

UNIVERSITI PUTRA MALAYSIA

A STRUCTURAL STUDY ON THE SPECIFICITY OF F1 PROTEASE

AZIRA MUHAMAD.

FBSB 2005 34



A STRUCTURAL STUDY ON THE SPECIFICITY OF F1 PROTEASE



Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

December 2005



Abstract of the thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Master of Science

A STRUCTURAL STUDY ON THE SPECIFICITY OF F1 PROTEASE

By

AZIRA MUHAMAD

December 2005

Chairman : Assoc. Professor Raja Noor Zaliha Raja Abd. Rahman, PhD

Faculty : Biotechnology and Biomolecular Sciences

Specificity studies of a thermostable alkaline serine protease F1 with its substrates were carried out through computational docking method. Structures of a series of synthetic peptide substrates were docked to the active site of the homology modelled F1 protease using AutoDock 3.0.5. The resulting clusters of the substrates that were docked were analysed by inspecting the energetic results and the orientation of each cluster to determine the arrangement of productive binding. The amino acids of the binding site that participated in the hydrophobic and hydrogen-bond interactions were also determined. Docking results showed that all substrates tested bound near the catalytic residues with SucAAPFpNA, the biggest substrate, showing the most negative docked energy value ($E_{docked} = -18.75$ kcal/mol). Smaller substrates such as GpNA and AApNA showed higher docked energy ($E_{docked} = -7.77$ kcal/mol and -8.77 kcal/mol, respectively). The



best docked structure of each substrate was determined from the clusters. It was found that most of the lowest E_{docked} conformations display the best docked orientations with respect to the least distance calculated between the carbonyl carbon of the substrate P1 residue and γ -oxygen of the Ser226 catalytic triad. From the results, it also demonstrated that S1, S2 and S4 subsites of the enzyme play a critical role in determining the substrate specificity of F1 protease from the point of view that bigger-sized substrates such as SucAAPFpNA and SucAAPLpNA showed more favourable E_{docked} . This work also support the hypothesis that the catalytic serine and histidine residues were essential in catalysis as well as in stabilizing the enzymesubstrate complex for binding.

Validation of computational study was carried out through biochemical assay. It was found that SucAAPFpNA was the most preferred substrate for the enzyme with specific activity of 3.079 U/mg followed by SucAAPLpNA at 1.016 U/mg. SucAAPFpNA was also observed to show the highest binding affinity towards the protease ($K_m = 1.26$ mM) and the highest catalytic ratio (1.226 min⁻¹.mM⁻¹) compared to the other substrates tested. Similar to computational observations, smaller peptides showed lower specific activity and binding affinity towards the protease. Rank-order of the substrates tested for the docking and experimental methods were found to be similar for the top two substrates, with lesser agreement for the other substrates.



KAJIAN STRUKTUR SPESIFIKASI F1 PROTEASE

Oleh

AZIRA MUHAMAD

Disember 2005

Pengerusi : Profesor Madya Raja Noor Zaliha Raja Abd. Rahman, PhD

Fakulti : Bioteknologi dan Sains Biomolekul

Suatu penyelidikan spesifikasi terhadap serine protease F1 yang stabil suhu tinggi dan bersifat alkali bersama substrat telah dijalankan melalui kaedah percamtuman pengkomputeran. Struktur-struktur peptida sintetik telah dicantum ke tapak aktif F1 protease yang telah dimodelkan secara homologi dengan menggunakan AutoDock 3.0.5. Kelompok yang telah dihasilkan oleh substrat-substrat yang telah dicantum, telah dianalisa keputusan tenaga dan orientasi setiap kelompok untuk menentukan susunan cantuman yang produktif. Asid amino di tapak pengikat yang menyertai interaksi hidrofobik dan ikatan hidrogen juga telah ditentukan. Percantuman substrat-substrat pada enzim ini menunjukkan yang kesemua substrat yang telah diuji mengikat berhampiran residu katalitik di mana SucAAPFpNA, iaitu substrat yang paling besar, menunjukkan nilai E_{docked} yang paling rendah, -18.75 kcal/mol. Substrat-substrat yang lebih kecil seperti GpNA dan AApNA menunjukkan nilai E_{docked} yang lebih tinggi, masing-masing pada -7.77



kcal/mol dan -8.77kcal/mol. Struktur yang paling sesuai telah ditentukan daripada setiap kelompok. Didapati bahawa kebanyakkan konformasi E_{docked} yang paling rendah mempunyai orientasi yang paling sesuai berpunca dari jarak di antara karbon karbonil substrat residu P1 dan γ -oksigen Ser226 triad katalitik yang paling rendah. Cantuman substrat-substrat itu juga menunujukkan subtapak S1, S2 dan S4 enzim memainkan peranan yang kritikal dalam penentuan spesifikasi substrat bagi F1 protease kerana didapati bahawa subtrat-substrat yang lebih besar seperti SucAAPFpNA dan SucAAPLpNA menunjukkan nilai E_{docked} yang lebih rendah. Kajian ini mengesahkan bahawa residu-residu katalitik seperti serine dan histidine adalah penting dalam katalisis dan dalam pengstabilan kompleks enzim-substrat untuk percantuman.

Kajian kesahihan telah dijalankan dengan melakukan asai biokimia. Didapati bahawa SucAAPFpNA adalah substrat yang paling digemari oleh enzim ini dengan aktiviti spesifik, 3.079 U/mg, diikuti SucAAPLpNA pada 1.016 U/mg. SucAAPFpNA juga didapati mempunyai afiniti pengikatan yang paling tinggi terhadap protease (K_m=1.26mM) dan ratio katalitik paling tinggi (1.226 min⁻¹.mM⁻¹) berbanding dengan substrat-substrat lain yang diuji. Peptida yang lebih kecil didapati mempunyai spesifik aktiviti dan afiniti pengikatan yang rendah terhadap F1 protease, seperti yang telah diperolehi dari analisis pengkomputeran. Persamaan yang didapati daripada urutan substrat-substrat yang diperolehi daripada percantuman pengkomputeran



V

dan kaedah eksperimental adalah pada tempat kedua teratas, substratsubstrat lain kurang menunjukkan kaitan.





ACKNOWLEDGMENT

I wish to express my deepest gratitude to my supervisor, Assoc. Prof. Dr. Raja Noor Zaliha Raja Abd. Rahman for her constant guidance and encouragement throughout the duration of this study. My sincere appreciation is extended to my co-supervisors, Prof. Dr. Abu Bakar Salleh, Prof. Dr. Mahiran Basri and Dr. Habibah A. Wahab. Their critical comments and support have been an invaluable source of motivation. My gratitude is also extended to Assoc. Prof. Dr. Mohd. Basyaruddin Abd. Rahman for his helpful comments and guidance.

I would like to thank all my lab mates, Ferrol, Kak Ain, Kak Ina, Kak Naz, Leow, Ghanie, Rofandi, Kak Lia, Kok Whye, Brother Mohd, Aiman, Sha, Shuk, Sue, Yob, Ada, Chee Fah and Tengku for their friendship and support during my stay at the lab. You guys have been really helpful. My sincere appreciation also goes especially to Bimo for his invaluable encouragement and comments. You have been a real source of inspiration. Thank you for being so patient and thank you so much for the long hours you've spent for me.

Last but not least, my heartfelt gratitude goes to my family, Bapak, Mak, Abang, Kak Yani and Alen. Thank you for your patience, encouragement and support.



TABLE OF CONTENTS

ABSTRACT ABSTRAK

APPROVAL

 $\left[\mathbf{C} \right]$

ACKNOWLEDGMENTS

DEC LIST LIST LIST	CLARA COFTA COFFI	FION ABLES GURES BBREVIAT	IONS	x xiii xiv xix
CHA	APTER			
1	INT	RODUCTIC		1
2	LITE	RATURE R	EVIEW	4
	2.1	Protease		4
		2.1.1 Sul	otilisin-like serine protease	5
		2.1.2 Str	uctural properties of serine protease	8
		Sul	ostrate Specificity	11
		The	eoretical Approaches	14
		2.1.3 The stee	ermostable protease from <i>Bacillus</i> arothermophilus F1	16
		2.1.4 Pu	rification of proteases	17
		2.1.5 Kin	netic studies of proteases	18
		2.1.6 Ap	plications of proteases	20
	2.2	Molecular	Modelling	22
		2.2.1 Do	cking	24
		Do	cking Programs	25
		Ар	plications of docking	27
3	MA	TERIALS A	ND METHODS	29
	3.1	Chemical	S	29
	3.2	Compute	r Hardwares and Softwares	30
	3.3	Compute	r Simulations	31
		3.3.1 Mo	odelling of enzyme	31
		3.3.2 Mo	odelling of Peptides	31
		3.3.3 Do	cking Calculations	37
		3.3.4 Ar	alysis	40
	3.4	Experime	ntal procedures	40
		3.4.1 Sci	eening for Positive Colonies	40
		3.4.2 As	say of Protease Activity	40
		3.4.4 Pu	rification of Bacillus stearothermophilus	
		pro	otease F1	41
		Cr	ude Enzyme Preparation	41



ii

iv

vii

viii

			Heat Treatment	41
			Ultrafiltration	42
		3.4.5	Substrate Specificity	42
			Specific Activity	42
			Kinetic Study	43
4	RESU	JLTS		44
	4.1	Comp	No. 1. 11: Contractions	44
		4.1.1	Modelling of enzyme	44
		4.1.2	Flexible docking of substrates	44
			SUCAPFPNA	4 8
			SucAAPLpNA	51
			VLKpNA	51
			AAFpNA	54
			GltFpNA	59
			AceFpNA	59
			FpNA	62
			AApNA	67
			GpNA	70
	4.2	Expe	rimental Procedures	73
		4.2.1	Purification of Bacillus stearothermophilus protease F1	73
		4.2.2	Substrate Specificity	73
			Specific activity	73
			Kinetic study	75
5	DICC			01
3		.05510	JN	ð1
	5.1	Com	Flacible leading of substantian	81 01
	5.0	5.1.1	Flexible docking of substrates	81
	5.2	Expe	Imental Procedures	85
		5.2.1	Purification of Bacillus stearothermophilus protease F1	85
		5.2.2	Substrate Specificity	86
	5.3	Expe	rimental Validation of Docking Results	90
6	CON	ICLUSI	ON AND RECOMMENDATION	97
				•
REFE	RENC	CES		99
APPE	ENDIC	CES		107
BIOI	DATA	OF TH	E AUTHOR	147



LIST OF TABLES

Table		Page
1	Substrates chosen for docking. The curly arrows indicate the rotatable bonds defined in the flexible docking. The bold residue is the P1 amino acid	34
2	Docking parameters	39
3	Results obtained for docking of various substrates to F1 protease. The numbers in brackets are the cluster rank of the conformation chosen	47
4	Purification table of recombinant F1 protease	74
5	Substrate specificity of F1 protease with peptide substrates	74
6	K_{m_r} , V_{max_r} , k_{cat} and specificity of F1 protease with several substrates	s 80
7	The ΔG_{ES} calculated for the three substrates	80



LIST OF FIGURES

Figure		Page
1	Serine protease mechanism (Voet and Voet, 1995)	7
2	Schematic representation of substrates/inhibitor binding to a subtilisin-like serine protease. Nomenclature P4- P2' and S4 – S2' according to Schechter and Berger (1967)	10
3	Three-dimensional structure of the predicted F1 protease. The residues rendered as sticks represent the catalytic triads	32
4	Active site of F1 protease. The enzyme is represented as atomic surfaces. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.	45
5	Orientation of the lowest E_{docked} of SucAAPFpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.	49
6	Yellow lines indicate the distance of 3.88Å measured between the carbonyl carbon of the SucAAPFpNA P1 residue and γ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type	49
7	Hydrogen-bond and hydrophobic interactions of the lowest E_{docked} of SucAAPFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds	50
8	Orientation of the lowest E_{docked} of SucAAPLpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad	52



9	Yellow lines indicate the distance of 3.71Å measured between the carbonyl carbon of the SucAAPLpNA P1 residue and y-oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured	
	according to the atom type	52
10	Hydrogen-bond and hydrophobic interactions of the lowest E_{docked} of SucAAPLpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds	53
11	Orientation of the lowest E_{docked} of VLKpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad	55
12	Yellow lines indicate the distance of 5.11Å measured between the carbonyl carbon of the VLKpNA P1 residue and γ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured	
13	according to the atom type Hydrogen-bond and hydrophobic interactions of the lowest <i>E</i> _{docked} of VLKpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate	55
14	the length of the hydrogen bonds Orientation of the lowest E_{docked} of AAFpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents	56
	the catalytic triad	57
15	Yellow lines indicate the distance of 3.67Å measured between the carbonyl carbon of the AAFpNA P1 residue and γ-oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type	57
16	Hydrogen-bond and hydrophobic interactions of the lowest <i>E</i> _{docked} of AAFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts	



- 17 Orientation of the lowest *E*_{docked} of GltFpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.
- 18 Yellow lines indicate the distance of 3.86\AA measured between the carbonyl carbon of the GltFpNA P1 residue and γ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type
- 19 Hydrogen-bond and hydrophobic interactions of the lowest E_{docked} of GltFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds.
- 20 Interactions of AceFpNA with the active site of F1 protease (a) Orientation of the lowest E_{docked} of AceFpNA docked to the active site of F1 protease. (b) Hydrogen-bond and hydrophobic interactions of the lowest E_{docked} of AceFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds.
- 21 Interactions of AceFpNA with the active site of F1 protease (a) Orientation of the best docked conformation of AceFpNA docked to the active site of F1 protease. (b) Hydrogen-bond and hydrophobic interactions of the best docked of AceFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds. 6-
- 22 Orientation of the lowest E_{docked} of FpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.



58

60

61

63

64

65



23	Yellow lines indicate the distance of 5.12Å measured between the carbonyl carbon of the FpNA P1 residue and γ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type	65
24	Hydrogen-bond and hydrophobic interactions of the lowest E_{docked} of FpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds.	66
25	Orientation of the lowest E_{docked} of AApNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.	68
26	Yellow lines indicate the distance of 4.01Å measured between the carbonyl carbon of the AApNA P1 residue and γ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type	68
27	Hydrogen-bond and hydrophobic interactions of the lowest E_{docked} of AApNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds.	69
28	Interactions of GpNA with the active site of F1 protease (a) Orientation of the lowest E_{docked} of GpNA docked to the active site of F1 protease. (b) Hydrogen-bond and hydrophobic interactions of the lowest E_{docked} of GpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds.	71
29	Interactions of GpNA with the active site of F1 protease (a) Orientation of the best docked structure of GpNA to the active site of F1 protease. (b) Hydrogen-bond and hydrophobic interactions of the best docked structure of GpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length	





of the hydrogen bonds	72
30 Time-course measurement of enzyme activity	76
31 Michaelis-Menten plot for the velocity data of SucAAPFpNA	76
32 Michaelis-Menten plot for the velocity data of SucAAPLpNA	77
33 Michaelis-Menten plot for the velocity data of AApNA	77



 \bigcirc



LIST OF ABBREVIATIONS

$\Delta G_{ extbf{binding}}$	Free energy of binding
Å	Angstrom
dH ₂ O	distilled water
$E_{ m docked}$	Docked energy
h	hour
HF	Hartree-Fock
К	Kelvin
kcal/mol	kilocalorie per mol
М	Molar
mg	milligram
ml	milliliter
mM	milliMolar
min	minute
RESP	restrained electrostatic potential
RMSd	root mean square deviation (s)
μg	microgram
μΙ	microliter
v/v	volume per volume
w/v	weight per volume



(()

CHAPTER 1

INTRODUCTION

Proteases are enzymes that cleave peptide bonds at points within the protein and remove amino acids sequentially from either N or C-terminus. There are four mechanistic classes of proteases, they are serine proteases, cysteine proteases, aspartic proteases and metallo proteases. Two distinct families can be observed for serine proteases, which are the chymotrypsin family and the subtilisin family which include the bacterial enzymes such as subtilisin (Rao *et al.*, 1998). An example of subtilisin serine protease is *Bacillus stearothermophilus* protease F1. *B. stearothermophilus* was isolated from decomposed oil palm branches and its protease showing the characteristics of an extremely thermostable alkaline protease. *B. stearothermophilus* was shown to have a high degree of thermostability, able to grow up to 80° C with a pH range from 8.0 to 10.0. It has an optimum growth rate at 70° C and pH 9 (Rahman *et al.*, 1994).

Proteases differ in their properties such as substrate specificity, active site and catalytic mechanism. Structure-based mutational analysis of serine proteases specificity has produced a large database of information useful in addressing biological function and in establishing a basis for targeted design efforts (Perona and Craik, 1995). Despite extensive research on proteases, relatively little is known about the factors that control their



specificities. These questions that are difficult to answer experimentally, might be resolved theoretically.

Molecular modelling methods have been used in many ways to address problems in structural biology. The overall aim of modelling methods will often be to try to relate biological activity to structure (Forster, 2002). Molecular modelling is a general term that covers a wide range of molecular graphics and computational chemistry techniques used to display, build, manipulate, simulate and analyze molecular structure and to calculate properties of these structures. Through modelling, the understanding of properties, three-dimensional conformations, and enzymatic mechanism can also be improved.

One of the most useful areas of application of molecular modelling is the approach of fitting together, or docking, a protein to a second molecule. Molecular docking is important in understanding possible interactions between a protein and a ligand in the formation of a biologically important protein-ligand complex. It is also a suitable tool to estimate interaction energy of the binding step by means of rigid energy minimization methods (Teodoro *et al.*, 2001).

There are many docking software available and most of these programs are free for public. One of the widely used programs is AutoDock.



AutoDock is an automated docking program of flexible ligands to receptors. It was developed by Olson's group in 1990 (Goodsell and Olson, 1990). AutoDock uses free energy of the docking molecules using threedimension potential-grids. The search algorithms implemented include Simulated Annealing (SA), Genetic Algorithm (GA) and Lamarckian GA (LGA) with GA + local search (LS) hybrid. The search and score method in AutoDock involves exploration of the configuration spaces available for interaction between ligand and receptor and it evaluates and ranks configurations using a scoring system which is the binding energy (Morris *et al.*, 1998).

In order to further understand the molecular basis of substrate specificity of *B. stearothermophilus* F1 protease, the following objectives were implemented in view of this research:

- To find the best substrate based on binding energies obtained from docking procedure
- 2. To study the interaction of F1 protease with its substrates
- 3. To validate results obtained through computational studies by experimental methods



CHAPTER 2

LITERATURE REVIEW

2.1 Proteases

Proteases comprise two groups of enzymes, the endopeptidases and the exopeptidases. The exopeptidases act only near the ends of polypeptide chains at the N or C terminus while the endopeptidases act preferentially in the inner regions of peptide chains away from the N and C termini. These proteases are classified according to their catalytic mechanisms. There are four mechanistic classes recognized by International Union of Biochemistry and Molecular Biology. They are serine proteases, cysteine proteases, aspartic proteases and metallo proteases (Rao *et al.*, 1998).

Serine proteases have been grouped into six distinct families of which the chymotrypsin-like clan and the subtilisin-like clan are the two largest family (Siezen and Leunissen, 1997). They have the same active site geometry and the catalysis proceeds via the same mechanism but the general three-dimensional structure is different in the two families. Three residues (histidine, aspartic acid and serine) that form the catalytic triad are essential in the catalytic process of the enzyme.



Plant proteases such as papain and bromelain are included in the cysteine proteases family. The essential cysteine and histidine residues play the same role as serine and histidine residues in serine proteases during the catalytic process. Most of aspartic proteases belong to the pepsin family. The pepsin family includes digestive enzymes such as pepsin and chymosin, fungal proteases and processing enzymes such as renin. A second family comprises viral proteases such as the proteases from the HIV also called retropepsin. Metallo proteases are found in bacteria, fungi as well as in higher organisms. They differ widely in their sequences but the great majority of enzymes contain a zinc atom which is catalytically active. In some cases, zinc may be replaced by another metal such as cobalt or nickel without loss of activity (Rao *et al.*, 1998).

2.1.1 Subtilisin-like serine protease

Subtilisin-like serine proteases (EC 3.4.21.62), termed "subtilases" occur in archaea, bacteria, fungi, yeasts and higher eukaryotes. Based on sequence homology, a subdivision into six families is proposed (Siezen and Leunissen, 1997). They are subtilisin family, thermitase, proteinase K, lantibiotic peptidase, kexin and pyrolysin family. Subtilisins are enzymes that are secreted by members of the genus *Bacillus* with subgroups of true subtilisins, high-alkaline proteases and intracellular proteases. The reactions of subtilisin are hydrolysis of proteins with broad specificity for peptide bonds, and showed a preference for a large uncharged residue in

