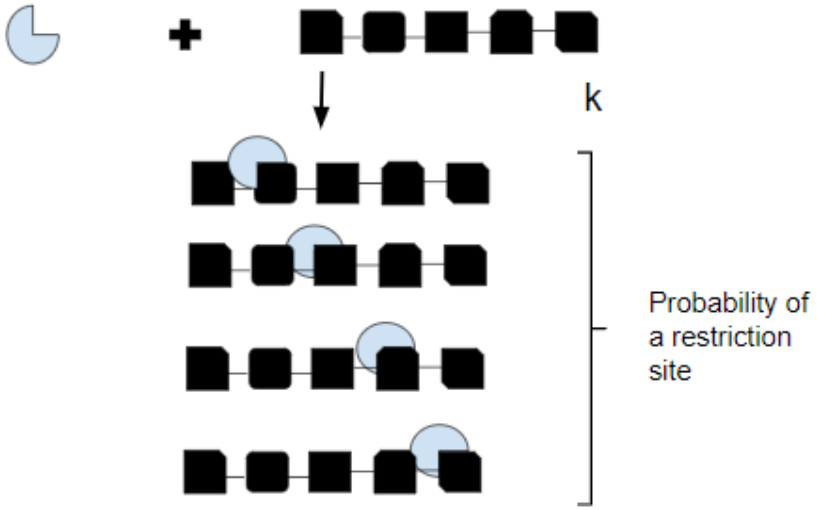


Figure 3-2: Probability of a restriction site equal throughout a chain with length k.

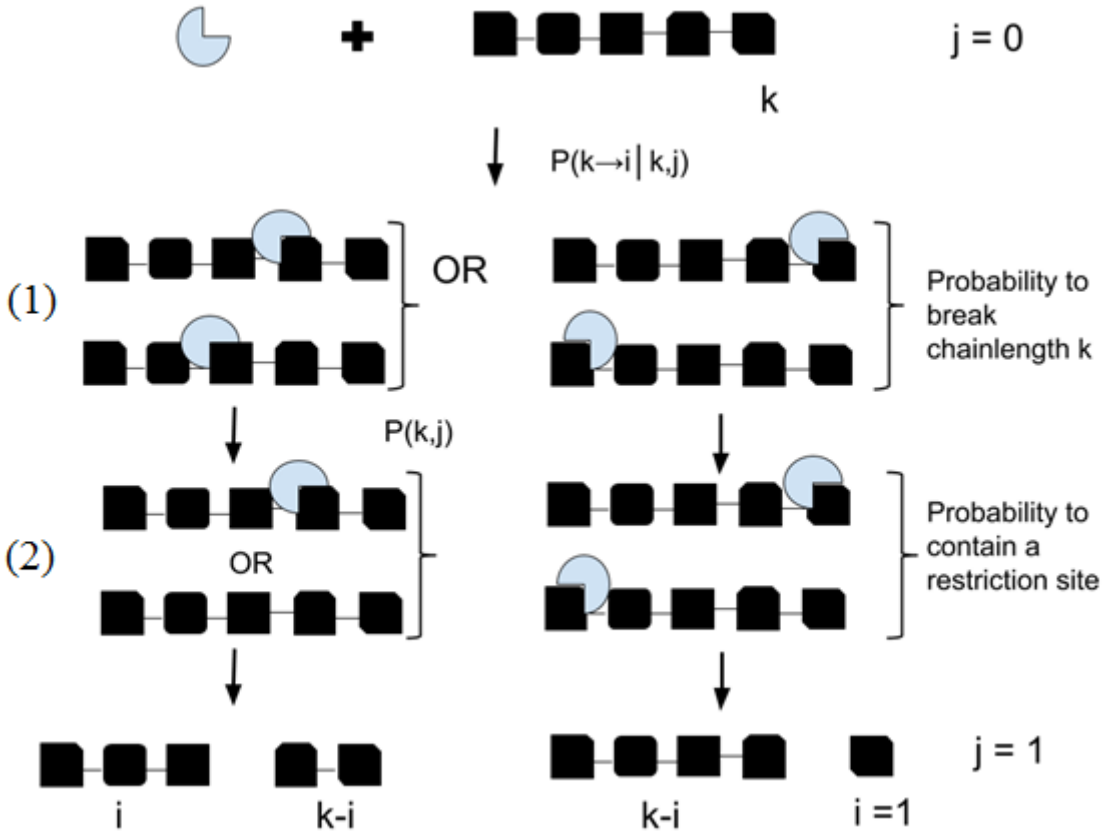


Figure 3-3: Enzymatic protein hydrolysis mechanism.

The study represented the chains being broken with length k as shown in Figure 3-2 and the polypeptides or chains being produced with length i as shown in Figure 3-3. Initially, there are $k-1$ peptide bonds in a chain with length k in which an enzyme can break to form a chain with length i. This implies that there are $k-1$ different places in a chain with length k that can be broken by an enzyme. Therefore, there exists the probability to which bond will be broken by the enzyme. If we assume the enzyme has no preference as to which bond to break, then the probability of each bond broken by the enzyme is equal as shown in Figure 3-2. In reality, an enzyme binds or prefers a peptide bond before or after a certain amino acid. This was introduced in the model as restriction sites.

During hydrolysis, an enzyme binds to the restriction site in the chain with chainlength k in two possible ways to form an enzyme-chain complex as shown in Figure 3-3(1). An enzyme can bind to outer or internal bonds in the chain with length k which was represented in the model by $P(k \rightarrow i|k,j)$. However, the enzyme can bind to the peptide bond in the chain with length k during protein hydrolysis if a restriction site preferred by the enzyme is present in the

chain and this stage is denoted by $P(k,j)$. $P(k \rightarrow i|k,j)$ defines the probability that a chain of length k breaks to form a chain of length i on the condition that the chain of length k contains a restriction site. $P(k,j)$ defines the probability that chain k might contain a restriction site. The presence of a restriction site in a chain of length k depends on the catalytic action of an enzyme.

If an enzyme breaks the external bonds, this implies that the enzyme will always bind to the bonds since a chain always contains two ends and thus releasing a monomer. On the other hand, if an enzyme breaks internal bonds, this means that the enzyme can bind or not bind depending on whether the chain being broken contains a restriction site as shown in Figure 3-3(2). The restriction sites were defined by tracking the number of broken bonds during protein hydrolysis by using generation, j. In other words, if one bond is broken $j = 1$ as shown in Figure 3-3, the number of restriction sites available are $J - j$. Further, in generation $j = 1$, multiple chains can form including a chain with length k which can form other chains if there are restriction sites. This means that protein hydrolysis will stop when there are no longer restriction sites.

As the first step to develop the mathematical model, enzymatic protein hydrolysis mechanism is described, followed by the derivation of the reaction rate expression. Furthermore, we present a model for endopeptidase, exopeptidase and combination of both using population balance approach. Moreover, the energy balance of enzymatic protein hydrolysis process was derived. Finally, we present the model inputs and how the model was implemented.

3.2.2 Enzymatic protein hydrolysis mechanism

The reaction mechanism for the process is described by Michaelis–Menten(M-M) equation. Several studies reported a single substrate, single enzyme M-M mechanism for protein hydrolysis as shown in eq. 3-4.

$$S + E \overset{k_m}{\leftrightarrow} ES \overset{k_{cat}}{\longrightarrow} E + P \tag{3-4}$$

However, enzymatic protein hydrolysis involves multiple polypeptide chains broken by a single enzyme to form smaller peptide chains. In this section, a modified Michaelis-Menten equation to describe the effect of the multiple substrates on the reaction rates.

To derive multiple substrates and single enzyme M-M reaction rate equation, enzyme, polypeptide chains, products and enzyme-substrate complex will hereafter be denoted as E, S, P and ES respectively. Then S_{ij}, ES_{ij} and P_{ij} denote a peptide with chain length i in generation j, enzyme-peptide complex and products or polypeptides or amino acids with chain length i in generation j respectively. Generations are defined as the number of broken bonds. The following chemical equation describes the reaction mechanism with multiple chains:

$$S_{ij} + E \overset{k_{m,ij}}{\longleftrightarrow} ES_{ij} \overset{k_{cat,ij}}{\longrightarrow} E + P_{ij} \quad i = 1,2,3,\dots M \;\; j = 1,2,..J \tag{3-5}$$

Where $k_{m,ij}$ represents the M-M parameter for the enzyme binding to substrate S_{ij} and $k_{cat,ij}$ denotes the kinetic rate constant. To derive the reaction rate equation, it was first assumed that ES_{ij} concentration is at equilibrium, which implies that ES_{ij} complex is constant. Then the concentrations of the reactants over the concentration of product is written as in eq. 3-6.

$$k_{mij} = \frac{[E][S_{ij}]}{[ES_{ij}]} \quad i = 1,2,\dots M \quad j = 1,\dots J \tag{3-6}$$

Where E is defined by the initial enzyme concentration (E_o) and ES_{ij} as presented in eq. 3-7.

$$\mathrm{E_o} = \mathrm{E} + \sum_{i=1}^{M}\sum_{j=1}^{J} ES_{ij} \tag{3-7}$$

By rearranging eq. 3-6, E and ES complex concentrations in eq. 3-7 are calculated by using eq. 3-6 as shown below.

$$[ES_{ij}] = \frac{[E][S_{ij}]}{k_{mij}} \tag{3-8}$$

By rearranging eq. 3-8 and substituting into eq. 3-7, E_O can be calculated using the following equation.

$$E_o = \frac{k_{mij}[ES_{ij}]}{[S_{ij}]} + \sum_{i=1}^{M}\sum_{j=1}^{J} ES_{ij} \tag{3-9}$$

Then rearranging eq. 3-9 into eq. 3-10. the rate of product formation is defined as shown in eq. 3-11.

$$[ES_{ij}] = \frac{E_o[S_{ij}]}{k_{mij} + \sum_{i=1}^{M}\sum_{j=1}^{J}[S_{ij}]} \tag{3-10}$$

$$R_{ij} = \frac{d[P_{ij}]}{dt} = k_{cati}[ES_{ij}] \tag{3-11}$$

Substitution of eq. 3-10 into eq. 3-11 gives the following general rate expression corresponding to a multiple-substrates, single-enzyme M-M model may be written as:

$$R_{ij} = \frac{k_{cat}E_oS_{ij}}{k_m + \sum_{i=1}^{M}\sum_{j=1}^{J}S_{ij}} \tag{3-12}$$

Where S_{ij} represents the molar concentration of the peptides and an amino acid with chain length i and 1 of generation j. Assuming M-M constants, k_{mij} , are equal then $k_m = k_{mij}$.

3.2.3 Population balance model

The kinetic model for enzymatic hydrolysis was developed using a population balance approach, by adapting an existing model developed for cellulose hydrolysis. Cellulose is a polymer which is made up of only one type of monomer (glucose), while proteins consist of different types of amino acids. Enzymatic hydrolysis of a protein with chain length k and restriction sites J leads to formation of peptide chains with length i ($i < k$) in generation j because of an endopeptidase and exopeptidase. To model enzymatic protein hydrolysis, each enzyme action was considered separately and then as a combination of the two actions.

Population balance describes any system with population distribution. In this study, we consider populations of polymers or peptides chains. This means that we can take a protein as a population of polymer chains. Then, for a continuous system undergoing binary break ups and constant growth of polymer chains, chains distribution is described by the kinetic eq. 2- 10 (Lebaz, et al., 2015).

After discretization and substituting the growth terms as zero in eq. 2-10, Kumar and Ramkrishna (1996) found out that breakage results in the death of one large chain and the birth

of two smaller ones as shown in eq. 2-11, rewritten below. The left-hand side of eq. 2-11 is the rate of change in the total number of polymer chains with chainlength i ($1 \leq i \leq M$ M is the total initial number of peptide bonds in the protein). The polymer chains or monomers are considered as classes or groups of the chain length as suggested by Kumar and Ramkrishna (1996). On the right hand, the first term describes the number of chains of length i hydrolysed at a given time, which depend on the breakage frequency, T_i. The second term defines formation of new chains in the i^{th}interval.

$$\frac{dN_i}{dt} = -T_i N_i(t) + \sum_{k=i+1}^{M} n_{ik}\, T_k N_k(t) \qquad (3\text{-}13)$$

Where N_i is the total number of polypeptide or amino acid with chain length i, t denotes the reaction time, T_i is the breakage frequency of the polypeptide with chain length i, T_k defines breakage frequency of the polypeptide with chain length k which break to form new chains length i, and the term n_{ik} refers to the probability of breaking a chain with length k to form a new chain of length i.

In the context of polymer systems and reaction kinetics, molar concentration is used instead of the number of molecules. The number of molecules is related to the molar concentration through the Avogadro number. Then, eq. 3-13 can be rewritten as follows:

$$\frac{dC_i}{dt} = -R_i + \sum_{k=i+1}^{M} n_{ik} R_k \qquad i = 1, 2, \ldots M \qquad (3\text{-}13)$$

Where C is the concentration and R is the reaction rate.

3.2.3.1 Modelling endopeptidase

As stated above, a model for endopeptidase, exopeptidase and endopeptidase-exopeptidase will be proposed. In this section, a model for protein hydrolysis due to an endopeptidase is presented. The hydrolysis of proteins due to endopeptidase is considered a pure breakage process. This means that there are no growth of new chains and the total number of monomers remains the same.

An endopeptidase binds to the internal bonds of a chain with length k as shown in Figure 3-4. There are $k-1$ different places in which a chain of length k can break if we assume no preference is given as to where the chain break. Therefore, the study modelled the system such that the probability of a bond being a restriction site is equal throughout a chain. However, there are only two possible bonds along a chain with length k which an enzyme can break (as shown in Figure 3-4) to form chain i on the condition that chain k contain at least one restriction site, $P(k,j)$. Thus, the term $P(k \rightarrow i|k,j)$ is defined as in eq. 3-14.

$$\mathrm{P(k \rightarrow i|k,j)} = \frac{2}{\mathrm{k-1}} \qquad \mathrm{i} < \mathrm{k} \quad j = 1,2 \ldots J \tag{3-14}$$

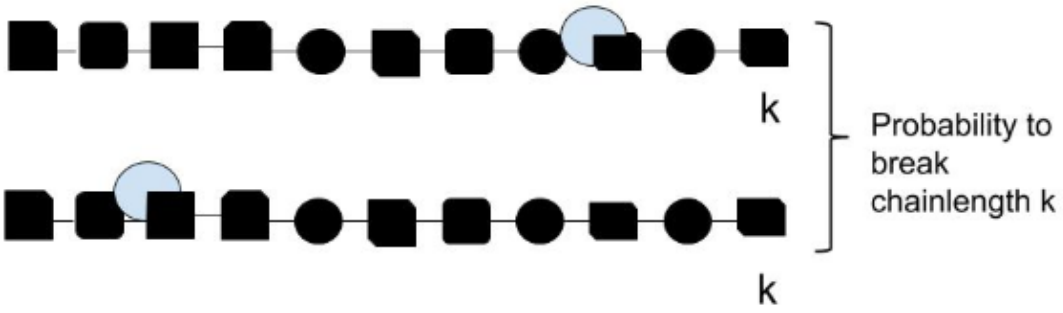


Figure 3-4: Probability to break a chain of length k to form two new chains of length i and $k-i$ during protein hydrolysis due to endopeptidases activity.

During protein hydrolysis, not all the bonds are broken by the enzyme. Hence, the probability of a chain with length k to contain a restriction site was introduced in the model. As stated in section 3.2.1, the study described a protein before adding an enzyme as a long chain with chainlength M and J number of bonds breakable by the enzyme. Further, in generation j, multiple chains can form including a chain with length k. The formed chain with length k can be broken if the chain contains a restriction site. Since there are multiple chains with multiple bonds there exists a probability to which bond will be broken by an enzyme.

To define the probability that a single chain with length k will contain a restriction site out of multiple chains. First, the probability that chain k does not contain a breakable bond was described. To derive the expression, the probability of the first, second and $k-1$ bond in chain k not being broken was defined as the total non-breakable (M-J) divided by total number

of bonds available in generation, j $(M-j)$ by using the statistical method of sampling without replacement as shown in eqns. 3-15 - 3-17.

$$P_1 = \frac{M-J-1}{M-j-1} \tag{3-15}$$

$$P_2 = \frac{M-J-2}{M-j-2} \tag{3-16}$$

$$P_{k-1} = \frac{M-J-k-1}{M-j-k-1} \tag{3-17}$$

Then, the total probability of chain with length k to not contain any breakable bond is defined as shown in eq. 3-18.

$$P_T = P_1 * P_2 \ldots * P_{k-1} \tag{3-18}$$

The probability of a formed or broken chain to contain a restriction site, $P(k,j)$ can now be represented as shown in eqns. 3-19 and 3-20.

$$P(k,j) = 1 - P_T \tag{3-19}$$

$$P(k,j) = 1 - \prod_{l=1}^{k-1} \frac{M-J-l}{M-j-l} \tag{3-20}$$

Where $P(k,j)$ refers to the probability that a chain with length k and generation j contains a restriction site or a breakable bond.

In the description of an initial long chain with length M and J restriction sites that can be cleaved and formation of new products or polypeptides with different chain lengths, the following equation is proposed. A new variable was added in eq. 3-13 to predict a real system which includes specificity or restriction sites. The left-hand side in eq. 3-21 defines the concentrations of a polypeptide i in generation j as time progress. The first term on the right-hand side is the rate at which a polypeptide with chain length i in the current generation j breaks, and the second term describes the rate at which a polypeptide i is formed from all possible polypeptides with chain length k in the previous generation $j-1$.

$$\frac{dC_{ij}}{dt} = -R_{ij}P(i,j) + \sum_{k=i+1}^{M} P(k \rightarrow i|k,j-1)\, P(k,j-1)\, R_{k,j-1} \qquad (3\text{-}21)$$

Where C_{ij} is the molar concentration of the polypeptide with chain length i of generation j which means number of broken bonds, R_{ij} defines breakage reaction rate of the polypeptide with chain length i from generation j, $R_{k,j-1}$ represents breakage reaction rate of the polypeptide with chain length k from generation $j-1$, $P(k \rightarrow i|k,j-1)$ is defined as the probability to break chain length k to form a new chain i if it is hydrolysed, $P(i,j)$, is the probability that a polypeptide of chain length i and generation j contains a restriction site, $P(k,j-1)$, is the probability that a polypeptide of chain length k and generation $j-1$ contains a restriction site and J refers to the total number of breakable bonds. The reaction rates R_{ij} and $R_{k,j-1}$ are described by the multiple-substrates M-M as in eq. 3-12.

3.2.3.2 Modelling exopeptidase

The exopeptidase catalytic action is also known as a chain-end process. The enzyme acts on the two ends of the protein chains (Figure 3-5) and releases an amino acid. This means that during hydrolysis only monomers with chain length $i = 1$ and polypeptides of chainlength $k-1$ are formed when breaking polypeptides with chainlength k. This implies that the probability of an enzyme to bind on the chain of length k is independent on the chain length. This is because there are two preferred links where each chain breaks as shown in Figure 3-5 and two ways each chain can break to form a monomer. The resulting probability expression corresponding to exopeptidase is shown in eq. 3-22.

$$P(k \rightarrow i|k,j) = 1 \quad i = 1 \; or \; i = k-1 \qquad (3\text{-}22)$$

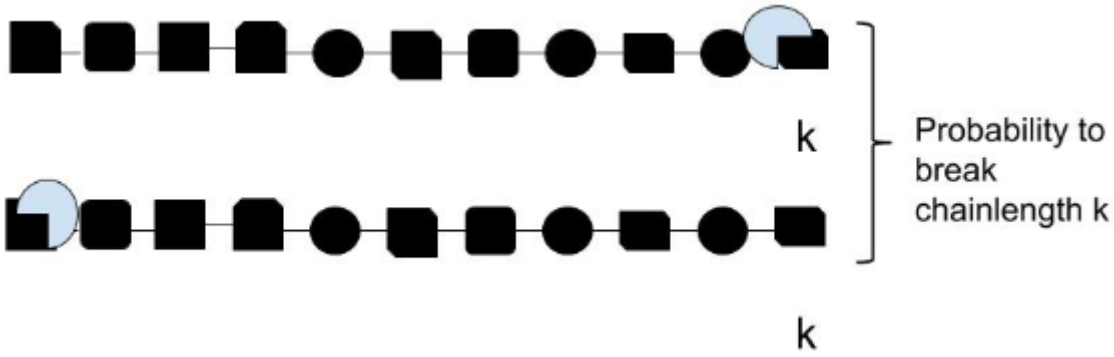


Figure 3-5: Probability to break a chain of length k to form two new chains of length 1 and $k-1$ during protein hydrolysis due to exopeptidases activity.

$P(k \rightarrow i|k,j)$ describes the conditional probability that a chain with length k will break to form chain with length i on the condition that a restriction or cleavage site is present $P(i,j)$. For an exopeptidase, $P(k,j)$ is equal to one since a given chain will always contain two ends (restriction sites) as shown in eq. 3-23.

$$P(k,j) = 1 \tag{3-23}$$

The following eq. 3-24 is proposed for an exopeptidase. The equation is like eq. 3-21 for an endopeptidase. However, the second term on the right-hand side of eq. 3-24 is described as the rate at which a polypeptide with chain length i is formed from all possible polypeptides with chain length k in the current generation j. This is because during hydrolysis only a single chain continuously breaks to form a monomer or chain with length $i = 1$. Hence, $R_{k,j}$ is used in eq. 3-24 not $R_{k,j-1}$ as in eq. 3-21.

$$\frac{dC_{ij}}{dt} = -R_{ij}P(i,j) + \sum_{k=i+1}^{M} P(k \rightarrow i|k,j)\, P(k,j)\, R_{k,j} \tag{3-24}$$

Where $R_{k,j}$ represents breakage reaction rate of the polypeptide with chain length k to form a chain with length i in generation j, $P(k \rightarrow i|k,j)$ is defined as the probability to break chain length k to form a new chain i in generation j, $P(k,j)$ is the probability that a polypeptide of chain length k and generation j contains a restriction site.

3.2.3.3 Modelling endopeptidase and exopeptidase

The equation for combined endopeptidase-exopeptidase is shown in eq. 3-25:

$$\frac{dC_{ij}}{dt} = \left(\frac{dC_{ij}}{dt}\right)_{endo} + \left(\frac{dC_{ij}}{dt}\right)_{exo} \tag{3-25}$$

In the equation, each catalytic activity has separate kinetic constants.

3.2.4 Degree of hydrolysis

The degree of hydrolysis (DH) is the percentage of the total number of peptide bonds in a protein, which has been cleaved during hydrolysis (Adler-Nissen, 1986). The degree of hydrolysis is calculated as in eq. 3-26 according to the proposed model.

$$DH = 1 - \frac{\sum_{i=1}^{M-1}\sum_{j=1}^{J}(i-1)C_{ij}(t)}{M\ C_{MJ}(0)} \tag{3-26}$$

3.2.5 Energy balance

The general law of energy conservation or First Law of Thermodynamics can be expressed as

$$\begin{Bmatrix}\text{rate of energy}\\ \text{accumulation}\end{Bmatrix} = \begin{Bmatrix}\text{rate of energy}\\ \text{entering system}\\ \text{by inflow}\end{Bmatrix} - \begin{Bmatrix}\text{rate of energy}\\ \text{leaving system}\\ \text{by outflow}\end{Bmatrix} + \begin{Bmatrix}\text{rate of heat}\\ \text{added to system}\end{Bmatrix} + \begin{Bmatrix}\text{rate of work}\\ \text{done on system}\end{Bmatrix} \tag{3-27}$$

The total energy of a thermodynamic system is the sum of its internal energy, potential energy and kinetic energy. The following assumptions are made enzymatic hydrolysis process:

1. Change in kinetic energy and potential energy is negligible compared to changes in internal energy.
2. The rate of work is negligible compared to rates of heat added to the system.
3. The rate of energy entering the system by inflow and rate of energy leaving the system by outflow are equal to zero because enzymatic hydrolysis is a batch process.

For these assumptions, the energy balance in eq. 3-27 can be written as

$$\frac{dU_{int}}{dt} = h_{rxn} + Q \tag{3-28}$$

Where U_{int} is the internal energy of the system, h_{rxn} is the rate of energy due to a chemical reaction and Q is the rate of heat transfer to the system. Since the system is isothermal, the rate of change of internal energy is equal to zero as shown below

$$dU_{int} = 0 \tag{3-29}$$

An expression for the rate of energy due to a chemical reaction can be written as

$$h_{rxn} = (-H_{RXN}) \sum_{i=1}^{M} \sum_{J=1}^{J}, P(i,j)\, R_{ij} v \tag{3-30}$$

Where H_{RXN} is the total heat of reaction and R_{ij} is defined previously as the reaction rate to form a chain length i. The resulting energy balance expression predicting the heat flow from the system or reaction can be written as in eq. 3-31.

$$Q = (-H_{RXN}) \sum_{i=1}^{M} \sum_{J=1}^{J} P(i,j)\, R_{ij} v \tag{3-31}$$

To account for the neutralisation energy, $-H_{RXN}$ was rewritten as $-H_{APP}$ as shown in eq. 3-32.

$$Q = (-H_{APP}) \sum_{i=1}^{M} \sum_{J=1}^{J} P(i,j)\, R_{ij} v \tag{3-32}$$

Where H_{APP} is the total heat of reaction.

3.2.6 Model inputs

The experimental conditions, constants and model parameters are required in the proposed model. The experimental conditions are specified in the experimental design and the model parameters are obtained using the experimental data. Table 3-6 shows the model constants for implementation of the proposed model. As indicated in Table 3-6 , the total number of breakable (J) was determined using experimental data.

Table 3-6: Model inputs

Description of the constant	Symbol	Value	Reference
Total Chain length	M	201	(Spellman, et al., 2003)
Total number of breakable bonds	J	Estimated from experimental data	Appendix A.2.1
Molecular weight of whey	Mw	23644 Da	(Spellman, et al., 2003)

3.2.7 Model implementation

The model was implemented in Python as shown in Appendix B. The model was solved simultaneously using the ODEINT function. ODEINT function integrates a system of ordinary differential equations. The model constants were determined using the optimize. least_squares function. Optimize. least_squares function minimises the error between the estimated model and the experimental data points. The regressed constants were used to determine the determination coefficient, R^2 was determined using r2_score function to describe the goodness of fit.

4 Results and discussion

In this section, a comparison between the degree of hydrolysis data obtained using spectrophotometry and heat flow data found using an isothermal microcalorimeter is made. Also, the effects of temperature, pH, enzyme-to-substrate ratio and substrate concentration on the degree of hydrolysis as well as the optimum conditions of the enzymes are presented. In addition, the effect of the buffer type and enzyme concentration at the optimum conditions of the enzymes is discussed. Lastly, the proposed model for enzymatic protein hydrolysis was fitted to experimental data obtained at the optimum conditions. The results were obtained by performing protein hydrolysis using papain and bromelain as explained in section 3.1. The results are given separately for each type of enzyme.

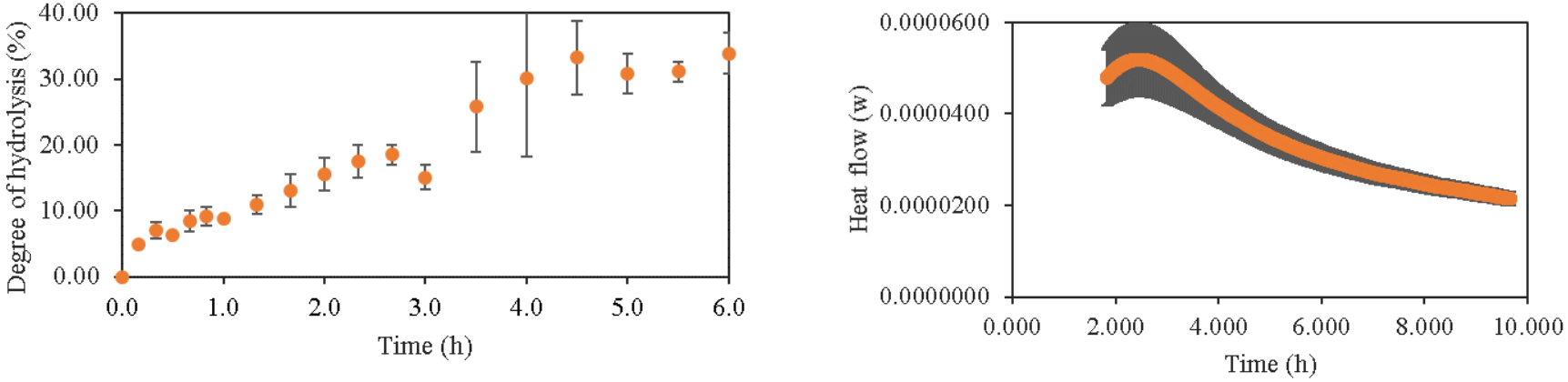
4.1 Comparison between degree of hydrolysis and heat flow data

Two types of methods were used to measure enzymatic protein hydrolysis: Spectroscopy method which measures the degree of hydrolysis and calorimetry which measure the heat flow during the reactions. In this section, a comparison between the two methods is made.

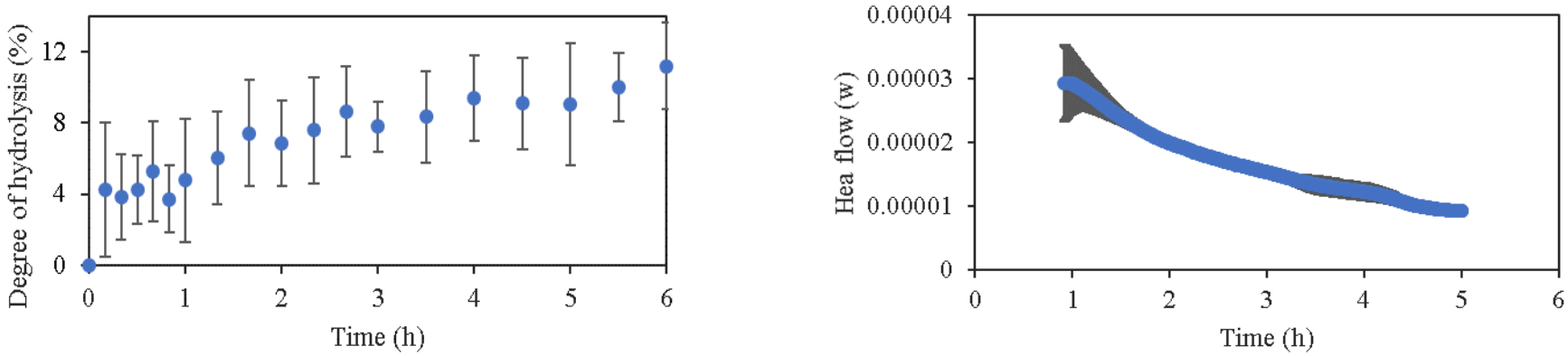
Figure 4-1 presents the degree of hydrolysis and heat flow as a function of time when performing protein hydrolysis using papain and bromelain. The degree of hydrolysis curves for both papain and bromelain shows a rapid increase in the peptides formed during protein hydrolysis and a slow increase as time progresses. This behaviour agrees with what has been established in the literature (Valencia, et al., 2015) (Valencia, et al., 2014) (Kristinsson & Rasco, 2000) (Martinez-Araiza, et al., 2012) (Spellman, et al., 2003). On the other hand, the heat flow graphs show an increase followed by a decrease in the total heat released with time. This is consistent with the degree of hydrolysis data because initially enough peptide bonds are available for hydrolysis hence causing an increase in the reaction rate and heat released thus an increase in the number of bonds broken. However, as the reaction progress, more chains are broken, and new smaller ones are formed, the number of peptide bonds starts to limit the reaction rate hence decreasing the reaction rate and increasing the degree of hydrolysis.

The errors bars in Figure 4-1 indicate that calorimetry technique is more accurate, repeatable and interpretable than the spectroscopy method. This is because the calorimetry method gives accurate and permits the rate of the reaction to be monitored continuously. In addition,

experiments carried out in a microcalorimeter can be left in the machine for a longer period without being laborious. Furthermore, calorimetry method is efficient since multiple experiments can be performed at the same time. On the other hand, the spectroscopy method is laborious because samples need to be individually marked, diluted, mixed with reagent and introduced to the detector by hand before acquiring data. Also, the experiments can only be performed for a certain amount of time because samples need to be taken out after every time interval which can be laborious. Lastly, the data obtained using spectroscopy needs further processing to obtain the degree of hydrolysis while direct data is obtained from calorimetry method. In summary, Isothermal microcalorimetry is useful in predicting enzymatic protein hydrolysis. However, the capital costs associated with Isothermal microcalorimetry are much higher.



a) Papain



b) Bromelain

Figure 4-1: Degree of hydrolysis and heat flow experimental data obtained when hydrolysing whey protein using (a) papain and (b) bromelain.

4.2 Effects of the process conditions on enzymatic protein hydrolysis

The effect of enzyme-to-substrate ratio, pH, temperature and substrate concentration on the degree of hydrolysis of whey protein using papain and bromelain was investigated using a central composite design. The degree of hydrolysis data obtained as described in section 3.2 was statistically analysed using Statistica software. The effect estimates of the individual factors and their interactions on the degree of hydrolysis were evaluated by ANOVA. The final degree of hydrolysis data taken at the end of the experiment were used for ANOVA. The results are summarised in Table 4-1 and Figure 4-2 for papain, Table 4-2 and Figure 4-3 for bromelain.

Table 4-1: Effect estimates of the factors on the degree of hydrolysis of whey protein using papain at 120 min.

Factor	Effect	Standard error	p-value
Linear			
Temperature	3.6421	0.7021	0.0003
pH	0.9653	0.7021	0.197
S	-1.3992	0.7021	0.0717
E:S	2.1930	0.7021	0.0097
Quadratic			
Temperature	1.1267	0.8233	0.1520
pH	-1.9824	0.8233	0.0347
E:S	-0.7944	0.8233	0.3553
S	-0.9010	0.8233	0.2972
Interactions			
Temperature × pH	-2.3465	0.8599	0.0196
Temperature × S	-1.0050	0.8599	0.2672
Temperature × (E:S)	-1.1038	0.8599	0.2256
pH × S	-1.1886	0.8599	0.1943
pH × (E:S)	-0.7284	0.8599	0.4150
S × (E:S)	0.2939	0.8599	0.7389
R-Squared	0.8668		

Table 4-2: Effect estimates of the factors on the degree of hydrolysis of whey protein using Bromelain at 120 min.

Factor	Effect	Standard error	p-value
Linear			
Temperature	2.1679	0.8901	0.03372
pH	-2.0051	0.8941	0.04658
S	0.6606	0.8941	0.4755
E:S	1.7487	0.8941	0.07635
Quadratic			
Temperature	1.5176	1.0484	0.1756
pH	0.14057	1.0483	0.8958
E:S	0.4301	1.0484	0.6895
S	0.4713	1.0484	0.6617
Interactions			
Temperature × pH	-2.5803	1.0950	0.03805
Temperature × S	0.6432	1.0950	0.5688
Temperature × (E:S)	0.06691	1.0949	0.9524
pH × S	-2.0928	1.0950	0.08237
pH × (E:S)	-0.7312	1.0950	0.5181
S × (E:S)	0.7752	1.0950	0.4937
R-Squared	0.7196		

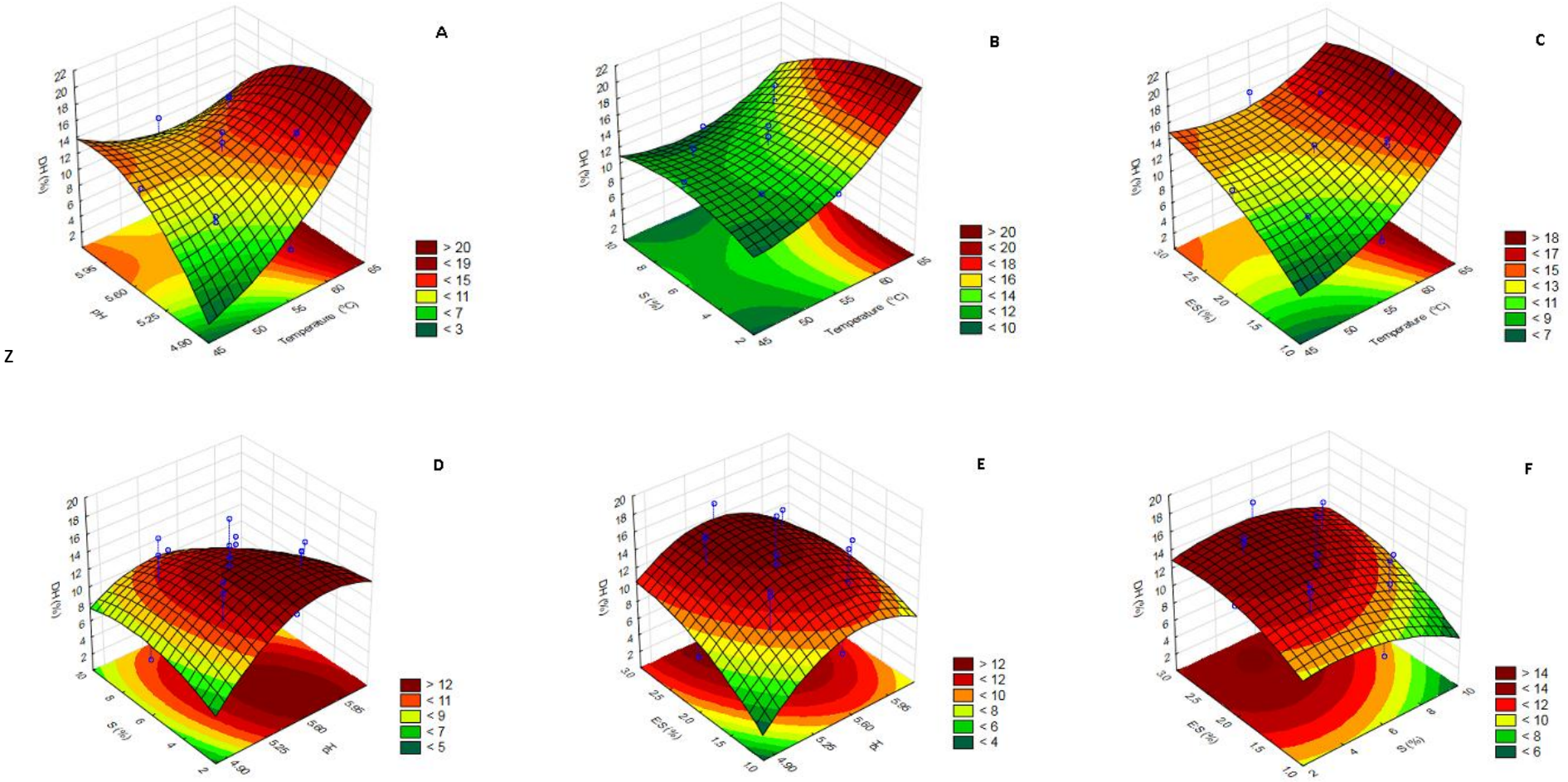


Figure 4-2: Response surfaces of the degree of hydrolysis (DH) obtained by hydrolysing whey protein using papain. (a) The effect of temperature and pH on the DH at substrate concentration (S) = 6 %(w/v), enzyme to substrate ratio (E: S) = 2 %(w/w) (b) the effect of S and temperature when (E: S) = 2 %(w/w), pH = 5.5 (c) the effect of (E: S) and temperature (d) the effect of S and pH when temperature = 55 ºC (e) the effect of (E: S) and pH (f) the effect of (E:S) and S.

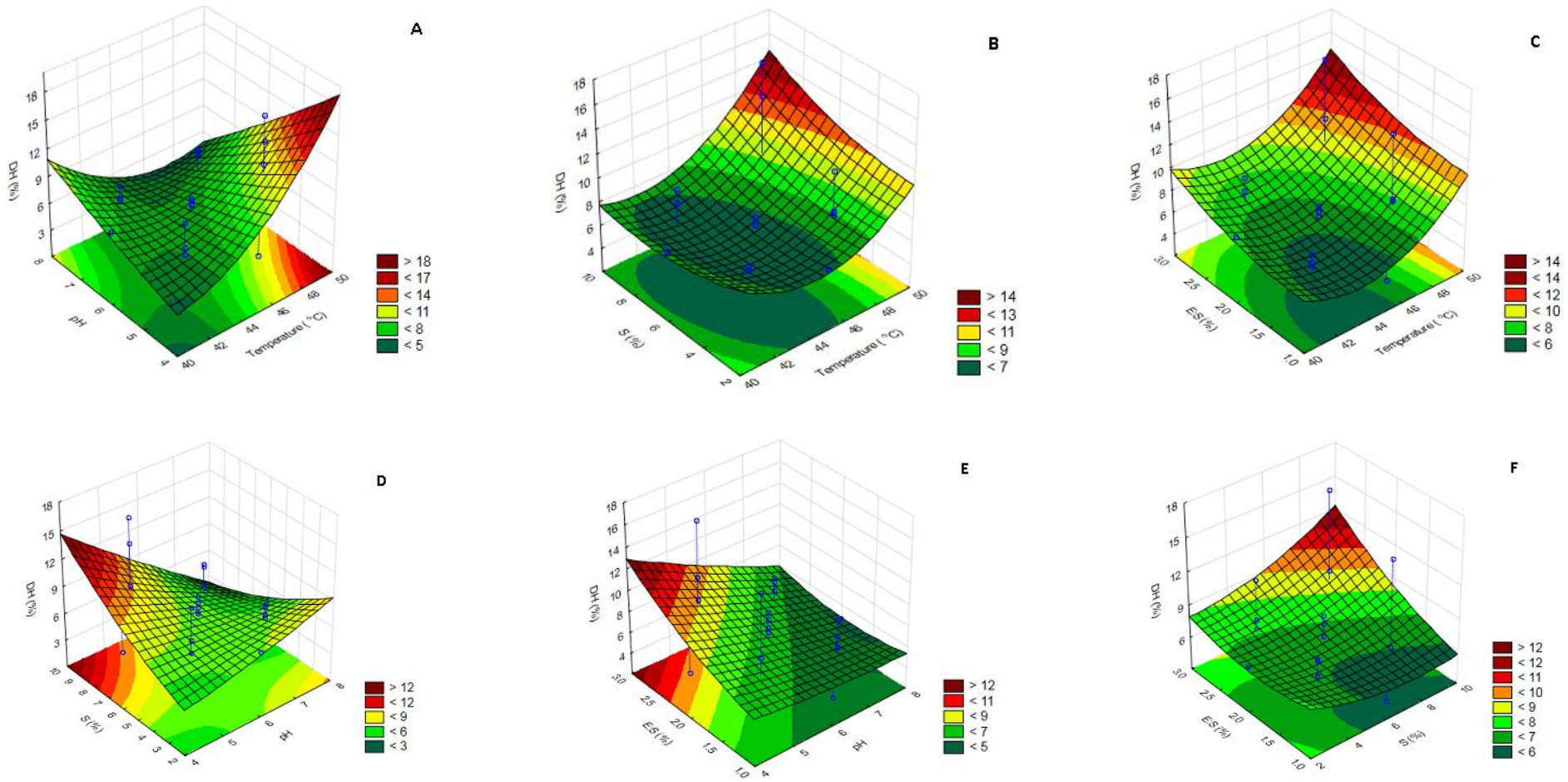


Figure 4-3: Response surfaces of the degree of hydrolysis (DH) obtained by hydrolysing whey protein using bromelain. (a) The effect of temperature and pH on the DH at substrate concentration (S) = 6 %(w/v), enzyme to substrate ratio (E: S) = 2 %(w/w) (b) the effect of S and temperature when (E: S) = 2 %(w/w), pH = 6 (c) the effect of (E: S) and temperature (d) the effect of S and pH when temperature = 45 ºC (e) the effect of (E: S) and pH (f) the effect of (E:S) and S.

The effects of the four factors on protein hydrolysis prepared by using papain as well as bromelain was determined from null hypothesis testing. As can be seen in Table 4-1and Table 4-2, the linear effects of temperature on the degree of hydrolysis using bromelain and papain are highly significant. However, the quadratic effects are not significant. This means that for both the enzymes, the optimal temperatures might lie outside the chosen ranges. This observation is evident in Figure 4-2a, Figure 4-2b, Figure 4-2c for papain and Figure 4-3a , Figure 4-3b, Figure 4-3c for bromelain.

Figure 4-2a, Figure 4-2b and Figure 4-2c show the effect of pH and temperature, substrate concentration (S%) and temperature, and enzyme to substrate ratio (E: S) and temperature, respectively, on the degree of hydrolysis at the end of the experiment(120 min). When the temperature is in between the range 45-65 ºC the degree of hydrolysis increases, indicating that temperature has a positive effect on the degree of hydrolysis. This is because below the deactivation temperature of enzymes, the enzyme activity is dependent on temperature (Ovissipour et al., 2009) (Cornish-Bowden, 2012). Also, when the temperature is at the maximum (65 ºC), a high degree of hydrolysis is achieved which suggests that papain is a heat resistant enzyme and a maximum reaction rate might lie outside the studied range. It is shown in Figure 4-3b and Figure 4-3c that when the temperature is between 40-50 ºC, there exists a minimum degree of hydrolysis. So above the minimal point, the temperature has a positive effect on whey protein hydrolysis using bromelain. Therefore, increasing the temperature above 50 ºC might get the true optimum for bromelain.

The linear effects of pH are significant for bromelain however not for papain over the studied pH range. But, the quadratic pH effects are highly significant for papain. This implies that the optimal levels of pH for papain are inside the studied experimental region. This is demonstrated in Figure 4-2a, Figure 4-2d and Figure 4-2e which present respectively the effect of temperature and pH, substrate concentration (S%) and pH, an enzyme to substrate ratio (E: S) and pH, on the degree of hydrolysis. From the response surface plots, the maximum degree of hydrolysis exists when the pH ranges from 4.8 to 6.2. Below and above the optimum pH, the degree of hydrolysis decreases which suggests that the enzyme activity of papain also decreases. This means that outside the pH of range 4.8 – 6.2 the enzyme might deactivate and the structure of the protein as well as the enzyme can be affected as illustrated by the significance of the quadratic effects.

Enzyme concentration effects are significant for papain however not for bromelain. It was expected that the enzyme to substrate ratio effects will be significant independent of the enzyme type since increasing the enzyme concentration increases the hydrolysis reaction rate (Valencia, et al., 2014). This is because in enzymatic protein hydrolysis, an enzyme acts as a catalyst, therefore, increasing its amount causes the reaction rate and degree of hydrolysis to increase (Valencia, et al., 2014) (Eisenthal & Danson, 2002). This might imply that an enzyme activity of bromelain is very low compared to that of papain. Thus, it suggests that the amount of an enzyme added was not enough to cause a statistically significant change in the degree of hydrolysis. The highest degree of hydrolysis in Figure 4-2c, Figure 4-2e, Figure 4-2f for papain and Figure 4-3c, Figure 4-3e, Figure 4-3f for bromelain occur at high E:S ratio.

Both linear and quadratic effects of substrate concentration are not significant for the two enzymes. This suggests that the effect of substrate concentration on the degree of hydrolysis might be independent on the enzyme specificity. Figure 4-2b shows the effect of substrate concentration (S%) and temperature when the enzyme to substrate ratio (E: S) and pH are constant at 2 % (w/w) and 5.5. It can be clearly seen from Figure 4-2b an initial increase in the degree of hydrolysis at low substrate concentration followed by a decrease at high substrate concentration. This means that at a constant enzyme concentration there exists an optimum substrate concentration for the range over which the investigation was done where the degree of hydrolysis is at its maximum. This observation was reported previously by Adler-Nissen (1986). Adler-Nissen (1986) attributed the increase in the degree of hydrolysis at low substrate concentration due to a lot of enzyme active sites being occupied with substrate thus increasing the number of broken bonds and the decrease at high substrate concentration due to substrate inhibition (Adler-Nissen, 1986). Valencia et al. (2015) also reported that the decrease in the degree of hydrolysis is because the peptides produced during the reaction might be acting as the protease inhibitors.

Still, at substrate concentration, a different behaviour is observed for whey protein concentrate (WPC) hydrolysis using bromelain. Figure 4-3b and Figure 4-3d show that the maximum degree of hydrolysis when hydrolysis whey with bromelain is obtained at the highest studied substrate concentration (10 % w/v). This proposes that the optimum substrate concentration might lie outside the experimental range. This implies that bromelain might have low affinity for whey protein concentrate compared to papain where the maximum degree of hydrolysis is

obtained at the centre (6 % w/v). Butré, et al. (2012) suggested that the effect of substrate concentration on the degree of hydrolysis is due to a generic property of the system.

The effects of the four factors (temperature, pH, enzyme-to-substrate ratio and substrate concentration) on the degree of hydrolysis of whey protein using papain and bromelain within the studied conditions are observed in Table 4-1, Table 4-2, Figure 4-2 and Figure 4-3. Table 4-1 and Table 4-2 shows that the four factors affect the degree of hydrolysis when whey protein is hydrolysed using bromelain and papain. Temperature and enzyme-to-substrate ratio have a positive effect on the degree of hydrolysis. This means that increasing the factors within the limits increases the degree of hydrolysis. The tables also indicate that pH has a significant negative quadratic effect on papain and a significant negative linear effect on bromelain. This implies that for papain the optimal levels of pH are in the extremes of the experimental region and for both enzymes, increasing the pH in the studied ranges decreases the degree of hydrolysis. Figure 4-2and Figure 4-3 presents the response surfaces. It can be seen from the graphs that hydrolysing whey protein with papain at the chosen experimental conditions show optimal points. However, when hydrolysing whey with bromelain it does not show optimal points since the maximum degree of hydrolysis occurs at the boundaries.

The optimum values at the chosen experimental conditions of the four independent factors to yield maximum the degree of hydrolysis are shown in Table 4-3. The optimum values were indicated by the desirability graphs (given in Appendix A.1.2) obtained using Statistica Software.

Table 4-3: Optimum conditions of papain and bromelain

Enzyme	Factor			
	Temperature (^{o}C)	pH	E:S (% w/w)	Substrate concentration (% w/v)
Papain	65	5.15	3 %	6
Bromelain	50	6	3 %	10

4.3 Effect of enzyme concentration and buffer type on enzymatic protein hydrolysis

In the second part of the study, it was desired to see the effect of enzyme concentration and buffer type on enzymatic protein hydrolysis. This was achieved by monitoring the heat released upon mixing of the whey protein and enzymes (papain and bromelain) using an isothermal microcalorimetry (IMC). The experiments were performed for each enzyme type by keeping temperature, pH and substrate concentration constant at the optimum values obtained in the previous section, while varying the enzyme concentration and buffer type.

The effect of enzyme concentration and buffer type on heat released when hydrolysing whey protein with bromelain and papain was observed to be statistically significant from Table 4-4. The p-values represent the probability that the enzyme concentration and buffer type caused changes in the total heat released during protein hydrolysis. Table 4-4 shows the analysis of variance for the total heat released during enzymatic protein hydrolysis. Figure 4-4 presents the total heat flow during protein hydrolysis at different enzyme concentration and in different buffer solutions.

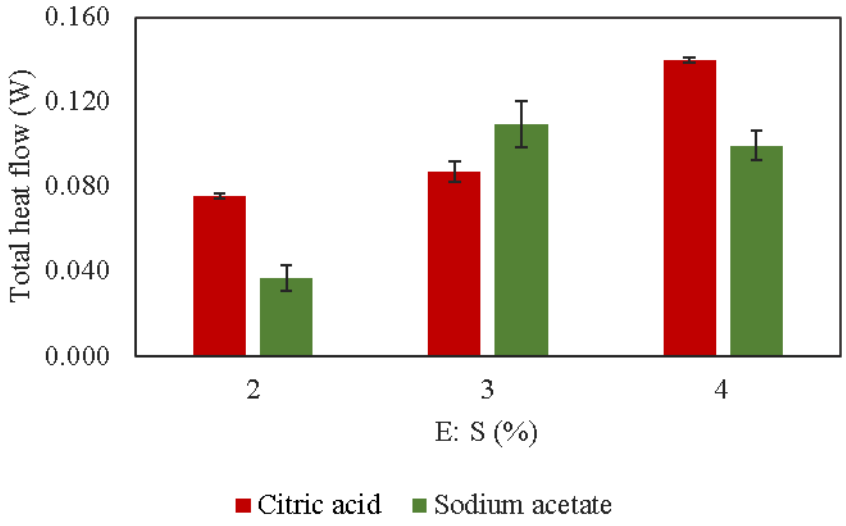
Table 4-4: Analysis of variance for enzyme concentration and buffer type when hydrolysing whey protein using papain and bromelain

Factor	Mean square	p-value
E:S (%)	0.004134	p< 0.01
Buffer type	0.001071	p< 0.01
E:S (%) *Buffer type	0.001285	p< 0.01

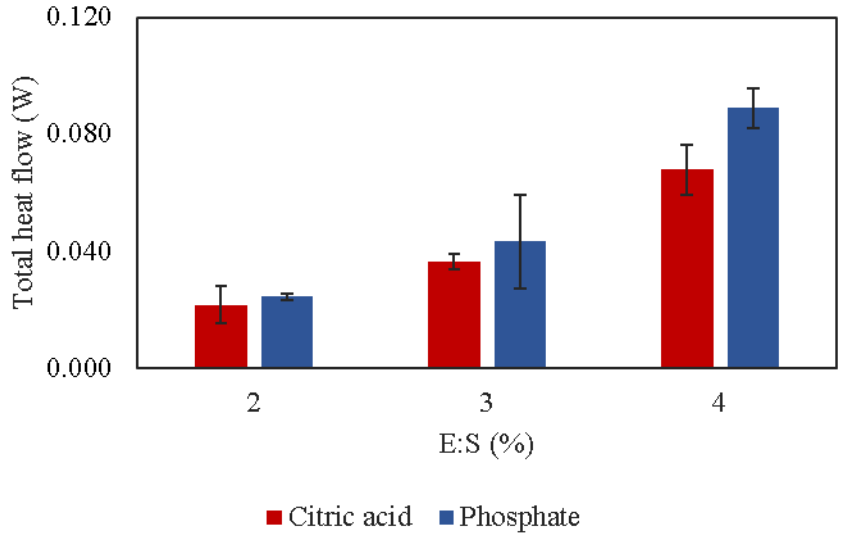
(a) Papain

Factor	Mean square	p-value
E:S (%)	0.004834	p<0.01
Buffer type	0.000481	p< 0.05
E:S (%) *Buffer type	0.000138	p> 0.05

(a) Bromelain



a) Papain



b) Bromelain

Figure 4-4: The effect of buffer type and enzyme concentrations (E:S) on the heat flow when hydrolysis whey protein using (a) papain and (b) bromelain.

The p-values shown in Table 4-4 indicate that enzyme concentration and buffer type have a significant effect on the total heat released during protein hydrolysis, since p is less than 0.05. This means that increasing enzyme concentration increases the number of peptides bonds being broken thus increasing the heat flow as can also be seen in Figure 4-4. The direct proportionality between the enzyme concentration and the number of bonds broken during protein hydrolysis was also observed in the results obtained in Section 4.1. The results imply

that the reaction rate and heat released during enzymatic protein hydrolysis is directly proportional to enzyme concentration.

It is commonly known that the heat measured by a microcalorimeter is the sum of all the heat effects taking place during the studied reaction (Bianconi, 2007). During protein hydrolysis, the buffer solution takes up protons resulting from the formed peptides to maintain the pH constant. This implies that the buffer type affects protein hydrolysis process since the heat released during the reaction is significantly affected by the protonation of the buffer (Mazzei et al., 2016) (Freyer & Lewis, 2008). This is also evidence as can be seen in Figure 4-4.

For papain (Table 4-4a), the interaction between enzyme to substrate ratio (E:S) and the buffer type is significant. This means that the effect of the buffer type depends on the level of E: S when hydrolysing whey protein with papain. This suggests that for instance a high E: S ratio results in an increased heat released during the reaction. Therefore, to maintain the pH constant during the reaction, more protons would be taken up by the buffer conjugate base. However, when the E: S ratio is low, a small amount of heat is released during the hydrolysis reaction. Consequently, fewer protons are taken up by the conjugate base to maintain the pH constant.

Table 4-4b also shows that for bromelain, the interaction between buffer type and E: S are not significant ($p>0.05$). The reason for the different interactions observed for papain and bromelain is that at the optimum conditions in Section 4.1, the expected degree of hydrolysis is very low for bromelain compared to the one obtained when protein hydrolysis was performed using papain. This was deduced from the desirability plots in appendix A.1.2, Figure A-1, Figure A-2 and from section 4.1 results which indicate that an optimum was found for papain but not for bromelain. This suggests that for the results obtained using bromelain, whether the E: S ratio is high or low the heat released might be fairly constant causing no dependency of the buffer type on the E: S level.

In summary, both the buffer type and the enzyme to substrate ratio have a significant effect on the total heat released during enzymatic protein hydrolysis. This implies that the heat released measured by the IMC is the apparent heat flow rather than a reaction heat flow. Apparent heat flow is the sum of the heat of reaction and proton exchange heat flow.

4.4 Model validation

To determine whether the proposed model predicts enzymatic protein hydrolysis, a series of experiments were conducted by hydrolysing whey protein using two types of commercial enzymes, bromelain and papain. Bromelain exhibits endopeptidase properties while papain exhibits both endopeptidase and exopeptidase properties. Enzymatic protein hydrolysis was investigated by measuring the degree of hydrolysis as well as heat flow with respect to time during the experiments. The experimental data obtained were used to validate the model presented in section 3.2.

In this section, the results obtained from the model regression for each enzyme studied in this work are presented. The model predictions and experimental data of each enzyme are compared to the model found in the literature.

4.4.1 Endopeptidase

To validate the model for an enzyme with endopeptidase properties, the model developed in section 3.2.3.1 was compared with the experimental data obtained in section 3.1.5. To support the model proposed in this study, the current model and model found in the literature was also compared with the experimental data. The model found in the literature for enzymatic protein hydrolysis was first presented by Marquez-Moreno and Fernandez-Cuadrado (1993) and studied by Marquez and Vazquez (1999) as well as Valencia et al (2015). Finding from these works, Marquez and Vazquez (1999), and Valencia et al (2015), showed good agreement between the model and experimental data. However, the model was not derived from the fundamentals of enzymatic protein hydrolysis, thus does not have theoretical meaning. Therefore, the motive of the study was to propose a theoretical mathematical model derived from the basic principles of enzymatic protein hydrolysis.

The model input, the total number of breakable bonds (J) in whey protein by papain and bromelain were estimated as 55 and 34 as showed in appendix A.2.1. Regression analysis was performed on the experimental data to determine the model parameters in the proposed model and the model presented by Marquez-Moreno and Fernandez-Cuadrado (1993). The parameters were constrained by looking at the typical kinetic values found in the literature. The optimised model parameters are presented in Table 4-5. The regressed model constants were used to create the predicted reaction profiles as well as to examine the accuracy of the fit (R^2) between

the model predictions and experimental data. These are shown in Table 4-5, Figure 4-5 and Figure 4-6.

Table 4-5: Optimised model parameters and determination coefficient obtained for a) current model and b) literature model for two different buffers and various enzyme (bromelain) concentration.

E:S (%)	Buffer Type	k_{cat} (s^{-1})	k_m $(mol\ l^{-1})$	H_{App} $(J\ mol^{-1})$	R_Q^2	R_{DH}^2
3	Citric acid	0.319	0.000150	3860.4	-0.188	0.630
4	Citric acid	0.319	0.000150	3860.4	0.123	0.905
2	Phosphate	0.319	0.000150	2674.5	0.730	-0.208
4	Phosphate	0.319	0.000150	2674.5	-10.6	0.573

(a)

E:S (%)	Buffer Type	a	b	H_{App} $(J\ mol^{-1})$	R_Q^2	R_{DH}^2
3	Citric acid	0.0595	35.8	3860.4	-7.89	0.712
4	Citric acid	0.0800	35.8	3860.4	-10.7	0.762
2	Phosphate	0.0320	35.8	2674.5	0.474	-0.415
4	Phosphate	0.0800	35.8	2674.5	-13.2	0.866

(b)

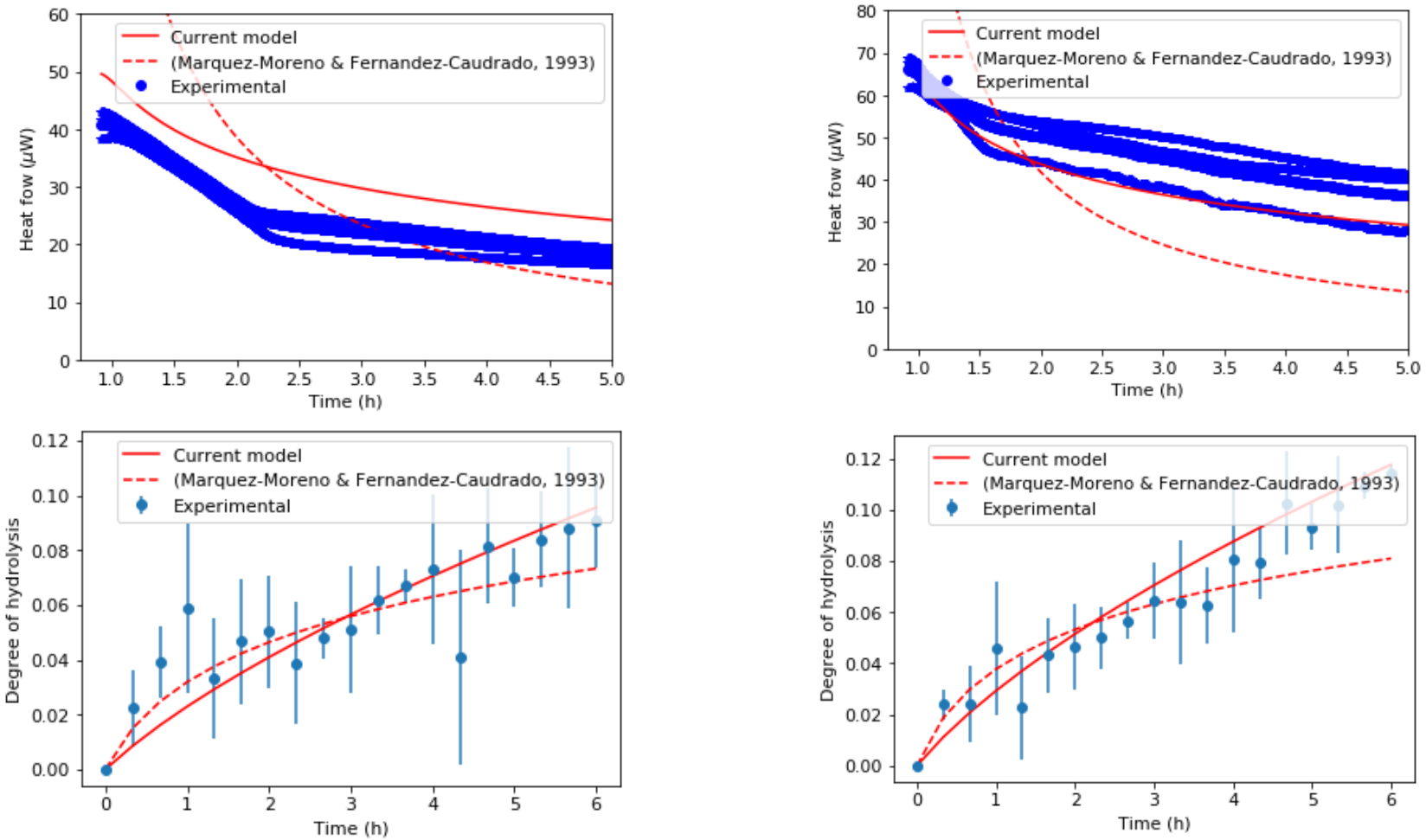
As shown in Table 4-5, the kinetic parameters (k_{cat} and k_m) for each enzyme type are constant. It should be pointed out that the constant kinetic parameters are because they are independent of the enzyme concentration and buffer type. The kinetic parameters changes with temperature and pH. However, in generating the experimental data in Figure 4-5 and Figure 4-6, the pH and temperature were kept at the optimum conditions of the enzyme. On the other hand, apparent

heat (H_{app}) is dependent on the buffer type as discussed in section 4.2 hence changes in H_{app} is observed when buffers are changed. As explained previously, this is because the heat measured by an isothermal microcalorimeter is the sum of all the effects taking place during the studied reaction. During enzymatic protein hydrolysis, the heat associated with the breaking of the peptide bond and the protonation of the buffer are expected (Mazzei, et al., 2016) (Freyer & Lewis, 2008).

It was expected that the coefficients of determination (R_Q^2 and R_{DH}^2) will be positive and in the range between 0 and 1. However, many values are negative because the model parameters were constrained during regression analysis using least squares method. The method works by generating the lowest sum error between the experimental data and model prediction. If the model parameters are constrained, the regression analysis will give a lowest sum squares however that does not mean that the fit is good. In addition, the negative R^2 values mean that the model residuals are much greater than the total variations in the experimental data. This implies that the proposed model is a poor fit to the experimental data.

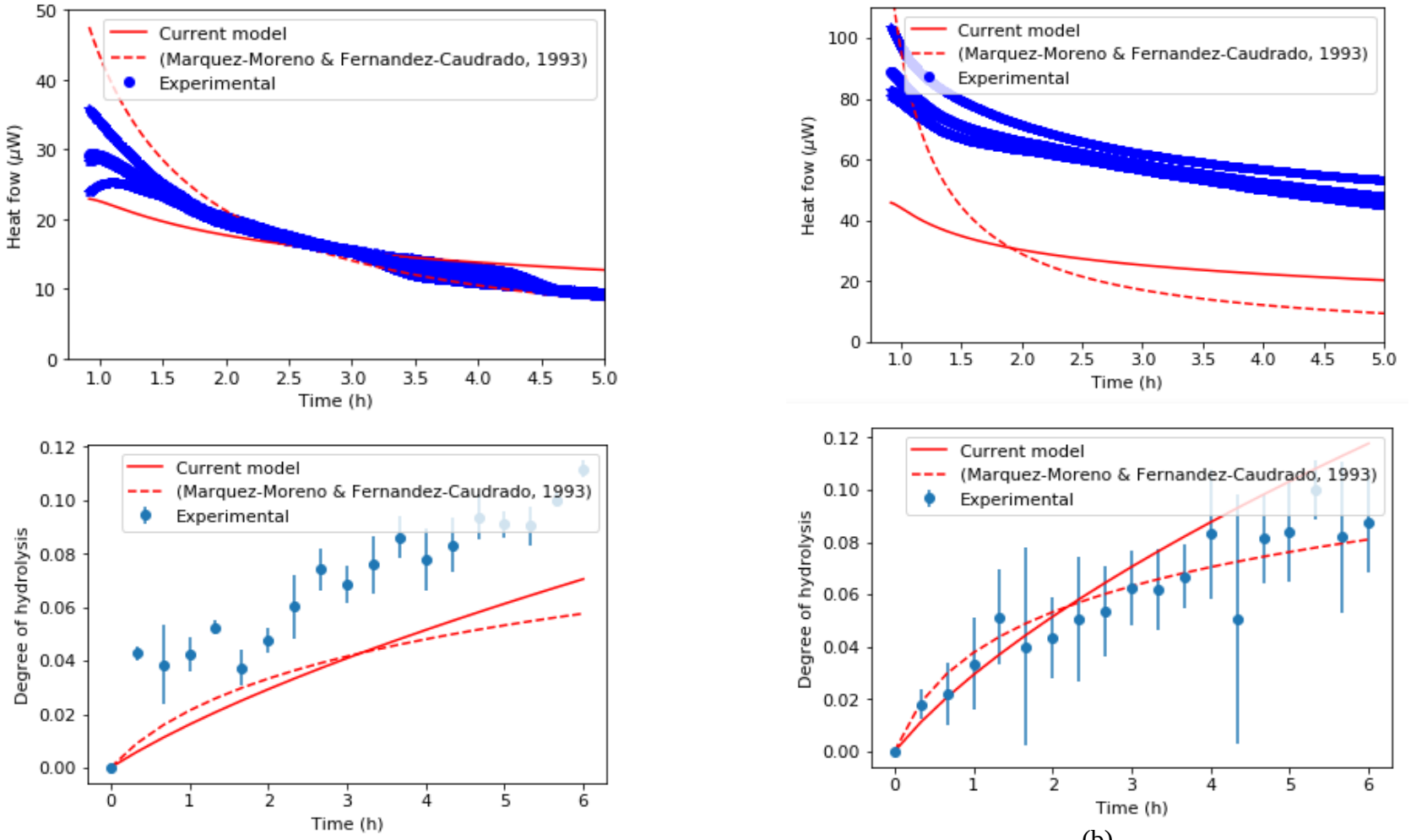
The kinetic constant k_m is interpreted as the substrate concentration that is required to obtain the reaction rate that is exactly half the maximum reaction rate (Cornish-Bowden, 2012). In Table 4-5a, the k_m constant is lower than the initial substrate concentration ($S_o = 0.00423\ mol\ l^{-1}$). This indicates that the initial reaction rate is close to the maximum rate. This suggests that increasing the substrate concentration above 10 % (w/v) will cause the reaction to approach the maximum reaction rate more rapidly thus reaching steady state.

As presented in Table 4-5b, the parameter a is dependent on the enzyme concentration while b is independent of the enzyme concentration and buffer type. The increase in enzyme concentration caused an increase in the value of parameter a. This means that the kinetic parameter a is directly proportional to enzyme concentration. Similar observations have been obtained by Marquez-Moreno and Fernandez-Cuadrado (1993), and Valencia et al. (2015), where the authors reported that kinetic parameter a is significantly affected by the enzyme concentration, temperature and pH while kinetic parameter b is only affected by the substrate concentration. Valencia et al. (2015) attributed the dependency of parameter b to substrate concentration due to the protein properties and its susceptibility to be broken by the enzyme.



(a) (b)

Figure 4-5: Predicted and experimental data for the hydrolysis of whey protein in citric acid buffer at substrate concentration = 10 %(w/v), temperature = 50 ºC, pH = 6 and different bromelain concentration (a) (E:S) 3 % and (E:S) 4 %.



(a) (b)

Figure 4-6: Predicted and experimental data for the hydrolysis of whey protein in phosphate buffer at substrate concentration = 10 %(w/v), temperature = 50 ºC, pH = 6 and different bromelain concentration (a) (E:S) 2 % and (E:S) 4 % .

The calculated regression results using the obtained model parameters for the degree of hydrolysis as well as heat released at different enzyme concentrations and buffer type are observed in Figure 4-5 and Figure 4-6. As can be seen in Figure 4-5 and Figure 4-6 the present model shows a poor prediction of the experimental data. However, the current model seems to predict the experimental data better than the model presented by Marquez-Moreno and Fernandez-Cuadrado (1993) as indicated by R^2 in Table 4-5 and the absolute error values in Table A-5a, appendix A.2.2. The proposed model has the smallest absolute error of 18.04 % and 14.67 % less than the model found in the literature for heat flow and the degree of hydrolysis predicted compared to experimental data. These observations were expected since empirical models are not based on the fundamentals of the process. Also, empirical models do not hold when applied outside their derived conditions. These results indicate that complex models such as population balance approach and multiple substrate Michaelis-Menten equation can be used to describe enzymatic protein hydrolysis better than over-simplified models without theoretical meaning. In summary, the proposed model for endopeptidase action predicted the experimental data better than the model found in the literature.

4.4.2 Endopeptidase-exopeptidase

This section val6idates a model for protein hydrolysis using an enzyme with endopeptidase as well as exopeptidase catalytic action. The parameters for an exopeptidase and endopeptidase model in eqn. 3-26 were determined by fitting the experimental data using least of squares optimization method. The experimental data was obtained by hydrolysis of whey protein with papain. As in the previous section, the current model study was compared with the model found in the literature. The model parameters, model prediction and experimental data are shown in Table 4-6, Figure 4-7 and Figure 4-8.

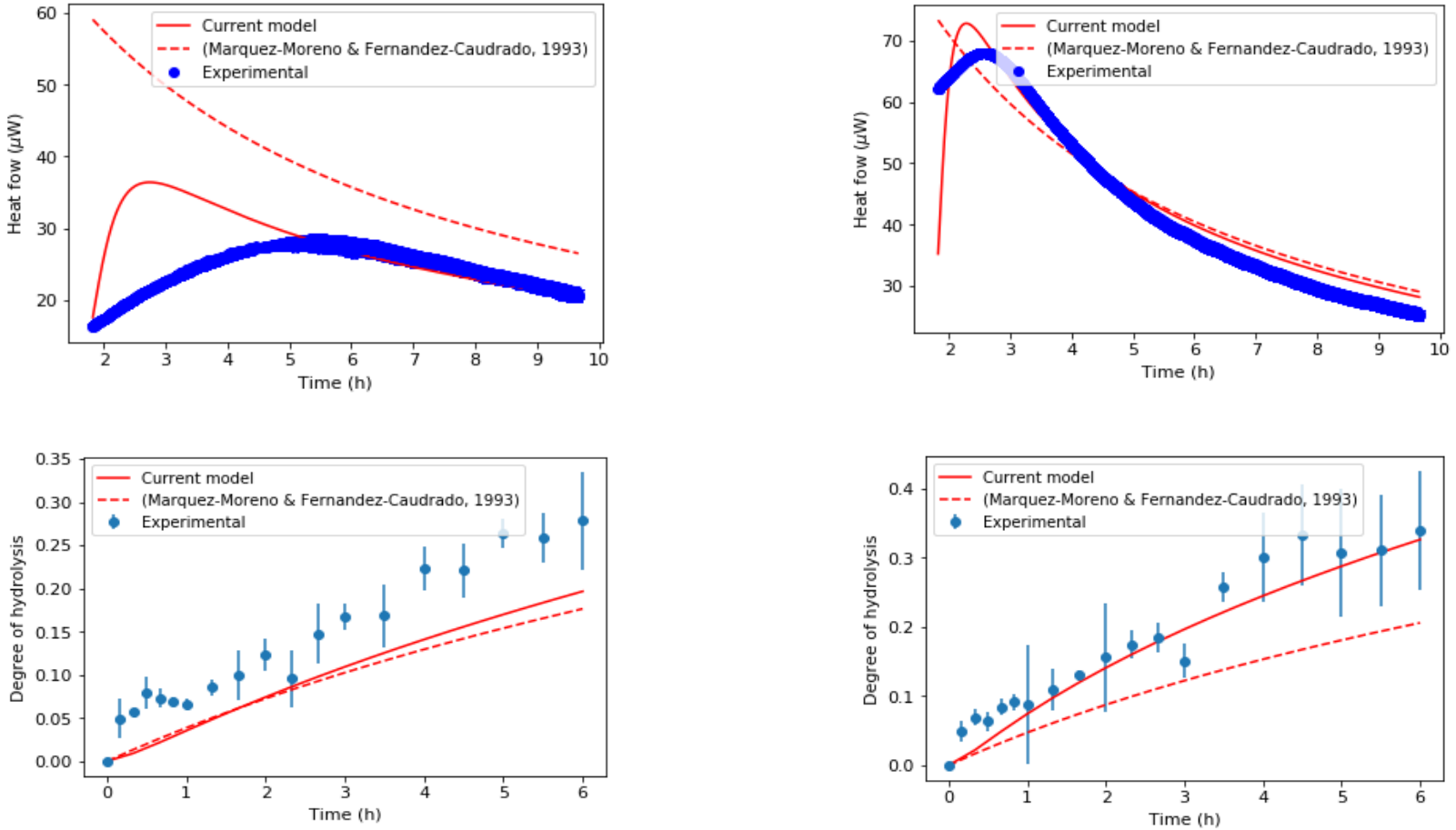
Table 4-6: Optimised model parameters and determination coefficient obtained for a) current model and b) literature model for two different buffers and various enzyme (papain) concentration.

E:S (%)	Buffer type	$k_{cat,endo}$ (s^{-1})	$k_{m,endo}$ $(mol\ l^{-1})$	$k_{cat,exo}$	$k_{m,exo}$ $(mol\ l^{-1})$	H_{App} $(J\ mol^{-1})$	R_Q^2	R_{DH}^2
4	Citric acid	1.804	0.01062	1.10E-05	47.28	2549.5	0.945	0.871
2	Citric acid	1.804	0.01062	1.10E-05	47.28	2549.5	-4.028	0.731
3	Sodium acetate	1.804	0.01062	1.10E-05	47.28	1872.7	0.329	0.828
4	Sodium acetate	1.804	0.01062	1.10E-05	47.28	1872.7	0.958	0.664

(a)

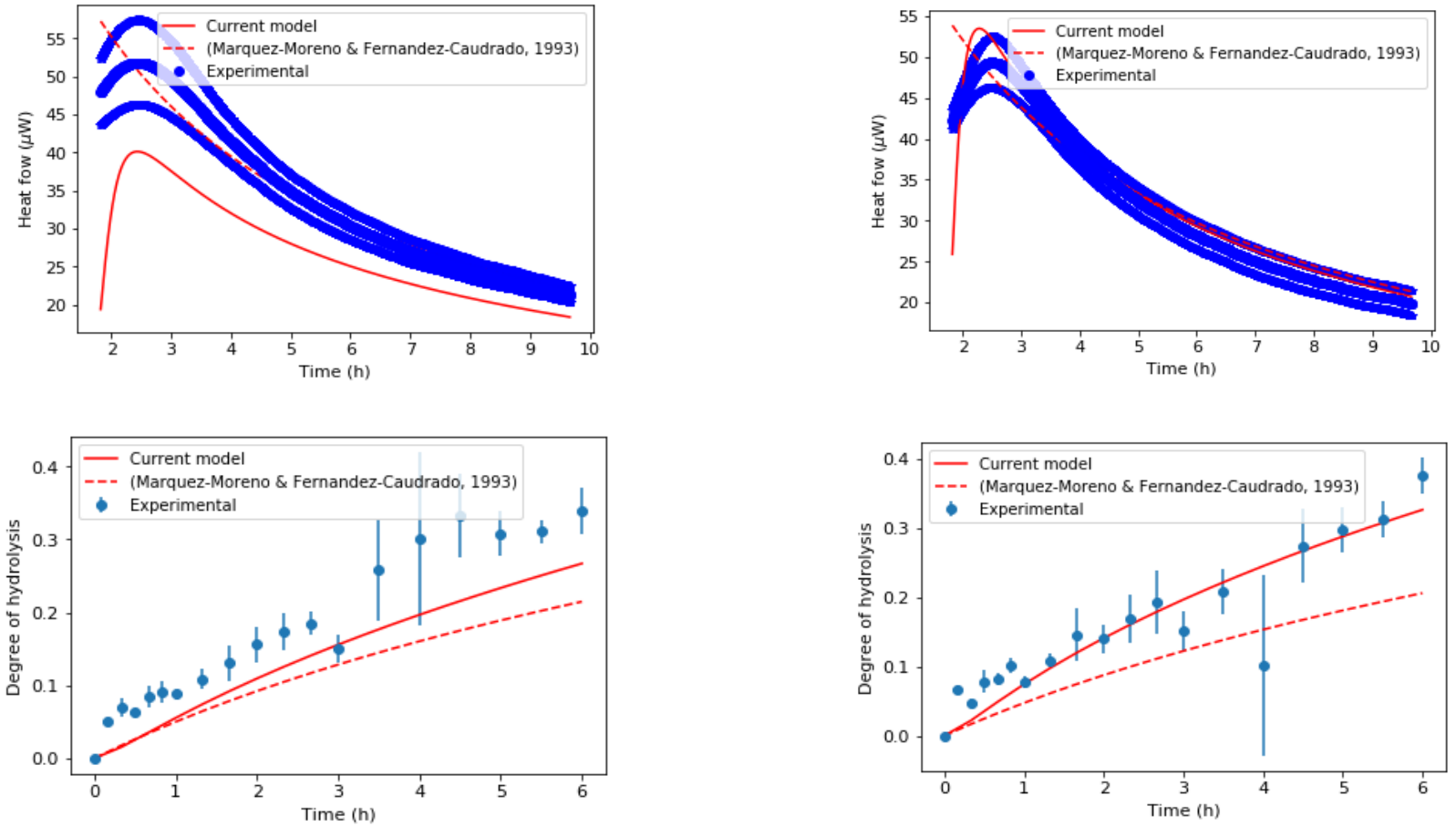
(E:S) %	Buffer type	a	b	H_{App} $(J\ mol^{-1})$	R_Q^2	R_{DH}^2
4	Citric acid	0.05501	3.75	2549.5	0.934	0.515
2	Citric acid	0.04170	3.75	2549.5	-35.5	0.612
3	Sodium acetate	0.05179	3.75	1872.7	-9.11	0.581
4	Sodium acetate	0.05501	3.75	1872.7	0.944	0.563

(b)



(a) (b)

Figure 4-7: Predicted and experimental data for the hydrolysis of whey protein in the citric acid buffer at substrate concentration = 6 %(w/v), temperature = 65 ºC, pH = 5.15 and different papain concentration (a) (E:S) 2 % and (E:S) 4 %.



(a) (b)

Figure 4-8: Predicted and experimental data for the hydrolysis of whey protein in sodium acetate buffer at substrate concentration = 6 %(w/v), temperature = 65 ºC, pH = 5.15 and different papain concentration (a) (E:S) 3 % and (E:S) 4 %.

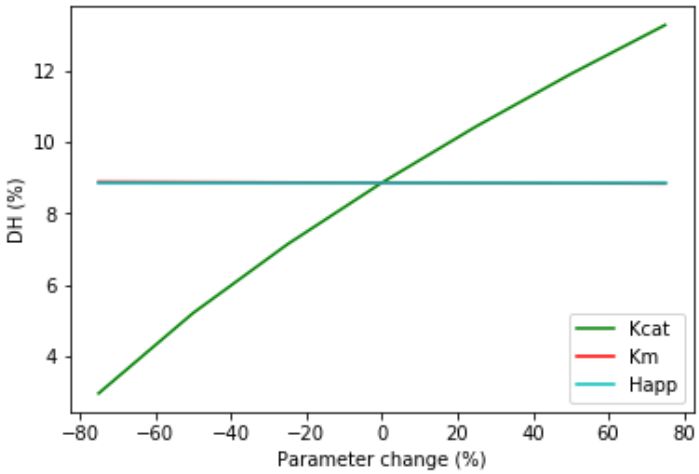
As can be seen in Table 4-6a, the k_m constants are higher than the initial substrate concentration ($S_o = 0.00317\ mol\ l^{-1}$). This indicates that the initial reaction rate is far from the maximum reaction rate. This implies that increasing the initial substrate concentration above 6 % will increase the initial hydrolysis rate thus, decreasing the time for the reaction to reach a steady state.

The determination of coefficients (R^2) in Table 4-6 indicates that the proposed model fits the experimental data. As can be seen in Table 4-6 and Table A-5b in appendix A.2.2, the current model gives a better fit than the model found in the literature. This is because the current model is derived from the fundamental of enzymatic protein hydrolysis while the other model is derived from the empirical data. Furthermore, the model for endopeptidase-exopeptidase activity fit the experimental data better than the model for endopeptidase.

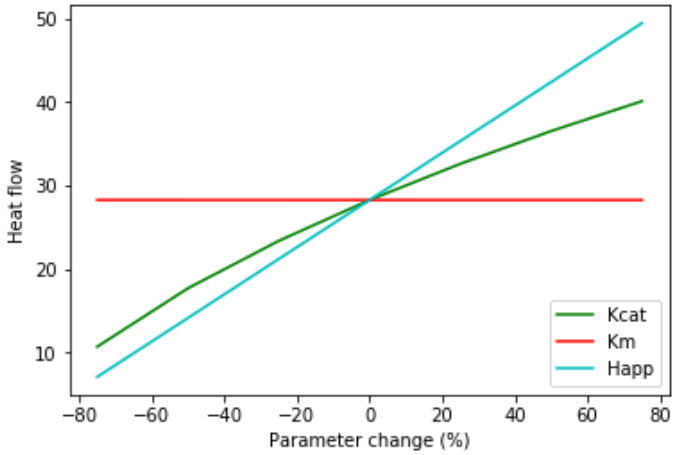
The aim of this section was to validate the proposed model for both endopeptidase and exopeptidase using heat flow data and predict the degree of the hydrolysis. It can be seen from Figure 4-7 and Figure 4-8 that the heat flow data predicts the model even though the fit is not perfect. The current model predicts the experimental data better than the model found in the literature. This implies that the proposed model closely approximate complex enzymatic protein hydrolysis reactions. This suggests complex models can be used to explain the fundamentals of enzymatic protein hydrolysis rather than empirical models.

4.5 Model sensitivity

A sensitivity analysis was performed to assess the validity of the proposed model and to reveal the parameter that has maximum impact on the degree of hydrolysis and heat flow. Local sensitivity analysis was performed in Python and the parameters were set at the fitted values. Then each parameter was increased and decreased by 25 %, 50 % and 75 % of its value. The results are shown in Figure 4-9.



(a)

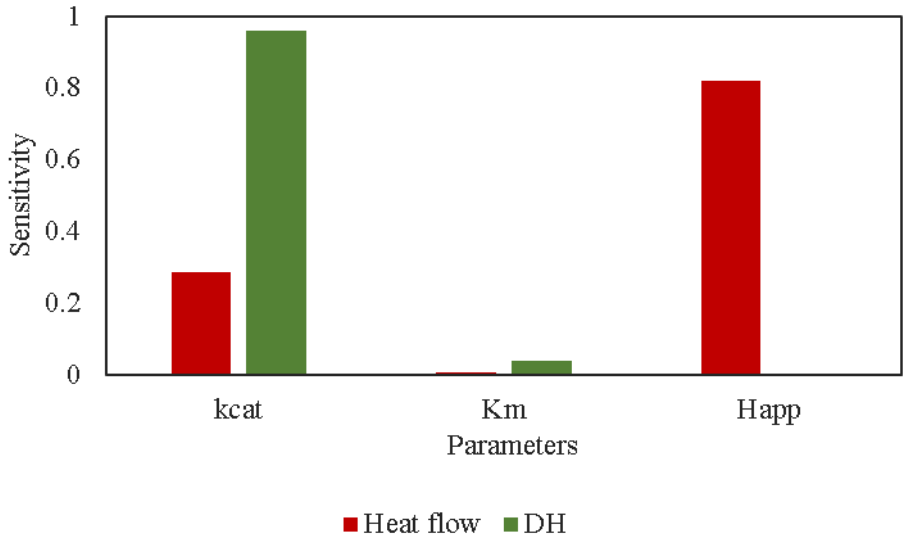


(b)

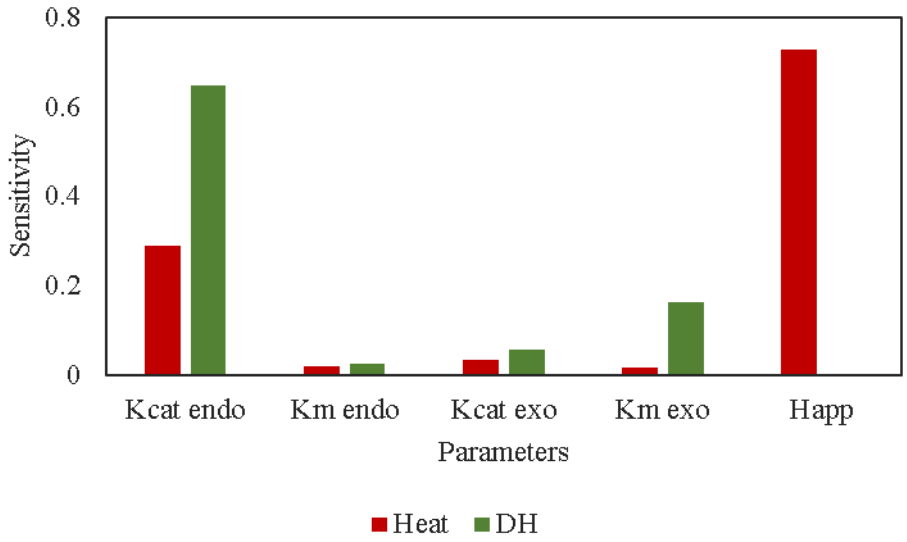
Figure 4-9: Local sensitivity analysis on the model parameters for endopeptidase

The sensitivity analysis indicated that the model is more highly responsive to changes in k_{cat} and H_{app}. The greatest impact on the degree of hydrolysis is changing the k_{cat}. The parameter is related to maximum the number of times each enzyme site converts substrate molecules into products. Parameter k_m had a less significant effect on the degree of hydrolysis. The increase in k_m caused a negative effect on the degree of hydrolysis. This is because the reaction rate expression is inversely related to parameter k_m. The local sensitivity approach involves varying one at a time parameter and checking the effects of the output. The analysis is quick and easy to implement; however, it does not investigate interaction effects between parameters hence it is misleading for nonlinear models (Saltelli, et al., 2008) (Homma & Satelli, 1996)

An additional global Sobol' sensitivity analysis was performed in Python using SAlib function as shown in appendix B.2 to calculate the overall sensitivity over the entire set range of values of all the parameters. The global approach varies the parameters of the model at the same time trying to explore the global input space and quantify effects of that upon the output (Saltelli, et al., 2008) (Homma & Satelli, 1996). As shown in the appendix A.2.3, Table A-6, the interaction effects are minimal as compared to the main effects. Figure 4-10 presents the total (main effects and interactions) Sobol' sensitivity indices. As can be seen in Figure 4-10, k_{cat} is the highest sensitive parameter on the degree of hydrolysis and H_{app} on heat flow. The analysis confirmed that parameter k_{cat} has the most influence on the model.



(a)



(b)

Figure 4-10: Global sensitivity of the model parameters for (a) an endopeptidase model and (b) exopeptidase model.

4.6 Model limitations and recommendations

Enzymatic protein hydrolysis is a very complex reaction. Protein hydrolysis is the breaking down of a peptide bond in a protein using an enzyme. During protein hydrolysis, an enzyme acts as the catalyst. Enzymatic protein hydrolysis involves multiple polypeptides that are formed during the reaction and can also break to form smaller polypeptides. This was

accounted for by considering the population balance approach for modelling the reaction. In addition, the model incorporated the fundamentals of enzyme kinetics. The model also accounts for different ways in which the enzyme breaks the bonds in the polypeptides. Furthermore, it is known that enzymes are specific which means that they prefer to break bonds next to certain amino acids. This implies that there are restriction sites during hydrolysis and not all the peptide bonds in the protein are broken during protein hydrolysis. This was also accounted for in the model.

However, there are model limitations. Some of the factors the model does not account for are as follows:

1. The exact initial chain length and restriction sites. Two of the model inputs are the total number of peptide bonds in the original protein and the total number of the amino acids the enzyme used for protein hydrolysis prefers to break or is specific to. In this study, the two inputs were obtained from the literature and estimated. It is recommended to determine the total number of the peptide bonds and the preferred number of bonds of the studied enzyme experimentally to achieve accurate results with the model.
2. Accurately determine the heat of reaction using an isothermal titration calorimetry.
3. Product and substrate inhibition. Demirhan et al. (2011) and Apar & Ozbek (2009) as well as Sousa, et al. (2006) reported that there exists inhibition during hydrolysis. Maximova and Trylska (2015) verified that there is inhibition of the enzyme by the products using an isothermal titration calorimetry. Valencia, et al. (2014) and Valencia, et al. (2016) studied enzyme kinetics with product and substrate inhibition. Therefore, it is recommended to include inhibition in the model.
4. Effects of pH and temperature. It is recommended to express the kinetic constants as the function of pH and temperature in order to test the predictability of the model.

The above factors affect the reaction rate and degree of hydrolysis. Therefore, considering these factors will allow for greater accuracy of the model.

5 Conclusions and recommendations

5.1 Conclusions

In this study, a population balance-based and modified Michaelis-Menten kinetic model has been developed for enzymatic protein hydrolysis. This approach has not been considered for the enzymatic protein hydrolysis in literature as yet. The proposed model considers the change in the peptides sizes (chain length) during protein hydrolysis reaction and predicts the degree of hydrolysis as well as the heat released during the reaction.

The two-catalytic enzyme action, that is endopeptidase and exopeptidase, have been modelled separately, then a combined model considering the two actions simultaneously. A model for protein hydrolysis by endopeptidase was proposed in which the peptide bond breakage depends on the chain length and peptide bonds specific to an enzyme. A model for the action of exopeptidase was proposed in which there is a constant rate of peptide bonds breakage.

The model has been tested for the two different commercial enzymes, namely bromelain and papain. The degree of hydrolysis and heat flow as a function of time was measured when performing protein hydrolysis using papain and bromelain. The degree of hydrolysis curves for both papain and bromelain shows a rapid increase in the hydrolysis rate and a decrease with time which was also observed from the heat flow data. The heat released during protein hydrolysis as a result of broken bonds to form peptides is directly related to the rate at which the peptides are produced.

For papain, the heat flow curve indicated that the rate of peptide formation increases followed by a decrease as time progresses. This is because initially, a single long chain of amino acids breaks into multiple smaller chains(peptides) causing an increase in peptide concentration thus increasing the reaction rate as well as heat flow. At a certain point in time, although the peptide concentration increases, the rate at which it increases is slow since the decrease in the peptide bonds starts to limit the process thus decreasing the overall rate of peptide formation.

For bromelain, the heat flow curve demonstrated that the heat released as the peptides are formed decreases with time. It was expected that the heat released with time during protein hydrolysis using papain and bromelain will increase in the beginning then decrease. The hypothesis was true for heat flow obtained using papain, but it was false for bromelain. It can

be argued the maxima or peak occurred earlier since the heat flow data was measured after a certain time due to the delay associated with an isothermal microcalorimeter. From the above findings, it can be proposed that the hydrolysis rate decreases due to the decrease in the peptide bonds available for protein hydrolysis and concluded that isothermal calorimetry technique can be used to quantify enzymatic protein hydrolysis.

Two methods were used to measure enzymatic protein hydrolysis. Spectroscopy technique was used to measure the degree of hydrolysis and isothermal microcalorimetry was used to obtain the total released data. The comparison between the two methods showed that isothermal calorimetry technique is easy to use for measuring enzymatic protein hydrolysis compared to Spectroscopy. In addition, the calorimetry method is interpretable, accurate, the data is repeatable and permits continuous monitoring of the reaction rate.

The model predictions determined from the current model and the model found in the literature were compared with the experimental data to validate the models. The determination of coefficient (R^2) showed that the proposed model describes the experimental behaviour better than the model found in the literature. The results proved that the population balance approach and multiple substrate Michaelis-Menten equation can be used to predict enzymatic protein hydrolysis.

The goal of the study was the proposal of a theoretical mathematical model that predicts the degree of hydrolysis over time for enzymatic protein hydrolysis and validation of the model with experimental data obtained using isothermal microcalorimetry. The model was developed using the population balance approach and the multiple substrate Michaelis-Menten equation. According to the results, the proposed complex model can be used to explain the fundamentals of enzymatic protein hydrolysis rather than oversimplified and empirical models without a theoretical background. Isothermal microcalorimetry is useful in predicting the degree of hydrolysis. Furthermore, the proposed model can be tested on other protein-enzyme systems.

6 References

Adler-Nissen, J., 1979. Determination of the degree of hydrolysis of food protein hydrolysates by trinitrobenzenesulfonic acid. *Agriculture food chemistry,* 27(5), pp. 1258 - 1262.

Adler-Nissen, J., 1986. *Enzymatic Hydrolysis of Food Proteins.* New york: Elsevier Applied Science.

Ahluwalia, V. K., Kumar, L. S. & Kumar, S., 2006. *Chemistry of Natural Products.* India: Taylor & Francis.

Ahmadifard, N. et al., 2016. Comparison the effect of three commercial enzymes for enzymatic hydrolysis of two substrate (rice bran protein concerntrate and soy-been protein) with SDS-PAGE. *Journal of food scince and technologies,* 53(2), pp. 1279-1284.

Apar, D. K. & Ozbek, B., 2009. A kinetic study on corn gluten hydrolysis. *Chemical Engineering Technology,* 32(4), pp. 673-675.

Barros, R. M. & Malcata, F. X., 2004. A kinetic model for hydrolysis of whey proteins by cardosin a extracted from cynara cardunculus. *Food Chemistry* , Volume 88, pp. 351-359.

Bhaskar, N., Benila, T. & Lalitha, R. G., 2008. Optimization of enzymatic hydrolysis of visceral waste ptoteins of (Catla catla) for preparation ptoteins hydrolysis using a commercial protease. *Bioresource Technology*, Volume 99, pp. 335-343.

Bianconi, M. L., 2007. Calorimetry of enzyme-catalysed reactions. *Biophysical Chemistry,* Volume 126, pp. 59-64.

Blanch, H. W. & Clark, D. S., 1998. *Biochemical Engineering.* New York: Marcel Dekker Inc.

Butré, C. I., Wierenga,, P. A. & Gruppen, H., 2014. Influence of water availability on the enzymatic hydrolysis of proteins. *Process Biochemistry,* Volume 49, pp. 1903-1912.

Butré, C. I., Wierenga, P. A. & Gruppen, H., 2012. Effects of ionic strength on the enzymatic hydrolysis of Diluted and. *Journal of Agricultural and Food Chemistry*, Volume 60, pp. 5644-5651.

Cornish-Bowden, A., 2012. *Fundamentals of Enzyme Kinetics.* 4th ed. Germany: Wiley-Blackwell.

Demirhan, E., Apar, D. K. & Ozbek, B., 2011. A kinetic study on sesame cake protein hydrolysis by alcalase. *Journal of Food Science,* 76(1), pp. 64-67.

Eftink, M. R., Johnson, R. E. & Biltonen, R. L., 1981. The application of flow microcalorimetry to the study of enzyme kinetics. *Analytical Biochemistry,* Volume 11, pp. 305-320.

Eisenthal, R. & Danson, M. J., 2002. *Enzyme Assays: a practical approach.* 2nd ed. New York: Oxford University Press Inc..

Elavarasan, K. & Shamasundar, B. A., 2016. Proteolysis of water washed meat from catla catla using bromelain and papain: Optimization of hydrolysis parameters. *Journal of Fisheries and Life sciences,* 1(1), pp. 1-7.

Engelen, M. P. et al., 2012. Casein protein results in higher prandial and exercise induced whole body protein anabolism than whey protein in chronic obstructive pulmonary disease. *Metabolism Clinical and Experimental,* Volume 61, pp. 1289-1300.

Falconer, R. J., 2016. Application of isothermal titration calorimetry-the research and technical development from 2011 to 2015. *Journal of Molecular Recognition,* Volume 29, pp. 205-515.

Fernandez, A. & Kelly, P., 2016. pH stat vs free-fall pH techniques in the enzymes hydrolysis of whey proteins. *Food Chemistry,* Volume 199, pp. 409-415.

Fountoulakis, M. & Lahm, H.-w., 1998. Hydrolysis and amino acid composition analysis of proteins. *Journal of Chromatography A,* Volume 826, pp. 109-134.

Freyer, M. F. & Lewis, E. A., 2008. Isothermal titration calorimetry: Experimental design, data analysis and probing macromolecule/ligand binding and kinetic interactions. *Method in cell Biology,* Volume 84, pp. 79-113.

Hammes, G. G., 1982. *Enzyme Catalysis and Regulation Student Edition.* New York: Academic Press.

Hathwar, S. C., Bijinu, B., Rai, A. K. & Narayan, B., 2011. Simultaneous recovery of lipids and proteins by enzymatic hydrolysis of fish industry waste using different commercial proteases. *Applied Biochemical Biotechnology,* Volume 164, p. 115–124.

Homma, T. & Satelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety,* Volume 52, pp. 1-17.

Hosseini, S. A. & Shah, N., 2011. Modelling enzymatic hydrolysis of cellulose part I: Population balance modelling of hdrolysis by endoglucanase. *Biomass and Bioenergy* , Volume 35, pp. 3841-3848.

Kristinsson, H. G. & Rasco, B. A., 2000. Fish protein hydrolysates: Production, biochemical, and functional properties. *Critical Reviews in Food and Nutrition,* 40(1), pp. 43-81.

Kumar, S. & Ramkrishna, D., 1996. On the solution of population balance equations by discretization - I. A fixed pivot technique. *Chemical Engineering Science,* 51(8), pp. 1333-1342.

Le Maux, S., Nongonierma, A. B., Barre, C. & FitzGerald, R. J., 2016. Enzymatic generation of whey protein hydrolysates under pH-controlled and non pH-controlled conditions: Impact on physicochemical and bioactive properties. *Food Chemistry,* Volume 199, pp. 246-251.

Lebaz, N., Cockx, A., Spérandio, M. & Morchain, J., 2015. Population balance approach for the modelling of enzymatic hydrolysis of cellulose. *The Canadian Journal of Chemical Engineering,* Volume 93, pp. 276-284.

Marquez, M. C. & Vazquez, M. A., 1999. Modelling of enzymatic protein hydrolysis. *Process Biochemistry,* Volume 35, pp. 111-117.

Marquez-Moreno, M. C. & Fernandez-Cuadrado, V., 1993. Enzymic hydrolysis of vegetable proteins: Mechanism and kinetics. *Process Biochemistry,* Volume 28, pp. 481-490.

Martinez-Araiza, G. et al., 2012. Modeling of enzymatic hydrolysis of whey proteins. *Food Bioprocess Technology,* Volume 5, pp. 2596-2601.

Maurer, H. R., 2001. Bromelain: biochemistry, pharmacology and medical use. *Celluar and Molecular Life Science,* Volume 58, pp. 1234-1245.

Maximova, K. & Trylska, J., 2015. Kinetics of trypsin-catalyzed hydrolysis determined by isothermal. *Analytical Biochemistry*, Volume 486, pp. 24-34.

Mazzei, L., Ciurli, S. & Zambelli, B., 2016. Isothermal titration calorimetry. *Methods in Enzymology*, Volume 567, pp. 215-235.

Montgomery, D. C., 1997. *Design and Analysis of Experiments*. 5th ed. New York: John Wiley & Sons.

Nielsen, P. M., Petersen, D. & Dambmann, C., 2001. Improved method for determining food protein degree of hydrolysis. *Journal of Food Science*, 66(5), pp. 642-646.

Ovissipour, M., Abedian, A., Motamedzadegan, A., Rasco, B. & Safari, R., 2009. The effect of enzymatic hydrolysis time and temperature on the properties of protein hydrolysates from Persian sturgeon (Acipenser persicus) viscera. *Food Chemistry*, 115(1), pp.238–242.

Palmer, T., 1995. *Understanding Enzymes*. 4th ed. Hertfordshire: Prentice Hall/Ellis Horwood.

Pelegrine, D. H. & Gasparetto, C., 2005. Whey proteins solubility as function of temperature and pH. *Lebensmittel-Wissenschaft & Technologie*, Volume 38, pp. 77-80.

Qi, W. & He, Z., 2006. Enzymatic hydrolysis of protein: Mechanism and kinetic model. *Front. Chem. China*, Volume 3, pp. 308-314.

Saltelli, A. et al., 2008. *Global Sensitivity Analysis: The Primer*. England: John Wiley & Sons Ltd.

Sinha, R., Radha, C., Prakash, J. & Kaul, P., 2007. Whey protein hydrolysate: Functional properties, nutritional quality and utilization in beverage formulation.. *Food Chemistry*, Volume 101, pp. 1484-1491.

Sousa, R. et al., 2004. Kinetic model for whey protein hydrolysis by alcalase multipoint-immobilised on agarose gel particles. *Brazilian Journal of Chemical engineering* , 21(02), pp. 147-153.

Spellman, D., McEvoy, E., O'Cuin, G. & FitzGerald, R. J., 2003. Proteinase and exopeptidase hydrolysis of whey protein:. *International Dairy Journal*, Volume 13, p. 447–453.

Stodeman, M. & Schwarz, F. P., 2002. Importance of product inhibition in the kinetics of the acylase hydrolysis reaction by differential stopped flow microcalorimetry. *Analytical Biochemistry*, Volume 308, p. 285–293.

Todd, M. J. & Gomez, J., 2001. Enzyme kinetics determined using calorimetry: A general essay for enzyme activity?. *Analytical Biochemistry,* Volume 296, pp. 179-181.

Tucker, G. A. & Woods, L. F., 1995. *Enzymes in Food Processing.* 2nd ed. London: Chapman and Hall.

Valencia, P. et al., 2015. Novel modeling methodology for the characterization of enzymatic hydrolysis of proteins. *Process Biochemistry,* Volume 50, pp. 580-597.

Valencia, P., Pintso, M. & Almonacid, 2014. Identification of the key mechanism involved in the hydrolysis of fish protein by alcalase. *Process Biochemistry,* Volume 49, pp. 258-264.

Walsh, G., 2002. *Proteins Biochemistry and Biotechnology.* New York: John Wiley & Sons.

Williams, B. A. & Toone, E. J., 1993. Calorimetric evaluation of enzyme kinetics parameters. *Journal of Organic Chemistry,* Volume 58, pp. 3507-3510.

Wisuthiphaet, N., Kongruang, S. & Chamcheun, C., 2015. Production of fish protein hydrolysates by acid and enzymatic hydrolysis. *Journal of Medical and Bioengineering,* 4(6), pp. 466-470.

Zapata-Montoya, E. J., Giraldo-Rios, D. E. & Baez-Suarez, A. J., 2018. Kinetc modelling of the enzymatic hydrolysis of proteins of viscera from red tilapia (Oreochromis sp.): Effect of substrate and enzyme concerntration. *Vitae Magazine of the Faculty of Pharmaceutical and Food Science,* 25(1), pp. 17-25.

Appendix A: Results

A.1 Statistica results

Hypothesis:

-Null hypothesis

$H_o: \mu_1 = \mu_2 = \cdots \mu_a$

-Alternate hypothesis

$H_1: \mu_i \neq \mu_j$ for at least one pair(i, j)

$\alpha = 0.05$ 95% confidence interval

A.1.1 Degree of hydrolysis CCD

Regression model

The following equation describes the response surfaces shown in Figure 4-2 and Figure 4-3.

$$y = b_o + \sum b_i x_i + \sum b_{ii} x_i^2 + \sum_{i<j} b_{ij} x_i x_j \qquad \text{(A-1)}$$

Where b_o is the constant, b_i, $b_{ii}$and b_{ij} are the regression coefficients for linear, quadratic and interaction terms respectively.y and x_i x_j are the depended variable and the operating conditions. Table A-1 present the regression constant values.

Table A-1: Regression coefficients for the regression model for predicting degree of hydrolysis using papain and bromelain.

Factor	Papain		Bromelain	
	b-coefficient	Standard error	b-coefficient	Standard error
Linear				
Temperature	2.007	2.3013	-7.83647	7.7418
pH	136.508	39.8437	25.97803	12.0159
S	8.141	4.3929	-1.07202	5.5408
E:S	31.255	17.9600	-0.83515	22.6623
Quadratic				
Temperature	0.025	0.0165	0.12141	0.0839
Ph	-8.091	3.3602	0.07029	0.5242
E:S	-1.589	1.6465	0.86022	2.0968
S	-0.113	0.1029	0.05892	0.1310
Interactions				
Temperature × pH	-0.670	0.2457	-0.51605	0.2190
Temperature × S	-0.050	0.0430	0.06432	0.1095
Temperature × (E:S)	-0.221	0.1720	0.02677	0.4380
pH × S	-0.849	0.6142	-0.52320	0.2737
pH × (E:S)	-2.081	2.4567	-0.73118	1.0950
S × (E:S)	0.147	0.4299	0.38759	0.5475

R-Squared	0.8668	0.7196
R-adjusted	0.6973	0.3628

A.1.2 Optimum conditions: Desirability plots

Papain

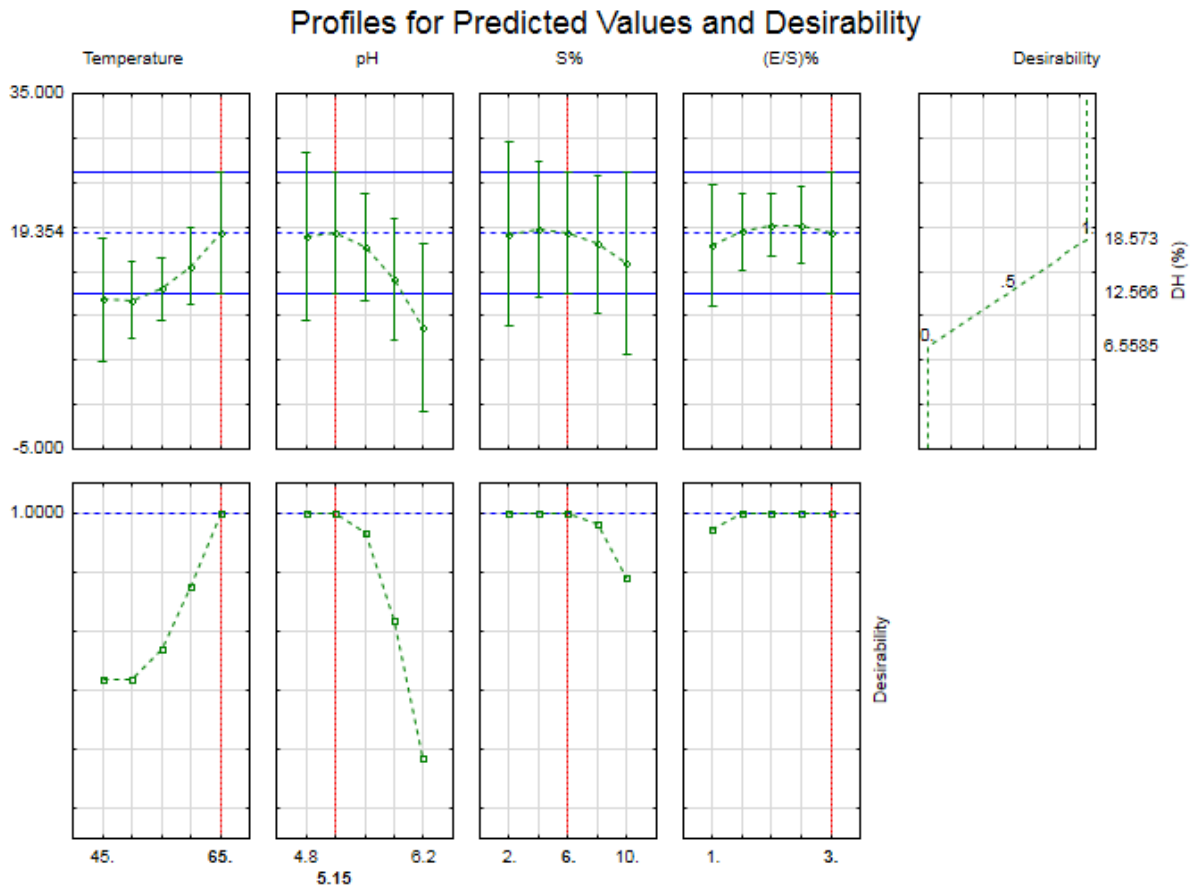


Figure A-1: Optimum conditions of papain.

Bromelain

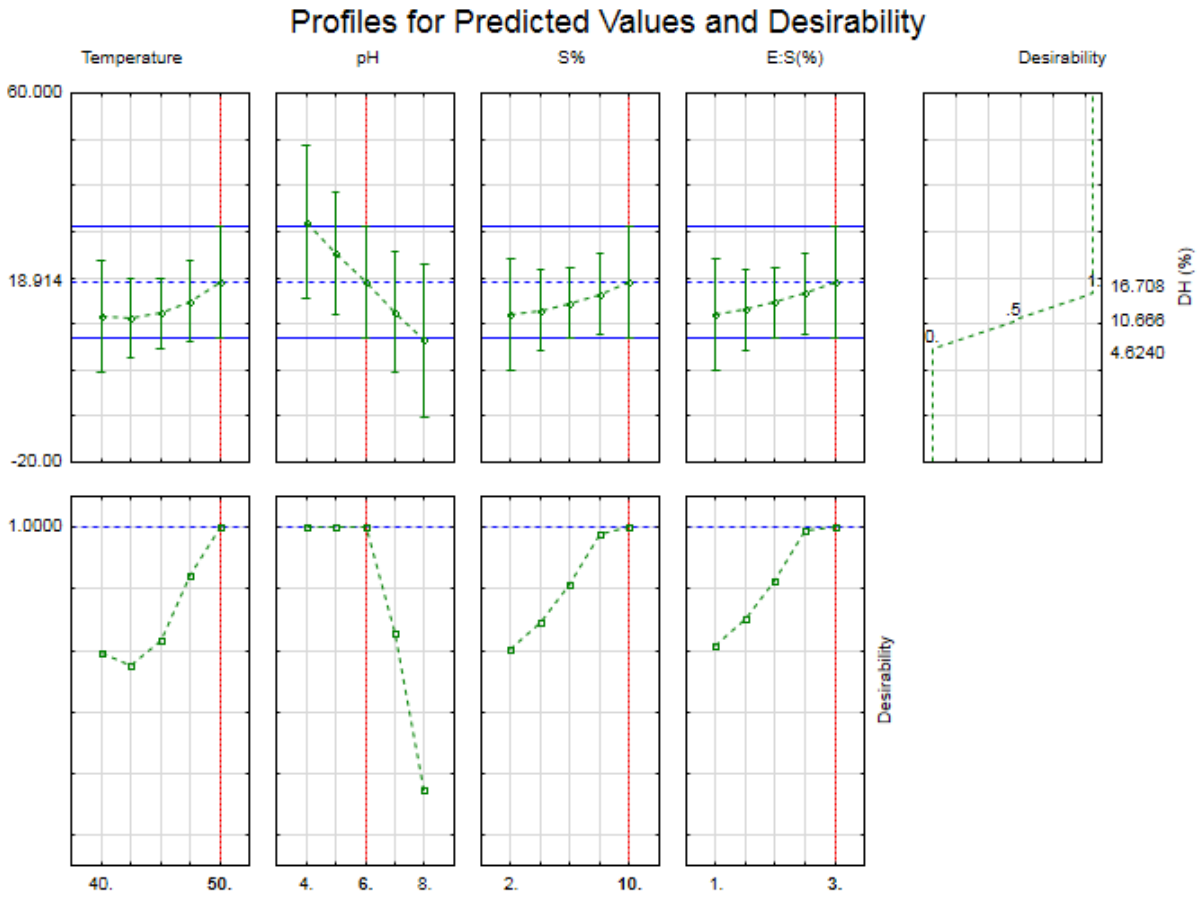


Figure A-2: Optimum conditions of bromelain.

A.1.3 Heat released data: Full ANOVA analysis

Table A-2: Analysis of variance for enzyme concentration and buffer type when using hydrolysis whey using a) papain and b) bromelain.

Factor	Sum of squares	Degrees of freedom	Mean square	F-test	p-value
E:S (%)	0.008269	2	0.004134	108.514	0.000019
Buffer type	0.001071	1	0.001071	28.109	0.001827
E:S (%) *Buffer type	0.002570	2	0.001285	33.727	0.000545
Error	0.000229	6	0.000038		

(a) Papain

Factor	Sum of squares	Degrees of freedom	Mean square	F-test	p-value
E:S (%)	0.009668	2	0.004834	67.9898	0.000000
Buffer type	0.000481	1	0.000481	6.7624	0.023206
E:S (%) *Buffer type	0.000276	2	0.000138	1.9396	0.186260
Error	0.000853	12	0.000071		

(b) Bromelain

A.2 Calculated results

A.2.1 Determining the number of restriction site

Table A-3: Amino acid composition of whey protein as determined by the Central Analytical Facility (CAF) Stellenbosch.

Amino acid	% m/m (according to dry mass of sample in g)	Composition
Alanine	4.31	0.048162
Arginine	2.36	0.026372
Aspartic acid	9.11	0.101799
Glutamic acid	15.24	0.170298
Glycine	1.67	0.018661
Histidine	1.99	0.022237
Isoleucine	5.63	0.062912
Leucine	9.54	0.106604
Lysine	6.92	0.077327
Methionine	3.02	0.033747
Phenylalanine	4.21	0.047044
Proline	4.26	0.047603
Serine	4.91	0.054866
Threonine	7.11	0.07945
Tyrosine	3.8	0.042463
Valine	5.41	0.060454
Sum	89.49	1

The number of breakable bonds in Table A-4 were calculated by adding the compositions of the specific amino acids and multiplying by the total number of peptide bonds.

Table A-4: The total number of breakable bonds by bromelain and papain in whey protein hydrolysis.

Enzyme	Specificity	J (Number of breakable bonds)
Bromelain	Ala, Gly, Leu	34
Papain	Arg, Gly, Leu, Lys, Phe	55

A.2.2 Model error

The literature and proposed model error in Table A-5 was calculated using eqns. A-1 and A-2.

$$error = \left|R^2_{Predicted} - R^2_{actual}\right| \quad \text{(A-2)}$$

$$\% = \frac{error_{literature} - error_{proposed}}{error_{literature}} \times 100 \quad \text{(A-3)}$$

Table A-5: Model error for the literature and proposed model for a) endopeptidase and b) for endopeptidase-exopeptidase.

Factors		Current		Literature		%	
E:S (%)	Buffer type	Error Q	Error DH	Error Q	Error DH	Q	DH
3	Citric acid	1.188	0.3696	8.889	0.2884	86.63	28.17
4	Citric acid	0.8766	0.0950	11.74	0.2385	92.53	60.18
2	Phosphate	0.2697	1.208	0.5263	1.415	48.75	14.67
4	Phosphate	11.62	0.4274	14.18	0.1335	18.04	220.1

(a) Endopeptidase

Factors		Current		Literature		%	
E:S (%)	Buffer type	Error Q	Error DH	Error Q	Error DH	Q	DH
4	Citric acid	0.05508	0.1293	0.06569	0.4847	16.16	73.33
2	Citric acid	5.028	0.2693	36.54	0.3879	86.24	30.58
3	Sodium acetate	0.6712	0.1720	10.11	0.4191	93.36	58.95
4	Sodium acetate	0.04182	0.3363	0.05635	0.4374	25.78	23.11

(b)Endopeptidase-exopeptidase

A.2.3 Sobol sensitivity analysis

Table A-6: First order and total sensitivity indices for a) endopeptidase model and b) endopeptidase-exopeptidase.

Parameter	First order		Total	
	Heat flow	DH	Heat flow	DH
k_{cat}	0.250425	0.965992	0.286111	0.960689
k_m	0.006349	0.033623	0.007961	0.037823
H_{app}	0.729178	0	0.821209	0

a) Endopeptidase

Parameter	First order		Total	
	Heat flow	DH	Heat flow	DH
$k_{cat,endo}$	0.249266	0.648598	0.289471	0.649053
$k_{m,endo}$	0.003689	0.01576	0.020293	0.027779
$k_{cat,exo}$	-0.00528	0.135135	0.036036	0.059345
$k_{m\,exo}$	-0.00154	0.149052	0.01735	0.164187
H_{app}	0.782769	0	0.729485	0

b) Endopeptidase-exopeptidase

A.2.4 Characterisation of peptide hydrolysates

Addition experiment was performed to characterise peptide hydrolysates and to check validity of the proposed model.

A.2.4.1 Method

In order to characterise peptides with different sizes, protein hydrolysates (section 3.1.5.3) at different time points obtained from hydrolysing whey with papain at 65 ºC, pH = 5.5, substrate concentration of 6 %(w/v) and enzyme-to-substrate ratio of 4%(w/w) were analysed using liquid chromatography and mass spectroscopy (LC-MS) method by Central Analytical Facility(CAF) Stellenbosch.

A.2.4.2 Results

Protein hydrolysates at different time intervals were analysed using LC-MS technique performed by Central Analytical facility(CAF) in Stellenbosch to examine the peptides formed during protein hydrolysis. Figure A-3 shows the results obtained. Figure A-4 presents the number of peptides at different chain lengths predicted using the proposed model.

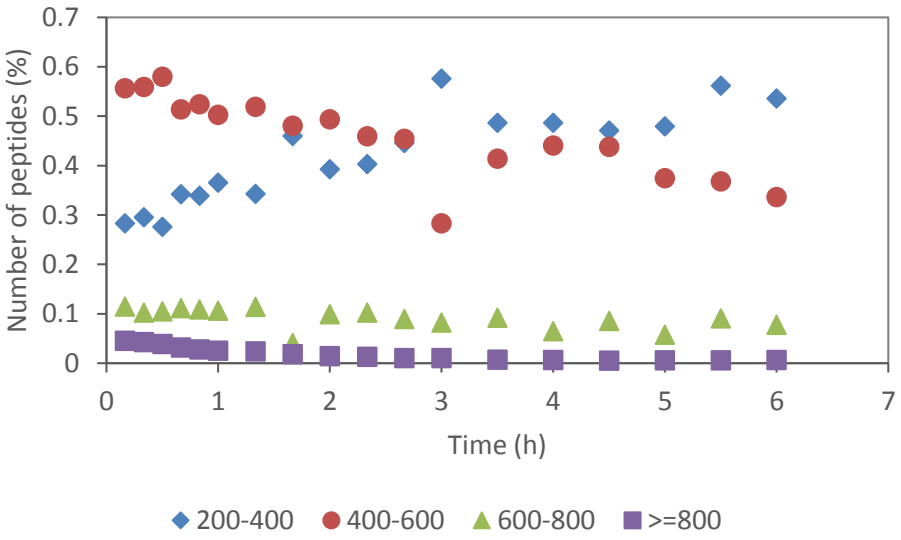


Figure A-3: Number of peptides with different mass to charge ratio as a function of time obtained using LC-MS method.

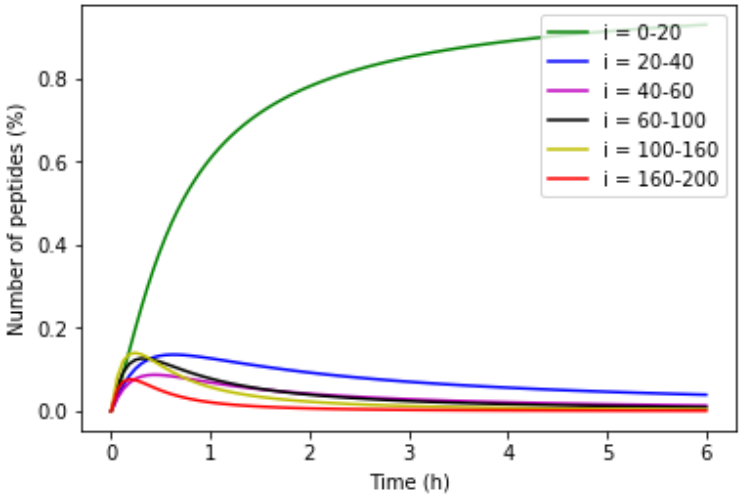


Figure A-4: Number of peptides with different chain lengths as a function of time obtained from the proposed model.

As can be seen in Figure A-3, the peptides with large mass to charge ratio decreases while the peptides with small mass to charge ratio increases with time. Similarly, in Figure A-4, the peptides with shorter chain length increase while the longer peptides decrease with time during protein hydrolysis. This implies that peptides with longer chains are more likely to be broken than shorter peptide chains, since peptides with longer chain length have a higher probability of containing a breakable bond during the hydrolysis process. This elucidate that the proposed model predicts enzymatic protein hydrolysis.

In comparing the two figures, the results from the model shows an increase followed by a decrease as time progress in the number of peptides for all the chainlength ranges except $i = 0$ - 20. The results obtained from LC-MS method shows a decrease with time for all the peptides with mass to charge ratio intervals except m: z = 200-400. The results from LC-MS were determined from samples taken from 10 minutes. It can be argued that the rapid increase in the longer chains observed from the proposed model might have occurred before 10 minutes.

Appendix B: Simulations and regression analysis

B.1 Regression analysis

Import python functions

```
from __future__ import division
import numpy as np
from scipy.integrate import odeint
import scipy.optimize
from sklearn.metrics import r2_score
import csv
from matplotlib import pyplot as plt
%matplotlib inline
```

Define Constants and initial values

```
v = 2/1000        #l  water volume
ms = 0.2              #g
Cso1 = ms/v          #g/l substrate concentration
MM = 23644           #g/mol
Cso = Cso1 / MM      #mol/l initial substrate concentration
M = 201              #no. of peptide bonds
J = 55
```

Calculating the probability to break chain length k to form a new chain i

Endopeptidase

```
n = np.zeros((M, M))
for k in range(2, M+1):
    for i in range(1, k):
        n[i-1][k-1] = 2/(k-1)
```

Exopeptidase

```
n1 = np.zeros((M, M))
for k in range(2, M+1):
```

```
    for i in range(1, k):
        if i==1 or i==k-1:
            n1[i-1][k-1] = 1
```

Calculating probability that a polypeptide of chain length *i* and generation *j* contains a restriction site.

Endopeptidase

```
p = np.zeros((M, J+1))
p [M-1][0] = 1
for j in range(1,J+1, 1):
    for i in range(2, M, 1):
        prod = 1
        for k in range(1, i,1):
            if j == 0 or (k>=M-j):
                prod = 1
            else:
                prod = prod*((M-J-k)/ (M-j-k))
        p[i-1][j] = 1- prod
```

Exopeptidase

```
p1 = np.zeros((M, J+1))
p1 [M-1][0] = 1
for j in range(1,J+1, 1):
    for i in range(2, M, 1):
        prod = 1
        for k in range(1, i,1):
            if (k>=M-j):
                prod = 1
            else:
                prod = 0
        p1[i-1][j] = 1- prod
```

Define functions

Endopeptidase main equation

```
def Derivatives(C_ij, tspan, Kcat, Km, CE):
    C_ij_matrix = np.reshape(C_ij, (rows,col), order ='C')          #reshapes C_ij vector to a C_ij
matrix
    r_ij = ((Kcat*CE*C_ij_matrix)/ (np.sum(C_ij_matrix) + Km))*p     #reaction rate
    dC_ijdt = np.zeros((M, J+1))
    for j in range(J+1):
        dC_ijdt[:,j] =  np.dot(n,r_ij[:,j-1]) - r_ij[:,j]
        dC_ijdt[:(M-1),0] = 0            #concerntrations of intermediate chain are 0 at generation 0
    dC_ijdt1 = np.reshape(dC_ijdt,rows*col,order ='C')              #reshapes matrix to a vector
    return dC_ijdt1
```

Model literature main equation

```
def model(dh, tspan, a, b):
    dHdt = a*np.exp(-b*dh)
    return dHdt
```

Endopeptidase-Exopeptidase main function

```
def Derivatives (C_i, tspan, Kcat, Km, Kcat1, Km1, CE):


    dC_ijdt1 = np.zeros((M, J+1))
    C_ij = np.reshape(C_i, (rows,col), order ='C')
    r_ij = ((Kcat*CE*C_ij)/ (np.sum(C_ij) + Km))*p
    r_ij_exo = ((Kcat1*CE*C_ij)/ (np.sum(C_ij) + Km1))*p1
    #r_ij[0,:] = 0        #rate of reaction breaking chainlength
    for j in range(J+1):
        dC_ijdt1[:,j] =  np.dot(n1,r_ij_exo[:,j]) - r_ij_exo[:,j] + np.dot(n,r_ij[:,j-1]) - r_ij[:,j]
    dC_ijdt1[:(M-1),0] = 0
    dC_ijdt = np.reshape(dC_ijdt1,rows*col,order ='C')
    return dC_ijdt
```

Energy balance

Endopeptidase

```
def energy_balance(C_ij,tspan, Kcat, Km, Hxn, CE):
    Q = np.zeros((len(tspan), 1))
     for t in range(len(tspan)):
        C_ij1 = C_ij[t]                                #all chains in all generations at time t
        C_ij2 = np.reshape(C_ij1, (rows,col), order ='C')       #reshapes a vector to a matrix
        r_ij = ((Kcat*CE*C_ij2)/ (np.sum(C_ij2) + Km))*p        #reaction rate
        Q[t] = (Hxn*((r_ij).sum())*v) /3600                     #Energy balance
      return Q
```

Literature model

```
def energy_balance1(model, tspan, a, b, Hxn1):
    Q = np.zeros((len(tspan), 1))
    for t in range(len(tspan)):
        r = a*np.exp(-b*model[t])
        Q[t] = Hxn1*r*v/3600
      return Q
```

Endopeptidase-Exopeptidase

```
def energy_balance(C_ij,tspan, Kcat, Km, Kcat1, Km1, Hxn, CE):
    Q = np.zeros((len(tspan), 1))
    for t in range(len(tspan)):
        C_ij1 = C_ij[t]                                #all chains in all generations at time t
        C_ij2 = np.reshape(C_ij1, (rows,col), order ='C')       #reshapes a vector to a matrix
        r_ij = ((Kcat*CE*C_ij2)/ (np.sum(C_ij2) + Km))*p        #reaction rate
        r_ij_exo = ((Kcat1*CE*C_ij2)/ (np.sum(C_ij2) + Km1))*p1
        Q[t] = (Hxn*((r_ij+r_ij_exo).sum())*v)/3600                     #Energy balance
         return Q
```

Model solution

Endopeptidase

```
def model_solution(C_ijo_new, tspan, Kcat, Km, CE):
   C_ij = odeint(Derivatives,C_ijo_new,tspan, args = (Kcat, Km, CE))
    return C_ij
```

Literature model

```
def model_solution1(dho, tspan, a, b):
   H = odeint(model, dho, tspan, args = (a, b))
    return H
```

Endopeptidase-Exopeptidase

```
def model_solution(C_ijo_new, tspan, Kcat, Km, Kcat1, Km1, CE):
    C_ij = odeint(Derivatives,C_ijo_new,tspan, args = (Kcat, Km, Kcat1, Km1, CE))
    return C_ij
```

Defining degree of hydrolysis

```
def DH(C_i, tspan):
    DH1 = np.zeros((len(tspan), 1))
    B = np.arange(0, M,1)
    Co = C_i[0,:]
    sum2 = (B*Co).sum()
    for t in range(len(tspan)):
      C = C_i[t,:]
       sum1 = (B*C).sum()
       DH1[t] = 1-(sum1/sum2)
       return DH1[:,0]
def func(C_ij, tspan):
    C_i = np.zeros((len(tspan), M))
    for t in range(len(tspan)):
       for i in range(0,M,1):
          C_ij1 = C_ij[t,:]
          C_ij2 = np.reshape(C_ij1, (rows,col), order ='C')
          C_i[t][i] = C_ij2[i].sum()
    return C_i
```

Loading experimental data from excel into python

Heat flow data

```
Q_dataa = []
with open('Papain_heat_data.csv','rt') as csvfile:            #open the file
    csvreader = csv.reader(csvfile)                    #initialise the csv.reader object
    for cnt, row in enumerate(csvreader):              #read each line from the file
        if cnt==0:
            headings=row                          #store headings
        else:
            Q_dataa.append(row)                    #store the data
 Q_dataa = np.array(Q_dataa)                       #data as string format
#storing data as a matrix with floats
rows2a, cols2a = Q_dataa.shape
Data_Qa = np.zeros((rows2a, cols2a))

for i in range(rows2a):
    for k in range(cols2a):
        Data_Qa[i][k] = float(Q_dataa[i][k])
 Time_exp2a = Data_Qa[:,0]
Q_expa = Data_Qa[:,5] #idex:1,2,3,4,5,6
#Time
starttime2 = min(Time_exp2a)
endtime2 = max(Time_exp2a)
dt2 = 0.01
points2 = len(Time_exp2a)
tspan2a = np.linspace(starttime2, endtime2, points2)
t_pred2a = np.linspace(0, endtime2, points2)
Range2 = 100/1000000
```

Degree of hydrolysis data

```
DH_dataa = []
with open('Papain_DH_data.csv','rt') as csvfile:            #open the file
```

```
    csvreader = csv.reader(csvfile)                 #initialise the csv.reader object
    for cnt, row1 in enumerate(csvreader):          #read each line from the file
        if cnt==0:
            headings=row1                           #store headings
        else:
            DH_dataa.append(row1)                   #store the data

DH_dataa = np.array(DH_dataa)
rows1a, cols1a = DH_dataa.shape
Dataa = np.zeros((rows1a, cols1a))

for i in range(rows1a):
    for k in range(cols1a):
        Dataa[i][k] = float(DH_dataa[i][k])
 Time_expa = Dataa[:,0]/60
DH_expa = Dataa[:,5] #index:1,2,3,4,5,6

# time points
starttime = min(Time_expa)
endtime = max(Time_expa)          #h
points = len(Time_expa)
tspana = np.linspace(starttime, endtime, points)
tpreda = np.linspace(0, endtime, 4000)
Range = 40/100
```

Parameters optimisation

Residuals

Endopeptidase

```
def residuals(parameters):
    Kcat = parameters[0]
    Km = parameters[1]
    HappC = parameters[2]
    HappP = parameters[3]
```

```
 #Predicted Q_3%C
C_ij3C = model_solution(C_ijo_new,tspan,Kcat,Km, 0.075)
Q_pred3C = energy_balance(C_ij3C, tspan, Kcat, Km, HappC, 0.075)
#Predicted Q_4%C
C_ij4C = model_solution(C_ijo_new,tspan,Kcat,Km, 0.1)
Q_pred4C = energy_balance(C_ij4C, tspan, Kcat, Km, HappC, 0.1)
#Predicted Q_2%P
C_ij2P = model_solution(C_ijo_new,tspan,Kcat,Km, 0.05)
Q_pred2P = energy_balance(C_ij2P, tspan, Kcat, Km, HappP, 0.05)
#Predicted Q_4%P
C_ij4P = model_solution(C_ijo_new,tspan,Kcat,Km, 0.1)
Q_pred4P = energy_balance(C_ij4P, tspan, Kcat, Km, HappP, 0.1)
#Experimental data_3%C
Q_exp4 = Data[:,1]
Q_exp5 = Data[:,2]
Q_exp6 = Data[:,3]
#Experimental data_4%C
Q_exp7 = Data[:,4]
Q_exp8 = Data[:,5]
Q_exp9 = Data[:,6]
#Experimental data_2%P
Q_exp10 = Data[:,10]
Q_exp11 = Data[:,11]
Q_exp12 = Data[:,12]
#Experimental data_4%P
Q_exp16 = Data[:,13]
Q_exp17 = Data[:,14]
Q_exp18 = Data[:,15]
#Errors_Q2_3%C
error_Q4 = ((Q_exp4)-Q_pred3C[:,0])/(Range*points)
error_Q5 = ((Q_exp5)-Q_pred3C[:,0])/(Range*points)
error_Q6 = ((Q_exp6)-Q_pred3C[:,0])/(Range*points)
```

```
#Errors_Q2_4%C
error_Q7 = ((Q_exp7)-Q_pred4C[:,0])/(Range*points)
error_Q8 = ((Q_exp8)-Q_pred4C[:,0])/(Range*points)
error_Q9 = ((Q_exp9)-Q_pred4C[:,0])/(Range*points)
#Errors_Q2_2%P
error_Q10 = ((Q_exp10)-Q_pred2P[:,0])/(Range*points)
error_Q11 = ((Q_exp11)-Q_pred2P[:,0])/(Range*points)
error_Q12 = ((Q_exp12)-Q_pred2P[:,0])/(Range*points)
#Errors_Q2_4%P
error_Q16 = ((Q_exp16)-Q_pred4P[:,0])/(Range*points)
error_Q17 = ((Q_exp17)-Q_pred4P[:,0])/(Range*points)
error_Q18 = ((Q_exp18)-Q_pred4P[:,0])/(Range*points)
#Predicted DH_2%C
C_ij1_2 = model_solution(C_ijo_new,tspan2,Kcat,Km, 0.05)
C_i2 = func(C_ij1_2, tspan2)
DH_pred2 = DH(C_i2, tspan2)
#Predicted DH_3%C
C_ij1_3 = model_solution(C_ijo_new,tspan2,Kcat,Km, 0.075)
C_i3 = func(C_ij1_3, tspan2)
DH_pred3 = DH(C_i3, tspan2)
#Predicted DH_4%C
C_ij1_4 = model_solution(C_ijo_new,tspan2,Kcat,Km, 0.1)
C_i4 = func(C_ij1_4, tspan2)
DH_pred4 = DH(C_i4, tspan2)
#Experimental data_3%C
DH_exp4 = Data1[:,1]
DH_exp5 = Data1[:,2]
DH_exp6 = Data1[:,3]
#Experimental data_4%C
DH_exp7 = Data1[:,4]
DH_exp8 = Data1[:,5]
DH_exp9 = Data1[:,6]
```

```
#Experimental data_2%P
DH_exp10 = Data1[:,10]
DH_exp11 = Data1[:,11]
DH_exp12 = Data1[:,12]
#Experimental data_4%P
DH_exp16 = Data1[:,13]
DH_exp17 = Data1[:,14]
DH_exp18 = Data1[:,15]
#Errors_DH_3%C
error_DH4 = (((DH_exp4/100)-DH_pred3)/(Range1*points1))
error_DH5 = (((DH_exp5/100)-DH_pred3)/(Range1*points1))
error_DH6 = (((DH_exp6/100)-DH_pred3)/(Range1*points1))
#Errors_DH_4%C
error_DH7 = (((DH_exp7/100)-DH_pred4)/(Range1*points1))
error_DH8 = (((DH_exp8/100)-DH_pred4)/(Range1*points1))
error_DH9 = (((DH_exp9/100)-DH_pred4)/(Range1*points1))
#Errors_DH_2%P
error_DH10 = (((DH_exp10/100)-DH_pred2)/(Range1*points1))
error_DH11 = (((DH_exp11/100)-DH_pred2)/(Range1*points1))
error_DH12 = (((DH_exp12/100)-DH_pred2)/(Range1*points1))
#Errors_DH_4%P
error_DH16 = (((DH_exp16/100)-DH_pred4)/(Range1*points1))
error_DH17 = (((DH_exp17/100)-DH_pred4)/(Range1*points1))
error_DH18 = (((DH_exp18/100)-DH_pred4)/(Range1*points1))
c = np.append(error_Q4,error_Q5)
e = np.append(c, error_Q6)
f = np.append(e, error_Q7)
g = np.append(f, error_Q8)
h = np.append(g, error_Q9)
i = np.append(h, error_Q10)
j = np.append(i, error_Q11)
k = np.append(j, error_Q12)
```

```
    o = np.append(k, error_Q16)
    p = np.append(o, error_Q17)
    q = np.append(p, error_Q18)
    u = np.append(q, error_DH4)
    v = np.append(u, error_DH5)
    w = np.append(v, error_DH6)
    x = np.append(w, error_DH7)
    y = np.append(x, error_DH8)
    z = np.append(y, error_DH9)
    i1 = np.append(z, error_DH10)
    a1 = np.append(i1, error_DH11)
    b1 = np.append(a1, error_DH12)
    f1 = np.append(b1, error_DH16)
    g1 = np.append(f1, error_DH17)
    h1 = np.append(g1, error_DH18)
    return h1
```

Endopeptidase-exopeptidase

```
def residuals(parameters):
    Kcat = parameters[0]
    Km = parameters[1]
    Kcat1 = parameters[2]
    Km1 = parameters[3]
    HappC = parameters[4]
    HappP = parameters[5]

    #Predicted Q4C
    C_ij2C = model_solution(C_ijo_new,tspan2,Kcat,Km, Kcat1, Km1, 0.03)
    Q_pred2C = energy_balance(C_ij2C, tspan2, Kcat, Km, Kcat1, Km1, HappC, 0.03)
    #Predicted Q6C
    C_ij4C = model_solution(C_ijo_new,tspan2,Kcat,Km, Kcat1, Km1, 0.06)
    Q_pred4C = energy_balance(C_ij4C, tspan2, Kcat, Km, Kcat1, Km1, HappC, 0.06)
    #Predicted Q5P
```

```
C_ij3P = model_solution(C_ijo_new,tspan2,Kcat,Km, Kcat1, Km1, 0.045)
Q_pred3P = energy_balance(C_ij3P, tspan2, Kcat, Km, Kcat1, Km1, HappP, 0.045)
#Predicted Q6P
C_ij4P = model_solution(C_ijo_new,tspan2,Kcat,Km, Kcat1, Km1, 0.06)
Q_pred4P = energy_balance(C_ij4P, tspan2, Kcat, Km, Kcat1, Km1, HappP, 0.06)
#ExpQ_2
Q_exp1 = Data_Q[:,5]
Q_exp2 = Data_Q[:,6]
#ExpQ_4C
Q_exp5 = Data_Q[:,3]
Q_exp6 = Data_Q[:,4]
#ExpQ_3P
Q_exp9 = Data_Q[:,11]
Q_exp10 = Data_Q[:,12]
#ExpQ_4P
Q_exp11 = Data_Q[:,9]
Q_exp12 = Data_Q[:,10]
#errorQ_C
error_Q1 = ((Q_exp1)-Q_pred2C[:,0])/(Range2*points2)
error_Q2 = ((Q_exp2)-Q_pred2C[:,0])/(Range2*points2)
#errorQ_C
error_Q5 = ((Q_exp5)-Q_pred4C[:,0])/(Range2*points2)
error_Q6 = ((Q_exp6)-Q_pred4C[:,0])/(Range2*points2)
#errorQ_P
error_Q9 = ((Q_exp9)-Q_pred3P[:,0])/(Range2*points2)
error_Q10 = ((Q_exp10)-Q_pred3P[:,0])/(Range2*points2)
#errorQ_
error_Q11 = ((Q_exp11)-Q_pred4P[:,0])/(Range2*points2)
error_Q12 = ((Q_exp12)-Q_pred4P[:,0])/(Range2*points2)
#Pred DH
C_ij12 = model_solution(C_ijo_new,tspan,Kcat,Km, Kcat1, Km1, 0.03)
C_i2 = func(C_ij12, tspan)
```

```
DH_pred2 = DH(C_i2, tspan)                                    #Predicted DH
#Pred DH
C_ij13 = model_solution(C_ijo_new,tspan,Kcat,Km, Kcat1, Km1, 0.045)
C_i3 = func(C_ij13, tspan)
DH_pred3 = DH(C_i3, tspan)
#Pred DH
C_ij14 = model_solution(C_ijo_new,tspan,Kcat,Km, Kcat1, Km1, 0.06)
C_i4 = func(C_ij14, tspan)
DH_pred4 = DH(C_i4, tspan)
#exp DH 2%C
DH_exp1 = Data[:,7]
DH_exp2 = Data[:,8]
DH_exp3 = Data[:,9]
#exp DH 4%C
DH_exp7 = Data[:,4]
DH_exp8 = Data[:,5]
DH_exp9 = Data[:,6]
#exp DH 3%P
DH_exp13 = Data[:,10]
DH_exp14 = Data[:,11]
DH_exp15 = Data[:,12]
#exp DH 4%P
DH_exp16 = Data[:,13]
DH_exp17 = Data[:,14]
DH_exp18 = Data[:,15]
#ErrorDH
error_DH1 = ((DH_exp1/100)-DH_pred2)/(Range*points)
error_DH2 = ((DH_exp2/100)-DH_pred2)/(Range*points)
error_DH3 = ((DH_exp3/100)-DH_pred2)/(Range*points)
error_DH7 = ((DH_exp7/100)-DH_pred4)/(Range*points)
error_DH8 = ((DH_exp8/100)-DH_pred4)/(Range*points)
error_DH9 = ((DH_exp9/100)-DH_pred4)/(Range*points)
```

```
    error_DH13 = ((DH_exp13/100)-DH_pred3)/(Range*points)
    error_DH14 = ((DH_exp14/100)-DH_pred3)/(Range*points)
    error_DH15 = ((DH_exp15/100)-DH_pred3)/(Range*points)
    error_DH16 = ((DH_exp16/100)-DH_pred4)/(Range*points)
    error_DH17 = ((DH_exp17/100)-DH_pred4)/(Range*points)
    error_DH18 = ((DH_exp18/100)-DH_pred4)/(Range*points)

    a = np.append(error_Q1, error_Q2)
    d = np.append(a, error_Q5)
    e = np.append(d, error_Q6)
    h = np.append(e, error_Q9)
    i = np.append(h, error_Q10)
    j = np.append(i, error_Q11)
    k = np.append(j, error_Q12)
    l = np.append(k, error_DH1)
    m = np.append(l, error_DH2)
    n = np.append(m, error_DH3)
    r = np.append(n, error_DH7)
    s = np.append(r, error_DH8)
    t = np.append(s, error_DH9)
    x = np.append(t, error_DH13)
    y = np.append(x, error_DH14)
    z = np.append(y, error_DH15)
    a1 = np.append(z, error_DH16)
    b1 = np.append(a1, error_DH17)
    c1 = np.append(b1, error_DH18)
    return k
```

Literature model

```
def residuals1(parameters):
    a2 = parameters[0]
    a3 = parameters[1]
```

```
a4 = parameters[2]
b = parameters[3]
HappC = parameters[4]
HappP = parameters[5]
#DH Predicted
DH_pred2 = model_solution1(0, tspan2, a2, b)          #Predicted DH
DH_pred3 = model_solution1(0, tspan2, a3, b)
DH_pred4 = model_solution1(0, tspan2, a4, b)
#Experimental data_3%
DH_exp4 = Data1[:,1]
DH_exp5 = Data1[:,2]
DH_exp6 = Data1[:,3]
#Experimental data_4%
DH_exp7 = Data1[:,4]
DH_exp8 = Data1[:,5]
DH_exp9 = Data1[:,6]
#Experimental data_2%P
DH_exp10 = Data1[:,10]
DH_exp11 = Data1[:,11]
DH_exp12 = Data1[:,12]
#Experimental data_4%P
DH_exp16 = Data1[:,13]
DH_exp17 = Data1[:,14]
DH_exp18 = Data1[:,15]
#Errors_DH_3%
error_DH4 = (((DH_exp4/100)-DH_pred3)/(Range1*points1))
error_DH5 = (((DH_exp5/100)-DH_pred3)/(Range1*points1))
error_DH6 = (((DH_exp6/100)-DH_pred3)/(Range1*points1))
#Errors_DH_4%
error_DH7 = (((DH_exp7/100)-DH_pred4)/(Range1*points1))
error_DH8 = (((DH_exp8/100)-DH_pred4)/(Range1*points1))
error_DH9 = (((DH_exp9/100)-DH_pred4)/(Range1*points1))
```

```
#Errors_DH_2%
error_DH10 = (((DH_exp10/100)-DH_pred2)/(Range1*points1))
error_DH11 = (((DH_exp11/100)-DH_pred2)/(Range1*points1))
error_DH12 = (((DH_exp12/100)-DH_pred2)/(Range1*points1))
#Errors_DH_4%
error_DH16 = (((DH_exp16/100)-DH_pred3)/(Range1*points1))
error_DH17 = (((DH_exp17/100)-DH_pred3)/(Range1*points1))
error_DH18 = (((DH_exp18/100)-DH_pred3)/(Range1*points1))
#QpredC
DH_pred2C = model_solution1(0, tspan, a2, b)[:,0]
DH_pred3C = model_solution1(0, tspan, a3, b)[:,0]
DH_pred4C = model_solution1(0, tspan, a4, b)[:,0]
Q_pred3C = energy_balance1(DH_pred3C, tspan, a3, b, HappC)
Q_pred4C = energy_balance1(DH_pred4C, tspan, a4, b, HappC)
#QpredP
Q_pred2P = energy_balance1(DH_pred2C, tspan, a2, b, HappP)
Q_pred4P = energy_balance1(DH_pred4C, tspan, a4, b, HappP)
#Experimental data_3%
Q_exp4 = Data[:,1]
Q_exp5 = Data[:,2]
Q_exp6 = Data[:,3]
#Experimental data_4%
Q_exp7 = Data[:,4]
Q_exp8 = Data[:,5]
Q_exp9 = Data[:,6]
#Experimental data_2%P
Q_exp10 = Data[:,10]
Q_exp11 = Data[:,11]
Q_exp12 = Data[:,12]
#Experimental data_4%P
Q_exp16 = Data[:,13]
Q_exp17 = Data[:,14]
```

```
Q_exp18 = Data[:,15]
#error 3%C
error_Q4 = ((Q_exp4)-Q_pred3C[:,0])/(Range*points)
error_Q5 = ((Q_exp5)-Q_pred3C[:,0])/(Range*points)
error_Q6 = ((Q_exp6)-Q_pred3C[:,0])/(Range*points)
#error 4%C
error_Q7 = ((Q_exp7)-Q_pred4C[:,0])/(Range*points)
error_Q8 = ((Q_exp8)-Q_pred4C[:,0])/(Range*points)
error_Q9 = ((Q_exp9)-Q_pred4C[:,0])/(Range*points)
#Errors_2%P
error_Q10 = ((Q_exp10)-Q_pred2P[:,0])/(Range*points)
error_Q11 = ((Q_exp11)-Q_pred2P[:,0])/(Range*points)
error_Q12 = ((Q_exp12)-Q_pred2P[:,0])/(Range*points)
#Errors_4%P
error_Q16 = ((Q_exp16)-Q_pred4P[:,0])/(Range*points)
error_Q17 = ((Q_exp17)-Q_pred4P[:,0])/(Range*points)
error_Q18 = ((Q_exp18)-Q_pred4P[:,0])/(Range*points)

c = np.append( error_Q4, error_Q5)
e = np.append(c, error_Q6)
f = np.append(e, error_Q7)
g = np.append(f, error_Q8)
h = np.append(g, error_Q9)
i = np.append(h, error_Q10)
j = np.append(i, error_Q11)
k = np.append(j, error_Q12)
o = np.append(k, error_Q16)
p = np.append(o, error_Q17)
q = np.append(p, error_Q18)
u = np.append(q, error_DH4)
v = np.append(u, error_DH5)
w = np.append(v, error_DH6)
```

```
    x = np.append(w, error_DH7)
    y = np.append(x, error_DH8)
    z = np.append(y, error_DH9)
    i1 = np.append(z, error_DH10)
    a1 = np.append(i1, error_DH11)
    b1 = np.append(a1, error_DH12)
    f1 = np.append(b1, error_DH16)
    g1 = np.append(f1, error_DH17)
    h1 = np.append(g1, error_DH18)
    return h1
```

Solver

Solver functions are only defined for endopeptidase model to avoid repetition

```
def solver(residuals, guesses):
    res_1 = scipy.optimize.least_squares(residuals, guesses, bounds = ([0, 0.000, 0, 0], [10, 0.1, 6000, 6000]))#, method = 'dogbox') #Optimization
    parameters = res_1.x                                    #Optimization results
    return parameters
```

Literature model

```
#solver
def solver1(residuals1, guesses):
    res_1 = scipy.optimize.least_squares(residuals1, guesses,  bounds = ([0, 0, 0, 0, 0, 0], [2, 2, 2, 100, 5000, 5000]))
    parameters = res_1.x
    return parameters
```

Accuracy of the model fit

endopeptidase

```
def R2(parameters):
    Kcat = parameters[0]
    Km = parameters[1]
    HappC = parameters[2]
    HappP = parameters[3]
```

```
#Citric acid_Q
C_ij2C = model_solution(C_ijo_new,tspana,Kcat,Km, 0.05)
Q_pred2C = energy_balance(C_ij2C,tspana,Kcat,Km, HappC, 0.05 )
C_ij3C = model_solution(C_ijo_new,tspana,Kcat,Km, 0.075)
Q_pred3C = energy_balance(C_ij3C,tspana,Kcat,Km, HappC, 0.075 )
C_ij4C = model_solution(C_ijo_new,tspana,Kcat,Km, 0.1)
Q_pred4C = energy_balance(C_ij4C,tspana,Kcat,Km, HappC, 0.1 )
#Phosphate_Q
C_ij2P = model_solution(C_ijo_new,tspana,Kcat,Km, 0.05)
Q_pred2P = energy_balance(C_ij2P,tspana,Kcat,Km, HappP, 0.05 )
C_ij3P = model_solution(C_ijo_new,tspana,Kcat,Km, 0.075)
Q_pred3P = energy_balance(C_ij3P,tspana,Kcat,Km, HappP, 0.075 )
C_ij4P = model_solution(C_ijo_new,tspana,Kcat,Km, 0.1)
Q_pred4P = energy_balance(C_ij4P,tspana,Kcat,Km, HappP, 0.1 )
#DH
C_ij12 = model_solution(C_ijo_new,tspan2a,Kcat,Km, 0.05)
C_i2 = func(C_ij12, tspan2a)
DH_pred2 = DH(C_i2, tspan2a)
C_ij13 = model_solution(C_ijo_new,tspan2a,Kcat,Km, 0.075)
C_i3 = func(C_ij13, tspan2a)
DH_pred3 = DH(C_i3, tspan2a)
C_ij14 = model_solution(C_ijo_new,tspan2a,Kcat,Km, 0.1)
C_i4 = func(C_ij14, tspan2a)
DH_pred4 = DH(C_i4, tspan2a)
coefficient_of_determination =np.zeros((12, 1))
coefficient_of_determination[0] = r2_score(DH_exp1a2C/100, DH_pred2)
coefficient_of_determination[1] = r2_score(DH_exp1a3C/100, DH_pred3)
coefficient_of_determination[2] = r2_score(DH_exp1a4C/100, DH_pred4)
coefficient_of_determination[3] = r2_score(DH_exp1a2P/100, DH_pred2)
coefficient_of_determination[4] = r2_score(DH_exp1a3P/100, DH_pred3)
coefficient_of_determination[5] = r2_score(DH_exp1a4P/100, DH_pred4)
coefficient_of_determination[6] = r2_score(Q_expa2C, Q_pred2C[:,0])
```

```
    coefficient_of_determination[7] = r2_score(Q_expa3C, Q_pred3C[:,0])
    coefficient_of_determination[8] = r2_score(Q_expa4C, Q_pred4C[:,0])
    coefficient_of_determination[9] = r2_score(Q_expa2P, Q_pred2P[:,0])
    coefficient_of_determination[10] = r2_score(Q_expa3P, Q_pred3P[:,0])
    coefficient_of_determination[11] = r2_score(Q_expa4P, Q_pred4P[:,0])
    return coefficient_of_determination
```

endopeptidase-exopeptidase

```
def R2(parameters):
    Kcat = parameters[0]
    Km = parameters[1]
    Kcat1 = parameters[2]
    Km1 = parameters[3]
    HappC = parameters[4]
    HappP = parameters[5]

    C_ij2C = model_solution(C_ijo_new,tspan2a, Kcat, Km, Kcat1, Km1, 0.03)
    C_ij4C = model_solution(C_ijo_new,tspan2a, Kcat, Km, Kcat1, Km1, 0.06)
    Q_pred2C = energy_balance(C_ij2C,tspan2a,Kcat,Km, Kcat1, Km1, HappC, 0.03 )
    Q_pred4C= energy_balance(C_ij4C,tspan2a,Kcat,Km, Kcat1, Km1, HappC, 0.06)
    C_ij3P = model_solution(C_ijo_new,tspan2a, Kcat, Km, Kcat1, Km1, 0.045)
    C_ij4P = model_solution(C_ijo_new,tspan2a, Kcat, Km, Kcat1, Km1, 0.06)
    Q_pred3P = energy_balance(C_ij3P,tspan2a,Kcat,Km, Kcat1, Km1, HappP, 0.045 )
    Q_pred4P = energy_balance(C_ij4P,tspan2a,Kcat,Km, Kcat1, Km1, HappP, 0.06)

    C_ij1_2 = model_solution(C_ijo_new,tspana,Kcat,Km, Kcat1, Km1, 0.03)
    C_i2 = func(C_ij1_2, tspana)
    DH_pred_2 = DH(C_i2, tspana)
    C_ij1_3 = model_solution(C_ijo_new,tspana,Kcat,Km, Kcat1, Km1, 0.045)
    C_i3 = func(C_ij1_3, tspana)
    DH_pred_3 = DH(C_i3, tspana)
    C_ij1_4 = model_solution(C_ijo_new,tspana,Kcat,Km, Kcat1, Km1, 0.06)
```

```
    C_i4 = func(C_ij1_4, tspana)
    DH_pred_4 = DH(C_i4, tspana)
    coefficient_of_determination =np.zeros((8, 1))
    coefficient_of_determination[0] = r2_score(DH_expa2C/100, DH_pred_2)
    coefficient_of_determination[1] = r2_score(DH_expa4C/100, DH_pred_4)
    coefficient_of_determination[2] = r2_score(DH_expa3P/100, DH_pred_3)
    coefficient_of_determination[3] = r2_score(DH_expa4P/100, DH_pred_4)
    coefficient_of_determination[4] = r2_score(Q_expa2C, Q_pred2C[:,0])
    coefficient_of_determination[5] = r2_score(Q_expa4C, Q_pred4C[:,0])
    coefficient_of_determination[6] = r2_score(Q_expa3P, Q_pred3P[:,0])
    coefficient_of_determination[7] = r2_score(Q_expa4P, Q_pred4P[:,0])
    return coefficient_of_determination
```

Literature model

```
def R1(parameters):
    a2 = parameters[0]
    a3 = parameters[1]
    a4 = parameters[2]
    b = parameters[3]
    HappC = parameters[4]
    HappP = parameters[5]
    DH_pred2 = model_solution1(0, tspan2a, a2, b)
    DH_pred3 = model_solution1(0, tspan2a, a3, b)
    DH_pred4 = model_solution1(0, tspan2a, a4, b)
    #Q
    DH_pred2C = model_solution1(0, tspan, a2, b)
    DH_pred3C = model_solution1(0, tspan, a3, b)
    DH_pred4C = model_solution1(0, tspan, a4, b)
    Q_pred3C = energy_balance1(DH_pred3C, tspana, a3, b, HappC)
    Q_pred4C = energy_balance1(DH_pred4C, tspana, a4, b, HappC)
    Q_pred2P = energy_balance1(DH_pred2C, tspana, a2, b, HappP)
    Q_pred4P = energy_balance1(DH_pred4C, tspana, a4, b, HappP)
    coefficient_of_determination =np.zeros((8, 1))
```

```
    coefficient_of_determination[0] = r2_score(DH_exp1a3C/100, DH_pred3)
    coefficient_of_determination[1] = r2_score(DH_exp1a4C/100, DH_pred4)
    coefficient_of_determination[2] = r2_score(DH_exp1a2P/100, DH_pred2)
    coefficient_of_determination[3] = r2_score(DH_exp1a4P/100, DH_pred4)
    coefficient_of_determination[4] = r2_score(Q_expa3C, Q_pred3C[:,0])
    coefficient_of_determination[5] = r2_score(Q_expa4C, Q_pred4C[:,0])
    coefficient_of_determination[6] = r2_score(Q_expa2P, Q_pred2P[:,0])
    coefficient_of_determination[7] = r2_score(Q_expa4P, Q_pred4P[:,0])
    return coefficient_of_determination
```

Model parameters estimation

Endopeptidase

```
guesses = np.array([0.3, 0.00015, 1000, 1000])
parameters = solver(residuals,guesses)
R = R2(parameters)
R, parameters
```

Literature model

```
guesses = np.array([0.02,0.02, 0.02, 9.5, 3000, 4000])
parameters1 = solver1(residuals1,guesses)
R = R1(parameters1)
R, parameters1
```

Model predictions

endopeptidase

```
Kcat = parameters[0]
Km = parameters[1]
HappC = parameters[2]
HappP = parameters[3]
#Q_C
C_ij3C = model_solution(C_ijo_new, tspana, Kcat, Km, 0.075)
```

```
Q_pred3C = energy_balance(C_ij3C,tspana, Kcat, Km, HappC, 0.075 )
C_ij4C = model_solution(C_ijo_new, tspana, Kcat, Km, 0.1)
Q_pred4C = energy_balance(C_ij4C, tspana, Kcat, Km, HappC, 0.1)
#Q_C
C_ij2P = model_solution(C_ijo_new, tspana, Kcat, Km, 0.05)
Q_pred2P = energy_balance(C_ij2P,tspana, Kcat, Km, HappP, 0.05 )
C_ij4P = model_solution(C_ijo_new, tspana, Kcat, Km, 0.1)
Q_pred4P = energy_balance(C_ij4P, tspana, Kcat, Km, HappP, 0.1)
#DH
C_ij22 = model_solution(C_ijo_new, tspan2a, Kcat, Km, 0.05)
C_i2 = func(C_ij22, tspan2a)
DH_pred2 = DH(C_i2, tspan2a)
C_ij23 = model_solution(C_ijo_new, tspan2a, Kcat, Km, 0.075)
C_i3 = func(C_ij23, tspan2a)
DH_pred3 = DH(C_i3, tspan2a)
C_ij24 = model_solution(C_ijo_new, tspan2a, Kcat, Km, 0.1)
C_i4 = func(C_ij24, tspan2a)
DH_pred4 = DH(C_i4, tspan2a)
```

Literature model

```
a2 = parameters1[0]
a3 = parameters1[1]
a4 = parameters1[2]
b = parameters1[3]
HappC = parameters1[4]
HappP = parameters1[5]
#Q
DH_pred2_l = model_solution1(0, tspan2a, a2, b)
DH_pred3_l = model_solution1(0, tspan2a, a3, b)
DH_pred4_l = model_solution1(0, tspan2a, a4, b)
DH_pred2C_l = model_solution1(0, tspana, a2, b)
DH_pred3C_l = model_solution1(0, tspana, a3, b)
DH_pred4C_l = model_solution1(0, tspana, a4, b)
```

```
#Q
Q_pred3C_l = energy_balance1(DH_pred3C_l, tspana, a3, b, HappC)
Q_pred4C_l = energy_balance1(DH_pred4C_l, tspana, a4, b, HappC)

Q_pred2P_l = energy_balance1(DH_pred2C_l, tspana, a2, b, HappP)
Q_pred4P_l = energy_balance1(DH_pred4C_l, tspana, a4, b, HappP)
```

Plotting degree of hydrolysis and heat flow graphs

Endopeptidase-Literature model

Heat flow

```
plt.figure(1)
plt.errorbar(Time_expa, Q_expa3C*1000000, fmt = 'bo', label = 'Experimental')
plt.plot(Time_expa, Q_exp4*1000000,'b*')
plt.plot(Time_expa, Q_exp5*1000000,'b*')
plt.plot(Time_expa, Q_exp6*1000000,'b*')
plt.plot(tspana, Q_pred3C*1000000,'r-', label = 'Current model')
plt.plot(tspana, Q_pred3C_l*1000000,'r--', label = '(Marquez-Moreno & Fernandez-Caudrado, 1993)')
plt.legend(loc=1)
plt.xlabel('Time (h)')
plt.ylabel('Heat flow ')
plt.figure(2)
plt.errorbar(Time_expa, Q_expa4C*1000000, fmt = 'bo', label = 'Experimental')
plt.plot(Time_expa, Q_exp7*1000000,'b*')
plt.plot(Time_expa, Q_exp8*1000000,'b*')
plt.plot(Time_expa, Q_exp9*1000000,'b*')
plt.plot(tspana, Q_pred4C*1000000,'r-', label = 'Current model')
plt.plot(tspana, Q_pred4C_l*1000000,'r--', label = '(Marquez-Moreno & Fernandez-Caudrado, 1993)')
plt.legend(loc=1)
plt.xlabel('Time (h)')
plt.ylabel('Heat flow ')
```

```
plt.figure(3)
plt.errorbar(Time_expa, Q_expa2P*1000000, fmt = 'bo', label = 'Experimental')#yerr = Dataa[:,9]*1000000, fmt = 'o', label = 'Experimental')
plt.plot(Time_expa, Q_exp10*1000000,'b*')
plt.plot(Time_expa, Q_exp11*1000000,'b*')
plt.plot(Time_expa, Q_exp12*1000000,'b*')
plt.plot(tspana, Q_pred2P*1000000,'r-', label = 'Current model')
plt.plot(tspana, Q_pred2P_l*1000000,'r--', label = '(Marquez-Moreno & Fernandez-Caudrado, 1993)')
plt.legend(loc=1)
plt.xlabel('Time (h)')
plt.ylabel('Heat flow ')
plt.figure(4)
plt.errorbar(Time_expa, Q_expa4P*1000000, fmt = 'bo', label = 'Experimental')
plt.plot(Time_expa, Q_exp16*1000000,'b*')
plt.plot(Time_expa, Q_exp17*1000000,'b*')
plt.plot(Time_expa, Q_exp18*1000000,'b*')
plt.plot(tspana, Q_pred4P*1000000,'r-', label = 'Current model')
plt.plot(tspana, Q_pred4P_l*1000000,'r--', label = '(Marquez-Moreno & Fernandez-Caudrado, 1993)')
plt.legend(loc=1)
plt.xlabel('Time (h)')
plt.ylabel('Heat flow ')
```

Degree of hydrolysis

```
plt.figure(1)
plt.errorbar(tspan2a, DH_exp1a3C/100, yerr = Data1a[:,7]/100, fmt = 'o', label = 'Experimental')
plt.plot(tspan2a, DH_pred3, 'r-', label = 'Current model')
plt.plot(tspan2a, DH_pred3_l,'r--', label = '(Marquez-Moreno & Fernandez-Caudrado, 1993)')
plt.legend(loc=1)
plt.xlabel('Time (h)')
plt.ylabel('Degree of hydrolysis')
```

```
plt.figure(2)
plt.errorbar(tspan2a, DH_exp1a4C/100, yerr = Data1a[:,8]/100, fmt = 'o', label = 'Experimental')
plt.plot(tspan2a, DH_pred4, 'r-', label = 'Current model')
plt.plot(tspan2a, DH_pred4_l,'r--', label = '(Marquez-Moreno & Fernandez-Caudrado, 1993)')
plt.legend(loc=1)
plt.xlabel('Time (h)')
plt.ylabel('Degree of hydrolysis')
plt.figure(3)
plt.errorbar(tspan2a, DH_exp1a2P/100, yerr = Data1a[:,12]/100, fmt = 'o', label = 'Experimental')
plt.plot(tspan2a, DH_pred2, 'r-', label = 'Current model')
plt.plot(tspan2a, DH_pred2_l,'r--', label = '(Marquez-Moreno & Fernandez-Caudrado, 1993)')
plt.legend(loc=1)
plt.xlabel('Time (h)')
plt.ylabel('Degree of hydrolysis')
plt.figure(4)
plt.errorbar(tspan2a, DH_exp1a4P/100, yerr = Data1a[:,11]/100, fmt = 'o', label = 'Experimental')
plt.plot(tspan2a, DH_pred4, 'r-', label = 'Current model')
plt.plot(tspan2a, DH_pred4_l,'r--', label = '(Marquez-Moreno & Fernandez-Caudrado, 1993)')
plt.legend(loc=1)
plt.xlabel('Time (h)')
plt.ylabel('Degree of hydrolysis')
```

B.2 Model sensitivity

B.2.1 Global sensitivity

Endopeptidase model

Heat flow

```
def Heat(param_values):
    Y = np.zeros([param_values.shape[0]])
    for i, X in enumerate(param_values):
```

```
        Kcat = X[0]
        Km = X[1]
        HappC = X[2]
        C_ij = model_solution(C_ijo_new, tspan, Kcat, Km, 0.1)
        Q_pred = energy_balance(C_ij,tspan, Kcat, Km, HappC, 0.1)*1000000
        points = len(Q_pred)
        Y[i] = Q_pred[points-1]
    return Y
```

Degree of hydrolysis

```
def dh(param_values):
   Y = np.zeros([param_values.shape[0]])
    for i, X in enumerate(param_values):
        Kcat = X[0]
        Km = X[1]
        HappC = X[2]
       C_ij2 = model_solution(C_ijo_new, tspan1, Kcat, Km, 0.1)
        C_i = func(C_ij2, tspan1)
        DH_pred = DH(C_i, tspan1)

        points = len(DH_pred)
        Y[i] = DH_pred[points-1]
    return Y
```

Import functions

```
from SALib.sample import saltelli
from SALib.analyze import sobol
problem = {'num_vars': 3,
          'names': ['Kcat', 'Km', 'Happ'],
          'bounds': [[0, 10],
                     [0, 5],
                     [0, 10000]]
         }
# Generate samples
```

```
param_values = saltelli.sample(problem, 200,calc_second_order=True)
# Run model heat flow
Y_Q = Heat(param_values)
#Run model DH
Y_DH = dh(param_values)
# Perform analysis
Si = sobol.analyze(problem, Y_Q, calc_second_order=True)
#Total sensitivity
Si['ST']
```

B.2.2 Local sensitivity analysis

```
Parameter_change = np.array([-75, -50, -25, 0, 25, 50, 75])
factor = np.array([0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75])
Q_val = np.zeros((7, 3))
DH_val = np.zeros((7, 3))
sol = np.array([0.319, 0.00015, HappC])
parameters = sol
```

Changing Kcat

```
for i in range(7):
    Kcat = parameters[0]*factor[i]
    Km = parameters[1]
    HappC = parameters[2]
    C_ij = model_solution(C_ijo_new, tspan, Kcat, Km, 0.1)
    Q_pred = energy_balance(C_ij,tspan, Kcat, Km, HappC, 0.1)
    Q_val[i][0] = Q_pred[points-1]
    C_i = func(C_ij, tspan)
    DH_pred = DH(C_i, tspan)
    DH_val[i][0] = DH_pred[points-1]
```

Changing Km

```
for i in range(7):
    Kcat =parameters[0]
    Km = parameters[1]* factor[i]
    HappC = parameters[2]
```

```
    C_ij = model_solution(C_ijo_new, tspan, Kcat, Km, 0.1)
    Q_pred = energy_balance(C_ij,tspan, Kcat, Km, HappC, 0.1 )
    Q_val[i][1] = Q_pred[points-1]

    C_i = func(C_ij, tspan)
    DH_pred = DH(C_i, tspan)
    DH_val[i][1] = DH_pred[points-1]
```

Changing Happ

```
for i in range(7):
    Kcat = parameters[0]
    Km = parameters[1]
    HappC = parameters[2]*factor[i]
    C_ij = model_solution(C_ijo_new, tspan, Kcat, Km, 0.1)
    Q_pred = energy_balance(C_ij,tspan, Kcat, Km, HappC, 0.1 )
    Q_val[i][2] = Q_pred[points-1]

    C_i = func(C_ij, tspan)
    DH_pred = DH(C_i, tspan)
    DH_val[i][2] = DH_pred[points-1]
Q_values = Q_val*1000000
DH_values = DH_val*100
plt.figure(1)
plt.plot(Parameter_change, Q_values[:,0] ,'g',label ='Kcat')
plt.plot(Parameter_change, Q_values[:,1] ,'r',label ='Km')
plt.plot(Parameter_change, Q_values[:,2] ,'c',label ='Happ')
plt.legend(loc=4)
plt.xlabel('Parameter change (%)')
plt.ylabel('Heat flow')
plt.figure(2)
plt.plot(Parameter_change, DH_values[:,0] ,'g',label ='Kcat')
plt.plot(Parameter_change, DH_values[:,1] ,'r',label ='Km')
plt.plot(Parameter_change, DH_values[:,2] ,'c',label ='Happ')
```

```
plt.legend(loc=4)
plt.xlabel('Parameter change (%)')
plt.ylabel('DH (%)')
```

B.3 Characterisation of peptides

Endopeptidase-exopeptidase

```
C_ij1_4 = model_solution(C_ijo_new, tspana, Kcat, Km, Kcat1, Km1, 0.1)
C_i4 = func(C_ij1_4, tspana)


def peptides(C_i, y, chain, tspan):
    x = np.zeros((len(tspan), 1))
    for t in range(len(tspan)):
        x[t] = (C_i[t][y:chain]).sum()/(C_i[t,:M]).sum()
    return x
C_20 = peptides(C_i4, 0, 19, tspana)          #i = 20
C_40 = peptides(C_i4, 20, 39, tspana)         #i = 40
C_60 = peptides(C_i4, 40, 59, tspana)         #i = 60
C_100 = peptides(C_i4, 60, 99, tspana)        #i = 100
C_160 = peptides(C_i4, 100, 159, tspana)      #i = 160
C_180 = peptides(C_i4, 160, M-2, tspana)      #i = 180
```