Studies of enzymes from two protease families:
Tissue Kallikreins, ADAMs and MMPs.

A thesis submitted to the Queensland University of Technology for the degree of Master of Applied Science by Research.

Sergio Manzetti, B. Sc. (Chemistry)
Cluster for Molecular Biotechnology
Science Research Centre

School of Life Sciences, Queensland University of Technology, Brisbane, Australia.
June 2005

ABSTRACT

The human kallikrein family is a family of proteolytic enzymes, classified as serine proteases, that derive from chromosome 19, locus 13.3-13.4. These enzymes are widespread in pathophysiological processes such as cancer and neurodegenerative diseases; hence studies of catalytic sites and inhibitors are important in relation to the longer term of design of therapeutic drugs. One member of the family, human kallikrein 4 (hK4) which is thought to carry out crucial functions in the prostate, was expressed in this study as a secreted protein in a baculovirus expression system, bearing a His-tag and V5-epitope that were used for purification and detection respectively. Its mass was estimated to be 35kDa, ~2kDa less than the equivalent product expressed in monkey kidney cells. The protein was purified to 50-90% purity with a yield of 0.93mg/L-4.8mg/L based on methods derived from computational prediction of its properties, such as pI. Computational analysis was extended by applying high-performing computing techniques, such as molecular dynamics, and flexible ligand docking, to predict antigenic regions, the likely substrate specificity and putative inhibitors. These results show that hK4 has a loop, between Leu83-Ser94 that shows promise as a specific segment that can be exploited for generation of antibodies. Preferred substrates were also predicted to bear hydrophobic residues at the P'-region of the scissile bond and amphiphilic residues at the P-region. At the S-region, hK4 potentially involves its unique PLYH-motif in recognizing the P4/P5 position from the substrate. Flexible ligand-docking studies indicate that hK4 can be inhibited by inhibitors that carry a modified bulky hydrophobic sidechain with a guanidinium group at the P1-position and its own putative autoactivation region residues at the P2, P1' and P2' position.

The computational study was extended to other members of the kallikrein family, predicting distinctions between these that could be used for future studies. These results show that 8 of the fifteen kallikrein members are very homologous in terms of specificity bearing typical trypsin-like activity and specificity, except for hK2, hK3, hK4, hK5, hK7, hK9, hK15 that retain certain distinct signatures in the binding pocket in terms of secondary specificity.

The principles of substrate-specificity analysis that were developed were further applied on three metzincins, MMP-3, ADAM-9 and ADAM-10. These three enzymes are metalloproteases, which are involved in tissue remodeling, intracellular signalling and cell-to-cell mediation. The substrate-specificity analysis was carried out on all three metzincins using the structure of a crystallized complex of the MMP-3 enzyme with the TIMP-1 natural inhibitor as template. In this specific enzyme-substrate complex, the challenge was to model and suggest a possible orientation of the P-region, which is not known. The interactions on the P/S-region are therefore unclear and

need to be clarified. In order to suggest the arrangement of the enzyme-substrate complex and the undefined S-subsites, four new residues were added in an extended beta-sheet conformation to the P1' residue (derived originally from the TIMP-1 inhibitor) to create a full-length modeled substrate spanning P4'-P4. This new modeled region, in particular, was bound through backbone H-bonds with the enzyme at position 169 (MMP nomenclature) suggesting a new crucial residue for substrate binding, and satisfied steric and chemical restraints in the S'-region of the enzyme. This modeling approach also indicated a putative presence of an S2/S3-pocket on these metzincins which is composed of different residues for MMP-3, ADAM-9 and ADAM-10, and which could prove useful for future drug design projects. Furthermore, the data argue against the involvement of a polarizable water molecule in catalysis, a mechanism that has been postulated by various groups. A new catalytic mechanism is suggested to involve an oxyanion anhydride transition state.

This study is a demonstration of the power of combining bioinformatics with wet-lab biochemistry.

TABLE OF CONTENTS

LIST OF TABLES

7

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D = Three-Dimensional

ACT = Anti Chymotrypsin Inhibitor

ADAM = A Disintegrin And Metalloprotease domain

BM = Basement Membrane

CAPS = 3-[CyclohexylAmino]-1-PropaneSulfonic acid

CE = Cell Extract

CO = Coulomb potential

CPK = Corey, Pauling, Koltun

ECM = Extra Cellular Matrix

EMSP1 = Enamel Matrix Serine Protease

ES = Enzyme-Substrate

EVB = Empirical Valence Bond

HB-EGF = Heparin-Binding Epidermal Growth Factor

hK = Human Kallikrein

IGFBP = Insulin-like Growth Factor Binding Proteins

IR = InfraRed

KLK = Kallikrein

LJ = Lennard Joen potential

MD = Moleculr Dynamics

mk = Mouse Kallikrein

MMP = Matrix Metallo Protease

NES1 = Normal epithelial Cell-Specific 1

NGF = Nerve Growth Factor

NMR = Nuclear Magnetic Resonance

PBS =Phosphate Buffered Saline

PDB = Protein DataBank

PI-6 = Protease Inhibitor 6

pk = Pig Kallikrein

PSA = Prostate specific antigen

PTI = Pancreatic Trypsin Inhibitor

PVDF = PolyVinyliDene Difluoride

Px = Protein-specific sites

QM = Quantum Mechanics

RMSD = Root Mean Square Deviation

RT-PCR = Reverse Transcriptase Polymerase Chain Reaction

SDS-PAGE = Sodium Dodecyl Sulphate PolyacrylAmide Gel Electrophoresis

SN = SuperNatant

Solv = Solvent

STI = Soyabean Trypsin Inhibitor

Sx = Substrate-specific sites

TACE = Tumor necrosis factor Alpha-ConvertasE

TBS-T = Tris Buffered Saline Tween

TEMED = TEtraMEthylethyleneDiamine

TGF = Transforming Growth Factor

TIMP = Tissue Inhibitor for Metallo Proteases

TK = Tissue Kallikrein

UPA = Urokinase-type Plasminogen Activator

vdW = van der Waals

Å = Ångstrøm (0.1nm)

DECLARATION

The work in this thesis has not been previously submitted for a degree or a diploma at any other higher education institute. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due references are made.

_____   _____

Sergio Manzetti   Date

ACKNOWLEDGMENTS

CHAPTER 1

1.1   Prekallikreins.

1.1.1   Introduction to the kallikreins.

The kallikrein family (*KLK*) is a sub-family of the serine proteases, consisting of 15 known genes in humans, and 13-26 in mouse and rat (Diamandis *et al*, 2000, Clements, 1989, Harvey *et al*, 2000). This protein family is predominantly expressed in the prostate, kidney, salivary glands and breast, and is involved in post-translational activities, such as propeptide-cleavage, and possibly matrix degradation. One of these proteins, prostate specific antigen (PSA), has been detected in high concentrations in serum from cancer patients, especially those with prostate cancer (Partin *et al*, 1999, Rittenhouse *et al*, 1999, Wei *et al* 1997, Diamandis *et al*, 2000). There have been substantial studies on the relationship between the levels of PSA and detection of cancer since the late 80's. Subsequently, the US Food and Drug administration have approved PSA as a tumour marker, but unfortunately it is not a good predictor of staging (Diamandis, 1998). In some cases of prostate cancer, the serum concentration of PSA does not differ from that in men without prostate cancer: this argues for a role for better diagnostic indicators, perhaps related to PSA. If the involvement of the other kallikreins could be clarified, alternative detection methods may be developed, and possibly different therapeutics, such as novel inhibitors to these enzymes may be applied to treating such cancers.

The following information in this thesis will hopefully increase the knowledge about this new protein family, and supply both *in vitro* and *in silico* data for further studies in this field and possiblys contribute in establishing a foundation for rational drug design directed to some members of this family.

1.1.2   Nomenclature.

The first members of the kallikrein family to be recognised as relatives of tissue kallikrein were hGk3 and PSA (Baker and Shine, 1985, Riegman *et al*, 1992, Schedlich *et al*, 1987). Berg *et al* (1992) established the new nomenclature for these genes: *KLK1*, *KLK2* and *KLK3* for genes, and tissue kallikrein, hgK2 and PSA for the respective proteins. Berg also emphasized that the potential future members of the kallikrein family should be assigned numerically in the order of their publication. These very new members, were individually cloned in the period 1994 – 2000, and were not recognised as a gene family until early 1999 after the efforts of Stephenson *et al*

(1999), Yousef and Diamandis (1999a, d, 2000a), Yousef *et al* (1999b,c), Harvey *et al* (2000) and Diamandis *et al* (2000). These new kallikreins were therefore not nominated originally as kallikreins, but differently, as Protease M, Neuropsin and Stratum Corneum Chymotryptic enzyme (SCCE) (see Table 1.1). This nomenclature was first modified to the "kallikrein-like" nomenclature, by Yousef and Diamandis (1999a, b) and Yousef *et al* (1999a, b, c), who identified a certain familiarity between the original kallikrein family and these new genes and named these new genes for "kallikrein-like" genes. In turn, Harvey *et al* (2000) identified and proved the genetic familiarity between the old (*KLK1*, *KLK2*, *KLK3*) and the new kallikreins (Neuropsin, SCCE, NES, etc.), and proposed, after the previous efforts of Stephenson *et al* (1999), a nomenclature based on numbering these new genes, with increasing number from *KLK1* to *KLK14* in the telomeric direction. This nomenclature was applied, and, Yousef *et al* (2000a) subsequently found a new gene, which was designated *KLK15*, though located between *KLK1* and *KLK3* in the original locus. Table 1.1 on next page shows the current nomenclature of the various kallikrein members.

Table 1.1 Nomenclature of the kallikrein gene family.

The new and old nomenclature of the new kallikrein gene family members and the principal research groups that have contributed to this nomenclature.

Gene name	Harvey *et al* (2000)	Yousef *et al* (1999-2000)	Nelson *et al* (1999)	Hansson *et al* (1994)	Yamashiro *et al* (1997)	Anisowicz *et al* (1996)	Mitsui *et al* (2000)	Stephenson *et al* (1999)	Luo *et al* (1998)	Shimizu *et al* (1998)	Brattsand *et al* 1999	Little *et al* (1997)
KLK4	hK4	*KLK*-L1	Prostase					hK4				
KLK5	hK5	*KLK*-L2									SCTE	
KLK6	hK6	*KLK*-L3			Neurosin	Protease M						Zyme
KLK7	hK7	-		SCCE								
KLK8	hK8	*KLK*-L4								Neuropsin		
KLK9	hK9	-										
KLK10	hK10	-							NES1			
KLK11	hK11	-					TLSP					
KLK12	hK12	*KLK*-L5										
KLK13	hK13	*KLK*-L4										
KLK14	hK14	*KLK*-L6										
KLK15		hK15										

1.1.3   Gene structure

Until the mid 90's the kallikrein family was thought to consist of only 3 genes, the *KLK1*, *KLK2* and *KLK3* genes, encoding for tissue kallikrein, human kallikrein 2 and prostate specific antigen, respectively (Clements, 1997). Baker and Shine discovered *KLK1* cDNA in 1985, whereas the *KLK1* gene was identified by Evans *et al* (1985). Lundvall and Lilja (1987) and Schedlich *et al* (1987) sequenced and mapped the genomic region of the *KLK3* and *KLK2* genes respectively. *KLK* 1, 2 and 3 are all located in a 60kb region at 19q13.3-13.4 with a separation of 12kb between *KLK1* and 3 and of 31kb between *KLK2* and *KLK3* (Riegman *et al*, 1992). Kallikrein 1 is highly expressed in kidney, while *KLK2* and *KLK3* have expression patterns that are confined to the prostate, although some lower expression (at the RT-PCR level) has been detected in cancerous human endometrium and pituitary, as well as in breast cancer (Clements, 1997, Diamandis *et al*, 2000). The new genes (*KLK4*-15) have a very widespread expression pattern. These genes were recently found by Harvey *et al* (2000), Diamandis *et al* (2000) and Yousef and Diamandis (2000a) to lie in a ~250kb region located telomeric to the original cluster of *KLK1*, 2 and 3, on chromosome 19 at position q13.3-13.4. All of the new genes are transcribed in the direction telomere to centromere, while *KLK2* and 3 are transcribed in the opposite direction. The first, and parts of the second coding exons in the *KLK* genes encode for the prepropeptide region, while the remainder of the second exon through to the fifth exon encode for the rest of the protein. The conserved catalytic triad is encoded in the 2nd, 3rd and 5th exons (Clements *et al*, 2001). The majority of kallikreins have 5 coding exons and 4 introns, although some have up to 7 exons and 5 introns (Clements *et al*, 2001). All intron phases are completely conserved.

Several of the new kallikrein genes have multiple mRNA transcripts (Yoshida *et al*, 1998, Riegman *et al*, 1989, Liu *et al*, 1999, Mitsui *et al*, 2000), which is a trait not observed in rodent genes and also many pseudogenes, encoding for truncated proteins (Clements *et al*, 2001), possibly suggesting a potential high number of members as found in the rodent families. The sequence similarity within the entire kallikrein family is in the range of 25-49% at the protein level, resulting in roughly 5 phylogenetic branches: a) *KLK1, 2, 3,* b) *KLK4, 5, 7*, c) *KLK6, 13, 14*, d) *KLK15, 11, 8, 9* and e) *KLK10* and *KLK12* (Harvey *et al*, 2000).

As mentioned, the expression patterns of these newer genes are very widespread, possibly indicating their various functions. Some of the kallikrein genes are found to be highly expressed in the pancreas and prostate (Harvey *et al*, 2000) and others in tissues as brain, lung, thyroid gland and testis (Harvey *et al*, 2000, Diamandis *et al*, 2000 – see Table 1.2 for details).

1.1.4   Individual organization and expression patterns of the kallikreins.

The target protein of this project, human kallikrein 4, was originally discovered by Nelson *et al* (1999), which they named "prostase", due to its deduced prostate-specific expression. They reported 5 exons and 4 introns and the respective sequences of the intron/exon boundaries, as also found by Stephenson *et al* (1999). Stephenson *et al* (1999) found that the *KLK4* gene is transcribed on the reverse strand, in the same direction as *KLK1*, and opposite to *KLK2* and *KLK3*. Other groups that also studied this gene reported its transcriptional direction, and its position between *KLK2* and the Protease M-gene (*KLK6*) (Yousef *et al* 1999b, Myers and Clements, 2001, Dong *et al*, 2001).

Ekholm *et al* (2000) and Yousef and Diamandis (1999a) reported identification of the next kallikrein gene on the *KLK*-locus (*KLK5*) and found it to be composed of 5 exons and 4 introns as most of the other kallikreins. The *KLK6* gene, encoding Protease M (hK6), was initially reported by Anisowich *et al* (1996) with an expression pattern in both normal and LnCap cells. Yamashiro *et al* (1997) subsequently cloned the same gene from COLO201 adenocarcinoma cells, and mapped the non-coding and coding regions. The same gene was later described by Yousef *et al* (1999a) who reported its composition of 7 exons and 6 introns, which is unusual among the kallikreins.

Yousef *et al* (2000b) characterised and mapped also the *KLK7* gene and determined its tissue expression and hormonal regulation in human cell lines. According to Hanson et al (1994), who cloned the cDNA first, the gene was found to contain 6 exons (one non-coding) and 5 introns, and had completely conserved intron phases, like most of the other members of the kallikrein family. The *KLK8* gene was successfully cloned by Yoshida *et al* (1998), and found to be expressed predominantly in keratinocytes. As with the other newer kallikreins, it was composed of six exons and five introns and has 72% identity with mouse neuropsin and thought to be the human orthologue of this gene.

Yousef and Diamandis (2000b) analyzed the 300kb human kallikrein-gene region and identified a new kallikrein gene, the *KLK9* gene, and its chromosomal localisation between *KLK8* and *KLK10*. The *KLK9* gene was determined to be of 5 coding exons, and 4 introns. Luo *et al* (1998) were the first to report the existence of the *KLK10* gene, as NES1 (Normal Epithelial Cell-Specific 1). They mapped the genomic structure and position at 19q13.3-13.4, and concluded the presence of 6 exons (one untranslated) and 5 introns. Interestingly Luo *et al* (1998) observed that

the *KLK10* gene is located 0.73cM centromeric to the PSA and that the arrangement at this locus was; 19centromere-*KLK10*-*KLK1*-PSA-*KLK2*-19telomere. Their mapping was incorrect, considering the later results from Harvey *et al* (2000), Yousef and Diamandis (1999a, b), Yousef *et al* 1999a, b, c) and Diamandis *et al* (2000).

Yoshida *et al* (1998) defined the mRNA transcripts of the *KLK11* gene, but Yousef *et al* (2001) determined the precise exon-intron boundaries of the gene. This gene was deduced to span 5.3kb of the genomic sequence at 19q13.3-13.4, and contained 5 introns and 6 exons one of which was untranslated. Yousef and Diamandis (1999e) and Yousef *et al* (2000a) reported the genomic sequence of *KLK12* and *KLK13* with a detailed chromosomal localization, and their relative positions to each other, while Hooper *et al* (2000) reported the genomic structure and sequence of *KLK14*, which was found to be composed of the unusual number of 7 exons and 6 introns, like the *KLK6* gene. Finally, Yousef *et al* (2001) reported the *KLK15* gene, not previously identified by other groups, and found it to be between *KLK1* and 3, in the original kallikrein locus. As with most of the other kallikreins, this gene also contained 5 coding exons and 4 intervening introns. The expression patterns for these genes reported by these groups are presented below in Table 1.2.

Table 1.2 Expression patterns of the new kallikrein members.

	RT-PCR	NORTHERN BLOT	VARIOUS	VARIOUS
KLK1	(a) Kidney, pancreas and salivary gland.	Kidney, pancreas and salivary gland.		
KLK2	(a) Prostate, breast.	Prostate	Thyroid (i)	
KLK3	(a) Prostate, breast.	Prostate	Thyroid, Breast, Trachea, Salivary glands. (i)	
KLK4	(a) Prostate, breast, testis, uterus, thyroid.	Prostate	Testis, brain, lung, colon, mammary gland, adrenals, uterus, thyroid, and salivary glands (e)	Prostate (o)
KLK5	(a) Breast, brain, testis and skin.	Testis and mammary gland	Brain, nervous system. (f)	
KLK6	(a) Brain, breast, ovary, kidney and uterus.	Pancreas. Some expression in brain, testis, appendix, colorectal adenocarcinoma.	Mainly in salivary glands. Some in lung, fallopian tissue, breast, colon, kidney placenta and pituitary. (j)	Brain, kidney and pancreas. (l)
KLK7	(a) Skin, brain, kidney, breast, salivary glands and thymus	Pancreas	Cornifying epithelia (h)	Skin, brain and kidney. =Brain and some in spleen. (k) Brain, mammary glands, cerebellum, spinal cord, kidney and skin. Lower levels in salivary glands, uterus,

				thymus, thyroid, placenta and testis. **(p)**
KLK8	**(a)** Brain, skin.	Pancreas. Some expression in duodenum, liver, esophagus, stomach and salivary gland.		Hippocamidal pyrimidial neurons. **(n)** Limbic system in brain **(q)**
KLK9	**(a)** Skin, thymus, trachea, cerebellum, spinal cord.	Pancreas	Thymus, testis, spinal cord, cerebellum, trachea, mammary gland, prostate, brain, salivary gland, ovary, and skin **(g)**	
KLK10	**(a)** Breast, ovary, testis, prostate, small intestine, lung and *pancreas*.	Pancreas, some expression in esophagus, duodenum, trachea, colorectal adenocarcinoma.		
KLK11	**(a)** Brain, skin.	Pancreas. Some in heart, salivary gland, stomach, liver, and skeletal muscle.	Cerebellum, prostate, salivary glands, stomach, lung, thymus, small intestine, spleen, liver and uterus. **(d)** Brain and prostate, and moderately expressed in lung, heart and testis **(c)**	
KLK12		Pancreas, brain, duodenum, appendix.		
KLK13	**(a)** Prostate, breast, testis, and salivary gland.	Pancreas, esophagus and appendix. Some in duodenum,		

		stomach, brain, testis, prostate, salivary gland.		
KLK14	**(b)** Prostate and skeletal muscle.	Brain, bone marrow, fetal liver. Some in liver, pancreas fetal spleen and prostate.		
KLK15				Mainly thyroid gland. Some expression in the prostate, salivary and adrenal gland, colon, testis, and kidney. **(r)**
References	**(a)**Diamandis *et al*, 2000 **(b)** Hooper *et al* (2000)	Harvey *et al*, 2000	**(c)**Mitsui *et al* (2000) **(d)**Yousef *et al* (1999c) **(e)**Yousef *et al* (1999b) **(f)**Yousef and Diamandis (1999a) **(g)**Yousef and Diamandis (2000b) **(h)**Ekholm *et al* (2000) **(i)**Maglakara *et al* (2000) **(j)**Diamandis *et al* (2000a)	**(k)** Hanson *et al* (1994) **(j)** Anisowich *et al* (1996) **(m)**Yamashiro *et al* (1997) **(n)** Shimizu *et al* (1998) **(o)** Nelson *et al* (1999) **(p)**Yousef *et al* (2000c) **(q)**Kato *et al* (1997) **(r)**Yousef *et al* (2001)

1.1.5   Physiological functions of the kallikreins.

The function of hK1 is the most studied among the original members of the kallikrein family. Clements (1997) reviewed the current literature that showed that hK1 was involved in blood flow, vascular permeability, and sodium and water homeostasis. Tissue kallikrein (hK1) has also been implicated in hypertension and the inflammatory conditions of ischemia, renal nephritis and diabetic renal disease (Clements, 1997). Furthermore, it has been involved in inflammatory responses in the urogenital tract, and in male reproductive function, via enhancement of sperm motility and testicular blood flow (Clements, 1997). The substrate of hK1 is hepatic-derived kininogen that it hydrolyses to produce kinins (Clements, 1989), which are potent vasodilators and multipotent effectors in various systems as noted above (Raidoo *et al*, 1999).

hK2 and PSA were previously thought to be restricted to the prostate in entirely uncomplexed forms (Rittenhouse *et al*, 1998). PSA exists in the seminal fluid 50-70% as enzymatically active, and the rest as an inhibited form (in complex with Anti Chymotrypsin inhibitor (ACT), or in the pro-form). 50% of hK2 is inhibited by Protein C inhibitor in seminal fluid while the remainder is its active form (Rittenhouse *et al*, 1998). PSA is also expressed in breast, ovarian, and lung cancer (Yu *et al*, 1994, 1995).

PSA, and human kallikrein 2 (hK2), have been reported to participate in extracellular degradation (Lilja, 1985). hK2 is hypothesized to activate PSA (Lilja, 1987), extracellularly, which in turn degrades the seminal vesicle proteins seminogelins I and II and fibronectin leading to dissolution of the seminal clot (Lilja, 1987). Hsieh *et al* (2000) suggested specifically that PSA acts on the seminal coagulum secreted by seminal vesicle, leading to the release of the spermatozoa.   In addition, these kallikreins seem also to be related to pro-hormone processing (Clements, 1989). For instance, PSA has been shown to cleave and inactivate the insulin-like growth factor binding protein-3 (Cohen *et al*, 1992). Coombs *et al* (1998a) confirmed the reports that PSA can cleave laminin, fibronectin and parathyroid-hormone-related protein in such hormone processing pathways.

Lovgren *et al* (1997) purified PSA and hK2 from insect and hamster kidney cells and reported that hK2 and PSA were respectively expressed in their mature and zymogen form. ProPSA was then activated with hK2, at high rates, as previously suggested by Lilja (1985) and also shown by Takayama *et al* (1997) and Kumar and colleagues (1997). Lovgren *et al* (1997) reported also that

hK2 and PSA are cleaved (additional to their activation cleavage) internally, at K145 (PSA) and R145 and R101 for hK2. These cleavage products were found to be inactive. However Wolf *et al* (1995) documented that uPA is cleaved at an internal site (K158-I159) when truncated, but then retains enzymatic activity as a two-chain protein associated via disulphide bridges and internal non-covalent bonds, which could be the case for the kallikreins. Deperthes *et al* (1997) reported also that hK2 has some kininogenase activity, but at a very low level, suggesting that its role is not related with kininogen activation. Rittenhouse *et al* (1998) delineated that hK2 was not involved in hydrolysing structural semenogelin proteins in seminal plasma, like PSA and hK1. The same group reported that hK2 degrades IGFBPs *(*Insulin-like growth factor binding proteins) more rapidly than PSA and activates uPA *in vitro* (Takayama *et al*, 1997). In Rittenhouse *et al*'s (1998) view, this suggests that hK2 has an influence on the extracellular proteolytic cascade that occurs in prostate cancer cells since uPA is highly correlated with cancer aggressiveness in general, and especially with prostate cancer (Achbarou *et al*, 1994).

The new members of the kallikrein family are less studied in terms of physiological function. Nelson *et al* (1999) postulated the possible function of hK4 is to have a role in extracellular matrix degradation. Its closest relative, the enamel matrix serine protease 1 (EMSP1) (Fukae & Shimizu, 1998), is a porcine protein which is believed to be involved in remodelling of the organic matrix in tooth development (Simmer *et al*, 1998), a process requiring extracellular matrix degradation activity. Nelson *et al* (1999) postulated therefore that hK4 might have a similar role as the EMSP1 in matrix degradation, or in activating growth factors in cytokine networks. More recent studies show that hK4 readily activates scuPA, pro-uPa and proPSA *in vitro* (Takayama *et al*, 2001). Furthermore, hK4 was also recently shown to be expressed as two major transcripts, one of which was principally restricted to nuclear and perinuclear regions of the cell (Korkmaz, 2001).

Stratum corneum chymotryptic enzyme (hK7) was reported to catalyse the degradation of intracellular cohesive structures in the continuous shedding of cells from the skin surface (Egelrud *et al*, 1993, Sondell *et al*, 1995). Ekholm *et al* (2000) suggested that the hK7 might have a role in the proteolytic activation of the hK5 precursor, and is possibly inhibited by alpha1-antichymotrypsin. hK5 was later shown to activate hK7 (Brattsand, personal communication) and is expressed in skin, breast, salivary gland, and esophagus and has recently been proposed as a novel tumour marker (Dong et al, 2003, Yousef et al, 2003).

hK8 was reported to have amyloidic activity in the brain (Diamandis *et al*, 2000). Kishi *et al*, (1999) suggested that mouse neuropsin (the closest species orthologue to hK8) is involved or associated with neural plasticity. Diamandis *et al* (2000) suggest that hK8 might be involved in the production of cerebrospinal fluid, the formation of memory, and in some forms of epilepsy in humans. Interestingly, Shimizu *et al* (1998) detected strong proteolytic activity from mouse neuropsin upon fibronectin, an extracellular matrix protein, which is present in the nervous system, indicating its putative role in the proteolysis of the extracellular matrix.

Goyal *et al*. (1998) reported that hK10 (NES1) abolishes the ability of tumorigenic cells to grow in an anchorage dependent manner, thereby slowing down cell proliferation. In other words, Goyal *et al* (1998) suggested that the primary function of hK10 is as a tumour suppressor.

The physiological functions of the remaining hK9, hK11, hK12, hK13, hK14 and hK15 enzymes are still unknown.

1.1.6 The kallikreins in cancer.

Tissue kallikrein- (hK1) is involved in the kallikrein kinin system, which controls the processes of vasodilation and vascular permeability, in addition to other functions (Clements, 1997 Scholkens, 1996 and Hermann *et al*, 1999). The pathways of vasodilation seem to facilitate cancer progression according to Dlamini *et al* (1999) and Raidoo *et al* (1999). Raidoo *et al* (1999) reported that tissue kallikrein (TK) is involved in stimulating proliferation of tumour cells by increasing vascular permeability, and enhancing metastasis, based on their finding of high concentrations of hK1 in both poorly and well-differentiated giant tumour and mast cells. Indeed, the expression of tissue kallikrein and the bradykinin receptor 2, the effector receptor for bradykinin generated by the action of hK1 in endometrial cancer and prostate cancer, was later confirmed using RT-PCR detection (Clements & Mukthar, 1997). Raidoo *et al* (1999b) emphasised also the presence of other serine proteases in addition to tissue kallikrein, probably kallikreins, in facilitating tumour progression due to vasodilation. This suggestion is supported by the studies of Maeda *et al* (1999) and Tschope *et al* (1999), who noted the role of one of the kinins, bradykinin, released by hK1, in activating its receptor that in turn increases the release of arachidonic acid and other metabolites of eicosanoid synthetic pathway. These metabolites later enter the cyclooxygenase cascade that directly leads to expression of prostanoid, histamine, prostacyclin and release of nitric oxide, which all are potent vasodilators. The consequence of

these tissue kallikrein-facilitated pathways is increased vascular permeability, angiogenesis and vasodilation, facilitating cancer cell invasion (Maeda *et al*, 1999).

One of the most striking pieces of evidence was the report that the incidence of allelic loss on 19q was particularly high (81%) in oligodendrogial tumours (Reifenberger *et al*, 1994). These tumours are primary tumours of the nervous system that attack cerebral tissue. Reifenberger *et al* (1994) analysed the tissue from 37 brain tumour patients and documented loss of heterozygosity at, among others, location 19q13.2-13.4. The kallikrein genes are located at 19q13.1-13.4, which further emphasizes their probable involvement in cancer, especially brain tumours.

Diamandis *et al* (2000) reviewed PSA's role in cancer to be: an inducer of apoptosis, to reduce cell proliferation, to reduce tumorigenic potential of cancer cell lines, and to inhibit angiogenesis (blood vessel creation) thereby inhibiting carcinogenesis. PSA is downregulated in both breast and prostate cancer cells, and more aggressive tumours have lower PSA expression (Diamandis *et al*, 2000, Rittenhouse *et al*, 1998). As stated earlier, PSA is also involved in hormone processing (Coombs *et al*, 1998a and Clements, 1989), which indeed underscores PSA's involvement in cancer. hK2 and hK4, being prime activation candidates for the physiological form of PSA (Rittenhouse *et al*, 1998, Takayama *et al*, 2001) are therefore involved in these mechanisms as well. hK4 is expressed in prostate and breast tissue, and it is hormonally regulated (Diamandis *et al*, 2000). It is suggested to play a role in bony metastasis of breast and prostate cancers, and also regulate growth factor and cytokine networks, which are in turn related to tumour growth and extracellular degradation (Nelson *et al,* 1999, Bhat-Nakshatri *et al*, 1998, Raidoo *et al*, 1999).
hK6, is reported by Anisowicz *et al* (1995) to be dramatically downregulated in breast cancer, and expressed at higher levels in metastatic sites. Diamandis *et al* (2000a) detected hK6 in breast cyst fluid and in breast tumour cytosols (except for the normal tissues) at 74µg/L and 2.1µg/L respectively. The same group observed further that hK7 might activate Interleukin 1. Interleukin activation was proposed by Bhat-Nakshatri *et al* (1998) to induce uPA activation, through upregulation of NGF, a factor known to regulate uPa, and lastly, trigger extracellular matrix degradation, facilitating tumour progression (Raidoo *et al* 1999). In addition, the overproduction of hK7 in ovarian carcinomas was reported by Lowel *et al* (1999).

hK10 was reported to be expressed abundantly in a normal breast cell line, but was absent from the same cell line that had been irradiated to become tumorous, thus suggesting a down-regulation in some cancerous tissues (Diamandis *et al*, 2000). Besides, its expression is

downregulated in cancerous breast and prostate tissue, compared with normal tissue, and in a number of cell lines (Diamandis *et al*, 2000). hK10 is also reported to reduce the proliferation rates in cancer in nude mice (Diamandis *et al*, 2000). Goyal *et al* (1998) reported also that hK10 is dramatically downregulated in cancer cell lines, and that it seems to reduce tumorigenic potential of cell lines in a nude mouse model of carcinogenesis.

1.2   The kallikrein enzymes (serine proteases).

1.2.1   The catalytic mechanism of serine proteases.

The kallikrein enzymes are a sub family of the serine proteases, a superfamily of proteolytic enzymes that hydrolyse peptide bonds using a conserved catalytic triad (His57, Asp102 and Ser195 – chymotrypsin numbering). The enzymatic action is a simple acid-base catalysis carried out by the nucleophile oxygen ion from the hydroxyl group on serine 195, formed by the histidine that becomes an extremely potent base due to the charge-relay mechanism (Fersht, 1977). A detailed step-by-step explanation of this mechanism is presented in the figure 1.
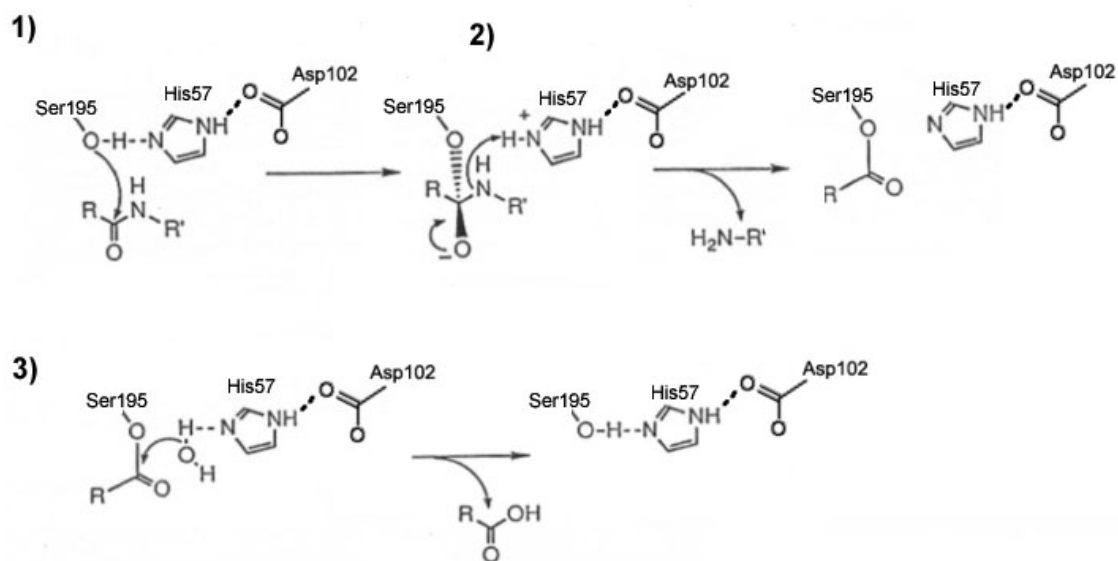
Figure 1.1. The catalytic mechanism of the serine proteases.

(1) The serine is deprotonated by His57, leading to the nucleophilic attack of the scissile peptidebond. (2) The nitrogen atom from the scissile peptide bond forms hydrogen bonds with the Histidine 57 nitrogen. This leads to the formation of the short-lived tetrahedral oxyanion complex (as seen in step 2), stabilized by H-bonds at the "oxyanion hole" (backbone atoms from Ser195 and Gly193). The oxygen from the scissile peptide transfers one electron back to the alpha carbon, leading it to disrupt the amidebond, due to the peptide-bond nitrogen's hydrogen-bond with the Histidine's nitrogen. The amine product is released, and a covalent bond formed between the enzyme and remaining part of the substrate creates the acyl enzyme. (3) Histidine 57 deprotonates a water molecule, leading it to nucleophilically attack the carbonyl group of the second part of the substrate protein, forming a new C-terminus. The equilibrium between the Histidine 57 and Serine 102 is re-established (active enzyme) (Fersht, 1977).

These proteases are divided into three subgroups, according to their primary specificity: the trypsins, the chymotrypsins and the elastases. To understand and classify the specificity of an enzyme (preferred amino acids side-chains in the hydrolysed peptide) an overview of the subsite-concept is outlined. The subsites on the enzyme (S-sites) are the residues that interact with the residues from the substrate (P-residues). The amino acid residues of the enzyme are designated …, S2, S1, S1', S2', … and the corresponding subsites on the substrate are termed …, P1, P2, P1', P2', … (Schecter and Berger, 1967). The S/P subsites are on the carbonyl side of the scissile peptide bond, towards the N-terminus of the substrate, while the S'/P' are the subsites toward the C-terminus.

Trypsin-like enzymes cleave substrates with the positively charged residues Arg or Lys at the P1 position, due to a salt-bridge established with Asp189 (S1) in the deep S1-cavity (Powers *et al*, 1993, Perona and Craik, 1995, Coombs *et al*, 1999). At the P2, P3 and P4 Asp, Glu and Arg residues occur normally, since this region is normally required to be acidic for proper interaction to occur (Maroux *et al*, 1971, Light & Janska, 1989). Trypsin and trypsin-like proteases prefer Arg at position P1, instead of Lys. The affinity for Arg is 2-10 fold higher than for Lys. This happens because Lys interacts solely by water mediated H-bonding with S1, while Arg interacts directly with Asp189 (Fig. 2) (Coombs *et al*, 1998b).

Chymotrypsin-like proteases prefer substrates with large aromatic and aliphatic side-chains, and have therefore a cavity lacking the negative charge (Perona and Craik, 1997). Chymotrypsins cleave polypeptides with a Tyrosine at P1, preferably, but also when a Phenylalanine is at P1 (Warshel, 1997), or Leucine and Tryptophan (Perona and Craik, 1997).
Elastase-like enzymes prefer small aliphatic residues at P1, such as Alanine and Glycine, and therefore have a shallow hydrophobic pocket (Perona and Craik, 1997).

This figure is not available online.
Please consult the hardcopy thesis
available from the QUT Library

Figure 1.2. The specificity pocket in trypsin.

The deep cavity is the S1 pocket that hosts the P1-side chain of the substrate. P2 is located to the left, while P1' is located upwards (printed with permission from Coombs *et al*, 1999).

1.2.2   Structural features of kallikreins.

The kallikreins are serine proteases. Serine proteases retain three typical folds: the subtilisin fold, the trypsin fold and the serine carboxypeptidase fold (Perona and Craik, 1995). These folds vary extensively from each other, and the kallikreins all belong to the trypsin-fold-group (which also include chymotrypsin proteases such as chymotrypsin and PSA). The classical trypsin-fold describes two large "sandwich-like" beta-barrels with the specificity pocket in the middle near the catalytic triad. (Perona and Craik, 1995). Two helices and a short helix are conventionally found in this group, but the presence of the latter may vary from member to member. The binding region is in a cleft between the two beta-barrels, where the catalytic triad lies in the middle. The size of these enzymes ranges from 200-300 residues. Eight separate surface loops have been implicated in distal site substrate or inhibitor contacts for different enzymes of the serine protease family (Perona and Craik, 1995). These are the regions that tend to vary extensively among members of the kallikrein family, which retain a homology to the mouse neuropsin structure (Fig 1.3).
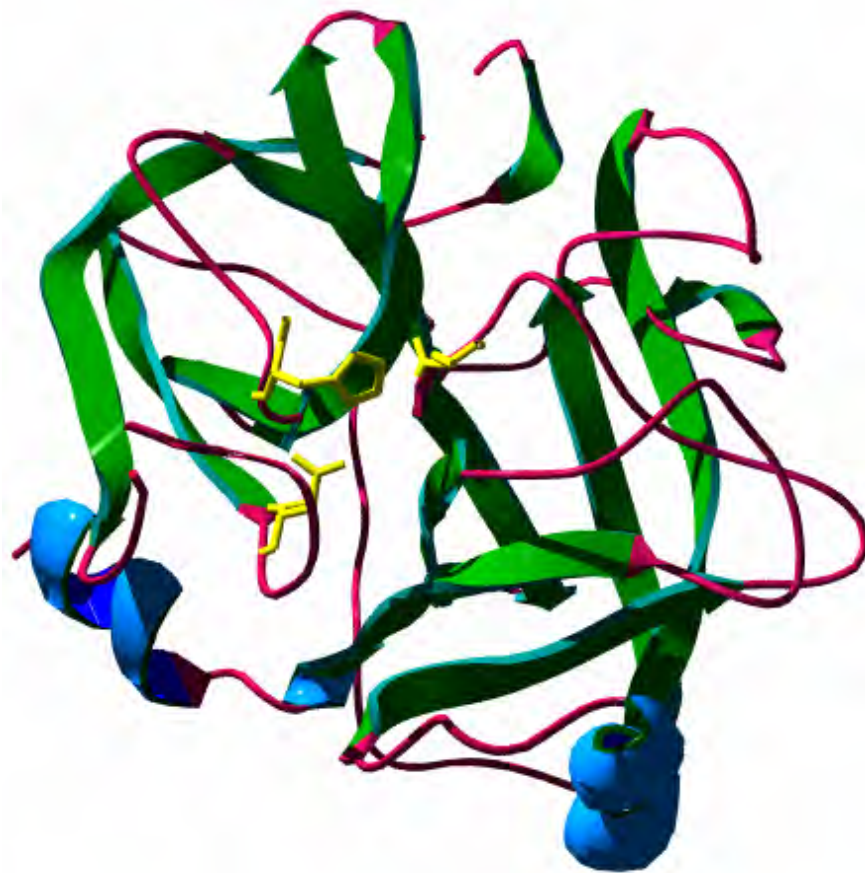
Figure 1.3. The structure of the mouse kallikrein 8.

Neuropsin (mouse kallirein 8) is shown here, and is one of the crystal structures with highest sequence homology to the kallikrein family (Kishi *et al*, 1999). All loops are coloured in red; beta-sheets in green and helices in blue. The catalytic triad is highlighted in yellow.

1.2.3 Conserved residues in serine proteases.

Margareth Dayhoff and co-workers (Dayhoff *et al*, 1972) mapped a series of residues that were found to be highly conserved within the serine protease superfamily. These residues include 6 of the 10 conserved cysteine residues in the kallikreins, as well as the catalytic triad and other subsites. These residues are Gly27, Pro36, Cys48, Gly49, Val58, Ala61, Gly62, His63, Cys64, Asp107, Leu113, Leu128, Gle143, Gle145, Leu158, Pro164, Cys171, Cys184, Gly186, Cys195, Gly197, Asp198, Ser199, Gly200, Gly201, Pro202, Gly211, Ser214, and Cys220, all part of the mature enzymes.

Pro36, Cys48, Ala61, His63 (catalytic), Cys64, Asp107 (catalytic), Leu113, Gly143, Gly145, Leu158, Cys171, Cys184, Cys195, Gly197, Asp198, Ser199 (catalytic), Gly200, Gly201, Pro202, Gly211, Ser214, Cys220 are all conserved in all kallikreins, trypsinogen and chymotrypsinogen. Gly49, Val58, Leu128, Gly186 are almost conserved in all members. The rest are either not conserved, or only conserved to a lesser extent in the kallikreins.

1.2.4 Natural inhibition mechanism of trypsins and chymotrypsins.

Serine proteases are inhibited by a wide range of inhibitors such as PTI (pancreatic trypsin inhibitor), STI (soyabean trypsin inhibitor), ACT (antichymotrypsin inhibitor), Protein C inhibitor, protease inhibitor 6 (PI-6), antithrombin, the Alzheimer`s disease amyloid A4 peptide, the hirustasin inhibitor and many more (Sticht *et al*, 1995, Schoofs *et al*, 2002). Kallikreins are to date reported to be inhibited by ACT, Protein C inhibitor, protease inhibitor 6 (PI-6), the hirustasin inhibitor (Deperthes *et al*, 1995, Mittl *et al*, 1997, Hsieh *et al*, 2002 and Saedi *et al*, 2001), but are probably also inhibited by many of the others listed above, being common trypsin-like and chymotrypsin-like serine proteases.

The mechanism of kallikrein-inhibition is well defined by the crystallized complex of the pig kallikrein 1 with the inhibitor hirustasin (Mittl *et al*, 1997), which designates a dimer that highly resembles the enzyme-substrate complex of serine protease (Wilmouth *et al*, 2001). The inhibitor docks 4 residues (Val27-Arg30) on the P-region that is predominantly attached to a conserved motif (Ser214-Gly216) on the enzyme by four backbone H-bonds. The "scissile bond" in the inhibitor forms a H-bond network similar to the oxyanion hole transition state, and the P'-region has two residues (Ile31 and Arg32) that provide their hydrophobic moieties in contact with the enzyme surface (see Fig 4). These contacts of interaction illustrate an inhibitory mechanism which is so similar to the enzyme-substrate binding mechanism (Wilmouth *et al*, 2001) that it presents an excellent opportunity to use the crystallized kallikrein-hirustasin complex as a

template for enzyme-substrate modelling studies for the new kallikrein family, as carried out in this study.
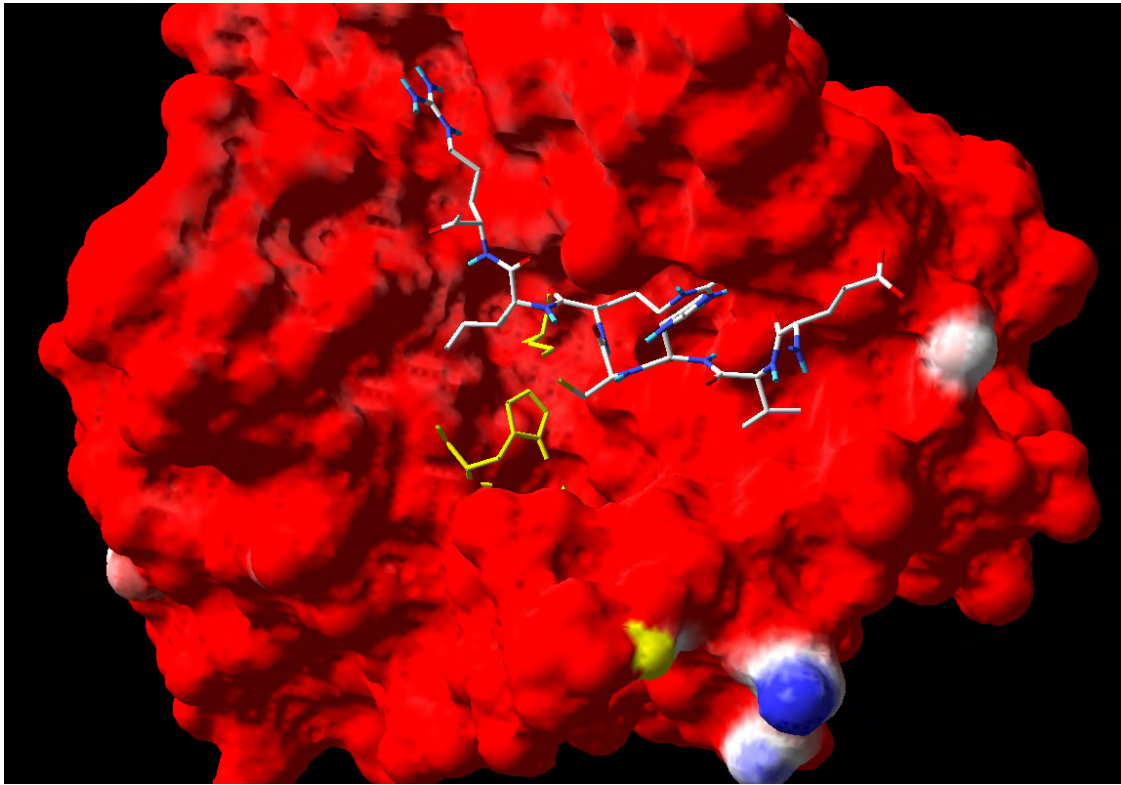
Figure 1.4. The binding interface between pig kallikrein 1 and the hirustasin inhibitor.

A molecular surface generated for pig kallikrein 1 with the interacting fragment of the hirustasin inhibitor (Glu26-Arg32) drawn in stick mode. The yellow catalytic diad (His 57, Ser195) is coloured yellow, the inhibitor in CPK, and the molecular surface by electrostatics (Generated in Swiss PDB Viewer, Guex and Peitsch, 1997).

1.3   Cultivation, expression and purification of kallikreins.

1.3.1   Expression systems used for kallikrein expression.

Only a few of the kallikreins, mainly PSA and hK2, have been thoroughly characterised biochemically. In order to express a recombinant protein of sufficient quality and quantity, the right expression system must be applied. Vlak and Keus (1990) reported on the quantities and qualities of recombinant protein expressed in bacterial, mammalian and insect cell systems. The most ideal expression system would appear to be the insect cell system. However bacterial systems are still useful since as an example, the activity of prokaryotic-expressed PSA was normal in the absence of glycan chains (Rittenhouse *et al*, 1998). Mammalian cells gave low yields of the recombinant protein, although all the required criteria, such as glycosylation, amidation, folding and acylation were present. In yeast cells the yield is again the essential problem, since the culture expressed only $1/30^{th}$ that of insect cells (Vlak & Keus, 1990). Lovgren *et al* (1997) expressed two mutants of hK2, with alterations in the propeptide bearing the cleavage site for factor X and one for enteropeptidase. This yielded expression levels up to 40 times higher than the wild type, in their insect cell system. Other useful approaches for increasing yield are those in which the N-terminal sequence composition is modified. Lovgren *et al* (1997) reported also that different expression systems might yield some different variants. Given the choice of a baculovirus system in this project, a post-translational pattern more similar to the human is expected, which facilitates the making of better and more specific antibodies for hK4 for immunohistochemical applications.

1.3.2   Expression of kallikreins in insect cells.

Shimizu *et al* (1998) cloned the mouse neuropsin gene (*mk*8) into a baculovirus system, a procedure that involves three steps: the pre-infection preparation step, which is the engineering step of the plasmid-enzyme construct and growing the insect cells, the infection and the post-infection harvesting. The cloning and preparation of the construct will not be discussed here, as it does not form part of my project.

For the preparation of the insect cells, the culture was grown at 27°C before infection with recombinant baculovirus at a multiplicity of infection (MOI) of 1-20. After infection, the incubation medium was harvested by removing the cells by centrifugation and then the medium containing the recombinant protein was purified as per section 1.3.3 (the yield was not reported in this study).

HK1 was expressed in insect cells (baculovirus system) under stirred conditions, in a 2L bioreactor, at pH 6.2 at an oxygen saturation of 40% and a cell density of 10^7 cells/mL (Angerman *et al*, 1992). The infection was carried out at a MOI of 10 and a yield as high as 5mg/L was achieved after 96hr incubation.

1.3.3   Purification procedures.

Reviewing past procedures in the literature, purification of kallikreins from a recombinant cell culture is carried out in three main steps - centrifugation, dialysis and chromatography. In the isolation of kallikrein 2 from seminal fluid, and neuropsin from mouse cerebellum, several centrifugation steps were applied, in the region of 10,000xg for 10-15 min., in order to remove cellular components such as cell wall and larger complexes. After this, dialysis was required for about 24h at 4°C, in order to exclude larger proteins, such as extracellular structure proteins and larger carbohydrates from the supernatant (Shimizu *et al*, 1998, Frenette *et al*, 1998). At this stage, size exclusion chromatography was applied to remove cytosolic and nuclear proteins. Finally, anion-exchange chromatography was applied for further fractionation, and the kallikrein eluted with a linear gradient of 0-1M NaCl (Shimizu *et al*, 1998).

Another alternative, which will be applied in this project, is using Nickel- or Cobalt-affinity chromatography, instead of size exclusion chromatography. The recombinant protein carries a hexa-histidine tag at the C-terminal end, which interacts with the metal resin and requires competing concentration of an equivalent ion (imidazole usually) upon elution. As described by Yarski *et al*, (2000) Nickel-NTA resin was added directly to the supernatant from the cell extracts and incubated for 3 to 4hrs at 4°C with rocking. The resin was then packed with gravity flow into a 1cm diameter column and washed with a minimum of 10 column volumes with phosphate solution. The protein was then eluted using 100mM imidazole solution (Yarski *et al*, 2000).

Yet another option is using antibody affinity chromatography. Lovgren *et al* (1999) purified PSA using an affinity chromatography column, using 4 mg of immobilised monoclonal anti-PSA antibody. The column was washed with ionic solutions and the retained fraction was eluted with stronger buffers.

Following chromatography, the fraction containing the target protein must be concentrated, preferably using the Centricon kit (Frenette *et al*, 1998, Shimizu *et al*, 1998).

1.4 The metzincins (Zinc-endopeptidases).

1.4.1 Introduction to the metzincins: functions and expression patterns.

A second topic of study in this thesis evolves around two subgroups of the metzincin superfamily, which is one of the four (currently known) groups of Zn-endopeptidases. The metzincins encompass several subgroups, such as the snake-venom adamalysins, the meprins, the astacins, the serralysins, the MMPs (matrix metalloproteases) and the ADAMs (A Disintegrin And Metalloprotease domain). The two latter groups can be subdivided further into classes, such as the stromelysins, the matrixins, the reprolysins and the disintegrin bound adamalysins (Stöcker *et al*, 1995, Basbaum and Werb, 1996, Nagase, 1996, Primakoff and Myles 2000,). Many of the MMPs and ADAMs have the ability to cleave, activate and solubilize important growth factors such as tumour necrosis factor α (TNF α), transforming growth factor β (TGF β), heparin-binding epidermal growth factor (HB-EGF) and interleukins as well as components of the basement membrane (BM) and extracellular matrix (ECM), including collagen type IV and stromelysin (Izumi *et al*, 1998, Vincent *et al*, 2001, Kiyama *et al*, 1999, Rosendahl *et al*, 1997, Roghani *et al*, 1999). Both families are involved in several normal and pathological processes including embryonic development, cell growth and proliferation, inflammation responses, wound repair, multiple sclerosis, arthritis and cancer progression and metastasis (Basbaum and Werb, 1996, Nagase, 1996, Primakoff and Myles, 2000). To date, 20 distinct MMPs and 33 ADAMs have been described, although not all ADAMs are catalytically active proteases. The additional presence of a disintegrin domain within the ADAMs family, which has been variously shown to bind to integrins on cell surfaces further highlights the potential for these ADAMs proteases to be involved in regulation of cell-cell/cell-ECM contacts as well as ECM degradation. Both of these functions are critical for regulating cell migration and invasion – two key stages in the complex process of cancer cell metastasis.

Both MMPs and ADAMs are synthesised as zymogens, and activated by a cysteine switch mechanism (Springman *et al*, 1990); importantly, for those catalytically active members, they share a homologous catalytic motif (consensus HEXXHXXGXXHX) in their N-terminally-located metalloprotease domain. This motif contains a His-triad responsible for the stabilisation of a catalytic Zn-ion (Stocker *et al*, 1995), and a conserved catalytic glutamate (Reinemer *et al*, 1994). Previous studies by Browner and colleagues (1995) suggest the catalytic mechanism of zinc metalloproteases to be a general base-catalysis, carried out by the Zn-ion, glutamate 202 and a putative water molecule. A second, non-catalytic Zn-ion, which is found in the MMPs, is

thought to be structurally important but not catalytically active (Lovejoy *et al*, 1994, Dhanaraj *et al*, 1996a, b).

MMP-3 is central to the enzyme-substrate building approaches carried out in the studies reported in this thesis, and additionally ADAM -9 and -10 are chosen as targets for this study as they share significant proteolytic substrate specificity, and structural and activation characteristics/mechanisms with MMPs. Like several MMPs, bovine ADAM-10 has been shown to cleave wild type collagen type IV in vitro (Millichip *et al*, 1998), and it is likely that human ADAM-10 will also degrade the same substrate. ADAM-9 recombinant metalloprotease domain has recently been shown to successfully cleave fibronectin, gelatins and the insulin B-chain, again demonstrating specific substrate overlap with the MMPs (Roghani *et al*, 1999).

MMP-3, ADAM-9 and ADAM-10 belong to the matrixins, adamalysin group and reprolysin groups, respectively. All these groups retain a common fold (Fig 1.5), built on five antiparallel beta-sheets, a long traversal helix and a central helix which carries the catalytic motif, adjacent to the Met-turn (the structural support for the Zn-chelating system - Stöcker *et al*, 1995, Dhanaraj *et al*, 1996a, b). The reprolysins and adamalysins share also a common lateral amphiphilic helix that supports the beta-sheet quintet, which in the matrixins is replaced by an S-loop that is sustained by a structural Zn-ion (Dhanaraj *et al*, 1996a, b). Indeed the matrixins require three additional structural calcium ions to enable extracellular stability (see figure 1.5), which adamalysins and reprolysins do not need (Dhanaraj *et al*, 1996a, b, Maskos *et al*, 1998). On the other hand, matrixins are totally devoid of cysteine bridges, while adamalysins and reprolysins have 3 to 4 of these (Dhanaraj *et al*, 1996a, b).
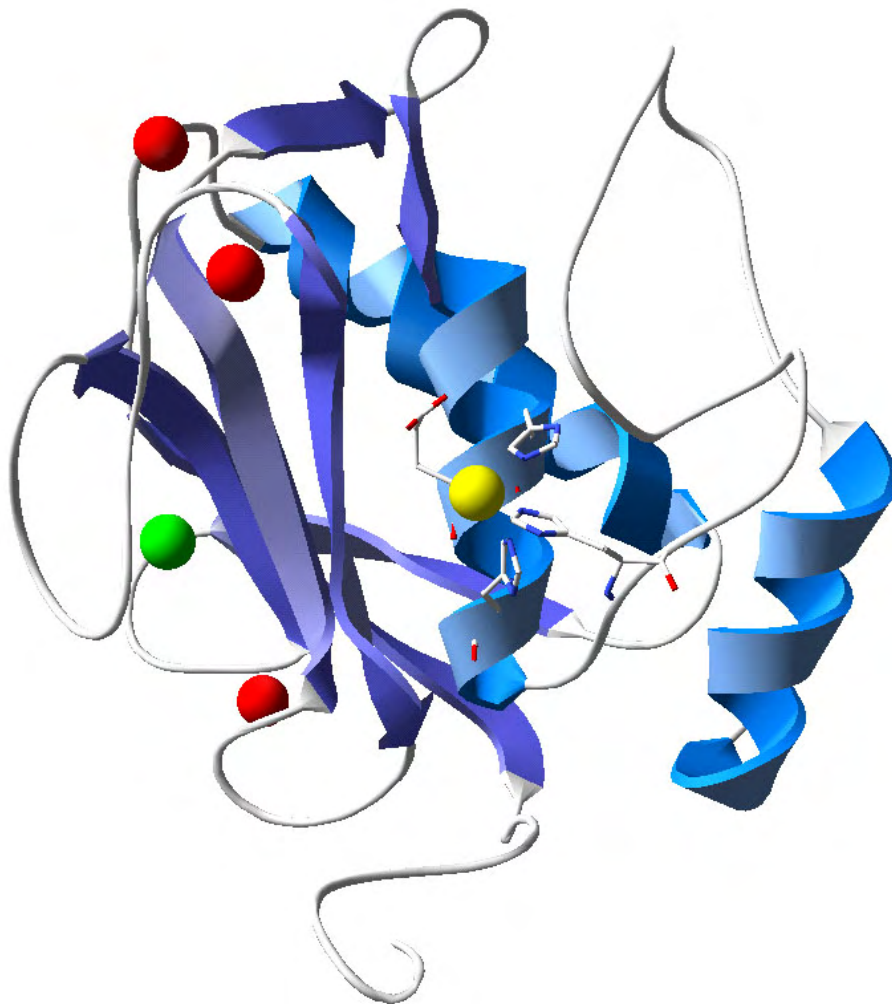
Figure 1.5. MMP-3 shown in ribbon view.

Yellow ion, catalytic Zn chelated by three histidine residues; green ions, structural Zn-ion sustaining the S-loop; red ions, structural calcium ions that sustain the beta-sheet quintet at its turns.

1.4.2   Natural inhibition of metzincins.

The cellular and structural basis of the natural inhibitory mechanism of metalloproteases is far less studied than that of serine proteases. From a cellular basis, metzincins (as for most other proteases) are regulated by a dynamic association and balance with naturally-occurring inhibitors. For MMPs these inhibitors are the tissue inhibitors of the MMPs (TIMPs), which bind to MMPs in a 1:1 stoichiometric manner (Murphy and Docherty, 1992) and also were recently found to inhibit certain members of the ADAMs, specifically ADAM-10 and –17 (Amour *et al*, 1998 & 2000). Throughout the last 12 years inhibition studies have been extensively employed to design a variety of effective synthetic inhibitors specifically against the MMPs,, some of which have been assessed in clinical trials, with limited success, for subsequent use as therapeutic agents (Heath and Grochow, 2000). These inhibitors target mainly the S1' pocket, which is the center of substrate recognition (Dhanaraj *et al*, 1996a, b) and some of these further extend to S2' and S3' positions (Maskos *et al*, 1998). The S-side has not been exploited so far in the reviewed studies, and still remains elusive as to how it acts on the substrate during binding (Johnson *et al*, 2000, Fasciglione *et al*, 2000).

In order to develop a protocol for subsite investigation for metzincins, the only existing enzyme-inhibitor crystal structure of a metzincin, MMP-3&TIMP-1 (Gomis-Rüth *et al*, 1997), was investigated. This dimer (Fig 1.6) shows a totally different inhibitory mechanism from serine proteases, revealing a bidental inhibition (two-fragment inhibition) of one beta-sheet from the inhibitor (Cys1-Pro5), acting like the P'-region of a substrate, and one loop (Cys67-Glu70) which represents a part of the P-region (Gomis-Ruth *et al*, 1997). The inhibitory beta-sheet (Cys1-Pro5) binds in an antiparall manner, via H-bonds, to the substrate-binding beta-sheet, and parallel to a short 2-residue beta-anchor found in all metzincins (Gomis-Rüth *et al*, 1997, Johnson *et al*, 2000). The P-region has not been investigated in an extended beta-sheet conformation.

This mechanism of inhibition presents difficulties in deducing where the substrate orients from the P1'-residue towards the P-region, and how this region is associated with the enzyme. Finding the favourable orientation of the substrate aids in determining specificity-subsites on the S-region of the enzyme. These subsites will be important to map for future inhibition studies.
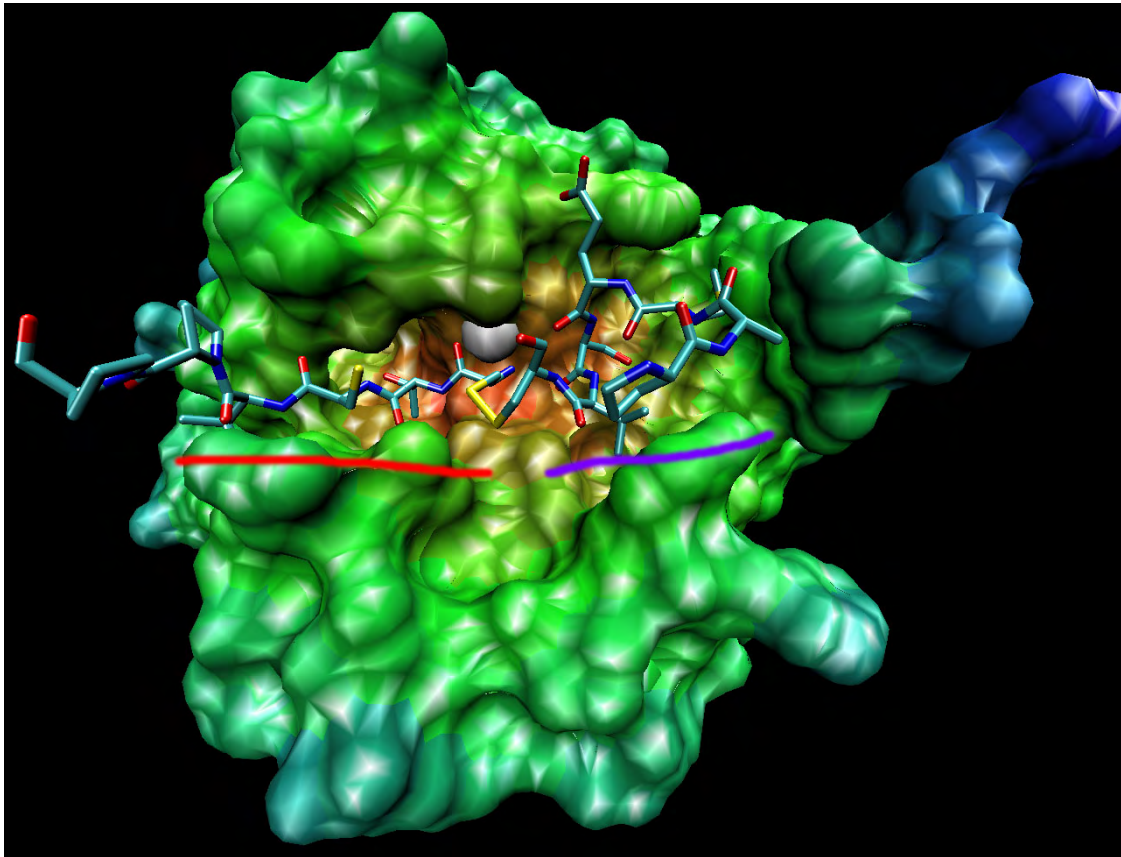
Figure 1.6. Binding mechanism of TIMP-1 to MMP-3.

The inhibiting residues from TIMP-1 are shown in stick mode docked on the enzyme surface of MMP-3 (coloured by depth). The red line designates the P'-region of the inhibitor, while the blue line designates the P'region on the N-terminal side of the scissile bond. The scissile bond does not really exist in this complex: it is replaced by a disulphide bridge (yellow) right in front of the grey coloured Zn-ion. Only the inhibiting residues from TIMP-1 are shown for graphical purposes.

1.5   Computational analysis of proteins

1.5.1   *Homology modelling.*

The central computational procedure in structural-bioinformatical analysis is homology modelling, which is a commonly applied procedure to produce the tertiary structure of a sequence using a template crystal structure. Building a model of a sequence requires a certain percentage of sequence identity between the sequence and the template. Depending on the position of the conserved residues, some models may be modelled down to 25% sequence identity, as for serine proteases (Martì-Renom *et al*, 2002). However, when the sequence identity is as low as 25-40% the correct positioning of the residues from the sequence onto the template (threading) is absolutely vital, through the sequence-to-structure alignment. Indeed, in these cases the superimposition algorithms tend to neglect the secondary structure propensity of the residues and try to satisfy phylogenetic criteria, which do not always overlap with structural information. However, since sequence diversity does not necessarily mean structural diversity (Webster, 2000) the conservation of the secondary structure elements is the primary priority during the threading process (sequence to structure alignment), because of their role in the original folding process of the protein (Hinds and Levitt, 1992). In other words, when threading a sequence on a structure, the conserved helices and beta-sheet observed throughout the relative protein family of the target sequence must be prioritised and maintained in their typical conformation (beta sheet or helix). The fold-variation within protein families is primarily located in the peripheral loops, creating unique signatures for each member of a protein family. These signatures aid in long-range and short range-electrostatic recognition of substrates, and additionally in cellular localization mechanisms in the cell.

In this context, more efficient threading algorithms have recently arisen. These methods take extensively into account secondary structure information from PDB templates from a database (Protein Databank, Westbrook *et al*. 2002), and score the unfamiliar residues according to their propensity for occurring in either beta-sheet or helix, according to the set in the database (Karplus *et al*, 1998, Gough et al, 2001). Applying these threading-engines can be very efficient when the sequence identity is low, although the manual threading is the most important feature during a homology modelling process. This underscores the visual recognition of conserved secondary structure motifs, and their interaction with the environment (typical approaches: Manzetti, 2002). When the alignment is prepared properly, the model building can be initiated.

The model building of a protein sequence applies either the rigid body assembly (Katchalski-Katzir *et al*, 1992), the segment matching (Hart and Read, 1992) or modelling by satisfaction of spatial restraints (Sali, 1995). The first constructs a model from a few core regions and from loops and side chains that are obtained from dissecting related structures. The second method approximates positions of conserved atoms from the templates to calculate coordinates of other atoms. The third uses distance geometry or optimisation techniques to satisfy spatial restraints obtained from the alignment of the target sequence with the template structures (similar to the second method).

1.5.2   *Enzyme-substrate modelling*

Constructing models of enzyme-substrate (ES) complexes relies on using an exemplary enzyme-inhibitor structure as template given that the natural inhibitor mimics strongly the natural substrates. In general, the inhibitory mechanism of natural serine protease inhibitors (similar to most mechanisms of inhibition in nature) mimics strongly the enzyme-substrate complex (Wilmouth *et al*, 2001). Specifically in the case of trypsin-inhibitors, the inhibitor docks the P4-P3' positions to the enzyme surface and avoids cleavage through an internal architecture that prevents the scissile bond from reaching a transition state with the catalytic residues (Mittl *et al*, 1997, Wilmouth *et al*, 2001). Therefore, after homology modelling of a candidate kallikrein, the final construction of the enzyme-substrate model relies on the structural similarity between the modelled holoenzyme and the enzyme from the crystal structure of the template enzyme-inhibitor that are superimposed. Superimposition for serine proteases should rely on the three catalytic residues and the S1-residue from the primary specificity pocket as superimposition targets, which prevents incorrect fit of the enzyme in accord with the binding interface.

The mechanism of inhibition of ADAMs and MMPs is quite different from serine proteases, but still shares the same secondary structure of the main inhibitory segment, the extended beta-sheet conformation. The mechanism of inhibition for metzincins requires a different approach given the presence of a catalytic Zn-ion, and is well described by the complex of MMP-3/TIMP-1 (crystallized by Gomis-Ruth *et al*, 1997). According to their results, the inhibitor bidentally inhibits the binding interface of the metalloprotease, mimicking a cleaved substrate, where the P'-region is aligned in an antiparallel orientation with the substrate binding beta-sheet through backbone H-bonds (Johnson *et al*, 2000), and a part of the P-region is solely represented by a loop from the TIMP-1 inhibitor (this loop is connected to the P'-region through one disulphide bridge near the catalytic centre). This bidental inhibition creates a challenge in modelling

enzyme-substrate complexes for metzincins, where the scissile bond and the entire P-region must be modelled *ab initio.*

1.5.3   Force Fields

Once the model has been built, the structure is subjected to an energy minimization (geometrical optimisation according to the criteria given in a *force field*). Force fields such as CHARMM (MacKerell *et al*, 2001), GROMOS (van Gunsteren *et al*, 1995) and AMBER (Kollman *et al*, 1986) are libraries of physical properties of distances between atoms and angles and torsions around bonds derived from IR and resonance experiments. The physical constants in a force field reproduce therefore the physical qualities of amino acids and other molecules in a multi-atomic system *in silico* (computationally), following the *additivity principle* (Dill, 1997) which is used to compute the energy of the system, whether it is a protein/lipid/glycan *in vacuo* or *in aquo* (simulated in vacuum or solvent). The additivity principle assumes addition of energies of the components of a system to estimate the total energy of the system, a principle that is very successful in most bio-computational approaches, but shows some weaknesses in non-homogeneous systems (as water-lipid interfaces, and protein-lipid interfaces) (Wang *et al*, 2001).

A force field can be described as:

$$E = \sum_{bonds} \frac{a_i}{2}(l_i - l_{i0})^2 + \sum_{angles} \frac{b_i}{2}(\theta_i - \theta_{i0})^2 + \sum_{torsions} \frac{V_n}{2}\left[1 + \cos(n\omega)\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq1}^{N} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq1}^{N} \frac{q_i q_j}{r_{ij}}$$

The two last components are the Lennard Jones potential (vdW interactions) and the Coloumb potential (electrostatic interactions), respectively. The first three components describe the harmonic functions that give the bonds, angles and torsions the allowed intervals (l_i-l_{i0}; length2-length1, etc…) to oscillate as the equivalent oscillation frequencies determined using IR-spectroscopy techniques (Leach, 2001).

Consider for instance the angles: if the angle between two bonds is too small, there will be a geometrical clash between the electrons, which make up the bond. This is not allowed in nature, exactly because electrons repel each other. Therefore, a low angle yields a high-energy state, where repulsion is not energetically optimal. This applies for all the other terms as well, and is the basis for all energy-monitoring force-field libraries that are built on physical data, which are reproduced using mathematical functions. All approaches applied in computational sciences are therefore *knowledge-based theoretical approaches,* except for *ab initio* methods.

A force field can therefore be simplified to the form:

$$E_{tot} = E_{bonds} + E_{angles} + E_{torsions} + E_{LJ} + E_{Coul}$$

This describes what an energy minimization (EM) aims to minimise, E_{tot}. Using a simple statistical stochastic search (familiar with the least-square method) an EM-algorithm approximates a lower energy state, where bonds, angles and atomic positions relative to others are carefully modified according to empirical physical constants. Mainly, there are three methods applied for energy minimization, the steepest descent, the conjugate gradient and the Newton-Raphson method. The difference between these is at the statistical level, i.e., the statistical function used to find the optimal configuration of an atomic system is different (Warshel, 1997). The steepest descent is a faster method, but does not make large changes on the structure in the energy-landscape of a protein or a molecule: it has no or low chance of escaping a local minimum and just refines the position within the local minimum. The conjugate gradient and the Newton-Raphson method are more "brute-force" methods that are used initially to find a local minimum in the energy landscape of the atomic system, these are computationally more expensive and have the ability to "go back to a previous minimum" if no lower minimum is found with the given geometrical alterations (Warshel, 1997).

1.5.4   Loop modelling

Loop modelling is a manual or semi-manual refinement of loops that is applied on regions that are constructed in a model based on zero homology with the template, or parts that have high diversity from a template segment. The ideal conformation is searched either through a database of loops that derive from crystallized structures, or using a stochastic or alternatively an exhaustive search algorithm which varies the torsion angles of the backbone atom in a given set of probable regions on the Ramachandran plot (2D plot of the degrees of torsion around the alpha carbons in a polypeptide – Ramakrishnan & Ramachandran, 1965). The final choice of the loop is structurally optimised using EM, as described in the force field section of the previous paragraph.

1.5.5   Docking

Docking is a procedure that predicts the interactions between a ligand and a receptor. It's a statistical search (a search that excludes unreasonable alternatives defined by a Monte-Carlo approach based on the force-field) for an ideal conformation between the two bodies where

solvation is included on the surface and at the interface between the receptor and ligand (Webster, 2000). The ligand is moved while the receptor is kept immobilized. The ligand is translated and rotated to alternative fits on the receptor's surface. The search for "good" conformations is normally carried out using the genetic algorithm, also called the evolutionary algorithm. This algorithm chooses the best orientation of the ligand to the receptor using a "Darwinistic approach", where the fittest survive. The high-energy conformers where atomic clashes between the two bodies and non-bonded repulsion occur are totally discarded and never used. Only the energetically favourable conformers, where optimal H-bonds and electrostatic interactions are fulfilled are stored and these orientations combined with each other to "produce a stronger offspring" (Goldberg, 1989). Such searches can be confined to an active site if it is known.

Docking can either apply a flexible ligand or a rigid ligand. For the flexible ligand option a given set of rotamer values for the ligand dihedrals are included in a library. Due to increased processing efficiency of modern computers, flexible ligand docking is increasingly used. However, rigid body docking is still used to dock monomers together in order to explore potential dimeric conformations between larger biomolecules and lipid-protein interactions (Vakser, 1996). Docking can be performed at low-resolution for protein models of lower quality, and at high resolution for crystal structures and high quality model (>50% identity) (Vakser, 1995). The topology of the receptor and the ligand is generated, where charges, radii and atom masses are assigned to the coordinates. Once the topology is set, the assembly of the grid is carried out, which represents the units that the ligand can be translated/rotated around the receptor in a 3D space. The grid defines the resolution, few grid units (50*50*50 – normal resolution, 75*75*75 high resolution). The ligand rotates in 3 directions, and is translated in 6 directions. Every time the ligand is translated/rotated, an energy scoring function determines its thermodynamical and free energy of the ligand and the receptor and the added water molecules, and free energy of the complex (Goodsell *et al*, 2002, Osterberg *et al*, 2002); this data is saved to an output file that is at a final stage used to produce the PDB structures of the complexes at the end of the analysis.

1.5.6   *Molecular design*

Inhibitors that are designed for clinical purposes are usually of the competitive class. Designing non-competitive inhibitors requires a) the target enzyme to be allosteric and b) in-depth knowledge of the allosteric mechanism. Therefore the conventional approach is to direct an inhibitor to the active site of a target enzyme (Rich *et al*, 2001a, b), which carries chemical properties that will bind it reversibly or irreversibly to the binding pocket. This requires a thorough knowledge of the binding pocket; an identification of chemical properties that can be

used to distinguish homologue binding-pockets of the same enzyme family, in order to avoid toxic effects (non-specific binding). However, it is not only the quality of the pocket that must be considered for the design of an inhibitor, but also the inhibitor itself; its rigidity & flexibility, hydropathicity, and other chemical and structural factors that affect its behaviour in solution and in a physiological environment.

Depending on the inhibitor, it is known that cysteine residues can be used for cross-linking anchor points on the inhibitor, to aid certain structural motifs in maintaining a fairly rigid shape. Double-bonds within the inhibitor can also be used instead of single bonds, to reduce flexibility of the torsion angles, although the higher susceptibility to oxidation of double-bonds can lead to unexpected effects, inducing a weaker potency (inhibition constant) and most importantly possible unwanted interactions that may have side-effects in the clinical setting. Most importantly, the secondary structure of the inhibitor should be in consensus with the secondary structure of a true substrate in order to yield maximum specificity (Glenn *et al*, 2001). For serine proteases and metalloproteases, beta-sheet conformation is preferred (Perona and Craik, 1995, Gomis-Rüth *et al*, 1997). Therefore the choice of the amino residues, if it is a peptidomimetic inhibitor, should rely on their propensity to occur in such a configuration, and indeed on their electrostatic contact with the surface.

1.5.7   Molecular dynamics

Molecular dynamics (MD) was originally developed as astro-dynamics by NASA in the early 60s to simulate the motion of groups of cosmic bodies such as planets and comets of solar systems, and is heavily used to predict solar and lunar eclipses and possible collisions between comets and planets (Sagui and Darden, 1999). MD is *deterministic,* which means that the original start configuration gives one output after a simulated interval of time, depending on the start motion impulse from the Maxwell distribution (Leach, 2001, Warshel, 1997). In other words, the initial forced directions of movement on the atoms will yield the same final assembly of atoms in the protein/molecule, because of the principle of Newtonian mechanics, which states the force of a moving body remains constant until another force influences it to change direction (Newton, 1687). This is where energy minimization techniques differ from molecular dynamics, which is molecular mechanics based on classic physics. EM methods work on statistical probability (see previous paragraph), while MD works on classical Newtonian mechanics. The only statistical factor is the input (crystal structure or a protein model). Therefore, applying classical physics that is used in every-day calculations in physics, astronomy and mechanics is the core of the strength

of molecular dynamics, which makes it the future of theoretical endeavours in the fields of physics, chemistry and molecular biology (Sagui and Darden, 1999).

A MD simulation requires the input structure to be represented as Cartesian coordinates (x,y,z) or as a Z-matrix (internal coordinates – a tridimensional matrix). A topology of the protein/molecule is generated based on the force field for the given set of molecules. The topology is a file, generated using the force field library, contains charges, atomic sizes and masses, coordination numbers (metal ions for instance) and other physical factors required to simulate the inputted multi-atomic system. The topology is usually confined in a cubic box of customisable dimensions (known as periodic boundary conditions). If simulated *in vacuo* the periodic boundary conditions are irrelevant because there is only vacuum around the protein; however when the simulation is intended with solvent phase, the box must be of conditioned dimensions, in order to limit the number of solvent molecules to a confined space homogenously around the protein.

The positions of the water molecules are optimised with the protein surface using a conventional energy minimization (EM) procedure. The simulation of the atoms can be carried out with restraints around their original positions, using position restraints in order to avoid positional fluctuations of the system in a periodic system. The bonds are normally simulated using algorithms as the SHAKE (Miyamoto and Kollmann, 1992) and LINCS algorithms (Hess *et al*, 1997), which build more on *ab initio* estimations rather than simple mathematical functions (morse or harmonic function). The morse function and harmonic function are mathematical series that represent the distance oscillation between two bonded atoms during the simulation, the difference between the two lies in that the Morse potential (Morse, 1929) takes into account a larger distance interval between the bonded atoms, and therefore requires more time in computation while the harmonic potential is more simplistic and works over a narrower interval, making it more efficient. However neither of these are precise enough (van der Spoel, 2001), and therefore the bond algorithms are preferred. Specifically, in simulation of protein models or X-ray structures, the LINCS and SHAKE algorithms are commonly used (Faraldo-Gómez *et al*, 2002). In refinement of X-Ray or NMR where flexible loops occur, the loop-data generated from the X-ray scattering is often not precise enough, and therefore MD is applied by freezing the well-defined regions in the structure, and simulating the flexible parts with the bond-algorithms *in vacuo*.

Water molecules are usually simulated as rigid molecules; there are mainly 4 common types of waters, SPC, SPC/E, TIP3P and TIP4P (van der Spoel *et al*, 1998). The differences between these are how many, and where the charges of the water atoms are located on the water molecule. Flexible water molecules have also been developed, although these are not suitable for molecular simulations of large systems and only restricted to pre-MD calibration of the solvent-protein interface (Leach, 2001). Time is simulated using an integration step size, which is usually 1 or 2 femtoseconds, over this integration step the motions are simulated in accord with the Newtonian equation of motion, where the temperature for the atoms is simulated using a temperature-coupling algorithm (Berendsen *et al*, 1984) yielding the atomic velocity following a Maxwell-distribution. Proton transfers and electron-excitation are not simulable with MD [but with MD/EVB (empirical valence bond) and MD/QM (quantum mechanics) Warshel, 1997].

Non-bonded electrostatics is usually calculated using the Lennard Jones potential (and occasionally with the Buckingham potential), while the interaction between charged (non-bonded) atoms are simulated using the Coulomb potential (Leach, 2001). LJ and Coulomb potentials are hyperbolic functions that describe respectively the van der Waal interaction and the charge-to-charge interaction over a given distance between two non-bonded atoms. Computationally, these functions depend on distance cut-offs, meaning that atoms outside a given distance are not included in the "electrostatic neighbourhood" at the specific time-frame, however the neighbourhood is normally updated every 5-10 femtoseconds in a MD simulation, so new atoms can easily create new electrostatic environments if it is a "constructive step" towards an energetic equilibrium. Cut-offs are bad and imprecise (Sagui and Darden, 1999), and methods based on the Ewald summation (Ewald, 1921), as particle mesh Ewald (Sagui and Darden, 1999), and particle-particle-mesh Ewald (Hockney and Eastwood, 1981) give much higher quality in the final results, but require longer computational time (Sagui and Darden, 1999). Eventually, a reaction field (Schreiber and Steinhauser, 1992) can be applied instead of the mentioned electrostatic methods, which have shown to be computationally inexpensive though still quite precise (van Gunsteren *et al*, 1995, Daura *et al*, 1998, 2001).

Other integrators of dynamics optional to molecular dynamics are velocity Langevin dynamics and Brownian dynamics, which respectively take into account friction in the Newtonian equation of motion, and diffusion for non-homogenous systems, although these are used to a lesser extent in computational analysis of biomolecular motions (van der Spoel *et al*, 2001).

1.6 Aims

In this study the aim was to purify, study and report structural and functional properties of kallikrein 4 in particular, and other members of the new kallikrein gene family, using biochemical approaches and high-performance computing techniques. The computational methodology applied to the kallikreins is further extended to matrix metalloproteases, which require more information in enzyme-to-substrate binding and present different challenges. This is explained by the fact that the only matrix metalloprotease-inhibitor complex crystallized to date (Gomis-Rüth *et al*, 1997), shows a non-typical inhibition which is more complex when compared to the inhibitory mechanism of serine proteases.

The results gained in this study are intended to provide a broader basis for the understanding of the structure/function relationship of the kallikreins, with an emphasis on substrate affinity (for future *in vitro* assays) and on antigenic regions for *in vitro* and *in vivo* detection.

CHAPTER 2.

2.1 Material and Methods: Biochemical experiments.

2.1.1. Expression system.

Recombinant hK4 was prepared from a construct of the zymogen (pro-hK4), which included the pre-region as well as a V5-epitope and a hexa-His tag at the C-terminus, expressed in a baculovirus expression system (prepared by Willemsen N). The cell culture media containing the extracellularly secreted recombinant hK4 was centrifuged at 1000rpm for 5 min and the supernatant saved. The cells remaining in the culture flasks were scraped off and diluted in 5ml PBS (phosphate buffer saline solution) and subsequently centrifuged at 2000rpm for 5 min; clarified media and cell pellets were stored at –80 °C.

2.1.2 Exclusion of contaminating proteins

Thawed media containing pro-hk4 was dialyzed against 0.1M Acetate buffer pH 4.8 for 20 hrs at 4 °C using dialysis tubes with a size cut-off of 12kDa. After removal of precipitated material (centrifugation for 2 mins at 13000rpm), the supernatant was re-dialyzed against 0.1M Tris buffer pH 8.0, for 20 hrs at 4 °C, and the insoluble pellet removed.

2.1.3 Equilibration and incubation with metal affinity resin.

100μL of Cobalt-resin (CLONTECH) was equilibrated by washing two times by vortexing, centrifugation and re-suspension with 10 column volumes (1cv = 100μL) of wash buffer (50mM sodium phosphate pH 8.0, 300mM NaCl). When using Ni-NTA-resin (QIAGEN) the manufacturers' recommended wash buffer was used in the same procedure (50mM NaH_2PO_4 +300mM NaCl + 20mM Imidazole, adjusted to pH 8.0 with conc. NaOH). 20mL of supernatant from 2.1.2 was incubated at 4 °C with gentle stirring with 100μL of equilibrated resin (CLONTECH TALON or QIAGEN Ni-NTA) for 3 ½ hours.

2.1.4 Elution of hK4 from the column.

The incubated resin (100 μL – column volume) and sample were poured into a chromatography column of 15mL volume. Flow through fractions were saved and examined (see 2.1.5) for residual (unbound) kallikrein, the resin was washed ideally 3 times with 10 column volumes of wash buffers (see 2.1.3) (two and four wash steps were also performed to assess the ideal number of washes). Elution was achieved with 5 column volumes of elution buffer A (50mM Phosphate,

300mM NaCl, 150mM Imidazole) for Cobalt-resin or elution buffer B (50mM NaH_2PO_4+300mM NaCl + 250mM Imidazole, pH 8.0) for Ni-NTA resin.

2.1.5 Analysis of fractions.

All fractions (flow-through, wash-fractions and elutions) were dialyzed against ddH_2O, vacuum-dried to powder and dissolved in minimum volume of sterile water.

2.1.5.1 SDS PAGE

Samples were separated on 15% gels (4ml H_2O, 3mL 40% 29:1 Bisacrylamide, 2.5mL 1.5M Tris pH 8.8, 100μL 10% Ammonium persulphate solution, 100μL 10% Sodium dodecyl sulphate solution and 20μL TEMED). Samples were prepared by mixing 15μL sample with 2μL of mercaptoethanol and 15μL of sample buffer (4mL Distilled water, 1mL 0.5 M Tris-HCl pH 6.8, 0.8mL Glycerol, 0.8 mL. 1.6 mL 10% SDS, 0.4mL beta-mercaptoethanol, 0.2mL 0.05% (w/v) bromophenol blue). The samples were then separated on the 15% PAGE gels at 200V, until tracking dye reached the lower tank buffer. The electrotransfer to nitrocellulose membranes was carried out at 4°C for 45 minutes at 250mA in CAPS buffer (2.2g CAPS, 850mL H_2O, adjusted to pH 10.8 and 150mL methanol). After electrotransfer, the membranes were routinely blocked in 5% Skim Milk overnight (5g Skim Milk dissolved in 100mL tris-buffer saline tween solution), for Western Blots, or were stained (without blocking) with Ponceau Red stain solution to check sample loading.

2.1.5.2 Western blot and stripping of membranes..

The nitrocellulose membranes were incubated on a rocker with primary antibody (mouse) [1/2000 dilution for His-antibody (LifeTech, Melbourne, 3170 Vic, Australia) and 1/5000 for V5 Ab (LifeTech) with 5% skim milk/tris-buffer saline-solution] for 2 hours. After incubation with primary mouse antibody, washing was performed using a tris-buffer saline tween solution (TBS-T) with rocking, twice for two minutes each and three times for five minutes each. The membrane was then incubated for an hour with the secondary antibody labeled with horseradish peroxidase and subsequently washed for 3x5minutes in TBS-T. Finally, the nitrocellulose membrane was stained with Femto-Solution (Genotech, St. Louis, U.S.A.) for 1 minute, and developed against X-ray film for 10sec, 60sec and 2 minutes.

Stripping of membranes was carried out using a solution of 4g SDS and 1.54g Tris dissolved in 200mL dH_2O, with addition of 1.4mL mercaptoethanol immediately before use. The membranes were incubated in the solution for 1hr at 50 degrees in rotating motion.

2.1.5.3 Silver-staining.

A mixture of 0.5 ml 40% sodium citrate, 0.4 ml 20% ferrous sulphate, 9 ml H20 was stirred for 30 minutes. 0.1mL of 20% silver nitrate solution was then carefully added drop wise. The nitrocellulose membrane was immersed in this solution for 5 minutes, washed with dH_2O and dried in the fume-hood overnight (van Oostveen *et al*, 1997).

2.1.5.4 Protein estimation and densitometry.

Protein estimation was carried out using the Bradford method (Bradford, 1976), with a protein staining solution purchased from Bio-Rad (Sydney, Australia).

The silverstained membranes were scanned and densitometric analyses of the TIF images were carried out using ImageQuant® software from Molecular Dynamics (Sunnyvale, CA, USA).

2.2 Material and Methods: Computational experiments.

2.2.1 *Sequence analysis and alignments*

All sequences were analysed according to Manzetti (2002). This method aids in identifying pre/pro-regions, so the sequence of the mature enzyme can be isolated for modelling purposes. Antigenicity plots were generated according to the method by Hopp&Woods (1981). Alignments were computed using the sequence alignment program CLUSTAL W (Thompson *et al*, 1994).

The sequence-based subsite-identification was carried out using sequence-comparison with the alignment between the kallikreins and the mouse neuropsin which has most of its subsites mapped (Kishi *et al*, 1997).

2.2.2 *Protein modelling*

2.2.2.1 Modelling the Kallikreins

All modelling (including loop modelling) approaches were carried out using the SWISS Model suite (Guex and Peitsch, 1999); structural alignments were generated with the SAM-T99 interface (Karplus, 1998) and finally optimised manually. Model building was carried out automatically at the SWISS MODEL server, with ProMod II (Guex and Peitsch, 1999). All energy minimizations were carried out using the GROMOS force field (Scott *et al*, 1999). Templates from the PDB database, for homology modelling of the various kallikreins are shown in Table 2.1.

Table 2.1. List of PDB codes for the templates used for homology modelling of kallikrein members.

Most sequences required 2 templates given their lower sequence identity to existing primary template.

Sequence	Template 1 (PDB id)	Template 2
HK2	1A05	-
HK3	1SGF	-
HK4	1NPM	-
HK5	1NPM	1QBO
HK6	1A0J	1QQU
HK7	1NPM	1SGF
HK8	1NPM	-
HK9	1NPM	1A0J
HK10	1NPM	1SGF
HK11	1NPM	1SGF
HK12	1NPM	1DF2
HK13	1NPM	1SGF
HK14	1NPM	1SGF
HK15	1NPM	1A0J

2.2.2.2 Modelling ADAM-9 and ADAM-10.

The sequences of ADAM-9 and ADAM-10 were retrieved from the Entrez database, accession numbers NP_003807 and NP_001101, respectively. The sequences of the catalytic domains for ADAM-9, L208-I405 (Weskamp *et al*, 1996), and ADAM-10, E217-V452 were modelled using the optimal approach mode in the SWISS Model suite (Guex and Peitsch, 1997), with an enhancement of the structural alignment using the SAM-T99 alignment tune-up interface (Karplus *et al*, 1998) at http://www.cse.ucsc.edu/research/compbio/HMM-apps, followed by a manual threading optimisation. ADAM-9 was modelled in two sessions; a) where 1QUAa was used as singular template for correct disulphide and substrate-binding β-sheet modelling and b) where 3AIG was used to construct another model with a correctly modelled Met-turn. The two resulting models where used to build the final candidate of ADAM-9 with correct disulphides, the substrate-binding cleft and the Met-turn.

Due to a smaller population of suitable templates for ADAM-10 only the structure of TACE (1BKCi) was used for the modelling session. All regions except the reprolysin-loop were modelled with homologue regions.

The catalytic Zn-ion was added to ADAM-9 and -10 by iteratively superimposing each model with the crystal structure of atrolysin-C (PDBid: 1HTDb), and merging it to one layer.

2.1.3   Enzyme-substrate modelling and subsite analysis

2.1.3.1 Kallikreins.
All modelling approaches were performed using the SWISS Model suite (Guex and Peitsch, 1997). The modelling of enzyme substrate complexes requires a different template; the selected kallikrein models were superimposed with the crystal structure of pig kallikrein 1 (with the catalytic triad as target) in complex with the hirustasin inhibitor (Mittl *et al*, 1997). Enzyme-substrate modelling was carried out by merging the Glu26-Arg32 segment from the hirustasin inhibitor to a common layer separately with the chosen kallikrein models. Residues from this segment were changed using a rotamer library incorporated in Swiss PDB Viewer to complementary residues.

2.1.3.2 Metzincins
The structure of the MMP-3/TIMP-1 complex (Gomis-Rüth *et al*, 1997) (PDBid: 1UEA) was retrieved from the PDB databank Westbrook *et al* (2002). Because the enzyme in 1UEA contained heteroatoms, it was replaced with the equivalent structure crystallized by Chen and co. (1999) (PDBid: 1B3D). All residues from TIMP-1 except for the first four N-terminal residues

were then intentionally removed. These four residues were then merged to a common layer with 1B3D, reproducing all details of binding between the N-terminal segment of TIMP-1 and the MMP-3 structure as previously reported (Gomis-Rüth *et al*, 1997).

The substrate was modelled by preserving the P'-region from TIMP-1 and adding five P-residues to its P1'-cysteine residue. The amino acid composition at the P-region was chosen on the basis of the work of Smith *et al* (1995). This experimental P-segment was modelled by adjusting its Phi/Psi angles in an antiparallel β-sheet conformation H-bonded to the adjacent substrate binding β-sheet from the enzyme (exact residue donating H-bonds: Ala167). The substrate for MMP-3 consisted therefore of the following subsites; GPLA↓TCVP (where the GPLA-motif at the P-region derives from Smith *et al.* (1995), and the rest from TIMP-1).

ADAM-9 and ADAM-10 were separately superimposed with the pre-refined MMP-3-substrate complex (see next paragraph). The substrate from the dimer of MMP-3-substrate was then merged to a single layer with ADAM-9, and ADAM-10, respectively. Using a database of rotamer values, the residues on the substrate chain were mutated to create the reported optimal substrate residues: ADAM-9; AALY↓LVCG (Insulin B-chain - Roghani *et al*, 1999) where alanine at the P4 position, derived from other well-hydrolysed peptides from the same study (this combination was applied because charged residues at those positions affected Zn-coordination at the catalytic centre during preliminary simulations). The substrate for ADAM-10 was LAGA↓VMSS (Pro-TNF - Rosendahl *et al*, 1997), where the methionine at the P2'-position was chosen from the results supplied by Vincent and co-workers (2001), since an arginine at this position (found in Pro-TNF) disturbed the catalytic centre by forming a salt-bridge with the catalytic glutamate, and the Gly at the P2 position substituted the original Gln, due to spatial problems. Because of the structural differences between the two ADAMs and MMP-3, a manual adjustment of the substrate was performed to approach the reported H-bonds at the P'-region found in metalloproteases (Johnson *et al*, 2000). Changes in the Phi/Psi angles of the P-segment were applied to the β-sheet area on the Ramachandaran plot to extend the H-bonding antiparallel with the substrate binding β-sheets.

2.1.4   Molecular design

All inhibitors were designed using ISIS Draw from www.mdl.com. These 2D models were translated to 3D coordinates and subjected to a preliminary *in vacuo* simulation for 10ps using the ChemSite package (www.chemsite.net) with the AMBER 5.0 force field (Weiner *et al*, 1986) at a

constant temperature of 298 K. Charges were assigned using the MOPAC package (Stewart, 1990). This generated a plausible start configuration of the drug for the docking purposes.

2.1.5   Docking

Docking experiments on hK4 were carried out using AUTODOCK 2.4 (Olson *et al*, 1996). The system was confined to a box [55.3, 47.7, 23.8], with 60^3 grid units (GU) and a grid spacing of 0.375Å. The ligand was confined to a volume defined by a box centred at the hydroxyl oxygen of the catalytic serine residue (size 5*5*5 GU). 10 runs of the genetic algorithm were done with the default parameters generated by autogrid (Olson *et al*, 1996). The 10 highest-ranking results were visually inspected and analysed with LIGPLOT (Wallace *et al*, 1995).

2.1.6   Unfolding of antigenic segments and the C-terminal construct.

The antigenic segments, the C-terminal and the S-peptide were unfolded from their native state (as found in the kallikrein 4 model) in order to start the folding simulations from an unbiased point. The unfolding of these three oligopeptides was carried out in a simulated annealing session by bringing the temperature from 1000K down to 0K *in vacuo* in a 100ps interval. The C-terminal construct (containing the enterokinase cleavage segment, the V5-antibody and the His-tag) was not modellable with homology modelling and was simply constructed in a helical conformation in SWISS PDB Viewer (Guex and Peitsch, 1997). In order to start from an unbiased conformation for this segment as well, the polypeptide model was unfolded in a 50ps simulation at 700K *in vacuo.*

2.1.7   Refolding of antigenic segments

Re-folding of the antigenic segments (C-terminal and the S-peptide) was carried out in a box of 90nm^3 filled with ~800 SPC water molecules at 300K for 50 nanoseconds, with an integration step size of 2fs. All electrostatics were simulated using a Lennard-Jones and Coulomb cut-off of 1.3 and 0.8 nm. No restraints were used, and chemical bonds were simulated using a harmonic potential.

2.1.8   Molecular simulations of hK4, ADAM-9, ADAM-10 and MMP-3.

MD simulation of hK4 in complex with its substrate was carried out at 300K, in a box of 125nm^3 filled with ~3000 single-point-charge (SPC) water molecules using the GROMACS 3.1 package (Lindahl *et al*, 2001) with the GROMOS 43a2 force field (Scott *et al*, 1999). Kallikrein 4 -plus-substrate was simulated for 1 nanosecond, with short- and long- range electrostatics simulated

using a coulomb cut-off of 0.9nm, a van der Waal cut-off of 0.9nm and a dielectric constant of 1. The integration step size was set to 1femtosecond (fs), and energies were computed once per 100fs. Neighbour searching was updated every 5 fs. All N- and C-termini were assigned neutral state, and histidine residues protonated at the delta-nitrogen. All bonds where restrained using the LINCS algorithm (Hess *et al*, 1997), and temperature and pressure coupling was carried out using the Berendsen scheme (Berendsen *et al*, 1984).

The model of the holoenzyme of kallikrein 4 with its C-terminal construct was simulated with six 500ps intervals as described above for the enzyme-substrate complex of kallikrein 4. The repetition was carried out so a statistically stronger estimate could be drawn of the orientation of the C-terminal under the influence of the kallikrein 4 surface.

MMP-3, ADAM-9 and ADAM-10 in complex with their respective substrates were confined in virtual 125nm^3-boxes filled with ~3200 SPC-water molecules and minimized with the solvent with 100 steps (SD). The simulation was carried out at 300K in a 2.1ns-interval with a time-step of 1fs using the GROMOS96 43a2 force field in the GROMACS 3.1 package. The positions of the water molecules were optimised with a 100-step SD energy minimization. All N- and C-termini were assigned a neutral state and the Zn-chelating histidine residues were protonated at the δ-nitrogen, as observed in crystal structures of MMPs. Distance restraints were applied for Zn- and Ca^{2+}-ions with their respective chelating atoms: Zn^{2+}-εN: 2-2.1 Å, Ca^{2+}-C=O: 2.2-2.5Å. The choice of calcium-chelating atoms in the distance restraints derives from experimental data of the crystal structures of MMP-3 (Chen *et al*, 1999). Chemical bonds were simulated using the LINCS algorithm (Hess *et al*, 1997) and temperature/pressure coupling was carried out using the Berendsen scheme (Berendsen *et al*, 1984). Short- and long- range electrostatics were simulated using Particle-Mesh Ewald method (Darden *et al*, 1993) with a distance to neighbours of 1nm, a van der Waal cut-off of 1nm,, an electrostatic cut-off of 1nm and a dielectric constant of 1.

2.1.8.1 Analysis of trajectories

Subsite-interactions were investigated using the accessory program g_mdmat, random mean square deviations (rmsd) were generated with g_rms and energies computed with g_energy in the GROMACS 3.1 package (Lindahl *et al*, 2001). Residual structural fluctuations were computed with the program g_rmsf from the GROMACS package 3.1.

CHAPTER 3

3.1   Preliminary modelling and purification of hK4

3.1.1   Preliminary computational inspection of human kallikrein 4.

The purification of kallikrein 4 was initially aided by computer modelling to predict characteristics of the protein that could be used to simplify the purification procedure and strategy.

Initially, the effect of the C-terminal construct containing the enterokinase signature (DDDD), the V5-antigen (KGKPIPNPLLGLDST) and the hexa-his tag (HHHHHH) was investigated for its potential in being recognized on the surface by the applied antibodies, and its effects on the enzyme's fold and stability. The risk of hydrophobic clustering would most likely have given a compacted tail that could result in it being inaccessible from the antibodies, which would complicate the detection procedure.

The second task was carried out by estimating the electrostatic potential of the enzyme-surface of kallikrein 4, given by the exposed character of its negatively charged and positively charged residues (27 vs 12) from the model. This approach is commonly applied when the binding pocket of an enzyme or the binding interface between dimers is searched (Swiss PDB Viewer manual, GlaxoSmithWellcome, Geneve, Switzerland), and proves useful in this case to assess the electrostatic surface of the protein for purification purposes.

The first task was carried out by performing six (A-F Fig 3.1) 500ps molecular dynamics simulation in a water phase with the denatured C-terminal construct freely hanging from the enzyme's C-terminal residue. The simulations indicated that the construct has the tendency to stretch itself over the molecular surface of the enzyme, exposing an extended conformation toward the solvent. A graphical assessment of the accessibility of this polypeptide shows this (Figure 3.1) and also indicates that the V5-region tends to cluster itself quite tightly packed due to the composition of several hydrophobic residues. The His-tag and the enterokinase cleavage site on the other hand, occur mostly in exposed and non-clustered conformations. This suggests, in accord with Hopp & Woods (1981), that the His-tag antibody can in most cases be more reliable than the V5-antibody in the detection procedures and presents therefore a higher antigenic potential for immunohistochemical detection.

The second issue, the electrostatic potential, indicated a highly negatively charged surface, as seen in Fig. 3.2. Having predicted the full exposure of the acidic and basic residues to the solvent with the electrostatic potential map (except for Asp189, which is in the S1 cavity), the pI can be directly estimated by accounting for all these acidic and basic residues. The isoelectric point is estimated to be 5 for the kallikrein 4 model *with* the C-terminal extensions of the construct, and 4.71 without. As a direct consequence of this, isoelectric dialysis was investigated for efficacy in yielding a cleaner cell extract to apply on the metal-affinity column, in order to reduce non-specifically bound contaminating proteins.
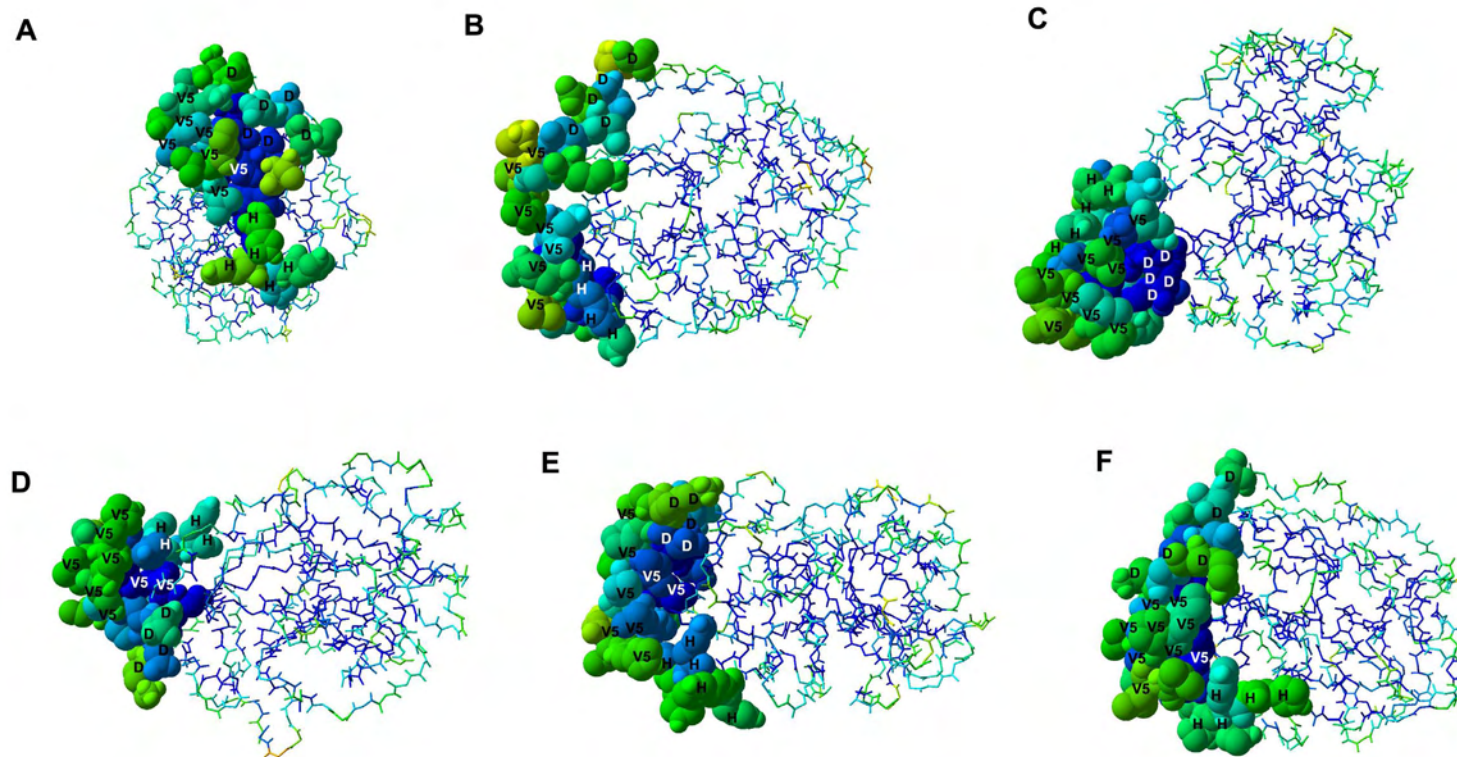
Figure 3.1. Representation of the predicted orientations of the C-terminal construct.
Six molecular simulations of 500ps resulted the following models of kallikrein 4 with the C-terminal construct attached (A-F). The holoenzyme is depicted in backbone representation, while the C-terminal construct is represented in vdW-spheres, all coloured by accessibility in Swiss PDB Viewer (Guex and Peitsch, 1997). Regions marked with "D" indicate the enterokinase cleavage site; the regions marked with V5 indicate the V5 segment, while the region marked with H indicates the hexa-His-tag.
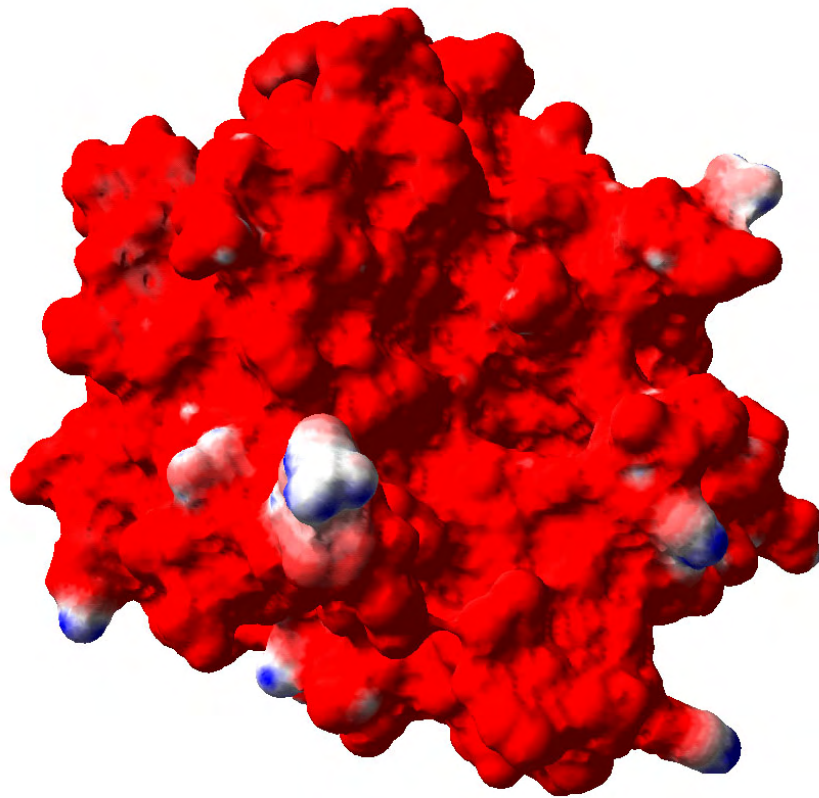
Figure 3.2. Electrostatic surface of the human kallikrein 4 model.

Red regions indicate negative electrostatic environments; blue regions indicate positive and white neutral environments.

3.1.2   Detection of recombinant human kallikrein 4 and antigenic regions.

Figure 3.3 shows the results of an anti-V5 Western Blot analysis of cell culture extracts harvested from a baculovirus system after 24, 48, 72 and 96 hours induction. The size of the protein is estimated around 35kDa, which deviates by about 2kDa from a COS cell transient expression control (personal communication Dr. T. Harvey). This deviation can be explained by alternative glycosylation patterns between insects and mammals. The pre-peptide alone is estimated to have a mass of 2.2kDa (calculated with Protparam – Appel *et al*, 1994). Figure 4 also shows that 72hrs post-infection is the optimal incubation time for highest production of recombinant protein under the given conditions. Interestingly, after 96hrs post-infection no recombinant protein was found, suggesting that other degradative enzymes are released into the cell medium after 72hrs - perhaps removing the His-tag.

The recombinant protein could also be detected with an anti His-tag antibody (data not shown), and its ready availability in the lab and better blotting quality on Western-Blots resulted in it being used subsequently. A hK4-specific antibody, generated against a C-terminal oligopeptide sequence – SEEVCSKLYDPLYHPS – was tested but showed non-specific results in baculovirus expressed hK4 on western blots (N. Willemsen, Y. Dong personal communication and data not shown). Because the identification of antigenic segments of hK4 has proven useful for immunohistochemical detection of hK4 in our lab (Clements, personal communication), and this hK4-specific antibody gave good results in a study assessing endogenous levels (Dong, personal communication), a more detailed analysis of novel antigenic sequences of hK4 was computationally assessed in the next section. This computational analysis of antigenicity may prove useful to identify and subsequently generate better and more specific antibodies.
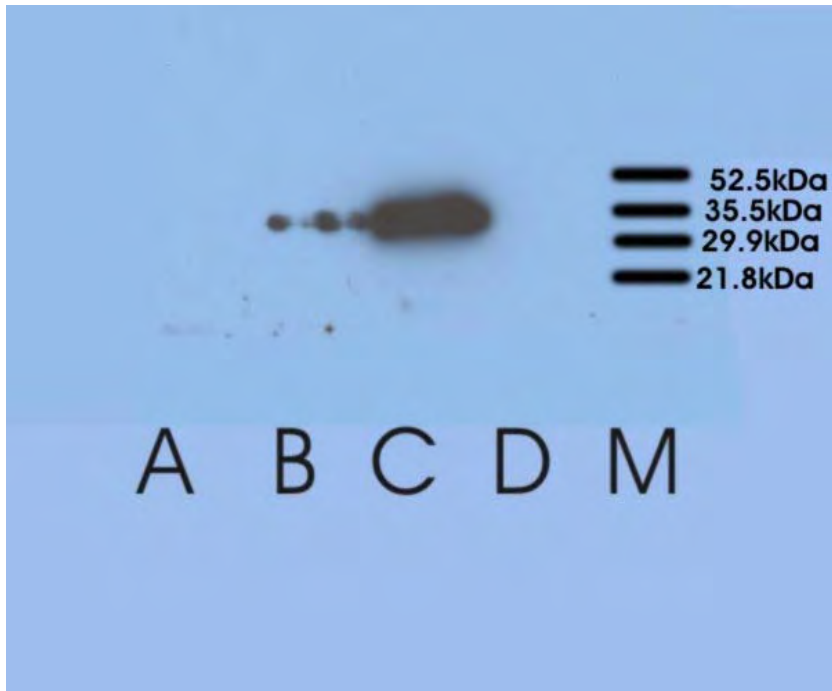
Figure 3.3. Recombinant human kallikrein 4 detected on western blots

A western blot developed after incubation with the V5-antibody. The lanes represent protein expressed by Sf9 cells after various periods of induction: A: 24 hrs; B: 48hrs; C: 72hrs; D: 96hrs; M: Protein marker.

3.1.2.1 Identifying the antigenic regions on hK4 – proposing new antigenic segments for specific antibody generation

hK4-specific antibodies (Dr. T. Harvey and Dr. Y. Dong) against peptides derived from three segments of the human kallikrein 4 sequence where raised in mouse and rabbit and used to distinguish hK4 from other kallikrein proteins in cell extract, especially from PSA and hK2. The antigenic regions chosen for this purpose were Ile31-Gln42 (N-terminal), Leu101-Leu113 (mid-region) and Ser174-Ser189 (towards C-terminus). Due to some negative results while using these antibodies during *in vitro* detection experiments with hK4 (Dr. Y. Dong personal communication), a thorough predictive analysis of the antigenic potential of these segments and an investigation of other possible antigenic regions has been carried out.

In terms of accessibility, these regions are mostly buried; the N-terminal region that gave the worst results *in vitro* is 70% inaccessible according to molecular modeling calculations for a human kallikrein 4. Its sequence composition (**IING**ED**C**S**P**HSQ) shows that it is weakly antigenic - residues in italic are weakly antigenic; bold residues are non-antigenic while others have antigenic potential (Hopp & Woods, 1981). On a similar analysis, the two other regions, **L**SVR*H*PEY*N*R**PLL** and *S*EE**VC**S**KL**YDPLY*H*PS have a higher antigenic potential, but are not believed to be good antigens. This is because: the first is half helical and half coiled, which is not necessarily easy to reproduce *in vitro* as an oligopeptide, while the second has a linear conformation in the enzyme, which is also not easy to reproduce considering its high hydrophobic content (which yields a clustering and aggregative potential in water disrupting structural linearity which is required for binding, Wilmouth *et al*, 2001). Most importantly, neither of these have higher sequence variance from PSA and human kallikrein 2 than 46%, which may yield non-specific interaction with the produced antibodies.

In terms of sequence, the most diverse region in hK4 from all other kallikreins is the PLYH-motif (Nelson *et al*, 1999), however this segment is far too short to produce a functional antibody, and contain only one antigenic residue. A structural investigation of the kallikrein 4 model, showed that the segments Leu83-Ser94 (**L***H*SLEADQE**PG**S) and Ser139-Gly145 (*S*Q**CP**TAG) fulfil requirements to produce suitable antigenic oligopeptides: they have 80% and 85% sequence diversity from hK2 and PSA; further they have 60% and 50% accessible residues and most importantly Leu83-Ser94 has an O-like motif which makes it ideal for structural linking with a disulphide bridge at the N- and C-terminus in order to produce a circular oligopeptide,

structurally similar to the native segment (see Fig 3.4). This attempt to conserve its structure is important for *in vivo* detection purposes.

The second motif (*SQ**CP**TA**G***) has a linear conformation and is indeed very short. However it has 85% sequence diversity from hK2 and PSA, and could be considered as an alternative.

Here, an analysis and comparison is carried out for the suggested antigenic segment Leu83-Ser94 (further named S-peptide) and the C-terminal antigenic region Ser174-Ser189, which gave successful results *in vitro (*Dr. Y. Dong, personal communication).
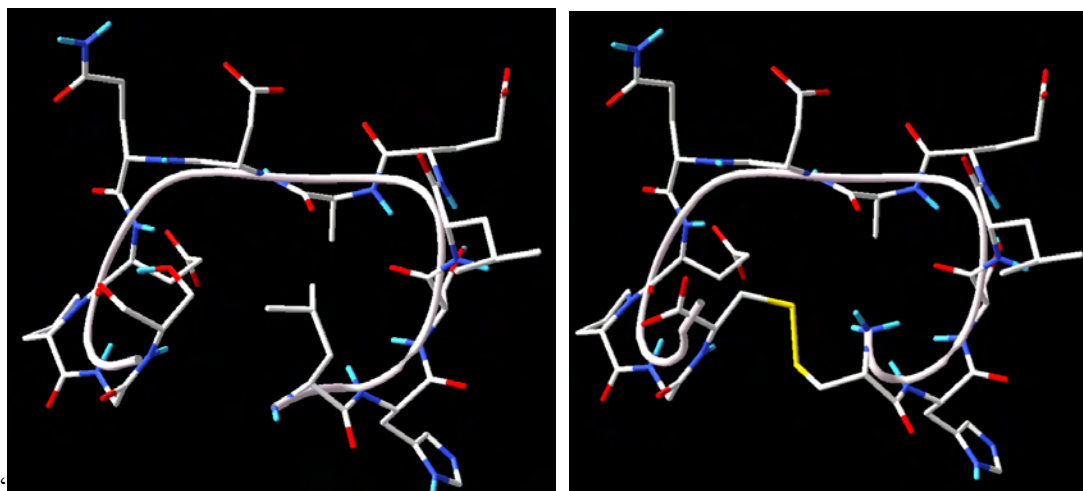
Figure 3.4. Model of the native S-peptide from hK4 (L*H*SLEADQEPG*S*-motif).

The models, left, before energy minimization; right, after energy minimization shows the C/N-ends linked through a disulphide bond after the native leucine and serine were by mutated to cysteine residues.

This analysis relies on assessing the two criteria which are required to successfully anticipate the antigenicity of an antigenic segment: a) the composition of antigenic residues (Hopp & Woods, 1981) and b) the structural flexibility of the oligopeptide in a solvated condition, given that the conformation of an antigen is determinant for antibody-recognition (Manca *et al*, 1993).

Using the (empirically functional) C-terminal antigen as a control here, the assumption made is that the suggested S-peptide will be useful for generating hK4-specific antibodies *only if* a) it is *more* antigenic than the empirically tested C-terminal peptide and b) it is *more* rigid in solution than the C-terminal peptide.

The first criterion is quite uncomplicated to determine: antigenicity plots show that the S-peptide *is* more antigenic than the C-terminal antigen (Fig 3.5). The second criterion is preferably assessed with NMR, but because this is a costly and time-requiring procedure for such a task and this is a preliminary analysis, the predictive power of molecular dynamics simulations is applied here based on other successful methodologies (van Gunsteren *et al*, 2001). The 50-nanosecond simulation shows (by calculating the random-mean-square-deviation (RMSD_ of the oligopeptides) that the S-peptide is approximately half as flexible as the C-term antigen (RMSD plots, Fig 3.6), mainly because of its disulphide bridge linking it to form a circular oligopeptide. Furthermore, Figure 3.6 shows also that after a 50 ns simulation, the S-peptide retains a structural conformation more similar than the C-terminal segment to its native conformation from the human kallikrein 4 model (which indicates structural nativity in solution, important for replicating structural recognition). At last, Figure 3.7 shows that the mean atomic displacement of the C-terminal antigenic segment is 1.5-2Å higher for each residue than for the S-peptide, yet again indicating a predicted higher stability for the suggested S-peptide in solution. **This peptide (LHSLEADQEPGS, preferably in a cross-linked form with N- and C-terminal cysteine) is here suggested as a new candidate for higher selectivity between kallikreins given its higher sequence diversity from hK2 and hK3 (PSA).**
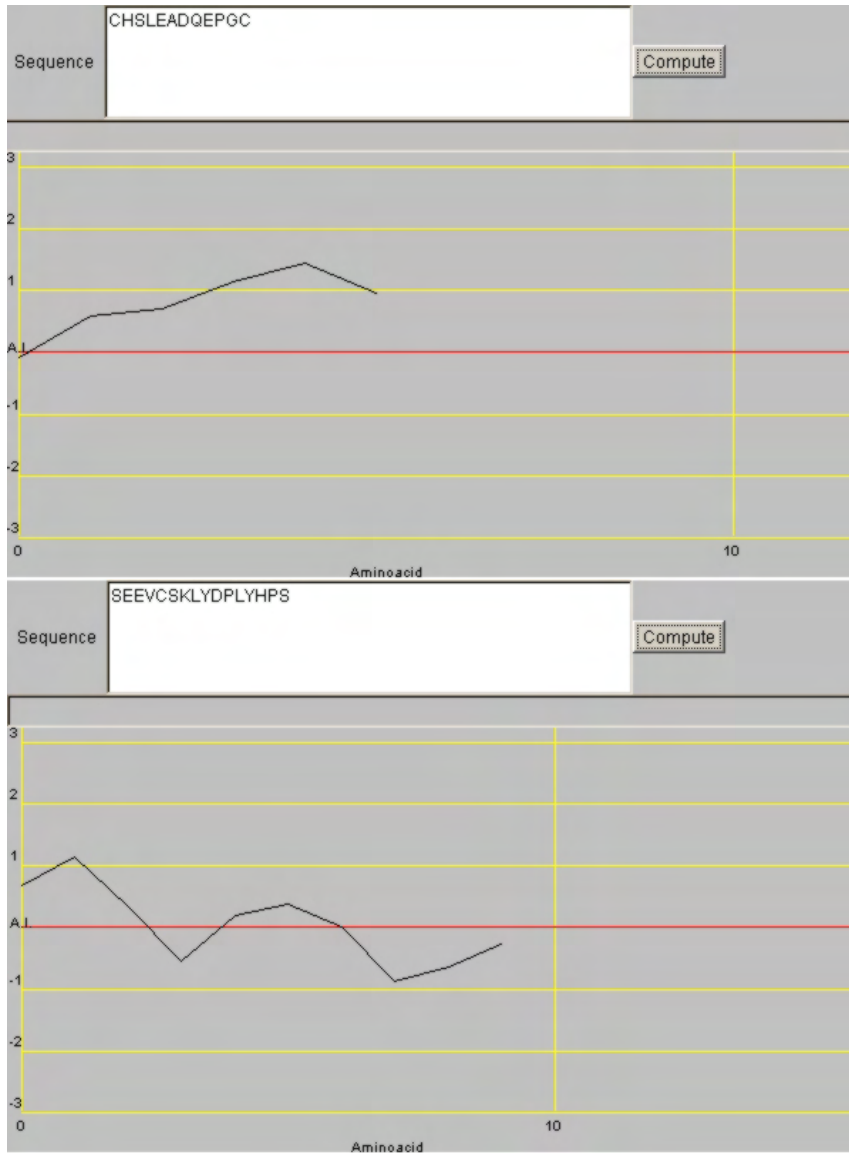
Figure 3.5. Antigenicity plots of two antigenic segments from hK4.

Top, the S-peptide; bottom, the C-terminal antigen (empirically applied oligopeptide for and *in vitro* detection of hK4). Antigenic strength is indicated with the positive values above the null-threshold, which indicates that the S-peptide (top) has a considerably higher overall antigenic potential (Hopp and Woods, 1981).
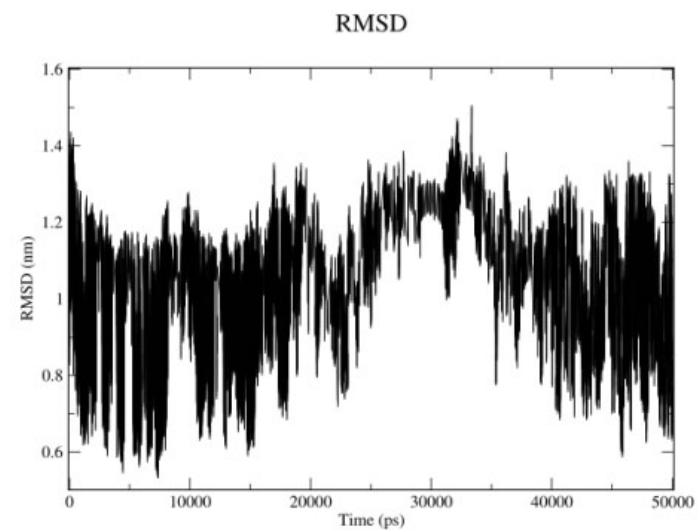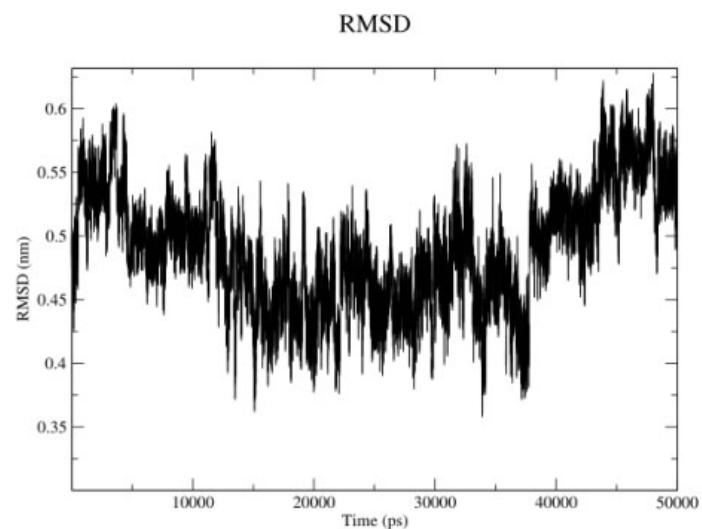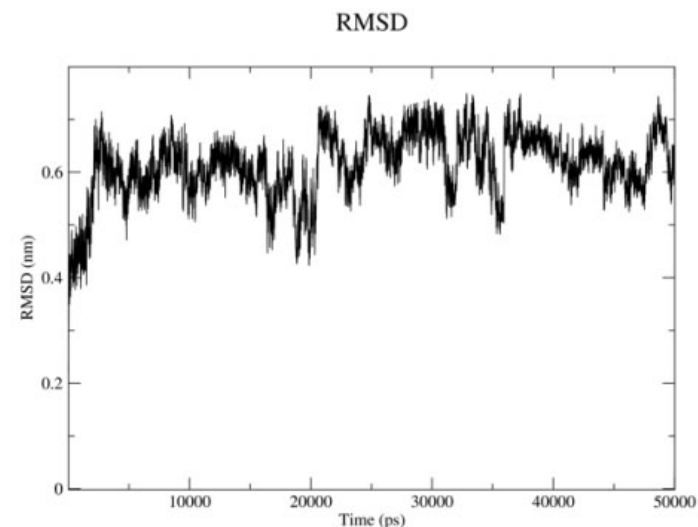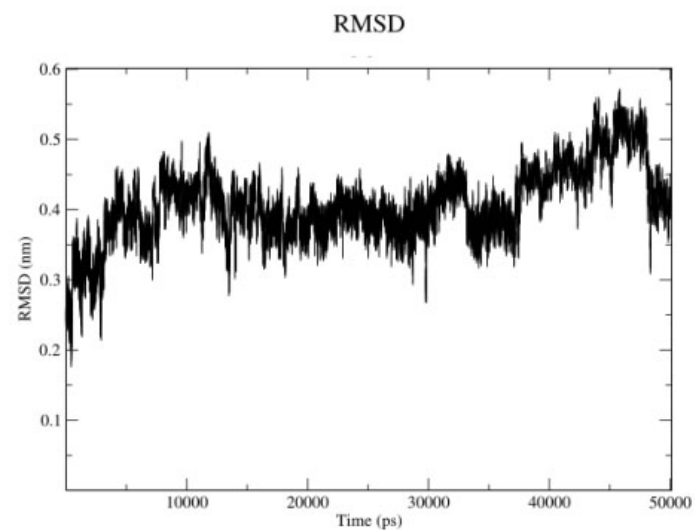
Figure 3.6. RMSD plots for the 50ns simulations of the antigenic segments of rhK4.

Left, S-peptide; Right half C-terminal peptide. Upper plot, the structural fluctuations in RMSD of the S-peptide from its denatured form which was the input of the simulations; Lower plot, structural fluctuations of the simulated oligopeptide from the original conformer as modelled in kallikrein 4. Note the values on the y-axis.

RMS fluctuation
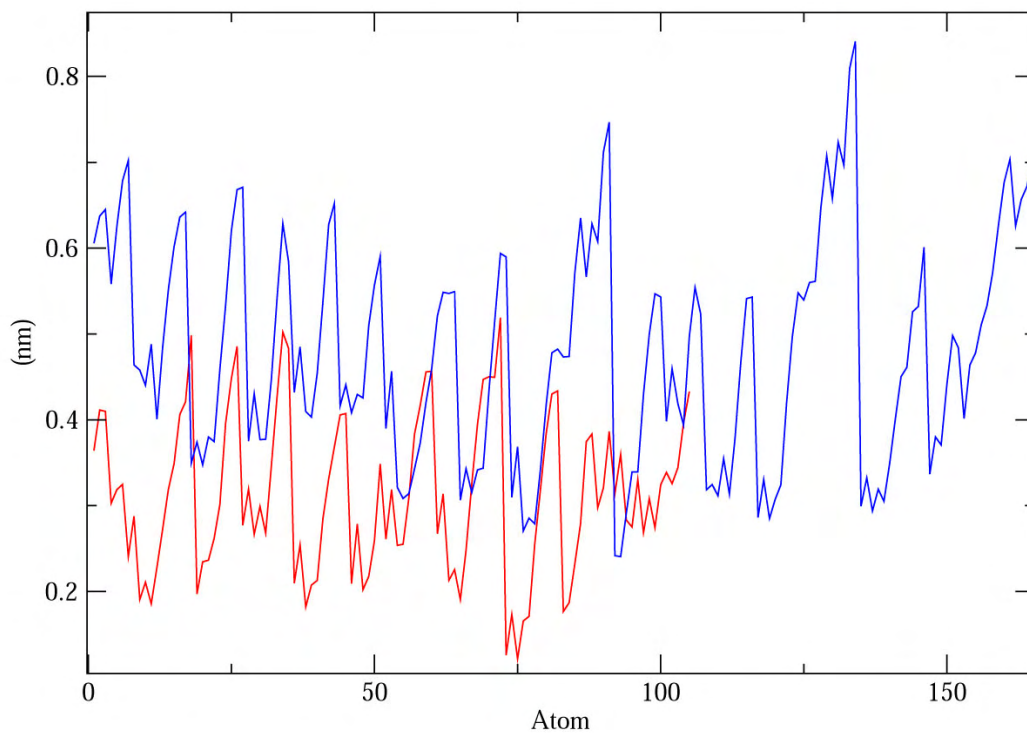
Figure 3.7. Atomic fluctuations computed for the two antigens.

This picture shows how much each atom of the C-terminal antigen (Blue line - 158 atoms on the x-axis) and the S-peptide antigen (Red line – 106 atoms on the y-axis) has fluctuated in nm (y-axis) over the 50-nanosecond simulation. The blue line has a considerably higher structural fluctuation.

3.2   Purification of hK4

3.2.1   *pH fractionation of crude extracts.*

In section 3.1 it was indicated that there was an opportunity for removing superfluous proteins from the crude cell-extract before affinity chromatography, based on the estimated low isoelectric point of kallikrein 4. Indeed, the crude extract (CE) from the media and cells contained a variety of cellular and extracellular proteins from the baculovirus system (see lane 4.8, 8.0 in Fig 3.8 and Table 3.1), and many of these proteins bound non-specifically to the affinity resin, possibly because they are His-rich or negatively charged proteins forming salt-bridges with Co^{2+} from affinity resin. In order to avoid high non-specific binding and exploiting the estimated low pI of this protein (estimated *without* the His-tag), a common strategy in purification of proteins by precipitation at the isoelectric point (Stone and Hess, 1965) was applied to remove superfluous proteins and keep the target protein in solution. Acetate buffer was chosen as buffer for this pH, given its pKa at 4.5 (Solomons, 1996), and the cell extract was dialysed against 0.1M Acetate pH 4.8 buffer for 24 hours. The visible precipitant and soluble proteins were separated by centrifugation and inspected using SDS-PAGE and western blotting with His-antibodies for localisation of the target protein. As seen in Fig. 3.9, the target protein remained in the supernatant while many contaminating proteins were removed by precipitation. The efficacy of purification of this simple step was investigated, using spectroscopic determination of total protein content, where the difference between the absorbance at 320nm and 280nm corresponds roughly to the overall concentration of protein in mg/L in the sample. Table 3.1 shows that dialysis to pH 4.8 + re-dialysis to pH 8.0 removed about 75% of the protein from the CE without loss of target protein.
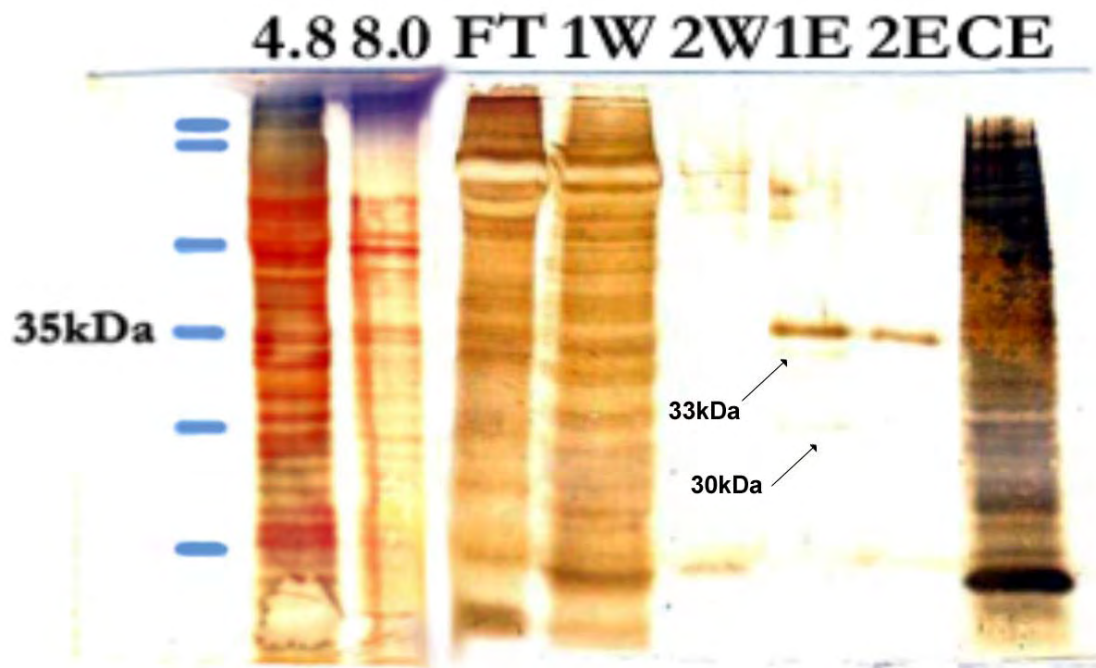
Figure 3.8. Typical silver-stained blot of a purification profile.

This composite membrane shows a colloidal silver-stain of the lanes from the crude extract (CE) though the pellets precipitated at pH 4.8 (4.8) and pH 8 (8.0) and the fractionation process FT→ 2E. FT, Flow-through; 1W, first wash; 2W, second wash; 1E, first elution; 2E second elution. The expected size of purified hK4 is ~35kDa. Bands of this size were observed in the first and second elutions (1E and 2E).
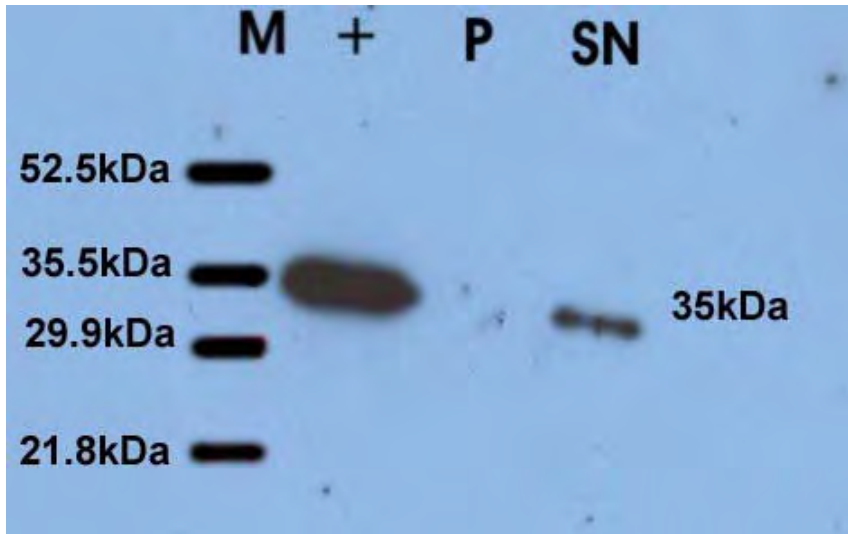
Figure 3.9. Analysis of fractions from dialysis at pH 4.8.

The membrane was stained with His-tag antibody (as the computational approaches in section 3.1 advised), to detect hK4 after 24 hours of dialysis against 0.1M acetate buffer at pH 4.8. hK4 is conserved in the soluble fraction (SN), and absent in the pellet (P). Markers are indicated on the left under "M", showing the detected band at 35kDa aligned with the positive control (+) (COS expressed hK4).

Table 3.1. Typical purification of His-tagged hK4.

Absorbance at 280nm, corrected for light scattering (A_{320}), of original preparation (crude extract) compared with those of various stages though elution from metal ion affinity column. FT, flow through (unbound) fraction; 1W, 2W, wash fractions; 1E, 2E, eluted (high imidazole) fractions.

Sample		$A_{280}\text{-}A_{320}$
Crude Extract		1.06
Supernatant from dialysis 4.8		0.45
Supernatant from dialysis 8.0		0.25
Metal Affinity fractions		
	FT	0.16
	1W	0.07
	2W	0.02
	1E	0.09
	2E	0.008

3.2.2   Affinity chromatography.

The partially purified cell extract from the dialysis treatment was used for affinity chromatography fractionation. The computed exposure of the His-tag from the native enzyme structure indicates that the His-tag will retain an exposed conformation and not be buried between the V5-region and the enzyme.

In initial experiments, several pH values for the wash buffer, which is also used to equilibrate the resin and cell extract prior to fractionation, were tested. Cobalt resin (TALON) was equilibrated with phosphate buffer at pHs of 6.5, 7 and 8. After incubation of the resin in the crude cell extract for 3½ hrs, the efficacy of binding was assessed by comparison of western-blot band intensities. The blots (Figure 3.10) showed that pH 8 yielded best binding conditions, where the flow-through fraction had virtually no recombinant protein. After two wash steps, the protein remained bound and was finally eluted by addition of imidazole. At pH 7.0 the protein still bound, but a stronger band was observed in the flow-through and the first wash step, and less hK4 appeared in the elutions. At pH 6.5 all protein flowed through the resin and virtually no hK4 bound to the resin.

Ni-NTA resin (QIAGEN) was tested at pH 8.0, but showed no binding at all (see Fig 3.11). Although the cobalt affinity column tested at the same time did have some activity in the flow through fraction, a considerable amount of protein was present in the first elution (see Fig 3.10).
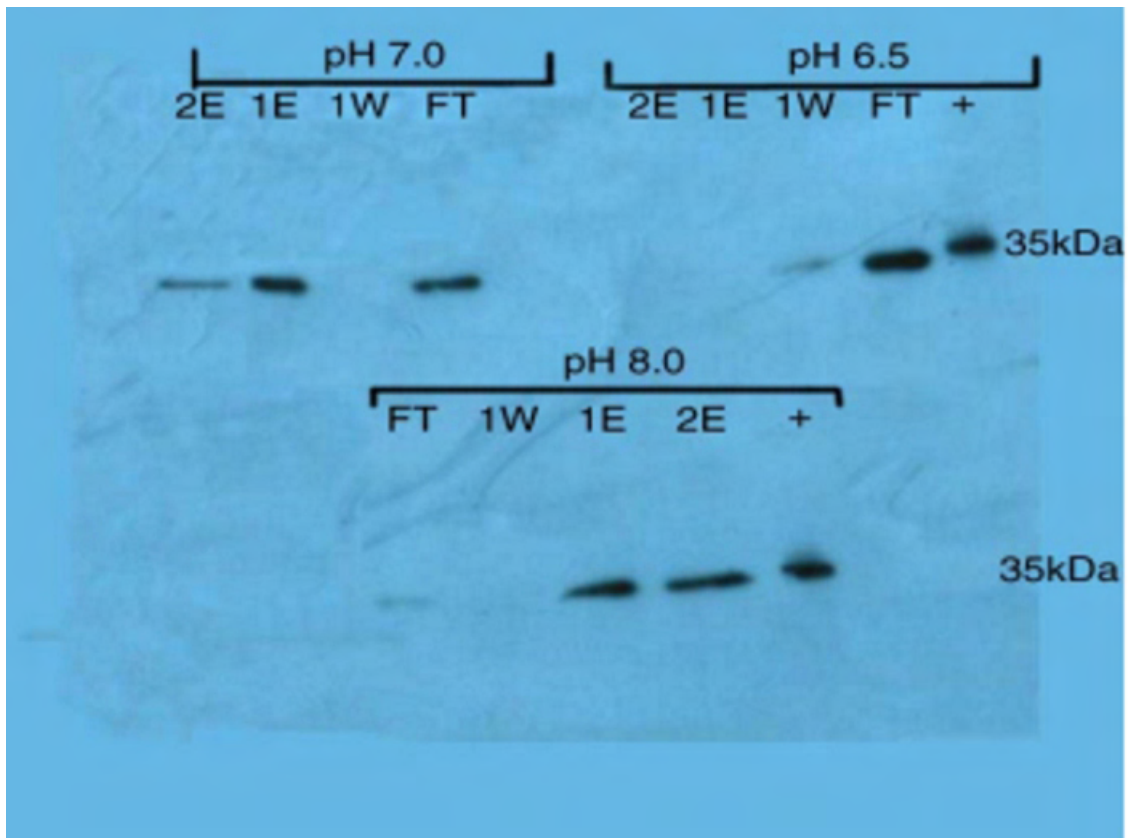
Figure 3.10. Comparison of pHs for a TALON cobalt-resin purification.

Anti-V5 stained composite blot. Note better apparent yields of recombinant protein at higher pH. +, positive control: FT, Fall-through; 1W, first wash; 2W, second wash; 1E, first elution; 2E, second elution.
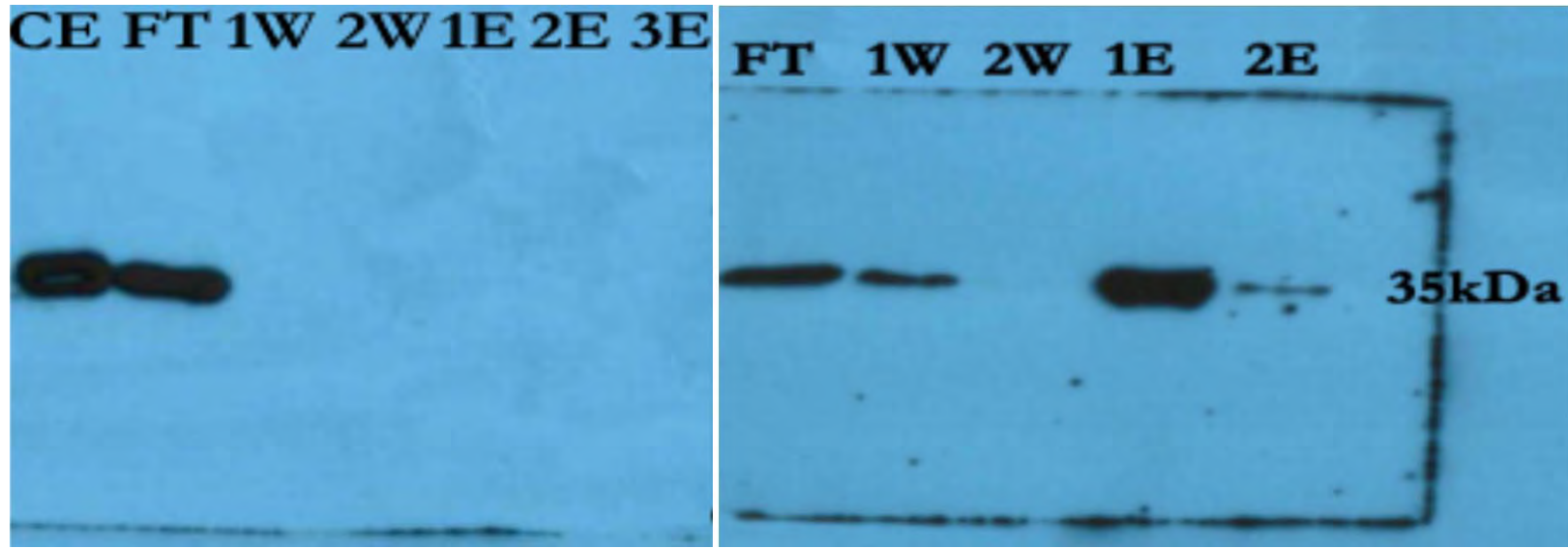
Figure 3.11. Affinities of Ni-NTA and Cobalt resin for rhK4 compared by western blots.

Left blot, fraction from Ni-NTA affinity chromatography (pH 8.0); right blot, fraction from cobalt affinity chromatography (pH 8.0). Abbreviations: CE, crude extract; FT, flow-through; 1W, 1st wash; 2W, 2nd wash; 1E, 1st elution; 2E, 2nd elution; 3E, 3rd elution. Both blots are stained with anti-His-tag antibody.

3.2.3   Analysis of chromatography fractions.

Protein content and purity of the recombinant protein was analysed following silver staining (van Ostveen *et al*, 1997). Densitometric measurements of the silverstained fractions from the affinity chromatography procedure showed a concentration estimate of 90.7% recombinant protein in the first elution (see Table 3.2 and Fig 3.12). The samples were then concentrated by vacuum drying in order to achieve a more precise estimate of the impurities in the elutions. After this step the samples showed a decrease of the overall relative concentration of rhK4 at 35kDa down to 31.2% and the impurities increased to 60-70% (Table 3.2). This trend of decreasing purity with increasing concentration of the samples was observed in the two experiments (Table 3.2), and emphasises the need for another purification step after metal-affinity chromatography, as the content of non-specifically bound proteins is still too high for applications such as generation of antibodies or crystallization.

The concentration of the recombinant protein in the original cell culture extract was also estimated. This was done using two approximations in order to get an estimation range, the absorbance method (A_{280}-A_{320}) and the Bradford method (Bradford, 1976). The first method indicated an approximate concentration of 0.93mg/L of recombinant protein. The Bradford method, using a standard curve based on BSA standards, gave an estimate of 4.81mg/L. Given the densitometric results (Table 3.2), the purity ranges between 90% and 36%, and the concentration ranges therefore between 0.93mg/L - 4.81mg/L (as absorbance method and Bradford method idicate). Compared to Angermann *et al*'s (1992) yield of 5mg/L (Bradford method) with the same expression system makes the estimate quite reasonable.

A questioned to be answered is also how much resin is required to be used with the cell extract. According to the manual for the cobalt – resin (Clontech Labs, Palo Alto, CA), 0.5mL of metal resin binds 2-4mg of hexa-his-tagged recombinant protein. When purifying hK4, the isoelectric dialysis procedure was used to reduce any contaminating proteins in order to keep a minimal amount of resin. Less than 50µL of resin can be used, as low as 5-10µL of resin per experiment for 38mL of crude cell culture extract would chelate the moderate estimated mass of 40-192µg of rhK4 in the given volume of crude cell extract [40mL of CE (0.93µg/mL to 4.8µg/mL of rhK4).

Given these remarks, it is reasonable to conclude that the given concentrations of recombinant protein in the cell culture extract span a reasonable range compared to previous results (Angerman *et al*, 1992) using a baculovirus system. Large-scale systems (10-15L) will provide

enough protein if the conditions are replicated correctly, for crystallization purposes and antibody generation.
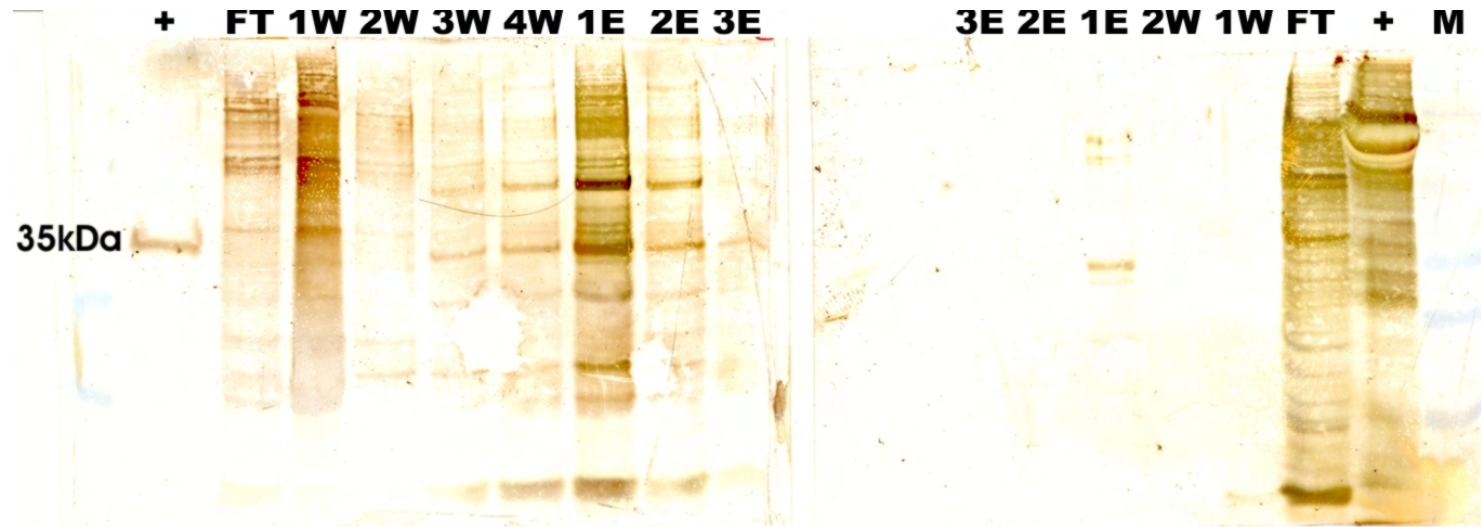
Figure 3.12 Comparison of concentrated and diluted samples from the fractionation procedure.

Left, silver stained membrane shows that the affinity chromatography samples that were concentrated after chromatography and redissolved in a minimum volume (100γL) of sterile water. These samples show the presence of contaminating proteins, are not visible when the samples are not concentrated after chromatography (right). M, marker; +, positive control, FT, flow-through; 1W, 1st wash; 2W, 2nd wash; 3W, 3rd wash; 4W, 4th wash, 1E, 1st elution; 2E, 2nd elution; 3E, 3rd elution. (The two positive controls differ, given that these are two different samples, one is purified hK4 from COS cells, and the other is from unpurified COS cell extract)

Table 3.2. Densitometric intensities of the major bands from the concentrated and diluted fractions from the affinity chromatography.

Recombinant kallikrein 4 is at 35kDa; both samples derive from the first elution of two different experiments. The samples from experiment 1 are equivalent to figure 3.8. The bands from experiment 2 are not shown.

Exp./Sample	Major bands	Intensity
Experiment 1, 1st elution	35kDa	90.7%
	33kDa	4.7%
	30kDa	4.5%
Experiment 2, 1st elution	37kDa	31.2%
	35kDa	36.8%
	30kDa	31.9%

CHAPTER 4

4.1 Computer modelling of the kallikrein families

An understanding of the topology of the kallikreins and specific surface properties that could be involved in interactions is critical for a full understanding of the family. The following particulars develop around substrate affinity, which is the key element in anticipating correct substrates for future enzymological studies, as well as for cellular and inhibitor design.

To begin with, a broader knowledge of the folding of the most distinct members of the kallikrein is introduced, and this knowledge is used as a basis for choosing specific members for future studies. This section is followed by an extensive sequence analysis to map conventional subsites and activation signatures. These results are also important to distinguish in greater detail between typical trypsin/chymotrypsin-like members of this pivotal protease family. However, because sequence analysis alone is often not enough for anticipating substrates, a structural computation of enzyme-substrate models, and an assessment of chemical and spatial complementarity at the binding interface is carried out to increase the precision of substrate-binding preferences, especially for kallikrein 4. Finally, these results are further used to dock inhibitors, which are designed here on the basis of the surface properties, which may function to inhibit kallikrein 4 *in vitro*.

4.1.1   *Modelling distinct kallikrein members: Loops and diverse motifs.*

The loops in serine proteases play major roles in long-range and close-range electrostatic interactions with the substrate, and interaction with other factors (Perona and Craik, 1995). The identification of the chemical composition and location of the loops in the kallikrein family aids in distinguishing these from each other, in terms of their roles in substrate binding. In order to examine similarities and differences between the most divergent members of the kallikreins, models of these sequences where built. Kallikreins 1, 2 and 3 (PSA) have the distinct kallikrein loop that precedes the catalytic Asp102 (see Fig 4.1 and 4.2), and which plays a role in the substrate selectivity at the P-region (Coombs *et al*, 1998a).
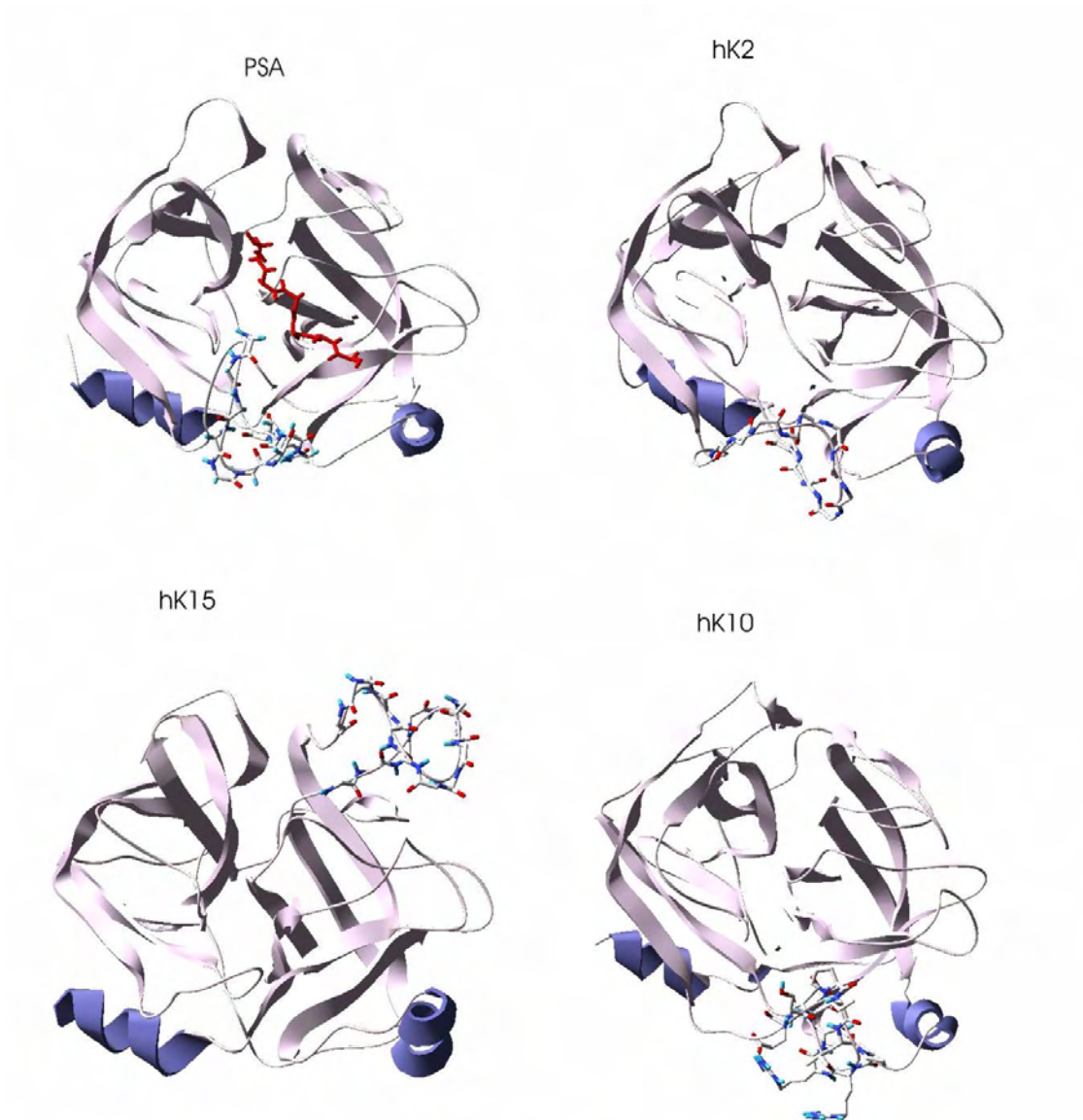
Figure 4.1. A representation of the structures of the most different kallikrein sub-folds.

The distinct loops are represented with a ball-stick representation for each of the listed members. PSA (Coombs *et al*, 1998a) is represented with its hypothetical substrate (coloured red), because this illustrates its adjacent position to the kallikrein loop. The substrate position is the same for all trypsin- and chymotrypsin-like proteases.

Figure 4.2. The 15 members of the kallikrein family aligned with chymotrypsin and trypsin (two bottom rows).
Black line under alignments delineates high or low identity. Pink-coloured triplets indicate the initiation of the mature sequences; the light-blue regions show the most pivotal subsites; the red dotted residues indicate the catalytic triad; the yellow regions confine the named loops; green block shows the conserved residues that provide their backbone atoms to freeze the P2-P4 residues in place on the S-side of the enzyme.

The kallikrein loop is not conserved in the other kallikreins, although hK10 has a similar, but shorter insert at this position. The composition of this segment (SGPILPRR) is quite different from the conventional residues in the classic kallikrein-loop (2-3 basic residues + 4-6 polar residues in hK1, 2 and 3). The residues from this segment may, in an equivalent manner to the kallikrein loop, influence the substrate selectivity at the P3/P4 positions. The remaining members of the kallikrein, which are not shown in Fig 4.1, do not have any distinct traits at this position (Fig 4.2 – kallikrein loop region).

A unique motif is the kallikrein 15-loop, inserted between the conserved Gly145 and Leu 158 of kallikrein 15 (see Fig 4.1, 4.2). This loop, which is examined further in the next section has no homology with the loops from the SWISS loop database (Guex and Peitsch, 1997), and given its absent homology with loop-templates, its true orientation may only be solved with experimental approaches. Using modelling results (see Fig 4.1), it is predicted to overhang the P'-region and may therefore play a role in substrate recognition, but proof of this awaits crystallographic examination.

The target protein of this project, hK4, has no unique or large deletions/inserts compared to the conventional kallikrein fold, except for a very short PLYH-motif (P171-H174) (also noted by Nelson *et al*, 1999 – Fig. 4.2), which, as we shall see in section 4.1.4, plays a role in sidechain interaction with P4 from the substrate.

4.1.2   Modelling subsites of the kallikreins using multiple sequence alignments.

An in-depth analysis of the residues that play a role in substrate specificity will hopefully lay more knowledge for therapeutic approaches . As a general comparison, the high similarity among the kallikreins with their relatives (trypsin and chymotrypsin) (see Fig 3.14) is used here to initially map some differences based on sequence alignment and comparison with existing structural data.

S1 subsite: The aligned kallikrein sequences (Fig 4.2, see also Fig 1.2) show that Asp189 at the bottom of the S1-pocket is very conserved among the kallikreins. hK15 has a glutamic acid residue at this position, which is slightly longer than the normal Asp in the S1 cavity. Considering the chemical complementarity of sidechains in proteins, glutamic acid interacts well with both arginine and lysine, but slightly better with the latter (Hinds and Levitt, 1992), which may indicate a preference for Lys at P1 in the natural/optimal substrate of kallikrein 15.

hK9 has a glycine at the S1 position, which could host a phenylalanine or a tryptophan side chain by hydrophobic attraction. Because hK9 has also a glycine at 216 and a serine at 190, which are semi-conserved with Atlantic crab collagenase in addition to the rare glycine at S1, the specificity of a kallikrein with these features is expected to be broader than for normal trypsin-like enzymes (Perona and Craik, 1995) and probably directed to structural proteins. Another unusual subsite is found in hK7 that has an asparagine at S1 position. Considering the molecular architecture of asparagine, it would make a good H-bond with a tyrosine's hydroxyl group, indicating chymotryptic selectivity, which is also supported by Hinds and Levitt (1992), Skytt *et al* (1995) and Hanson *et al* (1994). The activities and subsites are listed in Table 4.1.

S1' subsite: The S1' position in the kallikreins is dominated by residue 40 and from the alignment seems to be quite variant among the fifteen members. Indeed this residue creates subgroups within the family, and is the second most important specificity determinant. The interactions at S1' are analysed extensively in section 4.5.

The non-specific subsites on the enzyme, S214, W215 and G216, which bind the substrate at the P2-P4 residue via backbone H-bonding, are naturally also conserved among the kallikreins (Fig 4.2). These three residues confirm a preserved S-region with trypsin and are consistent with binding of the substrate for the kallikreins that is predominantly fixed via H-bonds at the P-

region. However, the central sidechain interactions that determine the selectivity of the enzyme are mapped in the enzyme-substrate complexes reported in section 4.4.

Table 4.1. Subdivision of the kallikrein family according to their primary specificity.

Conventional residues that are used to identify the activity of the serine proteases are used here to subdivide the kallikreins according to their specificity.

Protein	Primary specificity (S1)	Perona and Craik's (1997) elastase check.	Activity	Interacting P1 (Coombs *et al* ,1999)	Ref.
Trypsin	**D at S1**	**G at 216, G at 226**	**Trypsin**	**R or K**	**Koivunen *et al* (1989)**
HK1	D at S1	G at 216, G at 226	Trypsin-like	R or K	Bridon and Dowell (1994)
HK2	D at S1	G at 216, A at 226	Trypsin-like	K pred. and R	Bridon and Dowell (1994)
HK4	D at S1	G at 216, G at 226	Trypsin-like	R or K	Nelson *et al* (1999)
HK5	D at S1	G at 216, G at 226	Trypsin-like	R or K	
HK6	D at S1	G at 216, G at 226	Trypsin-like	R or K	Yousef *et al* (1999b)
HK10	D at S1	G at 216, A at 226	Trypsin-like	K pred. and R	
HK11	D at S1	G at 216, G at 226	Trypsin-like	R or K	Mitsui *et al* (2000)
HK12	D at S1	G at 216, G at 226	Trypsin-like	R or K	
HK13	D at S1	G at 216, G at 226	Trypsin-like	R or K	
HK14	D at S1	G at 216, G at 226	Trypsin-like	R or K	
HK8	D at S1	G at 216, G at 226	Trypsin-like	R or K	Shimizu *et al* (1998)
HK15	E at S1	G at 216, G at 226	Trypsin-like	K pred. and R	
Chymotrypsin	**S at S1**	**G at 216, G at 226**	**Chymotrypsin**	**Y, W, F or L**	**Tomita *et al* (1989)**
HK3	S at S1	G at 216, S at 226	Chymotrypsin-like	Pred. Y, W, F or L	Lilja and Lundwall (1987, Kurkela *et al* (1995) Cooombs *et al* (1998), Robert *et al* (1997), Peter *et al* (1998), Hinds and Levitt (1992).
HK7	N at S1	G at 216, G at 226	Chymotrypsin-like	Pred. Y. Also W, F or L.	Skytt *et al* (1995) and Hanson *et al* (1994)
Collagenase	**G at S1**	**G at 216, G at 226**	**Collagenase**	**L, A, I**	Perona and Craik, (1995)
hK9	G at S1	G at 216, A at 226	Collagenase-like	**L, A, I**	Hinds and Levitt (1992) chemical complementarity matrix.

4.1.3   Cleavage/activation regions and patterns of the kallikrein zymogens.

Serine proteases generally have a zymogen peptide, which is processed (removed) by other proteases at certain cleavage sites to produce the mature enzyme. The kallikreins follow the same pattern of activation, although most have unknown cleavage sites and only a few have been confirmed experimentally (see Table 4.1). Bioinformatical methods have been applied to predict such sites, by especially Yousef *et al* (1999a,b,c,d) and Nelson *et al* (1999). The reliability of such methods is questionable, because these methods rely solely on sequence and, for instance, do not take into account conserved phylogenetic patterns within a protease family as in the case of the kallikreins, which considerably simplifies such an estimate.

Relying on experimentally confirmed activation patterns (Table 4.1) and using these as reference points, a conserved cleavage pattern was identified in the alignment. At the cleavage sites of chymotrypsinogen, hK1, hK2 and hK3 an "SRI-triplet" was found (Ser-Arg-Ile). At the cleavage site of trypsinogen a DKI-triplet (Asp-Lys-Ile) was found, see Table 4.2. These activation sequences are typical P2-P1' positions for activation by trypsin-like enzymes (Light and Janska, 1989). Accounting for that, the activator of most of the kallikrein zymogens must be a trypsin-like enzyme, and most of the kallikreins carry this selectivity. It is conceivable but not proven that kallikreins activate each other in a cascade pathway, a theory which has been proposed (Willemsen, personal communication).

It can be interpreted that two cleavage patterns for these proteins have evolved during evolution: the chymotrypsin-like triplet, and the trypsin-like triplet, respectively Ser-Arg-Ile, and Asp-Lys-Ile. The SRI- triplet places an arginine in the specificity pocket (P1 = Arg) of the cleaving enzyme, and is well conserved in the alignment. The serine residue (P2) is the most conserved in the triplet, which is a typical P2-subsite for activation by trypsin enzymes (Light and Janska, 1989). Isoleucine as the P1' residue, is also preserved, invoking the crucial role of the hydrophobic P1 residue, also a typical signature of trypsin substrates (Perona and Craik, 1995). The varying residues around the preserved S*R*I-triplet (Fig 4.2) function to diversify the P4, P3, P2', P3' positions, thereby yielding a more sensitive and specific control of activation hand in hand with expression patterns and cellular localization of these enzymes and their activators. The variations follow an acceptable pattern in the case of serine to threonine, isoleucine to valine, alanine to leucine.

Table 4.2. Conserved activation patterns in the kallikrein family.

The P2-P1-P1' region is the SRI-triplet in chymotrypsinogen, hK1, hK2, hK3 and DKI-triplet in trypsinogen. Some acceptable variation has been found at the P2 and P1' position, in the other kallikreins, but the P1 position (the most crucial for specificity) is preserved in 7 out of the 8 unknown cases. The asterisks in the reference indicate speculation from previous groups that support this conserved cleavage pattern.

Protein	SRI-triplet (P2-P1-P1')	Residues	Reference
Chymotrypsinogen	SRI	32,33,34	Tomita *et al*, 1989.
HK1	SRI	23,24,25	Baker and Shine, 1985.
HK2	SRI	23,24,25	Riegman *et al*, 1992.
HK3	SRI	23,24,25	Gaultier *et al*, 1993.
HK5	SRI	65,66,67	Eckholm *et al* (2000)
HK11	TRI	20,21,22	
HK9	TRA	21,22,23	
HK10	TRL	40,41,42	Luo *et al*, 1998.*
HK4	SQI	29,30,31	Stephenson *et al* , 1999.*
	DKI-triplet (P2-P1-P1')		
Trypsinogen	DKI	22,23,24	Koivunen *et al*, 1989.
HK7	DKI	28,29,30	
HK8	DKV	31,32,33	Shimizu *et al*, 1998, and Kishi *et al*, 1999.
HK15	DKL	20,21,22	
HK14	NKI	23,24,25	
HK6	NKL	20,21,22	
HK12	PKI	20,21,22	
HK13	SKV	23,24,25	

In the second group seen in Table 4.2, the DKI-triplet has a conserved lysine at P1. A lysine residue at P1 is also susceptible to interaction with an S1 residue in a trypsin-like enzyme (Davis *et al*, 1985, Coombs *et al*, 1999). The kallikreins placed under trypsinogen in table 4.2 all have this conserved lysine, and vary to some extent in the P2 and P1' positions. The Isoleucine residue is more preserved as the P1' in the DKI-triplet than in the SRI-triplet group, indicating a possible affinity for the activating enzymes of this group toward the isoleucine residue as P1'. Furthermore, the mutations follow an acceptable pattern, where Isoleucine has mutated to valine and leucine in hK8, hK6 and hK15 respectively. Both these residues are hydrophobic, but have a rather different molecular architecture than isoleucine, which emphasises small changes of the P1' residue's affinity toward the activating enzyme.

Kallikrein 10 has also a conserved lysine at P1 (cleavage reported by Luo *et al*, 1998), but has an interesting feature. This kallikrein has a propeptide very different from the others. On the alignment (Fig 4.2), it is obvious that this kallikrein has an insert within the actual signal peptide that the others don't have. This forces the SRI-triplet to be moved 7 residues upstream from the standard SRI/DKI triplets region (where all the other kallikreins, as well as chymotrypsinogen and trypsinogen are placed). This special feature might suggest a very different secretion pattern for hK10, which indeed correlates with its special structural features near the "kallikrein-loop" (section 4.1).

In the case of hK4 a glutamine residue has replaced an arginine during evolution, which is a non-conserved mutation. This position, previously noted by Stephenson *et al* (1999) is another unusual detail, since cleavage at Gln as P1 is not typical for chymotryptic/tryptic enzymes. However, this residue was investigated and shown to be a rare but feasible P1-residue for activation by a chymotryptic enzyme, and was suggested to be candidate for cleavage by PSA (Coombs *et al*, 1998a and Isaacs*et al*, 1997).

On reflection, it is very likely that all kallikrein-zymogens except for hK4, are activated by trypsin-like enzymes (Table 4.3), possibly each other in a cascade-fashion (Willemsen, personal communication). Ideal fits between activation segments on the zymogens and the binding cleft from the processing enzyme, in concert with localization experiments would supply useful information to investigate and eventually propose such a cascade. However, more than relative sequence analysis can be done *a priori*, in order to predict more about substrates and the

differences of substrate affinities among some kallikreins. The next sections illustrate such methodologies and results gained specifically around the target protein, hK4.

Table 4.3. Table of activation positions and compatible processing enzymes.

The conserved P2-P1-P1' pattern shows common cleavage patterns, distinguishing two groups as being trypsinogen-like, and chymotrypsinogen-like cleaved.

Protein	Cleavage	Activated by	Primary indication	Secondary Indication	Ref.
Chymotrypsinogen	R33/I34	Trypsin-like	Chymotrypsg-triplet	Alignment	Tomita *et al* (1989),
HK1	R24/I25	Trypsin-like	Chymotrypsg-triplet	Alignment	Baker and Shine (1985) Beduschi *et al* (1998) Podlich *et al* (1999) Lovgren *et al* (1997)
HK2	R24/I25	Trypsin-like	Chymotrypsg-triplet	Alignment	Riegman *et al* (1989), Schedlich *et al* (1987), Beduschi *et al* (1998) Lovgren *et al* (1997). Angermann *et al* (1992)
HK3	R24/I25	Trypsin-like	Chymotrypsg-like-triplet	Alignment	Lundwall, A. (1989)
HK4	Q29/I30	Chymotrypsin-like?	Unconserved mutation at P1, Chymotrypsg.-like.	Alignment	Stephenson *et al* (1999)
HK5	R66/I67	Trypsin-like	Chymotrypsg-triplet	Alignment	
HK9	R22/A23	Trypsin-like	Chymotrypsg-like-triplet	Alignment	
HK10	R42/L43	Trypsin-like	Chymotrypsg-like-triplet	Alignment	
HK11	R21/I22	Trypsin-like	Chymotrypsg-like-triplet	Alignment	
HK13	R24/V25	Trypsin-like	Chymotrypsg-like-triplet	Alignment	
Trypsinogen	K23/I24	Trypsin-like	Trypsg-triplet	Alignment	Koivunen *et al* (1989) Perona and Craik (1995)
HK7	K29/I30	Trypsin-like	Trypsg-triplet	Alignment	
HK12	K21/I22	Trypsin-like	Trypsg-like-triplet	Alignment	
HK14	K24/I25	Trypsin-like	Trypsg-like-triplet	Alignment	
HK6	K21/L22	Trypsin-like	Trypsg-like-triplet	Alignment	
HK15	K21/L22	Trypsin-like	Trypsg-like-triplet	Alignment	
HK8	K32/V33	Trypsin-like	Trypsg-like-triplet	Alignment	

4.1.4   Molecular modelling of enzyme-substrate complexes: mapping the subsites of kallikrein 4 and other members of the human kallikrein family.

The knowledge of how an enzyme binds its substrate and which subsites it supplies in substrate-recognition was anticipated here by using a familiar enzyme-inhibitor complex. The enzyme-inhibitor complex suitable for the kallikrein family is the crystal structure of the pig tissue kallikrein (pk1) in complex with the hirustasin inhibitor (Mittl *et al*, 1997), which binds the substrate in an analogous way to the true enzyme-substrate complex for chymotrypsin (Wilmouth *et al*, 2001).

After a structural superimposition between hK4 and pk1 and the construction of the substrate, the model of the enzyme-substrate complex of hK4 had the conventional binding mode at the interface, as trypsin and pig kallikrein 1. The composition of the substrate (Ser-Gly-Ser-Cys-Ser-Arg-Ile-Ile-Asn-Gly (P6-P4'), was chosen on the basis of the activation region of hK4 (section 4.3), assuming autoactivation of this trypsin-like enzyme (Silverberg and Kaplan, 1982) and by using typical substrate-residues for trypsins around P3-P6 and P3'-P4' (Light and Janska, 1989). The backbone H-bonding seen in figure 4.3, shows the strongest part in binding (four H-bonds) between P2-P3 residues from the substrate and the Ser222-Phe223-Gly224 motif (equivalent to the SWG-motif from trypsins, marked yellow in Fig 4.2). From P4-P6 no backbone H-bonds occur. The P'-region has one H-bond at the scissile bond, and one between the backbones of P2' and Phe55, while P3' and P4' have no H-bonds with the enzyme. The catalytic centre is in a plausible conformation with the substrate, forming the transition state oxyanion hole (see Fig. 4.4) in a similar fashion to the crystallized enzyme-substrate complex of chymotrypsin (Wilmouth *et al*, 2001).

Because the binding mode of the computer model of hK4 and substrate shows preserved structural features with the two crystallized complexes (Mittl *et al*, 1997, Wilmouth *et al*, 2001) it is feasible to compare the enzyme-substrate model with the crystal structure of mouse kallikrein 8 (Kishi *et al*, 1999), which is the closest structural homologue to the human kallikreins and also has its subsites characterized. The comparison concerted with use of the sequence alignment in Fig. 4.2 (which aids in deducing further binding subsites for other kallikreins), shows that the subsites of human kallikrein 4 are identifiable (Fig 4.5). These residues are listed in Table 4.4 along with the equivalent residues from the other kallikrein members, and show when compared to the other members to provide chemical invariance at the centre of the pocket (S2-S1') and

increasing variance at the peripherical regions of the binding pocket, S3, S4 and S2'. These differences at the periphery of the binding pocket designate the uncommon specificities between the fifteen kallikrein enzymes, which determine their substrate selectivity. The S-regions of hK2, hK3, hK4, hK7, hK9, hK10 and hK15 are quite special when it comes to their specificities in comparison with the other members of the family; these regions show different residues especially at the S2 and S1' positions (see Table 4.4). hK4 varies from all other members by having a hydrophobic dominance at the S-region, which differs from the typical presence of charged residues as aspartate and histidine at the S2/S3 positions present in the other kallikrein members listed in table 4.4. HK2, hK3 and hK9 present a higher population of charged residues at the S-region compared to the other enzymes. However, all these members have one common binding feature: the typical hydrophobic preference at the S'-region, typically populated by hydrophobic residues (see table 4.4).

Now that enough is computationally anticipated regarding specific sub-sites, the next section presents what this knowledge can be used for: suggesting substrates for kallikrein 4.

This analysis is carried out as inspired by the work of Brinkworth *et al* (2002) applied on phosphokinases. The method relies on a force-field based assessment of chemical and geometrical complementarity between the interacting sidechains at the enzyme-substrate interface. This approach was specifically tested on serine proteases by Manzetti and Barnard (2001), and is shown in the next section to correlate with *in vitro* results.
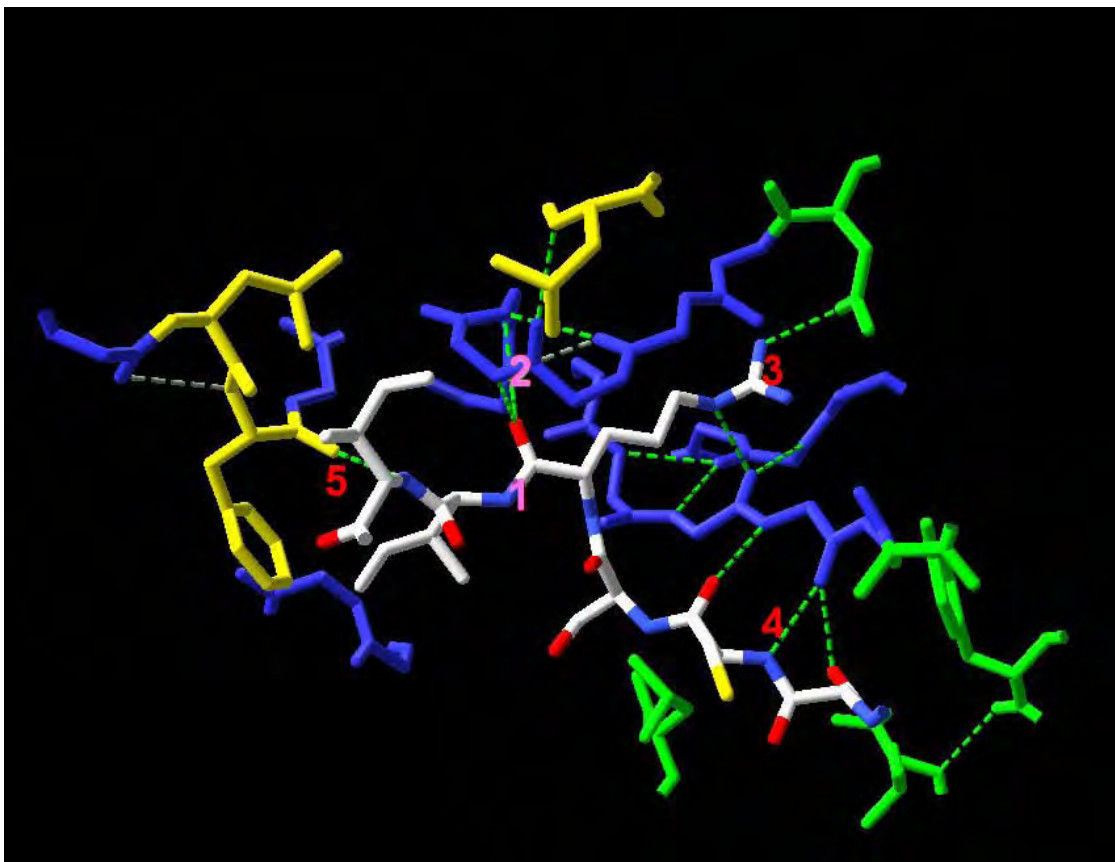
Figure 4.3. The binding architecture of kallikrein 4 in complex with a suitable substrate.

Yellow and green residues are respectively the predicted S'- and S-subsites on the enzyme. The blue-coloured residues are weakly invariant residues that support the binding site of the enzyme, and also provide backbone H-bonding of the substrate. The substrate is coloured in CPK (atom colouring), where: 1) the scissile bond; 2) the oxyanion hole; 3) the salt-bridge between Arg-Asp (P1-S1); 4) the backbone H-bonding at the P-region as determined for trypsin-like proteins (Perona and Craik, 1995); and 5) the only backbone H-bond at the P'-side, between P2' and Phe55.
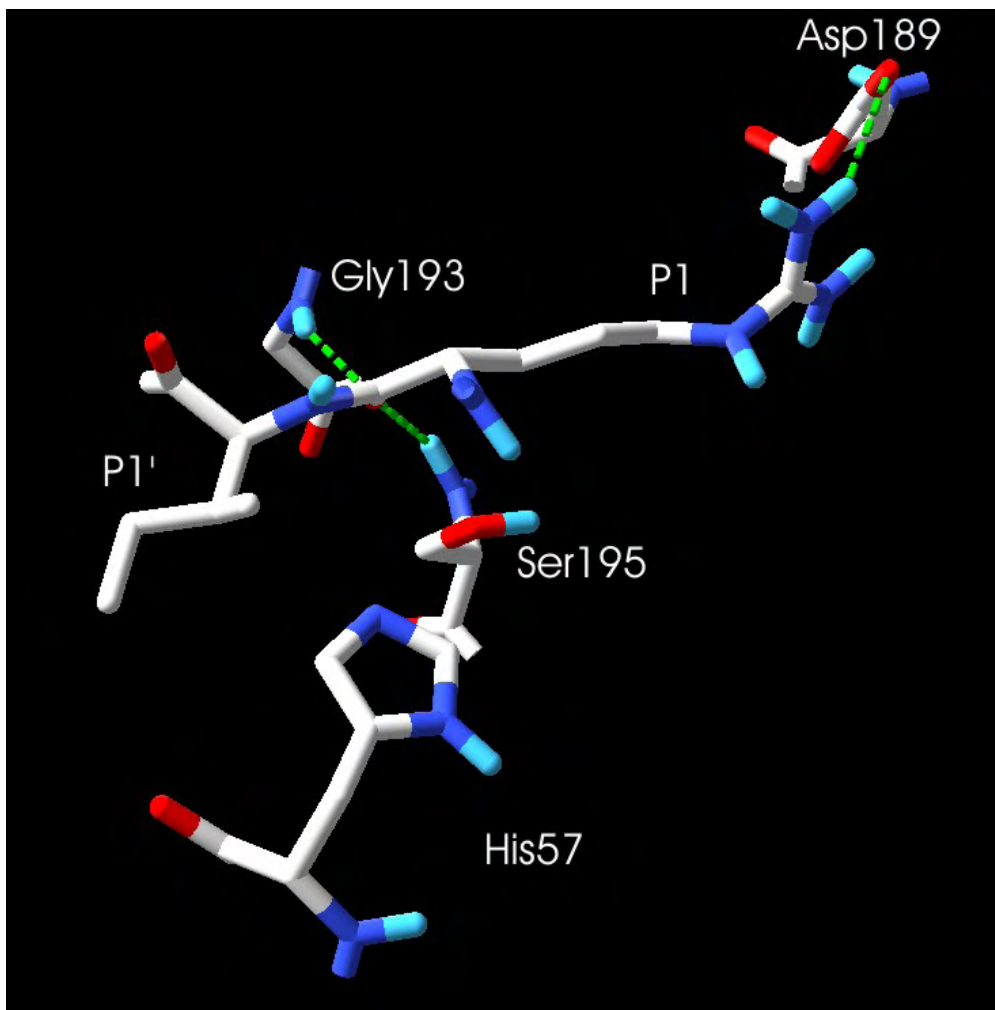
Figure 4.4. A modelled oxyanion transition state complex.

Shown is the catalytic centre of the modelled kallikrein 4 with the oxyanion hole between the carbonyl at P1 and the backbone atoms of Gly193 and Ser195, as previously described (Fersht, 1977) and recently experimentally proved (Wilmouth *et al*, 2001). Additionally, the salt-bridge at the bottom of the primary specificity pocket between Arg at P1 and Asp189 is shown.
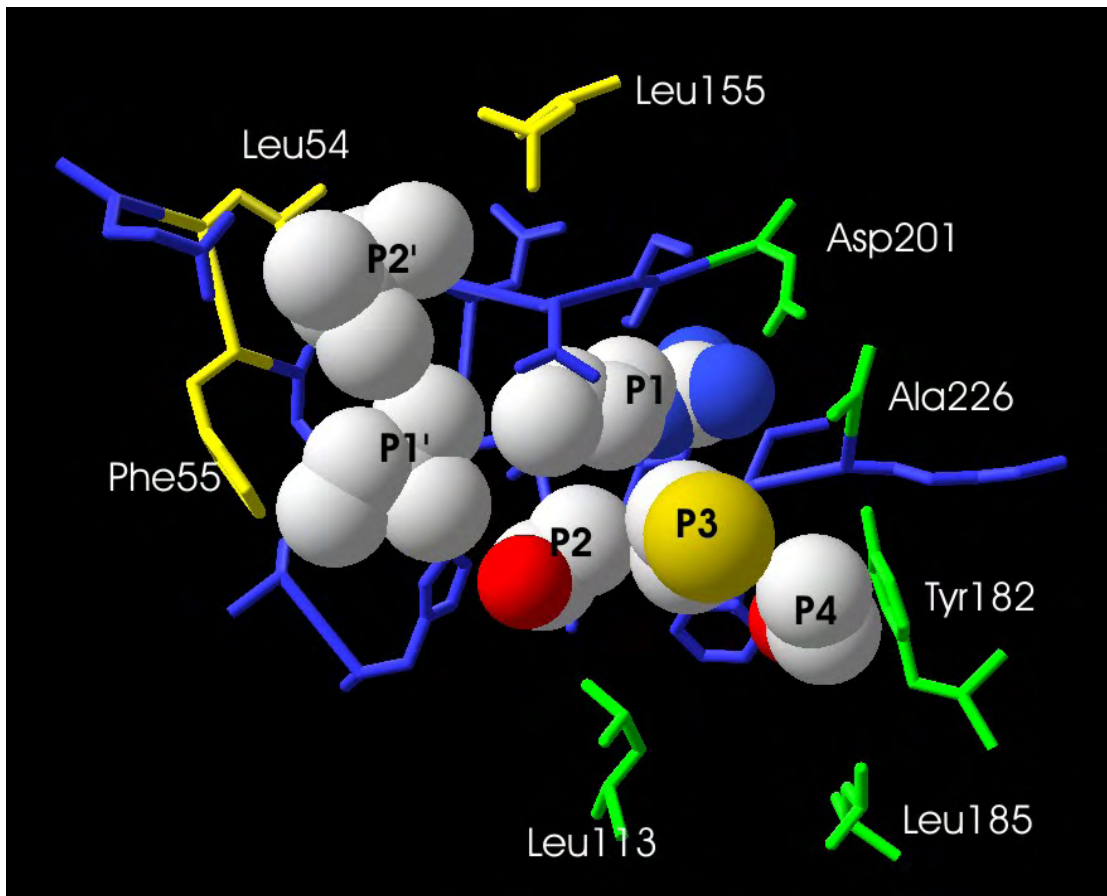
Figure 4.5. Illustrated interaction between the enzyme and the substrate.

All backbone atoms and hydrogens are hidden for graphical purposes, since these do not participate in substrate-specific interactions (Coombs *et al*, 1998a). Tyr 182 is from unique PLYH-motif of kallikrein 4.

Table 4.4. Listed subsites of the entire kallikrein family.

Top two rows indicate the subsites supplied by structural comparisons between the crystal structures of pk1 (Mittl *et al*, 1997) and mk8 (Kishi *et al*, 1999). Those in bold have been structurally examined, the others are deduced by alignment, and the italics are vague due to the flexible kallikrein15 loop dominating this region in unknown ways (putative high B-factor). Residues of the human kallikrein members are numbered in accord with their native sequences. The two crystal structures (top two rows) are numbered by conventional chymotrypsinogen numbering. The un-numbered residues derive from the alignment and have not been structurally confirmed. All human kallikrein members are numbered with individual numbering.

	S2'	S1'	S1	S2	S3	S4
pk1	**Phe40/Phe151**	**Gln41**	**Asp189**	**Tyr99**	**Thr218**	**His172/Lys175**
mk8	**Leu40/Phe151**	**Ile41**	**Asp189**	**His99**	**Asp218**	**Tyr172/Lys175**
hK1	Phe/Phe	Gln	Asp	Tyr	Val	Val/Gln
hK2	**Ala48/Arg167**	**His49**	**Asp207**	**Lys107**	**Glu332**	**Ser189/Glu190**
hK3	**Ala48/Thr167**	**Val49**	**Ser207**	**Ser117**	**Glu332**	**Pro189/Gln190**
hK4	**Leu54/Met161**	**Phe55**	**Asp183**	**Leu113**	**Ala226**	**Tyr182/Leu185**
hK5	**Leu91/Phe200**	**Tyr92**	**Asp239**	**His150**	**Tyr264**	**Tyr221**
hK6	Leu/Phe	Leu	Asp	His	Ile	Pro/Gly
hK7	**Leu53/ Phe159**	**His54**	**Asn199**	**His109**	**Phe224**	**Tyr180, Leu183**
hK8	Leu/Phe	Leu	Asp	His	Asp	Pro/Gly
hK9	**Leu46/Phe158**	**Phe47**	**Arg197**	**His108**	**Glu223**	**Ala222/Tyr179**
hK10	Phe/Tyr	His	Asp	Asp	Tyr	Pro/Gly
hK11	Leu/Leu	Leu	Asp	His	Asp	Pro/Gly
hK12	Leu/Phe	Arg	Asp	His	Gly	Pro/Gly
hK13	Leu/Tyr	Leu	Asp	His	Phe	Pro/Gly
hK14	Phe/Tyr	Leu	Asp	His	Glu	Pro/Arg
hK15	*Phe/Leu*	*Asn*	Glu	His	Val	Pro/Gly

4.1.5 Predicting the subsite specificity of kallikrein 4 based on force field energy scoring – proposed substrates for kallikrein 4.

4.1.5.1 Estimating substrate-affinities by scoring of non-bonded electrostatic energies.

The scoring of interaction energies relies on the bonds, angles, torsions, van der Waal and Coulomb energies (see section 1.12.3) of the substrate and the enzyme and between these two entities.

The following derivation and argument formed the basis of results summarized and published by Manzetti and Barnard (2001), where trypsin was assessed against *in vitro* parameters.

The correlation between *in vitro* binding and *in silico* scoring relies on:

The total solvation energy ($\Delta\Delta G$solv w→cmplx) of a complex (protein + ligand) when water is added to the protein surface is:

$$\Delta\Delta G_{\text{solv w}\rightarrow\text{cmplx}} = \Delta G_{\text{solv, cmplx}} - \Delta G_{\text{solv, w}} \qquad \text{(Warshel, 1997)} \qquad [1]$$

Where *w* stands for water, and *cmplx* is protein+ligand. Because, the GROMOS force field calculates the energy *in vacuo* and there is no solvent, it can be simplified that:

$$\Delta G_{\text{solv, cmplx}} = \Delta G_{\text{bind cmplx}} \qquad \textit{(in vacuo)} \qquad [2]$$

The energy computed in the GROMOS force field of the protein-substrate complexes $\Delta U_{in\ silico}$, can be interpreted as an enthalpic value, which when simulated *in vacuo* (where the pressure is zero) can be equalized as $\Delta H_{in\ silico}$.

Now, since $\Delta G_{\text{protein}} = \Delta H_{\text{protein}} - T*S_{\text{protein}}$, and the entropic contributions are not taken into account in the computation of energy *in vacuo* (Scott *et al*, 1999), we can deduce that:

$$\Delta G_{\text{protein}} \propto \Delta U_{in\ silico} \qquad [3]$$

Low negative values of $\Delta U_{in\ silico}$ designate complexes in a more favorable energy state than complexes in an unfavorable conformation with higher energetic values.

The link to associate $\Delta U_{in\ silico}$ with *in vitro* parameters is given by,

$$\Delta G_{protein} = -RT \ln K_{eq} \qquad [4]$$

Which means that we can associate $-RT \ln K_{eq}$ to the force field energy $\Delta U_{in\ silico}$ as:

$$RT \ln K_M. \propto \Delta U_{in\ silico} \qquad [5]$$

Where K_{eq} is the equilibrium constant for the formation of the activated complex enzyme+substrate (ES), in the reaction,

$$E + S \underset{\leftarrow}{\overset{k_M}{\rightarrow}} ES \xrightarrow{k_{cat}} EP \rightarrow E + P \qquad \text{(Warshel, 1997)} \qquad [6]$$

where E is the protease, and S is the substrate, generating the cleaved product P, at the constant k_{cat}.

The outlined reasoning shows that a correlation can be expected between a force field based theoretical energy ($\Delta U_{in\ silico}$) and the *in vitro* kinetics at 300K for a serine protease with trypsin fold.

After applying the force-field based scoring onto a series of different residues at each subsite (from the substrate), a complementary substrate is shown with its interaction energies ($\Delta U_{in\ silico}$) in tables 4.5 and 4.6. Using this scoring method, alanine was seemingly more compatible with the S1' cavity (ΔU= -32.414kJ/mol) than isoleucine (ΔU= 0.332kJ/mol). Incompatible residues with the S1'-cavity of hK4 were basic, acidic and aromatic residues.

The interaction between P2' from the substrate and the S2'-subsites shows favourable interaction between isoleucine (substrate, Table 4.5, ΔU= -19.99kJ/mol) and Leu155 and Leu54 from the enzyme. Alternative residues here are bulky and aromatic residues. The P2-P4 residues need to be amphiphilic (Ser, Thr, Cys, Gln, Tyr) in order to have favourable interaction with the enzyme and allow solvation of the substrate. The P5, P6, P3' and P4' residues were not in contact with the surface and were therefore not included in the specificity tables.

Uncomplimentary residues were also assessed, and an example is presented (Table 4.7) by the use of a substrate composed of proline residues since this peptide is not cleaved by trypsins (Perona and Craik, 1995). The unfavourable, large positive values in table 4.7 show the high-energy state at the binding interface, which reports a repulsive non-bonded interaction at P1' and unfavourable torsion of proline in the extended beta-sheet conformation at the P-region. This shows that it is possible to exclude certain residues (Table 4.8) in this computational *a priori* assessment, and also suggest complementary residues based on existing data and force field data (Table 4.9).

The knowledge presented about the preference of kallikrein 4 lays a preliminary foundation for *in vitro* assays, which prove useful given that its crystal structure is unavailable. The reliability of these results is lower than the reported reliability with trypsin and its substrate (Manzetti and Barnard, 2001), because this is a protein model. Additionally, given that a solvated state *in vitro* can only be "truly" predicted with NMR there are limitations in these assumptions alone. The conformation of the substrate before it docks to the enzyme, which can present preferable orientations towards the surface of the enzyme (beta-sheet extended, Perona and Craik, 1995), cannot be assessed with the mentioned methodology in this section. Because van Gunsteren *et al* (2001) presented successful results in predicting structures of oligopeptides in solution with the use of molecular dynamics, a few other questions can be answered: How stable is the substrate on the surface of kallikrein 4 in the given conformation? How compatible is the substrate to the enzyme surface in a solvated state?

The next section will attempt to answer these questions by focusing on the dynamic and vibrational aspect of the binding interface, so more useful information can be gained to understand the mechanism of binding for this protein.

Table 4.5. Estimated ΔU values for the substrate in associated state with the human kallikrein 4 computer model.

"Resid" indicates the residues from the substrate, from top row to bottom row, P4 →P2' (the B and the number are negligible). The other red titles indicate the values described in section 1.11.2. Arg 4th row from top is the central P1 residue. 5th row from top is the P1' residue.

Resid #	Bonds	Angles	Torsion	Improper	non-bonded	electrostatic			Total
SER B 256	0.435	0.989	3.074	1.609	-21.77	-23.58	0.0000 //	E=	-39.248
CYSH B 257	0.121	1.311	7.557	0.244	-28.13	-13.91	0.0000 //	E=	-32.816
SER B 258	0.174	1.601	5.524	0.460	-26.56	-25.11	0.0000 //	E=	-43.915
ARG B 259	2.638	10.133	67.444	5.005	-53.90	-292.41	0.0000 //	E=	-261.094
ILE B 260	2.597	7.831	4.447	3.249	-13.22	-4.57	0.0000 //	E=	0.332
ILE B 261	1.519	6.350	2.152	4.848	-30.65	-4.21	0.0000 //	E=	-19.989

Table 4.6. Estimated ΔU values for a second substrate in associated state with the human kallikrein 4 computer model, with an alanine at the P1'.

Columns follow the same assignments as table 4.5.

SER B 256	0.218	0.659	5.433	0.683	-22.35	-10.99	0.0000 //	E=	-26.351
CYSH B 257	0.151	2.520	9.015	0.008	-28.64	-11.87	0.0000 //	E=	-28.822
SER B 258	0.200	2.413	2.998	0.244	-22.67	-19.43	0.0000 //	E=	-36.249
ARG B 259	3.207	12.772	77.037	1.448	33.94	-282.43	0.0000 //	E=	-154.027
ALA B 260	0.520	2.461	0.906	0.282	-31.51	-4.50	0.0000 //	E=	-31.842
ILE B 261	1.323	5.440	1.994	2.970	-25.68	-0.39	0.0000 //	E=	-14.353

Table 4.7. Estimated ΔU values computed for an unfavourable substrate in interaction with hK4.

High positive values indicate non-fitting residues with the computer model of kallikrein 4. Organized in the same way as figure 4.5 and 4.6.

PRO B 256	0.746	11.681	19.449	2.150	-15.49	-1.20	0.0000 //	E=	17.335
PRO B 257	1.078	16.558	18.952	8.623	-23.25	-10.07	0.0000 //	E=	11.893
PRO B 258	1.389	17.942	21.467	0.674	-16.28	-6.61	0.0000 //	E=	18.583
PRO B 259	4.882	30.968	25.735	5.288	-12.55	-23.79	0.0000 //	E=	30.527
PRO B 260	1.117	23.493	14.042	6.193	-12.08	3.83	0.0000 //	E=	36.595
PRO B 261	2.872	20.189	16.956	3.553	0.06	-12.96	0.0000 //	E=	30.670

Table 4.8. Worst fitting residues of an oligopeptide stretching P2'-P4 in interface-complementarity analysis with hK4.

P2'	P1'	P1	P2	P3	P4
Ser, Asp, Asn, Glu.	Glu, Asp, Tyr, Pro, Trp, Arg, Lys.	Glu, Asp, Tyr, Phe, Trp, Pro	His, Phe, Tyr, Pro	Trp, Leu, Pro	Arg, Lys

Table 4.9. Best fitting residues of an oligopeptide stretching P2'-P4 in interface-complementarity analysis with hK4.

Bold marked residues are the preferred residues.

P2'	P1'	P1	P2	P3	P4
Trp/Ile/**Phe**	**Ala**, Ile, Thr, Val	**Arg**/Lys	Cys/Ser/Met/**Thr**	Gln, **His**	**Ser**/His/Thr

4.1.5.2 Molecular Dynamics of hK4 in complex with a substrate.

To further investigate the enzyme-substrate interface of kallikrein 4, a 1ns-simulation of kallikrein 4 with the oligopeptide from the previous section, Ser-Gly-Ser-Cys-Ser-Arg-Ile-Ile-Asn-Gly (P6-P4'), was carried out in a box of ~3200 water molecules (Fig 4.6).
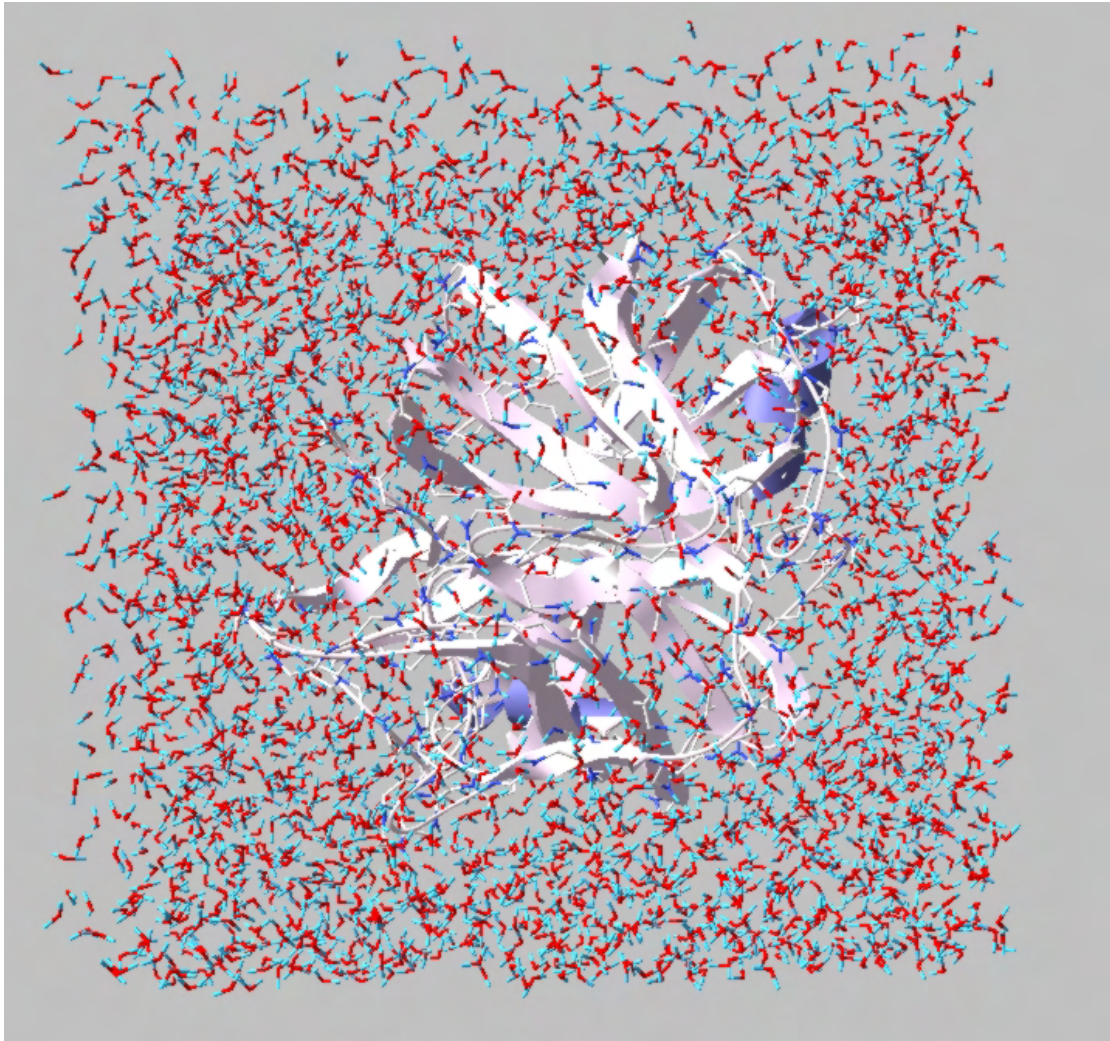
Figure 4.6. Kallikrein 4 with substrate in a box of ~3200 water molecules at 990ps during the molecular simulation.
The protein is shown in ribbon mode, solvated by the water molecules shown in ball&stick mode.

The structural stability of the enzyme in water was monitored using a continuous calculation of the random mean square deviation (RMSD) of the enzyme over the simulated time-interval (commonly used approach, Heymann and Grubmuller, 2001, Hayward et al, 2001). High RMSD indicates unfolding of proteins from their original state in a simulation, which can occur when the models are of low quality (given poorly modelled hydrophobic cores structurally destabilized by solvent penetration – data not shown). The RMSD of hK4 (Fig 4.7) was converging within previously published intervals (Hayward et al, 2001), which indicates a gradual approximation to a structural equilibrium.

The stability of the substrate to the enzyme interface was monitored after taking snapshots of it at 0ps, 500ps and 1000ps and by calculating the RMSD between these three variants. The H-bond network at the P-region is completely preserved after a 1ns simulation, and also the oxyanion-hole network (see Fig 4.8). The snapshots at three time-frames (Fig 4.9) indicate a stable substrate on the surface of the enzyme that converged to 1.22Å from its original modelled configuration. This moderate deviation was mainly contributed by the P4, P5 and P6 residues that do not have backbone H-bonding to the enzyme surface and freely move in the solvent.

Given this stability of the substrate, the subsite-interactions are less complicated to monitor. The contact maps (Fig 4.10) show the interaction "fingerprints" during the molecular simulation and aid to identify which residues from the enzyme did interact with the substrate in a dynamic and solvated system. The white clusters seen in figure 4.10 include both backbone interactions and sidechain interactions. Because of this mix of interaction classes, a combination of visual inspection and contact maps gives a better picture of which sidechain took place during the simulation (Figure 4.10 & 4.11).

After 1000ps in the solvated dynamical state, a new neighbour of the P1' position came in interaction distance with it: Phe73 (lower A region, on fig 4.11) shielded the isoleucine residue from the solvent. This could be one of the following, a) an artefact of the simulation, b) a weakness in the protein model and c) a feasible interaction. Whether this is an artefact or not is not easy to define based on the existing data, but because this segment follows two residues after the catalytic His57, which remains fixed during substrate binding (Wilmouth et al, 2001) along with its neighbours, this occurrence is probably one of the two first alternatives. Indeed, this phenylalanine residue does not belong to any unusual motif (as for instance the PLYH motif, noted by Nelson et al, 1999) and it has not been implicated in binding according to the thorough

review by Perona and Craik (1995) on trypsins and chymotrypsins. Phe73 in hK4 can therefore be excluded based on this evaluation with pre-existing data.

The S2'-residue (Ile) is at start of the simulation squeezed and kept firm by two "flaps" represented by Leu155 and Met167. This hydrophobic lock is the last firm lock on the substrate on the P'-side, because the P3' and P4' - residues move outwards in the solvent (data not shown for graphical purposes).

On the opposite side of the scissile bond, there is a hydrophobic dominance from Leu113 on the hydrophobic moiety of Ser (P2) (B-field on Fig 4.11). The hydroxyl group from the serine residue establishes H-bonds with the solvent, and its backbone remains in H-bonded contact with the SFG-motif from enzyme (Fig 4.8 and Fig 4.2). The P3-residue (Cys) orients its sidechain towards Ala226, which lies on the F-region as seen in Figure 4.13, which indicates yet another amphiphilic complementarity on the S-region. Serine at P4 is the last residue to remain in contact with the S-region from the enzyme surface, and it has its hydrophobic moiety in contact with Leu185 (from the PLYH-motif, hk4-specific motif noted by Nelson *et al*, 1999) while its hydrophilic hydroxyl tip established an H-bond to Tyr182. The remaining two residues at the P-region (P5 and P6) do not interact with the enzyme in terms of sidechain interactions, and as mentioned above, are detached from the surface in terms of backbone H-bonding as well (as also the two latter from the P'-region). This binding trend suggests that hK4 may not need more than P2'-P4 when assayed with oligopeptide substrates, which differs from the serine protease collagenase that binds 6 residues on the P-side (Perona and Craik, 1995). These binding results can also be interpreted speculatively in that hK4 does not process extracellular substrates such as collagen, which specifically required prolonged binding at the P-region (Perona and Craik, 1995). Finally, the interaction energies which lie in the negative region (as they should) indicate a stable conformer, given the Coulomb and vdW interactions between the chains (fig 4.12).

The interactions postulated here are plausible for two main reasons. They follow empirical trends (Mittl *et al*, 1997, Kishi *et al*, 1999, Wilmouth *et al*, 2001) and they are based on a stable model, assessed in the MD-simulation. Therefore the interactions between the substrate and the enzyme can be reproduced with high plausibility in the case of the kallikreins, and most importantly, they can be verified and compared to structural data at the end.

Now that a significant preliminary knowledge base is built, its application can be directed either to enzymatic assays, or drug design projects. Given the importance in supplying anticipative results for future studies, the drug design potential from these results is presented in the next
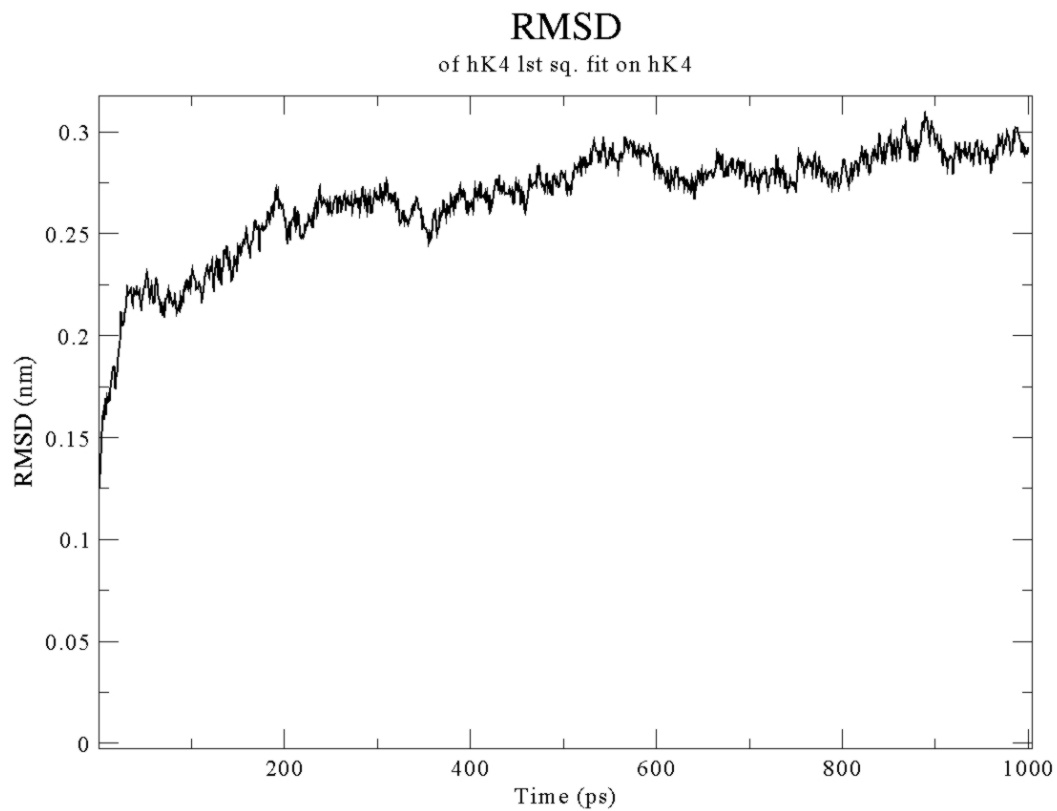
Figure 4.7. RMSD-plot of the random mean square deviation of kallikrein 4 enzyme, converges towards a stable conformer.

The substrate is not taken into account, since the model stability is the crucial factor to supply a stable substrate. (0.3nm=3Å).
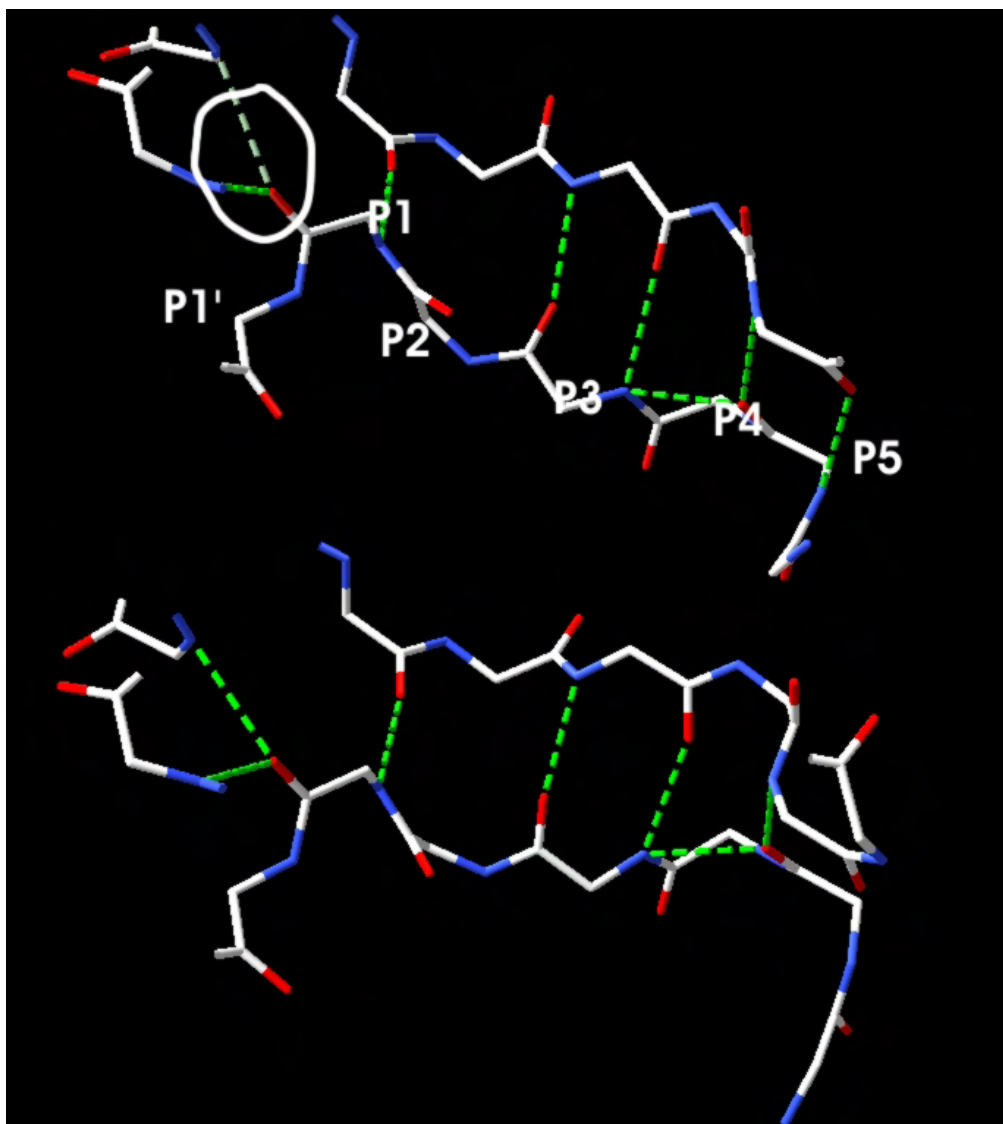
Figure 4.8. Conservation of H-bond network between enzyme and substrate in molecular simulation.

The reported backbone H-bonding between P2-P4 and the SFG-motif (Perona and Craik, 1995 and Wilmouth *et al*, 2001) is well conserved after 500ps (top) and after 1000ps (bottom). The circle indicates the oxyanion whole, which has been visibly strengthened (color of H-bond on top at circle shows weak H-bond) after 1000ps. All sidechains are hidden for graphical purposes.
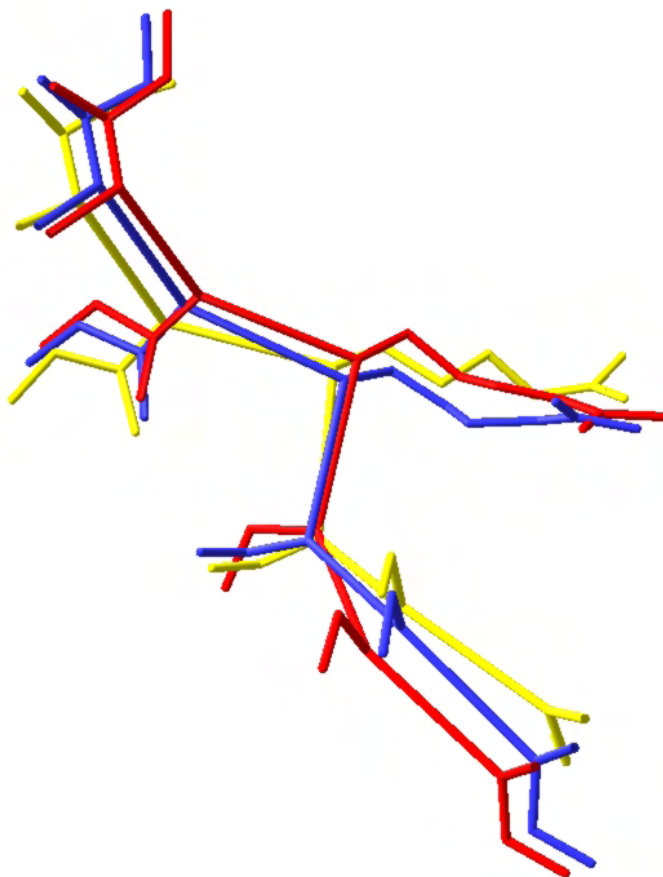
Figure 4.9. Dynamics of the hK4 substrate when complexed to the enzyme's binding site (enzyme not shown).

The substrate showed in C-alpha mode: yellow is the start configuration; red is at 500ps and blue is at 1000ps. The rmsd between the two last and the initial configuration is 1.66A and 1.22A respectively, which indicates a convergence to the initial state during this simulation.
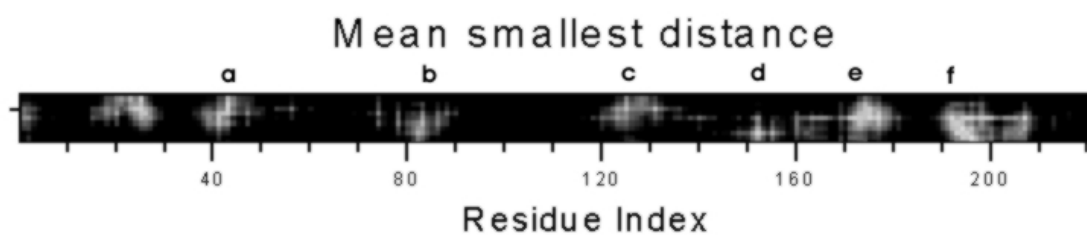
Figure 4.10. A contact map generated over a 1000ps interval during the simulation of hK4 and the docked substrate.

The horizontal numbering indicates the residues on the enzyme, while the vertical direction indicates the residues on the substrate. Each white dot represents a residue that is in contact with another from the other chain. Each region is designated by a letter: a) P2'-Leu54, P1'- Phe55; b) P2- Leu1113; c) P2'- Leu155; d) P4- Tyr182, P4- Leu185; e) P1- Asp201; f) P3-Ala226. (The numbering on the figure is with Ile31 on the enzyme as Ile1).
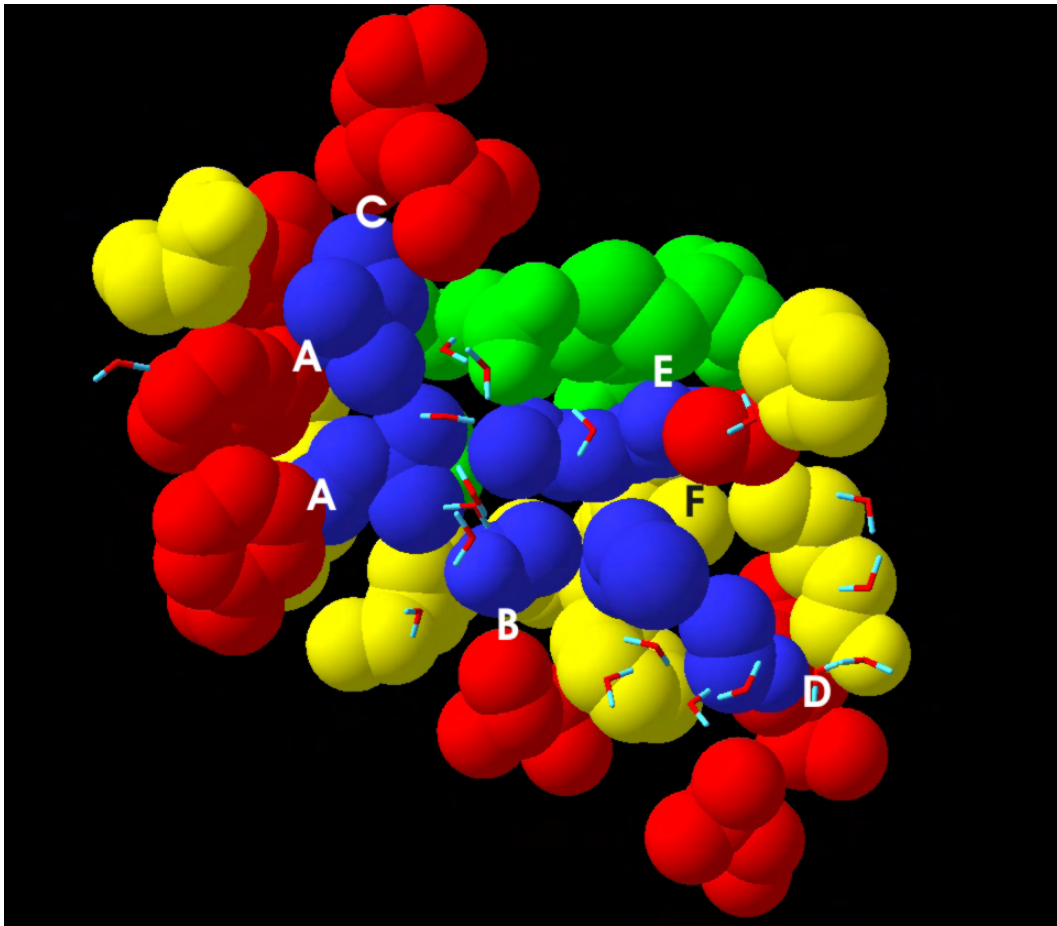
Figure 4.11. Graphical representation of the interacting residues from hK4 and the substrate after a 1000ps simulation in solvent.

The equivalent positions as observed in the contact maps in Fig 4.10, noted with letters: the blue sidechains are from the substrate; the red sidechains are the interacting subsites; the green are the S1-hydrophobic residues which interact with the hydrophobic moiety from arginine (conserved in trypsins); and the yellow are adjacent conserved residues in kallikreins
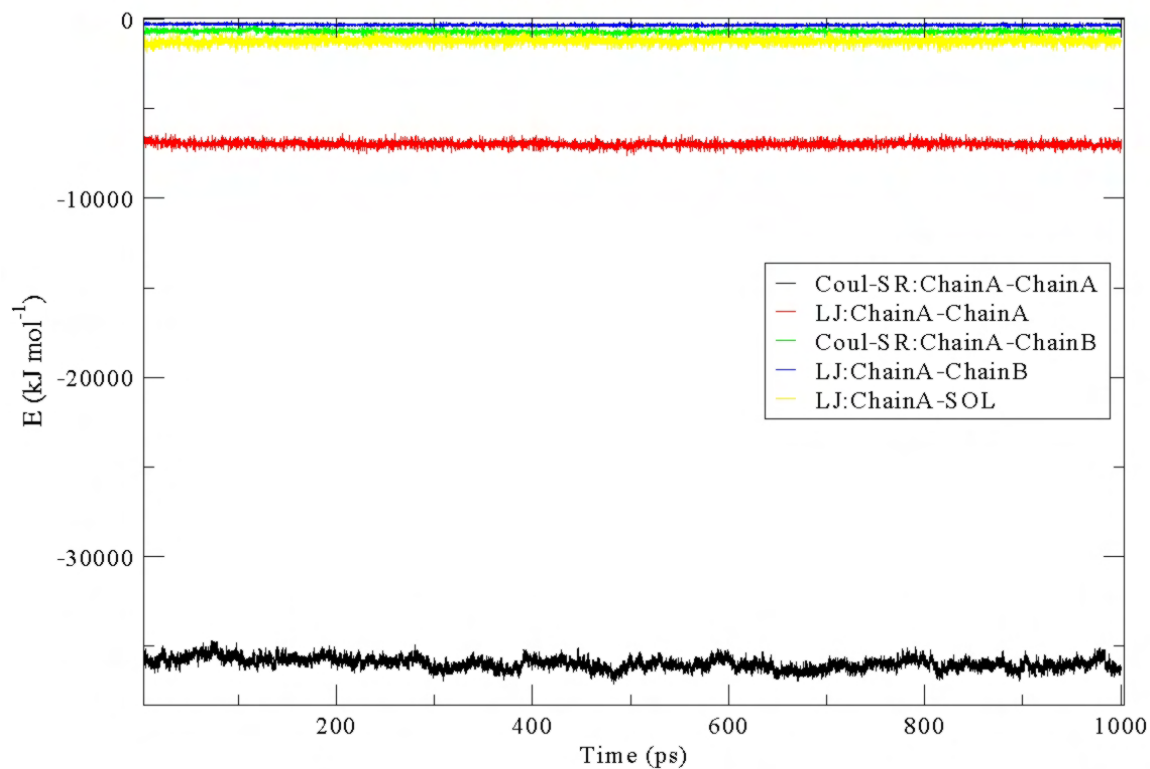
Figure 4.12. Force field energies in the simulated system consisting of the enzyme-substrate complex and the solvent phase.

Van der Waal interaction with the solvent was continuously in a negative energetic state (favourable) as all the other interaction energies were. Coul stands for Coulomb energy (electrostatics) between the given groups (see legend), and LJ stands for Lennard Jones, which describes the van der Waal -interactomic contacts.

4.1.6   Molecular design and docking of putative hK4 inhibitors.

Three inhibitors directed to human kallikrein 4 were built for flexible ligand docking experiments (Fig 4.13) on the basis of the gained information around the binding pocket of hK4. These inhibitors had modifications of the P1 sidechain with a typical S1-pocket inhibitor for trypsins, N-acetyl-D-glucosamine, which was originally directed towards mouse kallikrein 8 (Kishi *et al*, 1999), the closest structural homologue to human kallikrein 4. The inhibitors span from the P2-P2' and are targeted to inhibit the binding site of human kallikrein 4 by using the residues equivalent to its own activation region, since autoactivation of trypsinogens is a known phenomenon (Silverberg and Kaplan, 1982, Rittenhouse *et al*, 1998).

The choice of modifying the sidechain at P1 (Mod-P1, Fig 4.13) was made to increase the hydrophobic moment, which is much stronger than the equivalent carbon skeleton of an arginine or lysine side chain (typical for primary trypsin specificity). Increasing the hydrophobic moment yields an entrapment of the side chain mimicker in the P1 pocket, thereby inhibiting the protein to host its natural substrate (A common procedure with trypsin like enzymes, Glenn *et al*, 2001). An approach that was used as inspiration to deduce the hK4 inhibitors is reported by Llinas-Brunet *et al* (2000), who produced a potent inhibitor of the Hepatitis C virus protease which contains several of the natural cleavage subsites for this enzyme. The model of kallikrein 4 and its own activation peptide shows an energetically favourable complementarity. The inhibitor bound well in accord with a substrate, the hydrophobic moiety of the P1-sidechain was locked by strong hydrophobic attractions to the pocket (Fig 4.14), where the ring atoms interacted with Phe223, Cys203, Gly 204 and Cys 228. The guanidinium tip established a salt bridge with the primary S1-determinant, Asp201 and Ser 202 (see Fig 4.15, 4.16). At the S1' position, Gly205 and Asn 204 interacted with Isoleucine. These are located roughly between the "upper" A and the E region seen on figure 4.11, implying that the side chain of the isoleucine oriented itself differently than from a substrate P1' residue. Ser at P2 bound closer to the original P2 orientation from the substrate, with Leu 113 (as also predicted by subsite analysis), by hydrophobic attraction with its carbon moiety. The backbone did however not establish the conventional H-bonds with residue 213 as it occurs in substrates. This trend of weak backbone binding in inhibitors was also observed in docking experiments performed on PSA by Coombs *et al* (1999). However, later, Klebe *et al* (2000) and Klebe and Shafferhans (2001) docked successfully to a test set of modelled serine proteases (down to 40% homology) to 1.4Å RMSD for the inhibitors.

The results show that the inhibitor is very likely to interact with most kallikreins, and possibly even trypsin. This makes it a non-specific inhibitor, although can be used for *in vitro* assays. The

length of the inhibitor may yield better specificity (Llinas-Brunet *et al*, 2000), although flexible docking with such large ligands is computationally very expensive and the conformations tend to be less plausible with the increasing size of the ligand. This was observed when a P3-P3' segment was tested on kallikrein 4, which resulted in a highly clustered binding mode, perpetuated by internal ligand interactions (data not shown).
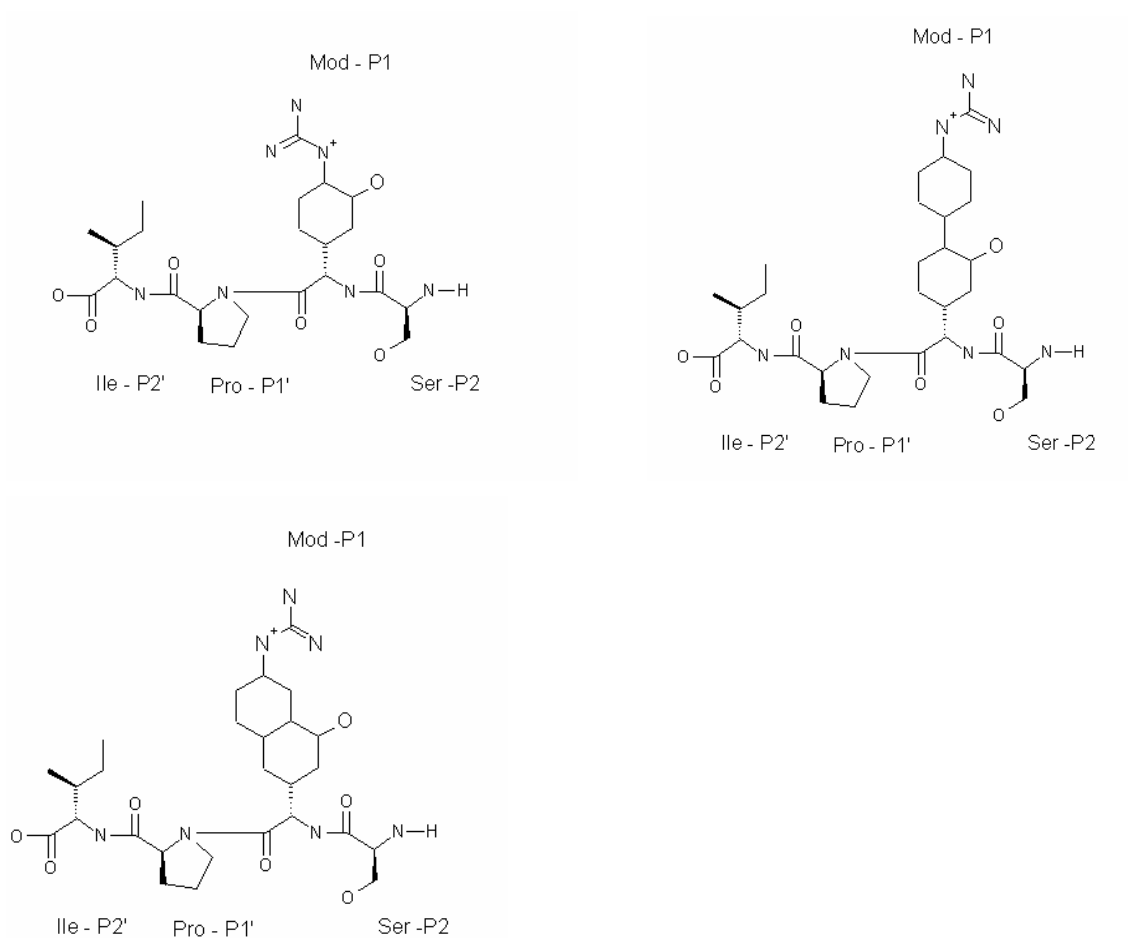
Figure 4.13. 2D-drawings of the inhibitors used for docking experiments with kallikrein 4 as the receptor molecule.

Top left, homologue to n-acetyl-glucosoamine as modified P1 sidechain; Top right, a guanidine-cyclohexyl group added to the previous conformer; Bottom left, another modification of a double-cyclohexyl unit as the hydrophobic unit of P1, with the guanidinium tip.
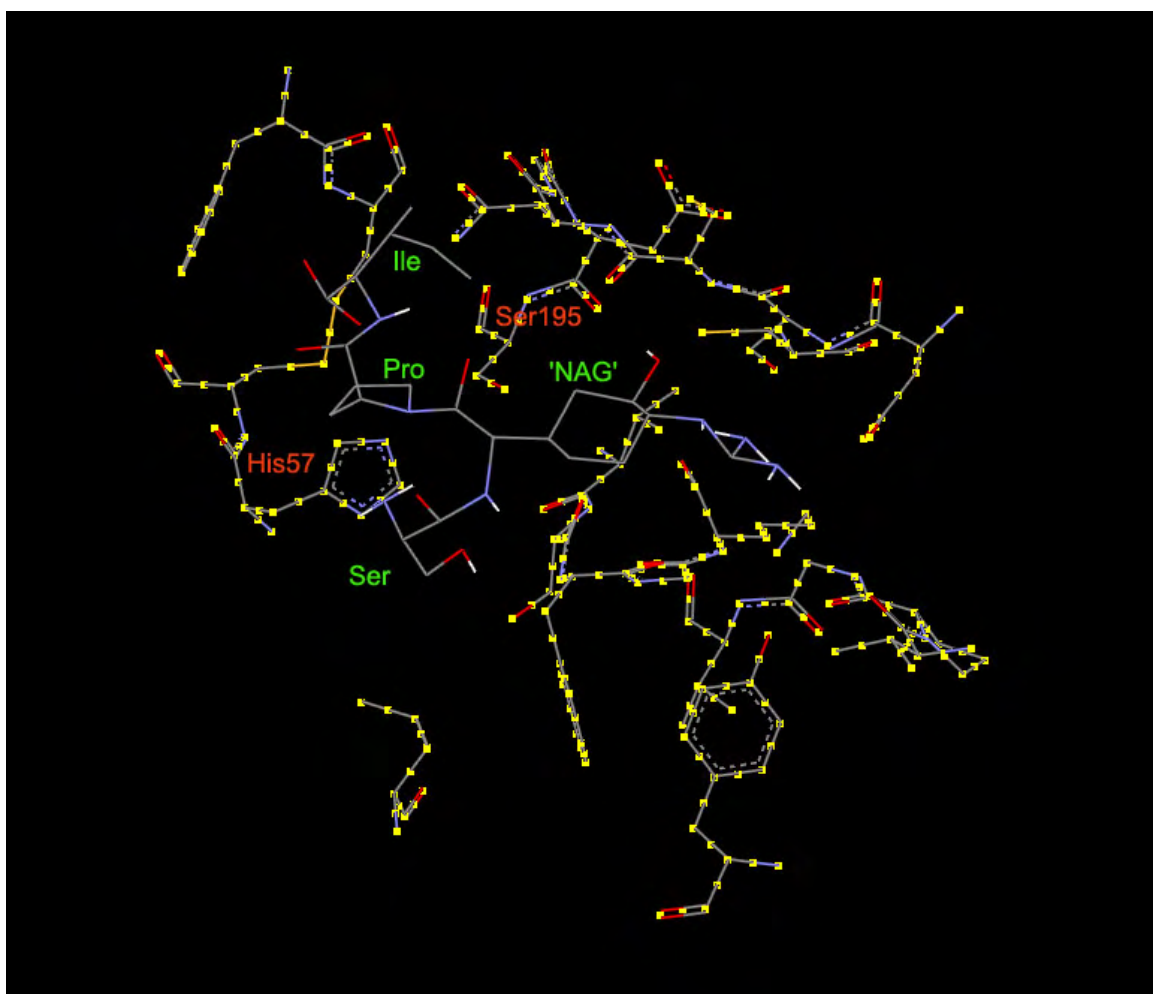
Figure 4.14. Binding modus between the first inhibitor shown in figure 4.15, after a docking session using AUTODOCK.

Red-labelled residues indicate the catalytic diad, green indicates the inhibitor residues. Yellow-dotted residues indicate the binding residues from the enzyme.
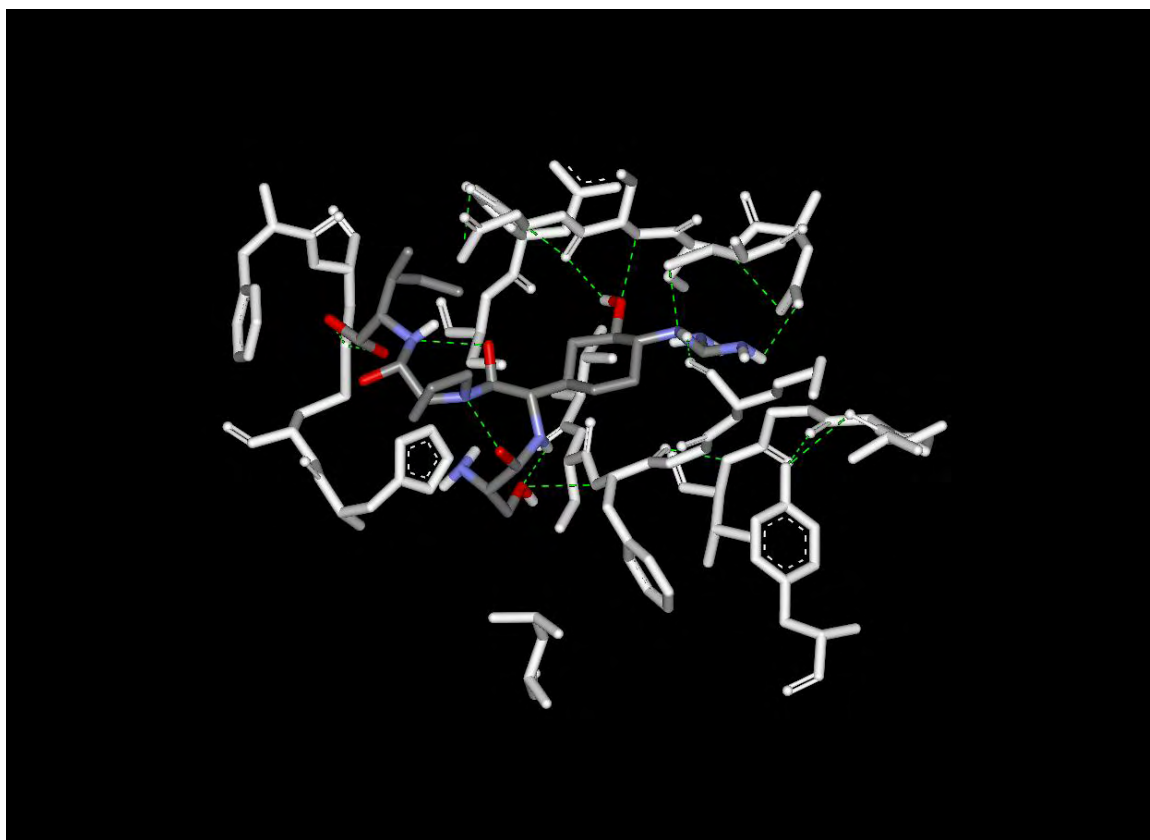
Figure 4.15. H-bond network of the inhibitor in complex with hK4 model.

Same as figure 4.16, however hydrogen bonds are shown in green, white residues are enzyme-residues, while CPK-colored residues are from the inhibitor. Far right H-bond between the inhibitor and the enzyme (grey) show the P1-S1 salt-bridge.
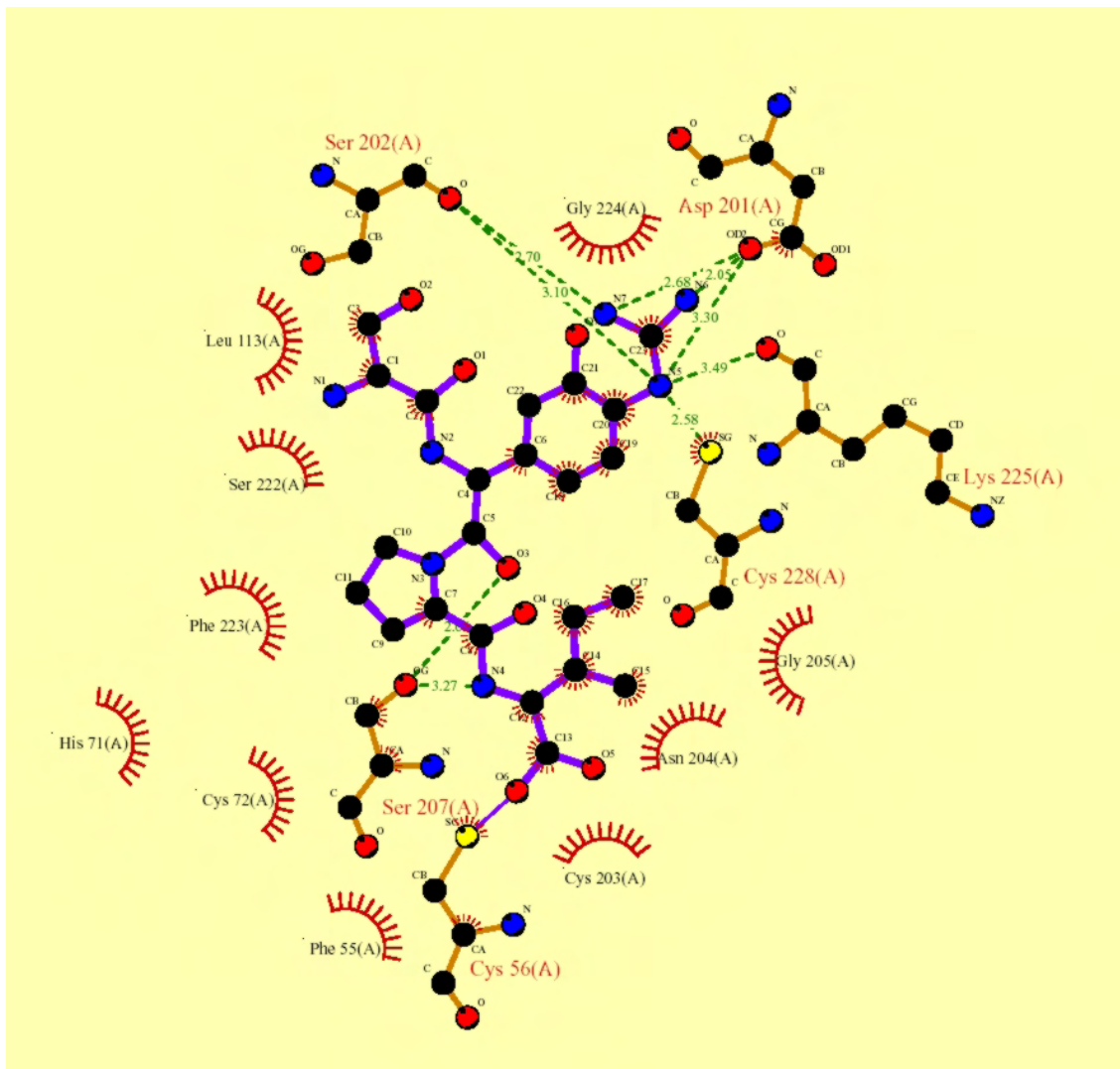
Figure 4.16. LIGPLOT diagram for the enzyme-inhibitor complex of kallikrein 4 and a tested inhibitor.

Red-circled atoms and residues indicate groups and atoms involved in hydrophobic interactions, green lines indicate putative H-bonds. Asp201 corresponds to Asp189 from trypsins; Leu113 is the reported S2 subsite in the previous chapter; Phe55 is the reported S2' residue, which interacts with the P1' position due to torsional changes from a beta-sheet conformation found in normal substrates, Gly205 is a glycine conserved in all kallikreins plus trypsin and chymotrypsin, taking part of the S2' pocket as well.

CHAPTER 5

5.1 Application of enzyme-substrate modelling to three metzincins: MMP-3, ADAM-9 and ADAM-10.

The results from the previous section highlighted to three essential procedures for assessing a substrate's complementarity to an enzyme *a priori:* a) enzyme-substrate modelling b) assessing residue complementarity and c) molecular dynamics simulation of the final complex. These procedures could be readily extrapolated to members of the kallikrein family, but the question remains whether it can be extrapolated to another protease family, such as the metalloproteases? There are two central obstacles to this: the inhibitory mechanism of the metzincins (section 1.2.4) is very different from serine proteases in that there is no explicit P2/P1-position and no coherent P-region found in the natural inhibitor, which poses a lack of empirical data to determine whether the enzyme substrate model is actually modelled correctly at the P-region. The second obstacle is at the catalytic center. As seen in the case of the serine proteases, the oxyanion hole was virtually identical between an inhibitor and a substrate (Mittl *et al*, 1997, Wilmouth *et al*, 2001), and could be easily modelled. In the case of the metalloproteases, the H-bonds and orientations between the scissile bond and the central catalytic residues are not as precisely known, as for serine proteases, because of the gap between the two inhibitory segments found in the natural inhibition situation (Gomis-Rüth *et al*, 1997). All that is known are the non-covalent bonds between transition-state-analogues and the catalytic center. The first obstacle requires a manual modelling of the P-region, and the second requires a thorough analysis of the coordination of the Zn-ion at transition state, with synthetic inhibitor analogues. Because there is much disagreement around the catalytic mechanism of metzincins (Mock, 1998, Browner *et al*, 1995, Matthews *et al*, 1988), which is important for modelling the ES-complex correctly and for designing better inhibitors against clinical targets as MMP-3, the modelling carried out in this chapter provides data in this regard as well, and aims also to suggests a new catalytic mechanism for metzincins.

The proteins targetted in this chapter are MMP-3, ADAM-9 and ADAM-10 which, to date have not been thoroughly analysed for sequence similarities with each other and with other metzincins. The following section introduces the sequence analysis of these target metalloproteases before the enzyme-substrate modelling is carried out, so a better understanding of these proteolytic enzymes can be achieved.

5.1.1   Sequences and alignment of selected metzincins.

In order to determine the properties of the sequences of the catalytic domains of ADAM-9 and ADAM-10, an alignment of these with the catalytic domains of adamalysin II, acutolysin C, TACE (ADAM-17), collagenase (MMP-1) and MMP-3 (Fig 5.1) was computed. Additionally, to get a broad phylogenetic perspective a phylogenetic tree of the catalytic domains of more distantly related metzincins, was computed (Fig 5.1).

The alignment shows a conserved Zn^{2+}-binding His-triad, the catalytic glutamate and the torsionally crucial glycine in the **HEXGHXXGXXHX**-motif in ADAM-9 and -10. The strictly conserved methionine residue, crucial for the Met-turn found in all metalloproteases to sustain the active site (Stocker *et al.*, 1995), is also present in both ADAMs. The second His-triad, **HXDXXPFDGXXXXLAHAXXPXXXXXGDXH** that chelates the second structural Zn^{2+}-ion which stabilizes the matrixin loop (Lovejoy *et al*, 1994, Dhanaraj *et al*, 1996a, b), is not found in either of the ADAMs, indicating a putative adamalysin-fold: where the matrixin-loop is replaced by a long amphiphilic helix commonly found in adamalysins (Dhanaraj *et al*, 1996a, b).

According to Stocker *et al.* (1995) who thoroughly mapped the various differences in primary structure among the metzincins, the residues found at the Z-position in the HEBXHXBGBXHZ-motif indicate the sub-family into which a metzincin protein can be classified. At the Z-position, adamalysins have an aspartic acid residue (Johnson *et al*, 2000), which is found in both ADAM-9 and ADAM-10. The B-positions indicate differences within sub-families, which tend to vary more (Stocker *et al*, 1995). At the first B-position, ADAM-10 has a valine residue, not found in any of the other metzincins aligned or in the eight matrixins aligned by Gall and colleagues (2001). Most interestingly, at the third B-position, ADAM-10 differs from all the other metzincins in the alignment by having a polar residue (Ser) instead of the expected hydrophobic residues found in all metzincins and matrixins (Stocker *et al*, 1995), indicating an exclusive divergence of ADAM-10. At the Met-turn motif, UBMOX, more diverse residues occur among the metzincins. However, a certain pattern has been suggested to distinguish these from each other (Stocker *et al*, 1995). In adamalysins the U-position is typically a cysteine residue, which is found in ADAM-9, while the O-position varies in the three adamalysins. In TACE and ADAM-10 a tyrosine residue occurs both at the U- and at the O-position while matrixins have an invariant alanine and proline respectively (Gall *et al*, 2001). These traits indicate that ADAM-9 is most

similar, sequence-wise, to snake venom adamalysins, while ADAM-10 is a reprolysin like ADAM-17.

The phylogenetic tree (Fig 5.2) shows similar features as deduced from the criteria given by Stocker *et al.* (1995). However, new evolutionary traits of this protease family arise; the entire adamalysin/MMP branch derives from a common ancestor with serralysin, one of the oldest of the metalloproteases. According to the analysis, ADAM-9 has developed in its own direction ancestral to the adamalysin-II/acutolysin-C branch, while TACE and ADAM-10 appear to be derived from the early stages of the evolution of the adamalysin group, where they converged in a common subgroup - the reprolysins. The differences between this subgroup and the remaining adamalysins can be seen mainly in a large insert ~20 residues before the catalytic Zn^{2+}-binding motif found in TACE and ADAM-10.   This region is equivalent to the segment that contains the S3'-subsite in MMPs (Holtz *et al*, 1999), suggesting a probable difference in specificity at this position. Another distinguishing feature about ADAM-10 and TACE is an insert right before the conserved phenylalanine (F125 on Fig 5.1). Adamalysin II, acutolysin-C and ADAM-9 have a conserved AQL-triplet, which is totally absent in the other metzincins, while ADAM-10 and TACE have an entirely different composition and a larger insert.

How this insert and other motifs presented in this section assemble when the protein is modelled are important to find out in order to understand as much as possible about the binding mechanism of these enzymes. The next section introduces the structures of the models of ADAM-9 and ADAM-10.

```
TACE-ADAM-17_1bkca      1    ..MKNTCKLLVVADHRFYRYMGRGEESTTTNYLIELIDRVDDIYRNTAWDNAGFKGYGIQIEQIRILKSP
ADAM-10_NP_001101       1    ..EKNTCQLYIQTDHLFFKYYGTREAVIAQ..ISSHVKAIDTIYQTTDFS..GIRNISFMVKRIRINTTA
Acutolysin-C_1quaa      1    PAPQTSIELFLIVDHSMYAKYNSNSSKITTT.LKARVNIMNAIYS.......SLNLVITLSGIEMWSAA
Adamalysin-II_3aig      1    NLPQRYIELVVVADRRVFMKYNSDLNIIRTR.VHEIVNIINEFYR.......SLNIRVSLTDLEIWSGQ
ADAM-9_NP_003807        1    LPQTRYVELFIVVDKERYDMMGRNQTAVREE.MILLANYLDSMYI........MLNIRIVLVGLEIWTNG
MMP-3_1b3d              1    ..........................FRTFPGIPKWRKTHLTYRIVNYTPDLPK........DAVDSAVEKALKVWEEV
Collagenase_AAB36941    1    ..............................GNPRWEQTHLTYRIENYTPDLPR........ADVDHAIEKAFQLWSNV


TACE-ADAM-17_1bkca     69    QEVKPGEKHYNMAKSYPNEEKDAWDVKMLLEQFSFDIAEEASKVCLAHLFTYQDFDMGTLGLAYVGSPRA
ADAM-10_NP_001101      65    DEKDPTNPFR.....FPN.....IGVEKFLELNS...EQNHDDYCLAYVFTDRDFDDGVLGLAWVGAP.S
Acutolysin-C_1quaa     62    DLITVQSSSRNTLKLFAS.....WRETDLLKRTS.........NDNAQLLTATNFNGNTVGLAYLKTMCN
Adamalysin-II_3aig     62    DFITIQSSSSNTLNSFGE.....WRERVLLTRKR.........HDNAQLLTAINFEGKIIGKAYTSSMCN
ADAM-9_NP_003807       62    NLINIVGGAGDVLGNFVQ.....WREKFLITRRR.........HDSAQLVLKKGFGG.TAGMAFVGTVCS
MMP-3_1b3d             46    TPLTFSRLYEGEADIMIS........................FAVREHGDFYPFDGPGNVLAHAYAPGP
Collagenase_AAB36941   41    TPLTFTKVSEGQADIMIS........................FVRGDHRDNSPFDGPGGNLAHAFQPGP


                                                                      HEXXHXXGXXH
TACE-ADAM-17_1bkca    139    NSHGGVCPKAYYSPVGKKNIYLNSGLTSTKNYGKTILTKEADLVTTHELGHNFGAEHDPDGLAECAPNED
ADAM-10_NP_001101     121    GSSSGGICEKSKLYSDGKK.KSLNTGIITVQNYGSHVPPKVSHITFAHEVGHNFGSPHDSG..TECTPGES
Acutolysin-C_1quaa    118    SKY...................SVGLIQDH....SAIPLLMAVTMAHELGHNLGMNHD.GAGCS......
Adamalysin-II_3aig    118    PRS...................SVGIVKDH....SPINLLVAVTMAHELGHNLGMEHD.GKDCLRG....
ADAM-9_NP_003807      117    RSH...................AGGINVFG....QITVETFASIVAHELGHNLGMNHDDGRDCSCG....
MMP-3_1b3d             91    GING..................DAHFDDDEQWTKDTTGTNLFLVAAHEIGHSLGLFHS............
Collagenase_AAB36941   86    GIGG..................DAHFDEDERWTNNFREYNLHRVAAHELGHSLGLSHS............


                           UBMOX
TACE-ADAM-17_1bkca    209    ......QGGKYVMYPIAVSGDHENNKMFSQCSKQSIYKTIESKAQECFQERS.........
ADAM-10_NP_001101     188    KNLGQKENGNYIMYARATSGDKLNNNKFSLCSIRNISQVLEKKRNNCFV............
Acutolysin-C_1quaa    158    .......CATCIMAPVLSSG...PAKSFSDCSKHDYQSFLTIHKPQCLLN..........
Adamalysin-II_3aig    160    .......ASLCIMRPGLTPG...RSYEFSDDSMGYYQKFLNQYKPQCILNKP.........
ADAM-9_NP_003807      160    .......AKSCIMNSGASG.....SRNFSSCSAEDFEKLTLNKGGNCLLNIVAALYLVCGA
MMP-3_1b3d            131    ......ANTEALMYPLYHSLTDLTRFRLSQDDINGIQSLYGPPPD.................
Collagenase_AAB36941  126    ......TDIGALMYPSYTFSGDVQ...LAQDDIDGIQAIYG......................
```

Figure 5.1. Sequence alignment of the catalytic domains of the most distinct ADAM and MMP-members.

Red marked segments indicate completely preserved motifs, which of the catalytic HEXXH-motif. Collagenase is synonymous with MMP-2.

Figure 5.2. Phylogenetic tree for the various members of the metalloprotease superfamily.
Two distinct branches appear here, the thermolysin branch (Meprin, Flavastacin and thermolysin), and the serralysin branch, which is paternal to the reprolysins (ADAM-10 and ADAM-17), the adamalysins (ADAM-9, Adamalysins, Acutolysin and many more members) and the matrix metalloproteases (MMP-3 and Collagenase). (Computed with the Trace-Suite software, Innis *et al*, 2000).

5.1.2   Protein modelling

The modelling of ADAM-9 and ADAM-10 showed that these catalytic domains are most similar to the snake venom metalloprotease Atrolysin-C (rmsd 2.2Å) and TNF-alpha Converting Enzyme (rmsd, 3Å), respectively.

Available structural data of Adamalysin II, Atrolysin-C, Acutolysin-C, and TACE from the Protein Databank gives insight in the coordination of disulphide bridges in ADAM-9 and ADAM-10. ADAM-9 has the same coordination of disulphide bridges as found in its primary template Acutolysin-C (32.6% identity): a) Cys322-Cys401, b) Cys363-Cys385 and c) Cys365-Cys370, where a) connects the long loop (Phe317-Ala327) that follows the anti-parallel substrate-binding β-sheet (Table 5.1) with the N-terminal region, b) and c) act on a 15-residue segment between the third histidine from the Zn-chelating triad and the Met-turn that is pivotal for providing additional stability in this otherwise flexible region (simulation without disulphide bridges, data not shown).

The structure of TNF-alpha converting enzyme (TACE, ADAM-17) was the singular template for the modelling of ADAM-10 (30% identity, 49% similarity). Like its template, ADAM-10 has three disulphide bridges, a) Cys221-Cys312, b) Cys343-Cys450 and c) Cys398-Cys434; where a) connects the second and the third antiparallel β-sheet of the classic five β-sheets found in metalloproteases. This disulphide bond is found in neither adamalysins nor matrixins, and is a probable signature of reprolysins. The second bridge, b), connects the C-terminus with the loop that follows the substrate-binding antiparallel β-sheet as in adamalysins (Table 5.1), and the third, c) connects the start of the C-terminal helix with the turn that precedes the Met-turn, a disulphide bond also found in adamalysins.

Table 5.1. Residues from the β-sheets involved in substrate binding via H-bonds.

The residues of the substrate binding β-sheets are shown before and after 2 nanosecond molecular simulations. Residues marked with an asterisk indicate the corresponding residues from the reference simulation of MMP-3 crystallized by Gomis-Rüth *et al* (1997). The residues marked with ¤, indicate the singular H-bonding residues interacting with the substrate. Sim-1, first simulation; Sim-2, second simulation.

ADAM-9	Before 2ns	After 2ns
Substrate-binding beta-sheet	Gly311-Ala316	Sim-1: Gly311-Val318
		Sim-2: Gly314-Ala316
Substrate-binding beta anchor	Ser374¤	Sim-1: Asn373¤
		Sim-2: -
ADAM-10	**Before 2ns**	**After 2ns**
Substrate-binding beta-sheet	Gly325-Trp331	Sim-1: Gly325-Trp331
		Sim-2: Leu327-Trp331
Substrate-binding beta anchor	Ala421¤	Sim-1: Ala418-Ala420
		Sim-2: Ala418-Ala420
MMP-3	**Before 2ns**	**After 2ns**
Substrate-binding beta-sheet	Asn162-Ala169*	Asn162-Ala169
Substrate-binding beta anchor	Pro221-Tyr223*	Pro221-Tyr223

5.1.3   Enzyme-substrate building.

The modelling of the enzyme-substrate complex was built for MMP-3 on the structural basis of the MMP-3/TIMP-1 complex (Gomis-Rüth *et al*, 1997) and with the synthetic hydroxamate inhibitor (Chen *et al*, 1999). Preliminary binding pocket analysis with VIDA (Openeye Scientific Software, Santa Fe, USA) shows that the binding pocket runs from the P'-region downwards pass the catalytic center forming a P-region (Fig 5.3). This P-region contributed partially to binding TIMP-1 inhibitor in a hydrophobic manner (Gomis-Rüth *et al*, 1997) and indicates therefore where the natural substrate orients itself closer to its N-terminal side. However, the exact position of the scissile bond and the torsions of the residues is not known from P1' to the P-region. In order to find a natural orientation, a comparison between the synthetic and natural inhibition was carried out. This comparison shows that the catalytic center is virtually identical in the natural and synthetic inhibition situation (left and middle structure in Fig 5.4), and the backbone binding confirms that the P'-region is stabilized by the same residues through backbone H-bonds in both a synthetic and natural inhibition. This accommodation combined with the position of the N-terminal nitrogen from TIMP-1 (Gomis-Rüth *et al*, 1997) shows that the scissile bond (drawn black line on right model in Fig 5.4) is located slightly "above" the Zn-ion, orienting the carbonyl oxygen towards it, yielding a Phi/Psi angle of -125°/145°, which is in the beta-sheet region. From this point toward the P-region, no orientation of the backbone is deducible for the substrate by combining transition state analogue binding knowledge with natural inhibition. However there is evidence in the MMP-3/TIMP-1 complex (Gomis-Rüth *et al*, 1997) about the sidechain interactions and one backbone H-bond on the P-side of the catalytic center (left structure in Fig 5.4): Ala167 from MMP-3 establishes one H-bond with Ser68 from the TIMP-1 inhibitor and Phe210 attracts the hydrophobic moiety of Glu67 from TIMP-1 (left structure in Fig 5.4). Closer investigation of possible backbone torsions shows that the next two residues on the substrate (P2 and P3) can satisfy these empirical criteria through a beta-sheet torsion (right structure in Fig 5.4). The four experimental residues were therefore added in a beta-sheet manner, which provided favourable H-bonded interactions between the P2 backbone atoms and Ala167 (right structure Fig 5.4) and hydrophobic attractions between Phe210 and the P3 residue. This orientation was the only orientation that satisfied three points: a) the modelled scissile bond's orientation is in accord with the reported synthetic transition state analogue and the natural inhibitor (Gomis-Rüth *et al*, 1997, Chen *et al*, 1999), b) the substrate must/should be in an extended beta-sheet conformation between the P1-P1' residues for proteolysis to occur (Tyndall and Fairlie, 2001, Glenn *et al*,

2002) and c) the orientation satisfied all criteria posed by the natural inhibition case of MMP-3/TIMP-1 (Gomis-Rüth *et al*, 1997).

At this stage, the addition of complementary residues was carried out in accord with section 4.1.5. This enzyme-substrate complex was then used to build the equivalent enzyme-substrate models for ADAM-9 and ADAM-10. In this way 1 crystal structure (MMP-3) and two models (ADAM-9/10) could be assessed for substrate specificity on the basis of this sections' results.
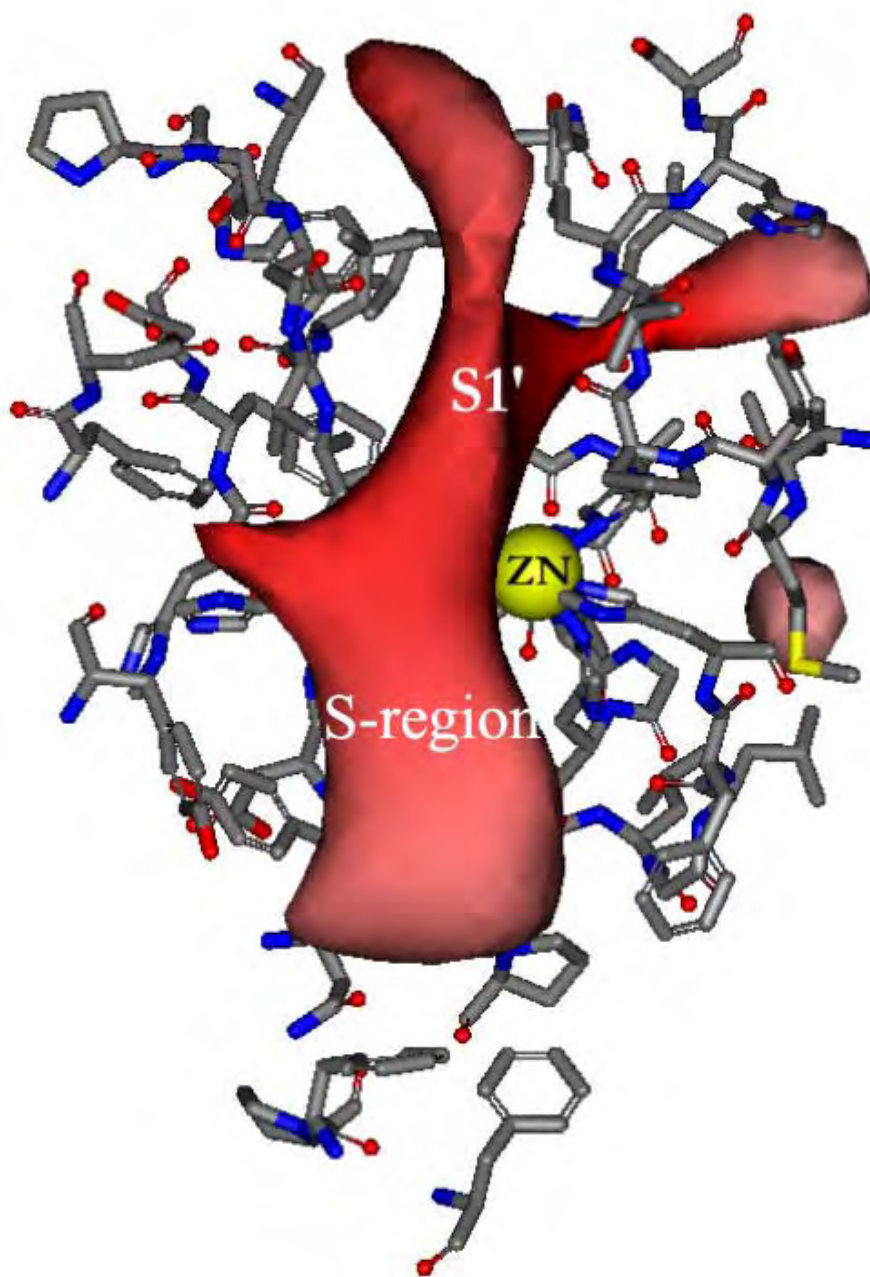
Figure 5.3. The hosting-space of the substrate-binding pocket of MMP-3.

This FRED-surface (Openeye Scientific Software, Santa Fe, USA) shows that there is a large binding space below the Zn-ion (yellow), which designates the P-region. This proves useful to determine where the substrate orients itself from the P'-region.
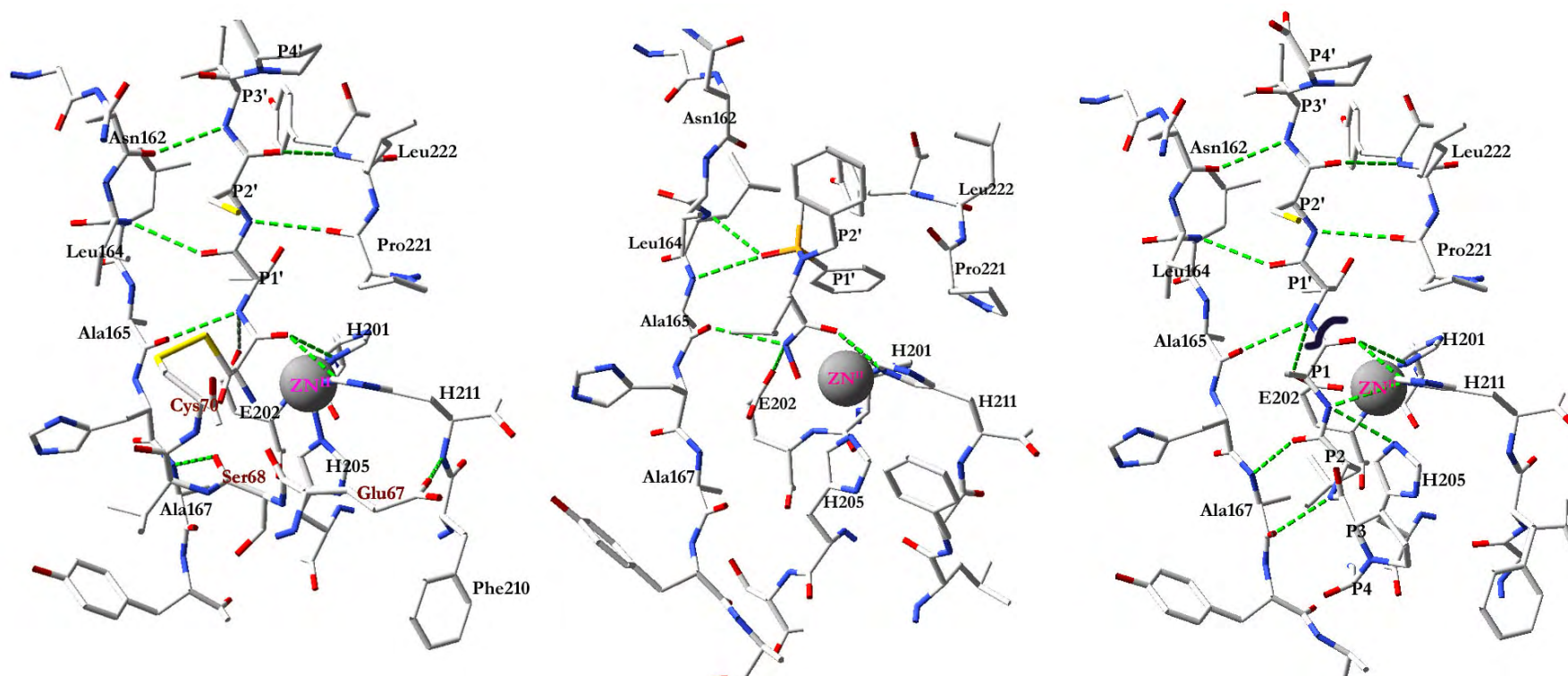
Figure 5.4. Building the enzyme-substrate complex for MMP-3.

Left, binding pocket and intramolecular H-bonds between the enzyme and the TIMP-1 interconnected P'&P-segments (noted with P1', P2', … on the P'-side and dark red residue names on the P-side) (Gomis-Rüth *et al*, 1997); middle, binding pocket of MMP-3 with hydroxamate inhibitor with intramolecular H-bonds (Chen *et al*, 1999); right, enzyme-substrate complex modelled on the basis of the two former complexes, scissile bond designated by black line. Residues of the catalytic center are E202, H201, H205, H211 and non-covalently binding residues are the trilateral noted residues.

5.1.4   Molecular simulations

The dimer of MMP-3&substrate had a stable conformation throughout the simulation (see Fig 5.5, 5.6): all rigid secondary structure elements remained well conserved and the final overall RMSD between the enzyme at the start and end of simulation was 1.84Å. During the simulation of MMP-3 and the hypothetical substrate, the substrate remained in its original β-sheet conformation at the P'- region (Fig 5.5). The backbone H-bonds of the experimental P-region on the substrate of MMP-3 experienced a slight flexibility at its very end (P3&P4), however P2 and Ala167 remained H-bonded via their backbone atoms throughout the simulation. Contact maps (Fig 5.7) and visual inspection (Fig 5.5) of the simulated MMP-3&substrate complex shows that after 2 nanoseconds the S1'-cavity remained confined by Leu164, Val198 and Pro221 (Dhanaraj *et al*, 1996a,b), which confirms that the dynamics around this pocket are in accord with the empirical results before and after simulation.

The experimental P-region showed interesting results. The P1-sidechain (Ala) had alternative orientations of its methyl group towards the solvent, His166 and between Pro221 and His211. Given this "indecisive" behaviour of the P1 sidechain, and the weak chemical variability among the MMPs in this region it is possible that no specific S1-position, or "pocket" occurs in these enzymes, something that may fit with the delicate dielectric environment required by the catalytic Zn-ion to function (being central in this region). Moving down the sequence, the P2 residue found a hydrophobic shelter, defined mainly by Ala169 and the hydrophobic moiety of Ser206, while the P3 residue experienced a hydrophobic dominance from Phe210, residues that are involved in the binding of TIMP-1 (Gomis-Rüth *et al*, 1997). The very end of the substrate, P4, interacted also with Phe210 to some extent but mostly with Phe86, which derives from the N-terminal coil. This can be seen by the dark region of the contact maps (Fig 5.7) at the first 12 residues of the simulations of MMP-3&substrate. The N-terminus was shown to be involved in binding TIMP-1 (Gomis-Rüth *et al*, 1997) and to be induced towards the binding cleft by peptidomimetics (Chen *et al*, 1999).

Conclusively for this crystal structure (with 5 added residues), the observed orientation of the sidechains from the P-region indicates that there is a large hydrophobic valley, which interacts with the P2/P3 groups. Residue Ala169, Ser206 and Phe210 define this S2/S3 valley. This region, as we shall see, varies in ADAM-9 and ADAM-10, and perfectly hosts the P2/P3 positions of their substrates in the same manner.

In the two simulations of ADAM-9 in complex with its substrate the RMSD profiles were quite similar, however they deviated from the crystal structure of MMP-3&substrate by ~1Å overall. This higher RMSD than the crystal structure was mainly caused by a flexibility of the region containing the Met-turn and the parallel β-anchor (red area figure 5.8), while the α helices and β-sheet were preserved well throughout the simulations (Fig 5.8).

The H-bonding network between the substrate and the enzyme remained intact in the simulation. The P4 backbone atoms established a "new" H-bond with the backbone atom of Val318, while the P2' residue exchanged a backbone H-bond with the parallel β-anchor for an H-bond with the sidechain of Asn373 from the same anchor. The remaining residues remained in the original H-bonded state, meaning that the substrate remained well preserved structurally, and its sidechains could be analysed for subsite interactions.

The subsite interactions at the S1' pocket in ADAM-9 were defined by Ala313, Ile344 and Asn373 (equivalent to Leu164, Val198 and Pro221 in MMP-3, Fig 5.5). The main specificity determinants here are Ala313 and Ile344 and given the hydrophobic identity of the P1' sidechain (Val), the interactions were favourable. Some polarity occurs in this pocket, represented by Ser343, which orients its sidechains inwards in the enzyme. However after analysing the two simulation outputs, this residue does not correspond to the main determinants noted for MMP-3 by Dhanaraj *et al* (1996a, b), but could still play a role in selectivity.

The S1 position from ADAM-9, which is methionine 315 (residue 166 by MMP#), experienced some interaction with the solvent, in addition the interaction to the tyrosine residue from the substrate (P1). However, at the end of both simulations the tyrosine residue had its sidechain inclined toward Met315, suggesting a slight flexibility of this P1-position as in the reference run, which indicates no specific P1-S1 interaction.

The new S2/S3 pocket was here quite different from MMP-3. Val318, Asn352 and Asn356 shaped this region in ADAM-9 analogously to Ala169, Ser206 and Phe210 in MMP-3. This trend of the P2/P3 residues orienting their sidechains in this wide but shallow cavity occurred in both simulations, in an analogous manner to the reference simulation (Fig 5.5).

The simulations of ADAM-10 produced two conformers with a relatively high difference from the input conformation (RMSD: 3.37Å and 3.33Å from the input, fig 5.9). This was mainly perpetuated by an interesting closure of the long reprolysin loop, which originally was modelled without a strong homologue template segment (red region Fig 5.9).

The substrate of ADAM-10 remained more firmly bound on the P'-region than the P-region as also observed in the reference run of MMP-3&substrate, but experienced a loss of one of the initial H-bonds on the P-region (between P2- and backbone atoms of Ala330). The reason for this was the P2 residue which, being a glycine without a sidechain, contributed with flexibility to this part of the substrate, and allowed an inverse β-sheet torsion exposing its miniscular sidechain towards the solvent. This was perpetuated by the positive charge field from the Zn-ion on the backbone oxygens from the central P-residues (P1 and P2) yielding two carbonyl oxygens in coordination with it.

The primary specificity pocket, the S1' pocket, was here mainly determined by the sidechains of Leu327, Thr379 and Ala418, with Thr379 inclining its hydroxyl group away from the hydrophobic P1' residue (Val) (Fig 5.5). The S1-position was represented by Leu329, which opposes the Zn-ion over the substrate and is equivalent to His166 from MMP-3/substrate and Methionine 315 from ADAM-9. Again, an interaction with the solvent by this side chain was also observed here, and shows that this position does not maintain specific interactions.

Given the inverse β-sheet torsion of the glycine at the P2 position, the P3 residue re-arranged and established an H-bond with the carbonyl oxygen from Ala330 (which was originally in H-bond with the P2). By consequence, Val332 dominated the attraction on the P3-residue in both simulations. However, the S2/S3 valley was shaped here by Val332, Asn387 and Pro391 (Fig 5.5).

By noting the subsites in Table 5.2 and Figure 5.5, it is evident that the catalytic domains of ADAM-9 and ADAM-10 share deep S1'-pockets with MMPs, and other enzymes of the metzincin family. This pocket is entirely hydrophobic in MMP-3 and, seemingly, in ADAM-9 as well. ADAM-10 has one amphiphilic residue that occurs at the "default" S1' subsite, Thr379, which distinguishes it from both MMP-3 and ADAM-9 in primary selectivity.

The S1'-cavity of ADAM-9 is also quite similar to the S1' pocket of MMP-3 in terms of space, and, given its hydrophobic properties, it would host analogous residues as does MMP-3. The S1'

cavity in ADAM-10 is smaller than the two others, and could host smaller sidechains such as serine and threonine, satisfying the potential H-bonding with Thr379.

Quite clearly there are differences in primary specificity between ADAM-10, ADAM-9 and MMP-3. For perceiving these differences better, a comparison to the serine protease family would place ADAM-10 away from MMP-3 and ADAM-9 in an analogous way as chymotrypsins and differ from trypsin (cleave small hydrophobic and bulky ones with a polar tip respectively - Perona and Craik, 1995).
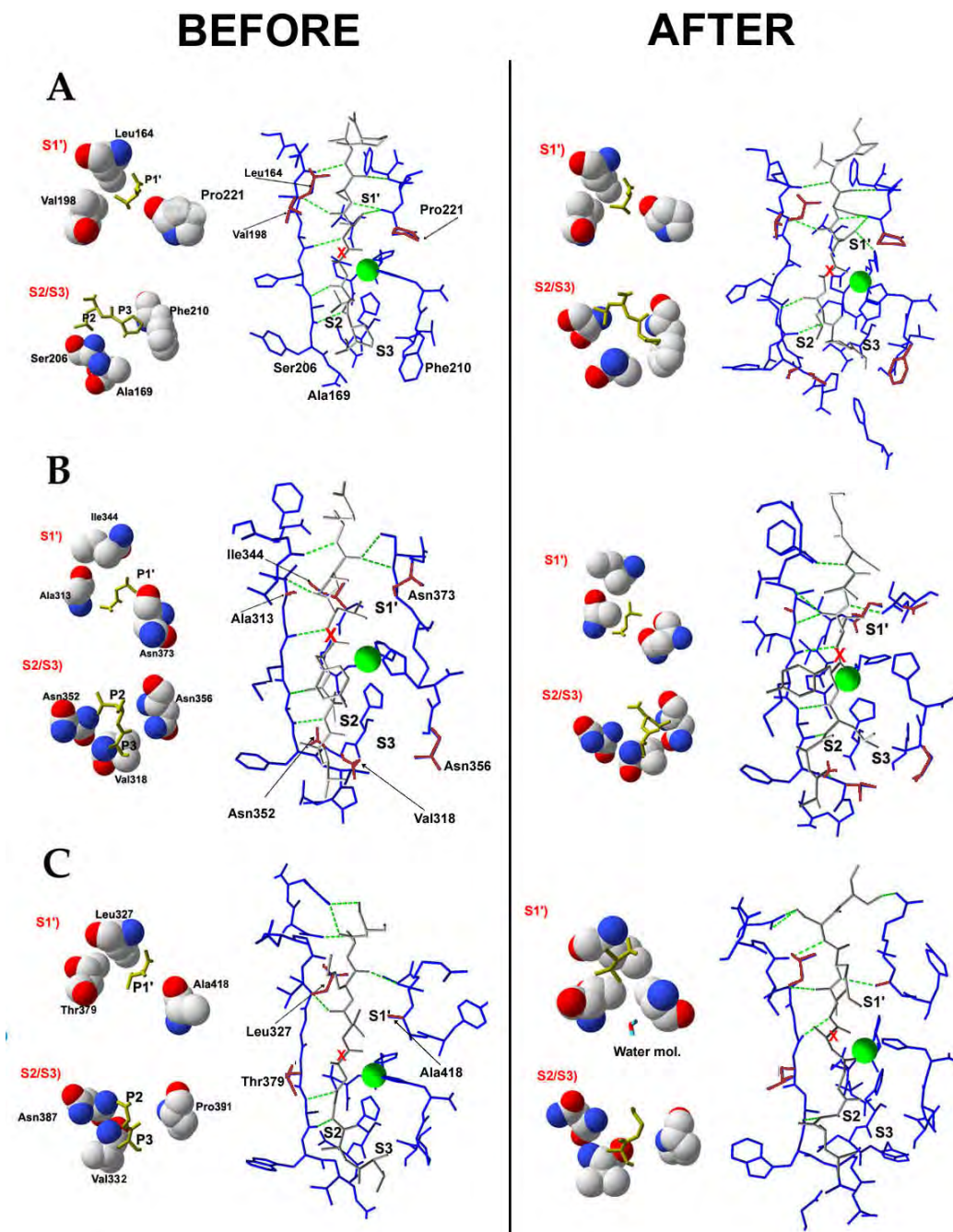
Figure 5.5. Structural changes at the binding pockets of MMP-3, ADAM-9 and ADAM-10.

A, MMP-3; B, ADAM-9; C, ADAM-10. Left half shows before simulation; right half, after simulation. The two columns show the binding pockets with the H-bond network in green, and the substrate recognizing subsites, denoted with S1' and S2/S3. A water molecule is found in the lower left S1' subsite, shown in stick mode. Green ball is Zn-ion. <u>Note</u>: For graphical purposes, ball&stick and CPK are *not* necessarily superimposed.
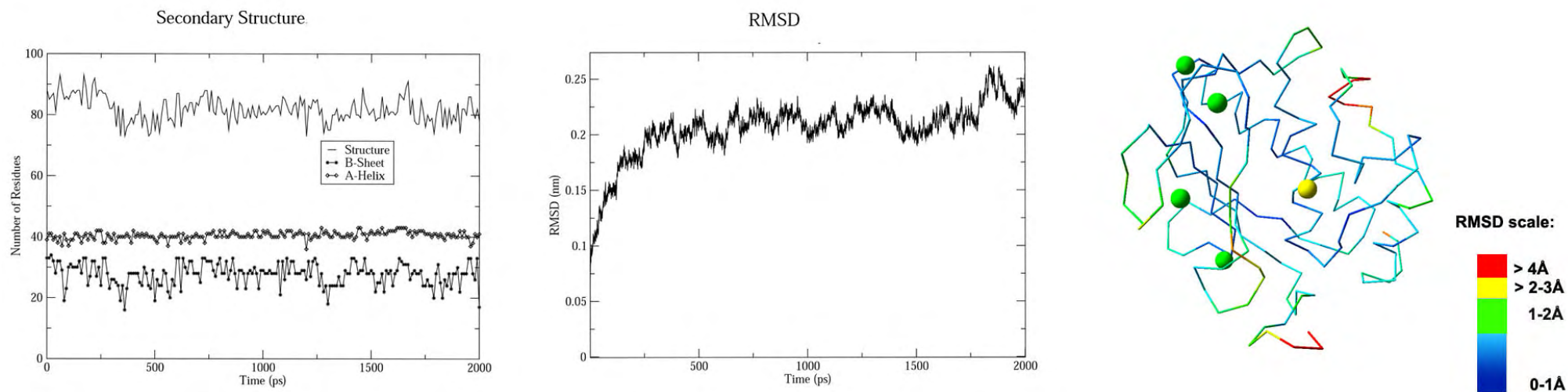
Figure 5.6 Structural fluctuations of the crystal structure of MMP-3 in complex with the hypothetical substrate.

Left graph shows the secondary structure-monitoring throughout the designated time-interval (x-axis) in number of residues (y-axis) in a given conformation (label box). Right graph illustrates the overall RMSD fluctuations from the original crystal structure of MMP-3&substrate, while the right C-alpha traced structure shows the structural fluctuations as colour-gradient on the structure, with the RMSD-scale next to it. Yellow ion is the catalytic Zn-ion, while the other green ions are structural ions.
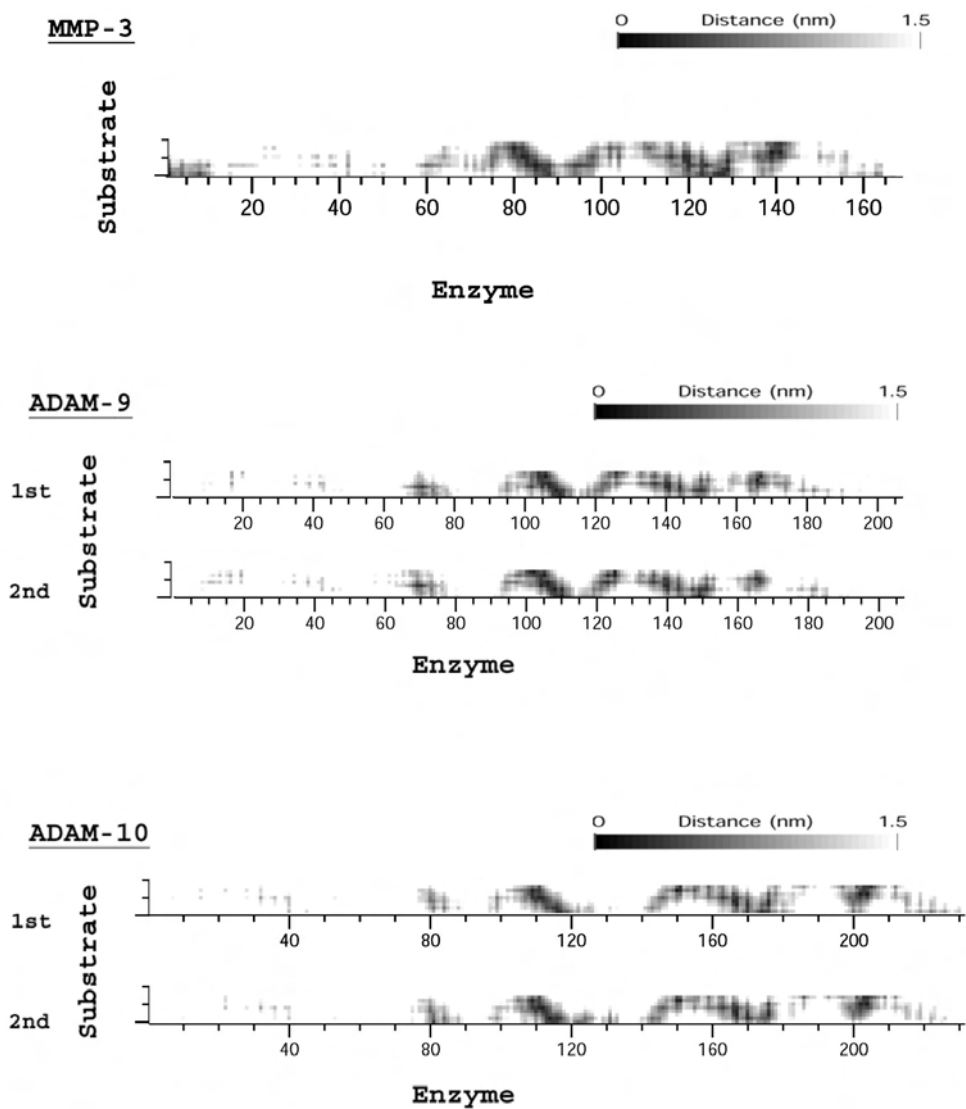
Figure 5.7. Contact maps for enzyme-substrate interactions.

Distance interaction matrix computed over the last nanosecond of the simulations for each dimer. The two matrices for each dimer illustrate the short distances (dark spots) and longer distances (lighter spots) between the interacting residues on the enzyme (horizontal) and on the substrate (vertical).
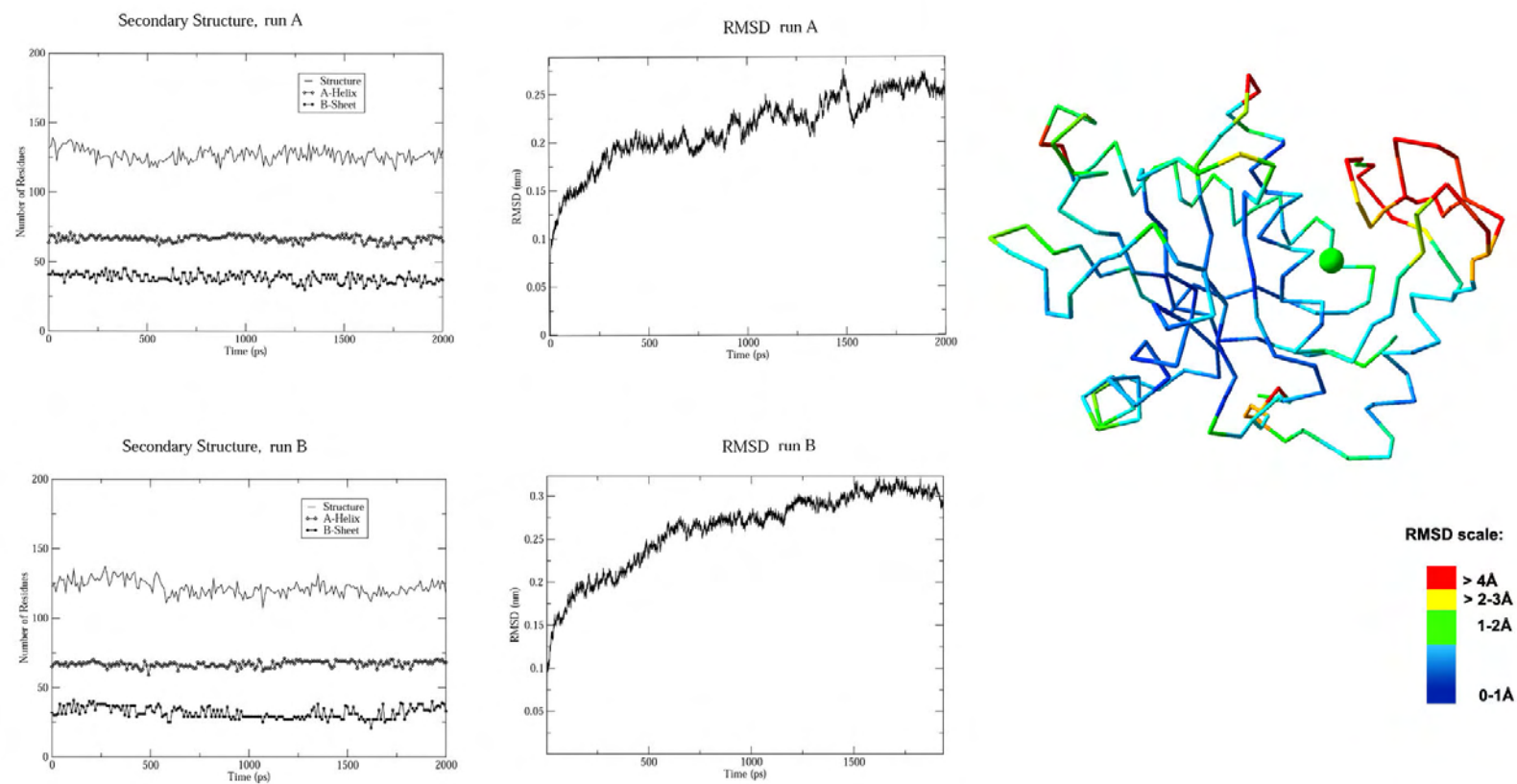
Figure 5.8. Fluctuations of the ADAM-9 model in complex with its hypothetical substrate.

Left graphs and right graphs, as in Fig 5.6; upper row and lower row, 1st and 2nd simulation. The illustration on the right shows the C-alpha trace of ADAM-9 coloured by RMSD-scale (bottom right) which shows those regions that fluctuated most. The green ion is not coloured by RMS scale but designates the catalytic Zn-ion.
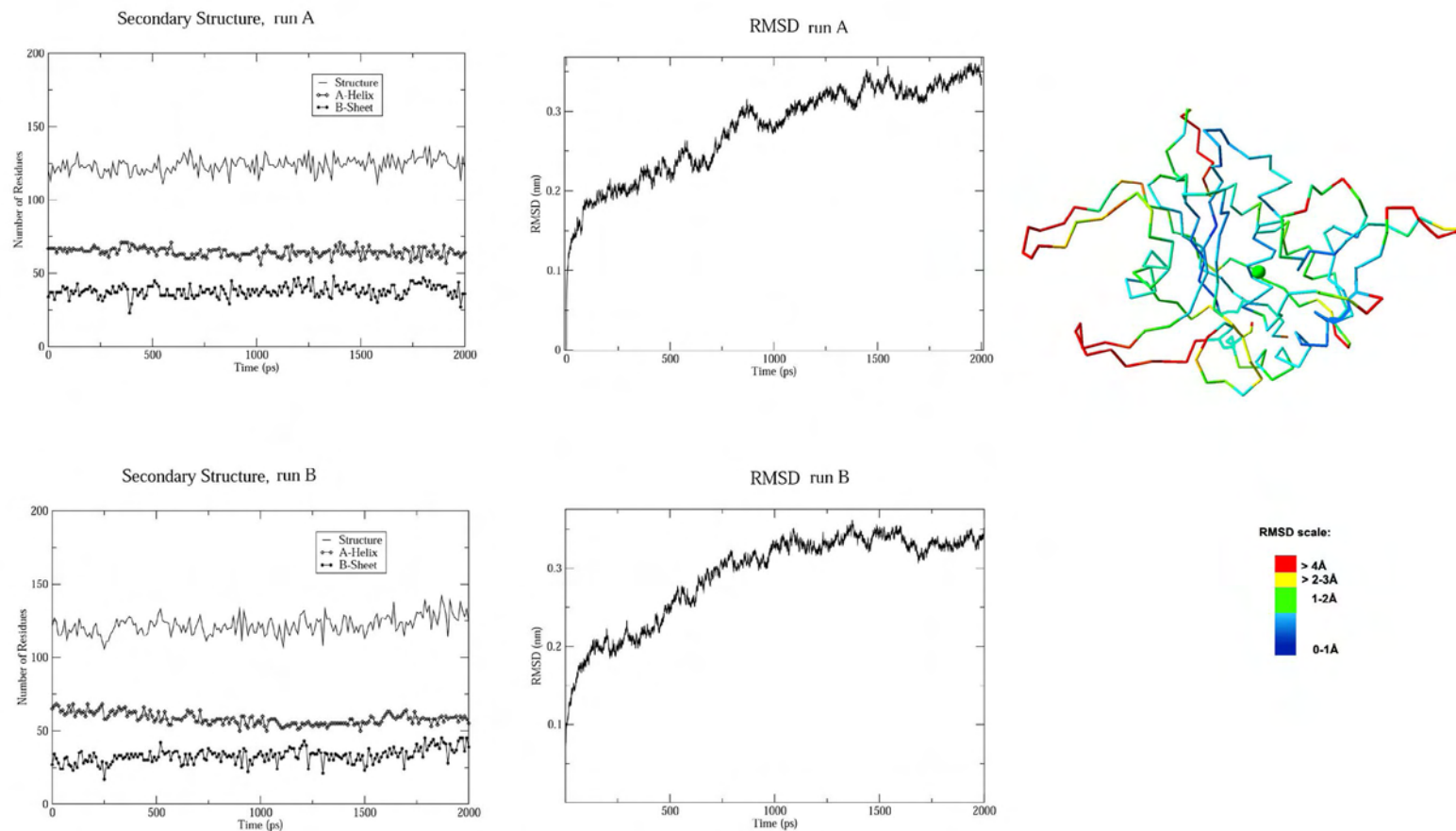
Figure 5.9. Fluctuations of the ADAM-10 model in complex with its hypothetical substrate.

Left graphs and right graphs, as in Fig 5.6; upper row and lower row, 1st and 2nd simulation. The illustration on the right shows the C-alpha trace of ADAM-10 coloured by RMSD-scale (bottom right) which shows those regions that fluctuated most. The green ion is not coloured by RMS scale but designates the catalytic Zn-ion.

Table 5.2. Determined subsites for MMP-3, ADAM-9 and ADAM-10.

The table represents the most occurring interacting residues from the enzyme, during the 2ns-simulation. Common residues that occurred twice are considered as subsites. The residues marked with an asterisk indicate the interactions that were computed based on the crystallized P'-region of MMP-3/TIMP-1 (Gomis-Rüth *et al*, 1997).

MMP-3	S4	S3	S2	S1		*S1'	*S2'	*S3'	*S4'
Sim-1	Phe210, Phe83	Phe210, Ala169	Ala167 Ser206	Solv, His166, His211		Leu164, Val198, Pro221	Solv, Val163	Solv, Leu164	Solv, Leu222

ADAM-9		S4	S3	S2	S1	S1'	S2'	S3'	S4'
Sim-1		Phe317, Val318	Val318 Met315	Val318, His351, Asn356	Met315, Solv	Ile344, Ala313, Asn373	Solv, Asn373, Thr312	Solv Ser374,	Solv, Gly310
Sim-2		His357, Solv	Tyr-P1, Met315	Val318, His351	Met315, Solv, His357	Ile344, Asn373, Solv	Solv, Asn373, Thr312	Solv, Ser374, Phe333	Solv

ADAM-10	S4	S3	S2	S1	S1'		S2'	S3'	S4'
Sim-1	Pro391, Val332	Val332, Pro391	Val332, Solv	Leu329, Solv	Leu327, Thr379, Ala418		Solv, Val326	Solv, Asn366	Solv, Thr421
Sim-2	Val332, Trp331	Val332, Asn387	Solv, His392	Leu329, His392, Solv	Leu327, Thr379, Ala418		Solv, Val326	Solv, Asn366	Solv, Arg421

All three simulated enzymes bound the substrate in a similar fashion. The familiar shape of the distance matrices, reproduced in all five simulations, illustrates this. In MMP-3, the N-terminus was shown to play a possible role in substrate binding, as can be seen by the dark region at the first 13 residues of the simulations of MMP-3&substrate, as also observed in the case with TIMP-1 by Gomis-Rüth *et al.* and with synthetic inhibitors (Chen *et al*, 1999). This interacting role of the N-terminus did not occur in either the simulations of ADAM-9&subtrate or ADAM-10&substrate, because of the different orientation of the N-terminus in adamalysins & reprolysins, which is related to the presence of the disintegrin and the cysteine rich domains, which MMPs do not have.

The H-binding mechanism in metzincins was shown here to be crucial to maintaining the substrate in position for a proper Zn-carbonyl (substrate) and Zn-Glu (catalytic) interaction. This was observed during preliminary simulations, which we carried out without optimised H-bonds at both the P- and P'-regions. The preliminary simulations of these experimental complexes *in vacuo* showed that the lack of H-bonds promoted a substrate that did not remain properly in contact with the enzyme and had flexible tendencies (data not shown). As a result, the weakly bound substrate without proper H-bonds inhibited the glutamate-Zn movement. Instead, the final complexes with an optimised H-bond network between the P'-region of the substrate and the antiparallel β-sheets (Table 5.1) showed a substantially "higher quality" of binding: the substrate was extended and maintained in a β-sheet conformation by the enzyme throughout the 2ns simulation in solvent.

Conclusively, the P-region from a substrate is likely to have the two H-bonds with residue 167 (MMP numbering) as modelled here (in a complementary manner to the substrate-binding β-sheet). The putative S2/S3-pocket is suggested to be the main non-covalent anchor of a natural extended substrate on the P-side. A similar anchoring by induced fit is reported in serine proteases, where the S1 and the S1' are the initial specificity sites for the substrate, and the oxyanion hole is the main stabilizer for catalysis (Perona and Craik, 1995, Kishi *et al*, 1997). Because there is no explicit S1 pocket here and a catalytic Zn-ion requires a delicate dielectricity, it is suggested that the main induced fit on both sides is perpetuated by the S1' and S2/S3 regions, and maintained by the backbone hydrogen bonds, between P1 and residue 165 (Johnson *et al*, 2000), and P2 and residue 167.

This S2/S3-pocket differs in composition in the three test cases, something which makes it attractive as a drug target pocket, and could be taken into account for future inhibitor design aspects, because targeting only one pocket raises the potential for lower selectivity and higher toxicity.

5.1.5   *The catalytic centre – a new catalytic mechanism*

During the MD simulations the catalytic Zn^{2+} attracted the side chain of the catalytic glutamate, from the Zn-binding motif in all 5 simulations. This glutamate movement has been previously observed in MMP-3 and carboxypeptidase A (Gomis-Rüth *et al*, 1997, Mustafi and Makinen, 1994).

Parkin (2001) reported that the Zn^{2+}-ion in MMPs is likely to be tetra-coordinated with the His-triad and a water molecule, at *ground* state, which is in accord with the results supplied by Alberts and colleagues (1998). However, some structures of MMP-holoenzymes show that the catalytic Zn^{2+}-ion is tri-coordinated with the three His-residues, even in the presence of adjacent water molecules (Chen *et al*, 1999, Dhanaraj *et al*, 1999). Interestingly, Gall *et al* (2001) reported that the catalytic zinc in MMP-11 is tetra-coordinated when complexed with an inhibitor mimicking the transition state, and all water molecules are expelled from the Zn-ion. This trend was repeatedly observed also by other groups who crystallized various metzincins in complex with other transition state analogues (Dhanaraj *et al*, 1996a, b, Seltzer *et al*, 1989, Pavlosvsky *et al*, 1999). This represents at least 6 different pieces of evidence that water is not bound to the Zn-ion during a transition state for these enzymes. Additionally, the (catalytic) glutamic acid side chain oxygen atoms are observed in various coordinations with the Zn, some with both oxygens coordinated, and others with one of the epsilon oxygen atoms. The Zn-ion is therefore suggested to be penta-coordinated at transition state due to the abolished water-molecule, primarily because of the large amount of evidence with expelled waters (Dhanaraj *et al*, 1996a, b, Seltzer *et al*, 1989, Borkakoti *et al*, 1994, Gall *et al*, 2001, Gomis-Rüth *et al*, 1997, Pavlosvsky *et al*, 1999), and also because such an intrusion was not observed in our simulations.

The coordinators of the Zn-ion at transition state are therefore proposed to be the two ε-nitrogens from the short spacer, the P1'-carbonyl and the two oxygen atoms from the catalytic glutamate – (based on the putative flexibility of the third histidine from the long spacer - Parkin *et al*, 2001). This set of empirical and computational results and observations, leads to the suggestion of a new catalytic mechanism for metzincins, based on the formation of an anhydride intermediate (observed for carboxypeptidases by Mustafi and Makinen, 1994), which supports a) the various

inclinations observed between the two oxygens (from the catalytic glutamate), the catalytic Zn-ion (Dhanaraj *et al*, 1996a, b, Seltzer *et al*, 1989, Borkakoti *et al*, 1994, Gall *et al*, 2001, Gomis-Rüth *et al*, 1997, Pavlosvsky *et al*, 1999), b) the various crystal structures of transition state analogues with expelled water molecules (Dhanaraj *et al*, 1996a, b, Seltzer *et al*, 1989, Borkakoti *et al*, 1994, Gall *et al*, 2001, Pavlosvsky *et al*, 1999) and c) the observed flexibility of the catalytic glutamate to approach the Zn-ion both in our simulations and empirically (Gomis-Rüth *et al*, 1997), and d) the hypothesized flexibility of the third histidine (Parkin *et al*, 2001).

In this mechanism (Fig 5.10), the catalytic glutamate moves in coordination with the Zn-ion where a polarization effect is induced on the glutamate epsilon oxygen, leading to the formation of a nucleophilic oxygen atom. The carbon atom from the scissile bond, which is polarized through its carbonyl oxygen by the Zn^{2+}-ion, becomes a weak carbocation, and accepts an electron from the polarized nucleophile ε-oxygen of the catalytic glutamate. This forms a transient bond between the enzyme and the substrate creating an oxyanion-tetrahedral anhydride intermediate, a formation investigated for carboxypeptidase A and discussed extensively (Mustafi and Makinen, 1994, Mock, 1998). Because of the established anhydride bond we believe this increases the polarization effect from the Zn-ion on the second epsilon-oxygen, which facilitates the electron transfer from the carbonyl carbon back to the glutamate first ε-oxygen, breaking the covalent E-S state. In concert with this, the third histidine establishes an H-bond with the amide nitrogen, and a water molecule close to the amide bond (observed in our simulations) acts as an acid upon the nitrogen atom, creating the leaving group. The resulting hydroxyl group acts as a base on the carbonyl thereby yielding the new C-terminus.

Some sources have suggested that a water molecule, polarized by the zinc-ion, forms an aggressive hydroxyl ion that nucleophilically attacks the scissile bond (Browner *et al*, 1995, Li Li, *et al*, 2000). On the other hand, Grams *et al*. (1995) suggested that the carbonyl carbon on the scissile bond, polarized by the Zn^{2+}-ion, becomes susceptible to nucleophilic attack from a nearby water molecule that is not in contact or polarized by the catalytic zinc ion, but rather, by the catalytic glutamate side chain. This is similar to the proposed mechanism for carboxypeptidase A (Mildvan, 1970). Our molecular simulations show that, in order to fit a water molecule between the substrate and the Zn^{2+}-ion at an optimal angle for both successful polarization [(through orientation and proximity effects (Koshland *et al*, 1972)] and maintenance of the crucial H-bonds between backbone atoms of the enzyme and the P'-region, an expansion of the binding pocket from 6.6Å to more than 11.4Å would be required. This is a dynamic movement comparable to

that seen in allosteric enzymes (Chothia *et al*, 1976). Most importantly, substrate analogues have been shown to expel a Zn-coordinated water molecule from coordination with the Zn-ion (Dhanaraj *et al*, 1996a, b, Seltzer *et al*, 1989, Borkakoti *et al*, 1994, Gall *et al*, 2001, Pavlosvsky *et al*, 1999). Additionally, the energy between Zn^{2+} and a water molecule is higher (+5kJ/mol) than the energy between Zn^{2+} and a carboxyl-group (as found in the catalytic glutamate residue), and 9kJ/mol higher than the interaction between Zn^{2+} and the ε-nitrogen from the histidine residues (Warshel, 1997). This indicates that the interaction between a water and zinc is unfavorable when there is one glutamate/aspartate-carboxyl group close enough to the Zn^{2+}-ion, as we have observed in the molecular simulations.
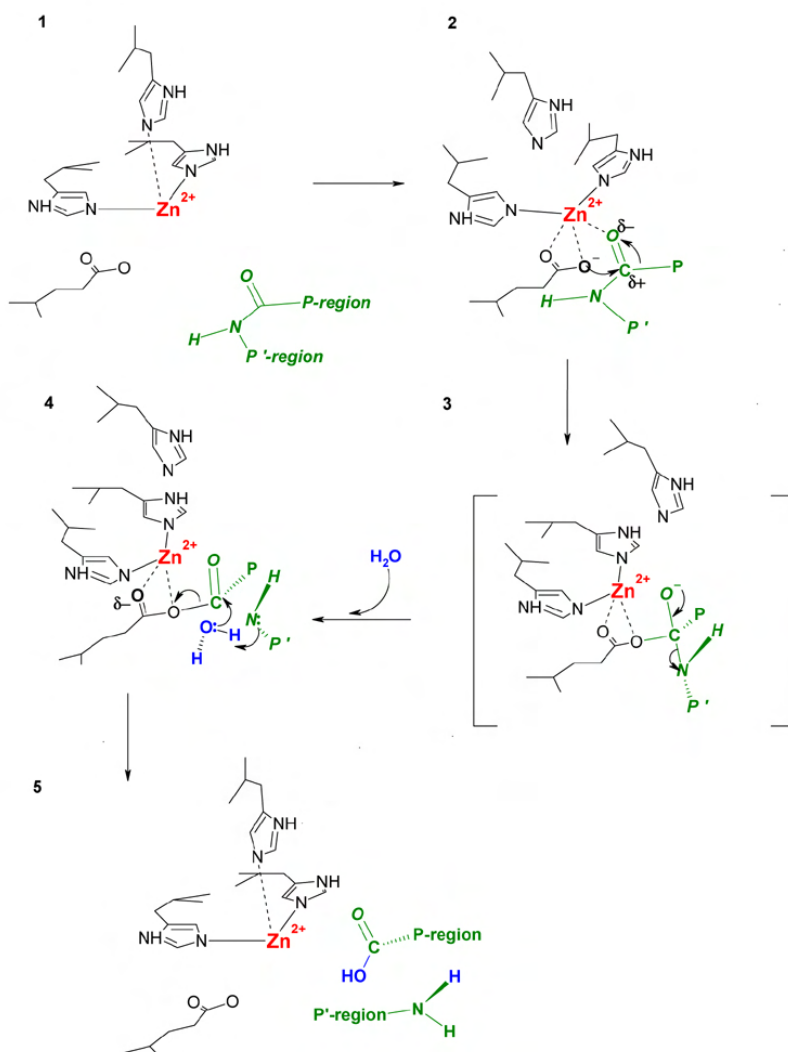
Figure 5.10. A suggested catalytic mechanism for the metzincin superfamily.
State 1: The Zn-ion is in its tri-coordinated resting state. *State 2*: The third histidine moves out of coordination from the Zn-ion [47], and the quasipenta-coordinated Zn-ion influences the Oε of the glutamate side chain to act as an aggressive base, which transfers an electron to the carbocation on the scissile bond (polarized by the Zn ion), creating an oxyanion transition state. *State 3*: The tetrahedral intermediate, the electron transfer leads the electron to the amide nitrogen given the high-energy state of the negatively charged oxygen, breaking the scissile bond. *State 4*: The anhydride bond breaks, given the inclination of the second glutamate oxygen to the Zn-ion, and a water-molecule protonates the leaving group and hydroxylates the new C-terminus. *State 5*: The Zn-ion goes back in coordination with the third histidine, and the cleaved substrate detaches.

CHAPTER 6

6.1   Importance of the subsite identification – is it useful?

The subsites identified on the enzyme in the previous sections for both kallikreins and metzincins are important to classify the specificity of these enzymes. Because of the strong sequence similarity between the 15 kallikreins (Harvey *et al*, 2000) the subsite positions have been deduced for the other members as well, using structure-to-structure comparison and alignment, which supplies further information regarding the various specificities of these other members. The subsites are of primary importance to anticipate correct oligopeptides or chromogenic substrates for *in vitro* assays. Combined with the knowledge of how oligopeptides behave in solution, the residual composition within the oligopeptide can also be suggested by properties in assembling rigid secondary structure (simple secondary structure prediction, see www.expasy.org for info). In this way, a beta-sheet conformation can easier be induced in solution, which is the default conformation of a substrate in serine proteases (Wilmouth *et al*, 2001). Although most oligopeptides behave quite irregularly in solution and are therefore harder to assay, typical beta-sheet-assembling oligopeptides do occur (van Gunsteren *et al*, 2001). The choice of the "right substrate" for an enzymatic assay can therefore be anticipated by the information supplied here and the residual prevalence of occurring in beta-sheets.

Indeed, Lu *et al* (2001) showed that the fold, sequence and mechanism of action of the serine proteases is so similar that they could predict the reactivities and free energies of binding for a oligopeptides to a serine proteases based on the *sequence alone*. At this stage, it is relevant to ask, why use the energy-based assessements reported here and the molecular simulations? First, the method supplied by Lu and colleagues (2001) was not successful in all cases. Second, the method does not take into account the dynamic effects on oligopeptides in solution and therefore quickly loses plausibility when this is not accounted for in the choice of a substrate. Third, and most important, it suggests nothing about the structural basis for the interaction, which is pivotal to understanding the binding mechanism of a protease and that can be anticipated with computer modelling.

Although the methods and results supplied here aid substrate predictions for *in vitro* assays, the strongest potential is to use these data to design selective inhibitors. These results were further

used to perform docking experiments (section 4.6) that lay a preliminary background for *in vitro* inhibition assays, which in turn provide results for clinical applications.

Physiological substrates have not yet been predicted solely on a computational basis. Some physiological substrates are usually other proteases that need to be activated or degraded, or structural proteins as collagen, titin etc. that need to be processed in the extra cellular matrix. The main difficulties with such predictions lie in the question of specificity: how much does specificity really tell about biochemical preference? This question can be answered by looking at the substrate-overlap for serine proteases in the regions except for the primary specificity pocket (P1-S1) (Coombs *et al*, 1998, Ding *et al*, 1995 and Corey *et al*, 1995). These data show that a high Michaelis-Menten constant (K_M) can be achieved in cases where the P2 and P2' positions are changed to 2-3 other residues. This indicates that although a subsite has a hydrophobic preference, and the physiological substrate candidate has its complementary part, another hydrophobic residue from another physiological substrate could really be suited at this position. This makes the identification of natural substrate *a priori* even more difficult and can be only kept on a weak-assumption level. In fact many enzymes do overlap in specificity but are kept away from metabolic interference through and tissue-specific expression. This is most likely the case for the kallikreins, given their presented similar subsites and their close sequence relationship.

Therefore the sequence of the putative substrate-segment alone is not sufficient for a good anticipation of a physiological substrate. Such a prediction would need to rely on additional factors as the exposure of the cleavage-segment to the solvent.

However, yet another problem arises at this stage. Many of the confirmed activation sequences found in the kallikrein family are often very hydrophobic (see Fig 4.2); because of their hydrophobic nature, these activation segments are therefore not exposed to the solvent. Indeed, by looking at the crystallized mature enzyme structure of mK8, the remaining part of the activation segment (P'-region) is buried 5-7Å in the enzyme (Kishi *et al*, 1999). This poses the question of whether the pre-pro-region is cut off before the mature peptide properly folds, or whether some quasi-allosteric changes are induced under the activator-zymogen interaction or eventually by chaperones during the proteolytic activation process.

In other words, unless a large amount of knowledge is present for a candidate physiological substrate and the enzyme (similar tissue expression, cellular localization, theoretical and empirical substrate affinity etc.), predicting <u>physiological substrates</u> on sequence alone, with either chromogenic techniques (oligopeptide as substrate, or chromogenic substrates) and *in silico* (as shown in this thesis), can be interpreted as a weak approach. The most straightforward way is an *in vitro* assay of the proposed physiological substrate, and a subsequent analytical confirmation of the simultaneous presence of the enzyme and substrate in the same tissue (for instance GFP-tagging, Korkmaz *et al*, 2001) or immunohistochemical labelling (Clements J, personal communication). The latter requires strong and specific antibodies. From the quantities and qualities of rhK4 purified here, the generation of antibodies is not advisable to date. However the predicted antigenic segment of hK4 (section 3.1.2.1) shows computational data that indicates better rigidity in solution and higher antigenicity compared to antipeptide hK4 antibodies currently in use in this lab and therefore a better antigen for accurate immunological detection of hK4.

6.2   Concluding remarks

Recombinant kallikrein 4 was expressed in a baculovirus system at a concentration estimated to 0.93mg/L-4.8mg/L. The ideal incubation time at a multiplicity of infection of 10 was 72 hrs post-infection. The baculovirus system provided a glycosylated recombinant in close similarity to the COS cell-expressed hK4, in terms of mass, by comparing the COS cell-variant and the baculovirus-expressed variant (~2kDa difference). The silver stained membranes showed that the recombinant target protein was purified to a range of 36-90% purity using only two steps; isoelectric dialysis anticipated using computer modelling, and affinity chromatography.

The use of computational methods showed promising results in this thesis. Conformational analysis of the C-terminal V5/His construct with hK4 was consistent with the resulting efficacy of the His-tag antibody ,which was used instead of the V5 antibody in the latter half of the purification studies. Antigenic analysis further designated the potential of a new peptide segment of hK4 to be applied for generating new and possibly more specific hK4 antibodies. Molecular modelling and ES- modelling of the target protein in complex with its presumed substrate gave some basic knowledge on how the subsite mechanism works in this protein in analogous fashion to other serine protease family members. The dimer modelling and the scoring of electrostatic

interaction between the enzyme and the substrate gives a basis to make a list of selective substrates for in depth analysis using classical mechanics (MD).

Molecular dynamics makes it possible to determine the chemical and geometrical fit between a hypothetical substrate and an enzyme. Factors as structural flexibility and chemical complementarity of the hypothetical substrate are clarified, so the hypothetical quality of the induced fit with the enzyme can be described by RMSD, energetics and visual inspections. Additionally MD simulations aid to estimate the impact of the solvent on the chosen substrate, and the quality of the protein models. Combining the assessment of the quality of a protein model, with the assessment of the induced fit between the substrate and the enzyme, a substantial amount of knowledge was gained on kallikrein 4 and the selected members of the metzincins. Because nature satisfies the functions of enzymes in the cell by complementarity and cells-specific expression and localization, predicting the complementarity is crucial for advances in biochemical sciences, specifically for design of clinically applicable inhibitors.

The knowledge gained of rhK4 mainly from an *in silico* aspect provided a sufficient basis, in concert with past knowledge on serine protease inhibitors, to design and test trypsin-like inhibitors *a priori*. The flexible ligand docking experiments performed here show that an inhibition can be predicted with 80-90% accuracy, given that the model was built on a sequence identity of 41%, (Klebe and Shafferhans, 2001). As a result three variants of the autoactivation region-mimickers are proposed for future experiments with kallikrein 4.

Furthermore ES-modelling was challenged by working with metzincins. This new methodology gave preliminary insight into the unknown composition of the primary specificity pocket of ADAM-9 and ADAM-10 (S1'). Additionally, a new S2/S3-pocket was identified, which in these computational experiments played role in stabilizing the substrate on the P-side. The S2/S3 pockets were quite homologous between the two ADAMs but differed more from MMP-3 and can prove useful for future drug-design approaches directed towards metzincins. The reported set of H-bonds on the P'-side and the computationally assembled H-bonds on the P-side allowed the catalytic glutamic acid to move in coordination with the Zn-ion during the simulations. Given the dynamic behaviour of the catalytic glutamate, a new catalytic mechanism based on the nucleophilic attack from the glutamate to the substrate, creating a covalent E-S complex is suggested. The role of water is thought to be involved without any direct polarization contact with the Zn-ion, but as a simple acid-base hydrolysis of the resulting broken scissile bond, by accessing it from the surface.

Combining computational sciences and empirical science is the core of this thesis, where the emphasis on using computational methods to predict and anticipate properties of proteins in biochemistry is evident. Unfortunately, the current commonly used methodologies are only limited to sequence analysis techniques which do not have the precision supplied by stochastic statistics and classical mechanics methods, as in theoretical energy estimations and MD. This thesis provides therefore new approaches that aim to further enlighten the upcoming generation of structural bioinformaticists.

APPENDIX

Atomic coordinates: The simulated dimers of ADAM-9, ADAM-10 and MMP-3 in complex with their substrates have been deposited on the protein databank (www.pdb.org) as 1M1V, 1M1I and 1M1W entries, respectively.

References

Amour A., Knight CG., Webster A., Slocombe PM., Stephens PE., Knauper V., Docherty AJ. and Murphy G., (2000). The in vitro activity of ADAM-10 is inhibited by TIMP-1 and TIMP-3. FEBS Letters 473:275-279.

Amour A., Slocombe PM., Webster A., Butler M. Knight CG., Smith BJ., Stephens PE., Shelley C., Hutton M., Knauper V., Docherty AJ. and Murphy G., (1998). TNF-alpha converting enzyme (TACE) is inhibited by TIMP-3. FEBS Letters 435:39-44.

Angermann A., Rahn HP., Hektor T., Fertig G. and Kemme M. (1992). Purification and characterization of human salivary-gland prokallikrein from recombinant baculovirus-infected insect cells. Eu. J. Biochem., 206, 225-233.

Anisowicz A., Sotiropoulou G., Stenman G., Mok SC. and Sager R. (1996). A novel protease homolog differentially expressed in breast and ovarian cancer. Molecular Medicine, 96, 1076-1551.B

Appel RD., Bairoch A. and Hochstrasser DF. (1994). A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends Biochem. Sci. 19:258-260.

Baker AR. and Shine J. (1985). Human kidney kallikrein: cDNA cloning and sequence analysis. DNA 4, 445-450.

Basbaum CB. and Werb Z. (1996). Focalized proteolysis: spatial and temporal regulation of extracellular matrix degradation at the cell surface. Curr. Opin. Cell. Biol. 8:731-738.

Beduschi MC. and Oesterling JE., 1998. Percent free prostate-specific antigen: the next frontier in prostate-specific antigen testing. Urology, 51, 98-109.

Berendsen HJC., Postma JPM., DiNola A. and Haak JR. (1984) Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684-3690.

Berg T., Bradshaw RA., Carretego OA., Chao J, Chao L., Clements JA., Fahnestock M., Fritz H., Gauthier F., MacDonald RJ., Margolius HS., Morris BJ., Richards RI. and Scicli AG. (1992). A common nomenclature for members of the tissue (glandular) kallikrien gene families. Recent Progress on Kinins, Birkhauser Verlag Basel, pp. 19-21.

Bhat-Nakshatri P., Newton TR., Goulet R. and Nakshatri H. (1998). NF-κB activation and interleukin 6 production in fibroblasts by estrogen receptor-negative breast cancer cell-derived interleukin 1α. Medical Sciences, 95, 6971–6976.

Bode W, & Huber R. (1992). Natural protein proteinase inhibitors and their interactions with proteinases. Eu. J. Biochem., 204, 433-451.

Borkakoti N., Winkler FK., Williams DH., D'Arcy A., Broadhurst MJ., Brown PA., Johnson WH. and Murray EJ. (1994). Structure of the catalytic domain of human fibroblast collagenase complexed with an inhibitor. Nat. Struct. Biol. 1:106-110.

Bradford MM. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248-54.

Brattsand M. and Egelrud T. (1999). Purification, molecular cloning, and expression of a human stratum corneum trypsin-like serine protease with possible function in desquamation. J. Biol. Chem. 274, 30033-30040.

Bridon DP., Dowell BL. (1995). Structural comparison of prostate-specific antigen and human glandular kallikrein using molecular modelling. Urology, 45, 801-806.C

Brinkworth RI., Horne J. and Kobe B. (2002). A computational analysis of substrate binding strength by phosphorylase kinase and protein kinase A. J Mol Recognit.15:104-11.

Browner MF., Smith W. and Castelhano AL. (1995). Matrilysin-inhibitor complexes: common themes among metalloproteases. Biochemistry 34:6602-6610.

Chen L., Rydel TJ., Dunaway CM., Pikul S, Dunham KM., Gu F. and Barnett BL. (1999). Crystal Structure of the Stromelysin Catalytic Domain at 2.0 A Resolution: Inhibitor-Induced Conformational Changes. J. Mol. Biol. 293:545-557.

Chothia C, Wodak S, Janin J. (1976). Role of subunit interfaces in the allosteric mechanism of hemoglobin. Proc. Natl. Acad. Sci. U S A 73:3793-3797.

Clements J. (1989). The glandular kallikrein family of enzymes: Tissue-specific expression and hormonal regulation. Endocrine Reviews, 10, 393-418

Clements J. (1997). The molecular biology of the kallikreins and their roles in inflammation. Biological Chemistry Review.

Clements J., Hooper J., Dong Y., Harvey T. (2001). The expanded human kallikrein (KLK) gene family: genomic organization, tissue-specific expression and potential functions. Biol Chem. 382:5-14.

Clements J. and Mukhtar A. (1997). Tissue Kallikrein and the Bradykinin receptors are expressed in endometrial and prostate cancers. Immunopharmacology 36: 217-220.

Cohen P., Peehl DM and Lamson G. (1992). Prostate-specific antigen (PSA) is an insulin-like growth factor binding protein-3 proteases found in seminal plasma. J. Clin. Endoc. Met. 76, 1046-1053.

Coombs GS., Bergstrom RC., Pellequer JL., Baker SI., Navre M., Smith MM., Tainer JA., Madison EL. and Corey DR. (1998a). Substrate specificity of prostate-specific antigen (PSA). Chemistry & Biology, 5, 475-488.

Coombs GS., Rao MS., Olson AJ., Dawson PE. and Madison EL. (1998b). Revisiting catalysis by chymotrypsin family serine proteases using peptide substrates and inhibitors with unnatural main chains. J Biol Chem. 20:24074-24709.

Coombs G., Rao MS., Olson AJ., Dawson PE. and Madison E. (1999). Revisiting catalysis by chymotrypsin family serine proteases using peptide substrates and inhibitors with unnatural main chains. The Journal of Biological Biochemistry 274, 24074-24079.

Corey DR., Willett WS., Coombs GS. and Craik CS. (1995). Trypsin specificity increased through substrate-assisted catalysis. Biochemistry 34:11521-7.

Darden TA., York DM. and Pedersen LG. (1993). Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. J. Chem. Phys. 98:10089–92

Daura X., Gademann K., Schafer H., Jaun B., Seebach D. and van Gunsteren WF. (2001). The beta-peptide hairpin in solution: conformational study of a beta-hexapeptide in methanol by NMR spectroscopy and MD simulation. J Am Chem Soc 123:2393-404

Daura X., Jaun B., Seebach D., van Gunsteren WF. and Mark AE. (1998) Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 280:925-32.

Davis CA., Riddell DC., Higgins MJ., Holden JJ. and White BN. (1985). A gene family in Drosophila melanogaster coding for trypsin-like enzymes. Nucleic Acids Research 13, 6605-6619.

Dayhoff MO. (1972). National Biomedical Research Foundation. 5, Supplemental, 79-81.

Deperthes D., Marceau F., Frenette G., Lazure C., Tremblay RR. and Dube' JY. (1997). Human kallikrein hK2 has low kininogenase activity while prostate-specific antigen has none. Biochimica et Biophysica Acta, 1343, 102-106.

Deperthes, D., Chapdelaine, P., Tremblay, R.R., Brunet, C., Berton, J., Hebert, J., Lazure, C. and Dube, J.Y. (1995). Isolation of prostatic kallikrein hK2, also known as hGK-1, in human seminal plasma. Biochim. Biophys. Acta 1245:311-316.

Dhanaraj V., Ye QZ., Johnson LL., Hupe DJ., Ortwine DF., Dunbar JB., Rubin JR., Pavlovsky A., Humblet C. and Blundell TL (1996a). X-ray structure of a hydroxamate inhibitor complex of stromelysin catalytic domain and its comparison with members of the zinc metalloproteinase superfamily. Structure 4:375-386.

Dhanaraj V., Ye QZ., Johnson LL., Hupe DJ., Ortwine DF., Dunbar JB. Jr., Rubin JR., Pavlovsky A., Humblet C. and Blundell TL. (1996b). Designing inhibitors of the metalloproteinase superfamily: comparative analysis of representative structures. Drug. Des. Discov. 13:3-14.

Diamandis EP., Yousef GM., Luo LY., Maglakara A. and Obiezu CV. (2000). The new human kallikrein gene family: Implications in carcinogenesis. TEM, 11, 54-59.

Diamandis EP., Yousef GM., Soosaipillai AR., Grass L., Porter A., Little S. and Sotiropoulou G., (2000a). Immunoflourometric assay of human kallikrein 6 (Zyme/Protease M/Neurosin) and preliminary clinical applications. Clinical Biochemistry, 33, 369-375.

Dill, KA., 1997. Additivity principles in biochemistry. J Biol Chem 10:701-704

Dlamini K., Raidoo D. and Bhoola K. (1999). Visualisation of tissue kallikrein and kinin receptors in oesophageal carcinoma. Immunopharmacology 43, 303-310.

Dong Y, Kaushal A, Brattsand M, Nicklin J, Clements JA. (2003). Differential Splicing of KLK5 and KLK7 in Epithelial Ovarian Cancer Produces Novel Variants with Potential as Cancer Biomarkers. Clinical Cancer Research 9:1710-20.

Egelrud T., Regnier M., Sondell B., Shroot B. and Schmidt R. (1993). Expression of stratum corneum chymotryptic enzyme in reconstructed human epidermis and its suppression by retinoic acid. Acta Dermatologica Venereologica, 73, 181-184.

Ekholm E., Bratsand M. and Egelrud T. (2000). Stratum Corneum Tryptic Enzyme in Normal Epidermis: a missing link in the desquamation process? The Society of investigative Dermatology 114, 56-63.

Ewal P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann. Phys.8:211-217.

Faraldo-Gomez JD., Smith GR. and Sansom MS. (2002). Setting up and optimization of membrane protein simulations. Eur Biophys J. 31:217-27.

Fasciglione GF., Marini S., D'Alessio S., Politi V. and Coletta M. (2000). pH- and temperature-dependence of functional modulation in metalloproteinases. A comparison between neutrophil collagenase and gelatinases A and B. Biophys. J. 79:2138-2149.

Fersht A. (1977). Enzymes structure and mechanism. pp.27. W.H. Freeman and Company Ltd.

Frenette G., Tremblay RR., Dube' JY. (1998). Simple purification procedure for human prostatic kallikrein hK2 in its active form. Journal of Chromatography, 713, 297-300.

Gall AL., Ruff M., Kannan R., Cuniasse P., Yiotakis A., Dive V., Rio MC., Basset P. and Moras D. (2001). Crystal structure of the stromelysin-3 (MMP-11) catalytic domain complexed with a phosphinic inhibitor mimicking the transition-state. J. Mol. Biol. 307:577-86.

Gauthier ER., Chapdelaine P., Tremblay RR. and Dube JY. (1993). Characterization of rhesus monkey prostate specific antigen cDNA. Biochimica et Biophysica Acta 1174, 207-210.

Glenn MP., Pattenden LK., Reid RC., Tyssen DP., Tyndall JD., Birch CJ. and Fairlie DP. (2001). Beta-strand mimicking macrocyclic amino acids: templates for protease inhibitors with antiviral activity. J Med Chem 45:371-381.

Goldberg DE. (1989). Genetic algorithms in search, optimisation and machine learning. Reading, MA, Addison-Wesley.

Gomis-Rüth FX., Maskos K., Betz M., Bergner A., Huber R., Suzuki K., Yoshida N., Nagase H., Brew K., Bourenkov GP., Bartunik H. and Bode W. (1997). Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. Nature 389:77.

Gough J., Karplus K., Hughey R. and Chothia C. (2001). Assignment of homology to Genome Sequences using a Library of Hidden Markov Models that Represent all Proteins of Known Structure. J. Mol. Biol. 313:903-919.

Goyal J., Smith KM., Cowan JM., Wazer DE., Lee SW. and Band V. (1998). The role for NES1 serine protease as a novel tumour suppressor. Cancer Research. 58, 4782–4786

Grams F., Reinemer P., Powers JC., Kleine T., Pieper M., Tschesche H., Huber R. and Bode W. (1995). X-ray structures of human neutrophil collagenase complexed with peptide hydroxamate and peptide thiol inhibitors. Implications for substrate binding and rational drug design. Eur. J. Biochem. 228:830-841

Guex N., and Peitsch MC. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18: 2714-2723.

Hanson L., Stromquist M., Backman A., Wallbrandt P., Carlstein A. and Egelrud T. (1994). Cloning, expression and characterization of stratum corneum chymotryptic enzyme. Journal of Biological Chemistry, 269, 19420-19426.

Hart TN., Read RJ. (1992). A multiple-start Monte Carlo docking method. Proteins 13, 206-222.

Harvey T., Hooper J., Myers S., Stephenson SA., Ashworth LK., Clements JA. (2000). Tissue specific expression patterns and fine mapping of the human kallikrein (KLK) locus on proximal 19q13.4. J. Biol. Chem. 320, 2365-5689.

Heath EI. and Grochow LB. (2000). Clinical potential of matrix metalloprotease inhibitors in cancer therapy. Drugs 59:1043-1055.

Hermann A., Arnhold M., Kresse H., Neth P., Fink E. (1999). Expression of components of the kallikrein-kinin system in human cell lines. Immunopharmacology 45, 135-139.

Hess B., Bekker H., Berendsen HJC., Fraaje JGEM. (1997). LINCS: A linear contraint solver for molecular simulations. J. Comp. Chem. 18:1463-1472

Heymann B, and Grubmuller H. (2001). Molecular dynamics force probe simulations of antibody/antigen unbinding: entropic control and nonadditivity of unbinding forces. Biophys. J. 81:1295-1313.

Hinds DA. and Levitt M. (1992). A lattice model for protein structure prediction at low resolution. Proc. Natl. Acad. Sci. USA, 89, 2536-2540.

Hooper DJ., Bui LT., Rae FK., Harvey TJ., Myers SA., Ashworth LK. and Clements J. (2000). Identification and characterization of KLK14, a novel kallikrein serine protease gene located on human chromosome 19q13.4 and expressed in prostate and skeletal muscle. Genomics 73, in press.

Hopp TP. and Woods KR. (1981). Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA, 78, 3824-3828

Hsieh MC. and Cooperman BS. (2000). The preparation and catalytic properties of recombinant human prostate-specific antigen (rPSA). Biochimica et Biophysica Acta, 1481, 75-87.

Hsieh MC. and Cooperman BS (2002). Inhibition of prostate-specific antigen (PSA) by alpha(1)-antichymotrypsin: salt-dependent activation mediated by a conformational change. Biochemistry. 41:2990-7.

Innis CA., Shi J. and Blundell TL. (2000). Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. Prot. Eng. 13:839-847.

Isaacs JT., Denmeade SR., Wei L., Lovgren J., Malm J., Lilja J. (1997). Specific and efficient substrates for assaying the proteolytic activity of prostate specific antigen. Cancer Research, 57, 4924-4930.

Izumi Y., Hirata M., Hasuwa H., Iwamoto R., Umata T., Miyado K., Tamai Y., Kurisaki T., Sehara-Fujisawa A. and Ohno S. Mekada E. (1998). A metalloprotease-disintegrin, MDC9/meltrin-gamma/ADAM9 and PKCdelta are involved in TPA-induced ectodomain shedding of membrane-anchored heparin-binding EGF-like growth factor. EMBO Journal 17:7260-7272.

Johnson LL., Pavlovsky AG., Johnson AR., Janowicz JA., Man CF., Ortwine DF., Purchase CF. 2nd., White AD., and Hupe DJ. (2000). A rationalization of the acidic pH dependence for stromelysin-1 (Matrix metalloproteinase-3) catalysis and inhibition. J. Biol. Chem. 275:11026-11033.

Karplus K., Barrett C., and Hughey R. (1998). Hidden Markov Models for Detecting Remote Protein Homologies. Bioinformatics 14:846-856

Katchalski-Katzir E., Shariv I., Eisenstein M., Friesem AA., Aflaflo C. and Vakser IA. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. PNAS USA 89:2195-2199.

Kishi T., Kato M., Shimizu T., Kato K., Matsumoto K., Yoshida S., Shiosaka S., Hakoshima T. (1997). Crystallization and preliminary X-ray analysis of neuropsin, a serine protease expressed in the limbic system of mouse brain. J. Struct. Biol., 118, 248-251.

Kishi T., Kato M., Shimizu T., Kato K., Matsumoto K., Yoshida S., Shiosaka S., Hakoshima T. (1999). Crystal structure of neuropsin, a hippocamal protease involved in kindling epileptogenesis. J. Biol. Chem., 274, 4220-4224.

Kiyama R., Tamura Y., Watanabe F., Tsuzuki H., Ohtani M. and Yodo M. (1999). Homology modeling of gelatinase catalytic domains and docking simulations of novel sulfonamide inhibitors. J. Med. Chem. 42:1723-1738.

Koivunen E., Huhtala ML. and Stenman UH. (1989). Human ovarian tumour-associated trypsin. Its purification and characterization from mucinous cyst fluid and identification as an activator of pro-urokinase. . J. Biol. Chem. 264, 14095-14099.

Korkmaz KS., Korkmaz CG., Pretlow TG. and Saatcioglu F. (2001) Distinctly different gene structure of KLK4/KLK-L1/prostase/ARM1 compared with other members of the kallikrein family: intracellular localization, alternative cDNA forms, and Regulation by multiple hormones. DNA Cell Biol. 20:435-45.

Koshland DE. Jr, Carraway KW., Dafforn GA., Gass JD. and Storm DR. (1972). The importance of orientation factors in enzymatic reactions. Cold Spring Harb. Symp. Quant. Biol. 36:13-20.

Kumar A., Mikolajczyk SD., Goel AS., Millar LS. and Saedi MS. (1997). Expression of pro form of prostate-specific antigen by mammalian cells and its conversion to mature, active form by human kallikrein 2. Cancer Research 57, 3111-3114.

Kurkela R., Herrala A., Henttu P., Nal H. and Vihko P. (1995). Expression of active and secreted human prostate-specific antigen by recombinant baculovirus-infected insect cells on a pilot-scale. Biotechnology, 13, 1230-1233.

Leach A. (2001). Molecular Modelling: Principles and applications. 2nd Ed. Pearson Education Limited. pp. 305.

Li Li., Binz T., Niemann H. and Singh BR. (2000). Probing the mechanistic role of glutamate residue in the zinc-binding motif of type A botulinum neurotoxin light chain. Biochemistry 39:2399-2405.

Light A. and Janska H (1989). Enterokinase (enteropeptidase): comparative aspects. Trends in Biochemical Sciences. 14, 110-113.

Lilja H. (1985). A kallikrein-like protease in prostatic fluid cleaves the predominant seminal vesicle protein. J. Clin. Investig., 76, 1899-1903.

Lilja H. (1987). Seminal vesicle-secreted proteins and their reactions during gelation and liquefaction of human semen. J. Clin. Investig., 80, 281-285.

Lindahl E., Hess B. and van der Spoel D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. J. Mol. Mod. 7:306-317.

Little SP., Dixon EP., Norris F., Buckley W., Becker GW., Johnson M., Dobbins JR., Wyrick T., Miller JR., MacKellar W., Hepburn D., Corvalan J., McClure D., Liu X., Stephenson D., Clements J., and Johnstone EM. (1997). Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. J. Biol. Chem. 272, 25135–25142.

Liu XF., Essand M., Vasmatzis G., Lee B. and Pastan I. (1999). Identification of three new alternate human kallikrein 2 transcripts: evidence of long transcript and alternative splicing. Bioch. Biophys. Res. Comm., 264, 833-839.

Llinas-Brunet M., Bailey M., Fazal G., Ghiro E., Gorys V., Goulet S., Halmos T., Maurice R., Poirier M., Poupart MA., Rancourt J., Thibeault D., Wernic D. and Lamarre D. (2000). Highly potent and selective peptide-based inhibitors of the hepatitis C virus serine protease: towards smaller inhibitors. Bioorg Med Chem Lett, 10:2267-2270.

Lovejoy B., Cleasby A., Hassell AM., Longley K., Luther MA., Weigl D., McGeehan G., McElroy AB., Drewry D., Lambert MH. and Jordan SR. (1994). Structure of the catalytic domain of fibroblast collagenase complexed with an inhibitor. Science 263:375-377.

Lovgren J., Rajakoski K., Karp M., Lundwall A. and Lilja H. (1997). Activation of the zymogen form of prostate-specific antigen by human glandular kallikrein 2. Bioch. Biophys. Res. Comm., 238, 549-555.

Lovgren J., Tian S., Lundwall A., Karp M. and Lilja H. (1999). Production and activation of recombinant hK2 with propeptide mutations resulting in high expression levels. Eu. J. Biochem., 266, 1050-1055.

Lu SM., Lu W., Qasim MA., Anderson S., Apostol I., Ardelt W., Bigler T., Chiang YW., Cook J., James MN., Kato I., Kelly C., Kohr W., Komiyama T., Lin TY., Ogawa M., Otlewski J., Park SJ., Qasim S., Ranjbar M., Tashiro M., Warne N., Whatley H., Wieczorek A., Wieczorek M., Wilusz T., Wynn R., Zhang W. and Laskowski M. Jr. (2001). Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. PNAS USA 98:1410-5

Lundwall A. (1989). Characterization of the gene for prostate-specific antigen, a human glandular kallikrein. Bioch. Biophys. Res. Comm., 161, 1151-1159.

Lundwall A. and Lilja H. (1987). Molecular cloning of human prostate specific antigen cDNA. FEBS Lett. 214, 317-322.

Luo L., Herbrick JA., Scherer S., Beatty B., Squire J. and Diamandis E. (1998). Structural characterization and mapping of the normal epithelial cell-specific 1 gene. Bioch. Biophys. Res. Comm., 247, 580-586.

MacKerell AD. Jr., Banavali N. and Foloppe N. (2001). Development and current status of the CHARMM force field for nucleic acids. Biopolymers 56:257-65.

Maeda H., Wu J., Okamoto T., Maruo K. and Akaike T. (1999). Kallikrein-kinin in infection and cancer. Immunopharmacology 43, 115-128.

Maglakara A., Cheung CC., Asa SL. and Diamandis EP. (2000). Expression of prostate-specific antigen and human glandular kallikrein 2 in the thyroid gland. Clinica Chimica Acta, 300, 171-180.

Manca F., Habeshaw JA., Dalgleish AG., Fenoglio D., Li Pira G. and Sercarz EE. (1993). Role of flanking variable sequences in antigenicity of consensus regions of HIV gp120 for recognition by specific human T helper clones. Eur J Immunol 23:269-74.

Manzetti S. and Barnard R. (2001). An analysis of reliability of in silico automated modelling using the GROMOS96 force field, in prediction of kinetic parameters for trypsin and its substrates. Combined Conference Abstracts, COMBIO2001, Pos-2-018.

Manzetti S. (2002). Taking the complexity out of protein sequence analysis. Drug Discov Today 7:172-175.

Maroux S., Baratti J. and Desnelle P. (1971). Purification and specificity of porcine enterokinase. J. Biol. Chem., 246, 5031-5039.

Martí-Renom MA., Fiser A. and Šali A. (2002). Comparative modelling, methods and applications. Predicting the structure of difficult proteins, Bologna Winter School 2002. University of Bologna.

Maskos K., Fernandez-Catalan C., Huber R., Bourenkov GP., Bartunik H., Ellestad GA., Reddy P., Wolfson MF., Rauch CT., Castner BJ., Davis R., Clarke HR., Petersen M., Fitzner JN.,

Cerretti DP., March CJ., Paxton RJ., Black RA. and Bode W. (1998). Crystal Structure of the Catalytic Domain of Human Tumor Necrosis Factor-Alpha- Converting Enzyme. Proc. Nat. Acad. Sci. 95:3408-3412.

Mildvan AS. (1970). The Enzymes. Paul D. Boyer editor. Academic Press 2:246.

Millichip MI., Dallas DJ., Wu E., Dale S. and McKie N. (1998). The metallo-disintegrin ADAM10 (MADM) from bovine kidney has type IV collagenase activity in vitro. Biochem. Biophys. Res. Commun. 245:594-598.

Mitsui S., Yamada T., Okui A., Kominami K., Uemura H. and Yamaguchi N. (2000). A novel isoform of a kallikrein-like protease, TLSP/Hippostasin, (PRSS20) is expressed in the human brain and prostate. Bioch. Biophys. Res. Comm., 272, 205-211.

Mittl PR.., Di Marco S., Fendrich G., Pohlig G., Heim J., Sommerhoff C., Fritz H., Priestle JP., Grutter MG. (1997). A new structural class of serine protease inhibitors revealed by the structure of the hirustasin-kallikrein complex. Structure 5: 253.

Miyamoto S. and Kollman PA. (1992). SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. J. Comp. Chem. 13:952-962.

Morse PM. (1929). Diatomic molecules according to wave mechanics vibrational levels. Phys. Rev. 34:57-64.

Murphy, G. and Docherty, JP. (1992) The matrix metalloproteases and their inhibitors. Am. J. Res. Cell. Mol. Biol. 7:120–125.

Nagase H. (1996). Matrix metalloproteinase. In: Hooper N, ed. Zinc metalloproteinases in health and disease. pp. 153-204 London: Taylor and Francis.

Nelson P.S., Gan L., Ferguson C., Moss P., Gelinas R., Hood L., Wang K. and (1999). Molecular cloning and characterisation of prostase, an androgen-regulated serine protease with prostate-restricted expression. Proceedings of the National Academy of Sciences, 96, 3114-3119.

Newton I. (1687). Philosophiae Naturalis Principia Mathematica. London, Royal Society. Trans. by A. Motte and edited by F. Cajori. University of California Press: Berkeley, 1934.

Osterberg F., Morris GM., Sanner MF., Olson AJ. and Goodsell DS. (2002). Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 46:34-40.

Parkin G. (2001). Synthetic analogs of zinc enzymes. Met. Ions Biol. Syst. 38:411-460.

Partin AW., Catalona WJ., Finlay JA., Darte C., Tindall DJ., Young CYF., Klee GG., Chan DW., Rittenhouse HG., Wolfert RL,. Woodrum DL. (1999). Use of human glandular kallikrein 2 for the detection of prostate cancer: preliminary analysis. Urology 54, 839-845.

Perona JJ. and Craik C. (1995). Structural basis of substrate specificity in the serine proteases. Protein Science, 4, 337-360.

Perona JJ. and Craik C. (1997). Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. J. Biol. Chem., 272, 29987-29990.

Peter A., Lilja H. and Malm J. (1998). Semenogelin I and II, the major gel-forming proteins in human semen are substrate of transglutaminase. Eu. J. Biochem., 252, 216-221.

Podlich D. and Kemme M. (1999). Different susceptibility of human tissue prokallikrein to cleavage by neutrophil proteinases. Immunopharmacology, 45, 95-101.

Primakoff P. and Myles DG. (2000). The ADAM gene family: surface proteins with adhesion and protease activity. Trends in Genetics 16: 83-87

Raidoo D. Dlamini Z. and Bhoola K. (1999). Visualisation of tissue kallikrein and kinin receptors in oesophageal carcinoma. Immunopharmacology 43, 303-310.

Raidoo DM., Sawant S., Mahabeerm R. and Bhoola KD. (1999b). Kinin receptors are expressed in human astrocytic tumour cells. Immunopharmacology 43, 255-263.

Ramakrishnan C. and Ramachandran GN. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. Biophys J. 5:909-33.

Reifenberger J., Reifenberger G., Liu L., James D., Wechsler W. and Collins VP. (1994). Molecular genetic analysis of oligodendroglial tumours shows preferential allelic deletions on 19q and 1p. American Journal of Pathology, 145, 1175-1189.

Reinemer P., Grams F., Huber R., Kleine T., Schnierer S., Piper M., Tschesche H. and Bode W. (1994). Structural implications for the role of the N terminus in the 'superactivation' of collagenases. A crystallographic study. FEBS Lett, 338:227-233.

Riegman PH., Vlietstra RJ., Klaassen P., van der Korput JA., Geurts van Kessel A., Romijn JC. and Trapman J. (1989). The prostate-specific antigen gene and the human glandular kallikrein-1 gene are tandemly located on chromosome 19. FEBS Lett. 247, 123-126.

Riegman PH., Vlietstra RJ., Suurmeijer .L, Cleutjens CB. and Trapman J. (1992). Characterization of the human kallikrein locus. Genomics, 14, 6-11.

Robert M., Gibbs BF., Jacobson E. and Gagnon C. (1997). Characterization of prostate specific antigen proteolytic activity on its major physiologic substrate, the sperm motility inhibitor precursor / semenogelin I. Biochemistry 36, 3811-3819.

Roghani M., Becherer JD., Moss ML., Atherton RE., Erdjument-Bromage H., Arribas J., Blackburn RK., Weskamp G., Tempst P. and Blobel CP. (1999). Metalloprotease-disintegrin MDC9: intracellular maturation and catalytic activity. J. Biol. Chem. 274:3531-3540.

Rosendahl, MS., Ko SC., Long DL., Brewer MT., Rosenzweig B., Hedl E., Anderson L., Pyle SM., Moreland J., Meyers MA., Kohno T., Lyons D. and Lichenstein HS. (1997). Identification and characterization of a pro-tumor necrosis factor-alpha-processing enzyme from the ADAM family of zinc metalloproteases. J. Biol. Chem. 272: 24588-24593.

Saedi MS., Zhu Z., Marker K., Liu RS., Carpenter PM., Rittenhouse H. and Mikolajczyk SD. (2001). Human kallikrein 2 (hK2), but not prostate-specific antigen (PSA), rapidly complexes with protease inhibitor 6 (PI-6) released from prostate carcinoma cells. Int J Cancer. 94:558-63.

Sagui C. and Darden T. (1999). Molecular dynamics simulations of biomolecules: Long-Range Electrostatic Effects. Annu. Rev. Biophys. Biomol. Struct. 28:155-179.

Sali A. (1995). Comparative protein modeling by satisfaction of spatial restraints. Mol Med Today 1:270-7.

Schafferhans A. and Klebe G. (2001). Docking ligands onto binding site representations derived from proteins built by homology modelling. J Mol Biol, 307:407-427

Schechter I. and Berger A. (1968). On the size of the active sites in proteases. I. Papain. Biochem. Biophys. Res. Commun, 27:157-162.

Schedlich LJ., Bennetts BH. and Morris BJ. (1987). Primary structure of a human glandular kallikrein gene. DNA, 6, 429-437.

Schoofs L., Clynen E., Salzet M. (2002). Trypsin and chymotrypsin inhibitors in insects and gut leeches. Curr Pharm Des. 8:483-91. Review.

Schreiber H., Steinhauser O., (1992). Taming cut-off induced artifacts in molecular dynamics studies of solvated polypeptides. The reaction field method. J Mol Biol. 228:909-23.

Scott WRP., Hulnenberger HP., Tironi IG., Mark AE., Billeter SR., Fennen J., Torda AE., Huber T., Kru1ger P., and van Gunsteren F. (1999). The GROMOS Biomolecular Simulation Program Package. J. Phys. Chem. 103:3596-3607.

Seltzer JL., Weingarten H., Akers KT., Eschbach ML., Grant GA., Eisen AZ. (1989). Cleavage specificity of type IV collagenase (gelatinase) from human skin. Use of synthetic peptides as model substrates. J. Biol. Chem. 264:19583-19586.

Shimizu C., Yoshida S., Shibata M., Kato K., Momota Y., Matsumoto K., Shiosaka T., Midorikawa R., Kamachi T., Kawabe A. and Shiosaka S. (1998). Characterization of recombinant and brain neuropsin a plasticity-related serine protease. . J. Biol. Chem, 273, 11189-11196.

Silverberg M. and Kaplan AP. (1982). Enzymatic activities of activated and zymogen forms of human Hageman factor (factor XII). Blood, 60:64-70.

Simmer JP., Fukae M., Tanabe T., Yamakoshi Y., Uchida T., Xue J., Margolis HC., Shimizu M., DeHart BC., Hu CC., Bartlett JD. (1998). Purification, characterization, and cloning of enamel matrix serine proteinase 1. J. Dent. Res. 77, 377-386.

Skytt A., Stromqvist M. and Egelrud T (1995). Primary substrate specificity of recombinant human stratum corneum chymotryptic enzyme. Biochem. Biophys. Res. Comm. 211, 586-589.

Smith MM., Shi L. and Navre M. (1995). Rapid identification of highly active and selective substrates for stromelysin and matrilysin using bacteriophage peptide display libraries. J Biol. Chem. 270:6440-6449.

Springman EB., Angleton EL., Birkedal-Hansen H. and Van Wart HE. (1990). Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a "cysteine switch" mechanism for activation. Proc. Natl. Acad. Sci. U S A 87:364-368.

Stephenson SA., Verity K., Ashworth L. and Clements J. (1999). Localization of a new prostate-specific antigen-related serine protease gene, KLK4, Is evidence for an expanded human kallikrein gene family cluster on chromosome 19q13.3-13.4. . J. Biol. Chem, 274, 23210-23214.

Stewart JJ. (1990). MOPAC: a semiempirical molecular orbital program. J Comp Aid Mol Des 4:1-105

Sticht H., Bayer P., Willbold D., Dames S., Hilbich C., Beyreuther K., Frank RW. and Rosch P. (1995). Structure of amyloid A4-(1-40)-peptide of Alzheimer's disease. Eur J Biochem 233: 293

Stöcker W., Grams F., Baumann U., Reinemer P., Gomis-Rüth FX., McKay DB. and Bode W. (1995). The metzincins--topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. Protein Sci, 4: 823-840.

Stone SS. and Hess WR. (1965). Separation of virus and soluble noninfectious antigens in African swine fever virus by isoelectric precipitation. Virology. 26:622-9.

Takayama TK., Fujikawa K. and Davie EW. (1997). Characterization of the precursor of prostate specific antigen – activation by trypsin and by human glandular kallikrein. . J. Biol. Chem, 272, 21582-21588.

Takayama TK., McMullen BA., Nelson PS., Matsumura M and, Fujikawa K. (2001). Characterization of hK4 (prostase), a prostate-specific serine protease: activation of the precursor of prostate specific antigen (pro-PSA) and single-chain urokinase-type plasminogen activator and degradation of prostatic acid phosphatase. Biochemistry 40:15341-8.

Thompson JD., Higgins DG. and Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

Tomita N., Izumoto Y., Horii A., Doi S., Yokouchi H., Ogawa M., Mori T. and Matsubara K. (1989). Molecular cloning and nucleotide sequence of human pancreatic prechymotrypsinogen cDNA. Biochem. Biophys. Res. Comm. 158, 569-575.

Tschope C., Koch M., Spillman F., Heringer-Walther S., Mochmann HC., Stauss H., Bader M., Unger T., Schultneiss HP. and Walther T. (1999). Upregulation of the cardiac bradykinin B2 receptors after myocardial infarction. Immunopharmacology 44, 111-117.

Tyndall JD. and Fairlie DP. (2001). Macrocycles mimic the extended peptide conformation recognized by aspartic, serine, cysteine and metallo proteases. Curr Med Chem 8:893-907

Vakser IA. (1995). Protein docking for low-resolution structures. Protein Eng., 8:371- 377.

Vakser IA. (1996). Main-chain complementarity in protein-protein recognition. Protein Eng., 9:741-744.

van der Spoel D., van Buuren AR., Apol E., Meulenhoff PJ., Tieleman DP., Sijbers ALTM., Hess B., Feenstra KA., Lindahl E., van Drunen R. and Berendsen HJC. (2001). GROMACS user manual. Department of biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. www.gromacs.org.

van der Spoel D., van Maaren PJ. and Berendsen HJC. (1998). A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. J. Chem. Phys. 108:10220-10230.

van Gunsteren WF., Hünenberger PH., Mark AE., Smith PE. and Tironi IG. (1995). Computer simulation of protein motion. Computer Phys. Communications 91:305-319.

van Gunsteren WF. (2001). Computer simulations of biomolecular systems: possibilities, limitations and perspectives. Combined Conference Abstracts, COMBIO2001, Plenary 9B.

van Oostveen I., Ducret A. and Aebersold R. (1997). Colloidal silver staining of electroblotted proteins for high sensitivity peptide mapping by liquid chromatography-electrospray ionization tandem mass spectrometry. Anal Biochem 247:310-8.

Vincent B., Paitel E., Saftig P., Frobert Y., Hartmann D., De Strooper B., Grassi J., Lopez-Perez E. and Checler F. (2001). The disintegrins ADAM-10 and TACE contribute to the constitutive and horbol-esters-regulated normal cleavage of the cellular prion protein. J. Biol. Chem. 276: 37743-37746.

Vlak JM., Keus RJ. (1990). Baculovirus expression vector system for the production of viral vaccines. Advances in Biotechnology Processes, 14, 91-128.

Wallace AC., Laskowski RA., Thornton JM. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8:127-134.

Wang W., Donini O., Reyes CM., Kollman PA. (2001). BIOMOLECULAR SIMULATIONS: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid non-covalent interactions. Ann. Rev. Biophys. Biomol. Struct. 30:211-243.

Warshel A. (1997). Computer Modelling of chemical reactions in enzymes and solutions. John Wiley & Sons, Inc. pp 118-119.

Webster DM. (2000). Protein structure prediction: Methods and Protocols. Humana Press, vol 143, pp. 101-103.

Wei C., Willis RA., Tilton BR., Looney RJ., Lord EM., Barth RK. and Frelinger JG. (1997). Tissue-specific expression of the human prostate specific antigene-gene in transgenic mice: implications for tolerance and immunotherapy. Immunology, 94, 6369-6374.

Weiner SJ., Kollman PA., Nguyen DT. and Case DA. (1986). An all atom force field for simulations of proteins and nucleic acids. J. Comp. Chem. 7: 230-252.

Weskamp G., Kratzschmar J., Reid MS. and Blobel CP. (1996). MDC9, a widely expressed cellular disintegrin containing cytoplasmic SH3 ligand domains. J. Cell. Biol. 132:717-726.

Westbrook JZ. and Feng. (2002). The Protein Data Bank: unifying the archive. Nucleic Acids Res 30: 245-8.

Wilmouth RC., Edman K., Neutze R., Wright PA., Clifton IJ., Schneider TR., Schofield CJ., Hajdu J. (2001). X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. Nat Struct Biol 8:689-94.

Wolf BB. and Brown MD. (1995). Epidermal growth factor-binding protein activates soluble and receptor-bound single chain urokinase-type plasminogen activator. FEBS letters, 376, 177-180.

Yamashiro K., Tsuruoka N., Kodama S., Tsujimoto M., Yamamura Y., Tanaka T., Nakazato H., Yamaguchi N. (1997). Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain. Biochimica et Biophysica Acta 1350, 11-14.

Yarski MA., Bax BD., Hogue-Angeletti RA. and Bradshaw RA. (2000). Nerve growth factor alpha subunit: effect of the site-directed mutations on catalytic activity and 7S NGF complex formation. Biochimica et Biophysica Acta, 1477, 253-266.

Yoshida S., Taniguchi M., Suemoto T., Oka T., He X. and Shiosaka S. (1998). cDNA cloning and expression of a novel serine protease, TLSP. Biochimica et Biophysica. Acta 1399, 225-228.

Yoshida S., Taniguchi M., Hirata A. and Shiosaka S. (1998a). Sequence analysis and expression of human neuropsin cDNA and gene. Gene 213, 9-16.

Yousef GM., Luo LY., Scherer SW., Sotiropoulou C. and Diamandis EP. (1999a). Molecular characterisation of Zyme/Protease M, Nerosin (PRSS9), a hormonally regulated kallikrein-like serine protease. Genomics 62, 251-259.

Yousef GM, Obiezu CV, Luo LY, Black MH, Diamandis EP. (1999b). Prostase/KLK-L1 is a new member of the human kallikrein gene family, is expressed in prostate and breast tissue and is hormonally regulated. Cancer Research 59, 4252-4256.

Yousef GM., Luo LY. and Diamandis EP. (1999c). Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4. Anticancer Res. 19, 2843-2852.

Yousef GM., Chang A., Diamandis EP. (2000a). Identification and characterization of KLK-L4, a new kallikrein-like gene that appears to be down-regulated in breast cancer tissues. J. Biol. Chem., 275, 11891-11898.

Yousef GM, Scorilas A, Diamandis EP. (2000b). Genomic organization, mapping, tisue expression and hormonal regulation of trypsin-like serine protease (TLSP PRSS20), a new member of the human kallikrein gene family. Genomics 63, 88-96.

Yousef GM, Scorilas A, Maglakara A, Soosaipillai A, Diamandis EP (2000c). The KLK7 (PRSS6) gene, encoding for the stratum corneum chymotryptic enzyme is a new member of the human kallikrein gene family – genomic characterization, mapping, tissue expression and hormonal regulation. Gene, 254, 119-128.

Yousef MG, Scorilas A, Jung K, Ashworth L, Diamandis EP. (2001). Molecular cloning of the human kallikrein 15 gene (KLK15): Upregulation in cancer. J. Biol. Chem., 276:53-61.

Yousef GM. and Diamandis EP. (1999a). The new kallikrein-like gene, KLK-L2. J. Biol. Chem., 274, 37511-37516.

Yousef GM. and Diamandis EP. (1999b). Molecular Characterization, mapping, and tissue expression of KLK-L6, a hormonally regulated kallikrein-like gene. Unpublished

Yousef GM. and Diamandis EP. (2000a). The expanded human kallikrein gene family: Locus characterization and molecular cloning of a new member, KLK-L3 (KLK9). Genomics 65, 184-194.

Yousef GM, Polymeris ME, Grass L, Soosaipillai A, Chan PC, Scorilas A, Borgono C, Harbeck N, Schmalfeldt B, Dorn J, Schmitt M, Diamandis EP. (2003). Human kallikrein 5: a potential novel serum biomarker for breast and ovarian cancer. Cancer Research 63:3958-65.

Yu H, Diamandis EP, Levesque M, Asa SL, Monne M, Croce CM. (1995). Expression of the prostate-specific antigen gene by primary ovarian carcinoma. Cancer Research, 1995, 55, 1603-1606.

Yu H, Diamandis EP, Zarghami N, Sutherland DJ. (1994). Immunoreactive prostate-specific antigen levels in female and male breast tumours and its association with steroid hormone receptors and patient age. Clinical Biochemistry, 1994; 27; 75-79.