Engineering and design of aspartic proteinase inhibitors

by

Timothy John Winterburn

Being a thesis presented in accordance with the regulations governing the award of the degree of Philisophiae Doctor in the University of Wales.

Date:

Cardiff School of Biosciences,

Cardiff University.

UMI Number: U584021

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U584021
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

Aspartic proteinases are of considerable interest to the pharmaceutical, agrochemical and food industries. Many small molecule inhibitors have been synthesised chemically to block the action of individual members of this proteinase family. In contrast, only a handful of naturally-occurring, gene-encoded inhibitors of aspartic proteinases are known. One such inhibitor is IA3, an intrinsically unstructured 68-residue cytosolic protein from *Saccharomyces cerevisiae*, which operates through an unprecedented mode of action. On contacting the target enzyme, residues 2-32 form an amphipathic α-helix that occludes the active site cleft. The potency and specificity of this inhibitor were investigated by producing >80 variants in the form of recombinant proteins in *Escherichia coli* or as synthetic peptides; preliminary attempts were also made to examine the potential of directed evolution by phage display for the generation of further mutants. The inhibitory activity of these variant polypeptides at different pH values was measured against the natural target proteinase from *S. cerevisiae* and against vacuolar aspartic proteinases from *Pichia pastoris* and *Aspergillus fumigatus*, as well as the human lysosomal enzyme, cathepsin D.

New IA3-like sequences were also identified in five closely-related yeasts. One, from *Saccharomyces castellii*, differed substantially from *S. cerevisiae* IA3 and so was investigated in parallel. Key elements of the inhibitory sequence (i.e. residues 2-34) in the *S. castellii* IA3 were identified in relation to their counterparts in *S. cerevisiae* IA3 and the different effects of their respective C-terminal regions (i.e. residues 34 to C-terminus) were unravelled through the production of chimaeric polypeptides. In this way, potent IA3-derived inhibitors of the *P. pastoris* and *A. fumigatus* proteinases were produced, and an effective inhibitor of cathepsin D was identified. IA3 is shown to be remarkably adaptable as an inhibitor, and its potency may be re-directed to inhibit more distant enzymes of commercial interest.

Contents

Abbreviations and Conventions

Å	Angstrom
APP	β-amyloid precursor protein
A. fumigatus	*Aspergillus fumigatus*
BACE 1	β-site amyloid precursor protein cleavage enzyme-1
BACE 2	β-site amyloid precursor protein cleavage enzyme-2
bp	Nucleotide base pairs
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
ConA	Concanavalin A
CPK	Corey, Pauling and Kultun
C-terminal	Carboxy-terminal
CT(H)$_6$	C-terminal tag hexahistidine tag
D$_{600}$	Attenuance at 600 nm
Da	Daltons
DAN	diazoacetyl-norleucine-methyl-ester
DAPase™	diamino exopeptidase cathepsin C
DNA	Deoxyribonucleic acid
dATP (dCTP, dGTP, dTTP)	2′-deoxyadenosine 5′-triphosphate (and similarly for cytidine, guanosine, thymidine)
dNTP	Deoxynucleotide triphosphate
DMSO	Dimethyl sulphoxide
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetracetic acid

EPNP	1,2,-epoxy-3-nitrophenoxypropane
FPLC	Fast protein liquid chromatography
HIV	Human immunodeficiency virus
IgG	Immunoglobulin class G
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-thiogalactopyranoside
IVP	Isovaleryl-pepstatin
J	Phenyl-glycine
k_{cat}	Catalytic constant
K_i	Inhibitory constant
K_m	Michaelis constant
LB	Luria-Bertani
MALDI-TOF	Matrix-Assisted Laser Desorption Ionisation –Time-of-Flight
MS	Mass spectrometry
MES	2-[N-morpholino]ethane-sulphonic acid
Nle	Norleucine
Nph	*para*-nitrophenylalanine; [Phe(4-NO_2)]
NTA	nitrilotriacetic acid
N-terminal	Amino-terminal
NT(H)$_6$	N-terminal tag comprising M-K-H-H-H-H-H-H-H-M-Q
Pa	Pascal
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pfu	plaque forming unit(s)

phage	Bacteriophage
PheGly	Phenyl-glycine
P. pastoris	*Pichia pastoris*
PVDF	Polyvinylidene difluoride
Qcyclase™	glutamine cyclotransferase
***Q**	pyroglutamate
®	Registered trademark
rms	Root mean square
S	Substrate
Sv	Substrate variant
SAP	Secreted aspartic proteinase
S. castellii	*Saccharomyces castellii*
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulphate
Sta	Statine ((4s 3s)-4-amino-3 hydroxyl-6-methyl heptanoic acid
TAE	Tris, acetic acid, EDTA
TTBS	TBS containing 0.05% (v/v) Tween® 20
TBS	Tris-buffered saline
TEMED	N,N,N′,N′,-Tetramethylethylenediamine
TFA	Trifluroacetic acid
™	Trademark
Tris	2-Amino-2-(hydroxymethyl)-1,3-propandiol
v	Enzyme reaction velocity
V_{max}	Maximal limit of velocity

WT	Wild-type
YGWYSF	You get what you select for
Z	norleucine

The one and three letter abbreviations for nucleotides and amino acids are those recommended for use by the Biochemical Journal (Policy of the Journal and Instructions to Authors). Biochem. J. (1983), **209**: 1-27.

All molecular graphics figures (with the exception of Figures 5.1 and 5.2) were constructed by the author using the freeware program DeepView/SwissPDBviewer (version 3.7, http://www.expasy.org/spdbv) and rendered using the freeware program POV-Ray™ (version 3.6, http://www.povray.org).

Notes to the reader

This thesis should comprise three elements:

(i) VOLUME I: the main body of text

(ii) VOLUME II: all data tables from Chapters 5-10.

(iii) a flat-packed 3D/stereo viewer.

This is a thesis in 2 parts. Chapters 3 and 4 describe the genetic construction, and the expression and purification of the various recombinant forms of IA3 that were used in the experiments described in Chapters 5-10. These later chapters largely refer to variant IA3s by way of numbered references to the primary sequences detailed in the data Tables. This numbering system is consistent throughout the thesis so that, for example **1** is always wild-type *S. cerevisiae* IA3 and **42** is always is wild-type *S. castellii* IA3. The data tables can be found at the end of the relevant chapter, i.e. after the end of chapter discussion. However, the data tables from all chapters are collected together and reproduced in Volume II of this thesis, in an effort to reduce the need for "page flicking" backwards and forwards. Within the data tables, and in many of the figures, sequence elements originating from *S. cerevisiae* IA3 are coloured in yellow, whilst *S. castellii* IA3 sequence is coloured blue.

Finally, numerous molecular graphics figures in this thesis are parallel stereo pairs. These are compatible with all normal stereo viewers, such as the Taylor-Merchant Stereopticon® 707. In case these are not immediately available to the reader, please find enclosed the Loreo Lite 3D viewer. Instructions for its use can be found on the flat-packed viewer itself.

Chapter 1

Introduction

1.1 Aspartic proteinases and their sources

Based on the work of Hartley (1960), proteinases have been classified traditionally into four groups according to their chemical mechanism of catalysis (serine, cysteine, metallo and aspartic). However, the expansion of knowledge heralded by the "genomic era" has revealed many proteinases that do not fit into this conventional wisdom, thereby necessitating a refinement of the classification system. The MEROPS peptidase database (Rawlings *et al.*, 2004; http://merops.sanger.ac.uk) now records proteinases in a more sophisticated manner, although the basis of classification by catalytic mechanism is maintained. At least six types are now recorded, with the addition of the threonine and the glutamic peptidases (formerly known as carboxyl peptidases) to the four groups described above.

Aspartic proteinases (themselves formerly termed more generally as acid proteinases) are characterised by acidic pH optima and their use of two aspartic acid residues in the catalytic mechanism. Classically, an enzyme was classified as an aspartic proteinase if it was susceptible to inhibition by isovaleryl-pepstatin and by specific active site directed labels (see Section 1.3) such as diazoacetyl-norleucine-methyl-ester (DAN) and the epoxide 1,2,-epoxy-3-nitrophenoxypropane (EPNP) (Chen & Tang, 1972; Marcinisyn *et al.*, 1976; Kay, 1982). The classification process now recognises amino acid sequence and structural homology as important considerations as well, although these factors in themselves may lead to some complications. For example, Gustchina *et al.*, (2005) describe an aspartic proteinase that demonstrates little or no catalytic activity, a feature which on its own may have made classification by the traditional methods impossible. Nevertheless, the "clan" of aspartic proteinases contains all proteins that share a common evolutionary origin. This largest grouping includes those families with similar tertiary structures and which have the catalytic residues within a common sequence motif. The aspartic proteinases that are studied experimentally in this thesis are all from one family. However, >10 families of aspartic proteinases are currently (November 2005) reported in the MEROPS database, and their various roles are as diverse as the organisms and phyla in which they are found.

1.1.1 Vertebrates

Pepsin and gastricsin are extracellular proteinases in the gastric juices of mammals and both are involved in the digestion of dietary proteins. Both are classic examples of the pepsin-like family of aspartic proteinases, which includes all of those studied in detail in this thesis. This family was reviewed by Dunn (2002). Chymosin (previously known as rennin), is another gastric enzyme. It demonstrates a weak general proteolytic activity and a high milk clotting activity (Foltmann & Szecsi, 1998), but is found principally in the gastric juices of neonates from several species including cow, pig, rabbit and rat (Kageyama *et al.*, 2000). Notably, the human chymosin gene is inactive, as a consequence of three separate nucleotide deletions that shift the reading frame leading to premature termination of translation (Ord *et al.*, 1990).

Renin, another extracellular member of the family, is secreted into the blood where it has a role in the regulation of blood pressure by catalysing the first, rate-limiting step in the conversion of angiotensinogen into the hormone angiotensin II (Suzuki *et al.*, 1998; Fuchs *et al.*, 2002). This physiologically-important hormone modulates blood pressure directly by inducing vasoconstriction, and thus renin has proved an attractive target for the development of anti-hypertensive drugs (Abdelmeguid, 1993; Oefner *et al.*, 1999; Rahuel *et al.*, 2000). It has also been shown recently to have an important role in kidney development (Gribouval *et al.*, 2005).

In contrast, cathepsins D and E are both intracellular proteinases. Cathepsin D, which will be considered further in Chapter 10, is a lysosomal enzyme present in most cell types (Nakanishi *et al.*, 1997) and is believed to be involved in protein catabolism. Cathepsin E, which has a slightly different cellular localisation, is an endosomal proteinase with a distribution limited to certain cell types such as the gastrointestinal tract, lymphoid tissues and urinary organs. It appears to be contained within vesicular structures associated with the endoplasmic reticulum, endosome and Golgi complex (Kay & Tatnell, 1998). Although the exact physiological function of cathepsin E is unknown, it has been implicated in antigen processing (Bennet *et al.*, 1992; Chain *et al.*, 2005) and in the processing of biologically active peptides (Lees *et al.*, 1990; Athauda *et al.*, 1991). Tsukuba *et al.*, (2003) also reported that

mice lacking cathepsin E spontaneously develop skin lesions similar to those of humans with atopic dermatitis.

Another intracellular aspartic proteinase is napsin A. Currently, no specific function has been assigned to this enzyme. However, tissue distribution indicates it is predominantly in lungs and kidney (Chuman *et al.*, 1999; Schauer-Vukasinovic *et al.*, 1999) and it has become an excellent marker for distinguishing primary lung adenocarcinoma from either adenocarcinomas of other organs or other histological types of carcinoma (Suzuki *et al.*, 2005). In humans, there is a second napsin gene (napsin B), which is transcribed in spleen and immune cells but not in kidney or lung (Tatnell *et al.*, 1998). This gene lacks an in-frame stop codon which led the authors to postulate that napsin B may be a transcribed pseudo-gene, in a similar manner to human chymosin.

Several other aspartic proteinases have been discovered within the human genome that are thought to play important roles in the pathogenesis of Alzheimer's disease. The major pathological event in Alzheimer's disease is the formation of plaques that are composed of aggregated β-amyloid peptides, produced by cleavages in a protein called β-amyloid precursor protein (APP). These hydrolyses were thought to be catalysed by two proteinases, known as β and γ secretases. The activity of the γ secretase is currently considered to originate from an enzyme called presenilin-1 (Wolfe *et al.*, 1999), and possibly also presenilin-2 (Lai *et al.*, 2003; Saura *et al.*, 2004). Both presenilins, therefore, are potential therapeutic drug targets to combat Alzheimer's disease (Beher *et al.*, 2003). It has also been suggested that inhibitors of γ secretase may be of benefit in colorectal neoplastic diseases (van Es *et al.*, 2005). However, therapeutic targeting of the γ secretase(s) may be complicated by their similarity to Impas 1 endopeptidase (Weihofen *et al.*, 2003), a signal peptidase that has been shown by Casso *et al.* (2005) to be critical to embryonic development in *Drosophila*. Furthermore, double knock-out mice lacking both presenilins 1 and 2 showed a slow impairment of memory, synaptic plasticity and age-dependent neurodegeneration (Saura *et al.*, 2004), again raising questions over the clinical value of targeting these enzymes.

The principal origin of the β secretase activity is thought to be a proteinase termed BACE-1 (β-site amyloid precursor protein cleavage enzyme), although this nomenclature has been somewhat controversial with the proteinase also variously referred to as Asp 2 and Memapsin 2 in the literature. Sinha & Lieberberg, 1999 proposed that cleavage of APP by BACE-1 is the rate determining step in the formation of amyloid plaques. For this reason, BACE-1 has been the subject of much interest as a target for the treatment of Alzheimer's disease. However, the discovery of BACE-1 led to the identification of another aspartic proteinase, termed BACE-2 (or, by other nomenclatures, Memapsin 1) that has a high degree of identity with BACE-1. The BACE-2 gene was mapped to a location on human chromosome 21 associated with Downs Syndrome. This led to the suggestion that BACE-2 may also function as a second β secretase, and that it is associated with the pathology of both Downs Syndrome and Alzheimer's disease (Kim *et al.*, 2002). Certainly, BACE-2 also cleaves APP at the same β secretase site as BACE-1 (Farzan *et al.*, 2000) and, like BACE-1, it is highly expressed in the brain (Hussain *et al.*, 2000). However, BACE-2 and BACE-1 are not co-expressed. Combined with the observation that BACE-1 deficient mice have greatly reduced levels of β-amyloid in their brains, the inference was drawn that BACE-1 plays the major role as the brain β secretase (Luo *et al.*, 2001; Roberds *et al.*, 2001). Furthermore, in contrast to the γ secretases, the β secretases may provide a simpler target for therapeutic treatment of Alzheimer's disease because BACE-1 knock-out mice show no profound phenotypic defects (Roberds *et al.*, 2001).

1.1.2 Parasites

The human malarial parasite *Plasmodium falciparum* encodes 10 aspartic proteinase homologues, known as the plasmepsins (Klemba & Goldberg, 2005). Plasmepsins I, II and IV, together with the sequence- and structurally-related histo-aspartic proteinase, have all been shown to be critical to digestion of haemoglobin in the food vacuole of the parasite, making them attractive anti-malarial therapeutic targets (Ersmark *et al.*, 2004; Kesavulu *et al.*, 2005). However, the functions of the other six plasmepsins are not yet known.

Aspartic proteinases have been discovered in a variety of other parasites including *Ancylostoma caninum*, the common hookwork of dogs (Harrop *et al.*, 1996; Williamson *et al.*, 2002), *Schistosoma mansoni* (Wong *et al.*, 1997), and disease causing nematodes such as *Onchocerca volvulus* that is responsible for the condition known as "river blindness" (Jolodar & Miller, 1997). Aspartic proteinases have also been discovered in *Eimeria acervulina* (Laurent *et al.*, 1993) and from *Eimeria tennella* (Jean *et al.*, 2000). Parasites of the genus *Eimeria* have avian hosts and are the focus of much veterinary interest, as they are responsible for huge losses in the poultry industry.

1.1.3 Plants

A recent review of the structure and functions of aspartic proteinases from plants was provided by Simoes & Faro (2004). For many centuries, aspartic proteinase activity from plants has been used in the production of cheese. Traditionally, the enzyme was obtained from an aqueous extract of the flowers from the cardoon plant (*Cynara cardunculus*) and this clotted the milk prior to cheese making (Heimgartner *et al.*, 1990). Several candidate enzymes thought to be responsible for this process have been identified, namely, cyprosin, cardosin A and cardosin B (Heimgartner *et al.*, 1990; Verissimo *et al.*, 1996). In view of the commercial importance, these enzymes have been studied in detail (Ramalho-Santos *et al.*, 1996; Verissimo *et al.*, 1996; Faro *et al.*, 1999; White *et al.*, 1999). Although the exact physiological role of each enzyme *in planta* remains elusive, the cardosins have been shown to be involved in cell death events associated with plant senescence and stress responses (Castanheira *et al.*, 2005).

Aspartic proteinases have now been found in the genomes of many plants, although in several cases their roles remain unknown. Phytepsin, a close relative of the cardosins, is an aspartic proteinase from barley grains and other seeds and is involved in the hydrolysis of storage proteins upon germination (Sarkkinen *et al.*, 1992; Terauchi *et al.*, 2004; de Carvalho *et al.*, 2004). Genes encoding tens of putative aspartic proteinases have been identified in *Arabidopsis thaliana*, but few of these have been characterised yet (Chen *et al.*, 2002). An aspartic proteinase from the cells

of the extremely commercially-relevant *Nicotiana tabacum* has also been reported (Murakami *et al.*, 2000). The authors suggested that this enzyme has the potential to bind DNA and so postulated that it may play a role in the degradation of transcription factors.

In response to wounding, plants produce "pathogenesis-related" proteins at high levels, and several of these enzymes are aspartic proteinases (Rodrigo *et al.*, 1989; Kitajima & Sato, 1999). For example, one such enzyme is responsible for the selective hydrolysis of some of these pathogenesis-related proteins within tomato plants (Rodrigo *et al.*, 1989; Rodrigo *et al.*, 1991). It has been suggested that this enzyme may increase intracellular protein turnover and, therefore, contribute to increased amino acid pools for the synthesis of further defence proteins. Alternatively, it may have a more direct role in the hydrolysis of proteins secreted as virulence factors by invading pathogens (Schaller & Ryan, 1996).

An *et al.*, (2002) reported the discovery of five aspartic proteinases in the pitcher fluid and digestive glands of the carnivorous plant *Nepenthes alata*. The authors suggested that two of these enzymes in particular, NaAPs 2 and 4, enable the plant to utilise insect proteins as a nitrogen source by facilitating digestion of prey.

1.1.4 Retroviruses

Aspartic proteinases have also been shown to perform essential roles in the replication cycles of retroviruses including, amongst others, the human immuno-deficiency virus Types 1 and 2 (HIV-1 and HIV-2) (Ratner *et al.*, 1985), Rous sarcoma virus (RSV) (Schwartz *et al.*, 1983), simian immuno-deficiency virus (SIV) (Zhao *et al.*, 1993), feline immuno-deficiency virus (FIV) (Wlodawer *et al.*, 1995), and equine infectious anaemia virus (EIAV) (Powell *et al.*, 1996). They have also been implicated within the pararetroviruses from cauliflower mosaic virus (Torruella *et al.*, 1989), rice tungro bacilliform virus (Hay *et al.*, 1994) and cocoa swollen shoot virus (Hagen *et al.*, 1993).

1.1.5 Microorganisms

Microorganisms such as fungi produce a number of aspartic proteinases. The opportunistic pathogen, *Candida albicans*, the causative agent of thrush, uses as many as eight secreted aspartic proteinases (SAPs) during the process of host invasion (Morrison *et al.*, 1993; Monod *et al.*, 2002). It has also been observed that the infecting *Candida* cells use their arsenal of SAPs in a temporal and specific manner for the progression of this infection (Schaller *et al.*, 1998). Indeed, in the human oral candidosis model, the deletion of certain SAP genes was compensated for by the up-regulation of alternative and compensatory SAP genes (Schaller *et al.*, 1999). Aspartic proteinases secreted by fungi are in general very similar to pepsin.

The pathogenic role of aspartic proteinases extends to plant pathogens also (Hill & Phylip, 1998). The plant pathogenic fungi, *Botrytis cinerea* (Movahedi *et al.*, 1991; ten Have *et al.*, 2004) and *Glomerella cingulata* (Clark *et al.*, 1997) have both been shown to secrete aspartic proteinases during the early stages of plant infection.

However, not all fungal aspartic proteinases are secreted. The vacuolar aspartic proteinase from *Saccharomyces cerevisiae*, known as proteinase A, together with other fungal vacuolar aspartic proteinases (from *P. pastoris* and *A. fumigatus*) form the targets of much of the work described in this thesis. Therefore, *S. cerevisiae* proteinase A is described in more detail in Section 1.8.2, and the other fungal enzymes are introduced within the relevant chapter.

Finally, DNA encoding aspartic proteinase-like domains has been identified within the genomes of the prokaryotes *Escherichia coli* and *Haemophilus influenzae*. Expression of the recombinant domains in isolation showed that they could each function as active proteinases (Hill & Phylip, 1997). Unrelated to these, some integral membrane proteins in *E. coli* demonstrate aspartic proteinase activity and play a role in signal peptide processing (Barrett & Rawlings, 1995).

The selected examples listed above represent only a fraction of the range of aspartic proteinases. Nevertheless, they highlight both the extensive distribution and the wide

range of roles that the members of this family of enzymes perform. However, despite their functional diversity, the aspartic proteinases share certain common features that are characteristic of the enzymes in this family.

1.2 Precursors of Aspartic proteinases

All enzymes within this family are produced as quiescent precursors, known as zymogens. Zymogens contain an N-terminal "pro-segment" that maintains the enzyme in a dormant state, although the mechanisms whereby this is achieved are quite varied. These prosequences normally comprise 40-50 amino acids (Tang & Wong, 1987); however, the properts of some of the enzymes, such as plasmepsin II, are in excess of 120 amino acids in length (Dame *et al.*, 1994). In addition to regulating the enzyme activity, the propart may stabilise the structure at neutral pH (Tanaka & Yada, 2001; Dunn, 2002), or assist the protein to fold into a stable form (Klionsky *et al.*, 1988; van den Hazel *et al.*, 1993; Koelsch *et al.*, 1994). Activation of the zymogen to yield fully active proteinase is by limited proteolysis, sometimes of a single bond, and can occur by several mechanisms but all lead to a molecular rearrangement that results in exposure of the full catalytic activity of the enzyme. To achieve activation, the zymogen can totally self-activate, partially autoactivate or undergo processing by another proteinase. Porcine pepsinogen is a classic example of a zymogen that undergoes self-activation. At neutral pH values, the prosegment adopts a conformation, comprising one β-strand and three helices, that fits into the active site, thus preventing the enzyme engaging substrates (Sielecki *et al.*, 1991). However, the acidic pH environment within the stomach triggers the conversion of the zymogen to the active enzyme (Tang, 1998) by disrupting the salt bridges and the hydrogen bonds that maintain the pro-segment within the active site. Following a conformational change to the pro-segment, autocatalysis proceeds with the release of the propart residues (Dunn, 1997; Khan *et al.*, 1997; Khan *et al.*, 1999; Kageyama, 2002). Notably, aspartic proteinases that are pH-activated appear to share a Lys36-Tyr37 sequence in the propart that interacts with the pair of catalytic aspartic acids residues and displaces the catalytic water molecule (Section 1.4) and so provides a lock for the active site (Koelsch *et al.*, 1994). The events associated with

the activation of zymogens of gastric aspartic proteinases have been reviewed by Richter *et al.* (1998).

However, this occlusion of the active site is not the only method adopted by aspartic proteinase zymogens for self-inhibition. The propart of plasmepsin II interacts extensively with the C-terminal domain of the enzyme and induces a conformational reorientation of the two domains (or lobes, see Section 1.5) of the protein. This increases the separation of the catalytic residues by approximately 4Å (Khazanovich-Bernstein *et al.*, 1999) and thus renders the zymogen inactive even though its active site is fully accessible. Recombinant forms of proplasmepsin II catalyse unassisted activation and excise their own propart regions. However, the naturally occurring enzyme, as isolated from the parasite, lacks several of the N-terminal residues that are present in the "mature" form generated from recombinant preparations (Tyas, 1997). This implies that this zymogen undergoes assisted or complete processing by another enzyme.

The human precursor prorenin is an example of a zymogen incapable of performing autoactivation. Instead, the 46 residue propart of the enzyme is removed by (an)other proteinase(s). Both trypsin (Atlas *et al.*, 1985) and cathepsin B (Wang *et al.*, 1991) have this capability *in vitro*. Likewise, the aspartic proteinases encoded by many retroviruses are also produced as precursors, although with these enzymes the precursor is not as a proform but takes the form of a polyprotein that is processed by the virally encoded proteinase (Rothnie *et al.*, 1994).

Thus, proparts are a common feature of aspartic proteinases that helps to suppress the activity of the mature enzyme, promotes stability and may influence the folding pathway that enables the structure to achieve the native state. Additionally, the propart may carry information relating to the physiological function of the enzyme. For example, the 76-residue preprosegment of proteinase A from *S. cerevisiae* includes the vacuolar protein-sorting signal (Klionsky *et al.*, 1998). Deletion of the prosegment led to the proteinase being degraded (van den Hazel *et al.*, 1993).

1.3 The principal features of the primary sequence of aspartic proteinases

The explosion of genomic information from a wide variety of organisms has enabled the elucidation of thousands of aspartic proteinase sequences. Indeed, as of November 2005, the MEROPS database has 2307 entries under "aspartic peptidase". However, within this extended family of enzymes, the members share certain regions of sequence identity. One key region spans the first catalytic aspartate residue, Asp32 (numbering of porcine pepsin):

~Hydrophobic-Hydrophobic-Asp-Thr/Ser-Gly~

In all but the dimeric retroviruses (see later), this motif is repeated at the second catalytic aspartate (Asp215). In recent times, however, a few exceptions to this hallmark motif have been noted, such as the cockroach allergen Bla g 2 (Gustchina *et al.*, 2005) which demonstrates Asp^{32}-Ser-Thr and Asp^{215}-Thr-Ser motifs flanking each of these Asp residues.

Invariably, however, the two catalytic aspartates are conserved, confirming their importance in the catalytic mechanism. The key role of these residues was demonstrated experimentally by the use of two different active-site directed, affinity labels: diazoacetyl-norleucine methyl ester (DAN) and epoxide 1,2-epoxy-3-nitrophenoxypropane (EPNP). For example, these two compounds react covalently with the two catalytic aspartate residues in porcine pepsin resulting in a total loss of enzymatic activity (Chen & Tang, 1972).

Crystallographic studies on aspartic proteinases co-crystallised with inhibitors bound within their active sites have shed light on the significance of the Thr/Ser-Gly motifs that follow the catalytic aspartates in the primary sequences. The analysis by Pearl & Blundell (1984) of the structure of endothiapepsin revealed a symmetrical network of hydrogen bonding around the two catalytic aspartate residues, and that the carboxyl groups of these two aspartates lie orientated within a plane. Thus, one Asp-Thr-Gly partners with its spatially adjacent reciprocal pair of Asp-Thr-Gly in an arrangement likened to a "fireman's grip" (Figure 1.1A and B),

A

B

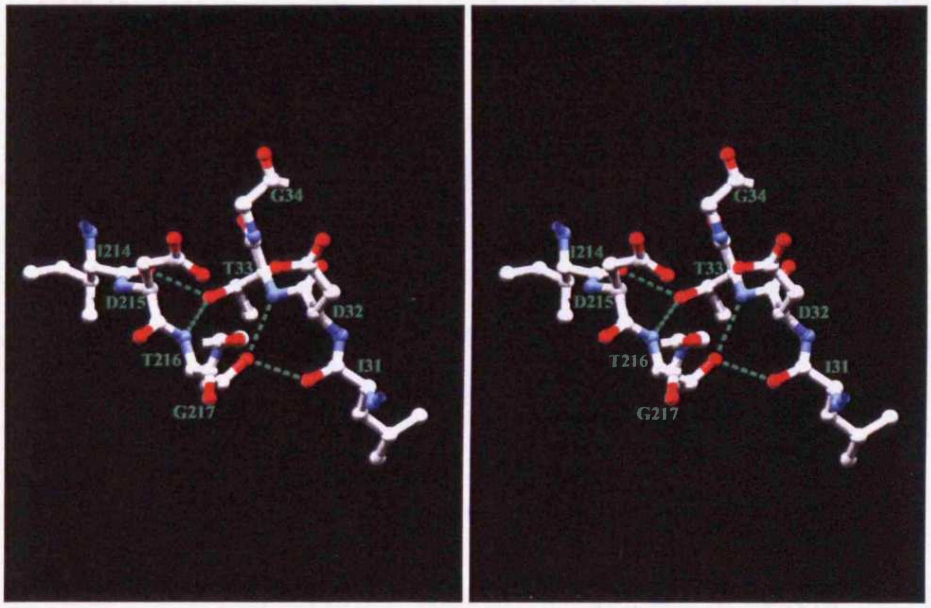


Figure 1.1 A and B Fireman's grip. Stereo image (**B**) of *S. cerevisiae* vacuolar proteinase A showing the fireman's grip formed by D32-T33-G34 with D215-T216-G217. (PDB code 1DPJ).

where "one hand holds the arm whose hand holds its arm". Using this analogy, the side chain oxygens of the two threonines represent the "hands", and the backbone of the respective loops to which they are attached form the "arms". In precise terms, the hydroxyl oxygen of Thr33 (porcine pepsin numbering) is in contact with the amido nitrogen of Thr216 and the backbone carbonyl oxygen of residue 214. In a reciprocal fashion, the hydroxyl oxygen of Thr216 interacts with the nitrogen of Thr33 and the carbonyl oxygen of the residue at position 31. This network of hydrogen bonds stabilises the centre of the active site of the proteinase.

1.4 The catalytic mechanism of aspartic proteinases

A peak of electron density identifiable during the X-ray crystallographic studies of aspartic proteinases is proposed to be a water molecule, positioned equidistant between the two catalytic aspartic acid residues (James & Sielecki, 1983; Pearl & Blundell, 1984). This information, taken in conjunction with the spatial arrangement of the two catalytic aspartate residues within the active site cleft, as observed in the structures of aspartic proteinases complexed with transition state analogues (Sugana *et al.*, 1987; James *et al.*, 1992; Parris *et al.*, 1992; Silva *et al.*, 1996; reviewed in Dunn, 2002), supplemented with kinetic data and isotope (^{15}N and ^{18}O) effects (Hyland *et al.*, 1991; Rodriguez *et al.*, 1993) led to a generally accepted model of the catalytic mechanism for aspartic proteinases (Veerapandian *et al.*, 1992). More recent high resolution structural evidence (Coates *et al.*, 2002) has provided further support to the mechanism outlined below and presented schematically in Figure 1.2.

Prior to the cleavage of a peptide bond, a water molecule is poised between a protonated Asp32 and unprotonated Asp215 (porcine pepsin numbering). The negatively charged form of Asp215 is stabilised by H-bonding to Thr218, so facilitating the role of this catalytic aspartate to act as a general base. Entry of the substrate enables the protonated oxygen of Asp32 to establish a hydrogen bond to the peptidic carbonyl oxygen of the substrate. The reaction is initiated by the delocalisation of a single negative charge (Asp215) onto the bound water molecule, caused by the presence of the central proton positioned between the "inner oxygens"

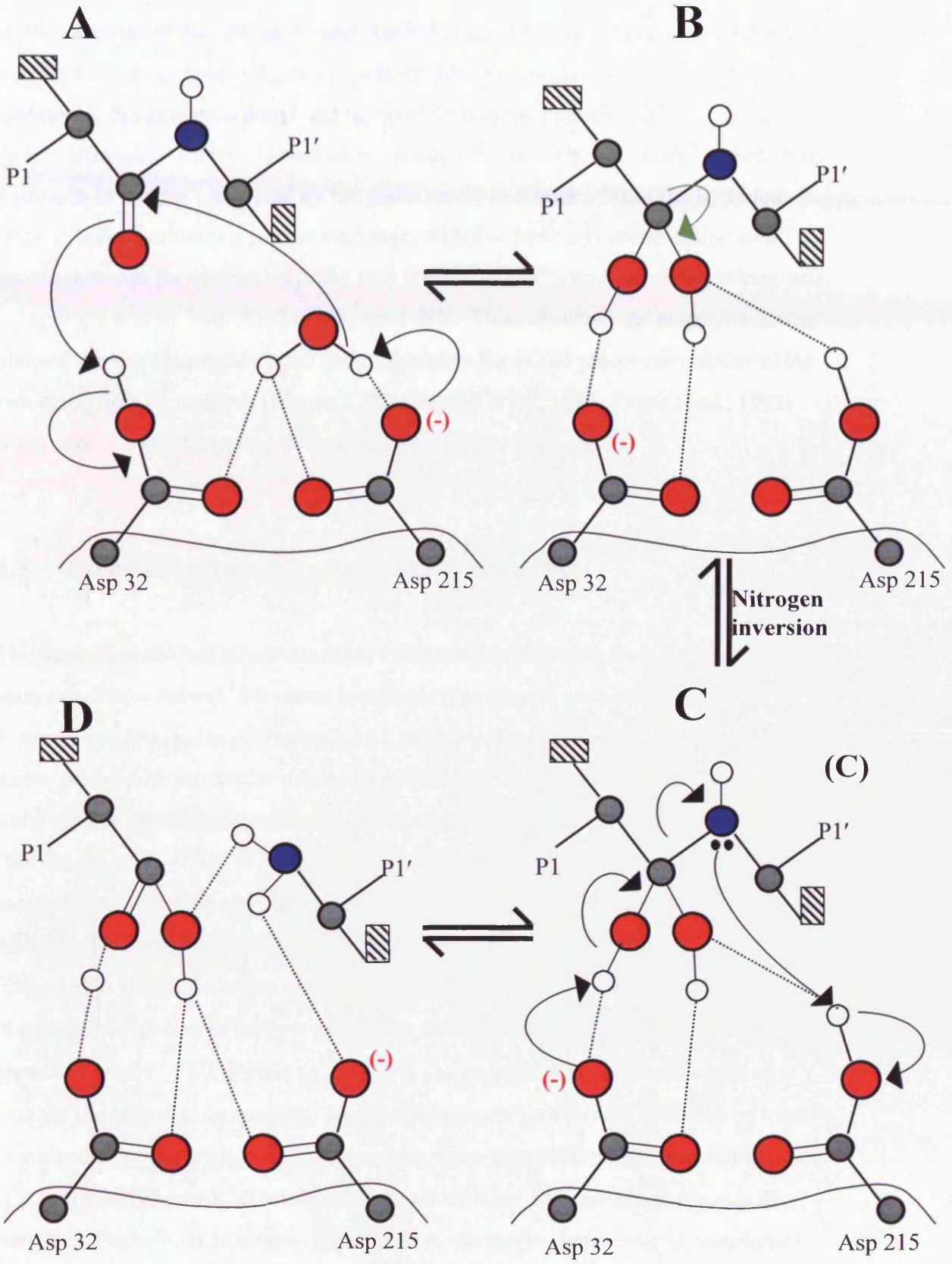


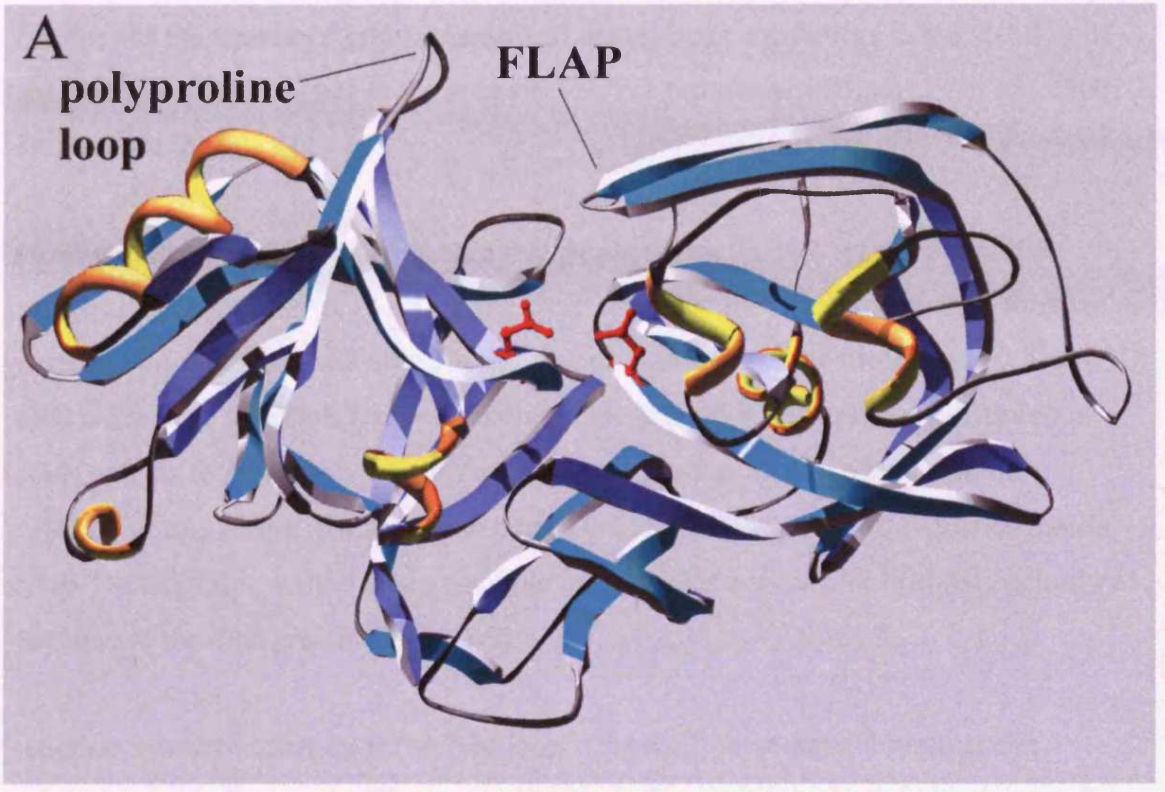
Figure 1.2 Proposed catalytic mechanism of aspartic proteinases. Adapted from Coates *et al.* (2002). See text for details.

of the carboxyl groups of Asp32 and Asp215 (Figure 1.2A). This confers enhanced nucleophilic properties to the bound water molecule, and facilitates the hydroxide addition to the amide carbonyl and the hydration of the peptidic carbonyl moiety to form a tetrahedral transition state intermediate (Figure 1.2B). A rearrangement (the "nitrogen inversion", depicted by the green arrow in Figure 1.B) of the hydrated peptide bond facilitates a proton exchange, with one hydroxyl group acting as a proton donor to the charged aspartic acid residue (Asp32), and the nitrogen lone pair accepting a proton from Asp215 (Figure 1.2C). This simultaneous proton exchange cleaves the scissile peptide bond and regenerates the initial protonation states of the two catalytic acid residues (Figure 1.2D) (Suguna *et al.*, 1987; Parris *et al.*, 1992; Silva *et al.*, 1996; Andreeva & Rumsh, 2001; Coates *et al.*, 2002).

1.5 Structural features of aspartic proteinases

The three-dimensional structures many of aspartic proteinases from a wide variety of sources are now solved. Monomeric aspartic proteinases, such as pepsin or *S. cerevisiae* proteinase A (Figure 1.3A), are typically single chain enzymes (~330 amino acids) with molecular masses of 35-43 kDa. However, two chain forms of some aspartic proteinases exist also, for example those from retroviruses (Wlodawer *et al.*, 1989; Wlodawer & Gustchina, 2000). The general form of aspartic proteinases is a bilobal structure created by two similar (or, in some cases, identical) domains (Figure 1.3A) with an approximate two-fold axis of symmetry (Tang *et al.*, 1978). In terms of secondary structure, each domain possesses a series of anti-parallel β-sheets that has a topology similar to the corresponding sheets in the opposite domain. This creates an extensive six-stranded anti-parallel β-sheet as a base for the enzyme. In contrast, the dimeric aspartic proteinases, typically of viral origin and sometimes known as retropepsins, form homodimers of a molecular mass of 20–30 kDa (Figure 1.3B) stabilised by non-covalent interactions at the subunit interface. Each dimer is structurally related to the single chain aspartic proteinases (Dunn *et al.*, 2002). Both monomeric and dimeric forms of these enzymes have secondary structure composed mainly of β-sheet, with very little α-helical content. The dimeric enzymes possess some unique structural characteristics, particularly at

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD



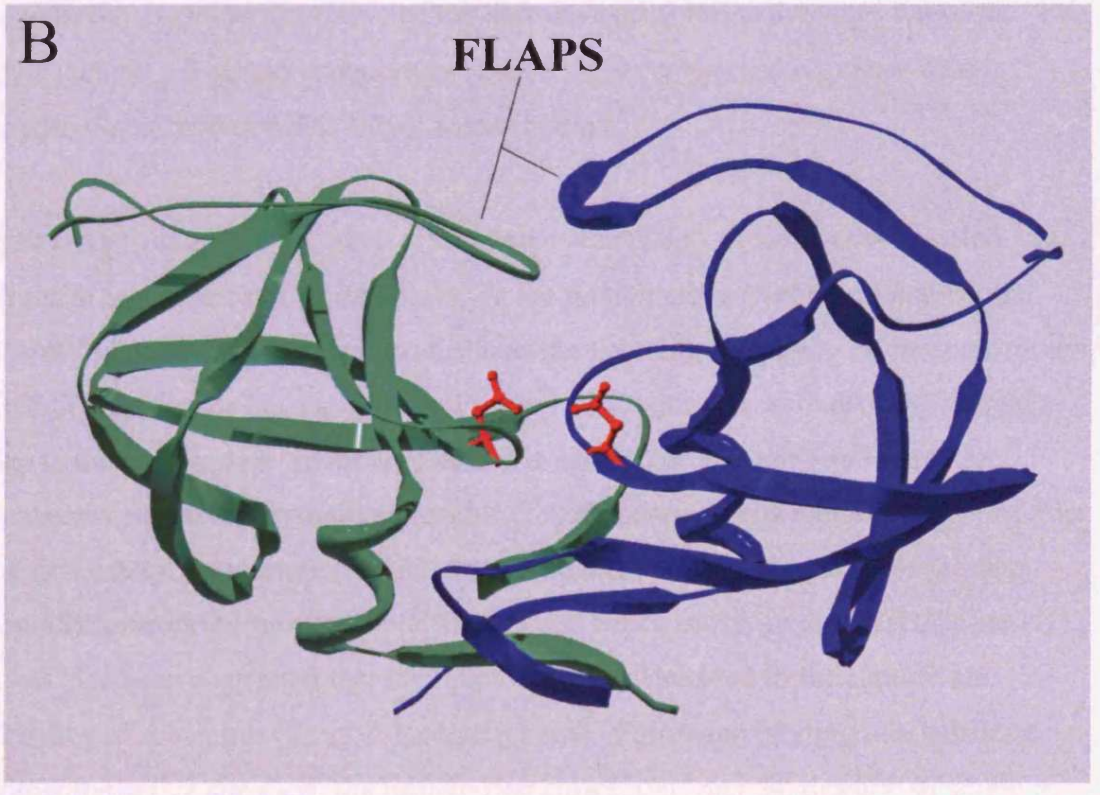


Figure 1.3 Aspartic proteinases of eukaryotic and viral origin. **A.** *S. cerevisiae* vacuolar proteinase A (PDB code 1DPJ). **B.** HIV proteinase-1 (PDB code 1HVP). The catalytic aspartate residues at the bottom of the active site cleft are in red.

the dimer interface. In this region, the N- and C-termini of one subunit interdigitate with the termini of the other to form a 4-stranded β-sheet. Furthermore, the Thr/Ser residues of the fireman's grip (Section 1.3) are critical contributors to the stability of the formation of the dimer in the retroviral HIV-1 proteinase (Strisovsky *et al.*, 2000; Dunn *et al.*, 2002).

However, there are also many shared features between the two types of aspartic proteinase, such as the large active site cleft formed at the junction between the two barrel-like domains (or subunits, in the case of dimeric aspartic proteinases). The cleft is 25-30 Å in length and can accommodate a peptide substrate of 7-8 amino acids, which, as in other types of proteinase, is bound as an extended β-strand (Tyndall *et al.*, 2005). Each domain contributes one of the catalytic sequence motifs (Asp-Thr/Ser-Gly), with the two catalytic aspartic acid residues positioned centrally at the base of the cleft (red in Figure 1.3).

Another conserved feature is the "Psi loop" (Figure 1.4), so named because the intersection of a key part of the structure takes a form resembling the Greek letter Psi (ψ). Each ~Asp-Thr/Ser-Gly~ (of the fireman's grip) forms the cross bar of the "Psi loop", whilst a β-strand containing a second highly conserved sequence motif, ~hydrophobic-hydrophobic-Gly~, passes behind.

A flexible β-hairpin loop, often referred to as the "flap", projects over the cleft and closes to trap substrates or inhibitors. In the monomeric aspartic proteinases, the N-terminal lobe of the enzyme contributes the flap (Figure 1.3A). In contrast, dimeric aspartic proteinases have two flaps (Figure 1.3B), with each subunit contributing a flap to the partnership. In most monomeric aspartic proteinases, an interaction between a conserved tryptophan residue (Trp39, porcine pepsin numbering) just after the first catalytic aspartate (Asp32) and a tyrosine located near the tip of the flap (Tyr75) controls the movement of the flap and hence access to the cleft (Figure 1.5). It has also been suggested that the tip of the flap is involved in the capture and cleavage of substrates (Tang & Koelsch, 1995). Following binding of a substrate within the active site cleft, the lid-like flap closes, thus maintaining the substrate in the correct orientation for hydrolysis (Wlodawer & Gustchina, 2000;

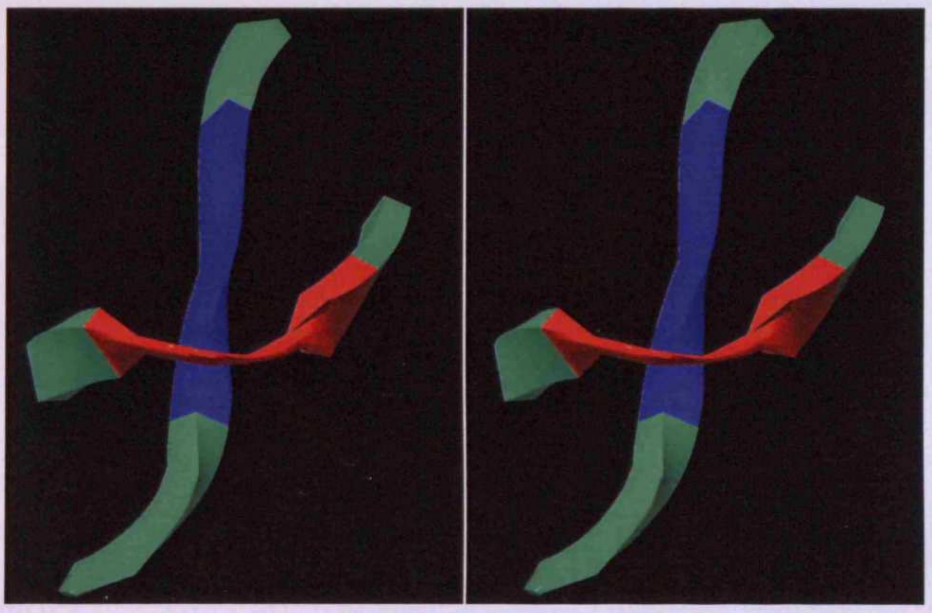


Figure 1.4 Stereo view ribbon representation of the residues constituting the "Psi-loop" of *S. cerevisiae* vacuolar proteinase A (PDB code 1DPJ). The cross of the Psi (red) is constructed of Asp32, Thr33 and Gly34 (right to left). The vertical element (blue) is formed from Ile120, Leu121 and Gly122 (top to bottom).

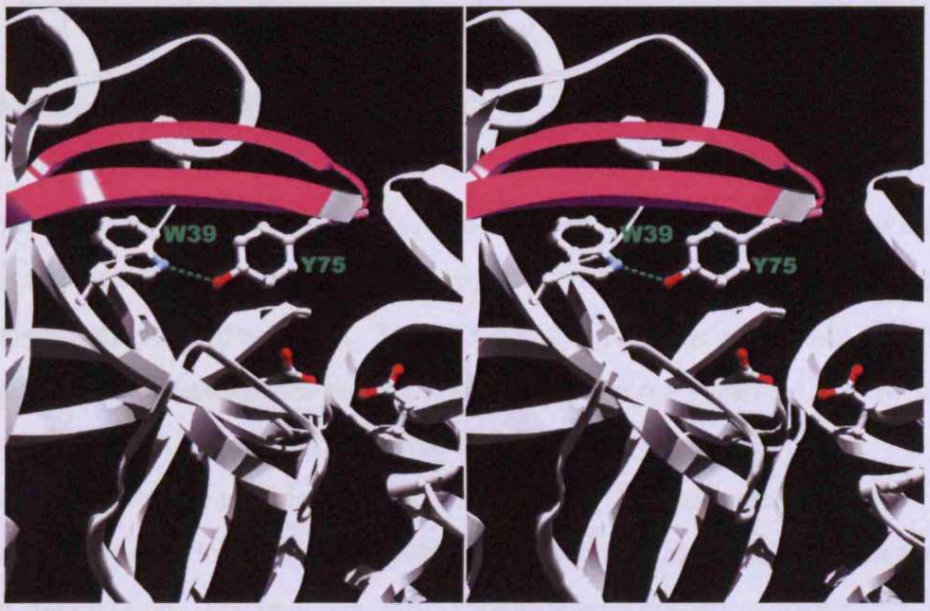


Figure 1.5 Stereo representation of the interaction between Tyr75 on the tip of the "flap" with Trp39 in *S. cerevisiae* vacuolar proteinase A (PDB code 2JXR). The flap (pink) acts as a lid over the active site cleft. The two catalytic aspartates are at the base of the cleft. The probable hydrogen bond between Tyr75 on the flap and Trp39 is depicted with a dotted green line.

Andreeva & Rumsh, 2001) and assisting in tightening the contacts between substrate and enzyme prior to catalysis (Wolfenden, 1999; Schramm, 2005).

Another structural feature, known as the "polyproline loop", can also be observed. In Figure 1.3A, showing the structure of *S. cerevisiae* proteinase A, this rather rigid reverse turn (type IV, according to the classification rules of Richardson, 1981) makes for quite an open active site cleft. In contrast, in other enzymes such as mouse and human rennin, the loop bends down over the active site cleft in an analogous manner to the "flap" (Aguilar *et al.*, 1997). In both scenarios, the location and constituent residues of the polyproline loop can affect the substrate specificity of the enzyme (Section 1.6).

1.6 Substrate specificity

On the basis of a recent extensive study, Tyndall *et al.* (2005) concluded that all families of endoproteinases, including the aspartic proteinases, bind the substrate in an extended β-strand conformation. Substrates bind within the active site of aspartic proteinases utilising interactions between each residue in the substrate and the amino acids that create individual subsites or pockets within the enzyme (Dunn & Hung, 2000). Inhibitors establish similar interactions. The nomenclature of Schechter & Berger (1967) is used to describe and identify these positions in the substrate and enzyme. The residues of the substrate are numbered sequentially from the scissile peptide bond (indicated by the asterisk).

$$H_2N\text{-}\mathbf{P_5}\text{-}\mathbf{P_4}\text{-}\mathbf{P_3}\text{-}\mathbf{P_2}\text{-}\mathbf{P_1}\text{*}\mathbf{P_1'}\text{-}\mathbf{P_2'}\text{-}\mathbf{P_3'}\text{-}\mathbf{P_4'}\text{-COOH}$$

The individual subsites in the active site cleft are identified similarly (S_4 to S_4'), with each subsite accommodating one residue of the peptide substrate (Figure 1.6). The spatial arrangement and the characteristics of the individual amino acids that contribute to each subsite determine the preference and, hence, the specificity of any given proteinase towards substrates or inhibitors. The summation of a complex set of interactions across all the subsites of an enzyme governs efficient binding and the acceptance and orientation of a substrate (Wolfenden, 1999). Subtle differences at

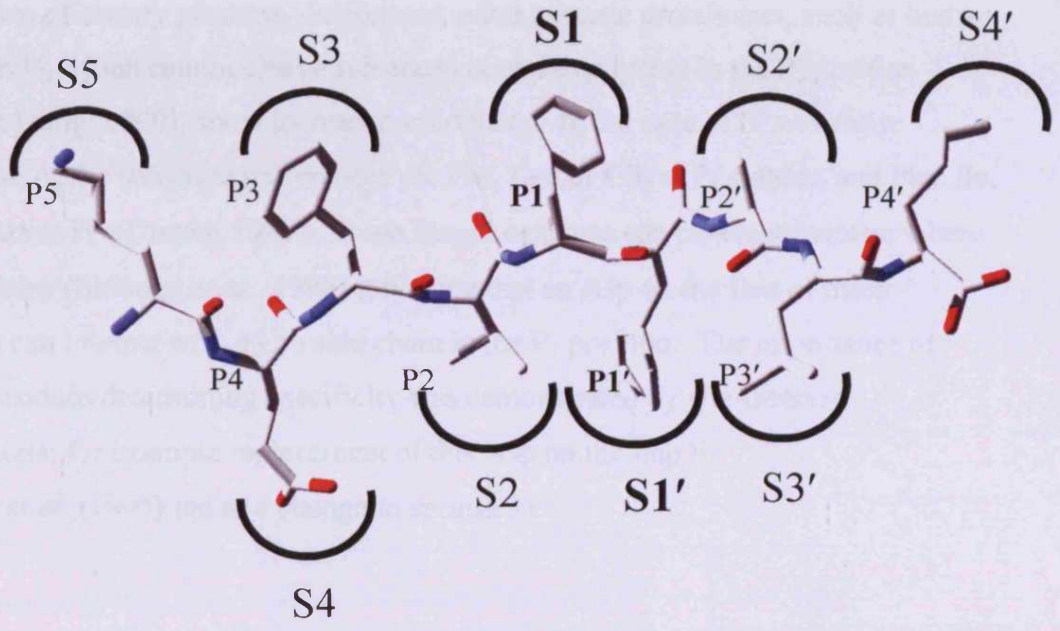


Figure 1.6 Schematic representation of the subsites S5-S4′ within the active site cleft of an aspartic proteinase. A substrate of sequence KEFVFFALK is shown. Nomenclature is that of Schechter & Berger (1967).

these subsites dictate the range of substrates that will be hydrolysed, and thus the wide range of physiological roles for this class of enzymes. At an enzyme kinetic level, these subtleties of structural complementarity between substrate/inhibitor and enzyme will be reflected in adjustments of K_m values or, in the case of inhibitors, K_i values.

Most eukaryotic aspartic proteinases have a primary specificity for hydrophobic (often aromatic) residues in the P_1 and P_1' subsites. Beyond these positions pepsin, for example, has a broad substrate preference, which possibly relates to its role in the breakdown of dietary proteins. In contrast, other aspartic proteinases, such as human cathepsin D, which cannot cleave substrates containing lysine in the P_2 position (Dunn & Hung, 2000), show increased specificity. In the case of *S. cerevisiae* proteinase A, the substrate preferences are Phe, Leu or Glu at P_1 subsite, and Phe, Ile, Leu or Ala at P_1' (Dreyer, 1989). Some fungal enzymes can cleave substrates where P_1 is a lysine (Shintani *et al.*, 1996) reflecting that an Asp on the flap of these enzymes can interact with a Lys side chain in the P_1 position. The importance of subsite residues determining specificity was demonstrated by site-directed mutagenesis; for example replacement of this Asp on the flap by Lowther *et al.* (1995) led to a change in specificity.

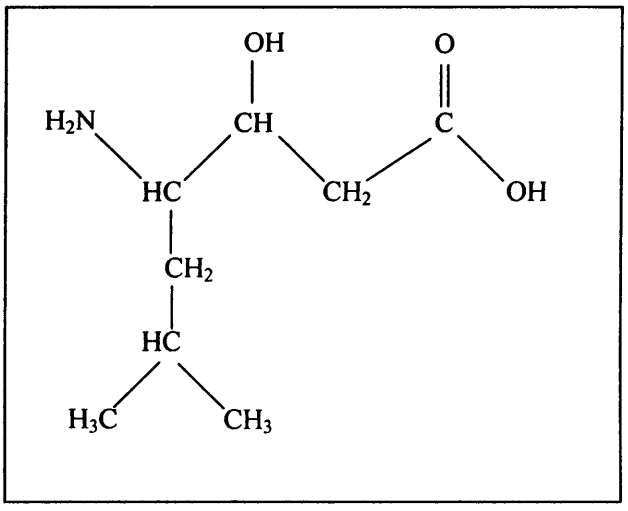
1.7 Inhibition of aspartic proteinases

As enzymes critically involved in many physiological processes and with proven roles in the pathogenesis of numerous diseases, aspartic proteinases are very attractive therapeutic drug targets (Cooper, 2002). Furthermore, many pathogenic fungi, microbial parasites and viruses rely on aspartic proteinases for their infectivity. Thus, inhibitors of aspartic proteinases are not only required to treat human conditions such as amyloidoses, peptic ulcers, cancer and hypertension, but they can also control human infections (such as HIV, malaria and thrush), protect commercial livestock and poultry, and prevent disease in cash-crops such as wheat, potato, rice and tobacco.

Aspartic proteinase inhibitors fall into two broad categories, small molecule inhibitors and proteins. These will be described in turn.

1.7.1 Small molecule inhibitors

In contrast to the widespread phenomenon of intrinsic inhibition of aspartic proteinases by the prosegment (Section 1.2), the aspartic proteinase family of enzymes has very few naturally-occurring *extrinsic* inhibitors (Kay, 1985). However, most of these enzymes are highly susceptible to inhibition by pepstatins, a series of non-proteinaceous natural products. Pepstatins are antibiotics, isolated from various species of *Actinomyces* (Gram-positive bacteria). The shared chemical feature is that they contain two residues of the unusual amino acid statine ((4s 3s)-4-amino-3-hydroxyl-6-methyl heptanoic acid).



Pepstatin occurs predominantly in three forms: isovaleryl-pepstatin (isovaleryl-Val-Val-Sta-Ala-Sta), isobutyryl-pepstatin (isobutyryl-Val-Val-Sta-Ala-Sta) and acetyl-pepstatin (acetyl-Val-Val-Sta-Ala-Sta). Other artifical forms that differ in the nature of the acyl group have been described. A more water soluble derivative, lactoyl-pepstatin (lactoyl-Val-Sta-Ala-Sta), was described by Kay (1982). Individual aspartic proteinases display different degrees of susceptibility to inhibition by pepstatins (Valler & Kay, 1985). Mammalian enzymes are usually equally sensitive to inhibition by acetyl- and isovaleryl-pepstatins, whereas acetyl-pepstatin is less potent against the fungal enzymes but more effective against HIV proteinase (Richards *et al.*, 1990).

Many crystal structures are available of aspartic proteinases complexed with isovaleryl-pepstatin. In each case, the backbone of the inhibitor adopts an extended

β-strand conformation, resembling the substrate and lying along the active site cleft of the enzyme. This positions the hydroxyl oxygen of the first statine residue between the carboxyl groups of the two catalytic aspartic acid residues, Asp32 and Asp215 (James *et al.*, 1982). Thus, this hydroxyl group replaces *and displaces* the catalytic water molecule located at this position within the native, free enzyme (compare Figures 1.7A and B). In this respect, the inhibitory mode is notably analagous to that of the prosegment in some zymogens (Section 1.2). For example, pepsinogen locates the ε-NH_3^+ group of a Lys sidechain in a similar position to achieve the same displacement of the catalytic water molecule (Khan & James, 1998; Kageyama, 2002). The bulky statine residue which occupies both the P1*P1′ positions, may act as a non-hydrolysable dipeptide mimetic, possibly with a structure closely reflecting the conformation of the transition state intermediate formed during normal peptide bond hydrolysis (Marciniszyn *et al.*, 1976; Leung *et al.*, 2000). Certainly, this would explain the potency of the interaction normally displayed.

The understanding gained from the studies with pepstatins, together with similar investigations of other enzyme systems, has informed the design of chemically synthesised small molecule inhibitors. The rationale underlying these strategies typically takes one of three forms:

(i) non-hydrolysable substrate analogues;
(ii) non-hydrolysable transition state analogues;
(iii) substrate variants.

It is considerably more difficult to identify, or to predict with confidence, the structure of a substrate at the transition state ($S^{\ddagger}$). This is a reflection of the intrinsically transient nature of this state (Schramm, 2005). Nevertheless, there are considerable benefits to be gained from designing inhibitors as transition state, rather than substrate (S), analogues (Wolfenden, 1999). For example, substrates in the ground state are normally bound to an enzyme quite weakly (K_m typically 10^{-3} to 10^{-6} M). In contrast, the transition state must be bound much more tightly by the enzyme, because this is the most unstable species on the reaction pathway, and classic dogma dictates that enzymes enhance the rate of reaction by lowering the equilibrium constant for

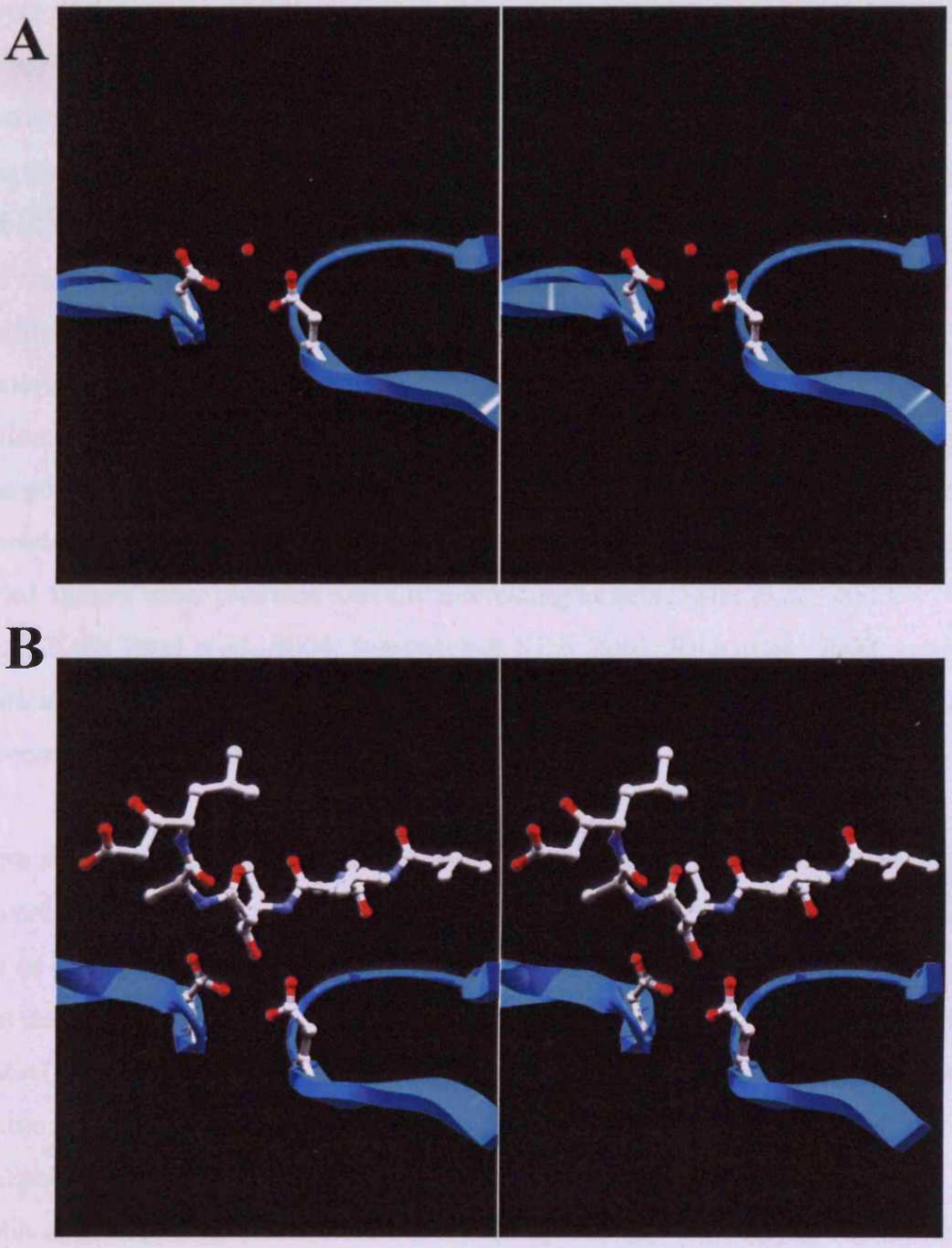


Figure 1.7A Stereo representation of plasmepsin II coordinating the catalytic water molecule. The catalytic aspartic acid residues (CPK) and water (red sphere) are shown. (PDB code 1LF4) **B. Stereo image of isovaleryl pepstatin binding to the active site of plasmepsin II.** (PDB code 1XDH) Pepstatin (shown in CPK) binding displaces the catalytic water molecule and juxtaposes the hydroxyl group of the statine moiety to the carboxyl groups of the two catalytic Asp residues. The orientation in **A** and **B** is the same.

attainment of the transition state (Pauling, 1946). This difference between uncatalysed (red) and enzyme-catalysed (blue) reactions is depicted as a reaction coordinate diagram (Figure 1.8A). The energy difference between $[S]^{\ddagger}$ and $[ES]^{\ddagger}$ is equivalent to the rate enhancement of the enzyme reaction, which typically is 10^{10} - 10^{15} for enzyme-catalysed reactions (Wolfenden, 1999). A non-hydrolysable substrate analogue would follow a similar reaction coordinate to the normal enzyme-catalysed hydrolysis, except that no cleavage of the analogue would occur. In comparison, a non-hydrolysable transition state analogue ($I^{\ddagger}$, green trace) is envisaged to pass through an initial $E \bullet I^{\ddagger}$ association complex before establishing a stable tightly bound $EI^{\ddagger}$ complex at a lower free energy level. Furthermore, whilst non-hydrolysable substrate analogues can be reasonably expected to inhibit any enzyme with which that substrate interacts, transition states are considerably more characteristic of an individual enzyme (Wolfenden, 1999; Hiratake, 2005). Thus, transition state analogues offer the opportunity for superior inhibitory specificity as well as potency. Several of the currently-available HIV proteinase inhibitors used as therapeutic drugs work on the transition state principle, as do several other inhibitors designed against other pharmaceutically interesting targets (Esler *et al.*, 2000; Bi *et al.*, 2000; Patel *et al.*, 2004; Bursavich & Rich, 2002; Rich *et al.*, 2002; Ersmark *et al.*, 2004). The design strategies for inhibitors against aspartic proteinases have been reviewed (Dash *et al.*, 2003).

For both substrate and transition state analogues, the key to their action is that they are non-hydrolysable. That is, the scissile bond has been replaced by an alternative that cannot be cleaved readily by the enzyme. A variety of methods have been employed for this task, such as the introduction of statine groups, like those found naturally in pepstatin (see above), or other substitutions/modifications to the constituent atoms of the amide (peptide) bonds of the molecule ("pseudopeptides") so as to render them uncleavable. In all cases, the catalytic water is also displaced, as seen with the pepstatin structures (Figure 1.7). However, the field of non-peptide peptidomimetics is arguably the most rapidly advancing (Rich *et al.*, 2002). Here, lead compound inhibitor "shapes" are identified, often by the use of a large library of peptide-based inhibitors, because this removes many of the difficulties associated with chiral centres in chemical syntheses. Subsequently, non-hydrolysable chemically synthesised

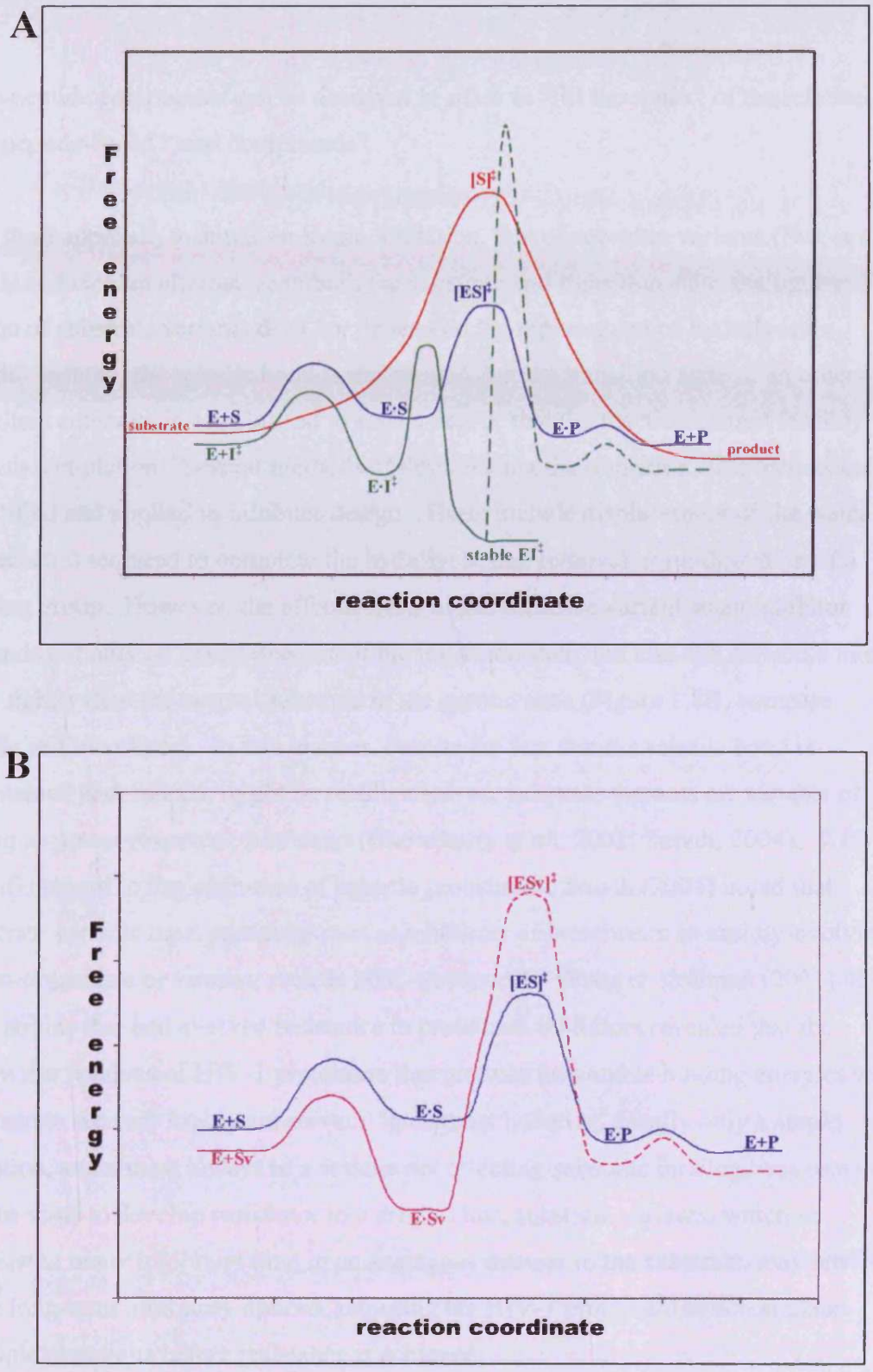


Figure 1.8A Schematic representation of the free energy profiles of uncatalysed (red) and enzyme (E)-catalysed (blue) reactions of a substrate (S) to form product (P). The green line shows the possible profile of a non-hydrolysable transition state analogue inhibitor, with the dotted line depicting the completion of the reaction, in the event that cleavage of the non-scissile bond could be achieved. **B. Comparison of the free energy profile of an enzyme catalysed reaction (blue) with that of the reaction with a substrate variant (Sv, purple).** The dotted line represents the theoretical completion of the reaction. Figures adapted from Schramm (2005) and Smyth (2004) respectively.

non-peptide compounds can be designed *in silico* to "fill the space" of the relatively few peptide-based "lead compounds".

The third approach to small molecule inhibition, that of substrate variants (Sv), is a subtle and elegant alternative to both the substrate and transition state analogues. The action of substrate variants does *not* depend on the replacement of hydrolysable bonds. Instead, the scissile bond is maintained, but the transition state of an otherwise excellent substrate is destabilised to such a degree that the reaction cannot feasibly pass to completion. Several methods of destabilising the transition state have been identified and applied to inhibitor design. These include displacement of the water molecule(s) required to complete the hydrolysis, and removal or modification of a leaving group. However, the effectiveness of the substrate variant as an inhibitor depends not only on destabilisation of the transition state, but also the molecule must bind tighter than the normal substrate in the ground state (Figure 1.8B, compare purple and blue lines). In this manner, despite the fact that the scissile bond is maintained and, indeed, might be readily cleaved, substrate variants are capable of acting as potent reversible inhibitors (Chowdhury *et al.*, 2002; Smyth, 2004). With specific regard to the inhibition of aspartic proteinases, Smyth (2004) noted that substrate variants have particular uses as inhibitors of proteinases in rapidly-evolving micro-organisms or viruses, such as HIV. A report by Wang & Kollman (2001) of HIV strains that had evolved resistance to proteinase inhibitors revealed that the active site residues of HIV-1 proteinase that produce favourable binding energies with a substrate are very highly conserved. In contrast however, usually only a single mutation, and almost always to a residue not affecting substrate binding, was required for the virus to develop resistance to a drug. Thus, substrate variants, which in contrast to many inhibitors bind in an analagous manner to the substrate, may provide more long-term inhibitory options, requiring the HIV-1 proteinase to accumulate multiple mutations before resistance is achieved.

1.7.2 Naturally-occurring gene-encoded aspartic proteinase inhibitors

In contrast to the three other principal classes of proteinase (serine, cysteine and metallo, Section 1.1), only a few naturally-occurring polypeptide inhibitors of aspartic proteinases have been reported (Phylip *et al.*, 2001; Rawlings *et al.*, 2004b). These include the mammalian kidney enzyme N-acetyl-D-glucosamine-2-epimerase, also known as renin-binding protein, which moonlights as an inhibitor of renin (Takahashi *et al.*, 1999), and a potent cathepsin D inhibitor, known as equistatin (Lenarcic & Turk, 1999), from *Actinia equina* (sea anenome). Notably, equistatin was first reported as subnanomolar inhibitor of the papain family of cysteine proteinases (Lenarcic *et al.*, 1997). However, subsequently it was discovered that the second of the three thyroglobulin type-1 domains in equistatin harbours subnanomolar inhibitory potency against the aspartic proteinase cathepsin D (Lenarcic & Turk, 1999; Strukelj *et al.*, 2000), although it did not inhibit either HIV proteinase or pepsin.

An 11 amino acid aspartic proteinase inhibitor called ATBI has been reported (Dash *et al.*, 2001a and 2001b) in prokaryotes, a noteworthy discovery in view of the fact that aspartic proteinases themselves appear to be very rare in prokaryotes (Section 1.1.5). ATBI, from the extremophilic bacteria *Bacillus sp.*, inhibited pepsin within subnanomolar potency, as well as an aspartic proteinase from the fungus *Aspergillus satoi* to a lesser degree.

Rather more is known about another pepsin-inhibiting protein, PI-3, from the nematode parasite, *Ascaris*. This was originally described by Abu Erreish & Peanasky (1974) in *Ascaris lumbricoides* and subsequently was sequenced (Martzen *et al.*, 1990) and cloned (Kageyama, 1998). A structure for the inhibitory complex of the 149-residue PI-3 with pig pepsin was provided by Ng *et al.*, (2000) (Figure 1.9). This inhibitor, which also inhibits gastricsin and cathepsin E, displays a novel protein fold and inhibits pepsin by partial occlusion of the active site. In particular, three important contact points were identified. The three N-terminal residues of PI-3 occupy the prime side of the active site cleft (sites P3′ to P1′) and the residues thereafter contribute to a multi-stranded β-sheet spanning both enzyme and

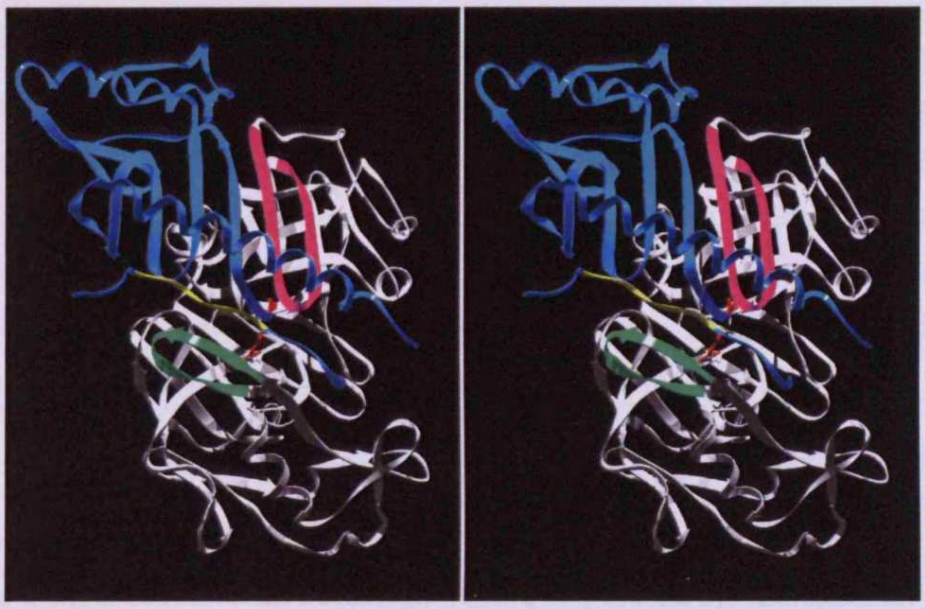


Figure 1.9 Stereo representation of the inhibitor PI-3 (blue) from *Ascaris* bound to pepsin (white) (PDB code 1F34). The N-terminus (dark blue) of PI-3 occupies the first 3 pockets to the prime side of the two catalytic aspartates (CPK) at the base of the active site cleft. PI-3 forms a multi-stranded β-sheet with the flap (pink) of pepsin. Four consecutive Pro residues in the C-terminal tail of PI-3 (yellow) interact with the polyproline loop (green).

inhibitor (blue) and involving the enzyme "flap" (pink) (Section 1.5). Over this expansive sheet, the inhibitor lays a long α-helix, although this does not directly occlude the cleft. Finally, the C-terminus of the inhibitor (yellow) interacts with the polyproline loop (in green, see also Section 1.5) of the enzyme.

Subsequently, related inhibitors have been discovered in other species of nematodes, including *Parelaphostrongylus tenuis* (Duffy *et al.*, 2002), *Trichostrongylus colubriformus* (Shaw *et al.*, 2003), *Ancylostoma* hookworms (Delaney *et al.*, 2005) and *Osteragia osteragi* (De Maere *et al.*, 2005). This prevalence of inhibitors has led Shaw *et al.* (2003) and Delaney *et al.* (2005) to adopt the name "Aspin" to describe these aspartic proteinase inhibitors unique to nematodes.

Other gene-encoded inhibitors of pepsin have also been reported, including at least two from plants. Galleschi *et al.* (1997) described a protein isolated from soft wheat bran, and Christeller *et al.* (1998) reported a 10 kDa proteinaceous inhibitor of pepsin from the phloem exudate of squash. The latter group subsequently characterised this inhibitor, termed SQAPI, in more detail. It inhibited pepsin, as well as one isoform (Sap4) of the *Candida albicans* secreted aspartic proteinases (Section 1.1.5) and the *Glomerella cingulata* secreted aspartic proteinase (Farley *et al.*, 2002), all with low nanomolar K_i values.

Various other aspartic proteinase inhibitors have been isolated from plants. For example, Roszkowska-Jakimiec & Lesniewska (2004) reported an inhibitor of cathepsin D in the seed coat of *Vicia sativa* (common vetch); and several isoforms of API-8, a 21 kDa cathepsin D inhibitor in *Solanum tuberosum* (potato), have been studied over the past three decades (Keilová & Tomášek, 1976; Ritonja *et al.*, 1990; Werner *et al.*, 1993). More recently, Cater *et al.* (2002) discovered that API-8, and the orthologue (>80% sequence identity) from *Solanum lycopersicum* (tomato), were in fact >20-fold more potent against the vacuolar proteinase A from *S. cerevisiae* (Section 1.1.5 and 1.8.2). Preliminary attempts to utilise these inhibitors as protective agents in transgenic crops have been reported (Brunelle *et al.*, 2004).

No structural data regarding the mode of action employed by API-8 against the *S. cerevisiae* vacuolar proteinase A is yet available. However, extensive work conducted within our laboratory, in collaboration with others, led to the elucidation of the crystal structure of another, much more potent, inhibitor in an inhibitory complex with this enzyme. The gene-encoded inhibitor, known as IA3, served as the basis for the studies described in this thesis, and is described in the next section.

1.8 IA3: A potent polypeptide inhibitor of *S. cerevisiae* vacuolar proteinase A

1.8.1 History of IA3

IA3 is a heat-stable 68 amino acid cytosolic protein from the yeast *S. cerevisiae*. First reported by Saheki *et al.* (1974), IA3 is so named because during a chromatographic purification step it was the third of 4 protein peaks that inhibited the vacuolar aspartic proteinase A from the same yeast. Subsequent re-examination of these activities (Meußdoerffer, 1980) suggested that the proteins contained within two of the other peaks, termed IA1 and IA4, were most probably artefacts resulting from various elements of the original purification process. However, IA2 and IA3 appear to be genuinely distinct, based on their differing isoelectric points (5.7 and 6.3, respectively). Whereas wild-type *S. cerevisiae* contained only IA3, the closely related brewer's strain *S. carlsbergensis* contained only IA2. In contrast, the strain of commercial baker's yeast (Plese Hefe) used in the original studies (Saheki *et al.*, 1974) contained both IA2 and IA3 forms, possibly because the preparation contained both strains of yeast. Thus, Meußdoerffer (1980) concluded that IA2 and IA3 are, in fact, different isoforms of the same protein. The sequence of *S. cerevisiae* IA3 (see Appendix I) was determined by Biedermann *et al.*, (1980) and the gene was isolated and cloned by Schu & Wolf (1991), but no sequence for IA2 exists in the public record.

1.8.2 The proteinases of the yeast vacuole and the role of their inhibitors

The vacuole of yeast is an acidic compartment containing multiple hydrolases (Schu & Wolf, 1991; van den Hazel *et al.*, 1996). Two endopeptidases, proteinases A and B, and also carboxypeptidase Y (originally termed "proteinase C", for example see Saheki *et al.*, 1974) have been widely reported in yeast vacuoles. As described above, proteinase A, which is an aspartic proteinase, appears to be the natural target of IA3. In contrast, proteinase B which is a subtilisin-type serine proteinase, itself also has a high-affinity, heat-stable, relatively short (74 residue) cytosolic polypeptide inhibitor known as IB2 (or IB1, again depending on the yeast strain). Much like IA3, this serine proteinase inhibitor also exists in different isoforms, and exhibits more potent inhibition at acidic pH values (Schu *et al.*, 1991). Furthermore, carboxypeptidase Y has its own cytosolic high affinity specific inhibitor, Ic, although this protein is considerably larger, 25 kDa (Bruun *et al.*, 1998). Proteinases A and B are essential for protein degradation during vegetative growth. In particular, proteinase A was shown to be critical to cell survival under starvation conditions (Teichert *et al.*, 1989). Furthermore, these two proteinases, together with carboxypeptidase Y, are necessary for sporulation.

The expression levels of proteinases A and B, and the levels of their respective inhibitors, appear to be coordinated, with compensatory increased levels of IA3 and IB2 expressed during the cell cycle or when the proteinase genes are upregulated (Schu *et al.*, 1991; Schu & Wolf, 1991). These results are intriguing because the inhibitors and enzymes occupy distinct compartments of the yeast cell and yet it suggests that they may interact with each other under "normal" cell conditions. It has been postulated that the inhibitors protect the cell against unwanted activities of these enzymes after accidental leakage during the many observed fragmentation and fusion processes of the vacuole (Holzer & Saheki, 1976). However, "knock-out" yeast with the genes for both IA3 and IB2 inactivated by homologous recombination did not demonstrate reduced viability, even under cell stress conditions such as starvation or heat-shock (Schu *et al.*, 1991; Schu & Wolf, 1991). This was despite a 20-50% increase in protein degradation *in vivo*. A more probable function is perhaps to regulate protein catabolism during autophagocytosis (Bruun *et al.*, 1998). Thus, the precise physiological role of the inhibitors remains something of an enigma.

However, IA3 has considerable commercial importance and has achieved a high level of personal relevance to the author of this thesis. The proteolytic activity of proteinase A damages the stability of beer foam and thus IA3, as the natural inhibitor of this enzyme, regulates the size of "head" on good beer (Dreyer *et al.*, 1983; Yokoi *et al.*, 1996; Wang *et al.*, 2005).

Proteinase A (EC 3.4.23.25; MEROPS classification: clan AA, family A1, peptidase A01.018), also known as saccharopepsin, is a classic example of an aspartic proteinase (Winther *et al.*, 2004). The enzyme, which is the product of the PEP4 gene, is synthesised as a 405 residue precursor that includes an N-terminal signal peptide (22 residues) and a propart (54 residues) (Ammerer *et al.*, 1986). Following processing in the endoplasmic reticulum the zymogen is transported via the Golgi, where the two Asn-X-Ser/Thr signal sequences are N-glycosylated (Asn67 and Asn267) with high-mannose type units, to the vacuole (Winther *et al.*, 2004). From a study of a wide range of mutants with substitutions in the propart, van den Hazel *et al.* (1995) noted that the transit time to the Golgi from translation on the endoplasmic reticulum was retarded, from which they concluded that the propart has a role in assisting the folding of the maturing proteinase.

In the vacuole, the zymogen autoactivates in a process that is accelerated under the acidic conditions in this compartment (Sorensen *et al.*, 1994; van den Hazel *et al.*, 1997). The mature, activated 329 residue (42 kDa) proteinase A in turn initiates the activation of the other vacuolar zymogens, including pro-proteinase B and pro-carboxypeptidase Y (Klionsky *et al.*, 1990; Rupp *et al.*, 1991; Sorensen *et al.*, 1994).

1.8.3 IA3 is an exquisitely selective inhibitor

Considering the role of *S. cerevisiae* proteinase A as one of the major *non-specific* degradative enzymes in yeast cells (Schu & Wolf, 1991), the exclusive nature of the inhibitory action exhibited by IA3 towards this enzyme is remarkable. IA3 showed no detectable activity against a wide range of aspartic proteinases, including several others from yeast and closely-related enzymes from the animal kingdom such as pepsin, gastricsin and cathepsin D (Dreyer *et al.*, 1985; Phylip *et al.*, 2001). Indeed,

in all instances, IA3 was shown to be cleaved as a substrate of these non-target enzymes (Phylip *et al.*, 2001).

1.8.4 The mode of inhibition employed by IA3 is apparently unique

For some time, the nature of the inhibitory interaction of IA3 with *S. cerevisiae* proteinase A was a mystery. Reports as early as the 1970s had suggested that IA3 was unstructured in free solution at a pH value of 6.9 (Núñez de Castro & Holzer, 1976). These findings were later confirmed by Li *et al.* (2000) at pH 6.0 and Green *et al.* (2004) who, using a range of spectroscopic techniques, described in detail the unstructured nature of this 68-residue inhibitor at a variety of weakly acidic pH values. The question as to how an unstructured protein might give rise to such potent (K_i = 7 nM at pH 3.1, 37 °C) inhibition (Dreyer *et al.*, 1985) was unclear.

However, the elucidation of the crystal structure of IA3 complexed with *S. cerevisiae* vacuolar proteinase A (Li *et al.*, 2000) at 2.2 Å resolution revealed an unprecedented mode of inhibition and one that often attracts comment as an exception within the general considerations of naturally-occurring proteinase inhibitors (Tyndall *et al.*, 2005). Although the inhibitor is unstructured in free solution, upon contact with the proteinase, residues 2-32 which have the sequence

```
1                 10                  20                  30
(M)NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KK
```

form a near-perfect α-helix (gold) along the active site cleft of the enzyme (Figure 1.10A). No other inhibitor of an aspartic proteinase has been shown to function by occluding the enzyme active site with an α-helix (Tyndall *et al.*, 2005). Furthermore, in the process, the flap (Section 1.5) is displaced by 8-9 Å from the observed position (shown in pink in Figure 1.10B) in other structures of the same enzyme containing small molecule inhibitors. Indeed, normal considerations of modes of inhibition for high affinity inhibitors suggest that such molecules are typically enclosed by the flap more tightly than the substrate, and do not leave the "lid" ajar.

A

B

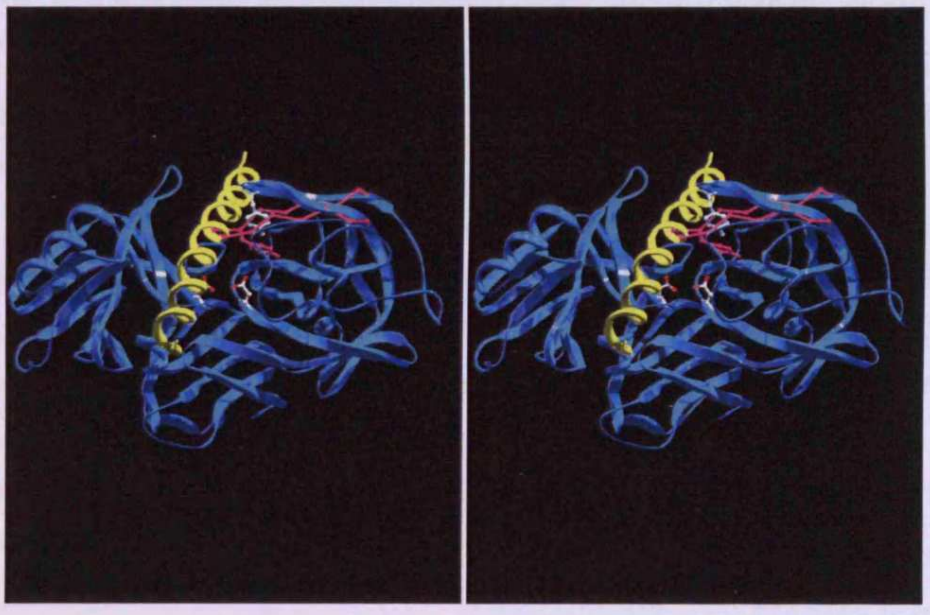


Figure 1.10A Inhibition of *S. cerevisiae* vacuolar proteinase A (blue) by IA3 (gold). Residues 2-32 of IA3 form a near-perfect α-helix in the active site cleft of the proteinase (PDB code 1DPJ). **B. Displacement of the flap region by IA3.** By occluding the active site cleft, IA3 displaces the flap region of the proteinase by up to 9 Å (stereo view). The structure of the flap with Tyr75 for the complex of the same enzyme with a small-molecule inhibitor is shown in pink (PDB code 2JXR) superimposed upon the IA3 complex structure (PDB code 1DPJ).

Notably, no residues of IA3 following Lys32 in the polypeptide sequence were apparent in the X-ray structure of the IA3-proteinase complex. This may explain the initial findings of Biedermann *et al.* (1980), that the N-terminal half of IA3 was the smallest peptide fragment of IA3 to show inhibitory activity against the proteinase. It would appear that the C-terminal half of IA3 either remains unstructured, or does not form a structure of sufficient regularity to produce an interpretable diffraction pattern.

1.8.5 Hydrophobic contacts make significant contributions to binding

The amphipathic nature of the inhibitory IA3 helix (Figure 1.11) was apparent from the X-ray structure, with almost every residue on the enzyme-enclosed face of the helix being hydrophobic, and charged or polar groups facing the solvent. It was suggested, therefore, that the enthalpic contributions to the interaction were generated predominantly by hydrophobic contacts. This finding was in remarkable agreement with the work of Núñez de Castro & Holzer (1976) almost 25 years earlier, who had reached a similar conclusion by the use of hydrophobic bond-disrupting reagents such as urea, guanidine.HCl, Triton and cholate, to weaken the inhibitory interaction.

The crystal structure of the enzyme-inhibitor complex, together with the subsequent work of Phylip *et al.* (2001), revealed that three specific regions of the inhibitory helix make significant hydrophobic contacts with the enzyme. These are the so-called front, centre and back "pins".

The front pin, constituted by Val^{8}-X-X-hydrophobic-Phe^{12}, interacts with the S4 pocket (Section 1.6) of the enzyme (Figure 1.12, and see also Appendix I for IA3 sequence), whilst the centre pin of Leu19 points towards the S2 pocket deep in the active site cleft (Figure 1.13). The IA3 residues constituting the pins are shown in gold and the principal residues of proteinase A forming the pocket are depicted in CPK colours and labelled in green.

The residues of the back pin, Val^{25}-Val^{26}-X-X-hydrophobic-Phe^{30}, also make substantial hydrophobic contacts (Figure 1.14). Of particular note is the interaction of Val26 and Phe30 with the region of the proteinase known as the polyproline loop (see Section 1.5 and visualised in green in Figure 1.14).

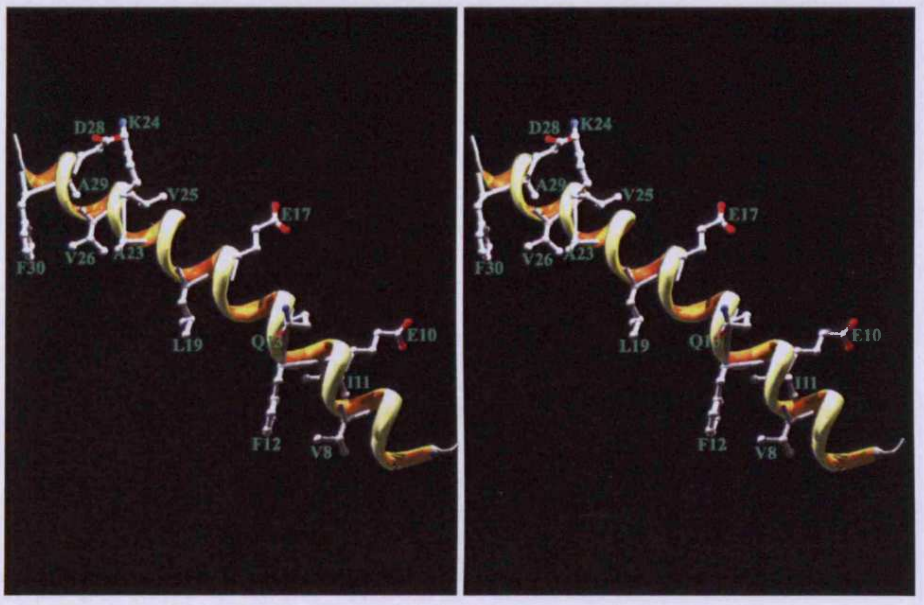


Figure 1.11 Stereo view of the amphipathic nature of the IA3 inhibitory helix (PDB code 1DPJ). Selected hydrophobic and hydrophilic residues are depicted, on opposite faces of the amphipathic helix. The solvent exposed face of the α-helix is on top and the hydrophobic surface interacting with the base of the cleft is below the helix.

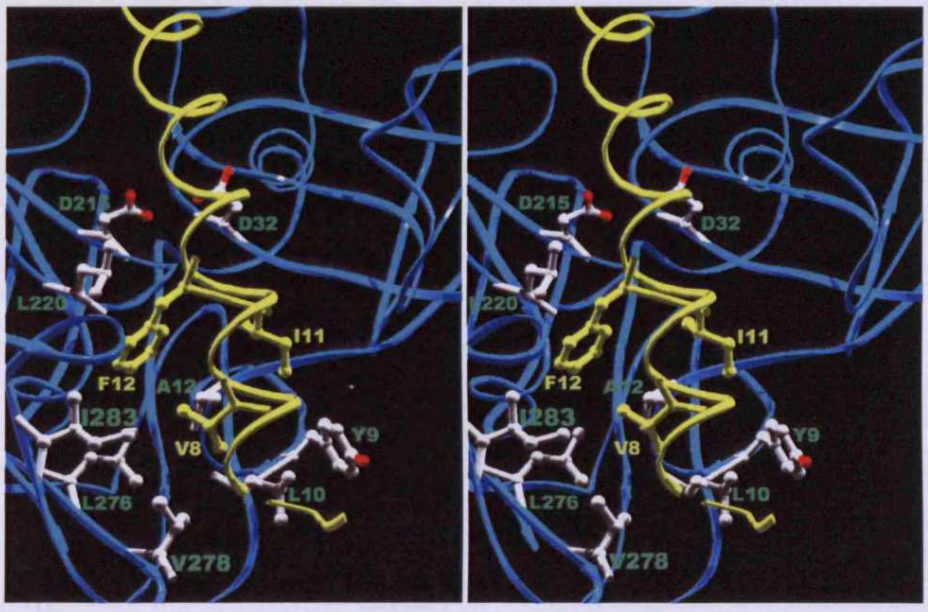


Figure 1.12 Stereo view of the interactions made by the front hydrophobic pin residues of IA3 with the active site of *S. cerevisiae* proteinase A. The IA3 helix and relevant residues (gold) are shown in the context of the proteinase A chain (blue). The principal residues (CPK) in proteinase A contributing to the front pin pocket are labelled (green). The catalytic aspartates (Asp32 and Asp215) are labelled for reference. (PDB code 1DPJ).

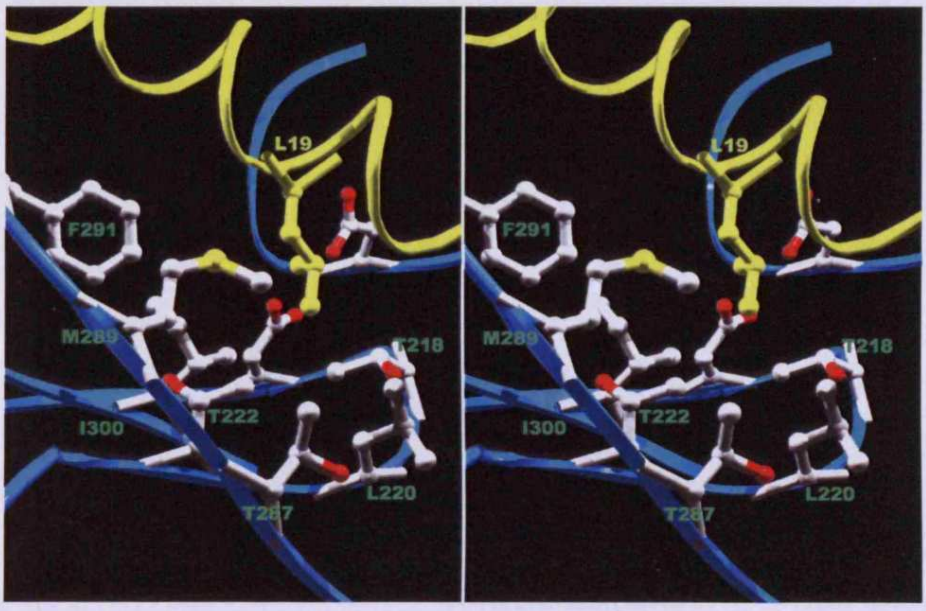


Figure 1.13 Stereo view of the interactions made by Leu19 of IA3 with *S. cerevisiae* vacuolar proteinase A. The IA3 helix (gold) including Leu19 (gold) is shown within the active site of proteinase A (blue). The proteinase residues (CPK) contributing to the centre pin pocket are labelled (green). The two unlabelled residues behind Leu19 are the catalytic aspartates (D32 and D215). (PDB code 1DPJ).

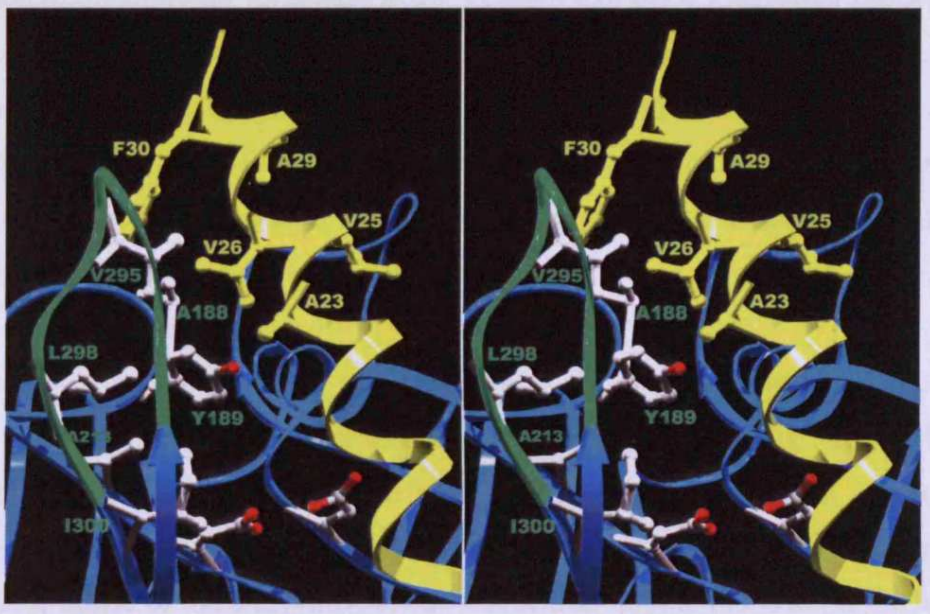


Figure 1.14 Stereo representation of the interactions of the back hydrophobic pin region of IA3 with *S cerevisiae* vacuolar proteinase A. The IA3 helix (gold) including back pin residues (gold) are shown within the active site of proteinase A (blue). The principal residues (CPK) in proteinase A contributing to the back pin pocket are labelled (green). The catalytic aspartates (Asp32 and Asp215) are labelled for reference. (PDB code 1DPJ).

The contributory residues of this feature of aspartic proteinases are particularly variable. Thus, it is possible that as well as providing binding energy, the interactions with the polyproline loop may also contribute to the exquisite *specificity* demonstrated by IA3 for this proteinase (Section 1.8.3).

1.8.6 The peptide backbone of IA3 is held clear of the catalytic Asp residues as a result of helix formation

Despite the cleavage of IA3 by non-target enzymes (Section 1.8.3), no digestion was detected of the inhibitor by *S. cerevisiae* vacuolar proteinase A, even upon prolonged incubation of the inhibitor for 3 days at 37 °C, pH 3.1 or pH 4.7, at molar ratios of inhibitor:proteinase as small as 5:1. Phylip *et al.* (2001) reported K_i values for a recombinant, C-terminally His-tagged form of IA3 against proteinase A at pH 4.7 of <0.1 nM (i.e. too tight binding for accurate determination on the available equipment) and 1.1 nM at pH 3.1. As well as emphasising the potency of the interaction, this also suggested a mild pH-dependence of the interaction.

IA3 avoids cleavage by the target enzyme as a result of the formation of the inhibitory α-helix, which holds the peptide backbone of IA3 clear of the catalytic jaws at the base of the active site cleft. No peptide bond in IA3 approaches within 5 Å of the catalytic aspartic residues. Remarkably, however, the water molecule required for the nucleophilic attack of the catalytic mechanism (Section 1.4) remains in precisely the same location, between the two catalytic Asp residues, as in the unliganded structure of Aguilar *et al.* (1997). This represents a unique distinction from all other reported inhibitors of aspartic proteinases, be they intrinsic pro-parts in zymogens (Section 1.2), pepstatins (Section 1.7.1), or peptidomimetic artificially-synthesised compounds (Section 1.7.1), all of which act by displacing the catalytic water molecule. Instead, the two residues of the IA3 "centrepiece" – Lys18 and Asp22 – form a complex series of interactions with both catalytic aspartic acid residues (Asp32 and Asp215) (Sections 1.3 and 1.4), the water molecule and also Tyr75 on the flap (see Figure 1.15).

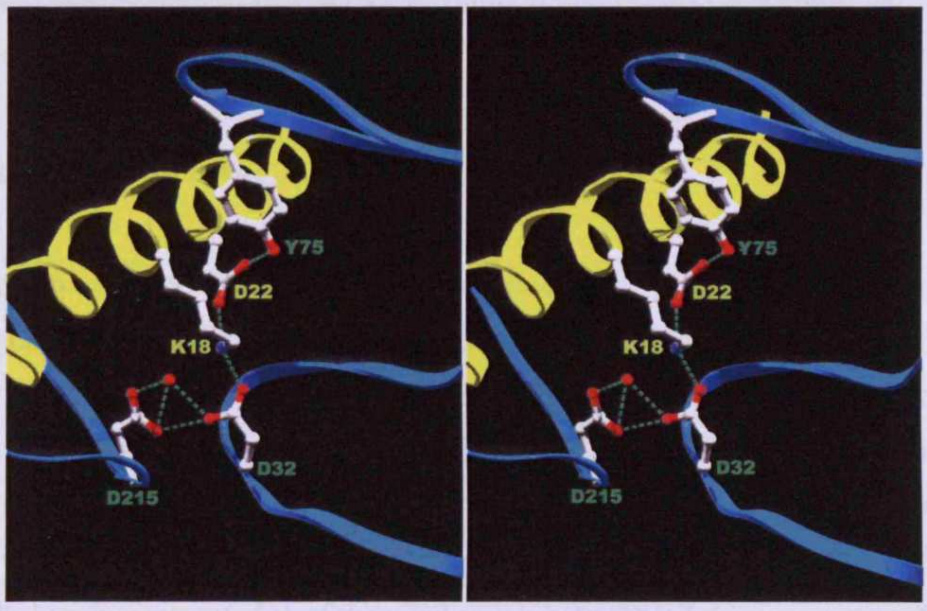


Figure 1.15 Stereo representation of the interactions made by the Lys18 and Asp22 centrepiece residues of IA3 with the catalytic apparatus of *S. cerevisiae* vacuolar proteinase A. The IA3 helix (gold) and the catalytic water (red sphere) coordinated by the two catalytic Asp residues (D32 and D215 of proteinase A) are shown within the active site of proteinase A (blue). Probable interactions are depicted with dotted lines (green). (PDB code 1DPJ).

The ε-NH_2 of Lys18 of IA3 makes a bond with the "outer oxygen" of Asp32 in the enzyme, as well as forming an (*i*, *i*+4) hydrogen bond with Asp22 on the adjacent turn of the inhibitory helix. In turn, the other carboxyl oxygen of Asp22 of IA3 makes a hydrogen bond with the phenolic hydroxyl group of Tyr75 on the enzyme flap overlaying the active site cleft. Thus, the location of these two charged residues on the otherwise hydrophobic face of the inhibitory helix appears precisely engineered to interact with the one region of the active site cleft that is more hydrophilic in character. The active sites of all the other enzymes tested (Dreyer *et al.*, 1985; Phylip *et al.*, 2001; Section 1.8.3) fail to provide a suitable folding template to induce the formation of the inhibitory helix. As a result, therefore, the peptide backbone binds as an extended β-strand and so is not held clear of the catalytic machinery of these non-target proteinases. The IA3 molecule is thus cleaved as a substrate.

1.9 Aims of the project

For several reasons, *S. cerevisiae* IA3 makes a fascinating research target. Even after the elucidation of the crystal structure of IA3 in complex with *S. cerevisiae* proteinase A (Li *et al.*, 2000) and the subsequent studies of Phylip *et al.* (2001), the precise origins of the exquisite specificity and potency of the interaction between inhibitor and enzyme were still not clear. The behaviour of IA3, specifically its transition from unstructured polypeptide in free solution to precisely-tailored α-helix in the enzyme active site, made interpretation of the relationship between primary sequence and function complex.

IA3 is not unique in displaying such a structural transition. An ever-increasing number of proteins involved in a wide range of cellular process are known similarly to be partially or fully unstructured in free solution (Schreiber & Serrano, 2005). Many of these polypeptides do assume structure, which is required for them to function, but only upon contacting the *correct* target. Thus, they mediate their effects in an extremely specific manner. Disordered regions of proteins are proposed to serve regulatory functions in transcription, cell cycle control, endocytosis and cell signalling (Dunker *et al.*, 2001; Tompa, 2002; Dyson & Wright, 2002; Fink, 2005). It appeared,

from the work preceding this thesis, that the list might be extended to include enzyme inhibition, with IA3 effective only when the "folding template" of the target proteinase active site directs the formation of the functional α-helix in IA3. Thus, information gained from the study of IA3 can be considered within this wider context of unstructured proteins as a whole.

However, to be an effective inhibitor, IA3 must not only be precisely compatible with its folding template. Once bound, it must make sufficiently strong interactions with the proteinase to remain in the active site cleft and, thus, inhibit the enzyme. Although closely related, these two requirements are distinct facets of IA3 – with the former linked to *specificity* (compatibility) and the latter to *potency* (binding strength). Mutations could affect either or both of these factors, but the limited body of evidence available made interpretation of some results difficult. The first aim of the project, therefore, was to improve our understanding of the interaction of IA3 with its natural target proteinase.

However, with just one enzyme target, distinguishing whether mutations affected specificity or potency was not possible. Therefore, the separation of potency from specificity required the identification of IA3 orthologues from other organisms or species, together with suitable new target proteinases. No reports of other IA3-like sequences can be found in publicly available scientific journals (November 2005). Nevertheless, the identification and characterisation of IA3-like sequences from new sources, and a comparison of these with *S. cerevisiae* IA3, would shed further light on the origins of potency and specificity within the IA3 sequence.

The most ambitious aim of the project was to re-direct the activity of variant forms of IA3 against new enzyme targets. The evident potential within IA3 to produce highly specific and potent inhibitory interactions of an apparently unique nature made the possibility of targeting more pharmaceutically- or commercially-relevant enzymes an extremely attractive but challenging proposition.

Two potential strategies for achieving retargeting were identified: **rational design**, the most well-established method of protein engineering, and **directed evolution**, the fastest-expanding technique.

Rational redesign would draw on the known structural data of the IA3-*S. cerevisiae* proteinase A complex. The increased understanding arising from the initial experiments with this natural target enzyme could be applied to the development of inhibitors for new targets. However, rational design is limited by the quality of the information available on which to base the design and is, by definition, dependent upon the capacity of the researcher to interpret and apply these data "rationally". Applying the available information to inhibitor design against new enzyme targets would have to be made on the assumption that the inhibitors would utilise the new "folding template" (i.e. the active site of the new target enzyme) in the same manner. Thus, rational design would also benefit from accurate structural details, such as X-ray crystallographic/NMR structures or homology models, of the active sites of the new target proteinases.

In recent years, directed (molecular) evolution has emerged as a major system for developing proteins with new characteristics, principally because it is not limited by our (lack of) understanding of protein structure, folding and chemistry, or the relationship of structure or sequence to function (Fersht, 1998; Chica *et al.*, 2005; Bloom *et al.*, 2005). Directed evolution bypasses these difficulties by reproducing fundamental elements of natural selection in a "test tube": diversity within a large pool combined with selection or screening to identify individuals with superior, desirable "phenotypic" characteristics. As a genotype-phenotype link is maintained, the selected or "fit" individuals can be identified and their genetic characteristics form the parental templates for the next generation. By careful control of the selection pressures applied to the pool, repeated iterative cycling in this manner "evolves" required traits or properties at the expense of undesirable ones, in an enormously accelerated timescale relative to Darwinian natural selection (Lehmann & Wyss, 2001). In contrast to rational design, therefore, directed evolution is limited by the size and quality of the library of variants and the development of a simple, effective and appropriate method of screening: it must always be remembered that YGWYSF – "you get what you select for" (Zhao & Arnold, 1997).

Based on the detailed information available from the crystal structure of the IA3-proteinase A complex, most work described in this thesis is based on rational design. Efforts, therefore, concentrated firstly on better understanding the origin of the specificity and potency of IA3, and the discovery and investigation of IA3-like molecules in other species. Subsequently, this wealth of information is applied to variant IA3 sequences with potent inhibitory activities against new, more pharmaceutically-relevant enzyme targets. In combination with this, the final aim of this work was to establish the initial framework of techniques for parallel investigations using the principles of directed evolution to select new IA3 sequences with novel inhibitory activities.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Bacterial cloning and expression vectors

The cloning plasmid pGEM®-T was purchased from Promega Ltd. (Southampton, UK). The *E. coli* expression plasmid pET22b was obtained from Novagen (AMS Biotechnology, Milton Keynes, UK). The variant of this vector, to enable attachment of an N-terminal Met-Lys-$(His)_7$-Met tag to the expressed protein if cloned into the *Nde*I site of the vector, was re-designed and constructed personally from the original material (Chapter 4). Plasmid pET22b containing the wild-type *S. cerevisiae* and *S. castellii* IA3 genes were gifts from Dr L. Phylip and Dr D. Wyatt (Cardiff School of Biosciences, Cardiff University, UK), respectively.

2.1.2 T7Select™ bacteriophage vector

The vector, 415-1b, and packaging extract were obtained from Novagen (AMS Biotechnology, Milton Keynes, UK).

2.1.3 *E. coli* host strains

All DNA manipulations and propagations were performed in DH5α (Pharmacia Biotechnology Ltd., Milton Keynes, UK). For expression studies, strains BL21(DE3)pLysS (Novagen, AMS Biotechnology, Milton Keynes, UK) or C41(DE3) pLysS (OverExpress™, Avidis SA, Saint-Beauzine, France) were used. The host strain for bacteriophage T7Select™ amplification was BL21 (Novagen, AMS Biotechnology, Milton Keynes, UK).

2.1.4 Media

Luria Bertani (LB) nutrient medium was prepared according to the manufacturer's instructions (Sigma Chemical Co., Poole, UK). Top agarose was prepared according to the protocol provided with the T7Select™ bacteriophage display system (Novagen, AMS Biotechnology, Milton Keynes, UK). SOC medium contained 20 g/l tryptone, 5 g/l yeast extract, 0.5 g/l NaCl, 10 mM $MgCl_2$, 2.5 mM KCl, 0.4 % (w/v) glucose. All media and constituents were sterilised either by autoclaving at 121 °C (975 kPa) for 20 min or by filtration through a 0.22 μm nitrocellulose filter (Millipore (UK) Ltd., Watford, UK) as appropriate. When required, antibiotics were used at the following concentrations: ampicillin 100 μg/ml, carbenicillin 50 μg/ml and chloramphenicol 34 μg/ml.

2.1.5 Antisera

The Ni-conjugated-alkaline phosphatase, which specifically recognizes hexa-histidine tags attached to proteins, was obtained from Qiagen Ltd. (Crawley, West Sussex, UK).

2.1.6 Molecular weight standards

DNA markers λ/*Hin*dIII and ϕX174/*Hae*III were purchased from Promega Ltd. (Southampton, UK). Prestained broad range molecular weight protein standards were purchased from New England Biolabs (Beverly, MA, USA).

DNA markers

λ/*Hin*dIII (bp)	ϕX174/*Hae*III (bp)
23,130	1,353
9,416	1,078
6,557	872
4,361	603
2,322	310
2,027	281
564	271
125	234
	194
	118
	72

Protein markers

MBP-β-galactosidase	175,000 Da
MBP-paramyosin	83,000 Da
Glutamic dehydrogenase	62,000 Da
Aldolase	47,500 Da
Triosephosphate isomerase	32,500 Da
β-Lactoglobulin A	25,000 Da
Lysozyme	16,500 Da
Aprotinin	6,500 Da

2.1.7 Oligonucleotides

Oligonucleotide primers used in standard and site directed mutagenic polymerase chain reactions were purchased from Operon Biotechnologies Inc. (Cologne, Germany). Primers used in double-stranded DNA sequencing (Section 2.2.11) from remote sites on plasmids pGEM®-T and pET22b were purchased from MWG-Biotech UK Ltd. (Milton Keynes, UK) and Operon Biotechnologies Inc., respectively. Bacteriophage DNA sequencing primers were purchased from Novagen (AMS Biotechnology, Milton Keynes, UK) or Operon Biotechnologies Inc.

2.1.8 Synthetic peptide substrates and inhibitors

Substrates and synthetic inhibitors were produced by Pepsyn Ltd. (University of Liverpool, Liverpool, UK) or Alta Bioscience (University of Birmingham, Birmingham, UK).

2.1.9 Aspartic proteinase enzymes

The vacuolar aspartic proteinases from *S. cerevisiae*, *Pichia pastoris* and *Aspergillus fumigatus* were kind gifts from Dr. J. Winther (Carlsberg Laboratory, Copenhagen, Denmark), Dr. L. Phylip and Dr. D. Wyatt (both Cardiff University), respectively. Human cathepsin D was provided by Dr. S. Gulnik (NCI, Frederick, MD, USA).

2.1.10 Other materials

Routine laboratory chemicals were purchased from Sigma Chemical Co., (Poole, UK), BDH Chemicals Ltd. (Poole, UK) and Fisher Scientific UK Ltd. (Loughborough, UK). Microfuge tubes (1.5 ml and 0.6 ml) were obtained from Anachem Ltd. (Beds., UK), Thermowell™ PCR tubes from Corning Inc. (Corning, NY, USA), and plastic 15 ml and 50 ml disposable tubes from Greiner Bio-One Ltd. (Stonehouse, Glos., UK). The suppliers of more specialised items are listed below:

Restriction endonucleases, DNA modifying enzymes and *Taq* DNA polymerase	Promega Ltd., Southampton, UK and New England Biolabs, Beverly, MA, USA.
Pfu Ultra™ DNA polymerase	Stratagene Europe, Amsterdam, The Netherlands
Wizard® *Plus* SV Miniprep DNA Purification System	Promega Ltd., Southampton, UK.
QIAquick™ Miniprep DNA Purification System	Qiagen Ltd., Crawley, UK
QIAquick™ gel extraction kit	Qiagen Ltd., Crawley, UK
Acrylamide/bisacrylamide	National Diagnostics, Manville, NJ, USA
Bacto-tryptone, yeast extract, bacto-agar	Difco Laboratories Ltd., Detroit, USA and Oxoid Ltd., Hants, UK
Agarose (electrophoresis-grade)	Bioline Ltd., London, UK.
Quikchange™ Site Direct Mutagenesis Kit	Stratagene Europe, Amsterdam, The Netherlands
Nunclon™ Δ Surface 96-well microtitre plates	Nunc GmbH, Weisbaden, Germany
Vivaspin 2 spin concentrators (3,000 and 5,000 molecular weight cut-off	Vivascience, Sartorius Ltd., Epsom, Surrey, UK

2 ml open clear ampoules for amino acid analysis samples	Adelphi Tubes Ltd., Haywards Heath, UK
Nickel-nitrilotriacetic acid (Ni-NTA) agarose or Superflow™ resin	Qiagen Ltd., Crawley, UK
TAGZyme™ kit	Qiagen Ltd., Crawley, UK
Polyvinylidene difluoride (PVDF) membrane	Bio-Rad Laboratories Ltd., Herts, UK
Nitrocellulose membrane	Amersham International plc, Bucks, UK
Concanavalin A-Sepharose 4B	Amersham Biosciences AB, Uppsala, Sweden
Sephadex G-25 and G-50 (fine)	Amersham Biosciences AB, Uppsala, Sweden
Endoproteinase Arg-C	Roche Diagnostics GmbH, Mannheim, Germany

2.2 Methods

2.2.1 Isolation of plasmid DNA

Plasmid DNA was isolated from 5 ml (pGEM®T) or 10 ml (pET22b) overnight *E. coli* cell cultures using either the Wizard® *Plus* SV (Promega Ltd., Southampton, UK) or QIAquick™ (Qiagen Ltd., Crawley, UK) miniprep DNA purification systems, according to the manufacturers' instructions. DNA was eluted in sterile nuclease-free

water. Both systems are a modification of the Birnboim & Doly (1979) alkaline lysis method. Where necessary, DNA was concentrated as described in Section 2.2.4.

2.2.2 DNA restriction digests

Typically, restriction enzyme incubations were performed in accordance with the manufacturers' instructions. Digests were analysed and purified by submarine agarose gel electrophoresis (Section 2.2.3).

2.2.3 Submarine agarose gel electrophoresis and DNA purification

Agarose gel electrophoresis was performed with 1, 1.5 or 2 % (w/v) agarose gels (12 cm) containing TAE buffer (40 mM Tris base, 40 mM acetic acid and 2 mM EDTA) and 0.5 μg/ml ethidium bromide. DNA sample loading buffer was TAE containing 0.02 % (w/v) bromophenol blue, 0.02 % (w/v) xylene cyanol and 5 % (v/v) glycerol. Electrophoresis was performed at room temperature at 90 V (constant) in running buffer (TAE). The DNA was visualised by short wavelength UV transillumination (254 nm) and fragment size was estimated by comparison with the mobility of known standards (Section 2.1.6). When appropriate, desired DNA fragments were excised with a sterile scalpel and extracted using the QIAquick™ gel extraction kit (Qiagen Ltd., Crawley, UK) according to the manufacturer's instructions.

2.2.4 Precipitation of DNA

DNA was precipitated by the addition of 0.1 vol of 3 M sodium acetate, pH 5.2, and 0.75 vol of isopropanol, followed by vigorous vortexing. DNA was harvested by centrifugation (14,000 × *g*, 45 min) and any co-precipitated salt was removed by washing the pellet with 80 % (v/v) ethanol (200 μl) before allowing it to air dry. The DNA pellet was resuspended in an appropriate volume of sterile nuclease-free water.

2.2.5 DNA amplification by the polymerase chain reaction (PCR)

Taq DNA polymerase (Promega Ltd., Southampton, UK) was used for qualitative or analytical experiments. High fidelity *Pfu Ultra*™ (Stratagene Europe, Amsterdam, The Netherlands) DNA polymerase was employed for sequence-sensitive applications (see Section 2.2.7 for technical details).

Generally, reaction mixtures (made in thin-walled Thermowell™ PCR tubes) contained Mg^{2+}free *Taq* DNA polymerase buffer (10 mM Tris-HCl pH 9.0, 50 mM KCl, 0.1 % (v/v) Triton X100), 1.5 mM $MgCl_2$, 0.2 mM each of dATP, dGTP, dCTP and dTTP, 1 μM each of the appropriate primers and 5 U *Taq* DNA polymerase (Promega Ltd., Southampton, UK). Template DNA, typically plasmid DNA, lysed bacterial colonies or bacteriophage lysed by heating to 70 °C in 10 mM EDTA for 10 min, was added and the volume was made up to 50 μl with nuclease-free water. After gentle mixing, the tubes were centrifuged briefly. The thermal cycling was performed using a Techne Flexigene (Techne, Cambridge, UK) PCR machine. Denaturation, annealing and extension conditions, together with cycle number, primer details and any variations from this protocol, are described in the appropriate section of text. PCR products were analysed and purified by submarine agarose gel electrophoresis (Section 2.2.3)

2.2.6 Quikchange™ site directed mutagenesis

The PCR-based Quikchange™ method of site directed mutagenesis (McClelland & Nelson, 1992) generated some of the described changes in the IA3 genes cloned from *S. cerevisiae* and *S. castellii*. The principles of the technique are described fully in the Stratagene Quikchange™ kit (Stratagene Europe, Amsterdam, The Netherlands), but briefly: utilising plasmid vector pET22b containing the appropriate IA3 gene as the template DNA, site specific mutations were introduced using two complementary oligonucleotide primers designed according to the guidelines provided with the kit. The primers, which annealed to opposite strands of the plasmid across the region to be

mutated, encoded the desired mutation(s) in their middle and contained sufficient flanking sequence (typically up to 18 bases) to ensure specificity of location. In the course of the PCR extension step the primers, and therefore the mutations, were incorporated into the new plasmid. The template DNA, which originated from a *Dam+ E. coli* bacterium (strain DH5α) and was therefore methylated, was subsequently removed by treatment of the PCR mixture with *Dpn*I (a restriction enzyme which specifically cleaves methylated DNA) leaving only the non-methylated mutant plasmid generated by the PCR.

Cycling conditions varied between mutagenic reactions, as dictated by the kit manufacturer's instructions (an example is described in Figure 3.1c). However, a typical (50 µl) reaction contained 100 ng of the appropriate pET22b-IA3 vector as template, 100 pmol of each primer and 0.2 mM of each dNTP in 1 × *PfuTurbo*™ reaction buffer (containing 10 mM $MgCl_2$) with 1 U *PfuTurbo*™ DNA polymerase. Following thermocycling, 20 U of *Dpn*I was incubated with the reaction at 37 °C for 2 h. The remaining DNA was precipitated (Section 2.2.4), resuspended in 5 µl nuclease-free water, and transformed into *E. coli* strain DH5α (Section 2.2.10). DNA was prepared from transformants, selected by their ability to grow on LB agar plates containing ampicillin, for double-stranded DNA sequencing (Section 2.2.11) using the T7 remote primers (sequences provided in Appendix II.E) that anneal to sites flanking the multiple cloning site of pET22b.

2.2.7 Site-specific mutagenesis and gene chimaera production by overlap extension PCR

Some *S. cerevisiae* – *S. castellii* IA3 gene chimaeras were produced, together with some site directed mutants and the *S. cerevisiae* domain swap mutant known as 3AI (Section 3.5), by the overlap extension PCR method (Ho *et al.*, 1989). Specific design rationales are indicated in the appropriate part of the text, but briefly this method requires the design of two central oligonucleotide primers containing the mutation (or desired overlap/complementary sequence), and two remote primers flanking the gene. Three separate PCRs are performed; the first amplifies the DNA

containing the mutation site/complementary sequence and the upstream region as far as the remote forward primer (upstream of the start of the gene), the second amplifies the mutation site/complementary sequence and the downstream region (to beyond the end of the gene). The central primers were always designed to contain at least 15 bp overlap. Because the mutated and/or complementary sequence is common to the products of both PCRs, these two overlapping fragments can be mixed, annealed and extended in a third PCR, using the two remote primers to amplify the full-length product.

To minimise the chance of accidental mutations, high fidelity *Pfu Ultra*™ DNA polymerase (Stratagene Europe, Amsterdam, The Netherlands) was employed. All reaction mixtures (made in thin-walled PCR tubes) contained 1 × *Pfu Ultra*™ buffer, 0.2 mM each of dATP, dGTP, dCTP and dTTP, 1 μM each of the appropriate primers. Template DNA, either plasmid DNA (approximately 100 ng) or gel purified PCR fragments (Section 2.2.3), were added and the volume was made up to 49 μl with sterile nuclease-free water. After gentle mixing, the tubes were centrifuged briefly and amplifications were commenced under "hot start" (reaction mixture minus enzyme incubated for 1 min at 95 °C) with addition of 1 μl (2.5 U) *Pfu Ultra*™ DNA polymerase. The cycling parameters were typically the same for all PCRs: 95 °C for 5 min followed by 30 cycles of 95 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min 45 s. Finally, the reaction mixture was left at 72 °C for 10 min. A Biometra T3 thermocycler (Biometra GmbH, Goettingen, Germany) was used for thermal cycling. All PCR products were analysed and purified by submarine gel electrophoresis followed by gel extraction (Section 2.2.3). The full length PCR product was digested with appropriate restriction enzymes (Section 2.2.2) and ligated into the appropriate plasmid vector (Section 2.2.8).

2.2.8 Ligation of DNA into plasmid vectors

Insert DNA and plasmid vectors (with the exception of pGEM®T) were cut with appropriate restriction enzymes (Section 2.2.2) and gel purified (Section 2.2.3) before ligation. Ligation reactions (10 μl) were prepared in ligation buffer (Promega Ltd.,

Southampton, UK; 30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 0.5 mM ATP) with a vector:insert (1:2 molar) ratio and 3 U T4 DNA ligase (Promega Ltd, Southampton, UK.). Incubation was for 16 h at 15 °C.

When appropriate, products from *Taq* DNA polymerase PCRs were ligated into the vector pGEM®T using the pGEM®T Vector Systems Kit™ (Promega Ltd., Southampton, UK.) according to the manufacturer's instructions. The pGEM®T vector is supplied in a linearised form, with overhanging 3′-thymidine residues. These single 3′-thymidine overhangs at the insertion point provide a universally compatible "sticky end" for PCR products generated by DNA polymerases, such as *Taq*, that add a single deoxyadenosine in a template independent fashion to the 3′-end of PCR products (Zhou & Gomez-Sanchez, 2000).

Ligation products were transformed (Section 2.2.10) into *E. coli* strain DH5α to amplify the plasmid for other applications.

2.2.9 Preparation of *E. coli* competent cells

Cultures of the appropriate *E. coli* strain (Section 2.1.3) were grown in LB medium, with or without antibiotics as appropriate, to an attenuance (D) at 600 nm of 0.4. After harvesting by centrifugation (2,000 × *g*, 10 min, 4 °C), the cells were resuspended in half the original culture volume of cold 100 mM $CaCl_2$ containing 15 % (v/v) glycerol and incubated on ice for 1 h. The cells were harvested as before and resuspended in $1/10^{th}$ culture volume of the cold $CaCl_2$/glycerol solution. Finally, the cell suspension was immediately divided into 100 µl aliquots, flash frozen in liquid nitrogen and stored at –80 °C.

2.2.10 Transformation of *E. coli* competent cells

Aliquots (100 µl) of competent cells (Section 2.2.9) were thawed on ice. The plasmid DNA (100 ng) to be transformed was added with gentle stirring using a sterile pipette

tip and the mixture was incubated on ice for 30 min. The cells were heat shocked for 100 s at 42 °C and returned to the ice for a further 10 min. Finally, 900 µl of sterile SOC medium (Section 2.1.4), pre-warmed to 37 °C, was added and the mixture was incubated for 1 h at 37 °C. Subsequently, the cells were harvested by centrifugation (3,000 × *g* for 3 min) and the pellet was resuspended in 100 µl of fresh, sterile LB medium (no antibiotic). The cell suspension was plated onto LB agar plates containing the appropriate antibiotic(s), allowed to dry, inverted and incubated at 37 °C for 16 h.

2.2.11 Double stranded DNA sequencing

Nucleotide sequence was determined by the chain-terminator method of Sanger *et al.* (1977). Automated sequencing was performed by Mr G. Lewis or Mr E. Linley of the Sequencing Facility, Cardiff School of Biosciences, Cardiff University, or by Lark Technologies Inc. (Takeley, Essex, UK). Sequence analysis was performed using the BioEdit® computer program (Tom Hall, Department of Microbiology, North Carolina State University, USA).

2.2.12 Ligation of DNA into bacteriophage vector T7Select™ 415-1b and packaging *in vitro* of ligation products

For ligation into the pre-cut (with *Eco*RI and *Hin*dIII) T7Select™ vector, 415-1b (Novagen, AMS Biotechnology, Milton Keynes, UK), the relevant DNA fragments (detailed in Section 10.5.1) were treated with *Eco*RI and *Hin*dIII and gel purified. According to the manufacturer's suggestion, the following ligation mix (5 µl) was constructed: 0.02 pmol each of fragment DNA and 415-1b vector arm DNA; 30 mM Tris-HCl, pH 7.8; 10 mM $MgCl_2$; 10 mM DTT; 1 mM ATP and 0.45 U T4 DNA ligase. After gentle stirring with a pipette tip, ligation mixtures were incubated at 15 °C for 8 h and stored at 4 °C prior to packaging *in vitro*.

Packaging *in vitro* was performed by the addition of ligation reaction mixtures to T7Select™ Packaging Extract in a 1:5 (v/v) ratio, followed by incubation at 22 °C for 2 h. The reaction was stopped with 270 µl sterile LB medium, together with 20 µl chloroform. Efficiency of packaging was assessed by plaque assay (Section 2.2.13), prior to plate lysate amplification (Section 2.2.14).

2.2.13 Estimation of T7Select™ bacteriophage titre by plaque assay

E. coli strain BL21 cells were grown at 37 °C with shaking (200 rpm) in sterile LB medium to a D_{600} of 1.0. Serial 10- or 100-fold dilutions (typically between $1:10^3$ and $1:10^7$ overall dilutions) of the phage were prepared in LB. Aliquots (100 µl) of these dilutions were mixed with 250 µl of cells, and molten top agarose (3.5 ml) at 45-50 °C was added to each. The contents of the tubes were mixed by inversion and immediately poured onto 100 mm LB agar plates, with swirling, to produce a uniform surface. Following agarose solidification, the plates were inverted and incubated at room temperature for 16 h. The plaques on each plate were counted and the phage titre (in plaque forming units per ml, pfu ml^{-1}) calculated thus:

Phage titre (pfu ml^{-1}) = n × 10 × dilution factor

(where n = number of plaques on the plate)

2.2.14 Plate lysate amplification of T7Select™ bacteriophage

E. coli strain BL21 cells were grown at 37 °C with shaking (200 rpm) in LB medium to a D_{600} of 0.7. Bacteria and phage packaging reactions (centrifuged at 3,000 × *g* for 1 min to separate the chloroform) were mixed in a ratio of 10^5 pfu per 1 ml of bacterial cells. Samples (400 µl) of these mixtures were combined with 4 ml top agarose (at 45-50 °C), and poured immediately onto a 100 mm LB agar plate and spread evenly over the surface by swirling. After the top agarose had solidified, the plates were inverted and incubated at 37 °C for 4 h, until the plaques were almost confluent. Phage particles were eluted by immersing the plate surface in 4.5 ml sterile

phage extraction buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 6 mM $MgSO_4$) and incubating at 4 °C for 16 h. After collection of the eluant into a sterile container, 200 µl chloroform was added with mixing, and the resulting solution was centrifuged at 3,000 × *g* for 5 min to clarify the lysate. The supernatant was transferred to a sterile tube and stored at 4 °C. The bacteriophage titre of the lysate was assessed by plaque assay (Section 2.2.13).

2.2.15 Liquid lysate amplification of T7Select™ bacteriophage

When larger total numbers of bacteriophage were required to ensure representation of rare individuals, such as during rounds of enrichment biopanning (Section 2.2.18), a liquid lysate amplification method was preferred to the plate method described in Section 2.2.14.

E. coli strain BL21 cells were grown at 37 °C with shaking (200 rpm) in 25 ml LB medium to D_{600} = 0.9 in a 50 ml disposable plastic tube. The culture was infected with 100 µl of an appropriate dilution of bacteriophage (estimated by plaque assay, Section 2.2.13) to result in approximately 0.01 – 0.001 pfu per bacterium. The infected culture was incubated as before until lysis was observed (usually 2 – 4 h) as a visible reduction in attenuance and strands of cell debris in the medium. The lysate was clarified by centrifugation (8,000 × *g*, 10 min) and the supernatant was transferred to a sterile tube. The titre of the lysate was determined by plaque assay (Section 2.2.13).

2.2.16 Coating of 96-well microtitre plates with aspartic proteinases and assessment by fluorimetric assay

Various conditions were tested for the immobilisation of the aspartic proteinases *S. cerevisiae* proteinase A, *P. pastoris* proteinase and human pepsin onto the surface of Nunclon™ Δ Surface 96-well microtitre plates. The optimum conditions determined are described below.

The enzyme to be coated was diluted to 300 nM in assay buffer (100 mM sodium acetate pH 4.1 for *S. cerevisiae* proteinase A and *P. pastoris* proteinase, 100 mM sodium formate pH 3.1 for human pepsin). Aliquots (100 µl) were incubated in the wells of the microtitre plate for 1 h at 24 °C to allow binding. The amount of enzyme was in excess of the reported maximum binding capacity of each well, and was shown to produce the maximal rate of substrate cleavage. The wells were washed three times with assay buffer containing 0.5 % (v/v) Tween 20, followed by three washes with assay buffer (no Tween) to remove excess enzyme, and were dried by patting the plate on to paper towels. The integrity of the enzyme after the coating step was confirmed by fluorimetric assay using the fluorogenic substrate:

Lys-(Dabsyl)-Thr-Ser-Val-Leu*Met-Ala-Ala-Pro-Glu-(Lucifer yellow)

(10 µM, diluted in assay buffer from a stock in DMSO). The cleavage of the scissile Leu-Met bond (marked with an asterisk) by the immobilised enzyme releases the reporter (Lucifer yellow, excitation wavelength 390 nm) from the quencher (Dabsyl) producing an increase in fluorescence at the emission wavelength (538 nm). Reaction progress was followed using a Biotek FL600FA fluorimeter, linked to a standard IBM-compatible PC running KC4 software.

2.2.17 Biopanning T7Select™ bacteriophage displaying inhibitors using *S. cerevisiae* proteinase A immobilised on 96 well microtitre plates

Various conditions were tested for biopanning using *S. cerevisiae* proteinase A immobilised in the wells of 96-well microtitre plates. The conditions described below represent the optimum conditions determined, as described in the relevant Section of text. After immobilisation of the proteinase (Section 2.2.16), on occasion, the wells were blocked for 1 h at 24 °C with either 150 µl of 2 mg ml^{-1} lima bean trypsin inhibitor, 2 mg ml^{-1} rabbit IgG or 5×10^7 pfu ml^{-1} of killed (by incubation at 60 °C in 200 mM citrate buffer, pH 2.2 for 2 h) control T7Select™ bacteriophage (termed "*+phage*", expressing S•Tag™) in the assay buffer.

T7Select™ bacteriophage displaying residues 2-34 of *S. cerevisiae* IA3 (termed "*ϕphage*") were diluted in assay buffer to produce a known titre (typically 10^6 - 10^9 pfu ml^{-1} (point **A**)). Samples (100 µl) of these dilutions were added to the enzyme-coated microtitre plate wells and incubated for 1 h at 24 °C to allow binding. The wells were washed 6-8 times with assay buffer (stage **B**, assay buffer sometimes contained 0.5% (v/v) Tween20 and/or block reagent (see above)) to eliminate non-specifically bound phage. The remaining phage were eluted by either alkaline denaturation of proteinase A with 100 µl 100 mM CAPS buffer pH 9.9 for 1 h at 37 °C or by competition with 100 µM isovaleryl pepstatin (100 µl, in assay buffer) or 50 µM (100 µl) of free synthetic peptide representing residues 2-34 of *S. cerevisiae* IA3 (point **C**). Aliquots of bacteriophage-containing solutions were taken at points **A** and **C**, and at various points during stage **B**, and were titrated by plaque assay (Section 2.2.13).

2.2.18 Biopanning and enrichment of T7Select™ bacteriophage using *S. cerevisiae* proteinase A immobilised on concanavalinA-Sepharose 4B (ConA-Sepharose)

Distinct experiments during the optimisation process involved different conditions. Some of the variations are described in the relevant section of text (Section 10.5.3). A summary of the general protocol is described here.

A mixture of approximately (precise titre calculated by plaque assay, Section 2.2.13) 1×10^8 pfu *+phage* (displaying S•Tag™) and 1×10^2 pfu *ϕphage* (displaying residues 2-34 of *S. cerevisiae* IA3) was prepared in binding buffer (for example 100 mM MES, pH 6.25, 1 M NaCl; other buffers are detailed in Figure 10.4B) and added to 1 nmol *S. cerevisiae* proteinase A. After incubation for 10 min at 24 °C, 450 µl of a 33 % ConA-Sepharose slurry, pre-equilibrated into the binding buffer was added. Incubation was for 3 min at 24 °C to allow *S. cerevisiae* proteinase A to bind to the ConA. The mixture was subjected to centrifugation (5,000 × *g*, 3 min) and the supernatant was removed. The ConA-Sepharose pellet was resuspended in 1.5 ml

binding buffer and transferred to a fresh tube containing 14 ml binding buffer. After gentle mixing the sample was centrifuged and the supernatant was removed as before. The ConA-Sepharose was washed a further seven times, in 15 ml binding buffer on each occasion. Finally, the ConA-Sepharose pellet was resuspended in 1 ml sterile LB and 10 µl was removed for plaque assay (Section 2.2.13). On the basis of the titre, a suitable volume of the remainder was added to 25 ml of *E. coli* strain BL21 cells at D_{600} = 0.9 for liquid lysate amplification (Section 2.2.15).

The amplified lysate, which represented the conclusion of the first round of enrichment, was titred by plaque assay (Section 2.2.13). Approximately 1×10^8 pfu of round 1 lysate was mixed with 1 nmol *S. cerevisiae* proteinase A as before, and the whole process repeated for a second round. Typically 4 or 5 rounds of enrichment were performed.

A representative number of plaques (from plaque assay plates) were screened by PCR (Section 2.2.5) at the conclusion of each round. Using oligonucleotide primers (PhageSeqF and R, see Appendix II.E) flanking the cloning region of the T7Select™ bacteriophage genome, *+phage* and *ϕphage* could be differentiated in this manner by virtue of their differing insert size (S•Tag™ is shorter).

2.2.19 Analysis of protein samples by SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed according to Laemmli (1970), with vertical slab gels using the Bio-Rad Mini Protean II system (Bio-Rad Laboratories Ltd., Herts, UK). Depending on the expected size of the protein(s) to be analysed, the running gel contained either 12.5 % or 20 % (w/v) acrylamide solution (acrylamide and N,N'-methylene bisacrylamide in a ratio of 37:1 (w/w)) in 375 mM Tris-HCl buffer, pH 8.8, 0.1 % (w/v) SDS, 0.225 % (w/v) ammonium persulphate and 13.2 mM N,N,N',N'-tetramethylethylenediamine (TEMED). The stacking gel consisted of 5 % (w/v) acrylamide solution in 125 mM Tris-HCl buffer, pH 6.8, 0.1 % (w/v) SDS, 0.45 % (w/v) ammonium persulphate and 13.2 mM TEMED. The electrolyte running

buffer contained 0.1 % (w/v) SDS, 192 mM glycine and 25 mM Tris-HCl, pH 8.3. The protein sample buffer consisted of 40mM Tris-HCl buffer, pH 6.8, 1 % (w/v) SDS, 5 % glycerol, 0.003 % (w/v) bromophenol blue. Immediately prior to use, 2 % (v/v) 2-mercaptoethanol was added. Samples were heated at 100 °C for 7 min and centrifuged for 3 min at 14,000 × *g* before loading. Gels were run at 200 V, fixed and stained using Coomassie Brilliant Blue R250 (0.03 % (w/v) dissolved in 50 % (v/v) methanol and 10 % (v/v) acetic acid) for 1 h and destained in a methanol (25 % (v/v)) and acetic acid (7 % (v/v)) solution. Pre-stained protein markers (Section 2.1.6) were subjected to simultaneous electrophoresis with the samples in order to determine molecular weights of samples.

2.2.20 Detection of Hexahistidine tagged proteins by Western blotting

Following SDS polyacrylamide gel electrophoresis (Section 2.2.19), the resolved proteins were transferred to a moistened nitrocellulose membrane (Amersham International plc, Bucks, UK), by electroblotting at 145 mA for 45 min using a semi-dry blotter (Sartorius Ltd., Epsom, Surrey, UK) according to the manufacturer's instructions. Following transfer, the membrane was washed 3 × 10 min in TBSHis (10 mM Tris-HCl buffer pH 7.5, 150 mM NaCl) and blocked for 1 h in TBSHis containing 3 % (w/v) bovine serum albumin, Fraction V (Sigma Chemical Co., Poole, UK). The membrane was washed 3 × 10 min in TTBSHis (20 mM Tris-HCl buffer, pH 7.5, 500 mM NaCl, 0.05% (v/v) Tween® 20) and the anti-hexahistidine Ni-conjugated alkaline phosphatase (Qiagen Ltd., Crawley, West Sussex, UK) was applied for 1 h at a 1:1000 dilution (in TTBSHis). The membrane was washed 3 × 10 min in TTBSHis to remove excess Ni-conjugated alkaline phosphatase before being stained with Nitro blue tetrazolium and 5-Bromo-4-chloro-3-indoyl phosphate (Bio-Rad Laboratories Ltd., Watford, Herts, UK) with continuous shaking until a colour reaction developed (usually within 5 min). Immersing the developed membrane in distilled water stopped the staining reaction and the membrane was allowed to air-dry.

2.2.21 Preparation of protein samples for N-terminal sequencing

Proteins were separated by SDS-PAGE as described in Section 2.2.19 except that 192 mM Tricine was used in place of glycine in the electrolyte running buffer. Resolved proteins were transferred onto polyvinylidene difluoride (PVDF) membrane (Bio-Rad Laboratories Ltd., Herts, UK), which had previously been re-hydrated by soaking in methanol followed by distilled water and then equilibrated into anode buffer 2, by electroblotting in an identical fashion to that described for Western Blotting (Section 2.2.20). Following transfer, the membrane was stained with 0.1 % (w/v) Coomassie Brilliant Blue R250 (in 50 % (v/v) methanol) for 5 min and destained in 50 % (v/v) methanol, 10 % (v/v) acetic acid for 2 min followed with three washes of deionised water over a period of 5 min. The membrane was allowed to air dry and the required band(s) were excised with a sterile scalpel blade. N-terminal sequencing was performed by Alta Bioscience, University of Birmingham, Birmingham, UK.

2.2.22 Expression of recombinant forms of *S. cerevisiae* and *S. castellii* IA3 in *E. coli*

The *E. coli* expression strain (Section 2.1.3) was transformed (Section 2.2.10) with the desired pET22b plasmid containing the IA3 gene to be expressed. A LB broth starter culture (100 ml, containing 100 μg/ml ampicillin, 34 μg/ml chloramphenicol) was established by inoculation with a single colony from the transformation plate and incubation for 16 h at 37 °C with shaking (300 rpm). A 15 ml sample of the starter culture was used to inoculate 750 ml (in a 2 l conical flask with baffles to improve aeration) of pre-warmed LB broth (containing ampicillin and chloramphenicol as before) and the culture was incubated at 37 °C, 300 rpm. Total culture volume was usually 1.5 l (two flasks) but up to 6 l (eight flask) cultures were prepared in some cases. Expression of the IA3 gene was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) when the D_{600} of the culture reached 0.8. Growth continued for a further 3 h after induction under the same conditions. Cells were harvested by centrifugation (3,000 × *g* for 3 min in a Beckman JS-4.3 rotor, 4 °C),

the pellet was resuspended in 10 ml (per 750 ml of culture) sonication buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl) and stored at –20 °C.

2.2.23 Purification of recombinant hexahistidine-tagged IA3s by immobilised metal affinity chromatography (IMAC)

The expression cells in sonication buffer (Section 2.2.22) were subjected to four rounds of freeze-thaw to induce cell lysis. The viscosity of the cell lysate was reduced by macrotip sonication on ice (4 × 20 s, 52 % amplitude, Sonics Vibracell cup-horn sonicator) and the cell debris was removed by centrifugation (40,000 × *g* for 15 min, 4 °C in Beckman JA-20 rotor). The supernatant was added to a suitable volume of 50 % Ni-NTA Superflow™ slurry (typically 50 µl of resin (100 µl of slurry) per 100 ml of expression culture) and mixed by gentle end-over-end rotation for 50 min at 4 °C. The resin/supernatant mix was loaded into a column and the resin bed was washed with approximately 100 - 500 bed volumes of sonication buffer, followed by 100 - 500 bed volumes of wash buffer (50 mM sodium phosphate, pH 6.0, 300 mM NaCl, 10 % (w/v) glycerol). IA3 was eluted in the same solution, but with its pH adjusted to 3.5. Fractions (typically 1.5 ml) containing IA3 were identified spectrophotometrically by their absorbance at 280 nm and by SDS-PAGE (Section 2.2.19). Fractions containing contaminant bands were heated twice at 100 °C for 5 min and the precipitated protein was removed by centrifugation. Alternatively, the protein was subjected to a second round of IMAC. The concentration of the purified, homogeneous IA3 was determined either by amino acid analysis (Section 2.2.26) and/or, if appropriate, by titration against a known amount of enzyme (Section 2.2.31).

2.2.24 Removal of N-terminal hexahistidine tags from purified recombinant IA3s using the TAGZyme™ method

The TAGZyme™ method of N-terminal histidine tag removal is described fully in the relevant section of text (Chapter 4). Briefly, removal relies on a diamino exopeptidase (DAPase™, cathepsin C). Degradation is performed in the presence of excess glutamine cyclotransferase (Qcyclase™) which converts an N-terminal Gln to pyroglutamate, a modified residue that is refractory to DAPase™ cleavage. The IA3 gene constructs in the modified pET22b vector encoding the N-terminal histidine tag (Section 2.1.1) replaced the Met^1 of the IA3 sequence with Gln. Therefore, DAPase™ removes only the His-tag before exposing an N-terminal Gln, which is rapidly converted to pyroglutamate by Qcyclase™, thus protecting the rest of the protein from cleavage.

The protocol was based on the enzyme supplier's suggestions and varied slightly between individual proteins but generally 100 mU (10 µl) DAPase™ was activated by incubation in an equal volume of 20 mM cysteamine-HCl for 5 min at room temperature. Activated DAPase™, together with 6 U (120 µl) Qcyclase™, was incubated (2 h, 37 °C) with samples of 1 – 1.5 mg (in less than 1 ml) of purified N-terminally tagged recombinant IA3, the pH of which had been adjusted to 7.0. EDTA (5 mM) was added to chelate free Ni^{2+} ions. DAPase™ and Qcyclase™, both of which are C-terminally His-tagged, together with any IA3 protein with intact His-tag and the DAPase™-cleaved di-histidine fragments, were removed by subtractive IMAC using a 1 ml Ni-NTA agarose column pre-equilibrated with TAGZyme™ buffer (20 mM sodium phosphate, pH 7.0, 150 mM NaCl). Flowthrough fractions (usually 8 × 0.9 ml) containing IA3 were identified by SDS-PAGE (Section 2.2.19), pooled and, if necessary, re-concentrated to the original sample volume by centrifugation in Vivaspin2 (3,000 molecular weight cut-off) spin concentrators (Section 2.1.10) according to the manufacturer's instructions. With the exception of the *S. castellii* IA3 K45 truncation mutant, the remaining dipeptide fragments produced by the DAPase™ treatment (i.e. those not removed by the subtractive IMAC step) were removed by gel filtration using a 28 ml (approximately)

Sephadex G-25 column that was equilibrated in 25 mM sodium phosphate buffer, pH 6.5, 50 mM sodium chloride. Fractions (0.9 ml) containing IA3 were identified by SDS-PAGE (Section 2.2.19), pooled and concentrated as before. His-tag removal was confirmed by amino acid analysis (Section 2.2.26) and MALDI-TOF MS (Section 2.2.28).

2.2.25 Digestion of recombinant IA3s with endoproteinase Arg-C

Conditions varied between IA3 proteins and during the optimisation process, but generally an IA3:endoproteinase Arg-C 50:1 (molar) ratio was constructed. Enzyme (20 pmol) was activated by incubation in 90 mM Tris-HCl buffer, pH 7.6, 8.5 mM $CaCl_2$, 5 mM DTT, 0.5 mM EDTA for 5 min at 37 °C before addition of purified recombinant IA3 (800 pmol, adjusted to pH 7). The mixture was incubated for 2 h at 37 °C, after which the endoproteinase Arg-C was inactivated by heating to 100 °C for 7 min or inhibited with 2 μM leupeptin.

2.2.26 Amino acid analysis

Chemically synthesised peptides (powder, Section 2.1.8) were dissolved in sterile distilled water or DMSO to an approximate concentration of 2 mM. Aliquots of peptides dissolved in DMSO were freeze-dried *in vacuo* prior to further processing. Recombinant proteins and peptides dissolved in water were analysed without drying. Samples for analysis (approximately 5 – 15 nmol) were hydrolysed in 0.5 ml HCl (6M, Aristar grade) in acid-washed 2 ml glass ampoules (Adelphi Tubes Ltd., Haywards Heath, UK) with norleucine (15 nmol), or arginine (15 nmol) if norleucine was present in the peptide, added as an internal standard. The tubes were flushed with nitrogen, evacuated and sealed before being placed in an oven at 110 °C and hydrolysed for 16 – 48 h. The tubes were opened and the contents were dried *in vacuo* over sodium hydroxide pellets. The samples were resuspended in 0.2 M sodium citrate buffer, pH 2.2, before being loaded into a capsule for analysis by Mr. M. O'Reilly or Mr. W. Edwards (Cardiff School of Biosciences, Cardiff

University) on a Biochrom 20 amino acid analyser (Pharmacia Biotechnology, St Albans, Herts, UK).

2.2.27 Analytical Fast Protein Liquid Chromatography (FPLC)

Cleavage of synthetic peptide substrates was monitored using a reversed phase column and FPLC. Assay tubes containing appropriate concentrations of buffer, substrate and enzyme in a final volume of 50 µl were incubated for a predetermined time at a defined temperature. Reactions were stopped by the addition of 450 µl 5 % (v/v) trifluoroacetic acid (TFA), followed by centrifugation (14,000 × *g*, 5 min) to pellet insoluble material. Samples were stored at 4 °C if necessary.

The samples were loaded on to a Pep RPC HR 5/5 reversed phase column (Pharmacia Biotechnology, Uppsala, Sweden), pre-equilibrated with buffer A (18 MΩ, 0.22 µm-filtered, de-gassed water containing 0.1 % (v/v) TFA) linked to a basic FPLC system (Pharmacia Biotechnology, Uppsala, Sweden). Substrate and cleavage products were eluted and resolved by a 0-50 % linear gradient of buffer B (0.22 µm-filtered, de-gassed acetonitrile (BDH Ltd., Poole, UK) containing 0.1 % (v/v) TFA) at a flow rate of 0.5 ml min^{-1}. Eluted peaks were detected at 214 nm with a Pharmacia Single Path Monitor UV-1/214 unit. Elution profiles were recorded on a Rec-482 chart recorder (Pharmacia Biotechnology, Uppsala, Sweden) enabling quantification of relative amounts by integration of the area under the relevant peaks. If required, the identity of the bond(s) cleaved could be established by collection of the products and either determination of their amino acid composition (Section 2.2.26) or by mass spectrometry (Section 2.2.28).

2.2.28 Mass spectrometry (MS)

Matrix-Assisted Laser Desorption Ionisation –Time-of-Flight (MALDI-TOF) MS was used to identify cleavage products and confirm protein identities. Typically, samples were prepared to produce a 2-10 µM peptide/protein concentration, to be compatible

with the sensitivity of the instrument. Specific details of the buffers and conditions of individual samples are provided in the relevant section of text. MALDI-TOF MS was performed by either Mr. M. Weldon and Dr. L. Packman of the Protein and Nucleic Acid Chemistry (PNAC) Facility, Department of Biochemistry, University of Cambridge, Cambridge, UK or Dr. D. Lamont and Dr. K. Beattie of the Proteomics Facility, University of Dundee, Dundee, UK. Data were analysed using the massXpert computer program (Rusconi & Belghazi, 2002).

2.2.29 Spectrophotometric determination of kinetic parameters for the enzyme-mediated hydrolysis of a synthetic chromogenic peptide substrate

Enzymatic reactions were monitored spectrophotometrically using either a Hewlett-Packard 8452A diode array spectrophotometer (HP, Palo Alto, CA, USA), or a Cary 50 Bio spectrophotometer (Varian Ltd., Oxford, UK). Both spectrophotometers were fitted with a temperature-regulated multi-cell carriage. The Hewlett-Packard system was linked to a Sysvol kinetic software package supplied with the spectrophotometer. An IBM-compatible PC running the Cary WinUV enzyme kinetics software controlled the Cary 50 spectrophotometer.

A chromogenic peptide substrate (RS6) Lys-Pro-Ile-Glu-Phe-Nph-Arg-Leu, was synthesised (Pepsyn Ltd., University of Liverpool, UK or Alta Bioscience, University of Birmingham, UK) to contain a *p*-nitrophenylalanine (Nph, above) reporter group in the P_1' position (nomenclature of Schechter & Berger, 1967; see Section 1.6). Cleavage of the peptide bond adjacent to the nitrophenylalanine generates a new amino-terminal residue in the peptide, i.e. the *p*-nitrophenylalanine itself. After cleavage, at pH values below 6.5, the α-amino group of the nitrophenylalanine is protonated. Protonated *p*-nitrophenylalanine has a shorter maximum absorbance wavelength, with the result that the rate of hydrolysis of this peptide bond can be monitored by measuring the decrease in absorbance of the solution at 300 nm. The concentration of RS6 substrate was determined by amino acid analysis (Section 2.2.26).

Assays were performed in semi-micro quartz cuvettes (1 cm pathlength) in a final volume of 800 µl in either 100 mM sodium formate buffer, pH 3.1 or 100 mM sodium acetate buffer, pH 4.7.

In enzyme/substrate reactions, all reagents except the enzyme were added prior to starting the reaction, and incubated for approximately 1 min to achieve the experimental temperature (typically 37 °C). The reaction was thus initiated by the addition of enzyme. In contrast, for inhibition studies enzyme and inhibitor were pre-incubated in the reaction buffer at 37 °C for approximately 1 min prior to reaction initiation by addition of the substrate.

For determination of kinetic parameters, initial rates (v) were measured at several different substrate concentrations [S] within an appropriate range spanning the Michaelis constant, K_m. The initial rates were measured by following the changes in absorbance as substrate cleavage occurred using the reaction rate software, which performed a linear-least squares regression analysis on the linear phase of the progress curve as the reaction proceeded. The substrate concentrations and their corresponding initial rates were entered into Sysvol, enabling v against [S] and 1/v against 1/[S] graphical productions. Sysvol plots curves using the equation:

$$v = \frac{V_{max}[S]}{(K_m + [S])}$$

and analyses the fit of each curve to the experimental data. A curve of best fit is generated by Marquardt iteration, enabling the program to calculate V_{max} (the maximal velocity of the enzyme) and K_m. All data were also analysed using the enzyme kinetics function of the Graphpad™ Prism software package. The results were universally coincident.

2.2.30 Determination of inhibition constants (K_i)

If the ratio of $[E_t]/K_i < 0.1$ (when $K_i \approx 10^{-5}$ to 10^{-8} M), where E_t = total enzyme concentration, the Michaelis-Menten equation can be assumed to be valid. In these cases, the Dixon (1953) construction can be applied and the K_i calculated using the competitive inhibition software of Sysvol. This section of the program is similar to that used to obtain V_{max} and K_m (Section 2.2.29), except that [I] against v data are entered also and theoretical curves are calculated using the competitive inhibition equation:

$$v = \frac{V_{max}[S]}{K_m(1 + ([I]/K_i)) + [S]}$$

If the $[E_t]/K_i$ ratio lies between 0.1-100 ($K_i = 10^{-8}$ to 10^{-10} M), the Michaelis-Menten equation cannot be applied because the assumption that $[I_{free}]$ (free inhibitor concentration) equals $[I_t]$ (total inhibitor concentration) is not valid. Under such circumstances, the method of Henderson (1972) was employed, taking account of the depletion of free inhibitor inherent in the behaviour of tight binding, pseudo-irreversible inhibitors. For the purpose of such analyses, the programme "Tightfitvar" was devised, written and kindly made available by Professor R.A. John and Dr. J.G. Jones (Cardiff School of Biosciences, Cardiff University, UK).

2.2.31 Determination of active enzyme concentration by titration

To determine the quantity of active enzyme present in spectrophotometric assays and, therefore, to calculate the concentration of individual enzyme stocks, a suitable tight binding inhibitor was used as a titrant. Isovaleryl-pepstatin (IVP, Section 1.7.1) was used to titrate all enzymes described in this study; the IVP-titrated concentration of *S. cerevisiae* proteinase A stocks was also confirmed by titration with the *S. cerevisiae* IA3 mutant V26A/A29G/F30M (numbered **52** in later chapters). The stock of IVP was dissolved in DMSO, and its concentration was determined by amino acid analysis (Section 2.2.26), as was the concentration of **52**. The active concentration of enzyme was obtained from a plot of percentage of enzyme activity

remaining (where 100 % activity is the rate in the absence of titrant) (y-axis) *versus* concentration of titrant (x-axis). Applying the approximation/assumption of directly proportional mutual depletion of enzyme and titrant, the active concentration of the enzyme is obtained by extrapolation of the linear portion of the plot to its intercept with the x-axis. The "Tightfitvar" programme (Section 2.2.30) enabled accurate calculation of this intercept value. With the knowledge of the active enzyme concentration, it was possible to reverse the extrapolation in order to calculate the concentration of other titrating inhibitors. This was especially useful in the instances where the concentration of an inhibitor was too low for determination by amino acid analysis.

Chapter 3

Site-directed mutation, expression and purification of recombinant forms of *S. cerevisiae* IA3 and *S. castellii* IA3

3.1 Introduction

This chapter, together with Chapter 4, describes the construction of the DNA encoding various modified recombinant forms of IA3 used in the subsequent studies. The reasoning and rationale behind individual mutants will not be dealt with in detail here, but in the relevant section of later chapters.

The general principles for producing recombinant *S. cerevisiae* IA3 in *E. coli* have been well established by previous work in this laboratory (Li *et al.*, 2000). The *S. cerevisiae* gene encoding IA3 (Schu & Wolf, 1991) was amplified from genomic DNA by Dr. L. Phylip (Cardiff University, Cardiff, UK) using PCR and was cloned into the *Nde*I-*Xho*I restriction sites of the *E. coli* recombinant protein expression vector pET22b. Cloning in this way attaches a Leu-Glu-$(His)_6$ tag to the C-terminus of IA3, facilitating the purification of the recombinant protein from bacterial cell lysates by immobilised metal affinity chromatography (IMAC, Section 2.2.23). During the course of the work described in this thesis, our colleague Dr. D. Wyatt (Cardiff University, Cardiff, UK) discovered another IA3, from the closely-related yeast *S. castellii*. This gene was subsequently cloned in the same manner. The *S castellii* IA3 is discussed in detail in Chapter 6.

In the studies described here, mutations of several different types were produced to investigate further the inhibitory properties of IA3, and to attempt to re-direct inhibitory specificities and improve potencies. As in previous studies (Li *et al.*, 2000; Phylip *et al.*, 2001; and unpublished data), site-directed single or multiple residue mutations were introduced into recombinant *S. cerevisiae* IA3. A similar process was also applied to the *S. castellii* IA3 gene. Furthermore, both *S. cerevisiae* and *S. castellii* IA3 forms were subjected to a series of site-specific C-terminal truncations. Chimaeric forms of IA3, containing sequence derived from both *S. cerevisiae* IA3 and *S. castellii* IA3, were also produced in various forms. Finally, a mutant of *S. cerevisiae* IA3 was manufactured where the C-terminal (i.e. residues 35-68) and N-terminal (residues 1-34) halves of the protein were transposed, so that the primary sequence of the protein read Naa35 → aa68-aa1 → aa34C.

Mutations were introduced by three different methods:

(i) Quikchange™ PCR;

(ii) Overlap extension PCR;

(iii) Oligonucleotide cassette mutagenesis.

In view of the large number of variants produced for this study, it is not practically feasible to describe in detail the design and manufacture of every IA3 mutant. Instead, this chapter follows the genetic manipulations made to produce at least one mutant of every type (site-directed residue(s), truncation, chimaera and domain-swap mutants), whilst representing all the mutagenic methods that were employed. Finally, specific examples are used to illustrate the expression and purification processes.

ADDITIONAL NOTES

- A complete list of all the mutants produced, together with the mutagenic strategy employed and the identity of any oligonucleotides required, is provided in Appendix II.

- During the course of research in this project, it became necessary to remove the His-tag from various recombinant IA3 forms. The design and purification of "tagless" IA3 forms are described in Chapter 4.

3.2 Site-directed mutation

As described in Section 1.8, previous X-ray crystallographic and kinetic assay data (Li *et al.*, 2000; Phylip *et al.*, 2001) had suggested distinct regions of *S. cerevisiae* IA3 that might be important for inhibitory potency. However, many questions remained unanswered, particularly regarding the origin of the exquisite specificity of *S. cerevisiae* IA3 for the target aspartic proteinase, *S. cerevisiae* vacuolar proteinase A. During the course of the studies described in this thesis, the contribution of individual or multiple residues to inhibitory potency was investigated by studying the effect of a wide variety of site-directed mutations. Some mutants were also constructed to attempt to re-direct inhibitory specificity against new enzyme

targets. Furthermore, the differing specificity and potency of IA3 from *S. castellii* was also investigated by site-directed mutation.

Site-directed mutations were produced by all three mutagenic methods employed in this study: the PCR-based methods of Quikchange™ and overlap-extension and also by oligonucleotide cassette mutagenesis. Each process will be illustrated by an example.

3.2.1 Site directed mutagenesis of *S. cerevisiae* IA3 Lys18→Arg (K18R) by Quikchange™ PCR

The Quikchange™ PCR method of site-directed mutagenesis (Section 2.2.6) is particularly attractive for the introduction of a small number of localised mutations because the mutated gene does not need to be re-cloned. The template DNA – plasmid pET22b containing wild-type *S. cerevisiae* IA3 – was amplified in its entirety using two complementary oligonucleotide primers (Figure 3.1A and B) encoding the desired mutation (in this case AAG to AGG at codon 18), together with sufficient flanking sequence to ensure specific annealing at the correct location on opposite strands of the plasmid. In the course of the PCR, therefore, both strands of the plasmid were amplified, generating a plasmid harbouring the mutation and containing staggered nicks on opposite DNA strands. The reaction conditions, detailed in Figure 3.1C, were as instructed by the Stratagene Quikchange™ kit. In order to minimise the chance of the introduction of other, accidental, mutations, high-fidelity *Pfu*Turbo™, a proof-reading DNA polymerase, was used in the PCR.

The template DNA was removed by treatment of the PCR mixture with the restriction enzyme *Dpn*I, which digests methylated DNA specifically (see Section 2.2.6), leaving only the non-methylated mutant plasmid generated by the PCR to be transformed into *E. coli* DH5α. The sequencing result of plasmid DNA prepared from transformant colonies confirmed that the site-directed mutagenesis had been successful, converting codon 18 of *S. cerevisiae* IA3 from Lys (AAG) to Arg (AGG), whilst introducing no unwanted mutations elsewhere in the IA3 sequence.

A

WT IA3 sequence		12 F	13 Q	14 S	15 S	16 K	17 E	18 K	19 L	20 Q	21 G	22 D	23 A	24 K	25 V
WT coding	5′-~~TTT	CAG	AGC	TCA	AAG	GAA	AAG	TTG	CAG	GGC	GAT	GCA	AAG	GTA~~-3′
K18R R:	3′-GTC	TCG	AGT	TTC	CTT	TCC	AAC	GTC	CCG	CTA	CGT	TTC	C-5′	
WT non-coding	3′-~~AAA	GTC	TCG	AGT	TTC	CTT	TTC	AAC	GTC	CCG	CTA	CGT	TTC	CAT~~-5′
K18R F:	5′-CAG	AGC	TCA	AAG	GAA	AGG	TTG	CAG	GGC	GAT	GCA	AAG	G-3′	
Mutant IA3 sequence		F 12	Q 13	S 14	S 15	K 16	E 17	R 18	L 19	Q 20	G 21	D 22	A 23	K 24	V 25

B

K18R F: 5′-CAGAGCTCAAAGGAAAGGTTGCAGGGCGATGCAAAGG-3′
K18R R: 5′-CCTTTGCATCGCCCTGCAACCTTTCCTTTGAGCTCTG-3′

Melting temperature (T_m) = 81.6 °C, based on the equation:

$$T_m = 81.5 + 0.41\,(\%GC) - 675/N - \%\ \text{mismatch}$$

Where **%GC** = percentage of G + C bases in oligonucleotide primer
N = number of bases

C

95 °C, 30 s
95 °C, 45 s ←
55 °C, 1 min] 17 cycles
68 °C, 12 min

Figure 3.1 Site directed mutagenesis of K18→R in *S. cerevisiae* IA3 by Quikchange™ PCR.

A. A representation of the IA3 sequence (in GREY) to which the mutagenic oligonucleotides (in **BLACK**) anneal. The base in the DNA codon (and the corresponding amino acid) in the wild-type (WT) sequence to be changed is in GREEN, the mutation is in **RED**. The two primers (F and R) are represented as they would anneal to the strands of the plasmid.

B. The two primers (mutation in **RED**) were designed, according to the guidelines of the Stratagene Quikchange™ kit, to have a T_m >78 °C, according to the equation illustrated.

C. Thermocycling conditions for the Quikchange™ PCR. The elongation time (in min) was 3 × the approximate size of the plasmid (in kb).

3.2.2 Site-directed mutagenesis of *S. castellii* IA3 Met18→Lys and Lys22→Met (M18K/K22M) by overlap extension PCR

Overlap extension PCR (Ho *et al.*, 1989) was a more efficient method of mutagenesis than Quikchange™ (Section 3.2.1) when several mutations were to be introduced simultaneously. The disadvantages of the process were that: (a) three separate PCRs were required; and, (b) the IA3 gene had to be re-cloned into pET22b after mutagenesis.

The methods employed, together with schematic details of the primers designed in this instance, are described in Figure 3.2. All three PCRs used the same conditions, described in Section 2.2.7. Plasmid pET22b, containing the recombinant *S. castellii* IA3 gene, served as the template DNA for the first two reactions. The first PCR amplified the portion of the gene corresponding to residues 1-23. The forward primer, F1, conserved the *Nde*I consensus sequence (**brown**) immediately upstream of the gene, whilst the reverse primer, R1, incorporated the mutations at codons 18 and 22 during the PCR.

The forward primer for the second PCR (F2) was designed to produce a 21 bp overlap with the 3′-end of the first PCR product, thereby also incorporating the mutations during the reaction. The second PCR therefore amplified from codon 17 to the end of the gene, using the reverse primer, R2, to add the *Xho*I consensus sequence (**purple**, required for cloning) immediately downstream of the final codon (81). The products of the first two PCRs were analysed (lanes 1 and 2, respectively, in Figure 3.3) and gel-purified by submarine agarose (2 % w/v) gel electrophoresis (Section 2.2.3). The sizes of the fragments were compared with the co-electrophoresed DNA marker ("ϕ" in Figure 3.3), which confirmed the expected sizes of the two fragments (79 bp and 201 bp for the front and back halves, respectively). The purified fragments from the two PCRs were mixed, denatured and allowed to anneal to each other by virtue of their overlapping complementary sequence. The *S. castellii* IA3 gene, containing the M18K/K22M mutations in the annealed overlap region, was therefore reconstituted in the third PCR, using the F1 and R2 oligonucleotides to prime the extension of the full-length product (Figure 3.2). This "readthrough" PCR was analysed by electrophoresis

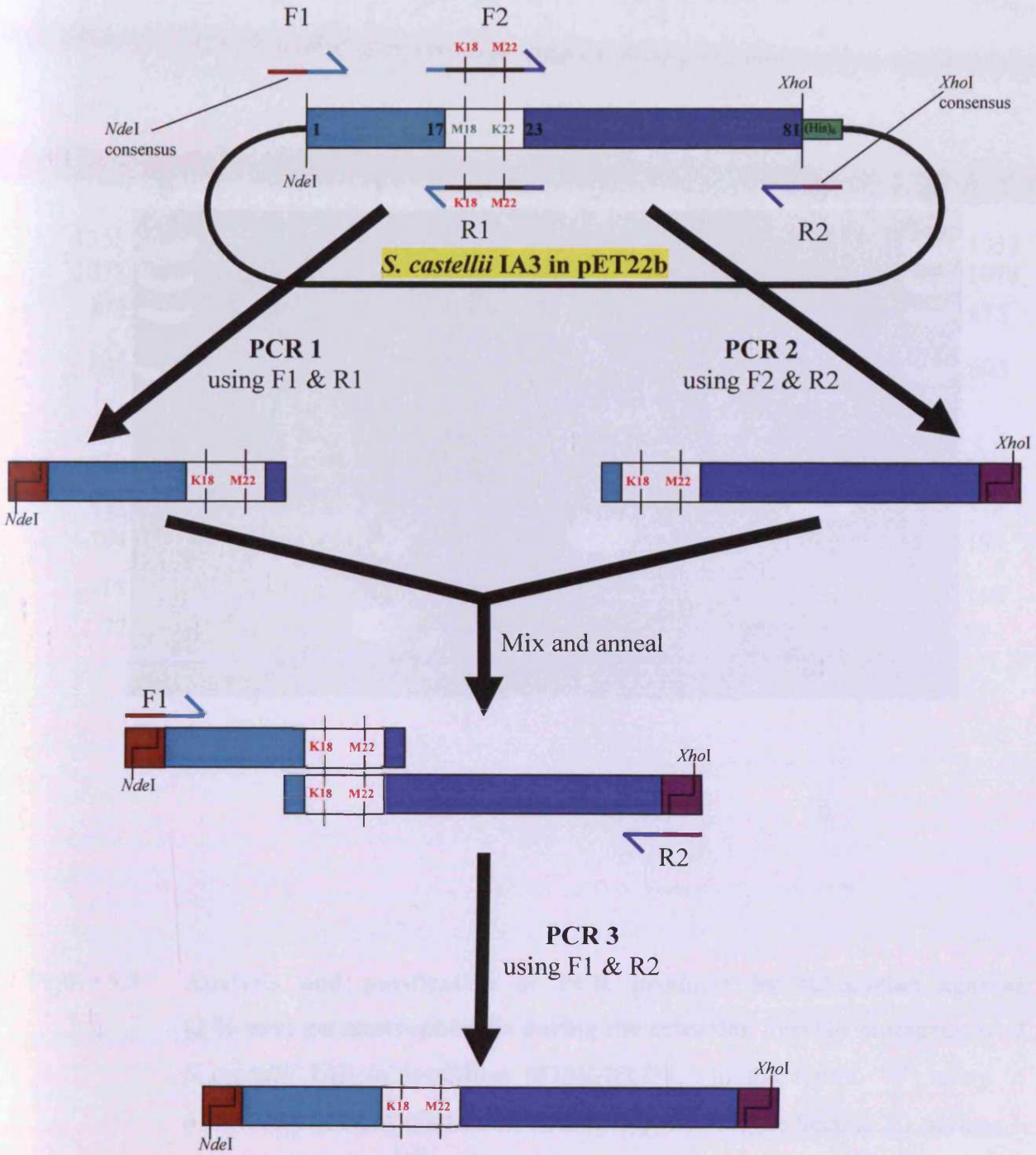


Figure 3.2 **Schematic explanation of the process to create the *S. castellii* IA3 mutant M18K/K22M by overlap extension PCR.**

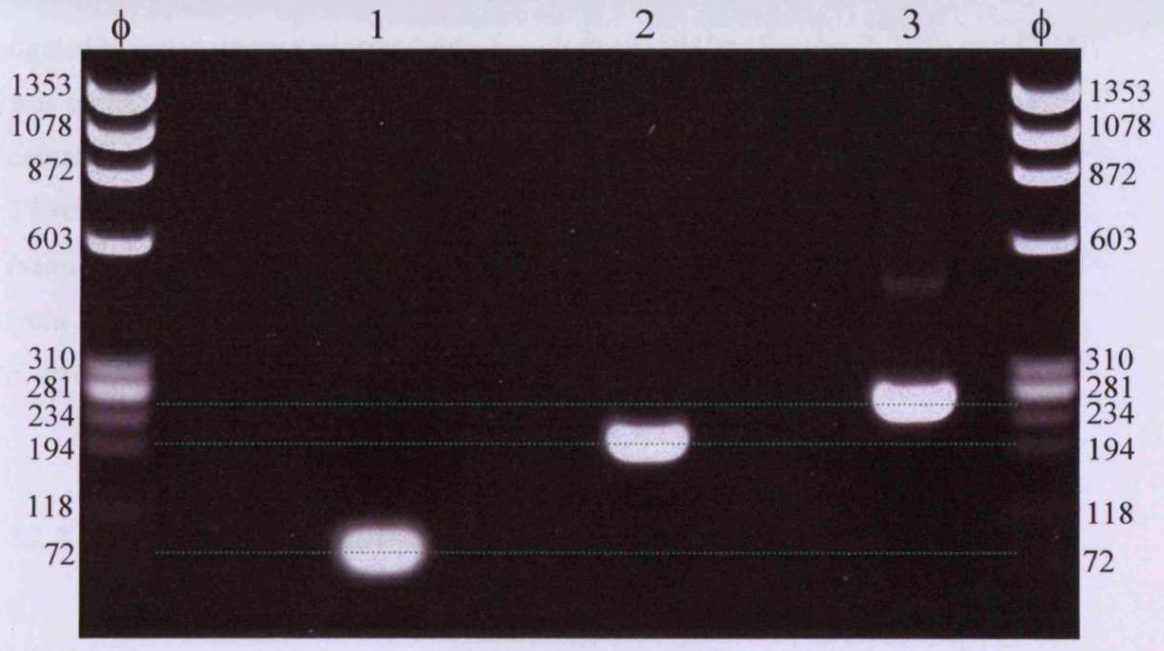


Figure 3.3 **Analysis and purification of PCR products by submarine agarose (2 % w/v) gel electrophoresis during the extension overlap mutagenesis of *S. castellii* IA3 to introduce M18K/K22M.** In the figure, "ϕ" refers to ϕ174X/*Hae*III DNA markers (New England Biolabs, see Section 2.1.6); lane 1 contains the product of the first PCR; lane 2 the second PCR product; and, lane 3 shows the product of the overlap extension "readthrough" PCR. All three products correlated well (green dotted lines) with their expected sizes of 79, 201 and 262 bp, respectively.

as before and the product confirmed qualitatively (lane 3 of Figure 3.3) by comparison with both DNA markers and the products of the first two PCRs. A fragment of approximately 260 bp was produced, which correlated with the expected size of 262 bp. The product was extracted from the gel (Section 2.2.3) and treated with *Nde*I and *Xho*I restriction enzymes (Section 2.2.2) to facilitate ligation (Section 2.2.8) into pET22b that had been prepared in the same manner. The re-ligated plasmid was transformed into *E. coli* strain DH5α (Section 2.2.10) and DNA was prepared from transformants, selected by their ability to grow on LB agar plates containing ampicillin. Double-stranded DNA sequencing (Section 2.2.11), using the T7 remote primers that anneal on either side of the multiple cloning site of pET22b (sequences provided in Appendix II.E), confirmed that codon 18 had been converted from Met (ATG) to Lys (AAG) and codon 22 from Lys (AAA) to Met (ATG) in *S. castellii* IA3, whilst introducing no other mutations.

3.2.3 Oligonucleotide cassette mutagenesis of *S. cerevisiae* IA3 Gln20→Glu, Val26→Ala, Ala29→Gly, and Phe30→Met (Q20E/V26A/A29G/F30M)

Some of the IA3 proteins prepared for this thesis contained multiple residue mutations spanning a wide region of the inhibitory sequence of *S. cerevisiae* IA3 (residues 1-34). The oligonucleotide primers required to produce such mutants in one step by either Quikchange™ or overlap extension PCR would have been very long. To avoid many potential problems with using such long primers in PCRs (for example annealing at the wrong point or secondary structure formation), a modified form of the *S. cerevisiae* IA3 gene in pET22b was produced, containing two unique restriction enzyme sites within the sequence. Thus, it was possible to insert oligonucleotide cassettes, engineered with appropriate "sticky ends" and harbouring the desired mutations, into the appropriately digested, modified IA3-pET22b vector.

3.2.3a Construction of the "antiSacJ35" IA3-pET22b vector

Inspection of the wild-type *S. cerevisiae* IA3 gene sequence (Schu & Wolf, 1991), shown in Figure 3.4A within a small section of the cloning vector, pET22b, revealed a *Sac*I restriction endonuclease site (in **blue**), ideally located close to the middle of the inhibitory region of IA3 (i.e. residues 1-34). However, although the intrinsic pET22b *Sac*I site had been removed as a result of the IA3 cloning process, a new *Sac*I site (**orange** in Figure 3.4A) was simultaneously generated, between the sequence encoding the final residue of IA3 (Glu, encoded by GAG) and the first three bases of the *Xho*I site (CTC). This unwanted *Sac*I site was removed by Quikchange™ site-directed mutagenesis (Sections 2.2.6 and 3.2.1) of the final Glu codon from GA**G** → GA**A** (details of the oligonucleotides for this process can be found in Appendix II.C).

Investigation of the pET22b sequence revealed no *Nhe*I restriction sites. Therefore, Quikchange™ site-directed mutagenesis of a single base was used to engineer a unique *Nhe*I site into the *S. cerevisiae* IA3 gene sequence corresponding to Ala34/Ser35 (GCT AG**T** → GCT AG**C**, **pink** box in Figure 3.4A, oligonucleotide sequences in Appendix II.C). It was now possible to split the inhibitory sequence (residues 1-34) of *S. cerevisiae* IA3 in two by simply digesting the modified IA3-pET22b vector with either *Nde*I and *Sac*I (thus removing the sequence corresponding to residues 1-14) or with *Sac*I and *Nhe*I (removing residues 15-34). This new version of *S. cerevisiae* IA3 in pET22b was termed "antiSacJ35"-pET22b.

3.2.3b Design of the oligonucleotide cassette and insertion into antiSacJ35-pET22b

Two complementary oligonucleotides (termed "EAGM TOP" and "EAGM BOT") were designed so that, when annealed, they re-constituted the region of the IA3 gene between the *Sac*I and *Nhe*I sites (including the appropriate "sticky ends") and encoded the desired mutations (in this case Q20E/V26A/A29G/F30M, see Figure 3.4B). The cassette was constructed by heat denaturing a mixture of the two oligonucleotides (approximately 5 nmol of each) at 94 °C for 10 min then cooling to 16 °C at 0.05 °C s^{-1} in a Biometra T3 thermocycler. The cassette was diluted appropriately

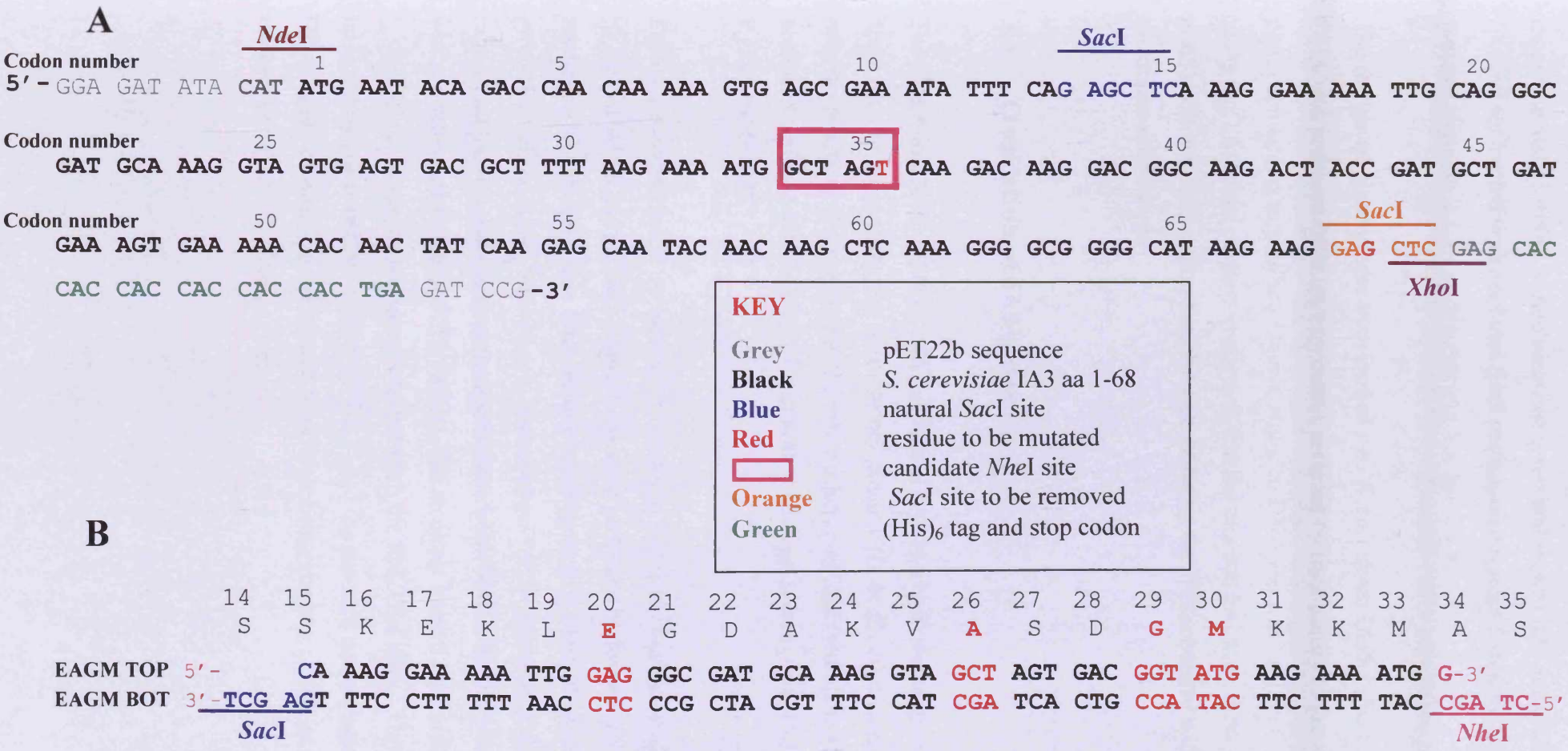


Figure 3.4A. Wild-type *S. cerevisiae* IA3 gene sequence cloned between *Nde*I and *Xho*I restriction sites of plasmid pET22b. Two separate Quikchange™ PCR mutageneses produced the "antiSacJ35" mutant form of the vector construct: (i) at codon 35 (to AGC), introducing an *Nhe*I restriction site; and (ii) codon 68 (to GAA), to remove the unwanted *Sac*I restriction site.

B. Demonstration of the production of the cassette from two synthetic oligonucleotides, including the "sticky" ends for ligation into antiSacJ35-pET22b digested with *Sac*I and *Nhe*I. Mutated codons are in red.

(approximately 1:1000) in nuclease-free water and ligated into antiSacJ35-pET22b (previously treated with *Sac*I and *Nhe*I restriction enzymes (Section 2.2.2) and gel purified (Section 2.2.3)) as described in Section 2.2.8.

The re-ligated plasmid was transformed into *E. coli* strain DH5α (Section 2.2.10) and DNA was prepared from transformants, selected by their ability to grow on LB agar plates containing ampicillin. Double-stranded DNA sequencing (Section 2.2.11) using the T7 remote primers confirmed that the cassette had been ligated into the antiSacJ35-pET22b vector, thereby reconstituting the IA3 sequence with the mutations introduced.

3.3 Truncation of IA3 genes

The C-terminal half, or "tail", of *S. cerevisiae* IA3 had been shown previously to make little or no contribution to inhibitory potency (Li *et al.*, 2000), raising questions over the function. To investigate the role further, two intermediate truncations were produced – after Ala45 and Tyr57 – in both wild-type *S. cerevisiae* IA3 and a mutant, K18I/L19M/D22I.

Since all recombinant *S. cerevisiae* IA3 genes in this study were cloned between *Nde*I (5′) and *Xho*I (3′) of pET22b, truncations were produced by inserting a *Xho*I restriction site after the new final residue (see Figure 3.5, which illustrates the production of the Ala45 truncation). Quikchange™ PCR (Sections 2.2.6 and 3.2.1) introduced the restriction site at the appropriate point within the complete IA3 gene whilst it remained in the pET22b vector. The resultant mutant plasmids were digested with *Xho*I to excise the portion of IA3 between the two *Xho*I sites. This fragment was removed by gel-purification (Section 2.2.3) and the plasmid was re-ligated with T4 DNA ligase, resulting in the desired truncation being introduced whilst maintaining the Leu-Glu-$(His)_6$ C-terminal tag.

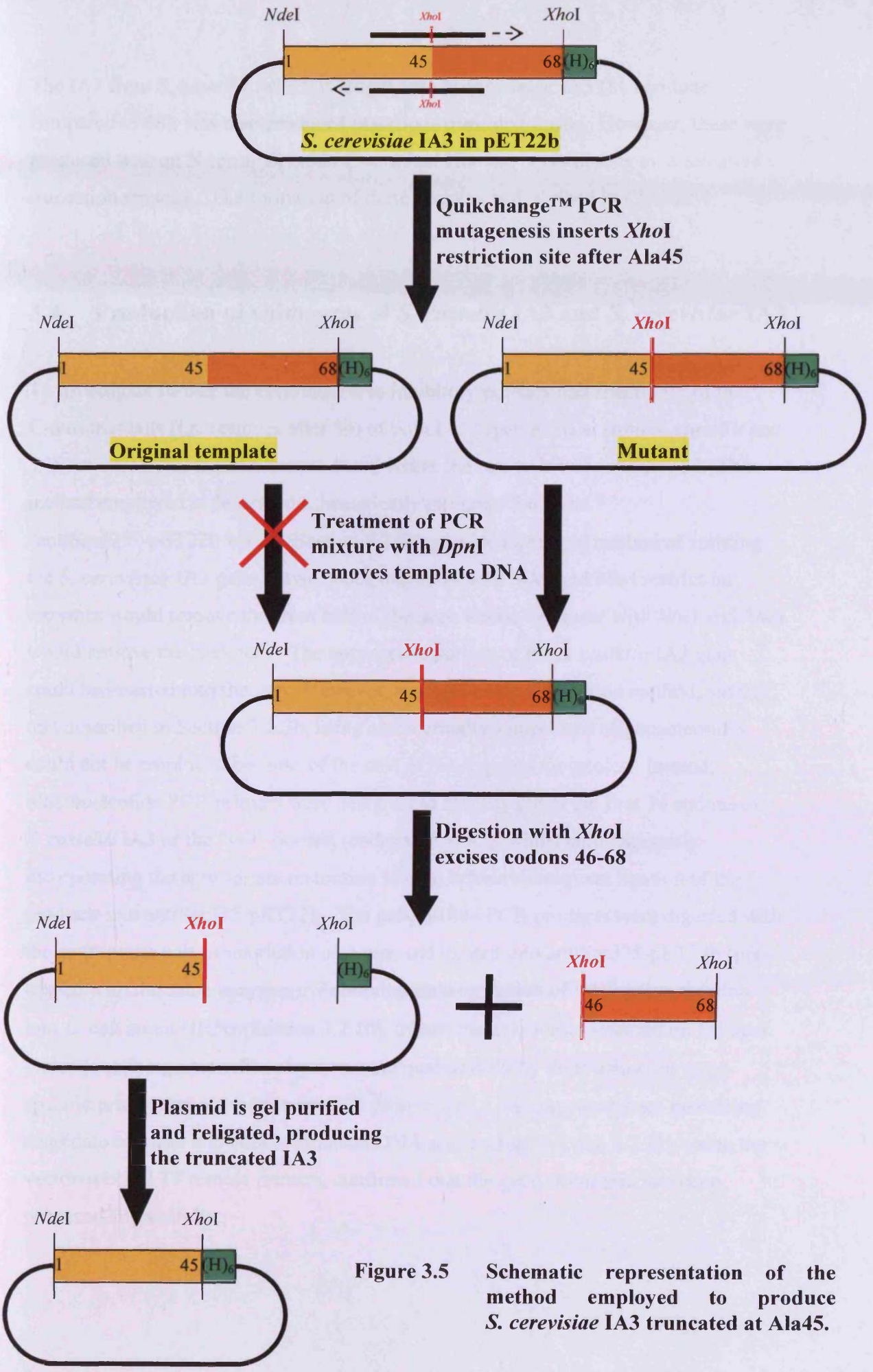


Figure 3.5 Schematic representation of the method employed to produce *S. cerevisiae* IA3 truncated at Ala45.

The IA3 from *S. castellii*, which is longer than *S. cerevisiae* IA3 (81 residues compared to 68), was also produced in various truncated forms. However, these were produced with an N-terminal rather C-terminal His-tag, necessitating an alternative truncation strategy. The formation of these proteins is described in Chapter 4.

3.4 Production of chimaeras of *S. castellii* IA3 and *S. cerevisiae* IA3

To investigate further the contribution to inhibitory potency and specificity of the C-terminal tails (i.e. residues after 34) of both IA3 "species" (i.e. from *S. castellii* and *S. cerevisiae*), chimaeras were produced where the two tails were exchanged. The method employed is described schematically in Figure 3.6. The "antiSacJ35"-pET22b vector (Section 3.2.3a) provided a simple method of splitting the *S. cerevisiae* IA3 gene in two, since digestion with *Nde*I and *Nhe*I restriction enzymes would remove the front half of the gene whilst treatment with *Nhe*I and *Xho*I would remove the back half. The appropriate portion of the *S. castellii* IA3 gene could be inserted into the gap. However, a simple cassette insertion method, such as that described in Section 3.2.3b, using commercially synthesised oligonucleotides could not be employed because of the cost of the required molecules. Instead, oligonucleotide PCR primers were designed to amplify either the first 34 codons of *S. castellii* IA3 or the "tail" portion (codons 35 to 81), whilst simultaneously incorporating the appropriate restriction sites to enable subsequent ligation of the products into antiSacJ35-pET22b. The gel-purified PCR products were digested with the appropriate pair of restriction enzymes and ligated into antiSacJ35-pET22b (pre-treated with the same enzymes). Following transformation of the ligation reaction into *E. coli* strain DH5α (Section 2.2.10), transformant colonies selected on LB agar plates containing ampicillin were screened qualitatively by PCR using one gene-specific primer and one vector-specific primer. DNA was prepared from promising candidate colonies and double-stranded DNA sequencing (Section 2.2.11), using the vector-specific T7 remote primers, confirmed that the gene chimaeras had been produced successfully.

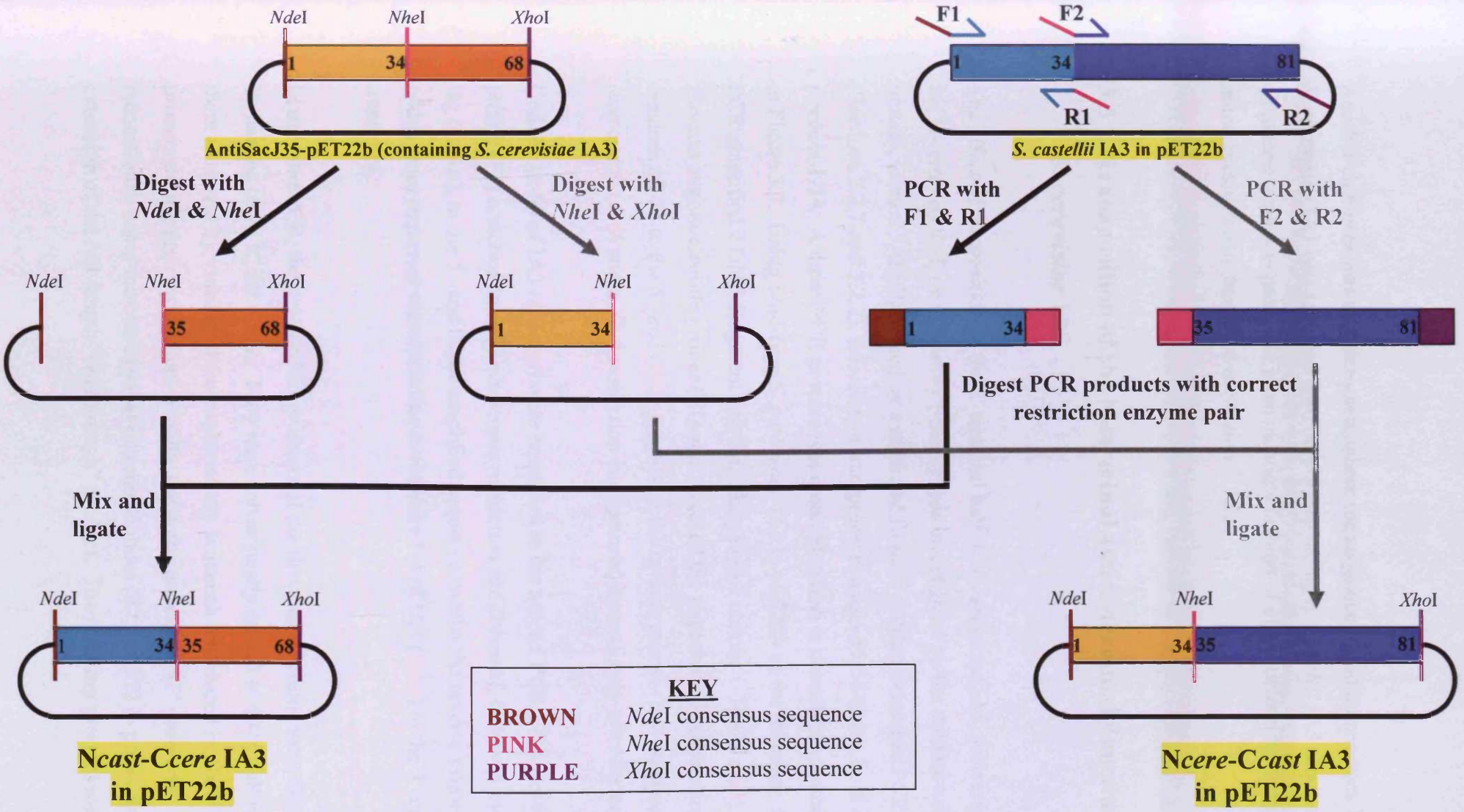


Figure 3.6 **Schematic representation of the production of the *S. cerevisiae*/*S. castellii* IA3 gene chimaeras.**

Another chimaera was also designed, where the sequence encoding residues 17-34 of *S. cerevisiae* IA3 was replaced with the corresponding *S. castellii* IA3 sequence. This chimaeric gene was produced in an identical fashion of oligonucleotide cassette mutagenesis to that described in Section 3.2.3.

3.5 Transposition of the N-terminal and C-terminal domains of *S. cerevisiae* IA3

The effect of the position of the C-terminal half of *S. cerevisiae* IA3, with respect to the N-terminal half, on inhibitory potency was investigated by the domain-swap mutant, termed 3AI. Produced by a modified form of overlap extension PCR (Sections 2.2.7 and 3.2.2), this mutant transposed residues 35-68 to the front of residues 1-34. A three-PCR process was required, which is described schematically in Figure 3.7. Using wild-type *S. cerevisiae* IA3 in pET22b as the template, the first PCR amplified a DNA fragment corresponding to IA3 codons 1-34 (in **gold**). The forward oligonucleotide primer (F1) also attached the sequence corresponding to residues 65-68 to the 5′-end (**orange**), whilst the reverse primer (R1) positioned a stop codon (**red**) and a *Xho*I restriction site (**purple**) immediately after codon 34.

Codons 35-68 of IA3 (**orange**) were amplified in the second PCR. The forward primer (F2) attached an *Nde*I consensus restriction site (**brown**), followed by a $(His)_6$-tag (**green**), to the 5′-end of the amplified sequence, whilst the reverse primer (R2) added extra sequence corresponding to residues 1-4 of IA3 (**gold**) to the 3′-end of the transcript.

In the third PCR, the gel-purified products of the first two reactions were denatured by heating to 95 °C for 5 min. They were subsequently cooled to 60 °C, allowing them to anneal by virtue of the complementary sequence introduced by the F1 and R2 primers during their production. Finally, the domain-swap "3AI" gene was reconstituted using the two external oligonucleotides (R1 and F2) to prime the extension of the full-length "readthrough" product. The resulting product was

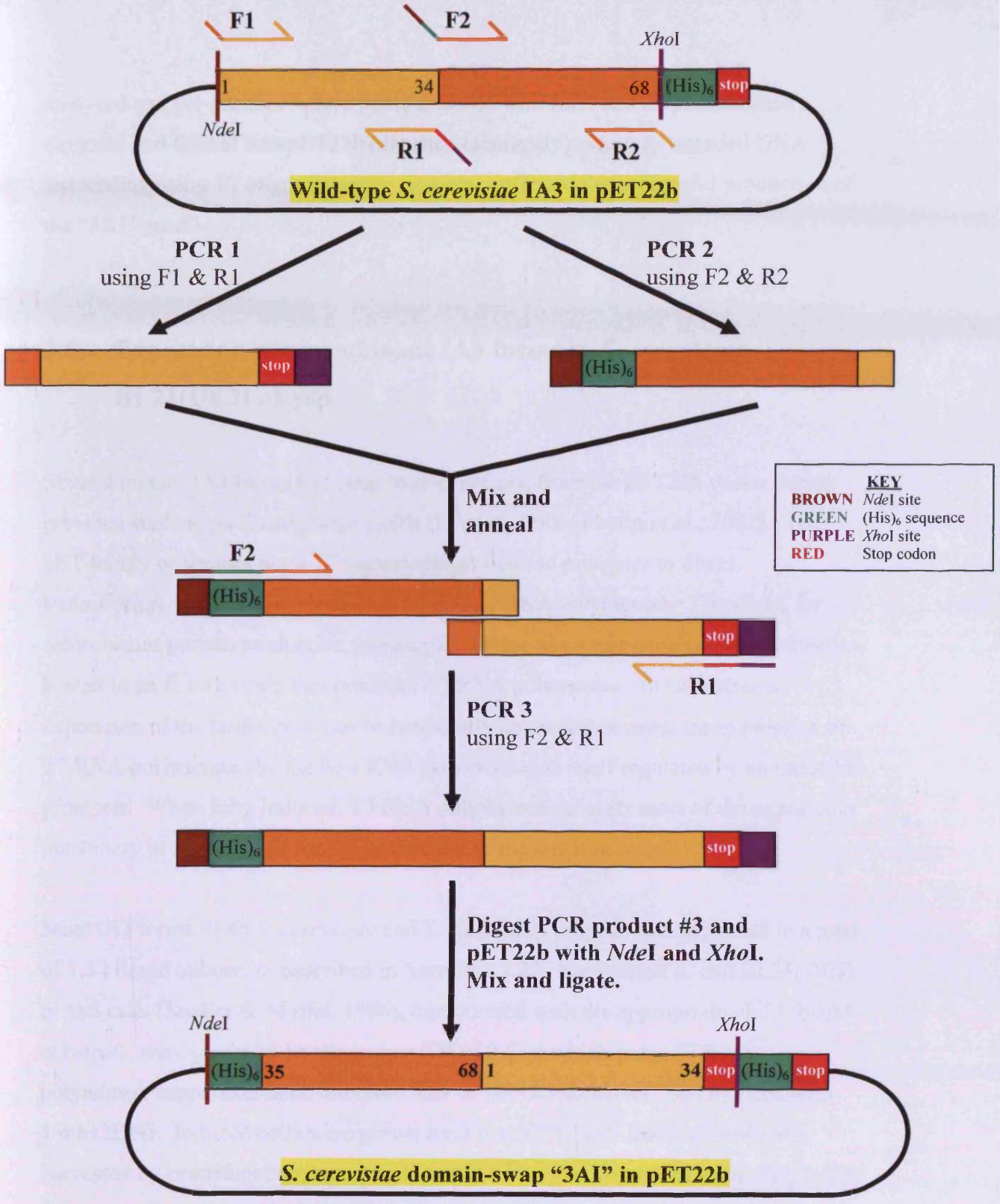


Figure 3.7 **Production of the *S. cerevisiae* domain-swap mutant "3AI" by overlap extension PCR.** Note that oligonucleotide primer F2 transfers the $(His)_6$ tag to the N-terminus of this mutant, and that primer R1 introduces a stop codon before the *Xho*I restriction site, preventing the addition of the C-terminal $(His)_6$ tag encoded by the pET22b vector. Colour coding of primers corresponds with the sequence thus attached to the PCR product.

analysed and gel-purified before being digested with *Nde*I and *Xho*I restriction enzymes and ligated into pET22b (digested identically). Double-stranded DNA sequencing using T7 oligonucleotide primers confirmed the successful production of the "3AI" gene.

3.6 Expression of recombinant IA3 forms in *E. coli* strain BL21(DE3) pLysS

Several mutant IA3 forms had been over-expressed from the pET22b vector during previous studies, producing large yields (Li *et al.*, 2000; Phylip *et al.*, 2001). The pET family of vectors use a T7 bacteriophage-derived promoter to direct transcription, which is not recognised by *E. coli* RNA polymerase. Therefore, for recombinant protein production using a pET vector, the expression construct must be hosted in an *E. coli* strain that produces T7 RNA polymerase. In such strains, expression of the target gene can be temporally controlled because the expression of T7 RNA polymerase (by the host RNA polymerase) is itself regulated by an inducible promoter. When fully induced, T7 RNA polymerase subverts most of the expression machinery in the host cell for the production of the target protein.

Most IA3 forms, both *S. cerevisiae* and *S. castellii* in origin, were expressed in a total of 1.5 l liquid culture, as described in Section 2.2.22. Expression *E. coli* BL21(DE3) pLysS cells (Studier & Moffat, 1986), transformed with the appropriate pET22b-IA3 construct, were grown to an attenuance (D) of 0.8, at which point T7 RNA polymerase expression (and, therefore, that of the IA3 construct) was induced with 1 mM IPTG. Induced cells were grown for 3 h at 37 °C with shaking (300 rpm), harvested by centrifugation, resuspended in 1/75th culture volume of sonication buffer (50 mM sodium phosphate, pH 8.0, 0.3 M NaCl) and stored frozen.

Some IA3 mutants, however, especially those containing two or more hydrophobic mutations, were produced in quite poor yields. In an effort to overcome this, some expressions were also performed in *E. coli* C41(DE3) pLysS cells (OverExpress™, Avidis SA, France). This bacterial strain, first described by Miroux & Walker (1996),

was selected specifically to improve the soluble expression of proteins containing a high proportion of hydrophobic residues. However, no significant improvement over the usual BL21(DE3) pLysS system was observed.

During the expression of the first few IA3 mutants constructed for the work described in this thesis, 1 ml samples of the culture were removed at the beginning and end of the induction, in an attempt to assess the level of induction of IA3 expression. However, interpretation by SDS-PAGE (Section 2.2.19) proved difficult because resolution of such a small protein from the large number of others of approximately the same size was not possible. Furthermore, Western blotting of the expression gel (Section 2.2.20) did not improve the resolution because the reaction of the Ni-conjugated-alkaline phosphatase with His-tagged IA3 proteins was poor, a result that had been noted previously in this laboratory (Dr. L. Phylip, unpublished data). It was decided, therefore, to continue with all subsequent purifications without first attempting to assess the level of expression.

3.7 Purification of His-tagged IA3 forms by IMAC

All recombinant IA3 forms produced in this study were expressed as hexahistidine-tag fusions. The His-tag facilitated purification from bacterial cell lysates by the ability to bind with high affinity to an immobilised metal affinity chromatography (IMAC) matrix. IMAC, first described by Porath *et al.* (1975), relies on a metal chelating agent, in this case the tetradentate nitrotriacetic acid (NTA), charged with Ni^{2+} ions (6-coordinate). The NTA is immobilised on an agarose or Superflow™ matrix and the two remaining unliganded positions of each chelated Ni^{2+} interact with the $(His)_6$-tag fused to the IA3, thus binding it to the column.

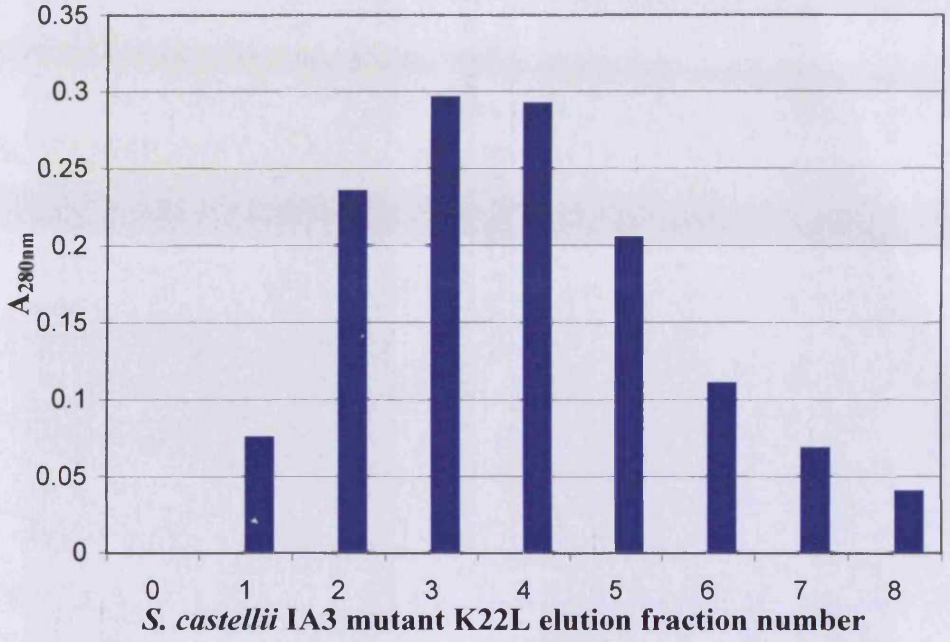
To extract the His-tagged IA3, the expression cells (suspended in sonication buffer, see Section 3.6) were lysed by repeated freeze-thaw cycles. After removing insoluble cell debris (Section 2.2.23), the soluble His-tagged IA3 was mixed with the Ni-NTA resin in a batch process for 50 min at 4 °C and applied to a sintered glass column.

After extensive washing (Section 2.2.23), the bound IA3 was eluted by adjusting the pH to 3.5, a value considerably below the pKa of histidine.

Identifying elution fractions containing IA3 proved difficult by spectrophotometric means because of the poor absorbance demonstrated by these proteins at 280 nm. The A_{280} of *S. castellii* IA3 is particularly low, because this protein contains fewer aromatic residues than *S. cerevisiae* IA3. For instance, Figure 3.8A shows the A_{280} elution profile of the *S. castellii* IA3 mutant, K22L. However, when 15 µl aliquots of each 1.5 ml fraction of this same elution were analysed by SDS-PAGE (Figure 3.8B), the preparation was shown to be both homogeneous and more concentrated than the A_{280} values might have suggested. For example, a concentration of 240 µM (approximately 2 mg ml^{-1}) was determined for fraction 3 in duplicate amino acid analyses (Section 2.2.26).

However, not all purifications of IA3 mutants resulted in such immediately homogeneous preparations. Some mutants (particularly *S. cerevisiae* IA3-derived sequences containing either two or more hydrophobic mutations or severe truncations, both of which reduced yields (Section 3.6)), eluted with many contaminant bands in the fractions analysed by SDS-PAGE (see, for example, Figure 3.9A). Possibly this reflected a combination of poor accumulation of the protein in a soluble form and an excess of Ni-NTA resin, resulting in insufficient His-tagged protein to saturate the column. Thus, non-specifically bound *E. coli* proteins consumed a large proportion of the binding capacity of the column and eluted with His-tagged IA3 at the lower pH. In such circumstances, most contaminant bands could be removed by heating the fractions twice to 100 °C for 5 min. The yield of IA3, which has little or no intrinsic structure (Green *et al.*, 2004), was not harmed significantly by this process (typically ≤15% loss of yield). However, almost all *E. coli* proteins were precipitated under such treatment and could be removed by centrifugation, resulting in a much cleaner SDS-PAGE profile (Figure 3.9B).

A



B

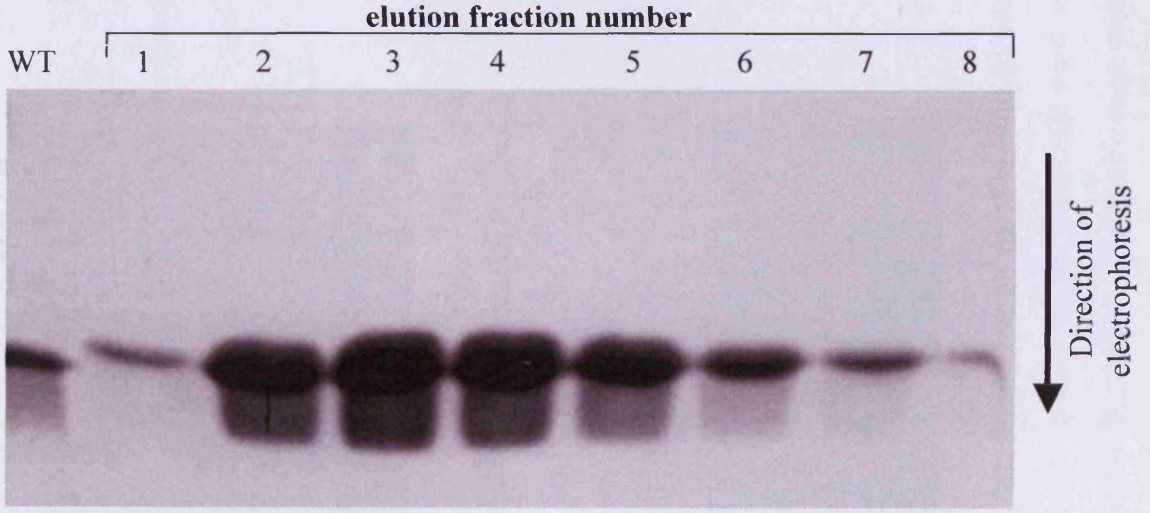


Figure 3.8 A. Elution profile for the *S. castellii* mutant IA3, K22L. The absorbances at 280 nm of 1.5 ml fractions eluted from the Ni-NTA purification column.

B. SDS-PAGE (20 % w/v acrylamide) analysis of eluted fractions. Aliquots (15 µl) of each fraction were electrophoresed, using wild-type *S. castellii* IA3 ("WT" in figure) as a marker.

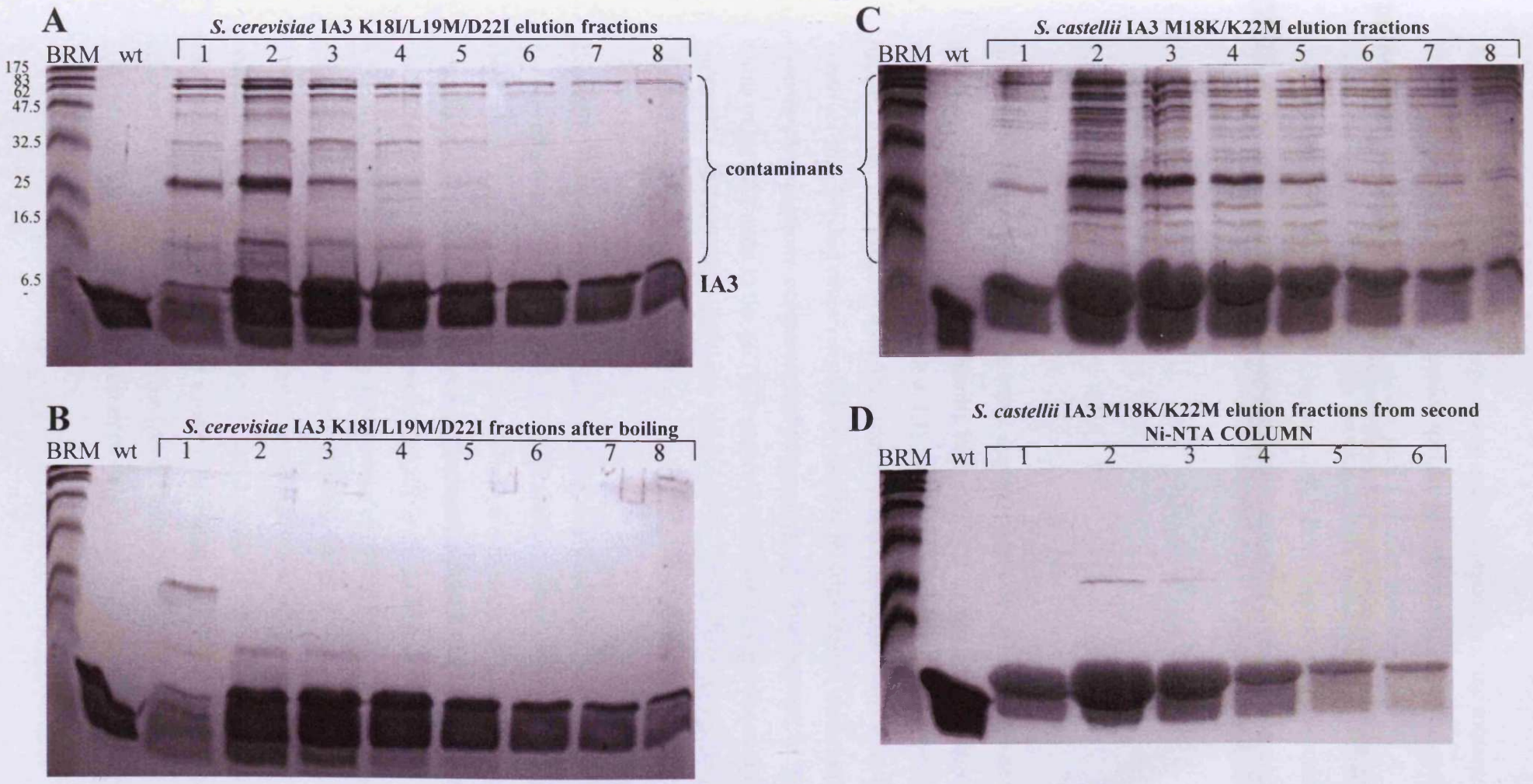


Figure 3.9 SDS-PAGE (20 % w/v acrylamide) analysis of elution fractions from Ni-NTA purification columns. Elution fractions from the Ni-NTA column purification of *S. cerevisiae* IA3 mutant K18I/L19M/D22I were analysed (15 µl) before (**A**) and after (**B**) boiling twice for 5 min. **C.** Elution fractions from the *S. castellii* IA3 (M18K/K22M) Ni-NTA purification were subjected to a second IMAC step on a smaller Ni-NTA column (**D**). "BRM" refers to New England Biolabs broad range protein marker (Section 2.1.6, sizes to the left of gel **A** are in kDa) and "wt" is wild-type *S. cerevisiae* IA3.

An alternative re-purification process was also employed for some mutants. Fractions from the original column were pooled and the pH adjusted to approximately pH 7 with NaOH. The pool was then mixed with a smaller volume of Ni-NTA Superflow™ resin in a similar batch process as before and the purification repeated. As can be seen by comparison of Figures 3.9C (first column elution fractions) and 3.9D (fractions from the second column), this method of contaminant removal was similarly effective.

3.8 Discussion

In total, during the course of this work, approximately 40 different IA3-derived proteins were produced recombinantly in *E. coli* in C-terminally His-tagged form (see complete listing in Appendices II & III). A variety of manipulation methods were applied to the IA3 genes originating from *S. cerevisiae* and *S. castellii*, in order to construct site-directed mutations of individual or multiple amino acid residues, C-terminal truncations, chimaeras and a domain-swap. Furthermore, in the process, forms of the IA3 gene in the pET22b expression vector were produced (for example antiSacJ35-pET22b) that will make future genetic manipulations to generate almost any conceivable mutant form of IA3 a much simpler proposition.

Generally, the yield of each IA3 mutant following purification by IMAC from a 1.5 l *E. coli* culture was >20 mg. Even when the original IMAC purification step did not yield a homogeneous preparation, the IA3 protein could be purified further by either re-application to another IMAC step or, through an exploitation of the unstructured, heat stable properties of these proteins (Phylip *et al.*, 2001), by boiling the sample. It was noticeable, however, that the introduction of hydrophobic mutations into the centrepiece region of the *S. cerevisiae* IA3 sequence often limited the yield. Typically, *S. cerevisiae* IA3 mutants containing two or more hydrophobic replacements at positions 18, 19 and 22 had low yields, although some individual residue changes appeared to have a more substantial effect than others. For example, whereas the yield of the *S. cerevisiae* IA3 mutant where Asp22 was mutated to Met (D22M) was almost identical to that of the wild type protein, the mutant with the

same residue instead changed to Leu (D22L) produced just 2% of the yield of the wild-type, from the same culture volume.

In notable contrast, the introduction of the equivalent mutation into *S. castellii* IA3 (i.e. Lys22 to Leu, K22L) did not reduce yield with respect to the wild-type. This is despite the fact that such a mutation results in a centrepiece combination of Met18 and Leu22, which in an *S. cerevisiae* IA3 sequence context (K18M/D22L) had resulted in one of the lowest yields (<1% of the wild-type). Therefore, although hydrophobicity is evidently a contributory factor in some way, other features within the sequence must also play an influential role.

A common feature of recombinant protein over-expression in *E. coli*, especially proteins of eukaryotic origin, is that many are produced in insoluble form, often as a result of mis-folding. This is frequently exacerbated by particularly hydrophobic elements of the expressed sequence. In free solution, *S. cerevisiae* IA3 is believed to be essentially unstructured (Phylip *et al.*, 2001; Green *et al.*, 2004), so it seems intrinsically unlikely that the protein would simply mis-fold. Furthermore, *S. cerevisiae* IA3 exhibits the typical characteristics of an unfolded protein in terms of amino acid composition as well (Tompa, 2002). At pH 7.0, *S. cerevisiae* IA3 has 39.7% charged residues and only 22% hydrophobic ones. The introduction of one hydrophobic residue in place of a charge one would not appear to upset this balance substantially. Moreover, the values for *S. castellii* IA3 are very similar: 39.5 % and 27% for charged and hydrophobic residues, respectively. Therefore, although the sequences of *S. cerevisiae* and *S. castellii* IA3 are quite different (Chapter 6), they are both very polar and small (<90 amino acid residues). In consequence, it seems unlikely that the introduction of a single, specific, hydrophobic residue substitution into *S. cerevisiae* IA3 would disturb the hydrophilic nature of the sequence to such an extent that the protein is rendered insoluble, particularly when an essentially similar protein (i.e the equivalent *S. castellii* IA3 mutant) is unaffected.

Nevertheless, the possibility existed that *S. cerevisiae* IA3-derived mutants containing hydrophobic substitutions were largely in the insoluble fraction after expression. For example, diseases such as sickle cell anaemia are caused by a single amino acid substitution leading to mis-folding or aggregation of proteins. However, the technical

difficulties (Section 3.6) of tracing recombinant IA3 proteins before the IMAC purification step prevented reaching a conclusive answer to this possibility. Expression using a strain of *E. coli* (Miroux & Walker, 1996) designed specifically to express highly hydrophobic sequences, such as membrane proteins, in soluble form did not improve the yield. One interpretation of this would be that the low yield of IA3 proteins containing hydrophobic substitutions from these cells cannot, therefore, simply reflect reduced solubility.

An alternative explanation is that the *S. cerevisiae* IA3 sequences containing certain hydrophobic mutations become more susceptible to cleavage by the host cell endopeptidases than the *S. castellii* IA3 sequences expressed with essentially the same mutations. The introduction of specific hydrophobic substitutions into this region of the *S. cerevisiae* IA3 sequence may also simultaneously produce a cleavage site for the host cell endopeptidases. In contrast, the small differences between the *S. castellii* and *S. cerevisiae* IA3 sequences in this region may mean that *S. castellii* IA3 remains spared from cleavage upon introduction of the same mutation(s). If this were the case, the *S. castellii* IA3 form would accumulate to normal levels in the expression cells, whilst the introduction of the equivalent mutations would lead to destruction of the recombinant *S. cerevisiae* IA3 form. Although this is a hypothetical situation, it appears to exhibit several parallels with the experimentally-observed reality.

Ultimately, however, all C-terminally His-tagged forms of IA3, even those with multiple hydrophobic mutations, were purified to homogeneity in more than sufficient quantities for the kinetic spectrophotometric assays described in this thesis. The next chapter describes the production of several N-terminally His-tagged and tagless forms of both *S. cerevisiae* and *S. castellii* IA3.

Chapter 4

His-tag removal from recombinant IA3 proteins

4.1 Introduction

All the recombinant IA3 proteins in this study were expressed as a histidine tag (His-tag) fusion to simplify their purification (see Chapter 3). Previous research conducted in this laboratory (Li *et al.*, 2000; Phylip *et al.*, 2001) had shown that C-terminally His-tagged recombinant wild-type *S. cerevisiae* IA3 had a very similar K_i value at pH 3.1 (1.1 nM) to the naturally occurring inhibitor (Dreyer *et al.*, 1985) when tested against *S. cerevisiae* vacuolar proteinase A. Thus, it had been concluded that the C-terminal His-tag had no significant effect on inhibitory potency.

However, during the course of the studies described later in this thesis, a series of results cast doubt upon the established dogma of the inert nature of both the C-terminal half of the protein and the histidine tag attached to it. To distinguish effects potentially arising from the C-terminal half of the protein itself from any caused by the His-tag, efforts were made to produce several recombinant tagless forms of IA3. However, due to the considerable simplification of the purification process (see Section 3.7), it was desirable to continue to express the proteins with a His-tag attached, and remove it after the IMAC purification step.

Two potential methods for His-tag excision were identified. The first relied on the introduction of an endoproteolytic cleavage site in front of the C-terminal His-tag. The second, more complicated process, required the redesign of both the pET22b vector and the IA3 constructs contained within it, in order to relocate the His-tag to the N-terminus. From here the tag could be removed by exoproteolytic cleavage.

Ultimately, both avenues were investigated. Therefore, this chapter describes not only the attempts to remove the C-terminal His-tag from IA3 proteins, but also the design, production and use of an N-terminal His-tag system that permitted subsequent His-tag removal.

4.2. Removal of the C-terminal His-tag from recombinant IA3 by endoproteolytic cleavage with clostripain

There are several well-established systems for the endoproteolytic removal of C-terminal tags. An example protocol, using the endopeptidase Factor Xa, is illustrated in Figure 4.1. Typically, an endopeptidase recognition sequence is introduced immediately before the desired cut site by site-directed mutagenesis. The native protein, purified as a His-tag fusion bound to an IMAC column, can then be released from the Ni-NTA resin by treatment with endopeptidase, leaving only the tag bound to the column. However, this results in the new C-terminal residues of the protein being those of the artificially-introduced endopeptidase recognition sequence (see Figure 4.1). Therefore, a typical endoproteolytic cleavage protocol was not appropriate because the purpose of His-tag removal in this project was to remove *all* non-native protein sequence.

However, inspection of the amino acid sequence of *S. castellii* IA3 (see Figure 4.2A) reveals two Arg residues (in red) in the C-terminal half, at positions 61 and 72. Clostripain, also called endoproteinase Arg-C, is a cysteine proteinase from *Clostridium histolyticum* that specifically cleaves peptide bonds on the C-terminal side of Arg residues (Labrou & Rigden, 2004). The two Arg residues therefore constituted endogenous clostripain cleavage sites and, theoretically, enabled the removal of the His-tag without the introduction of any exogenous sequence. Such a cleavage would also result in a moderate truncation of the C-terminus. However, this was not considered to be of concern because the activity of the molecule was localised to the first 34 residues (see later chapters).

4.2.1 Optimisation of clostripain digestion parameters for removal of the C-terminal His-tag from recombinant *S. castellii* IA3

The clostripain digestion conditions (see Section 2.2.25) were optimised in a series of trials. In the first, approximately 1 nmol of purified recombinant wild-type *S. castellii* IA3 was incubated with 20 pmol of enzyme at 37 °C for 1 h in a total volume of 100 µl. The progress of the digestion was monitored at 0, 10, 20 and

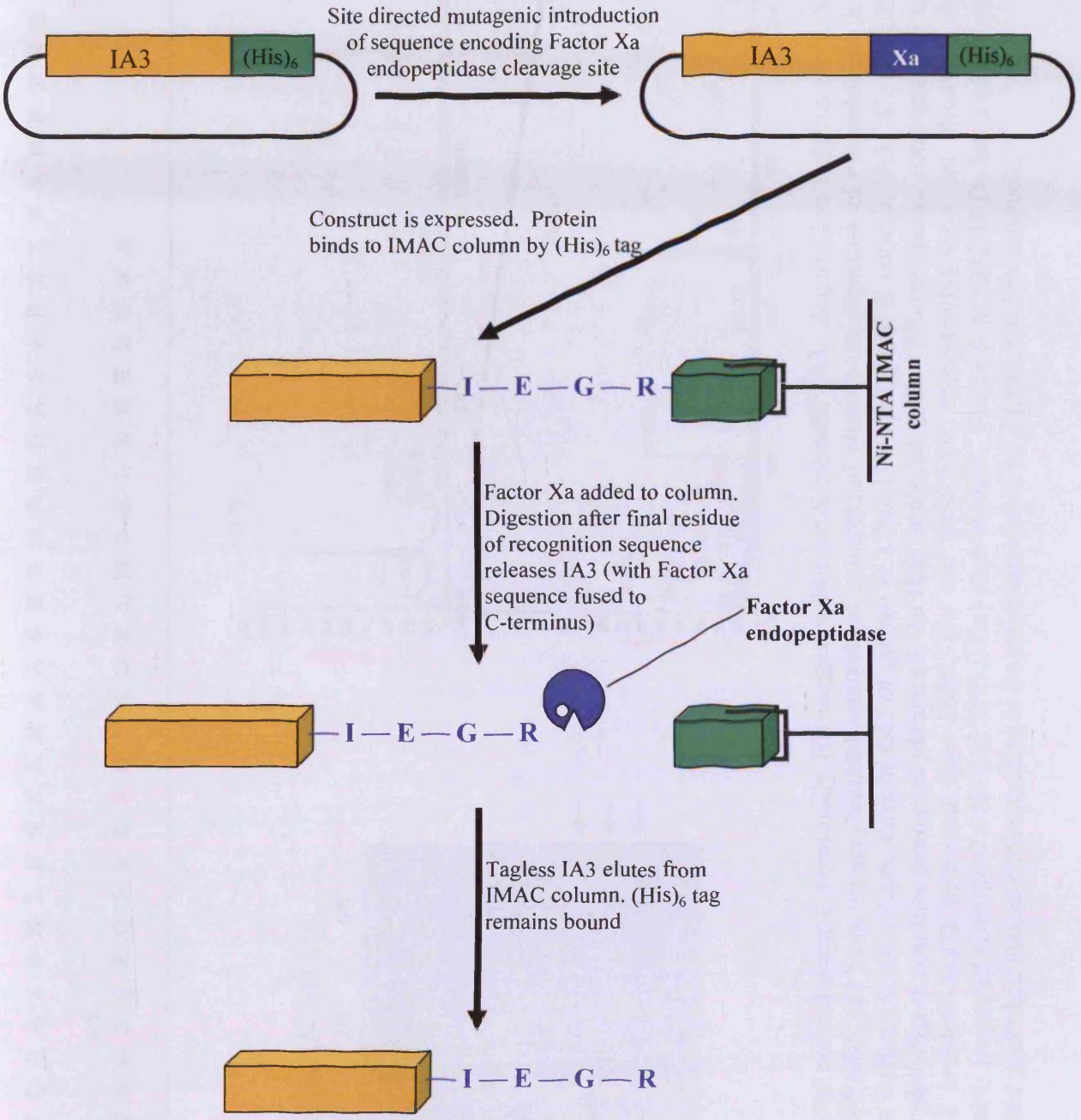


Figure 4.1 Schematic representation of how a C-terminal His-tag could be removed from an IA3 construct by the introduction of an endopeptidase (Factor Xa) cleavage sequence.

A

```
1         5         10        15        20        25        30        35        40        45        50
M S D K N A N V S E M F Q Q A K E M L E G K A N A A S E G M K N M A S P E S K E K M K T K G Q D M E

          55        60        65        70        75        80
K K A Q D S Y N K L R G A G K Q E Q T K E R Q D K D E L M D H L E H H H H H H
```

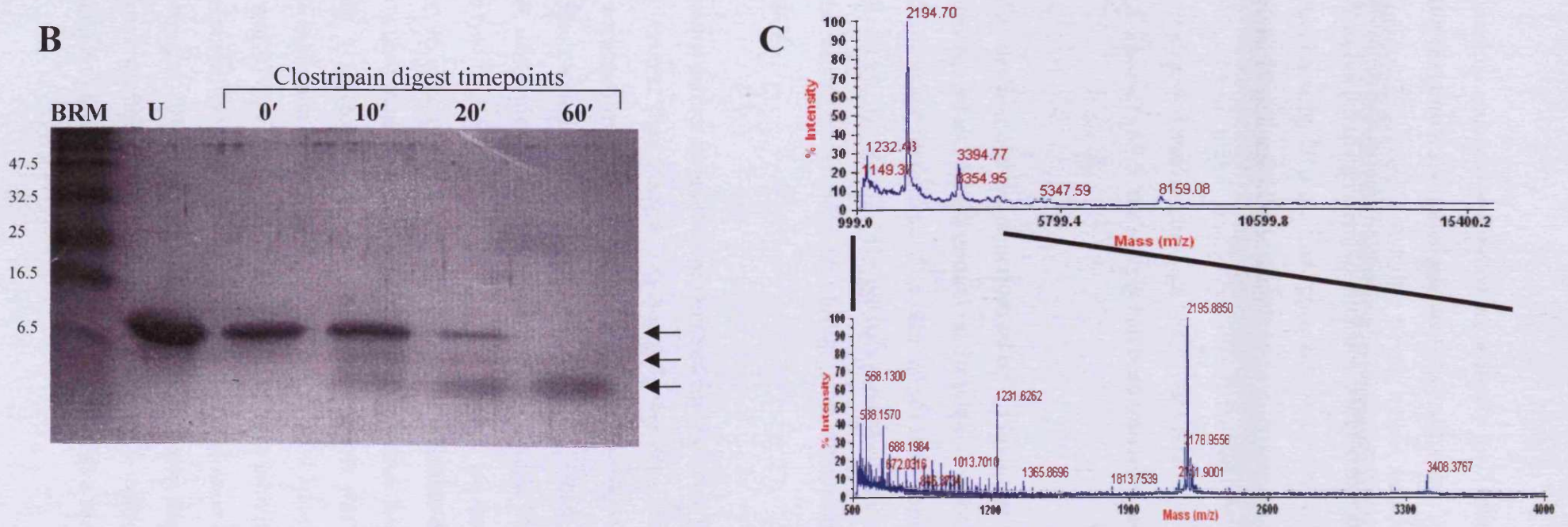


Figure 4.2A. Primary sequence of recombinant C-terminally His-tagged wild-type *S. castellii* IA3. Arg and Lys residues are highlighted in red and green, respectively. **B. SDS-PAGE (20 % (w/v) acrylamide) analysis of progress of clostripain digestion of recombinant *S. castellii* IA3 with time.** BRM = broad range marker (Section 2.1.6, sizes to side of gel are in kDa), U = undigested *S. castellii* IA3. **C. MALDI-TOF mass spectrometric analysis of a 40 min digest reaction prepared identically to that analysed in B.** Whereas peaks corresponding to N-terminal cleavage fragments Met1-Arg61 or Met1-Arg72 (expected sizes 6856.3 Da and 8068.9 Da, respectively) were not observed, signals for the corresponding C-terminal fragments (expected sizes 3407.6 Da and 2195.0 Da) were evident. The peak at 1232.48 Da has a mass corresponding to that of Gly62-Arg72. Several other peaks had masses corresponding to fragments produced by digestion at Lys residues.

60 min with 20 µl samples removed and boiled immediately for 7 min in 1 × SDS protein sample buffer (Section 2.2.19) to inactivate the clostripain. Analysis by SDS-PAGE (see Figure 4.2B) revealed that after 10 min a new, smaller band was apparent already. This band had increased in intensity, with a commensurate decrease in the starting product band, by 20 min. The appearance of a weak band of intermediate size in the 10 min and 20 min samples was interpreted to represent cleavage after Arg72, with the major, smaller band being the product of cleavage after Arg61. At 1 h, the undigested and intermediate bands were no longer visible, suggesting that all *S. castellii* IA3 in the sample had been truncated to the shorter form (1-61).

A second, essentially identical digest was performed to: (i) optimise the incubation period; (ii) investigate the effect of the digestion on the inhibitory potency of the IA3; and, (iii) identify the cleavage product(s). Samples (20 µl) were removed at 0, 10 and 40 min for analysis by SDS-PAGE. The gel (not shown) confirmed the disappearance of the starting material band by 40 min, with the appearance of a single smaller band as before.

At the same timepoints, further samples were removed for use in enzyme kinetic spectrophotometric assays. Upon removal, each sample was immediately diluted 5-fold in 100 mM sodium formate buffer, pH 3.1 and heated to 100 °C for 7 min. This process was designed to inactivate the clostripain rapidly, which has a narrow pH optimum of 7.6-7.9, whilst not causing the unstructured *S. castellii* IA3 any significant damage (see Section 3.7). The samples were tested for their ability to inhibit *S. cerevisiae* vacuolar proteinase A at pH 4.7, as described in Section 2.2.30. The 0 min timepoint determined the "maximal" level of inhibition for a given volume of the digestion mix. A suitable volume of diluted digestion mix was used to produce approximately 80% inhibition of the *S. cerevisiae* vacuolar proteinase A. Equal volumes of the 10 and 40 min samples were also tested and the inhibition at all three timepoints was compared. Notably, by 10 min the level of inhibition had fallen to 51.5 % and by 40 min just 30 % inhibition was observed. Clearly, the digestion process had a detrimental effect on the inhibitory potency of the molecule. As described in Section 4.2, it seemed unlikely that this would be as a result of a partial

shortening of the C-terminus. However, the observation was consistent with cleavage of *S. castellii* IA3 in a more critical region, i.e. a region *not* containing any target Arg residues. Further research into the original papers (Mitchell & Harrington, 1968) revealed that clostripain is also capable of minor levels of hydrolysis after Lys residues, of which *S. castellii* IA3 contains 14 (green in Figure 4.2A). Furthermore, four of these are found within the first 34 residues, the region thought to be responsible for the inhibitory potency of the molecule. In order to identify if cleavage at Lys residues was indeed responsible for the loss of potency from the clostripain-treated *S. castellii* IA3, the completed (40 min) digest was analysed by MALDI-TOF mass spectrometry (Section 2.2.28).

The spectrum, shown in linear and expanded views in Figure 4.2C, did not contain peaks of the expected masses for N-terminal fragments from digestion at either of the Arg residues (6856.3 Da (6725.2 Da without the initiator Met) and 8068.9 Da (7937.8 Da). However, the observed peaks of mass 3408.4 Da and 2195.9 Da coincided exactly with the expected masses of C-terminal fragments produced by such cleavages, and the peak with mass 1232.5 Da corresponded with the fragment between the two Arg residues (expected mass 1231.6 Da). However, several other peaks had masses that matched products from cleavage at Lys residues. Thus, it appeared that although cleavage was occurring primarily at the Arg residues, *S. castellii* IA3 was also being cleaved slowly elsewhere in the sequence.

4.2.2 Clostripain treatment of wild-type *S. cerevisiae* IA3

In parallel with the MALDI-TOF mass spectrometric tests described in Section 4.2.1, the possibility that clostripain was cleaving at Lys residues was also investigated by performing an identical digestion with *S. cerevisiae* IA3. As *S. cerevisiae* IA3 contains no Arg residues, it could not be cut during the incubation, unless clostripain was also capable of cleaving elsewhere within the sequence (presumably at any of the 13 Lys residues). Again, a sample of inhibitor was incubated with clostripain at a molar ratio of 50:1 for 40 min at 37 °C. Samples were removed, and the clostripain rapidly inactivated, for analysis by SDS-PAGE and enzyme kinetic

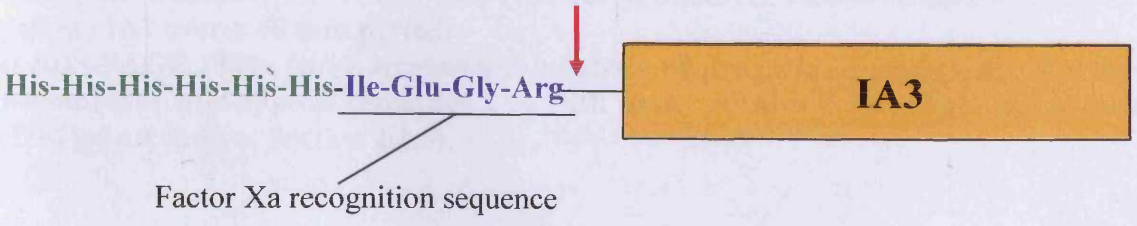
spectrophotometric assays as before, at the start and end of the incubation, together with an intermediate timepoint of 10 min.

To enable comparison of the kinetic inhibition data between the digestions of *S. castellii* and *S. cerevisiae* IA3 forms, the percentage inhibition levels for both 0 min samples were normalised to "100%", with the respective 10 and 40 min samples being adjusted as a percentage of their respective control value (Figure 4.3A). Whilst the loss in potency at the end of the timecourse was not as severe in *S. cerevisiae* IA3 (72% activity remaining as opposed to 39% in *S. castellii* IA3), it certainly suggested that the inhibitor was being digested, despite the lack of Arg residues. Indeed, this conclusion was subsequently confirmed by the SDS-PAGE result (Figure 4.3B), where cleavage was observed.

After this clear indication of cleavage in a sequence devoid of Arg, combined with the strong mass spectrometric evidence of digestion at Lys residues in *S. castellii* IA3, clostripain treatment was no longer considered as a suitable method for His-tag removal.

4.3 A change of tack: moving the His-tag to the N-terminus

After the complications of the clostripain digestions, other avenues of His-tag removal were sought. Investigations into various systems demonstrated that the simplest methods of completely removing the His-tag from a protein, to leave only the native protein sequence, require the tag to be located at the N-terminus rather than the C-terminus. As illustrated in Figure 4.1, most endopeptidases cleave to the C-terminal side of their recognition sequence. Therefore, in contrast to its use for removal of a C-terminal tag, the introduction of such an endopeptidase recognition sequence between an N-terminal His-tag and the IA3 sequence would leave only natural sequence after the cleavage process:



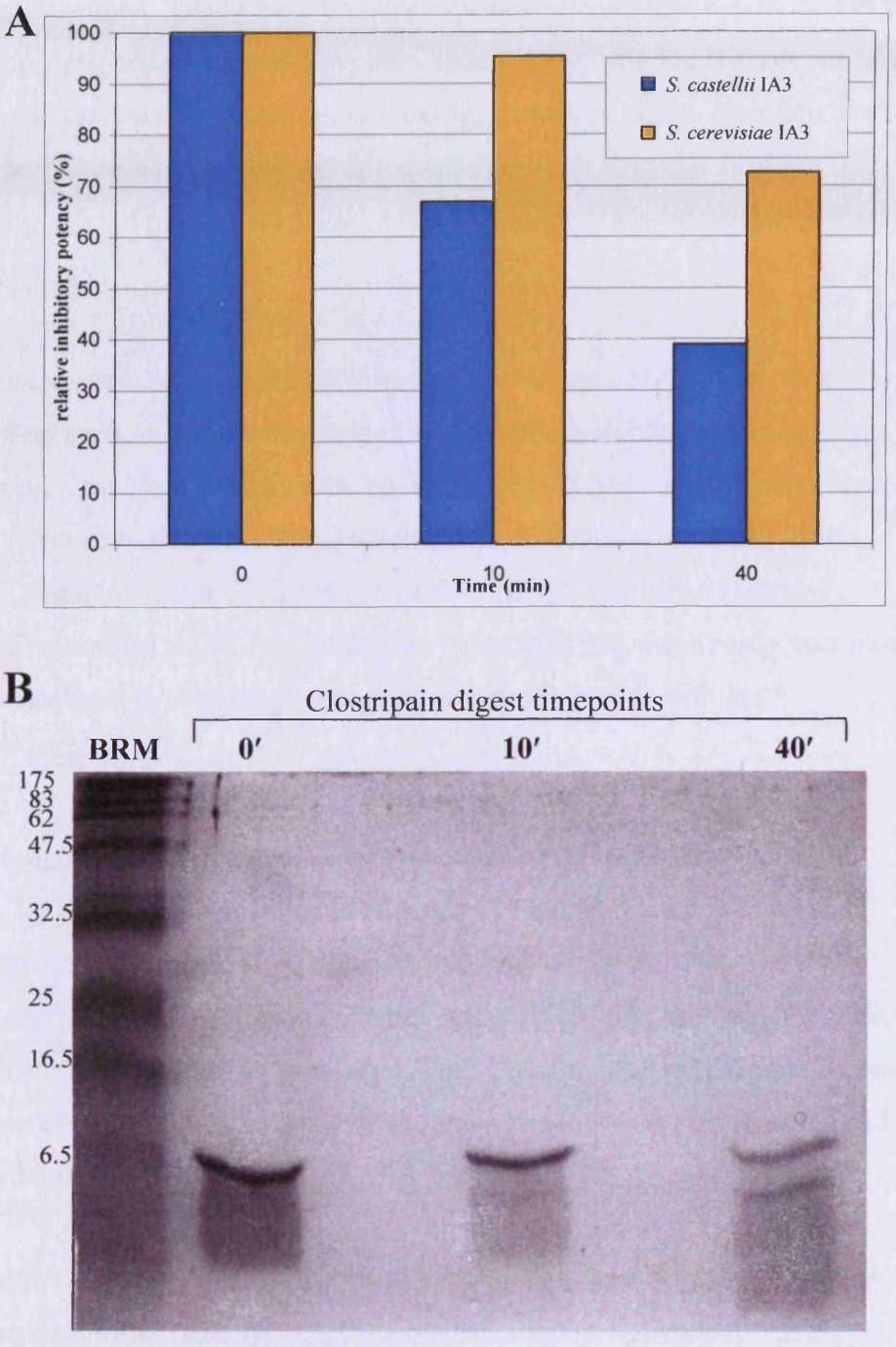


Figure 4.3A. Comparison of the relative inhibitory activity remaining in identical clostripain treatments of either wild-type recombinant *S. castellii* (blue) or *S. cerevisiae* (yellow) IA3 over a 40 min period.
B. SDS-PAGE (20% (w/v) acrylamide) analysis of progress of clostripain digestion of recombinant wild-type *S. cerevisiae* IA3 with time. BRM = broad range marker (sizes to left of gel are in kDa, Section 2.1.6).

Other systems, such as TAGZyme™ (first described by Pedersen *et al.*, 1999), had considerable attractions. Using the diamino exopeptidase cathepsin C (DAPase™), this protocol is designed to remove dipeptides sequentially from the N-terminus of the protein. Certain residues are refractory to cleavage, however, so the digestion can be halted at a specific point. Thus, by the introduction of such a residue between the His-tag and the start of the IA3 sequence, a His-tag can be removed whilst leaving the IA3 sequence intact.

The use of an exopeptidase was particularly attractive in light of the work with clostripain (Section 4.2), because the risk of non-specific endoproteolytic cleavage was eliminated. Therefore, a method based on the TAGZyme™ system was pursued. Expression vectors compatible with the system are available commercially (for instance the Qiagen plasmids TAGZyme™ pQE-1 and -2). However, because expression of recombinant IA3 forms from the vector pET22b was already optimised, it was instead preferable to redesign this vector to be compatible with the TAGZyme™ system.

4.3.1 Redesigning the pET22b vector to produce an N-terminal His-tag

Several considerations entered into the redesign of the vector, pET22b.
Firstly, because DAPase™ sequentially removes dipeptide units from the N-terminus, the tag had to contain an even number of residues. Also, all the residue pairs had to be compatible with rapid cleavage by DAPase™, although a stop point was needed immediately before the IA3 sequence.

As described in Chapter 3, all IA3 sequences were cloned into pET22b between *Nde*I and *Xho*I restriction sites, with the *Nde*I site contributing the translation initiating Met codon (see Figures 3.4A and 4.4A). It was desirable to transfer such a well-established system into the new context. However, because the sequence encoding the His-tag had to be introduced upstream of the IA3 gene, the Met encoded by the *Nde*I site could no longer serve as the translation initiator. Therefore, the region of the plasmid upstream of the *Nde*I site required redesign so as to introduce an initiator Met as well as the His-tag, whilst still maintaining the other elements of the original

sequence, such as the ribosome binding site. Furthermore, with the initiation now upstream of the cloning site, the dipeptide encoded by the *Nde*I sequence would also have to be removed. Fortuitously, this sequence - CATATG - encodes His-Met, a dipeptide that is compatible with DAPase™ cleavage.

One further complication remained. The expression cells, *E. coli* BL21(DE3)pLysS, contain an aminopeptidase that often removes the N-terminal Met residue from IA3 forms (Phylip *et al.*, 2001). Crucially, however, this is not always the case. For example, 6 cycles of Edman N-terminal sequencing (Section 2.2.21) of the *S. cerevisiae* IA3 mutant K18I/L19M/D22I revealed the sequence **M-N-T-D-Q-Q**, clearly demonstrating that the N-terminal Met of this particular mutant was present. As an even number of residues is required in the N-terminal tag in order to be certain of reaching the DAPase™ digestion stop-point "in-frame", the state of the initiator Met, be it present *or* absent, was important. The Qiagen TAGZyme™ vectors circumvent this problem by initiating translation with a Met followed by a Lys immediately upstream of the His-tag. As described by Hirel *et al.* (1989) and Dalbøge *et al.* (1990), a Lys at this position minimises processing of the initiator Met *in vivo*. Furthermore, in the event that the N-terminal Met residue is still removed by the endogenous *E. coli* endopeptidase, the Lys residue thus exposed is utterly refractory to DAPase™ cleavage. Any such "uncleavable" IA3 could then be removed by a subtractive IMAC step after the cleavage incubation.

In summary, therefore, the region to be introduced upstream of the *Nde*I cloning site had to encode Met-Lys-$(His)_6$ and position this initiator Met appropriately with respect to the ribosome binding site. With the need for such large changes to the vector, the simplest method appeared to be the excision of the relevant portion of the vector using appropriate restriction enzymes, followed by ligation of a replacement oligonucleotide cassette containing the required new sequence.

Inspection of the cloning region of pET22b (Figure 4.4A) revealed that the first unique restriction site upstream of the *Nde*I site (in **brown**) is *Xba*I (**red**). Two complementary artificially synthesised oligonucleotides (termed newpetTOP and newpetBOT) were designed so that, when annealed, they reconstituted the region (including the "sticky ends" produced by the digestion) between these two restriction

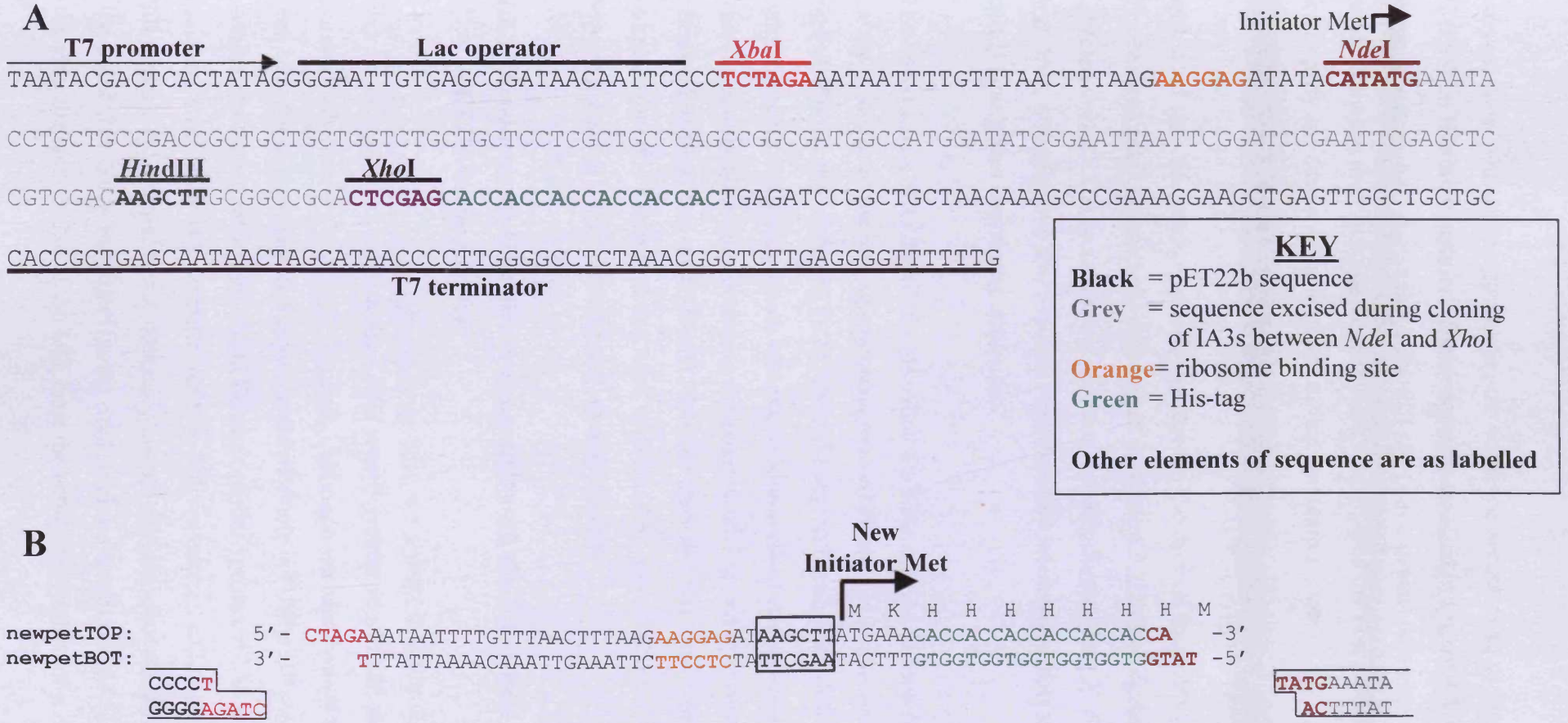


Figure 4.4A Representation of the cloning region of the expression plasmid vector pET22b. All IA3s were cloned between the *Nde*I (brown) and *Xho*I (purple) restriction sites, resulting in the replacement of the intervening vector sequence (in grey), which included a *Hin*dIII site (dark green). **B. Representation of the artificially synthesised oligonucleotides designed to replace the natural sequence between the *Xba*I site (red) and the *Nde*I site (brown) in the pET22b vector.** The oligonucleotides are represented as a cassette, annealed by their complementary sequence, producing appropriate "sticky ends" for ligation into suitably digested pET22b (represented in figure). The cassette also encoded a new *Hin*dIII site (boxed dark green).

enzyme sites, whilst also introducing the sequence necessary to produce the N-terminal His-tag construct. Their sequence, detailed in Figure 4.4B (and Appendix II.F), also shows that a *Hin*dIII site (**dark green**) was engineered into the region between the ribosome binding site (**orange**) and the new initiator Met residue, to simplify any future manipulations of this upstream region.

Therefore, the pET22b vector, containing wild-type *S. cerevisiae* IA3, was digested with *Xba*I and *Nde*I and gel purified to remove the excised fragment (data not shown). The cassette was constructed as described in Section 3.2.3b and ligated into the digested vector. The reconstituted plasmid was transformed into *E. coli* strain DH5α, and DNA was prepared from several transformants selected by their ability to grow on LB agar plates containing ampicillin.

The introduction of a *Hin*dIII site provided a simple method of screening preparations for the presence of the new upstream region. As illustrated in Figure 4.4A, the only natural *Hin*dIII site within pET22b lies in the region between the *Nde*I and *Xho*I sites (grey sequence in figure), which was removed as a result of cloning the IA3 gene. Therefore, only plasmids containing the successfully ligated upstream insert were linearised by treatment with *Hin*dIII (data not shown). The identities of positive samples from this qualitative test were confirmed by double-stranded DNA sequencing using the T7 forward and reverse primers.

4.3.2 Producing IA3 constructs compatible with the N-terminal His-tag pET22b vector by PCR

In order to be compatible with the TAGZyme™ system, modifications to the IA3 constructs themselves were also required. Although the Met-Lys-$(His)_6$ sequence had been introduced upstream of the *Nde*I restriction site, a DAPase™ stop site had to be introduced downstream of this point because neither species of IA3 encoded an intrinsic "stop" with its N-terminal residue. Furthermore, the C-terminal His-tag was still present. Therefore, it was necessary not only to introduce an artificial DAPase™ stop point after the N-terminal His-tag (and *Nde*I sequence) but also a translational stop after the last residue of the IA3, be it the actual last residue or a desired

C-terminal truncation point (see Section 3.3). Therefore, oligonucleotide primers were designed to attach the required sequence to the 5′- and 3′- ends of both *S. cerevisiae* and *S. castellii* (and chimaera) IA3 constructs during a PCR. A schematic representation of the process, from the initial PCR to the eventual production of a tagless IA3, is provided in Figure 4.5. Specific sequence details of the primers involved are provided in Appendix II.F, and the process is described below.

The forward PCR primers introduced an *Nde*I consensus sequence (**brown**), followed by a DAPase™ stop residue – Gln (**pink**) – before entering the appropriate IA3 sequence at codon 2. Meanwhile, the reverse primers encoded stop codons (**red**) after the final IA3 codon (in all three frames, in case of ribosome slippage), followed by the *Xho*I site (**purple**). All PCRs were performed with the high-fidelity DNA polymerase *PfuUltra*™ according to the conditions described in Section 2.2.7. After analysis and purification by 2 % (w/v) agarose gel electrophoresis and treatment with *Nde*I and *Xho*I, the PCR product was ligated into the N-terminal His-tag pET22b vector, which had been treated identically. Double stranded DNA sequencing using the T7 forward and reverse primers confirmed that all constructs correctly encoded a tag of M-K-$(H)_6$-H-M-Q before the second residue of the appropriate IA3, with a stop codon immediately after the desired final IA3 codon and *before* the *Xho*I site and C-terminal His-tag.

4.3.3 Production of tagless forms of IA3

The conditions for expression and purification of IA3 forms with an N-terminal tag were exactly the same as for the C-terminally tagged proteins described in Sections 3.6 and 3.7. It was noted that the N-terminally tagged proteins tended to produce superior yields to their C-terminal counterparts. After purification by IMAC (see Sections 2.2.23 and 3.7), elution fractions containing the N-terminally tagged IA3 were pooled. All N-terminally His-tagged proteins had their composition confirmed and their concentration accurately determined by amino acid analysis (Section 2.2.26). All were tested against the same range of aspartic proteinase enzymes as the C-terminally tagged IA3 forms (see Chapters 5-10).

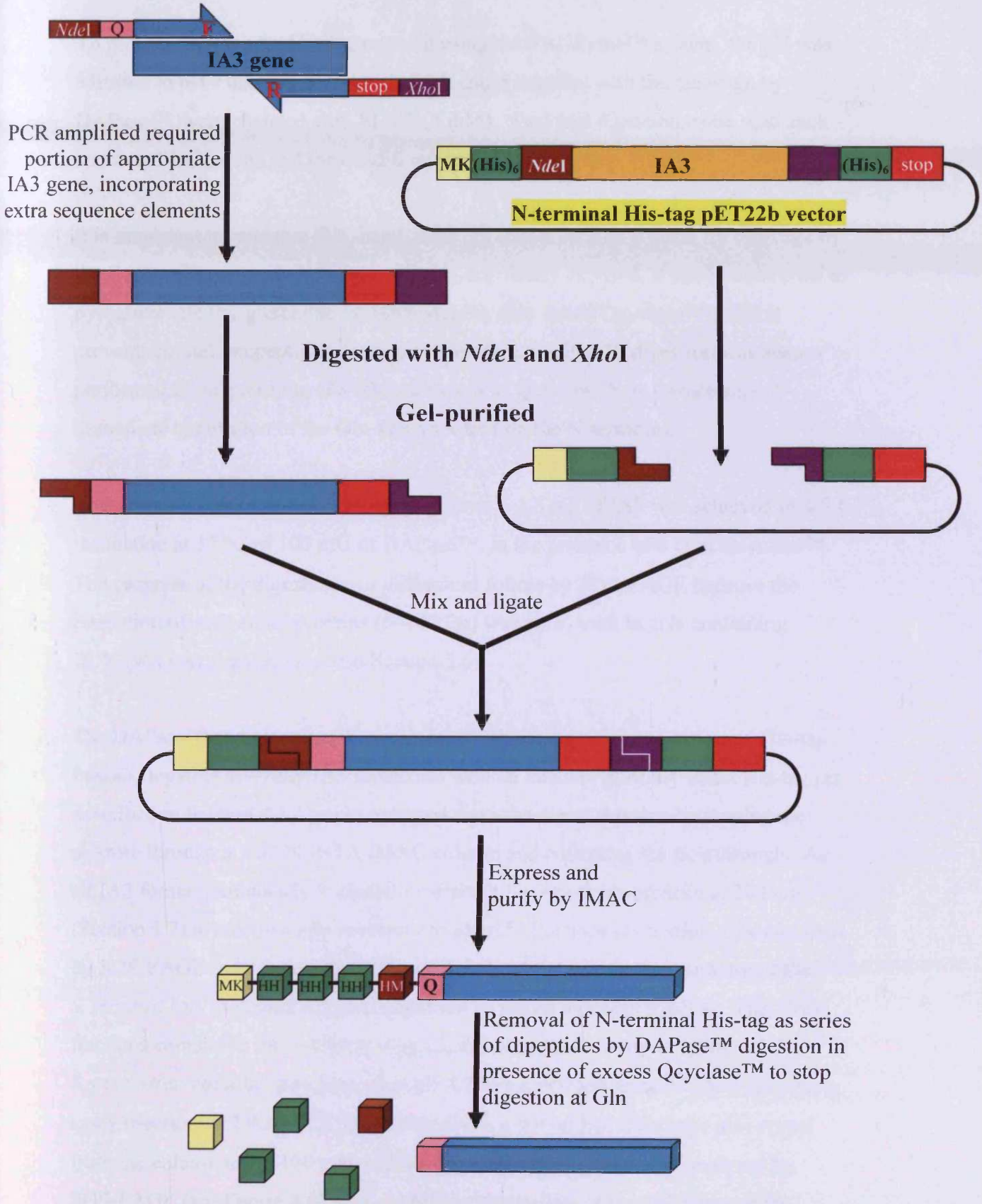


Figure 4.5 Preparation of IA3 genes for introduction into the converted N-terminal His-tag vector. The desired region of IA3 is amplified by PCR, with oligonucleotide primers incorporating the necessary extra sequence, including the cloning restriction sites, DAPase™/Qcyclase™ stop residue (Gln, in pink) and the translation stop codon (red) immediately after the last residue of IA3 sequence, to prevent expression of C-terminal His-tag after cloning of fragment into the vector.

To prepare samples for His-tag removal using the TAGZyme™ system, the pH was adjusted to pH 7 and any Ni^{2+} ions, which could interfere with the cleavage by DAPase™, were chelated with EDTA (5 mM). Two trial digestions, one with each species of IA3, identified the conditions described in Section 2.2.24.

It is important to note that Gln, itself, does not constitute a stop point for cleavage by the DAPase™ enzyme. However, when N-terminally exposed, it can be converted to pyroglutamate (by glutamine cyclotransferase, also called Qcyclase™), which prevents further exopeptidic cleavage. Therefore, DAPase™ digestion was always performed in the presence of a 60-fold excess of Qcyclase™, to ensure almost immediate cyclisation of the Gln as it appeared on the N-terminus.

Thus, removal of the N-terminal His-tag from 1–1.5 mg of IA3 was achieved by a 2 h incubation at 37 °C of 100 mU of DAPase™, in the presence of 6 U of Qcyclase™. The progress of the digestion was difficult to follow by SDS-PAGE because the resolution of such small proteins (6–10 kDa) was poor, even in gels containing 20 % (w/v) acrylamide (see also Section 3.6).

The DAPase™ and Qcyclase™ enzymes, which are supplied as C-terminal His-tag fusions, together with any IA3 molecules with an intact or partially intact His-tag (as described in Section 4.3.1) were removed from the digest reaction by passing the mixture through a 1 ml Ni-NTA IMAC column and collecting the flowthrough. As all IA3 forms (particularly *S. castellii* variants), have poor absorbance at 280 nm (Section 3.7) it was typically necessary to identify fractions containing unbound IA3 by SDS-PAGE (see, for example, Figure 4.6A). However, the tagless form of the *S. castellii* IA3 truncated at Lys45 could not be traced by either method. Therefore, fractions containing this inhibitor were identified by their ability to inhibit *S. cerevisiae* vacuolar proteinase A at pH 4.7 in a spectrophotometric enzyme kinetic assay (Sections 2.2.9 and 2.2.30). Generally, 6 × 0.9 ml fractions were also eluted from the column using 100 mM sodium formate buffer, pH 3.1, and analysed by SDS-PAGE (see Figure 4.6B). A qualitative assessment of the efficiency of the digestion could therefore be made by comparison of the intensity of the IA3 bands in the bound (His-tag remaining) and unbound (His-tagless) fractions.

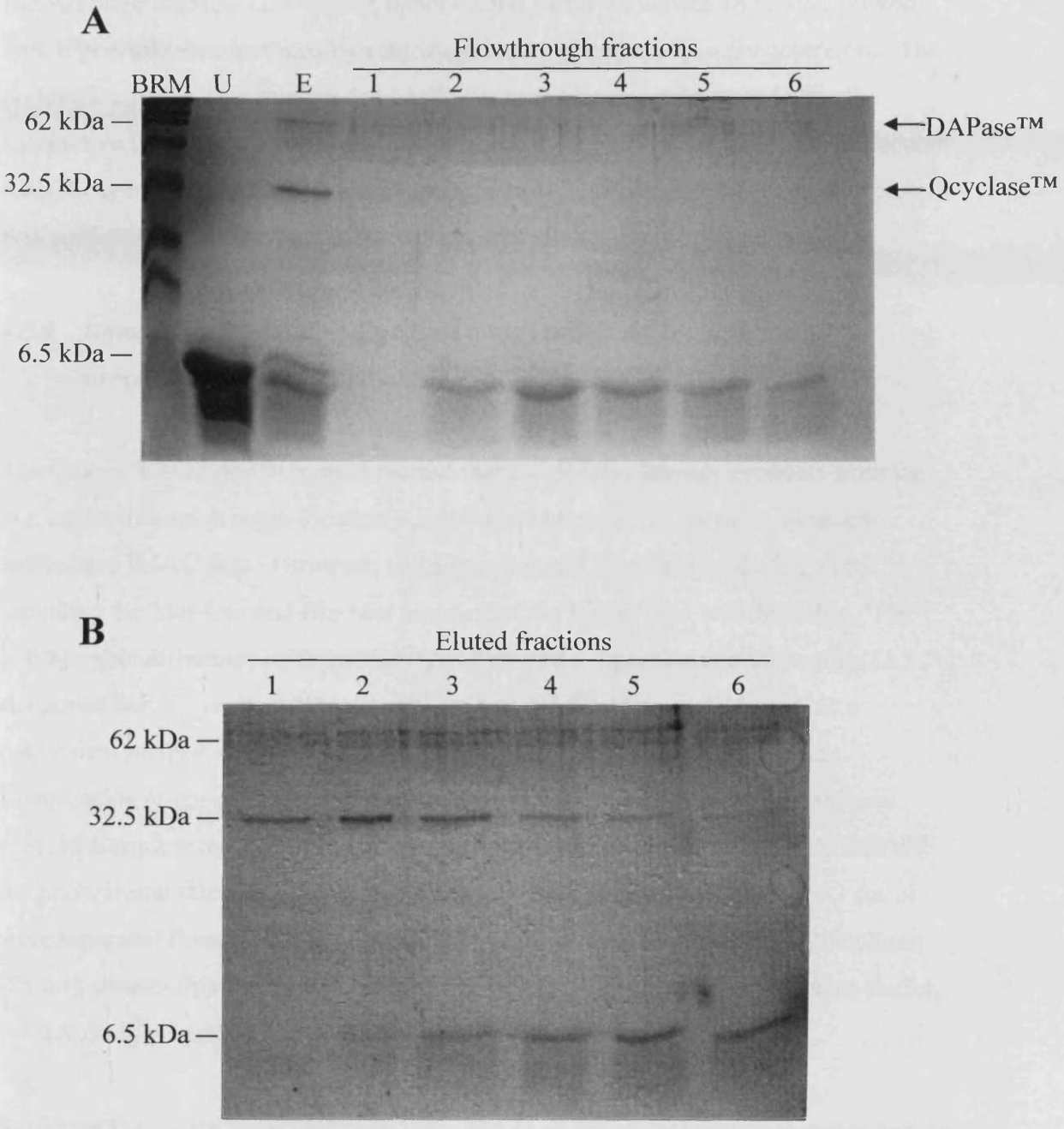


Figure 4.6 SDS-PAGE (20 % (w/v) acrylamide) analysis of fractions resulting from treatment of the *S. castellii*(1-34)/*S. cerevisiae*(35-68) IA3 chimaera with DAPase™ and Qcyclase™. A. Flow-through fractions 1-6. **B.** Fractions eluted by 100 mM sodium formate buffer, pH 3.1. BRM= broad range marker (Section 2.1.6). U= undigested N-terminally His-tagged chimaera IA3. E= end point of digestion, note the two high molecular weight bands corresponding to DAPase™ (52 kDa) and Qcyclase™ (33.5 kDa).

Flowthrough fractions, containing unbound and therefore tagless IA3, were pooled and, if possible, concentrated by centrifugation in Vivaspin2 spin concentrators. The tagless *S. castellii* IA3 truncated at Lys45 could not be re-concentrated after the subtractive IMAC step because it was too small to be retained by the 3,000 molecular weight cut-off membrane in the spin concentrator. Fortuitously, however, this protein was sufficiently concentrated for all subsequent kinetic assay applications.

4.3.4 Removal of remaining dipeptide fragments from the tagless IA3 preparation by gel filtration

The Qiagen TAGZyme™ protocol stated that the di-His cleavage products from the His-tag restriction process (Section 4.3.3) would be removed by the subsequent subtractive IMAC step. However, complete removal of all dipeptide fragments, including the Met-Lys and His-Met fragments (see Figure 4.5) was desirable. The considerable difference in molecular mass between a dipeptide and the tagless IA3 suggested that a size-exclusion system, such as gel filtration, would provide a convenient method of separation. Sephadex G-25 (Amersham Biosciences; fractionation range of 1000-5000 Da), packed into a column of sufficient volume enabled complete resolution of the void volume fractions (containing IA3) from the column volume fractions (containing the dipeptides). Therefore, tagless IA3 forms were separated from remaining dipeptide fragments using a Sephadex G-25 column (28 ml), pre-equilibrated in column running buffer (25 mM sodium phosphate buffer, pH 6.5, 50 mM NaCl).

Sufficient 0.9 ml fractions (36) were collected from the gel filtration column to ensure representation of the full bed volume. Attempts were made to identify those fractions containing tagless IA3 by measuring the absorbance of fractions at 280 nm. However, as described in Section 3.7, the paucity of aromatic residues, particularly in the *S. castellii* IA3 sequence, combined with the diluting effect of the gel filtration column, made precise identification difficult. For example, the absorbance elution profile for the *S. cerevisiae* (2-34)/*S. castellii* (35-81) IA3 chimaera is shown in Figure 4.7A. Aliquots (15 µl) of candidate fractions across absorbance peaks were analysed by SDS-PAGE as before, in order to trace the tagless IA3 (Figure 4.7B).

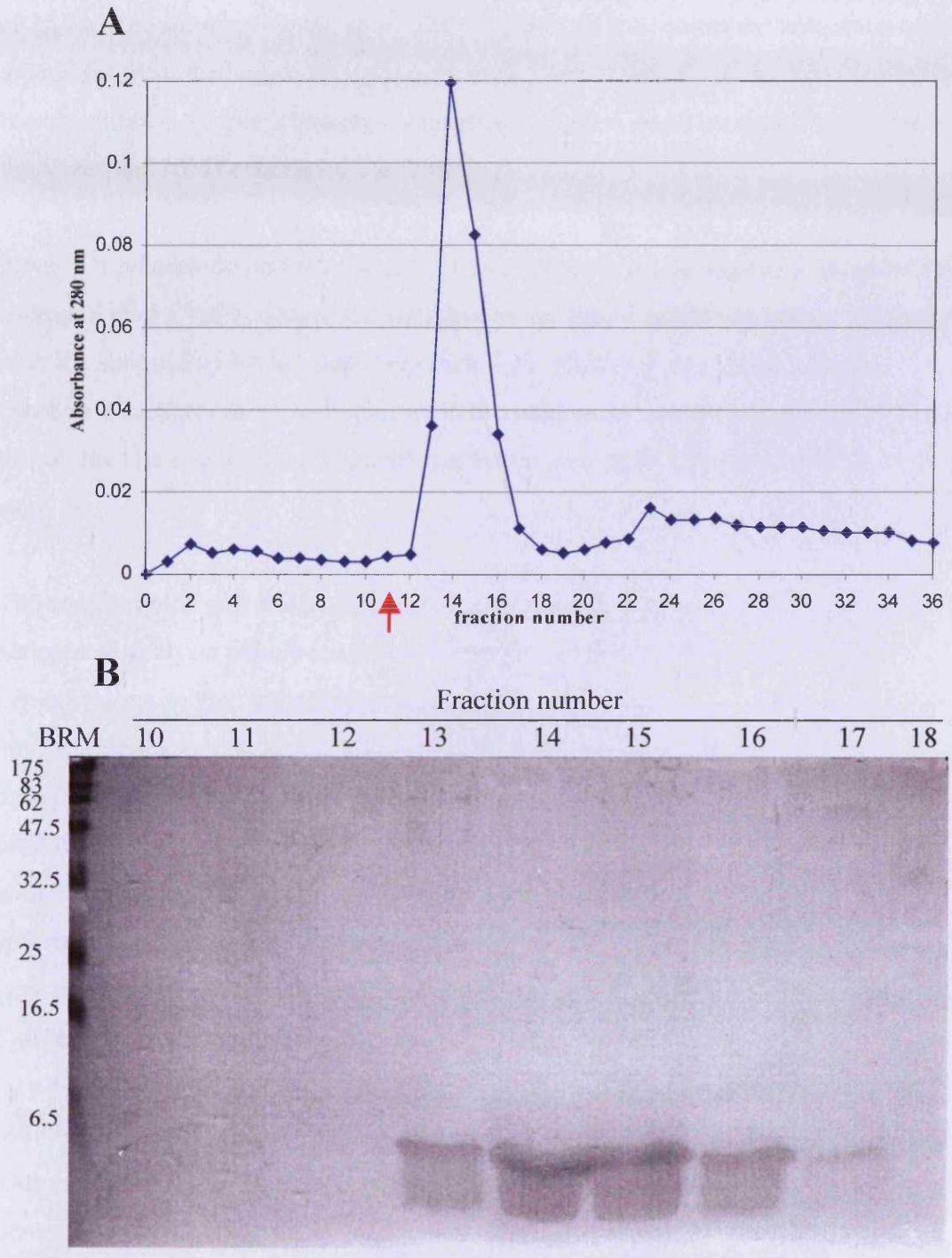


Figure 4.7 Gel filtration of tagless *S. castellii*(1-34)/*S. cerevisiae*(35-68) IA3 chimaera on Sephadex G-25. Pooled fractions from the subtractive IMAC step were applied to a 28 ml Sephadex G-25 column pre-equilibrated into 25 mM sodium phosphate buffer, pH 6.5, 50 mM sodium chloride. The red arrow marks the approximate position of the void volume of the column. Fractions (0.9 ml) were collected and analysed (**A**) by measuring their absorbance at 280 nm and by (**B**) SDS-PAGE (20 % (w/v) acrylamide). Fractions 13-17 (inclusive) were pooled.

As suggested by the elution profile, the tagless IA3 forms generally passed through the column slightly after the void volume, perhaps reflecting the unstructured nature of IA3 in free solution (Green *et al.*, 2004). Nevertheless, complete separation of the dipeptides from the tagless IA3 was achieved. Appropriate fractions were pooled, and re-concentrated to approximately the original sample volume by centrifugation using the Vivaspin2 spin concentrators as before.

It was not possible to perform the gel filtration process on the tagless *S. castellii* IA3 truncated after Lys45 because the sample was too dilute and in too large a volume after the subtractive IMAC step (Section 4.3.3). However, amino acid analysis (Section 2.2.26) revealed no histidines in the preparation whatsoever. This confirmed that all the His-containing dipeptides had been removed by the subtractive IMAC step.

The concentration and composition of all tagless IA3 preparations was determined by amino acid analysis (Section 2.2.26). The concentration of the tagless inhibitors demonstrating sufficient potency was also determined by titration against a known amount of enzyme (Section 2.2.31). Finally, MALDI-TOF mass spectrometry was used to confirm the successful removal of the His-tag. In each case the major mass peak matched the expected mass of a fragment corresponding to the appropriate IA3 with the initiator Met residue replaced by a pyroglutamate. As an example, the spectra for the *S. castellii* IA3 truncated at Lys45 are provided in Figure 4.8. The peak corresponding to the N-terminally His-tagged protein (6327.18 Da) in Figure 4.8A was missing from the digested sample in Figure 4.8B, and was replaced by a peak at 4966.71 Da, which corresponds to the expected mass (4967.3 Da) for residues 2-45 of *S. castellii* IA3, plus the N-terminal pyroglutamate and an associated sodium ion from the buffer (a common artefact of mass spectrometry). Thus, His-tag removal had been achieved.

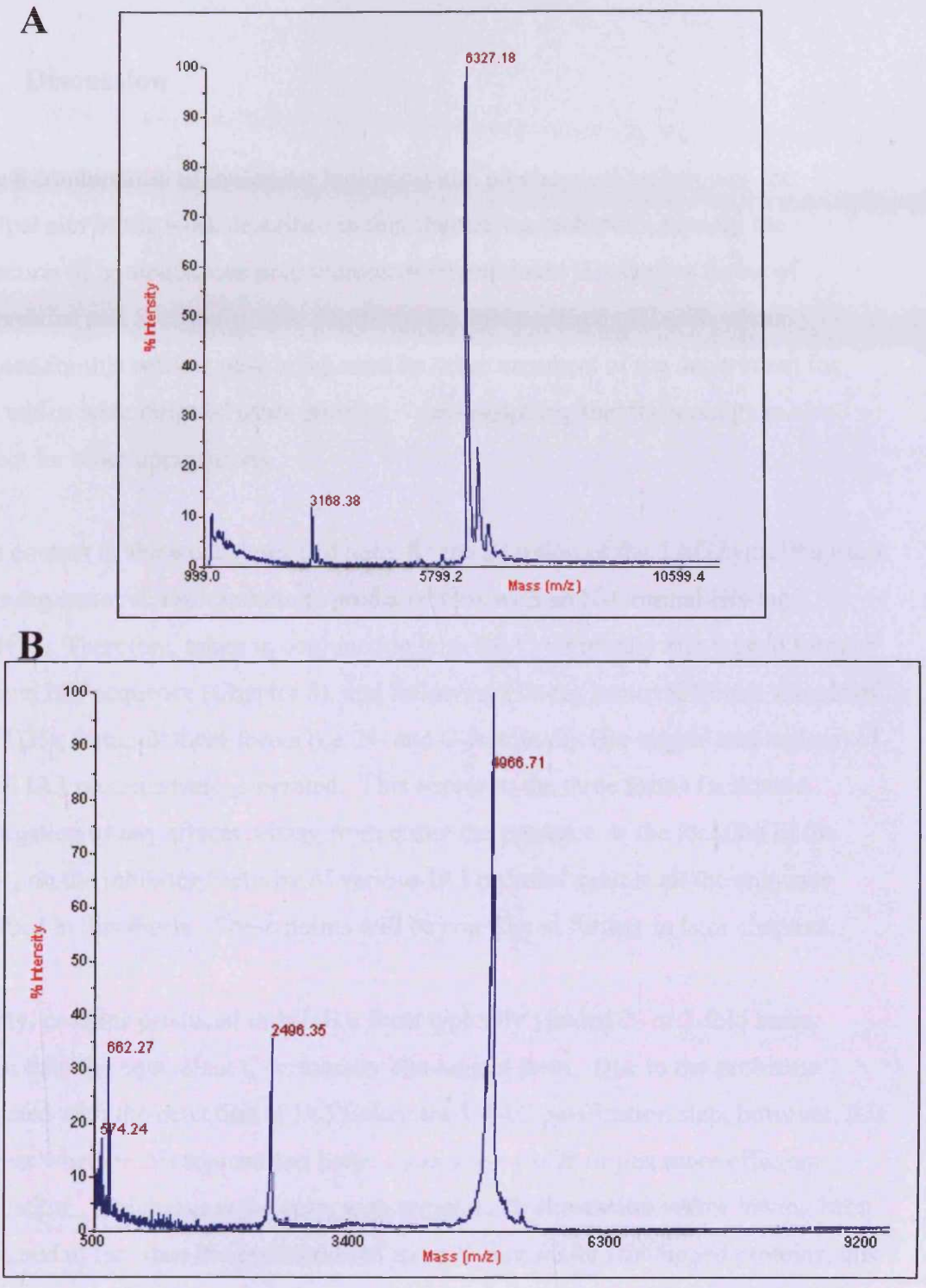


Figure 4.8. MALDI-TOF mass spectrometric analysis of *S. castellii* IA3 truncated at Lys45 before (A) and after (B) removal of the N-terminal His-tag by DAPase™ and Qcyclase™ treatment. The mass peak at 6327.18 Da (in **A**) corresponds to the expected mass (6311 Da) plus an associated sodium ion. The 3168.38 Da mass peak is probably the doubly-charged ion. The mass of the major peak in **B** corresponds with that expected for residues 2-45 of *S. castellii* IA3 plus pyroglutamate and an associated sodium ion (4967.3 Da). The peak at 2496.35 Da is probably the doubly-charged ion. No starting product is apparent.

4.4 Discussion

Using a combination of molecular biological and biochemical techniques, the principal aim of the work described in this chapter was achieved, namely the production of homogeneous preparations of recombinant His-tagless forms of *S. cerevisiae* and *S. castellii* IA3. Furthermore, the modified pET-22b vector designed for this work is now being used by other members of the department for work with a wide range of other proteins – demonstrating that the strategy is also suitable for other applications.

In the context of the work described here, by the adoption of the TAGZyme™ system of His-tag removal, each protein is produced first with an N-terminal His-tag ($NT(H)_6$). Therefore, taken in conjunction with the C-terminally His-tagged form of the same IA3 sequence (Chapter 3), and following His-tag removal from a sample of the $NT(H)_6$ form, all three forms (i.e. N- and C-terminally His-tagged and tagless) of several IA3 proteins were generated. This access to the three forms facilitated investigation of any effects arising from either the presence *or* the location of the His-tag on the inhibitory activity of various IA3 proteins against all the enzymes described in this thesis. These points will be considered further in later chapters.

Notably, proteins produced in $NT(H)_6$ form typically yielded 2- or 3-fold more protein than the equivalent C-terminally His-tagged form. Due to the problems associated with the detection of IA3 before the IMAC purification step, however, it is not clear whether this represented better expression yields or just more efficient purification. Whichever is the case, with the pET22b expression vector having been redesigned to facilitate the production of more N-terminally His-tagged proteins, this system may in the future provide a simple method of improving the yield of the problematic IA3 mutants, such as the *S. cerevisiae* IA3 constructs containing hydrophobic substitutions (Chapter 3).

Chapter 5

The inhibitory potency of S. *cerevisiae* IA3 for *S. cerevisiae* vacuolar proteinase A

5.1 Introduction

To increase our understanding of the basis of its inhibitory potency and specificity towards *S. cerevisiae* vacuolar proteinase A, an extensive collection of mutant forms of *S. cerevisiae* IA3 was constructed. These forms were designed, in particular, to investigate more precisely the primary locations of the activity, and the relative contributions of the four regions of the sequence identified by the X-ray crystal structure of Li *et al.* (2000) as critical to potency, i.e. the three hydrophobic "pins" and the so-called "centrepiece" (Section 1.8.5 and 1.8.6, respectively). The potency of all mutant inhibitors was determined in kinetic spectrophotometric assays at pH 4.7, as described in Section 2.2.30. Often, however, the potencies of the molecules tested were so high as to be beyond accurate determination with our experimental methodology. Therefore, the inhibitors were also assayed at the lower pH of 3.1, where the inhibitory activity was almost always accurately determinable. Thus, this chapter also recounts the various paths to the production of inhibitors with improved activity towards the natural target proteinase A of *S. cerevisiae* IA3, as well as examining the pH-dependency of the inhibitory activity.

5.2 Investigating the effect of the His-tag on the activity of recombinant *S. cerevisiae* IA3

In previous work on IA3 in this laboratory, recombinant forms of the protein were engineered to contain a hexahistidine tag at the C-terminus of the IA3 sequence (Li *et al.*, 2000; Phylip *et al.*, 2001). This facilitated the production of homogeneous preparations by IMAC (see Sections 2.2.23 and 3.7). To establish whether the positioning of the His-tag affected the behaviour of IA3, the His-tag was relocated to the N-terminus, and the recombinant protein was produced and purified to homogeneity as described in Section 3.7. The potency of this N-terminally His-tagged form (referred to as $NT(H)_6$) of IA3 towards *S. cerevisiae* vacuolar proteinase A at the two selected pH values (pH 3.1 and pH 4.7) was comparable with that of its C-terminally His-tagged counterpart (compare **1** and **2** in Table 5.1). The potency of both inhibitors at pH 4.7 was so tight that the K_i values lay at or beyond

the limit of accurate determination using the available assay methodology and so were estimated as < 0.1 nM, as described previously for the C-terminally tagged variant (Li *et al.*, 2000; Phylip *et al.*, 2001).

The N-terminal His-tag was designed so that it could be readily removed by N-terminal proteolytic "trimming" through five cycles of diaminopeptidase cleavage using the TAGZyme™ system described in Chapter 4. Following this procedure, the "trimmed" polypeptide, with a residual pyroglutamate residue at the N-terminus, was analysed by MALDI-TOF mass spectrometry. The observed mass peak of 7694.5 Da correlated with the theoretical mass (7684.8 Da) of "tagless" IA3, confirming the success of the procedure.

The potency of this tagless variant of IA3 (**3** in Table 5.1) at pH 3.1 and pH 4.7 was comparable to that of the two His-tagged versions, as well as the reported inhibitory constant (7 nM at pH 3.1) of the naturally-occurring IA3 obtained directly from yeast in earlier studies (Dreyer *et al.*, 1985). Thus, it was concluded that the presence and location of the His-tag does not substantially affect the inhibitory potency of the IA3 molecule. Based on the established protocol within the laboratory, all recombinant proteins described subsequently in this chapter were thus produced with an attached $LE(H)_6$ C-terminal His-tag.

5.3 The inhibitory activity of IA3 is located in the N-terminal half of the molecule

C-terminally His-tagged forms of IA3 truncated after either Tyr57 or Ala45 were constructed and purified as described in Sections 3.3 and 3.7, respectively. These shorter forms showed almost identical inhibitory potencies to the full-length recombinant molecule, which consists of 68 residues prior to the His-tag (compare **4** and **5** with **1** in Table 5.1). Attempts to produce more severely truncated IA3 variants in recombinant form in *E. coli* were not successful (Dr. L. Phylip, personal communication), so a peptide corresponding to residues 2-34 was obtained by solid phase synthesis (Section 2.1.8). Despite the complete absence of the C-terminal half

(residues 35-68) of the IA3 sequence, this peptide (**6** in Table 5.1) demonstrated comparable potency to the recombinant proteins. However, a shorter synthetic peptide that corresponded to residues 2-32 was found to have diminished potency at both pH values (**7** in Table 5.1) (Phylip *et al.*, 2001). Furthermore, Phylip *et al.* (2001) reported that peptides terminated at residue 30 or less became increasingly compromised in inhibitory capacity. However, the 2-32 length peptide was found to be invaluable because it permitted accurate measurement of K_i values for sequences that were otherwise beyond the detection limit at pH 4.7.

This strategy was exploited in an assessment of the importance of some of the residues that are known from the X-ray structures to lie on the solvent-exposed hydrophilic face of the inhibitory helix. Two peptides, with either 4 conservative or 7 less conservative substitutions were examined (**8** and **9**, respectively, in Table 5.1). The former was only marginally affected (compare with **7**) and even the latter (containing 4 Ala and 3 Gly residues as replacements) was still a reasonably effective inhibitor at pH 4.7. A sample of peptide **9** was incubated with *S. cerevisiae* vacuolar proteinase A at pH 4.7 and 37 °C for 16 h at a molar ratio of 50:1, prior to analysis by MALDI-TOF mass spectrometry. The single mass peak observed (3163.5 Da) corresponded perfectly to that of the intact peptide, indicating that despite the presence of the replacement Ala and Gly residues, with Gly being notoriously poorly accommodated in α-helices (Horovitz *et al.*, 1992; Fersht, 1998), this peptide had not been degraded by the proteinase. The data confirmed the evidence reported previously (Phylip *et al.*, 2001) that residues on the hydrophilic (outer) face of the α-helix contribute little if anything to the inhibitory potency of IA3. Therefore, no further studies were conducted on these residues.

5.4 Three hydrophobic "pin" regions of the IA3 sequence are crucial to inhibitory potency

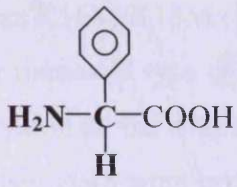
Most of the contributions to IA3 binding appear to originate from the hydrophobic face of the helix. Three clusters of residues, in particular, make significant hydrophobic contacts with the active site cleft of the enzyme: the front and back

"pins" formed from Val8/Ile11/Phe12 and Val25/Val26/Ala29/Phe30 respectively, together with Leu19 (Section 1.8.5 and Phylip *et al.*, 2001). In order to investigate the possible role of the front and back pins, attempts were made to disrupt the formation of the α-helix in these two regions. Peptides consisting of residues 2-32 were synthesised to contain proline replacements at either positions 7 or 26 or both. Proline is the most widely quoted helix-breaking residue because of the loss of two potential intrahelical hydrogen bonds and the conformational restriction of the *cis* peptide bond that results in a kink in the peptide backbone (Fersht, 1998). Despite this, however, the K7P mutant only showed a 3-fold reduction in potency (**10** in Table 5.1). Furthermore, the recombinant protein mutant $(Gly)_9$ (**11** in Table 5.1), which replaced residues 2-10 with Gly, also maintained an almost subnanomolar potency at pH 4.7. Even deletion of the N-terminal 6 residues in peptide WT 7-34 (**12** in Table 5.1) still yielded an inhibitor with a K_i value at pH 3.1 diminished by only 30-fold (compare **12** and **6** in Table 5.1). The implication of this result was that even the presence of the peptide backbone in this region, whilst desirable, was not strictly necessary.

In contrast, the V26P peptide mutant (**13** in Table 5.1) was a 60-fold weaker inhibitor than the equivalent wild type peptide at pH 4.7. However, truncation of the C-terminal end of the inhibitory sequence to the same position in WT 2-26 (**14** in Table 5.1) destroyed all inhibitory activity, confirming the importance of the back pin region. The simultaneous introduction of Pro residues in both front and back pin regions in the peptide K7P/V26P (**15** in Table 5.1) also resulted in total abolition of the inhibitory properties of the molecule. Indeed, following an incubation of this peptide (**15**) with *S. cerevisiae* vacuolar proteinase A at pH 4.7 and 37 °C for 15 h at a molar ratio of 300:1, a mass peak observed by MALDI-TOF MS analysis (2452.8 Da) correlated precisely with a fragment produced by cleavage of the E^{10}-I^{11} peptide bond. This bond was reported previously to be a cut site for non-target aspartic proteinases (Phylip *et al.*, 2001). It was concluded that the C-terminal 7 residues of the inhibitory sequence were more important than the equivalent residues at the N-terminus for the production, maintenance or stability of the helix.

The importance of the residue at position 19, the third hydrophobic "pin", was also established by mutation. A 6-fold reduction in potency at pH 3.1 was observed in a recombinant protein carrying the mutation of Leu19 to Met (L19M, **16** in Table 5.1), and a synthetic peptide containing a norleucine (an artificial isostere of Met) residue at position 19 (L19Z) showed a loss in potency of 3-fold at pH 4.7 (**17** in Table 5.1).

Similarly, replacement of Leu19 with the non-natural amino acid phenylglycine (PheGly; J):



(L19J, **18** in Table 5.1) reduced potency 4-fold, suggesting that the mere presence of a hydrophobic residue at position 19 was not sufficient, but that the shape, conformational flexibility and orientation were also critical factors. This conclusion was corroborated by the L19A synthetic peptide mutant (**19** in Table 5.1) which, with a K_i value of 700 nM at pH 4.7, effectively had lost all inhibitory potency despite the retention of the hydrophobic nature at position 19. Thus, just the deletion of an isopropyl group from the wild type residue (Leu) at a single position in the inhibitory helix had a profound effect on inhibition.

5.5 Hydrophobic replacement of the centrepiece residues

The only real exceptions to the generally amphipathic nature of the IA3 helix are the inhibitory "centrepiece" residues Lys18 and Asp22, both of which are on the hydrophobic face, and oriented down into the active site of the enzyme. Indeed, as described in detail in Section 1.8.6, the structure of the crystal complex (Li *et al.*, 2000) suggested that the two residues were within hydrogen bonding distance of each other and also interacted with one of the catalytic Asp residues, as well as the flap region of the enzyme. Although amphipathic helices have been reported often to contain stabilising salt bridges between residues at positions *i* and

i+4 (Khan & James, 1998), the location of such charged residues facing into a hydrophobic environment was remarkable.

A series of recombinant mutant proteins, containing hydrophobic substitutions at positions 18 and 22, was produced (Table 5.2). Replacement of Asp22 with Met (D22M, **20** in Table 5.2) had little impact on inhibitory potency. Similarly, the mutant D22L (**21** in Table 5.2), where Asp was replaced with an approximately isosteric Leu, demonstrated similar K_i values to wild type IA3. However, when Lys18 was replaced in the mutant K16M/K18M (**22** in Table 5.2) the K_i value at pH 3.1 was at least 5-fold better than wild type (**1** in Table 5.2), and brought it close to the limit of accurate determination on the available equipment. This suggested that hydrophobic replacements at position 18 were beneficial. Similarly, the K18I/L19M mutant (**23** in Table 5.2) was 10-fold more potent than the unfavourable single mutant L19M described in the last section (**16** in Table 5.1), again pointing to a beneficial role for a hydrophobic residue in place of Lys18.

Nevertheless, replacement of just one of the 18/22 centrepiece pair with a hydrophobic residue conceivably would have removed the potential hydrogen bond between the two positions, leaving at least one charge without a counteracting partner. To investigate if such an interaction could be replaced by further hydrophobic binding contributions, both Lys18 and Asp22 were replaced simultaneously with Met and Leu, respectively, their hydrophobic approximate isosteres. The resulting mutant, K18M/D22L (**24** in Table 5.2), was so potent at ***both*** pH values that it was not possible to record an accurate K_i value ($K_i < 0.1$ nM). Replacement of both residues with Ile (K18I/D22I, **25** in Table 5.2) was similarly effective. Interestingly, however, the additional Asp22→Ile mutation, when introduced in concert with the K18I/L19M mutations described above, appeared to reduce the potency of the earlier double mutant slightly at pH 3.1 (compare **23** and **26** in Table 5.2). This triple mutation (K18I/L19M/D22I) apparently returned the K_i value close to that of the wild-type, suggesting that the beneficial effect of the Ile residues at 18 and 22 was being offset by the detrimental L19M mutation.

Nevertheless, all of the mutants described above were so potent at both pH values that accurate interpretation of the results was difficult. As with the wild type IA3, sequential C-terminal truncation of the K18I/L19M/D22I mutant revealed that the C-terminal half made little or no contribution to inhibitory potency: both the Tyr57 and Ala45 truncated forms (**27** and **28**, Table 5.2) demonstrated identical K_i values to the full-length protein. Only when a synthetic peptide consisting of residues 2-32 of the inhibitor was evaluated could precise K_i values at both pH values be determined. In this form, the triple mutant peptide (**29** in Table 5.2) demonstrated an approximately 4-fold improvement upon the corresponding wild type peptide (**7** in Table 5.1). A 2-32 peptide form of the K18M/D22L mutant (K18Z/D22L, **30** in Table 5.2) fared even better, with a K_i value at pH 3.1 in the very low nanomolar range.

Thus, the benefits of hydrophobic replacements at the centrepiece positions seemed clear. To investigate further, a more extensive series (Table 5.3) of synthetic peptides of 2-32 length was constructed, containing multiple combinations of hydrophobic replacements at residues 17, 18, 19 and 22. The importance of selection of appropriate, compatible adjacent residues was emphasised by the K18J/L19J mutant (TJW5 in Table 5.3). Whereas the single introduction of a PheGly residue at position 19 had decreased potency by only 4-fold (see L19J 2-32, **18** in Table 5.1), this double change destroyed all the inhibitory activity of the modified peptide. The result contrasted starkly with the earlier improvement demonstrated for an L19M mutation upon the simultaneous introduction of an Ile at position 18. However, in a similar vein to that described for the K18I/L19M mutant (**23** in Table 5.2), the potency of a peptide containing the K18J/L19J double mutation could be "rescued" by the introduction of a further hydrophobic residue at adjacent position 17, although comparison of peptides TJW6 and TJW8 revealed that an Ile at this position was superior to a Leu. TJW7, in which the PheGly at position 19 was replaced with Nle, was moderately (2-fold) better than TJW6. This finding agreed with the mild preference for Nle over PheGly at this position noted earlier in the individual mutants L19J and L19Z (**18** and **17** in Table 5.1). However, all of the mutants described above remained considerably less potent than either the single L19J and L19Z mutants, or the corresponding wild type peptide (**7** in Table 5.1 and Table 5.3).

Maintenance of the residue at position 18 as the wild type (Lys) improved the K_i of TJW3 by almost 50-fold relative to the otherwise identical peptide TJW8. This emphasised clearly the unfavourable nature of a PheGly residue at 18. Previously, however, individually disadvantageous mutations (such as D22L, **21** in Table 5.2) had produced beneficial effects when in combination with a compatible residue at the corresponding (*i*, *i*+4) position. However, no inhibition by the mutant peptide TJW12 was observed, despite the introduction of an additional D22I substitution into the K18J/L19J variant (TJW5). In fact, the D22I mutation appeared incompatible with PheGly at both 18 and 19, even in combination with the hydrophobic replacements of Glu17 that had rescued it earlier (TJW6 and TJW8): neither TJW13 nor TJW15 showed any inhibition of *S. cerevisiae* vacuolar proteinase A at pH 4.7. Remarkably however, TJW14, which was identical to TJW13 except for a Nle in place of a PheGly at position 19, was at least as potent at pH 4.7 as the wild type peptide of the same length (**7** in Table 5.1) and the K_i value (8 ± 1 nM) measured at pH 3.1 was actually marginally (2-fold) better than that of the wild type. Therefore, exchange of all three charged residues at positions 17, 18 and 22 for hydrophobic residues (including accommodation of the unfavourable K18J mutation), together with an additional mutation of L19Z, was perfectly feasible without an overall detrimental effect upon the inhibitory potency of the peptide. Thus, both IA3 and the proteinase are capable of accommodating some substantial changes to the inhibitor sequence, even within such a critical region.

5.6 Centrepiece replacements by charged residues

Conservative changes of the Lys18 and Asp22 residues had little effect. Replacement of Lys18 by the more basic Arg (K18R, **31** in Table 5.4) or mutation of Asp22 for a longer Glu residue (D22E, **32** in Table 5.4) did not substantially affect the observed K_i values. Simultaneous introduction of these mutations in the double mutant protein K18R/D22E resulted in an inhibitor with identical potency to that of the wild type molecule (**33** in Table 5.4).

It was also possible to reverse the basic nature of the residue at position 18. The K18D replacement (**34** in Table 5.4) generated a mutant with acidic Asp residues at both positions 18 and 22, and this was within a largely hydrophobic active site region which was also populated by the two catalytic Asp residues of the enzyme. Despite this potentially unfavourable situation, the K_i recorded at pH 4.7 was still subnanomolar, and only a 6-fold decrease in potency was observed at pH 3.1.

However, the reciprocal change in which Asp22 was replaced with Lys (D22K, **35** in Table 5.4), decreased the activity of the inhibitor profoundly. The K_i value at pH 4.7 of this inhibitor with Lys residues at both 18 and 22 increased at least 500-fold with respect to the wild type (**1** in Tables 5.1 and 5.4), and all inhibitory activity was lost at pH 3.1. The hydrophobic centrepiece replacements (Section 5.5) had suggested that the derivation of an appropriate combination of residues might offset individually deleterious mutations such as D22K. Indeed, the simultaneous mutation of K18D with D22K produced an inhibitor of comparable subnanomolar potency to wild type IA3 at pH 4.7. This K18D/D22K mutant (**36** in Table 5.4) represented, effectively, a transposition of the centrepiece residues involved in the salt bridge. Another similar centrepiece transposition was acceptable: the mutant K18E/D22K (**37** in Table 5.4) maintained a K_i value at pH 4.7 that was 50-fold better than the D22K mutant on its own (**35**). In fact, even a simultaneous hydrophobic replacement of Lys18 with Val, producing the mutant K18V/D22K, resulted in a low nanomolar K_i value at pH 4.7 (**38** in Table 5.4).

That a hydrophobic replacement at position 18 could "rescue" the D22K mutation suggested another possible centrepiece. The K18M/D22K mutant (**39** in Table 5.4) was therefore constructed, which resulted in the reciprocal of the centrepiece in the D22M mutant (**20** in Table 5.2). The introduction of K18M was remarkably beneficial, with a superior K_i value at pH 4.7 to the other mutants tested with a D22K mutation (Table 5.4). This effect was reminiscent of two mutants already described - K16M/K18M and K18M/D22L (**22** and **24**, respectively, in Table 5.2).

However, in spite of the recovery to subnanomolar K_i values at pH 4.7, all of the mutants containing a Lys at 22 exhibited a curious phenomenon. Typically, when the K_i values at pH 4.7 and 3.1 were within the accurately determinable range then,

regardless of the length of the IA3-based inhibitor or the nature of the mutation(s) introduced:

$$\frac{K_i \text{ (pH 3.1)}}{K_i \text{ (pH 4.7)}} < 40$$

However, all the Lys22-containing IA3 mutants had a ratio of K_i values that was considerably greater (at least 10-times greater, except for **40**):

$$\frac{K_i \text{ (pH 3.1)}}{K_i \text{ (pH 4.7)}} > 40$$

An explanation for the apparent catastrophic loss of activity at pH 3.1 was provided by MALDI-TOF mass spectrometric analysis of the K18D/D22K form of IA3 after incubation with *S. cerevisiae* vacuolar proteinase A in a 40:1 molar ratio at 37 °C and pH 3.1 for 16 h. Two major peaks were observed with masses of 3211.3 Da and 5578.2 Da, consistent with cleavage of the A^{29}-F^{30} bond to produce fragments M1-A29 (expected mass 3210.6 Da) and F30-LE$(H)_6$ (expected mass 5576.7 Da) respectively. This bond in the IA3 sequence is one that is often hydrolysed preferentially by non-target aspartic proteinases (Phylip *et al.*, 2001). No such cleavage was observed in a parallel incubation at pH 4.7. Thus, a simple reversal of the centrepiece residues converted a potent inhibitor at both recorded pH values (**1** in Table 5.4) into a molecule that, whilst still a subnanomolar inhibitor at pH 4.7, was also cleavable as a substrate at pH 3.1. It appeared, therefore, that a pH-dependent transition or fundamental "switch" of behaviour had been introduced into IA3 by this mutation.

Such a "pH switch" phenomenon had not been observed in any of the previous IA3 mutants. However, replacement of the Lys22 with the other basic amino acid, Arg, also resulted in a pH switch, and this behaviour was shown in the presence of either an Asp (K18D/D22R) or Glu (K18E/D22R) at position 18 (**40** and **41** in Table 5.4, respectively). Mass spectrometric analyses were not performed on these Arg22-containing proteins, but the weight of previous evidence (including Phylip *et al.*, 2001, and much unpublished data) strongly suggested that the K_i values recorded at pH 3.1 for these proteins would be consistent with slow cutting as substrates. It appeared, therefore, that a basic residue at position 22 was producing

the effect, allowing for potent, subnanomolar inhibition of *S. cerevisiae* vacuolar proteinase A at pH 4.7, but "switching" the same molecule into a substrate of the enzyme at pH 3.1.

5.7 Discussion

Most of the binding energy contributions to the inhibitory activity of IA3 appear to come from hydrophobic interactions. The evidence accumulated in this chapter would suggest that the residues constituting the back pin region may be more critically important to potency than the front pin region. Sequence deletions, or inhibitory helix-disruption by the introduction of Pro, at the C-terminal end of the inhibitory sequence were certainly more detrimental to potency (Table 5.1). The importance of the central hydrophobic pin, Leu19, was also confirmed, although alternative residues (Met, Nle, PheGly) were identified that were only mildly detrimental to potency (Table 5.1). This confirms that the catastrophic loss of potency noted by Phylip *et al.* (2001) upon replacement of this residue with Ala was probably due to the shorter, less bulky, residue being unable to re-constitute the critical binding contacts.

The only real exception to the hydrophobic binding energy contributions are the centrepiece residues of Lys18 and Asp22, which establish a precisely-tailored network of hydrogen bond contacts between the two sidechains and with key features of the active site of the proteinase (Section 1.8.6 and Figure 5.1A). This charge dissipation not only contributes to the binding; it is also critical due to the essentially hydrophobic environment of the active site cleft. Indeed, as shown in Section 5.5, replacement of both residues with appropriate hydrophobic alternatives did result in very tight binding inhibitors (**24** and **25** in Table 5.2), most probably due to increased hydrophobic interactions. However, more extensive hydrophobic replacements in the region surrounding the centrepiece (in the TJW peptides in Table 5.3) revealed the importance of balance and residue compatibility in the search for extra hydrophobic binding contributions. Nevertheless, the mutational experiments, both within the centrepiece and throughout the sequence, indicated that *S. cerevisiae* IA3 is

remarkably tolerant of extensive changes to critical regions of the sequence with minimal effect on potency, as long as an appropriate combination of residues is found.

The complicated nature of the centrepiece interactions was also exemplified by the centrepiece reversal mutant K18D/D22K (**36** in Table 5.4). Superficially, this mutant might be expected to retain the (*i*, *i*+4) salt bridge between these positions, but with reversed polarity. However, considering the precise nature of the centrepiece interactions (Section 1.8.6), the discovery that the residues could be interchanged in this way, whilst retaining subnanomolar potency at pH 4.7, seems remarkable. Indeed, the significance of the salt bridge appeared to be confirmed when the various combinations of Lys or Arg at 18 and Asp or Glu at 22 (or vice versa) generated inhibitors of broadly comparable potencies (Table 5.4).

However, two observations indicate that this concept is too simplistic. Firstly, the mutant with Asp at both 18 and 22, and hence no apparent potential for salt bridge formation, retained most of the potency of the wild type inhibitor. This was despite the potential problems of accommodating a further two anionic groups in a predominantly hydrophobic environment also populated by the two catalytic Asp residues of the enzyme active site. Secondly, the conversion of all proteins containing a Lys at 22 from potent inhibitors at pH 4.7 (with the exception of **35**, the mutant with Lys at both 18 and 22) into apparent substrates of the same enzyme at the lower pH value of 3.1 represents a considerable departure from the behaviour of any other forms of IA3 studied previously. The investigations described in Section 5.6 revealed that only proteins with a basic residue at position 22 demonstrated this so-called "pH switch" (**36** - **41**), suggesting that this might be the critical progenitor of the phenomenon. Such a conclusion implies an indirect effect of this residue rather than a change in state of ionization, because both pH 4.7 and pH 3.1 are a very long way from the pKa of Lys.

No X-ray structure yet exists for an IA3 mutant with a basic residue at 22 complexed with *S. cerevisiae* vacuolar proteinase A. However, a model was constructed by our collaborator Dr. Daniel Bur (Actelion Ltd., Allschwil, Switzerland) of the K18M/D22K variant (**39** in Table 5.4) complexed with the enzyme, in an effort to explain the pH-dependency of the inhibitory activity. Two structural requirements

must be met when Lys is relocated from position 18 to 22 of IA3, i.e. one full turn of the inhibitory α-helix. Firstly, the hydrophobic contacts made by the Lys18 sidechain methylene groups must be re-established by the replacement residue. As an approximate isostere, the Met residue of **39** may well be best able to mimick the Lys sidechain. This offers a potential explanation for the superior potency of this mutant over the other Lys/Arg22-containing variants.

The second condition is that the newly-introduced positive charge at position 22 must be stabilised by an effective hydrogen bonding network. The modelling studies indicate that the optimum conformation for the Lys22 sidechain of **39** interposes the ε-NH_3^+ group between the carboxyl groups of the two catalytic Asp residues of the enzyme (Figure 5.1B, and compare with the wild type situation in Figure 5.1A). Notably, the insertion of the Lys residue at this position would also result in the displacement of the catalytic water molecule (Section 1.4). This would represent a related, but quite distinct, inhibitory mechanism to that of wild type IA3 - where the water molecule remains within the E-I complex (Li *et al.*, 2000 and compare Figures 5.1A and B).

The displacement of the catalytic water molecule is a well-established mechanism of inhibition of aspartic proteinases (Section 1.7.1). Furthermore, the insertion of a Lys residue between the catalytic Asp residues represents a striking parallel with one of the systems of intrinsic inhibition employed in Nature. Many aspartic proteinases, such as pepsin, are maintained as inactive zymogen precursors by insertion of a segment of the propart region into the active site so as to locate the ε-NH_3^+ group of a propart Lys (Lys36P in pepsinogen) between the catalytic aspartates (Section 1.2). Superposition of the modelled K18M/D22K IA3-proteinase A complex with the structure described for pepsinogen (3PSG in Khan & James, 1998) reveals that the ε-NH_3^+ group of Lys22 of the IA3 mutant adopts an almost identical position and orientation to that of Lys36P in pepsinogen (Figure 5.2). Therefore, both Lys22 of IA3 mutant **39** and Lys36P of pepsinogen can establish strong hydrogen bonds with both catalytic Asps (of the respective enzymes), at least one of which is unprotonated at pH 4.7. This maintains the stability of the zymogen and offers an explanation for the potency of **39** as an inhibitor of *S. cerevisiae* vacuolar proteinase A at pH 4.7. In

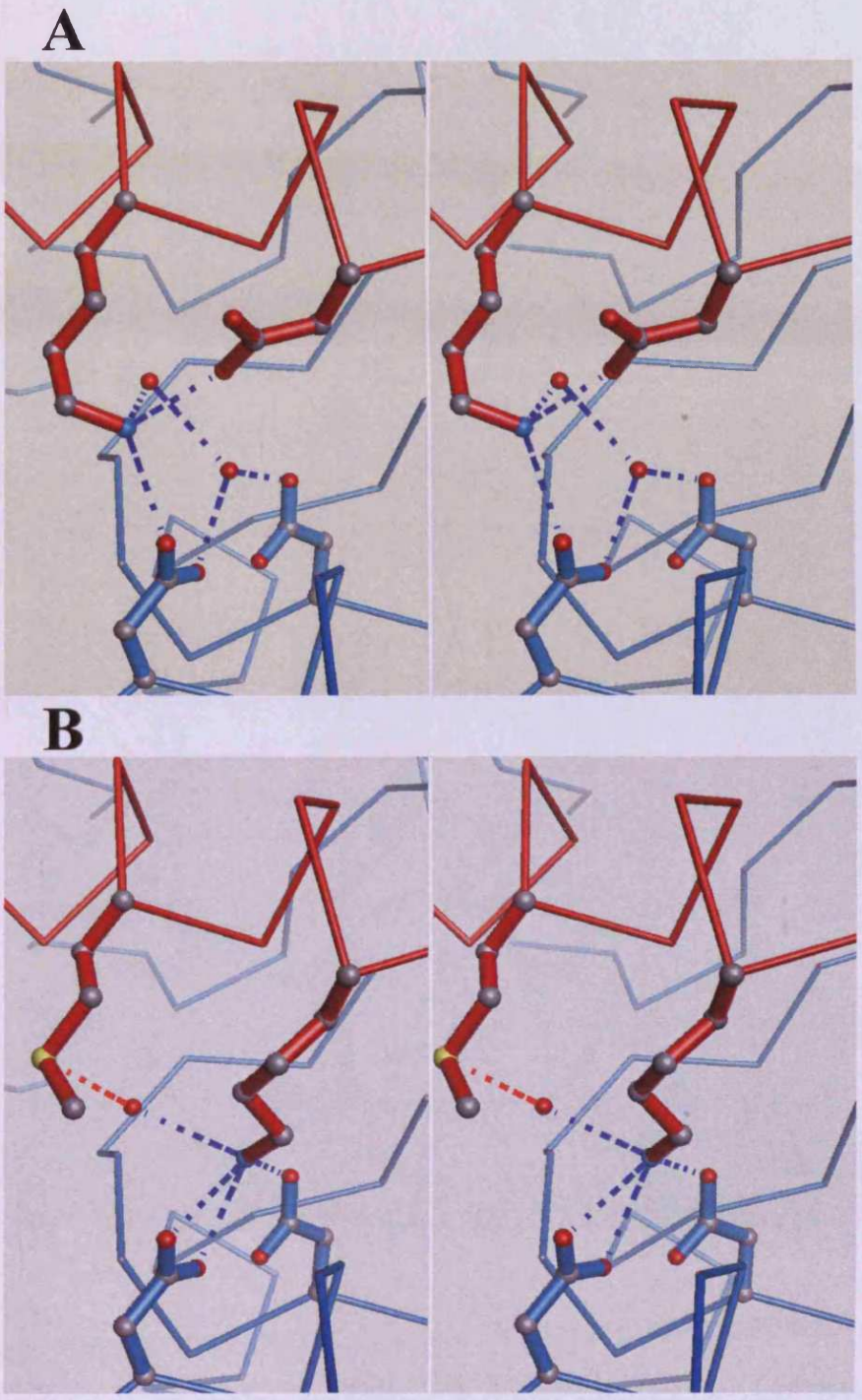


Figure 5.1A The hydrogen bond interactions between Lys18 and Asp22 residues from wild-type IA3 (red) and the two catalytic Asp residues (blue) of *S. cerevisiae* vacuolar proteinase A and two water molecules (PDB code 1DPJ). B. Model of the IA3 mutant K18M/D22K (39) complexed with the enzyme. The transposition of Lys to position 22 displaces the catalytic water molecule and produces a new hydrogen bonding network. Figure kindly produced by Dr. D. Bur (Actelion Ltd., Allschwil, Switzerland).

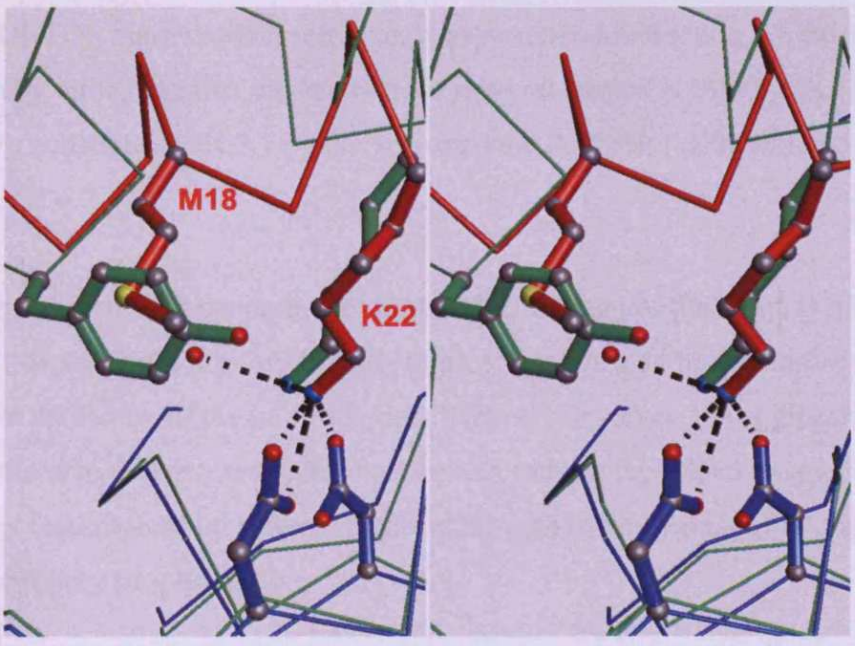


Figure 5.2 *S. cerevisiae* vacuolar proteinase A (blue) complexed with K18M/D22K mutant IA3 (red) is superimposed on pepsinogen (green; PDB code 3PSG) with side-chains of Tyr9 and Lys36P displayed. Figure kindly produced by Dr. D. Bur (Actelion Ltd., Allschwil, Switzerland).

the case of pepsinogen, zymogen activation is initiated by lowering the pH to ≤3. At this pH, both catalytic Asp residues become protonated. This destabilises the zymogen arrangement by disrupting the hydrogen bonding network and leads to propart release and, ultimately, the generation of active pepsin (Khan & James, 1998; Dunn, 2002).

The inhibitory data described in Table 5.4 suggest that an analogous situation might exist in the Lys/Arg22-containing IA3 variants. Based on the model, it can be proposed that, at pH 3.1, the ε-NH_3^+ group of Lys22 is unable to find either a negative counter-charge or appropriately positioned hydrogen bond partners from the now-protonated catalytic Asp residues. Therefore, the inhibitor is ineffective at pH 3.1. The MALDI-TOF mass spectrometric analysis described in Section 5.6 extends this conclusion by indicating that the centrepiece reversal mutant K18D/D22K (**36**) is cleaved as a substrate at pH 3.1 by the very enzyme it inhibits with subnanomolar potency at pH 4.7.

Thus, the behaviour of these particular mutant IA3 molecules that display a far more marked pH-dependent inhibitory binding to proteinase A may be indicative of a variation on the nature of the inhibitory mechanism. However, the engineering of these mutations in this extrinsic inhibitor appears to have replicated the mechanism that is a key feature used by Nature for silencing activity in zymogens by their intrinsic inhibitory propart region.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM) pH 4.7	K_i (nM) pH 3.1
1	CT$(H)_6$ WT IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	<0.1	1.1 ± 0.4*
2	NT$(H)_6$ WT IA3	MK$(H)_7$MQ NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	<0.1	0.6 ± 0.2
3	Tagless WT IA3	*QNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	0.1 ± 0.1	0.7 ± 0.3
4	WT Y57 IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~LE$(H)_6$[57]	<0.1	1.2 ± 0.4
5	WT A45 IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~LE$(H)_6$[45]	<0.1	2 ± 0.4
6	WT 2-34	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKZA	<0.1*	3 ± 0.6*
7	WT 2-32	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KK	2 ± 0.4*	15 ± 4*
8	DB1 2-32	NTDQ QKVSE IFQAS KEKLQ GDAKV VTDAF KK	5 ± 0.4	45 ± 4
9	JK1 2-32	NAGQ AKVSG IFQAS KEKLA GDAKV VADAF KK	35 ± 3	nd
10	K7P 2-32	NTDQ QPVSE IFQSS KEKLQ GDAKV VSDAF KK	7 ± 1	nd
11	$(Gly)_9$ IA3	MGGGG GGGGG IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	1 ± 0.7*	40 ± 10*
12	WT 7-34	KVSE IFQSS KEKLQ GDAKV VSDAF KKZA	30 ± 8*	85 ± 20*
13	V26P 2-32	NTDQ QKVSE IFQSS KEKLQ GDAKV PSDAF KK	125 ± 20	nd
14	WT 2-26	NTDQ QKVSE IFQSS KEKLQ GDAKV V	NI*	nd
15	K7P/V26P 2-32	NTDQ QPVSE IFQSS KEKLQ GDAKV PSDAF KK	NI	nd
16	L19M IA3	MNTDQ QKVSE IFQSS KEKMQ GDAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	nd	6 ± 1.2*
17	L19Z 2-32 IA3	NTDQ QKVSE IFQSS KEKZQ GDAKV VSDAF KK	6 ± 1	nd
18	L19J 2-32 IA3	NTDQ QKVSE IFQSS KEKJQ GDAKV VSDAF KK	9 ± 1	nd
19	L19A 2-32 IA3	NTDQ QKVSE IFQSS KEKAQ GDAKV VSDAF KK	700 ± 120*	NI*

Table 5.1 Interactions between *S. cerevisiae* vacuolar proteinase A and various forms of *S. cerevisiae* IA3. Mutations are highlighted in red. In all synthetic peptides Z = L-norleucine (an isosteric replacement for Met). All recombinant proteins show only the residues of the inhibitory sequence (1-34). Recombinant proteins were produced with either a C-terminal LE$(H)_6$ tag or N-terminal MK$(H)_6$HMQ tag. Removal of the N-terminal His-tag results in the exposure of an N-terminal pyroglutamate (*Q). The C-terminal "tail" of proteins is represented as ~~~~~~~~, according to the length of the molecule. For tail sequence, see Appendix I. *= From Phylip *et al.* (2001); nd = not determined; NI = no inhibition.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM) pH 4.7	K_i (nM) pH 3.1
1	WT IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	1.1 ± 0.4*
20	D22M	MNTDQ QKVSE IFQSS KEKLQ GMAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	2 ± 0.3
21	D22L	MNTDQ QKVSE IFQSS KEKLQ GLAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.6 ± 0.3*	8 ± 0.2*
22	K16M/K18M	MNTDQ QKVSE IFQSS MEMLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1*	<0.2*
23	K18I/L19M	MNTDQ QKVSE IFQSS KEIMQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1*	0.5 ± 0.3*
24	K18M/D22L	MNTDQ QKVSE IFQSS KEMLQ GLAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1*	<0.1*
25	K18I/D22I	MNTDQ QKVSE IFQSS KEILQ GIAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	<0.1
26	K18I/L19M/D22I	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	2 ± 0.3
27	K18I/L19M/D22I Y57 IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~LE(H)$_6$[57]	<0.1	2 ± 0.3
28	K18I/L19M/D22I A45 IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~LE(H)$_6$[45]	<0.1	2 ± 0.3
29	K18I/L19Z/D22I 2-32 IA3	NTDQ QKVSE IFQSS KEIZQ GIAKV VSDAF KK	0.6 ± 0.1	5 ± 1
30	K18Z/D22L 2-32 IA3	NTDQ QKVSE IFQSS KEZLQ GLAKV VSDAF KK	0.5 ± 0.1	1.5 ± 0.3

Table 5.2 The effect of replacement of the centrepiece residues of *S. cerevisiae* IA3 with hydrophobic residues. Mutations are highlighted in red. In all synthetic peptides Z = L-norleucine. Recombinant proteins were expressed with a C-terminal LE(H)$_6$ tag. Only residues in the inhibitory sequence (1-34) are detailed. The C-terminal "tail" of proteins is represented as ~~~~~~~~, according to the length of the molecule. *= from Phylip *et al.* (2001).

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM) pH 4.7
1	WT IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~~LE(H)$_6$	<0.1
7	WT 2-32	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KK	2 ± 0.4*
	TJW5	NTDQ QKVSE IFQSS KEJJQ GDAKV VSDAF KK	NI
	TJW8	NTDQ QKVSE IFQSS KLJJQ GDAKV VSDAF KK	625 ± 100
	TJW6	NTDQ QKVSE IFQSS KIJJQ GDAKV VSDAF KK	80 ± 10
	TJW7	NTDQ QKVSE IFQSS KIJZQ GDAKV VSDAF KK	35 ± 5
	TJW3	NTDQ QKVSE IFQSS KLKJQ GDAKV VSDAF KK	13 ± 2
	TJW12	NTDQ QKVSE IFQSS KEJJQ GIAKV VSDAF KK	NI
	TJW13	NTDQ QKVSE IFQSS KIJJQ GIAKV VSDAF KK	NI
	TJW15	NTDQ QKVSE IFQSS KLJJQ GIAKV VSDAF KK	NI
	TJW14	NTDQ QKVSE IFQSS KIJZQ GIAKV VSDAF KK	1.7 ± 0.4

Table 5.3 Further hydrophobic mutations to the centrepiece residues of *S. cerevisiae* IA3. Mutations are highlighted in red. Z = L-norleucine and J = L-phenyl glycine. NI = no inhibition. *= from Phylip *et al.* (2001).

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)		pH
			pH 4.7	pH 3.1	switch
1	WT IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	1.1 ± 0.4*	>11
31	K18R	MNTDQ QKVSE IFQSS KE**R**LQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	0.5 ± 0.1	>5
32	D22E	MNTDQ QKVSE IFQSS KEKLQ G**E**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	3 ± 0.3	>30
33	K18R/D22E	MNTDQ QKVSE IFQSS KE**R**LQ G**E**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	1.1 ± 0.3	>11
34	K18D	MNTDQ QKVSE IFQSS KE**D**LQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.3 ± 0.1	6 ± 0.8	20
35	D22K	MNTDQ QKVSE IFQSS KEKLQ G**K**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	50 ± 3	NI	∞
36	K18D/D22K	MNTDQ QKVSE IFQSS KE**D**LQ G**K**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.3 ± 0.1	120 ± 20	400
37	K18E/D22K	MNTDQ QKVSE IFQSS KE**E**LQ G**K**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	1.3 ± 0.2	610 ± 120	470
38	K18V/D22K	MNTDQ QKVSE IFQSS KE**V**LQ G**K**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	2 ± 0.4	1000 ± 200	500
39	K18M/D22K	MNTDQ QKVSE IFQSS KE**M**LQ G**K**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.2 ± 0.1	85 ± 15	425
40	K18D/D22R	MNTDQ QKVSE IFQSS KE**D**LQ G**R**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	5 ± 0.8	800 ± 300	160
41	K18E/D22R	MNTDQ QKVSE IFQSS KE**E**LQ G**R**AKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.7 ± 0.3	280 ± 30	400

Table 5.4 The effect of mutation of the centrepiece residues of *S. cerevisiae* IA3. Mutations are highlighted in red. All recombinant proteins were expressed with a C-terminal LE(H)$_6$ tag. Only residues in the inhibitory sequence (1-34) are detailed; the C-terminal "tail" (residues 35-68) is represented as ~~~~~~~~. The "pH switch", described in detail in Section 5.6, represents the relative fall in potency from pH 4.7 to pH 3.1. Large switch values are in red. NI = no inhibition. *= from Phylip *et al.* (2001).

Chapter 6

Inhibition of proteinase A from *S. cerevisiae* by IA3 from *S. castellii*

6.1 Introduction

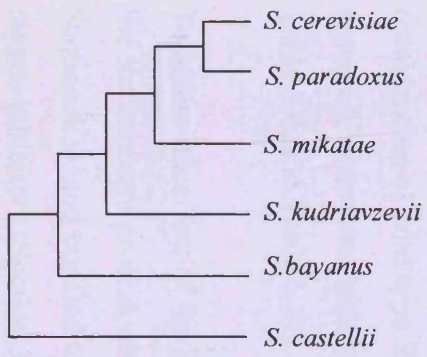
Numerous recent advances in genome sequencing projects facilitated attempts to identify IA3-like polypeptides from other species. Extensive BLAST searches were performed at frequent intervals by a colleague in our group (Dr David Wyatt). Five new orthologues were identified, all of which originated from yeast species closely related to *S. cerevisiae* (see Figure 6.1A, and Rokas *et al.*, 2003).

The alignment in Figure 6.1B reveals that four of the new IA3-like proteins were almost identical to *S. cerevisiae* IA3 in the inhibitory sequence region (residues 2-34), with no more than three changes. However, the fifth sequence, that from *S. castellii*, departed from this pattern substantially: 19 of the first 34 residues (over 55%) were altered. Furthermore, with 81 residues it was considerably longer than the other IA3 polypeptides, all of which consisted of 68 or 69 amino acids. The C-terminal halves (i.e. residues 35 to the C-terminus) of all the sequences were considerably more divergent than the N-terminal halves, suggesting a region of less critical sequence to the function of the molecule and, therefore, less constrained by natural selection. It appeared feasible, therefore, that these molecules were once again aspartic proteinase inhibitors that functioned in a similar fashion to *S. cerevisiae* IA3, with their inhibitory activity residing in the N-terminal half. As a result, these sequences were designated in this work as IA3 forms from the respective *Saccharomyces* species. The sequence from *S. castellii* was particularly noteworthy because of the substantial differences when compared with *S. cerevisiae* IA3. Therefore, this chapter describes investigations into the features of the IA3-like molecule from *S. castellii*.

6.2 Production and initial characterisation of *S. castellii* IA3

Attempts were made to purify the naturally-occurring inhibitor from cultures of *S. castellii* using methods similar to that of, for example, Biedermann *et al.* (1980) for the purification of naturally-occurring *S. cerevisiae* IA3. However, the inhibitor preparation was of insufficient purity and quantity for accurate quantitative investigation. Thus, in order to examine any inhibitory potency of the IA3-like

A



B

Species	Sequence 1 10 20 30 40 50 60 70 80
S. cerevisiae	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMAS QDKDG KTTDA DES-E KHNYQ EQYNK LKGAG HKKE
S. paradoxus	MNTDQ QKV E IFQSS KEKLQ GDAKV VSDAF K MAS KDKDG KVGIS D I-E KH YQ EQYNK LKGAG HK E
S. mikatae	MN DQ QKV E IFQSS KEKLQ GDAKV VSDAF K MAS KDENG NVVNT DET-E SPDYQ EQYNK LKGAG HKKE
S. kudriavzevii	MNTDQ QKV E IFQSS KEKLQ GDAKV V DAF K MAS KDKNG KDDNI SDNTR K DYQ EQYNK LKGAG H E
S. bayanus	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF K MAS KDKNS KSDSS GESSD K NYQ EQY K LKGA KREE
S. castellii	MSDKN ANVSE MFQQA KEMLE GKAKA ASEGM KNMAS PESKE KMKTK GQDME KKA-Q DSYNK LNGAG KQEQT KERQD KDELM DH

Figure 6.1A Evolutionary relationship between several *Saccharomyces* species (adapted from Rokas *et al.*, 2003).

B. Alignment of the five newly-discovered IA3 orthologues with *S. cerevisiae* IA3. Residues differing from those of *S. cerevisiae* IA3 (in yellow) are highlighted in red. Insertions are in blue, with deletions marked by a red hyphen. The larger orthologue from *S. castellii* has 55 differences from *S. cerevisiae* IA3, with 19 of these located in the inhibitory region of residues 1-34.

polypeptide from *S. castellii* in detail, it was necessary to produce the protein in recombinant form. Therefore, the intronless gene was amplified from genomic DNA prepared from *S. castellii* (by Dr David Wyatt) using appropriate primers. The resulting PCR fragment was cloned into pET22b between the *Nde*I and *Xho*I restriction enzyme cut sites, thereby placing the insert immediately upstream from the nucleotides encoding the C-terminal $LE(H)_6$ tag in the vector (see Section 3.1). Recombinant *S. castellii* IA3 was produced in *E. coli* and purified to homogeneity by IMAC with similar yields to wild type *S. cerevisiae* IA3 (Sections 2.2.23 and 3.7). MALDI-TOF mass spectrometry of the purified protein revealed a peak of 10122 Da, corresponding to the expected mass (10113.7 Da) of *S. castellii* IA3 lacking the N-terminal initiator Met residue.

The IA3-like protein from *S. castellii* was a potent inhibitor of *S. cerevisiae* vacuolar proteinase A at pH 4.7 although, unlike *S. cerevisiae* IA3, the K_i value was not subnanomolar (**42** in Table 6.1). The activity at pH 3.1 was very poor but the K_i value could be determined with sufficient accuracy to indicate that it demonstrated a similar "pH switch" to those in mutant forms of *S. cerevisiae* IA3 that contained a basic residue at position 22 (Chapter 5). The sequence of *S. castellii* IA3 (Figure 6.1B) also contains a Lys residue naturally at position 22, suggesting that once again this might be the origin of the phenomenon. Indeed, the Met18 and Lys22 centrepiece combination of wild type *S. castellii* IA3 is the same as that of the most potent of the mutant *S. cerevisiae* inhibitors developed with a basic residue at 22 (see **39** in Table 5.4, and Table 6.1).

To assess the effect of the location of the His-tag on the behaviour of *S. castellii* IA3, the recombinant protein was also produced in $NT(H)_6$ form (for details, see Section 4.3), and purified to homogeneity in the same manner as the C-terminally tagged inhibitors (Section 3.7). MALDI-TOF mass spectrometry of the protein revealed a peak of mass 10523.0 Da, which corresponded to the expected mass (10526.9 Da) of the *S. castellii* IA3 molecule plus the $MK(H)_6HMQ$ N-terminal tag.

The $NT(H)_6$ form of *S. castellii* IA3 (**43** in Table 6.1) was 10-fold more potent than the C-terminally tagged counterpart at pH 4.7, producing a subnanomolar K_i value equivalent to that demonstrated by the. *S. cerevisiae* IA3 mutant (K18M/D22K) with

the same centrepiece combination (**39** in Table 6.1). However, the pH switch remained. The improvement in potency at pH 4.7 might result from the removal of a detrimental effect of the His-tag at the C-terminus, or arise from a benefit brought by placing the His-tag at the N-terminus.

To investigate if the inhibitory activity of *S. castellii* IA3 resided in the same region as in *S. cerevisiae* IA3, a peptide corresponding to residues 2-34 of *S. castellii* IA3 was synthesised (Section 2.1.8). Remarkably, a subnanomolar K_i value was recorded at pH 4.7 for this peptide (**44** in Table 6.1). This represented a 10-fold improvement upon the potency of the full-length C-terminally His-tagged recombinant protein. As such, the K_i value was comparable to that of the $NT(H)_6$ form, suggesting that the "poor" performance of the C-terminally tagged form might be a consequence of the placement of the His-tag at the C-terminus. Nevertheless, this result confirmed that the inhibitory activity of *S. castellii* IA3 is located in the N-terminal half of the molecule, in a manner similar to that described for *S. cerevisiae* IA3.

The potency of peptide **44** at pH 3.1 was even more remarkable. The low nanomolar K_i value represented a >40-fold improvement upon the K_i of the C-terminally His-tagged recombinant form (**42** in Table 6.1). Furthermore, the pH switch had apparently been abolished, with the ratio of the inhibitory constants at the two pH values being reduced to just 20. This contrasted with all data for *S. cerevisiae* IA3 inhibitory sequences reported previously (Phylip *et al.*, 2001, and Chapter 5 of this thesis) where removal of the C-terminal half of the molecule did not improve potency at pH 3.1.

Therefore, deletion of the C-terminal "tail" (residues 35-81) of *S. castellii* IA3, together with the His-tag, not only resulted in an improved potency at both pH values but also removed the pH switch from a sequence containing a Lys residue at position 22. For purposes of clarity, and to aid interpretation, the rest of this chapter will describe the attempts to identify the origin of the improvements in potency at pH 4.7 demonstrated by the $NT(H)_6$ form of *S. castellii* and the 2-34 *S. castellii* IA3 peptide. A comparison of the regions critical to inhibitory potency in the two IA3 species is also described. Investigations into the pH switch are considered in the next chapter.

6.3 The role of C-terminal tail length and identity on the inhibitory potency of *S. castellii* IA3 against *S. cerevisiae* vacuolar proteinase A

Two differences were apparent from the alignment of the C-terminal tails (i.e. residues 35 to the C-termini) of *S. cerevisiae* IA3 and *S. castellii* IA3, illustrated in Figure 6.1B:

(i) the sequences were quite dissimilar, with few residues in common;

(ii) the *S. castellii* IA3 tail was 13 residues longer.

Chimaeric proteins were constructed in C-terminal His-tag form, where the tail regions of the two species were exchanged. N*cere*-C*cast* IA3 (**45** in Table 6.2), where residues 35-68 of *S. cerevisiae* IA3 were replaced with residues 35-81 of *S. castellii* IA3, had a K_i value below the limit of accurate detection at pH 4.7. Furthermore, the K_i value at pH 3.1 (4.0 ± 0.6 nM) was similar to that of wild type *S. cerevisiae* IA3 (1.1 ± 0.4 nM). Almost identical values were recorded for a further chimaeric protein that consisted of the *S. cerevisiae* IA3 inhibitory sequence containing the mutations K18I/L19M/D22I fused with the *S. castellii* IA3 tail (**46** in Table 6.2). Therefore, the *S. castellii* IA3 tail did not appear disadvantageous to the inhibitory sequence of *S. cerevisiae* IA3.

The reciprocal chimaera was also constructed in C-terminal His-tag form, where residues 35-81 of *S. castellii* IA3 were replaced with residues 35-68 of *S. cerevisiae* IA3. At pH 4.7, this protein – N*cast*-C*cere* IA3 (**47** in Table 6.2) - was 40-fold more potent than the C-terminally His-tagged wild type *S. castellii* IA3 (**42** in Table 6.1), with a subnanomolar K_i value at the detection limit on our equipment. As such, the potency at pH 4.7 was equivalent or even superior to that of both the synthetic peptide and the NT$(H)_6$ *S. castellii* IA3 (**44** and **43**, respectively, in Table 6.1), and also the *S. cerevisiae* IA3 mutant with the same centrepiece combination (K18M/D22K, see **39** in Table 6.1). However, with a K_i value at pH 3.1 of 180 ± 40 nM, the pH switch was still in evidence in **47**. This clearly separated the pH switch phenomenon from the poor performance of recombinant *S. castellii* at pH 4.7, a point that will be considered further in the next chapter.

The *S. cerevisiae* IA3 tail sequence (35-68) was attached also to the inhibitory sequence of a *S. castellii* IA3 mutant (**48**) that had demonstrated poor potency when its own tail sequence was present. The N*cast* M18V-C*cere* chimaera (**49** in Table 6.2) was 5-fold more potent at pH 4.7 than the *S. castellii* M18V IA3 (**48** in Table 6.2), although it was still not as effective as this centrepiece (K18V/D22K) in an all- *S. cerevisiae* IA3 context (see **38** in Table 5.4 and Table 6.2).

Consistently, therefore, the introduction of an *S. cerevisiae* IA3 tail in place of the endogenous *S. castellii* IA3 tail improved the potency of an *S. castellii* IA3 inhibitory sequence.

However, as stated above, the two tail species differed not just in sequence but also in length. Thus, a consequence of creating the N*cast*-C*cere* chimaeric inhibitors (**47** and **49**) was that both variables had been changed simultaneously. To examine these variables independently, a truncated form of *S. castellii* IA3 was produced with an identical length (68 residues) to the chimaeric proteins. However, for reasons detailed in Section 7.4, this protein – $NT(H)_6$ *S. castellii* Q68 IA3 (**50** in Table 6.2) - was produced in N-terminally His-tagged form, rather than with the C-terminal His-tag used in most of these studies. To provide a direct comparison, therefore, it was necessary also to produce an $NT(H)_6$ form of the "wild-type" N*cast*-C*cere* chimaera (**51** in Table 6.2).

The Q68-truncated form of $NT(H)_6$ *S. castellii* IA3 was almost as potent as the full-length counterpart (compare **50** in Table 6.2 with **43** in Table 6.1). However, the potency was essentially *identical* to that of the same-length $NT(H)_6$ N*cast*-C*cere* chimaera (**51** in Table 6.2). From this comparison of the K_i values of **50** and **51**, it appeared that the identity of residues 35-68 had little significance on the potency of the *S. castellii* inhibitory sequence (residues 2-34). Furthermore, the similarity in potency of **50** and **51** with the synthetic peptide (**44** in Table 6.1) suggested that the $NT(H)_6$ tag and residues 35-68 had little effect on the inhibitory potency at pH 4.7.

In creating the N*cast*-C*cere* chimaera **51**, the His-tag was relocated from C- to N-terminus. This translocation resulted in at least a 7-fold reduction in potency;

although the K_i value was still subnanomolar (compare **47** and **51** in Table 6.2). This loss of potency was in marked contrast to the findings when the His-tag was translocated in wild-type full-length *S. castellii* IA3, as described in Section 6.2, where the inhibitor was 10-fold *more* potent in $NT(H)_6$ form (compare **42** and **43** in Table 6.1).

From all of these results it would appear that, for an *S. castellii* IA3 inhibitory sequence at pH 4.7:

(i) in C-terminal His-tag form, the tail from *S. cerevisiae* IA3 was preferable to an *S. castellii* IA3 tail on an *S. castellii* inhibitory sequence (compare **42** and **47**), but this effect was negated by translocation of the His-tag from C- to the N-terminus (**43** and **51**);

(ii) in $NT(H)_6$ form, comparable inhibitory potency was observed in the presence of a full-length *S. castellii* IA3 tail (**43**), an *S. cerevisiae* IA3 tail (**51**) or an *S. castellii* IA3 tail truncated to the same length as that from *S. cerevisiae* IA3 (**50**);

(iii) all of the variants examined exhibited the pH switch effect, except for the synthetic peptide that contained no tail residues or His-tag (**44**).

Consequently, the interplay between the tail region, His-tag and the observed pH switch was of considerable interest. Further investigations into these aspects will be described in the next chapter.

Nevertheless, it was possible to produce recombinant forms of *S. castellii* IA3 with equivalent, or even superior, potencies to the synthetic peptide at pH 4.7.

6.4 Comparing the hydrophobic pins of *S. cerevisiae* and *S. castellii* IA3

As described in Chapter 5, most of the binding energy for the inhibition of *S. cerevisiae* vacuolar proteinase A by *S. cerevisiae* IA3 is derived from hydrophobic

interactions contributed by Leu19 and the two clusters of the front and back pins (see also Section 1.8.5).

Comparison of the *S. cerevisiae* and *S. castellii* IA3 sequences revealed that Leu19 is conserved in both species and the front pin region:

residue number	**8**	**9**	**10**	**11**	**12**	**13**
S. cerevisiae IA3:	V	S	E	I	F	Q
S. castellii IA3:	V	S	E	M	F	Q

has only a single, relatively conservative change at position 11. However, the back pin region exhibited more substantial differences in sequence:

residue number	**25**	**26**	**27**	**28**	**29**	**30**
S. cerevisiae IA3:	V	V	S	D	A	F
S. castellii IA3:	A	A	S	E	G	M

The most significant of these changes appeared to be V to A at position 26, and AF to GM at positions 29 and 30. Therefore, a recombinant C-terminally His-tagged *S. cerevisiae* IA3 mutant was produced incorporating the equivalent residues from *S. castellii* IA3 in these positions. The V26A/A29G/F30M mutant (**52** in Table 6.3) remained so potent at pH 4.7 as to preclude accurate comparison with wild-type *S. cerevisiae* IA3. However, at pH 3.1, the K_i value (30 ± 4 nM) was 30-fold poorer than the wild-type.

During the creation of the DNA encoding the V26A/A29G/F30M IA3 mutant (**52**) by oligonucleotide cassette mutagenesis (Section 3.2.3), one clone was revealed by double stranded DNA sequencing (Section 2.2.11) to contain a further, somewhat fortuitous Q20E mutation that replaced the Gln of *S. cerevisiae* IA3 with the Glu of the *S. castellii* IA3 sequence (**53** in Table 6.3). Despite the potential for an (*i*, *i*+4) interaction with either Lys16 or Lys24 thus introduced (Iqbalsyah & Doig, 2005), the inhibitor remained of comparable potency to **52**, as assessed at pH 3.1 (15 ± 3 nM;

30 ± 4 nM for **53** and **52**, respectively). This suggested that alteration of the sequence at residue 20 has little effect on potency.

The decline in potency at pH 3.1 of both **52** and **53** relative to the wild type inhibitor (1.1 ± 0.4 nM) was in contrast to the measurements obtained from a synthetic peptide containing only the V26A mutation (**54** in Table 6.3). In isolation, although no effect was observed at pH 4.7, the V26A mutation produced an inhibitor with a marginally improved potency at pH 3.1 (10 ± 2 nM), relative to the equivalent wild type sequence (**55**, 25 ± 8 nM). Therefore, the A29G and F30M mutations would appear to be disadvantageous at pH 3.1, at least in an otherwise *S. cerevisiae* IA3 sequence context.

The role of sequence context was examined further by the production of a recombinant form of *S. cerevisiae* IA3 where residues 18-34 were replaced with their counterparts from *S. castellii* IA3 (for details of the construction method see Sections 3.2.3 and 3.4). Therefore, both the centrepiece and back pin regions of *S. castellii* IA3 were introduced simultaneously into *S. cerevisiae* IA3. Individually, either of these sequence elements apparently had been well tolerated by *S. cerevisiae* IA3 at pH 4.7 (see **52**, **53** and **39** in Table 6.3). The only additional mutations introduced by this chimaeric interchanging of 17 residues were K24N, V25A and D28E, all of which appeared to be reasonably conservative changes to less critical residues.

However, the potency of the chimaera (**56** in Table 6.3) at pH 4.7 was at least 20-fold poorer than either wild type *S. cerevisiae* IA3 (**1**) or the back pin mutants (**52** and **53**). The K_i value was also 10-fold worse than the *S. cerevisiae* IA3 mutant containing the *S. castellii* M18/K22 centrepiece arrangement (**39**), and comparable to that demonstrated by recombinant C-terminally His-tagged *S. castellii* IA3 (**42** in Table 6.1). At pH 3.1, essentially all activity was removed (K_i = 500 ± 100 nM), although this was not entirely unexpected because Lys22 had been introduced into the centrepiece. Therefore, the simultaneous introduction of both the *S. castellii* IA3 back pin and the centrepiece residues into *S. cerevisiae* IA3 is somewhat disadvantageous.

To investigate whether the back pin residues of *S. cerevisiae* IA3 would be of benefit in a *S. castellii* inhibitory sequence, appropriate reciprocal mutations were introduced. However, in order to avoid the pH switch complication at pH 3.1 which occurs in recombinant protein forms of *S. castellii* IA3 (see Section 6.2, and Chapter 7), the study was carried out using synthetic peptides spanning residues 2-34.

Consistent with the data derived from the *S. cerevisiae* IA3 context just described, replacement of only the residues at 29 and 30 with those from *S. cerevisiae* IA3 reduced the potency by approximately 8-fold (compare **57** and **44** in Table 6.3). However, this effect could be partially ameliorated by the simultaneous introduction of a further mutation towards the *S. cerevisiae* IA3 back pin sequence at position 25, with the mutant A25V/G29A/M30F (**58** in Table 6.3) suggesting some cooperation favouring interaction among the 25 and 29/30 positions. Indeed, with a Val residue at position 25, either *S. cerevisiae*- or *S. castellii*-derived residues at positions 29 and 30 were equally favoured (compare **59** and **58** in Table 6.3). Furthermore, substitution of Ala25 with Ile instead of Val maintained almost exactly the same inhibitory values (see **60** in Table 6.3). Possibly, therefore, Ala at 25 compromised the inhibitory potency of *S. castellii* IA3 mildly, due to the small size of this residue.

In a final assessment of the effect of interspecies transposition of back pin residues, the entire segment of residues 24-34 of *S. castellii* IA3 was replaced with that of *S. cerevisiae* IA3. This inhibitor (**61** in Table 6.3) had comparable potencies at both pH values to those of peptide **58** (and **59** and **60**). Therefore the additional exchange of residues at positions 24, 26, 28 and 32 had little effect. Indeed, despite the introduction of no less than seven mutations, including all of the back pin residues, **61** retained a potency at both pH values that was comparable to the wild type *S. castellii* IA3 peptide (**44**) of the same length.

Therefore, whilst the introduction of extensive regions of *S. castellii* IA3 into the *S. cerevisiae* inhibitory sequence had been modestly detrimental to potency, the reciprocal change of *S. cerevisiae* residues into *S. castellii* IA3 appeared to have relatively little effect. Both sequences demonstrated a remarkable ability to accommodate the critical binding residues of the other, as long as the correct combination of residues at the appropriate relative positions was maintained.

6.5 Discussion

One of the most intriguing aspects of *S. cerevisiae* IA3 is the exquisite specificity of interaction, with activity apparently solely directed against the vacuolar proteinase from its own species. The discovery of new IA3 sequences presented an opportunity to investigate whether IA3 sequences are truly specific in this manner, or if activity could cross the species barrier. Of the five new sequences (Figure 6.1B), only that from *S. castellii* differed substantially from *S. cerevisiae* IA3 in the N-terminal 34 amino acids, although Leu19 and the residues corresponding to the front hydrophobic pin (Val^{8}-X-X-hydrophobic-Phe^{12}) were conserved, pointing to critical roles for these residues in the new IA3 sequences as well. IA3 from *S. castellii* was also significantly longer than the other inhibitors, although all five new sequences were considerably more diverse in the "tail" region after residue 35. Also, many of the differences in the *S. castellii* IA3 sequence were within regions identified as critical to the binding of *S. cerevisiae* IA3 to its own target proteinase. This provided a further insight into the potential for degeneracy within the IA3 sequence.

The sequences of the corresponding vacuolar aspartic proteinases for the new IA3 forms were also mined from the databases by my colleague, Dr. D. Wyatt (Cardiff University). With the exception of *S. castellii* proteinase, all the enzymes contained identical residues at all of the positions identified in *S. cerevisiae* proteinase A by Li *et al.* (2000) from the crystal structure to contact the inhibitory α-helix of *S. cerevisiae* IA3. This might explain the minimal differences in sequence amongst the IA3 molecules from *S. paradoxus*, *S. mikatae*, *S. kudriavzevii* and *S. bayanus*.

Taken together, these findings imply that: (a) the new sequences (including that from *S. castellii*) are indeed aspartic proteinase inhibitors, and (b) the mode of inhibition employed might be similar or comparable. As the most divergent of the sequences, the *S. castellii* IA3 was selected for production in various recombinant and synthetic peptide forms. Despite the extensive differences in the sequence, *S. castellii* IA3 displays very potent inhibition of *S. cerevisiae* vacuolar proteinase A. Furthermore,

in an equivalent manner to *S. cerevisiae* IA3, the activity resides in the N-terminal 34 residues.

However, at least two differences with *S. cerevisiae* IA3 are apparent:

(i) the C-terminal "tail" (residues 35-81) is not neutral in terms of the inhibition and appears detrimental to activity;

(ii) His-tag relocation to the N-terminus results in >10-fold improvement in potency at pH 4.7.

The discovery that the 2-34 synthetic peptide form of *S. castellii* IA3 is 10-fold more potent than the C-terminally His-tagged recombinant form at pH 4.7 is in stark contrast to *S. cerevisiae* IA3, where either form had been equally potent (Chapter 5). The *S. castellii* IA3 tail, therefore, modulates the activity of its inhibitory sequence, resulting in a less potent inhibitor, whereas the *S. cerevisiae* IA3 tail does not affect the potency of either inhibitory sequence. This suggestion that the *S. castellii* IA3 tail must be interacting in a distinct manner to the *S. cerevisiae* IA3 tail is supported by the chimaeric N*cast*-C*cere* IA3 (**47** in Table 6.2). In this mutant, where the *S. castellii* IA3 tail sequence replaces that of *S. cerevisiae* IA3, no detrimental effect of tail attachment is observed, and the K_i value is equivalent or even modestly superior to the tailless synthetic peptide (**44** in Table 6.1).

Rather than the *S. castellii* tail sequence itself, however, another possible origin of the detrimental effect could be the His-tag. Although *S. cerevisiae* IA3 is unaffected by a His-tag (Li *et al.*, 2000; and Chapter 5), the sequence of *S. castellii* IA3 may be sufficiently different to enable some form of interaction. As an unnatural addendum to the sequence, with a high charge density and considerable propensity to chelate transition metal ions or act as a hydrogen bond donor or acceptor, it certainly does not fulfil the characteristics of an inert tag. However, the potency of the N*cast*-C*cere* chimaeric IA3 (**47** in Table 6.2) appears to discount a role for a C-terminal His-tag. This chimaeric *S. castellii* IA3-derived protein with the *S. cerevisiae* IA3 tail is an inhibitor of at least equivalent potency to the 2-34 synthetic *S. castellii* IA3 peptide (**44** in Table 6.1), despite the presence of a His-tag at the C-terminus. However, the data are also consistent with the His-tag mediating these

effects only upon attachment to the *S. castellii* IA3 tail, possibly as a result of the increased length.

Relocation of the His-tag to the N-terminus improves potency by 10-fold, implying that the His-tag does, indeed, affect the inhibitory potency of *S. castellii* IA3 (compare **43** with **42** in Table 6.1). As the tailless and tagless synthetic peptide (**44** in Table 6.1) has an almost identical K_i value, one interpretation might be that the detrimental effect is removed by this relocation. However, the improvement can also be explained as an advantageous effect of the NT$(H)_6$ tag obscuring a persisting detrimental effect of the *S. castellii* IA3 tail. Notably, however, His-tag relocation to the N-terminus does NOT benefit the N*cast*-C*cere* chimaera (compare **47** and **51** in Table 6.2), suggesting that any benefit of relocation of the His-tag is also linked to the tail species.

Ultimately, the only available method to resolve this complicated situation is removal of the His-tag entirely from all of the variants, and comparison of the K_i values of the N- and C-terminally His-tagged forms with the tagless forms and the synthetic peptide. This aspect is also developed further in the next chapter.

However, regardless of the detrimental effect(s) afflicting some forms, the fact remains that the *S. castellii* IA3 inhibitory sequence is less potent than *S. cerevisiae* IA3 against *S. cerevisiae* vacuolar proteinase A, even when in the "ideal" peptide form. This may reflect that, of the two inhibitory species, only *S. cerevisiae* IA3 has encountered co-evolutionary pressures with specifically the vacuolar proteinase A of *S. cerevisiae* to ensure a harmonised interaction. Therefore, some regions of the *S. castellii* inhibitory sequence are likely to be less ideally suited to the task than their *S. cerevisiae* equivalents, providing an opportunity to investigate the constituent contributions in a new light. Arguably the back pin region displays the most substantial differences between the two inhibitory sequences.

The kinetic data certainly indicate that the individual residues of the *S. castellii* IA3 back pin are generally less suited to inhibiting the *S. cerevisiae* proteinase A than their *S. cerevisiae* IA3 counterparts (Section 6.4). Although the *S. castellii* IA3 back pin constituents retain the same hydrophobic nature of the equivalent *S. cerevisiae* IA3

residues, they are all smaller and, perhaps, less able to produce such strong hydrophobic interactions with the active site of the *S. cerevisiae* proteinase A.

With the reduced size of the back pin residues of its IA3, the possibility was considered that the partner *S. castellii* proteinase might contain compensatory larger residues in the appropriate region of its active site cleft. Throughout all the residues predicted to contact the inhibitory helix (based on its high sequence identity with the *S. cerevisiae* proteinase A), the *S. castellii* enzyme sequence contains only two differences (for an alignment, see Figure 8.1). Both changes are on the polyproline loop (Section 1.8.5), which contributes particularly to the binding pocket for the back pin residues at positions 26 and 30 of IA3. Figure 6.2 provides a stereo image of the relevant area of the *S. cerevisiae* proteinase A-IA3 complex. Both replacements in the *S. castellii* proteinase are, indeed, for relatively longer-reaching residues (both Val295 (green) and Leu298 (pink) are replaced with Met). The relatively open nature of the binding pocket at this end of the cleft in *S. cerevisiae* vacuolar proteinase A might explain the apparent preference for the "natural" Phe over the Met residue of *S. castellii* IA3 residue at 30. However, the replacement of Val26 with the smaller Ala from *S. castellii* IA3 appeared modestly beneficial in the kinetic experiments. It is possible that Tyr189 of the *S. cerevisiae* proteinase A (marked with ⇦ in Figure 6.2) is partially deflected from the location in the free enzyme location by Val26 when the *S. cerevisiae* IA3 binds. Inspection of the crystal structure (see Figure 6.2) indicates that the beneficial effect of the smaller Ala26 possibly reflects a decreased interference with Tyr189. This point will be dealt with in more detail in the General Discussion, Chapter 11.

On the opposite face of the active site from the poly-proline loop, the residues which line the main body of the *S. cerevisiae* proteinase A cleft (Figure 6.3) are exactly the same in the proteinase from *S. castellii*. Residues 25 and 29 (on adjacent turns of the IA3 helix) interact with this region and with each other. These intrahelical interactions may be an important contributory factor to potency in the back pin region.

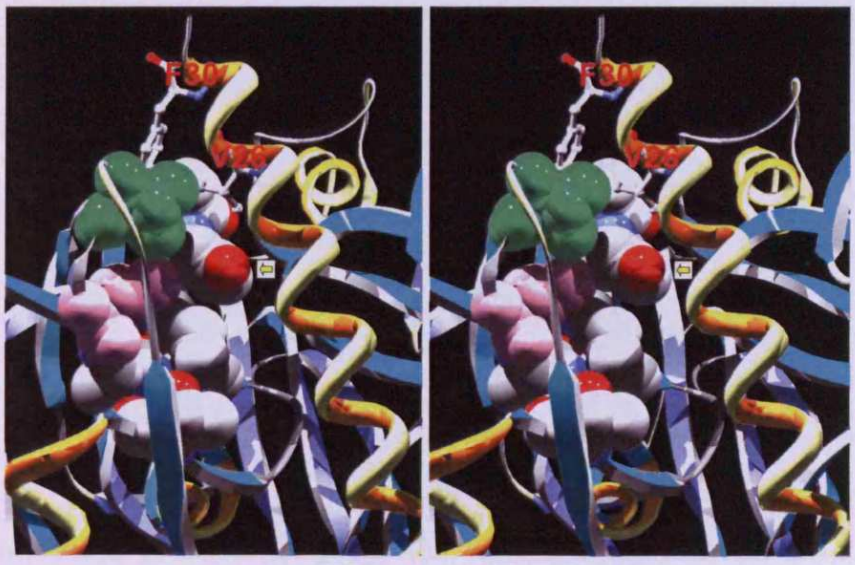


Figure 6.2 Stereo representation of the interaction of Val26 and Phe30 on adjacent turns of the *S. cerevisiae* IA3 inhibitory α-helix with the active site of *S. cerevisiae* vacuolar proteinase A. Val295 (green) and Leu298 (pink) on the poly-proline loop are both replaced by Met in the *S. castellii* proteinase. Tyr189 of the proteinase (marked with ⇦) is conserved in both proteinases.

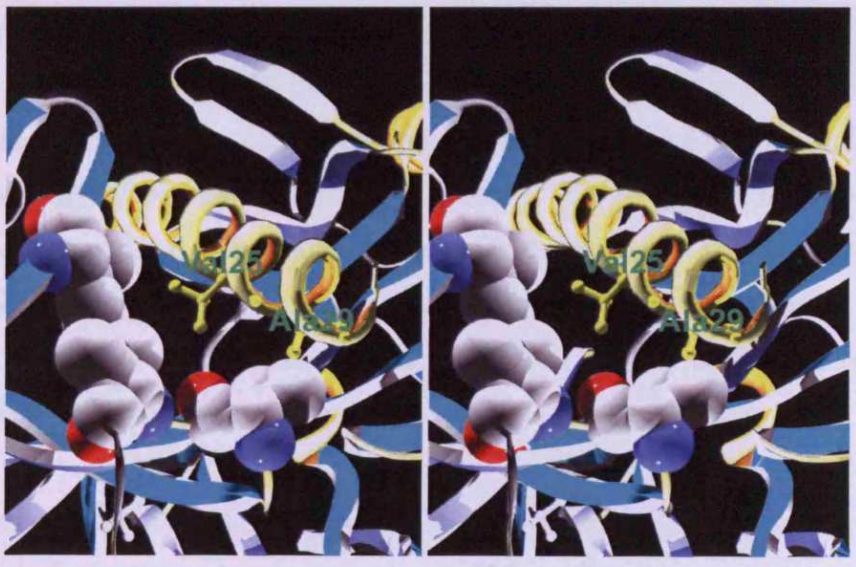


Figure 6.3 Stereo representation of the interaction of Val25 and Ala29 on adjacent turns of the *S. cerevisiae* IA3 α-helix with the active site cleft of *S. cerevisiae* vacuolar proteinase A. Val25 and Ala29 are represented as ball and stick and coloured yellow.

The *S. castellii* IA3-based peptide **57** (Table 6.3), which introduces the *S. cerevisiae* IA3 residues at positions 29 and 30 only, with the "wrong partners" at positions 25 and 26 appears to confirm this proposition. Peptide **57** is almost 10-fold less potent than the equivalent wild type *S. castellii* peptide which has the "correct" 25/29 and 26/30 pairings on the helix (**44** in Table 6.3). Similarly, the introduction of the "correct" *S. cerevisiae* IA3 25/29 and 26/30 pairings by replacement of the entire back pin region in peptide **61** (Table 6.3) returns the K_i value to almost that of wild-type peptide **44**. Therefore, it appears that the system can largely compensate, as long as compatible residues are placed in appropriate positions. That is, the *context* may be at least as important as the individual residues.

Such a conclusion might also explain the effect of the simultaneous introduction of both the *S. castellii* IA3 centrepiece and back pin regions into *S. cerevisiae* IA3. The resulting chimaeric protein (**56** in Table 6.3) is much less potent at pH 4.7 than might have been predicted from a consideration of the introduction of the two regions individually (**39**, **52** and **53** in Table 6.3). It seems, therefore, that inappropriate residues or sub-optimal interactions at one position, for example the back pin region, can largely be compensated for elsewhere, e.g. the centrepiece, and thus accommodated with minimal effect on the K_i value. However, anchoring both regions inappropriately makes such compensatory adjustments much harder.

Overall, though, both sequences appear to tolerate the introduction of some or all of the back pin region from the other sequence. This again reinforces the conclusion from Chapter 5 that, if compatible adjustments are made and the correct combination of residues is introduced, the IA3 sequence can accommodate substantial changes with minimal effect on potency.

New structural data would help identify the origin of several of the effects described in this chapter. Therefore, a complex of $NT(H)_6$ wild type *S. castellii* IA3 with *S. cerevisiae* vacuolar proteinase A was prepared on a large scale and has been supplied to our collaborating crystallographic laboratory (Dr Alex Wlodawer, National Cancer Institute, Frederick, MD, USA). At the time of writing, however, no crystallographic data have been obtained. Nevertheless, the kinetic data accumulated in this chapter strongly suggest that *S. castellii* IA3 inhibits *S. cerevisiae* vacuolar

proteinase A (and probably its own proteinase also) in an analogous manner to that of *S. cerevisiae* IA3, with the activity once again located in the N-terminal 34 residues of the molecule.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM) pH 4.7	K_i (nM) pH 3.1	pH switch
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	1.1 ± 0.4*	>11
39	*S. cerevisiae* K18M/D22K IA3	MNTDQ QKVSE IFQSS KEMLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.2 ± 0.1	85 ± 15	425
42	CT(H)$_6$ WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~~~LE(H)$_6$	4 ± 0.5	350 ± 70	88
43	NT(H)$_6$ WT *S. castellii* IA3	MK(H)$_6$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~~~	0.3 ± 0.2	125 ± 30	417
44	WT *S. castellii* IA3 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	0.4 ± 0.1	8 ± 1	20

Table 6.1 Inhibition of *S. cerevisiae* vacuolar proteinase A by different forms of wild-type (WT) *S. castellii* IA3. Wild-type *S. cerevisiae* IA3, and a mutant form of *S. cerevisiae* IA3 (**39**) with the same centrepiece residues as *S. castellii* IA3 are also included for comparison. Sequence derived from *S. cerevisiae* IA3 is coloured in yellow, *S. castellii* IA3 sequence is in blue. Where mutations directly swapped equivalent residues between species, the mutation is represented by the appropriate colour. Recombinant proteins were produced with either a C-terminal LE(H)$_6$ tag or N-terminal MK(H)$_6$HMQ tag. All recombinant proteins show only the residues of the inhibitory sequence (1-34), with the C-terminal tail region depicted as~~~~~~~~. The increased length of *S. castellii* IA3 (81 residues instead of 68) is reflected in this. For full sequence, see Appendix I. Z = L-norleucine. The "pH switch" represents the ratio of K_i values between the pH values. *= from Phylip *et al.* (2001).

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM) pH 4.7
45	N*cere*-C*cast* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~~~~~LE(H)$_6$	<0.1
46	N*cere*TM-C*cast* IA3	MNTDQ QKVSE IFQSS KEIMQ GDAKV VSDAF KKMA~~~~~~~~~~~~LE(H)$_6$	<0.1
47	N*cast*-C*cere* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	0.1 ± 0.1
48	*S. castellii* M18V IA3	MSDKN ANVSE MFQQA KEVLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	125 ± 20
49	N*cast* M18V-C*cere* IA3	MSDKN ANVSE MFQQA KEVLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	25 ± 2
38	*S. cerevisiae* K18V/D22K IA3	MNTDQ QKVSE IFQSS KEVLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	2 ± 0.4
50	NT(H)$_6$ *S. castellii* Q68 IA3	MK(H)$_7$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~[68]	0.8 ± 0.1
51	NT(H)$_6$ N*cast*-C*cere* IA3	MK(H)$_7$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	0.7 ± 0.2

Table 6.2 The effect of tail identity on inhibitory potency. Sequence derived from *S. cerevisiae* IA3 is coloured in yellow, *S. castellii* IA3 sequence is in blue. Mutations are highlighted in red. Recombinant proteins were produced with either a C-terminal LE(H)$_6$ or N-terminal MK(H)$_6$HMQ tag. All recombinant proteins show only the residues of the inhibitory sequence (1-34), with the C-terminal tail region depicted as~~~~~~~~. The increased length of *S. castellii* IA3 (81 residues instead of 68) is reflected in this.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM) pH 4.7
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1
39	*S. cerevisiae* K18M/D22K IA3	MNTDQ QKVSE IFQSS KEMLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.2 ± 0.1
52	*S. cerevisiae* V26A/A29G/F30M IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV ASDGM KKMA~~~~~~~~LE(H)$_6$	<0.1
53	*S. cerevisiae* Q20E/V26A/A29G/F30M IA3	MNTDQ QKVSE IFQSS KEKLE GDAKV ASDGM KKMA~~~~~~~~LE(H)$_6$	<0.1
54	*S. cerevisiae* V26A 2-30	NTDQ QKVSE IFQSS KEKLQ GDAKV ASDAF	6 ± 1*
55	WT *S. cerevisiae* 2-30	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF	5 ± 1*
56	N*cast Sac*I/*Nhe*I IA3	MNTDQ QKVSE IFQSS KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	2 ± 0.3
44	WT *S. castellii* IA3 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	0.4 ± 0.1
57	*S. castellii* G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEAF KNZA	3 ± 0.4
58	*S. castellii* A25V/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEAF KNZA	1.1 ± 0.1
59	*S. castellii* A25V 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEGZ KNZA	1.2 ± 0.3
60	*S. castellii* A25I/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANI ASEAF KNZA	1.4 ± 0.2
61	SEX 4	SDKN ANVSE ZFQQA KEZLE GKAKV VSDAF KKZA	0.8 ± 0.1

Table 6.3 The effect of mutating the back pin residues. Wild-type *S. cerevisiae* IA3, and a mutant form of *S. cerevisiae* IA3 (**39**) with the same centrepiece residues as *S. castellii* IA3 are also included for comparison. Sequence derived from *S. cerevisiae* IA3 is coloured in yellow, *S. castellii* IA3 sequence is in blue. Where mutations directly swapped equivalent residues between species, the mutation is represented by the appropriate colour. Other mutations are highlighted in red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag and only the residues of the inhibitory sequence (1-34) are depicted. The C-terminal tail region is represented as~~~~~~~~. Z = L-norleucine. *= from Phylip *et al.* (2001).

Chapter 7

Elimination of the pH switch from *S. castellii* IA3

7.1 Introduction

In the work described in Chapters 5 and 6, all recombinant protein forms of IA3 containing a basic residue at position 22 displayed the so-called "pH switch" phenomenon. An explanation for this catastrophic loss of potency at pH 3.1, relative to the K_i value at pH 4.7, was proposed for the *S. cerevisiae* IA3 forms in Section 5.7. Following the identification of a pH switch in recombinant *S. castellii* IA3, which naturally encodes a basic Lys residue at position 22, it was considered likely that the origin of the effect was the same in the new IA3 species.

However, the observation that the "tailless" synthetic peptide representing only residues 2-34 of *S. castellii* IA3 retained high levels of potency at ***both*** pH values (Section 6.2), thus removing the pH switch, suggested that the basis for the effect in recombinant *S. castellii* IA3 might differ from that in *S. cerevisiae* IA3. Furthermore, the effect of truncation of *S. cerevisiae* IA3 molecules displaying the pH switch had not been considered previously. Therefore, this chapter investigates possible contributions of the C-terminal tail region in the pH switch in *S. cerevisiae* IA3 mutants containing a basic residue at 22, and establishes the origin of the effect in *S. castellii* IA3.

7.2 The role of the C-terminal tail in the potency of *S. cerevisiae* IA3 molecules containing a basic residue at position 22

Before examining the effect of tail removal on the pH switch in *S. cerevisiae* IA3 molecules containing a basic residue at position 22, the situation in *S. castellii* IA3 was investigated more extensively. K_i values were determined at pH 3.1 for the five 2-34 synthetic peptides (**57** - **61** in Table 7.1) that were based upon the *S. castellii* IA3 inhibitory sequence (Section 6.4). In all cases, they retained reasonable potencies at the lower pH. The pH switch, therefore, was abolished from all five, confirming the findings of wild-type *S. castellii* peptide (**44** in Table 7.1) that the pH switch was removed from *S. castellii* IA3 in the absence of the tail region.

The *S. cerevisiae* IA3 mutant K18D/D22K (**36** in Table 7.1) had been shown previously to exhibit the pH switch phenomenon (Section 5.6). Therefore, to investigate the possible role of the C-terminal half in the pH switch observed in this molecule, a peptide corresponding to only residues 2-32 of the mutant IA3 sequence was tested against the *S. cerevisiae* proteinase A at both pH values. Removal of the tail residues from this mutant (**62** in Table 7.1) resulted in such a catastrophic loss of potency at both pH values (compare **62** with **36**) that it was not possible to assess the effect of the tail on the pH switch in *S. cerevisiae* IA3. This contrasts starkly with the effect of such truncations on the *S. castellii* IA3, and suggests fundamentally different roles for the C-terminal tails between the two IA3 species. Therefore, although the phenomenon was found in certain examples of both *S. cerevisiae* and *S. castellii* IA3-derived sequences, the origin of the pH switch was apparently distinct, prompting further investigation of the basis of the pH switch in *S. castellii* IA3.

7.3 The role of Lys22 in the pH switch in *S. castellii* IA3

In an *S. cerevisiae* IA3 sequence context, the introduction of a basic residue at 22 had been identified as a critical factor in the pH switch. To examine the potential role of Lys22 in the effect observed in recombinant proteins containing the *S. castellii* IA3 inhibitory sequence (residues 2-34), a series of proteins was produced with alternative residues at position 22. Transposition of the residues at positions 18 and 22 to produce the *S. castellii* IA3 mutant M18K/K22M resulted in a catastrophic loss of potency at both pH values (compare **63** with **42** in Table 7.2), and the corresponding mutant peptide consisting of only residues 2-32 was just as poor (**64** in Table 7.2). The mutant protein M18K/K22L (**65** in Table 7.2) was even worse. These results in the *S. castellii* IA3 sequence context were in complete contrast to the findings described earlier when these two centrepiece combinations (K18/M22 and K18/L22) were introduced into *S. cerevisiae* IA3. In the *S. cerevisiae* IA3-based sequences, both centrepieces were approximately equipotent to the wild type *S. cerevisiae* IA3 sequence (compare **1**, **20** and **21** in Table 5.2). Thus, the *S. castellii* IA3 would appear to be rather less tolerant of centrepiece manipulations than *S. cerevisiae* IA3,

at least for the inhibition of *S. cerevisiae* vacuolar proteinase A. However, the *S. castellii* IA3 mutant M18K/K22D, containing the *S. cerevisiae* IA3 centrepiece combination of K18 and D22, showed comparable potency to that of *S. castellii* wild-type IA3 at pH 4.7 (compare **66** and **42** in Table 7.2) but, perhaps more significantly, the pH switch was not evident.

Thus, the simple exchange of Met18 to Lys and Lys22 to Asp was sufficient to restore the potency of inhibition lost from the full-length wild type *S. castellii* sequence at pH 3.1. With this centrepiece combination, the presence of residues 34-81 together with the C-terminal His-tag appeared to be without influence because the synthetic peptide form of this M18K/K22D mutant consisting of only residues 2-32 (**67** in Table 7.2) showed comparable potencies at each pH value to the full length protein. Notably, however, the K_i values measured for **66** at both pH values were poorer by at least 30-50 fold compared to those derived for wild type *S. cerevisiae* IA3 (**1** in Table 7.1). Clearly, then, and perhaps not surprisingly, the inhibitory sequence from *S. cerevisiae* IA3 is a more effective inhibitor of the *S. cerevisiae* proteinase A than the counterpart sequence from *S. castellii* carrying the equivalent centrepiece (M18K/K22D).

The contribution to the pH switch of positioning a Lys residue at 22 was confirmed by introduction of the single mutation K22L into the *S. castellii* IA3 sequence (**68** in Table 7.2). At pH 4.7, this protein was marginally more potent than the wild-type counterpart (compare **68** and **42** in Table 7.2) and also the *S. castellii* M18K/K22D mutant IA3 (**66**). At pH 3.1, however, **68** was much more potent than the wild-type *S. castellii* inhibitor to the extent that, once again, the pH switch was removed. Indeed, in absolute terms, the respective K_i values recorded for **68** at pH 4.7 and pH 3.1 were each within an order of magnitude of the values derived for wild-type *S. cerevisiae* IA3 (**1** in Table 7.1).

Notably, the 2-32 synthetic peptide form (**69** in Table 7.2) of the K22L mutant was of comparable potency at pH 4.7 to the full length form (**68**). Furthermore, the potency of the peptide at pH 3.1 was, for the first time with *S. castellii*-based IA3 sequences, indistinguishable from the value measured at pH 4.7. This was in contrast to the full-length form (**68**), the potency of which fell by approximately one order of magnitude

between pH values. In fact, at pH 3.1 the potency of *S. castellii* K22L was approximately 5-fold better as a 2-32 peptide than in full-length recombinant form (compare **68** and **69**).

This result revealed that, even in the absence of Lys22 (and therefore the pH switch), an element of the *S. castellii* IA3 tail still prevented the recombinant protein (**68**) from achieving the potency of the tailless peptide at either pH. Such a finding was in clear contrast to *S. cerevisiae* IA3-based inhibitors where removal of the tail was never beneficial. This, again, highlighted the differing role of the C-terminal tail in the two inhibitory sequences.

7.4 Removing the pH switch from *S. castellii* IA3 without altering the inhibitory sequence

Although replacement of Lys22 eliminated the pH switch from recombinant forms of *S. castellii* IA3, this residue could not be solely responsible for the phenomenon because synthetic forms of *S. castellii* IA3 containing Lys22 did not display a pH switch. Therefore, another factor also had to have a role.

Comparison of the *S. castellii* IA3 synthetic peptides with all of the recombinant forms of *S. castellii* IA3 demonstrating the pH switch (i.e. those still containing Lys22) revealed three differences:

(i) all synthetic peptides contained the artificial amino acid Nle in place of Met;
(ii) the peptides were "tailless", i.e. lacked any sequence after residue 34;
(iii) the peptides did not have a His-tag at either N- or C-terminus.

Direct examination of the effect of replacing Met with Nle in peptides was not feasible because of the well-documented complications arising from oxidation of Met sidechains. Indeed, it was for this reason that Nle was introduced as a substitute for Met in peptides made by solid-phase synthesis. Nevertheless, numerous peptides based upon *S. cerevisiae* IA3 sequences which contained Nle in place of Met had

shown potencies at both pH 4.7 and pH 3.1 that were almost exactly coincident with the equivalent recombinant protein (see Chapter 5, and Phylip *et al.*, 2001). Furthermore, it was shown in the previous chapter that, at pH 4.7, the wild-type *S. castellii* IA3 synthetic peptide (**44** in Table 6.1) behaved very similarly to both the N-terminally His-tagged *S. castellii* IA3 (**43** in Table 6.1) and both His-tagged forms of the N*cast*-C*cere* chimaera (**47** and **51** in Table 6.1), all of which contained Met. From these data, and the consideration that Met and Nle both have closely similar hydrophobic, uncharged sidechains differing only by a methylene group in place of a S atom, it was considered unlikely that the presence of one over the other would mediate a pH-dependent effect.

Discrimination between the other two variables, however, was not possible whilst the His-tag was attached to the C-terminal tail region. Therefore, it was necessary to reposition the His-tag in the proteins to the N-terminus. This not only separated the variables but also permitted subsequent proteolytic removal of the tag using the TAGZyme™ system.

The possible roles of tail length and tail sequence identity in the pH switch were investigated using truncated forms of *S. castellii* and the N*cast*-C*cere* IA3 chimaera. The *S. castellii* IA3, which had been C-terminally truncated to the same length as the N*cast*-C*cere* chimaera without effect previously (**50** in Table 6.2, described in Section 6.3), was truncated after residues Tyr57 and Lys45. These positions, which were equivalent to those described in Chapter 5 for the truncated forms of *S. cerevisiae* IA3, were also applied to the truncation of the N*cast*-C*cere* chimaera (i.e. after Tyr57 and Ala45). Shortening the *S. castellii* IA3 polypeptide had only a small effect on inhibitory potency at pH 4.7 (compare **70** and **71** with **43** and **50** in Table 7.3) and, because all of these forms were very poor inhibitors at pH 3.1, the pH switch remained. Similarly, the tests of the N*cast*-C*cere* truncation series (**51**, **72**, and **73** in Table 7.3) revealed that the pH switch had not been removed by truncation of the alternative tail species either.

Comparison of the sequences of IA3 from *S. castellii* and *S. cerevisiae* (Figure 6.1B) reveals that the tail residues are quite dissimilar, except for a short conserved region between residues 57-64. However, this common sequence element did not, in itself,

appear to be responsible for the pH switch because the removal of this region from either *S. castellii* IA3 or the N*cast*-C*cere* chimaera by the respective Y57-truncations (**70** and **72** in Table 7.3) did not remove the phenomenon. Furthermore, truncation to within 12 residues of the length of the 2-34 synthetic peptide (**71** and **73** in Table 7.3) did not remove the pH switch either.

The His-tag therefore remained as the key difference between the peptides and the recombinant proteins. After five cycles of N-terminal proteolytic trimming using the DAPase™ enzyme of the TAGZyme™ system (Sections 2.2.24 and 4.3.3), the removal of the His-tag from both *S. castellii* IA3 and the N*cast*-C*cere* chimaera was confirmed by MALDI-TOF mass spectrometry (Section 2.2.28). For example, in the case of *S. castellii* IA3, the observed mass was 9162.8 Da which was consistent with the theoretical mass (9160.3 Da) of residues 2-81 of *S. castellii* IA3 plus the N-terminal pyroglutamate residue that represented the stop point of TAGZyme™ cleavage.

The tagless form of N*cast*-C*cere* (**74** in Table 7.3) had comparable potencies at both pH values to the counterpart containing the N-terminal His-tag (**51**). However, the pH switch still remained. Notably, removal of the N-terminal His-tag from *S. castellii* IA3 (**75** in Table 7.3) returned the inhibitory potency of the molecule at pH 4.7 to the value (K_i = 4 ± 0.5 nM) described earlier for the C-terminally His-tagged form (**42** in Table 7.1). Similarly, His-tag removal from the *S. castellii* K45-truncation mutant resulted in an identical 10-fold drop in potency at pH 4.7 (compare **76** and **71** in Table 7.3), resulting in a similar K_i value to the tagless full-length protein (**75** in Table 7.3). These results reprised the findings described for the various tagged and tagless forms of wild-type *S. cerevisiae* IA3 in Chapter 5 (Table 5.1), where it was concluded that a C-terminal His-tag did not modify the inhibitory potency of the molecule. Furthermore, this indicated that the observed improvement in potency at pH 4.7 upon relocation of the His-tag from C- to N-terminus of *S. castellii* IA3 (described in Chapter 6) represented a beneficial effect, rather than the removal of a detrimental effect mediated by a C-terminal His-tag.

Remarkably, however, the ratio of inhibitory constants at pH 3.1 to pH 4.7 was much reduced. Therefore, neither recombinant tagless form of *S. castellii* IA3 (**75** and **76**) displayed the pH switch. This suggested that the pH switch originated from disadvantageous interactions mediated by a His-tag at either the N- or C-terminus of an *S. castellii* IA3 molecule that contained a Lys residue at position 22.

7.5 Discussion

The initial aim of the studies described in this chapter was to examine the factors responsible for the observed "pH switch" in *S. castellii* IA3, where the inhibitor was potent at pH 4.7 but lost most, if not all, of its activity at pH 3.1. In the process, these studies also provided a greater insight into the detrimental effect of the C-terminal tail of *S. castellii* IA3 to its own inhibitory sequence (see Section 6.5). This discussion is therefore divided into two sections: the origin of the *S. castellii* IA3 pH switch and the interactions of the C-terminal tail of *S. castellii* IA3.

7.5.1 The role of the centrepiece, and eliminating the pH switch from *S. castellii* IA3

The centrepiece of *S. castellii* IA3 was seemingly more sensitive to substitutions than *S. cerevisiae* IA3. As discussed in Section 6.5, this probably reflects the fact that, unlike *S. cerevisiae* IA3, the *S. castellii* IA3 sequence has not been "optimised" by evolutionary selection to target the *S. cerevisiae* proteinase A, so further detrimental changes are more immediately noticeable.

A clear example of this critical importance of an appropriate centrepiece combination in *S. castellii* IA3 is provided by comparison of **65** (M18K/K22L) and **68** (K22L) with **66** (M18K/K22D) in Table 7.2. The K22L mutation (**68**) creates the most potent C-terminally His-tagged *S. castellii* IA3-derived inhibitor of *S. cerevisiae* vacuolar proteinase A, whilst the additional M18K mutation in **65** effectively destroys the activity at either pH value. However, protein **66**, which reconstitutes the *S. cerevisiae* IA3 centrepiece in *S. castellii* IA3, supports the proposition that Lys18 is

not solely responsible for this loss of potency, because it can be tolerated as long as it is accompanied by the simultaneous introduction of Asp at 22.

The fact that mutations to the centrepiece affect potency so greatly demonstrates the importance of this region for the inhibitory activity of *S. castellii* IA3. Moreover, the centrepiece combinations so detrimental to potency in *S. castellii* IA3 (K/L and K/M) are both approximately equipotent to the wild-type (K/D) in an *S. cerevisiae* IA3 context (compare **20** and **21** with **1** in Table 5.2). This suggests that the precision of fit contributed by other elements of the sequence, as well as the individual suitability of the residue in terms of the position relative to the active site cleft, may be important. This requirement for compatibility between residues on adjacent turns of the inhibitory helix is apparent in the context of the back pin region as well, in Section 6.4 (and discussed in Section 6.5). Taken together, the data accumulated in Chapters 6 and 7 support the view that *S. castellii* IA3 inhibits *S. cerevisiae* vacuolar proteinase A by an equivalent mode to that of *S. cerevisiae* IA3.

In *S. cerevisiae* IA3-based inhibitors, the pH switch had been noted only in proteins containing a basic residue at position 22. A similar effect is exhibited by *S. castellii* IA3: replacement of the wild-type Lys22 with an appropriate residue, such as Leu (**68** in Table 7.2), removes the phenomenon. In addition, the pH switch can be removed from *S. castellii* IA3 if the wild-type *S. cerevisiae* IA3 centrepiece of Lys18 and Asp22 (**66** in Table 7.2) is introduced. This suggests a shared element to the effect in the two species.

However, the His-tag is also crucially involved in the catastrophic loss in potency of recombinant *S. castellii* IA3 between pH 4.7 and pH 3.1, because removal of the His-tag causes the pH switch behaviour to disappear (compare the "pH switch" value of **75** with **43** in Table 7.3). Moreover, His-tag removal did ***not*** remove the pH switch from the N*cast*-C*cere* chimaera (compare the "pH switch" values of **51** and **74** in Table 7.3), which suggests that the identity of the tail also has a role in the pH switch in *S. castellii* IA3. Finally, although the excision of the His-tag and C-terminal tail (i.e. the synthetic peptide form, **44**) removes the pH switch from *S. castellii* IA3, an equivalent removal from a *S. cerevisiae* IA3 mutant displaying the pH switch in full-length form greatly reduced the activity at both pH values (compare **36** and **62** in

Table 7.1). Therefore, although possibly related by a shared role for Lys22, the phenomena in the two IA3 species also seem to have features that are distinct, with the cause in *S. castellii* IA3 appearing to be more complicated than that proposed for *S. cerevisiae* IA3 proteins containing Lys at 22 (Section 5.7).

Efforts to investigate these complications using NMR spectroscopy have been initiated by my colleague, Dr. D. Wyatt. Preliminary data have suggested that $NT(H)_6$ wild-type *S. castellii* IA3 in free solution undergoes a structural transition upon lowering the pH value of the solution from 6.35 to 3.0. The observed chemical shifts are >100-fold larger than might be expected as a consequence of the pH change alone. Together with the general peak broadening across the spectrum, this suggests that N-terminally His-tagged *S. castellii* IA3 displays definite elements of structure at the lower pH. Of particular note is the effect of the pH transition on the C-terminal residue of the protein. This residue has a characteristic position in protein NMR spectra (Dr. D. Wyatt, personal communication), and is clearly visible in the *S. castellii* IA3 spectrum at pH 6.35. However, the signals for this residue at pH 3.0 cannot be resolved from within the broader resonances, suggesting that the environment surrounding the residue has changed substantially. This is consistent with a change in conformation upon lowering the pH to 3.0. More extensive investigations are being undertaken by Dr. Wyatt, using an ^{15}N- and ^{13}C-labelled sample of $NT(H)_6$ *S. castellii* IA3 to attempt to elucidate the structures produced by this low-pH transition.

These findings represent a fundamental difference between the IA3 forms from *S. cerevisiae* and *S. castellii*, because no significant structure is evident in *S. cerevisiae* IA3 in free solution (Green *et al.*, 2004). This difference may offer some explanation for the distinct origins of the pH switch in *S. castellii* IA3 and *S. cerevisiae* IA3.

7.5.2 The detrimental effect of the *S. castellii* tail to its own inhibitory sequence

The tailless and His-tagless synthetic peptide form of *S. castellii* IA3 (**44** in Table 7.1) does not display either the pH switch or the sub-optimal performance at pH 4.7. However, because replacement of the *S. castellii* IA3 tail with that of the *S. cerevisiae* inhibitor (**47** in Table 6.2) had removed the detrimental effect at pH 4.7 also but *without* concomitant removal of the pH switch, the conclusion was drawn in Section 6.3 that the two phenomena were distinct.

In this chapter, the reciprocal situation is also shown to be the case. The tagless forms of recombinant *S. castellii* IA3 no longer show the pH switch, but they are also at least 10-fold less potent, at either pH, than the synthetic peptide possessing the same inhibitory sequence (compare **75** in Table 7.3 with **44** in Table 7.1). Therefore, the suggestion made in Section 6.2 (and discussed in Section 6.5) that the C-terminal His-tag might be responsible for the detrimental effect on inhibition can now be discounted. The alternative proposition, whereby the C-terminal His-tag is inert, is more consistent with the evidence and agrees with the findings for *S. cerevisiae* IA3 in Section 5.2. Thus, it would appear that the superior potency of the $NT(H)_6$ form of *S. castellii* IA3 reflects a *beneficial* effect of an N-terminal His-tag, which masks a remaining detrimental effect of the C-terminal *S. castellii* IA3 tail (compare **43** in Table 7.3 with **42** in Table 7.2). Moreover, the detrimental effect manifests itself at pH 3.1 as well because the potency of the tagless *S. castellii* IA3 (**75**) remains poor relative to peptide **44**, even though the pH switch has been removed by His-tag excision. A similar situation is also apparent (Section 7.3) when the pH switch is eliminated by replacement of Lys22 (compare **68** and **69** at pH 3.1 in Table 7.2)

The detrimental effect arising from the tail region is not removed by truncating the C-terminal tail of recombinant tagless *S. castellii* IA3 as far as Lys45. Only two differences between this protein (**76** in Table 7.3) and the synthetic peptide (**44** in Table 7.1) of "optimum" potency remain. The first, the N-terminal pyroglutamate residue that acts as a stop point for aminopeptidase removal of the $NT(H)_6$ tag, seems unlikely to mediate such a significant effect because this region of IA3 was shown in Sections 5.3 and 5.4 (albeit in an *S. cerevisiae* IA3 sequence context) to be tolerant of extensive alterations. The remaining distinction, therefore, is that **76** additionally

contains the tail residues 35-45 of *S. castellii* IA3. Therefore, because the detrimental effect is not witnessed with a *S. cerevisiae* IA3 tail, it seems reasonable to propose that an element of residues 35-45 of *S. castellii* IA3 is responsible for mediating an interaction that leads to this detrimental effect on the potency of the *S. castellii* IA3 sequence.

Any such interaction involving residues 35-45 would appear to be a total contrast to the situation with the *S. cerevisiae* IA3 tail. At the very least, this region of *S. cerevisiae* IA3 is ineffectual to the potency of the inhibitory sequence to which it is attached, be it of *S. cerevisiae* or *S. castellii* origin. Indeed, in some situations, the *S. cerevisiae* IA3 tail can even be a requirement, as in the case of the *S. cerevisiae* IA3 mutant K18D/D22K (compare **36** and **62** in Table 7.1). This positive role for the tail of *S. cerevisiae* IA3 may be consistent with the region making contacts with the proteinase, remote from the active site. Such effects might be largely undetectable in our kinetic assays, because of their relatively small contribution in comparison to the large binding energy provided by the *S. cerevisiae* inhibitory sequence. Furthermore, such presumably transient interactions may not result in a regularly structured conformer in the C-terminal tail of *S. cerevisiae* IA3, which would explain why no electron density appeared for this region in the X-ray crystal structure (Li *et al.*, 2000). In contrast, however, perhaps actively detrimental interactions mediated by an inappropriate C-terminal tail sequence are more observable kinetically, especially when the tail is attached to a sub-optimal inhibitory sequence such as *S. castellii* IA3.

However, the proposition must also be considered that the detrimental effect of residues 35-45 of *S. castellii* IA3 reflects an intramolecular event that does not occur, or is not possible, with the corresponding *S. cerevisiae* sequence. The 35-45 region in the two species is considerably different in sequence and character:

	40 45	
S. cerevisiae IA3:	SQDKDG KTTDA	**net charge -1**
S. castellii IA3:	SPESKE KMKTK	**net charge +2**

An intramolecular event would imply the formation of an inappropriate structure of some kind within the *S. castellii* IA3 residues. The structural prediction program PROF (Rost & Sander, 1993) was applied to the wild type sequences of *S. cerevisiae* IA3 and *S. castellii* IA3. This neural-network algorithm scores the sequence in a residue-by-residue manner for the relative probability (between: 1 = high; and 0 = low) of sequence elements adopting helix, sheet or loop structures. Notably, there are considerable differences in the predictions for the two sequences, with the *S. castellii* IA3 generally expected to show more helical content than *S. cerevisiae* IA3 across most of its sequence. In the context of the work described here, however, a significant contrast between the two sequences occurs at residues 35-45, the region identified by the truncation experiments to be responsible for the detrimental effect of the *S. castellii* IA3 tail (see Figure 7.1). The algorithm predicts only a loop for this region of *S. cerevisiae* IA3 (data not shown, but are reflected by the low helix scores on the graph in Figure 7.1). In contrast, however, the *S. castellii* IA3 prediction shows a "spike" of potential α-helical content over the same region.

As described in Section 1.8.4, *S. cerevisiae* IA3 has been shown by both CD and NMR studies to be essentially unstructured in free solution (Green *et al.*, 2004). Furthermore, the authors noted specifically the disordered nature of the sequence surrounding Gly40, which would appear to validate the suggestion of the PROF-prediction that this region of *S. cerevisiae* IA3 is unlikely to display structure. In contrast, however, preliminary NMR studies conducted by my colleague Dr. D. Wyatt (in parallel to the pH-effect examinations described in Section 7.5.1) suggest that, at pH 6.35, $NT(H)_6$ *S. castellii* IA3 is *not* totally unstructured in free solution. Some elements of the *S. castellii* IA3 sequence appear to explore slowly several conformations (unpublished data), with elements of structure appearing transiently in more than one region. It is conceivable that only a certain number of the resulting conformers might be able to bind to *S. cerevisiae* vacuolar proteinase A in an inhibitory manner, especially when the precise nature of the proposed inhibitory interaction is considered. However, the available data also do not exclude the possibility that any intramolecular structures may themselves mediate inappropriate intermolecular interactions with the proteinase. Such interactions could be at a site

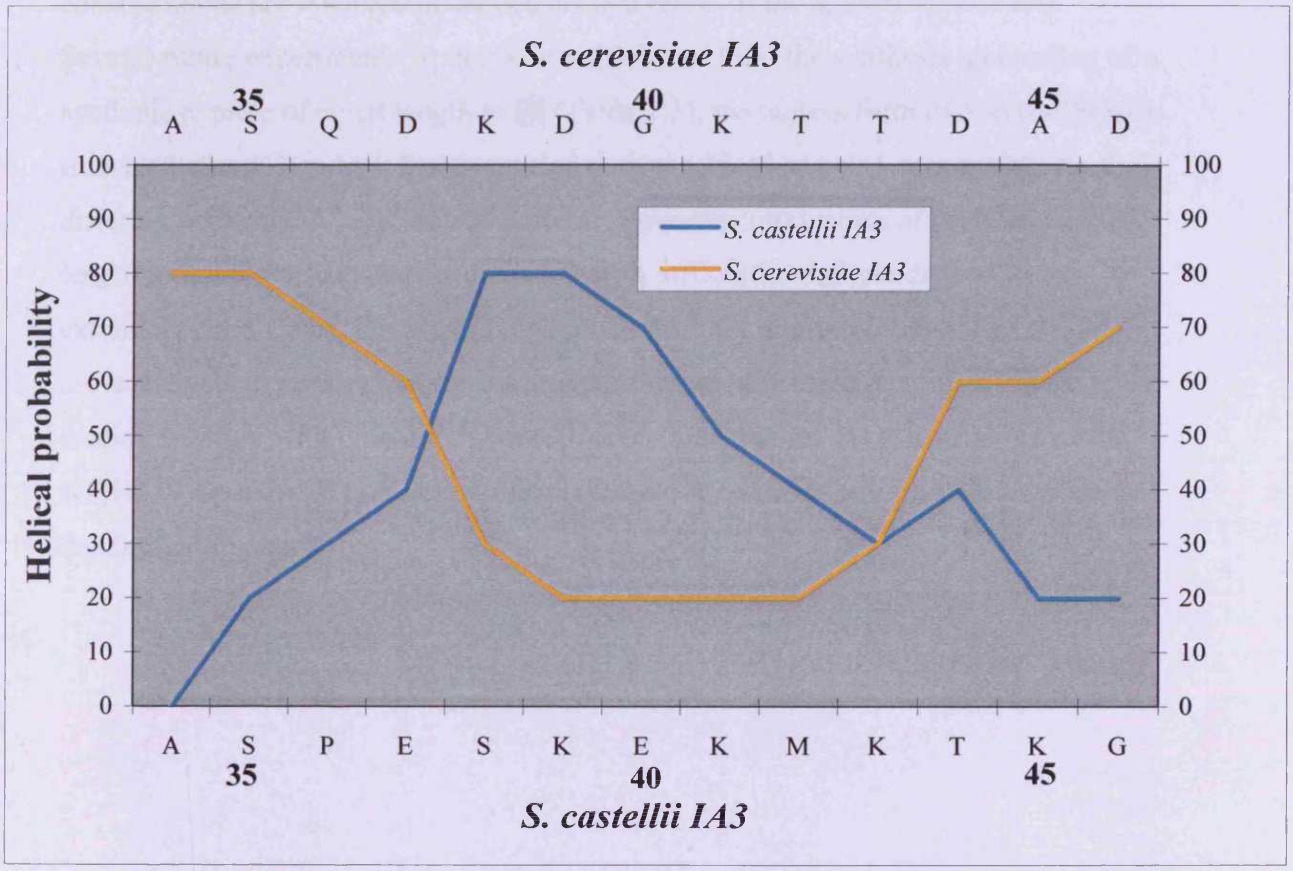


Figure 7.1 Prediction of the probability of helix formation in *S. castellii* IA3 versus *S. cerevisiae* IA3 using the PROF algorithm of Rost & Sander (1993). Although the entire sequence of both IA3 species was applied to the algorithm, only the results for residues 34-46 are presented. The amino acid sequence of *S. cerevisiae* IA3 is along the top axis, and corresponds to the gold line. The *S. castellii* IA3 sequence is along the bottom axis and corresponds to the cyan line. Both share the same vertical axis, which scores the probability for assigning helix to each residue in the sequence (100 being high, 0 being low)

remote from the active site cleft, leading to the detrimental effect observed with the *S. castellii* IA3 tail. These points will be considered further in the General Discussion (Chapter 11).

The nature and location of the structural elements in *S. castellii* IA3 in free solution have not yet been identified, so it is not possible to conclude whether these considerations are involved in the detrimental effect of the *S. castellii* IA3 tail. Several future experiments would be attractive, not least the synthesis and testing of a synthetic peptide of equal length to **76** (Table 7.3), the tagless form of *S. castellii* IA3 truncated after K45 which first identified residues 35-45 as being responsible for the differing behaviour of peptide and protein. However, production of peptides of this length by solid-phase synthesis is fraught with difficulties and yields tend to be extremely poor. Thus, the peptide may prove difficult to produce and/or to be preventatively expensive. In the meantime, it is hoped that the continued NMR studies, together with planned CD experiments and an X-ray crystal structure of the $NT(H)_6$ *S. castellii* IA3 - *S. cerevisiae* proteinase A complex will provide answers to this intriguing puzzle.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM) pH 4.7	K_i (nM) pH 3.1	pH switch
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	1.1 ± 0.4*	>11
42	WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~~LE(H)$_6$	4 ± 0.5	350 ± 70	88
44	WT *S. castellii* IA3 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	0.4 ± 0.1	8 ± 1	20
57	*S. castellii* G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEAF KNZA	3 ± 0.4	65 ± 8	22
58	*S. castellii* A25V/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEAF KNZA	1.1 ± 0.1	22 ± 1.1	20
59	*S. castellii* A25V 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEGZ KNZA	1.2 ± 0.3	26 ± 1	22
60	*S. castellii* A25I/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANI ASEAF KNZA	1.4 ± 0.2	26 ± 2	19
61	SEX 4	SDKN ANVSE ZFQQA KEZLE GKAKV VSDAF KKMA	0.8 ± 0.1	25 ± 4	31
36	*S. cerevisiae* K18D/D22K IA3	MNTDQ QKVSE IFQSS KEDLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.3 ± 0.1	120 ± 20	400
62	*S. cerevisiae* K18D/D22K 2-32	NTDQ QKVSE IFQSS KEDLQ GKAKV VSDAF KK	80 ± 7	>2000	N/A

Table 7.1 Comparing the pH switch effect in synthetic peptide *vs* recombinant protein forms of various IA3 molecules. Sequence derived from *S. cerevisiae* IA3 is coloured in yellow, *S. castellii* IA3 sequence is in blue. Where mutations directly swapped equivalent residues between sequence species the mutation is represented by the appropriate colour. Other mutations are coloured red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag. All recombinant proteins show only the residues of the inhibitory sequence (1-34), with the C-terminal tail region depicted as~~~~~~~~. The increased length of *S. castellii* IA3 (81 residues instead of 68) is reflected in this. Z = L-norleucine. The "pH switch" represents the ratio of K_i values between the pH values. *= from Phylip *et al.* (2001).

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)		pH
			pH 4.7	pH 3.1	switch
42	WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	4 ± 0.5	350 ± 70	88
63	*S. castellii* M18K/K22M IA3	MSDKN ANVSE MFQQA KEKLE GMANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	370 ± 50	NI	N/A
64	*S. castellii* M18K/K22Z 2-32	SDKN ANVSE ZFQQA KEKLE GZANA ASEGZ KN	350 ± 100	550 ± 50	1.6
65	*S. castellii* M18K/K22L IA3	MSDKN ANVSE MFQQA KEKLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	1000 ± 200	NI	N/A
66	*S. castellii* M18K/K22D IA3	MSDKN ANVSE MFQQA KEKLE GDANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	3 ± 0.8	50 ± 15	17
67	*S. castellii* M18K/K22D 2-32	SDKN ANVSE ZFQQA KEKLE GDANA ASEGZ KN	10 ± 1.5	35 ± 6	3.5
68	*S. castellii* K22L IA3	MSDKN ANVSE MFQQA KEMLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	1.1 ± 0.2	13 ± 3	12
69	*S. castellii* K22L 2-32	SDKN ANVSE ZFQQA KEZLE GLANA ASEGZ KN	2.5 ± 0.5	3 ± 0.8	1.2

Table 7.2 The effect on potency and the pH switch of mutations to the centrepiece of *S. castellii* IA3. Sequence derived from *S. castellii* IA3 is coloured in blue. Where mutations directly introduced the corresponding residue from *S. cerevisiae* IA3, the mutation is coloured yellow. Other mutations are in red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag. Only the residues up to 34 are detailed, with the C-terminal tail region depicted as~~~~~~~~. Z = L-norleucine. The "pH switch" represents the ratio of K_i values between the pH values.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM) pH 4.7	K_i (nM) pH 3.1	pH switch
43	NT(H)$_6$ WT *S. castellii* IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	0.3 ± 0.2	125 ± 30	417
50	NT(H)$_6$ *S. castellii* Q68 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~[68]	0.8 ± 0.1	910 ± 80	>1000
70	NT(H)$_6$ *S. castellii* Y57 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	2 ± 0.3	750 ± 70	375
71	NT(H)$_6$ *S. castellii* K45 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	0.9 ± 0.4	680 ± 70	755
51	NT(H)$_6$ N*cast*-C*cere* IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	0.7 ± 0.2	190 ± 80	270
72	NT(H)$_6$ N*cast*-C*cere* Y57 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	3 ± 0.3	460 ± 50	150
73	NT(H)$_6$ N*cast*-C*cere* A45 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	2 ± 0.6	570 ± 50	285
74	Tagless N*cast*-C*cere* IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	1.5 ± 0.2	240 ± 20	160
75	Tagless WT *S. castellii* IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	4 ± 0.3	150 ± 20	38
76	Tagless *S. castellii* K45 IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	10 ± 1.5	190 ± 30	19

Table 7.3 Investigating the effect of tail identity, tail length and the His-tag on the pH switch observed in recombinant *S. castellii* IA3. Sequence derived from *S. cerevisiae* IA3 is coloured in yellow, *S. castellii* IA3 sequence is in blue. Recombinant proteins were produced with an N-terminal MK(H)$_6$HMQ tag. All recombinant proteins show only the residues of the inhibitory sequence (1-34), with the C-terminal tail region depicted as~~~~~~~~, relative to its length on that molecule. Selected proteins had their His-tag removed, a process involving the cyclisation of the exposed N-terminal Gln to pyroglutamate (*Q). The "pH switch" represents the ratio of K_i values between the pH values.

Chapter 8

Inhibition of the vacuolar aspartic proteinase from *Pichia pastoris*

8.1 Introduction

The investigations into the binding of *S. cerevisiae* IA3 to the natural target, vacuolar aspartic proteinase A from the same organism, had provided an insight into the unprecedented mode of inhibition (Li *et al.*, 2000; Phylip *et al.*, 2001; Chapter 5 of this thesis). The specificity and high affinity (Dreyer *et al.*, 1985; Phylip *et al.*, 2001) of this naturally occurring inhibitor posed an attractive challenge: the re-design of IA3 to act against new, more pharmaceutically-relevant targets. As such a potent inhibitor of the natural target, much of the formative identification of residues critical to the action of IA3 had been derived from the effect of detrimental mutations (Phylip *et al.*, 2001; Chapter 5 of this thesis). These studies had examined only *S. cerevisiae* IA3 interacting with its own proteinase A, and therefore they had revealed little about the origin of the specificity of the molecule, only the potency.

The exquisite selectivity of the interaction made the identification of a realistically achievable new target a considerable first challenge in its own right. The methylotrophic yeast *Pichia pastoris* also has a vacuolar aspartic proteinase. Although not especially important as a target in itself, the *P. pastoris* proteinase was a feasible alternative enzyme, and it provided the opportunity to consider how mutations might alter specificity as well as potency. This chapter, therefore, describes the identification of mutant IA3 sequences with subnanomolar potencies against this new target enzyme.

8.2 The vacuolar proteinase from *Pichia pastoris*: a new target

The yeast *P. pastoris* is, perhaps, best known for the extensive use of this organism as a heterologous eukaryotic protein expression system (Lin Cereghino & Cregg, 2000; Ilgen *et al.*, 2004). Indeed, a *P. pastoris* expression system was employed to produce the vacuolar proteinase from *Aspergillus fumigatus* that will be described in Chapter 9.

The industrial importance of *P. pastoris* expression systems is reflected by the discovery that the sequence of the vacuolar proteinase from *P. pastoris* can only be found by a BLAST search if the patent record database is included, even when using a probe of relatively high sequence identity such as *S. cerevisiae* vacuolar proteinase A. The patent (Gleeson & Howard, 1996) identified the primary genes responsible for proteolytic processing in *P. pastoris* with the commercial benefit of enabling the production of strains deficient in such proteolytic activity, which in turn increased the yield of proteolytically sensitive recombinant products. It is interesting to note that the patent holders also originally found the gene encoding this proteinase by screening a bacteriophage λ-based *P. pastoris* genomic DNA library using the *S. cerevisiae* vacuolar proteinase A gene sequence as a probe.

The *P. pastoris* proteinase, like its counterpart from *S. cerevisiae*, is produced in the form of a proenzyme. The sequence of the mature *Pichia* proteinase is approximately 75% identical to that of *S. cerevisiae* vacuolar proteinase A (Figure 8.1). Based on this, a three-dimensional model was constructed by our colleague, Dr. D. Bur (Actelion Ltd., Allschwil, Switzerland) using the structures solved for the *S. cerevisiae* enzyme as the template.

The gene encoding the proenzyme form of the *P. pastoris* proteinase does not contain any introns, and therefore the coding sequence was amplified from genomic DNA by PCR, and cloned into an appropriate expression vector. The recombinant protein was produced in *S. cerevisiae* cells by my colleague, Dr. Lowri Phylip, according to the protocol described in Kondo *et al.* (1998). The mature *Pichia* proteinase was purified to homogeneity from the culture medium using the same purification protocol as that applied to *S. cerevisiae* vacuolar proteinase A (Kondo *et al.*, 1998). The N-terminal sequence determined through 12 cycles of Edman degradation was

Ala-Ser-His-Asp-Ala-Pro-Leu-Thr-Asn-Tyr-Leu-Asn~

which corresponds exactly to that predicted by the DNA sequence (Figure 8.1).

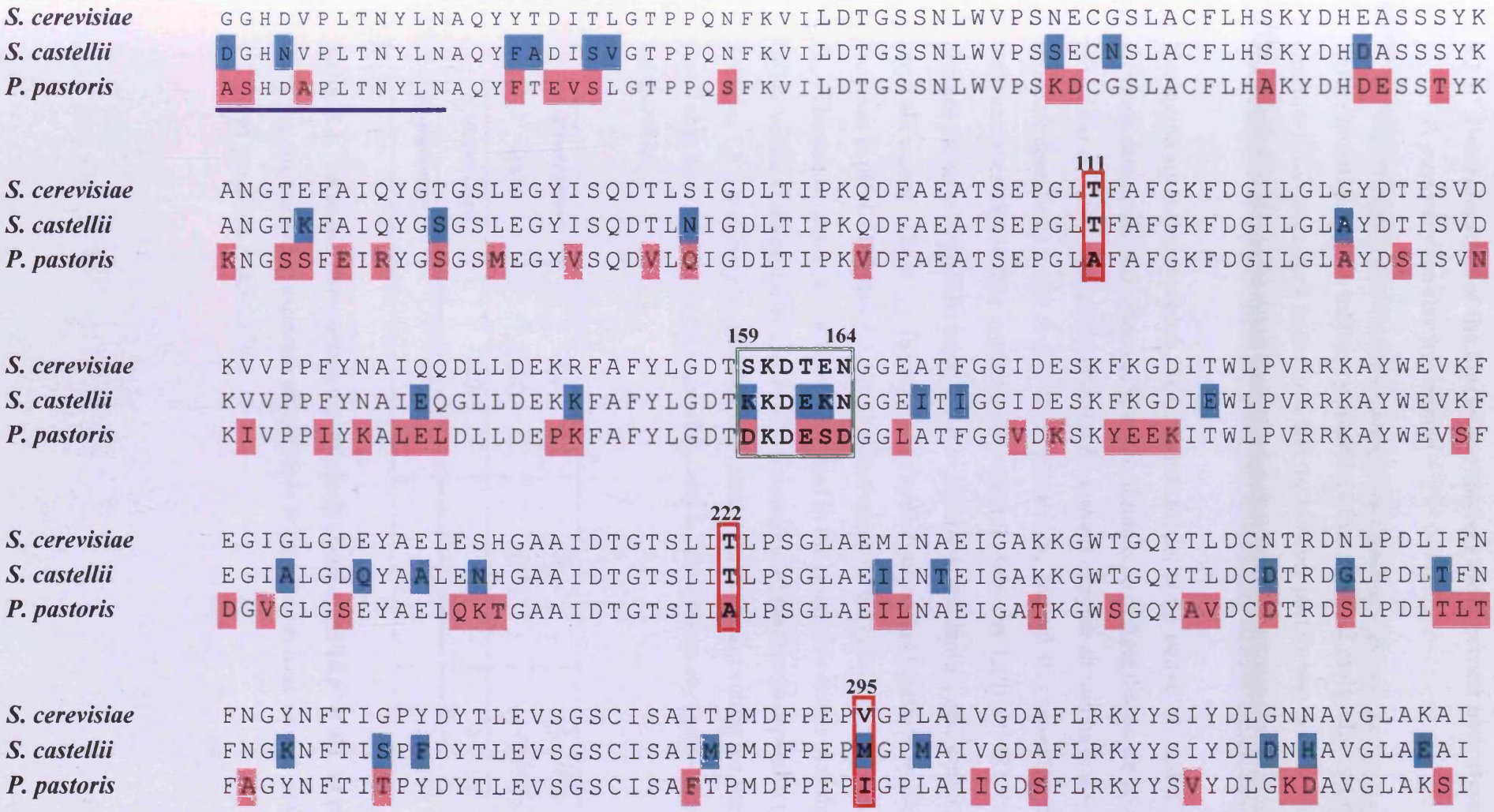


Figure 8.1 Amino acid sequence alignment of the mature vacuolar proteinases from *S. cerevisiae*, *S. castellii* and *P. pastoris*. Residue changes from *S. cerevisiae* proteinase A are highlighted in cyan (*S. castellii*) and pink (*P. pastoris*). Residues (111, 222 and 295, pig pepsin numbering) which contact the IA3 inhibitory helix directly and which differ between the *S. cerevisiae* and *P. pastoris* proteinase sequences are highlighted with red boxes. The differentially charged surface loop (Section 8.8) is marked with a green box, and the N-terminal sequence identified by Edman degradation (Section 8.2) is underlined in blue.

8.2.1 Determination of the kinetic parameters for substrate hydrolysis by *P. pastoris* vacuolar proteinase

Before investigating the inhibitory potential of various IA3 molecules against the enzyme, it was necessary to confirm that the purified protein was both active and had characteristics similar to other closely related aspartic proteinases.

The similarity in the sequence, the predicted model of the active site and the wealth of previous data on the substrate preferences of this type of fungal aspartic proteinase (Kondo *et al.*, 1998) suggested that the *P. pastoris* enzyme should have a very similar substrate specificity to the *S. cerevisiae* proteinase. Indeed, this proved to be the case. Both enzymes cleaved the synthetic chromogenic substrate KPIEF*F'RL (where * denotes the scissile peptide bond and F' = *p*-nitrophenylalanine) and were active at the same pH values. The *v* vs [S] curves for hydrolysis of this substrate by the two enzymes at pH 4.7 (Figure 8.2) enabled derivation of the kinetic parameters V_{max} and K_m. The concentration of each enzyme used in the respective assays was determined by active site titration using a known concentration of isovaleryl-pepstatin (IVP, Sections 1.7.1 and 2.2.31), enabling the calculation of the k_{cat} values for each enzyme (see Table 8.1, below). The values of K_m and k_{cat} for the two enzymes were comparable.

Proteinase from	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} M^{-1})
S. cerevisiae	140	44	3×10^5
P. pastoris	60	45	7×10^5

Table 8.1. Kinetic parameters for the hydrolysis of KPIEF*F'RL at pH 4.7, 37 °C by the vacuolar aspartic proteinases from *S. cerevisiae* and *P. pastoris*. The estimated precision is ±5-10 %.

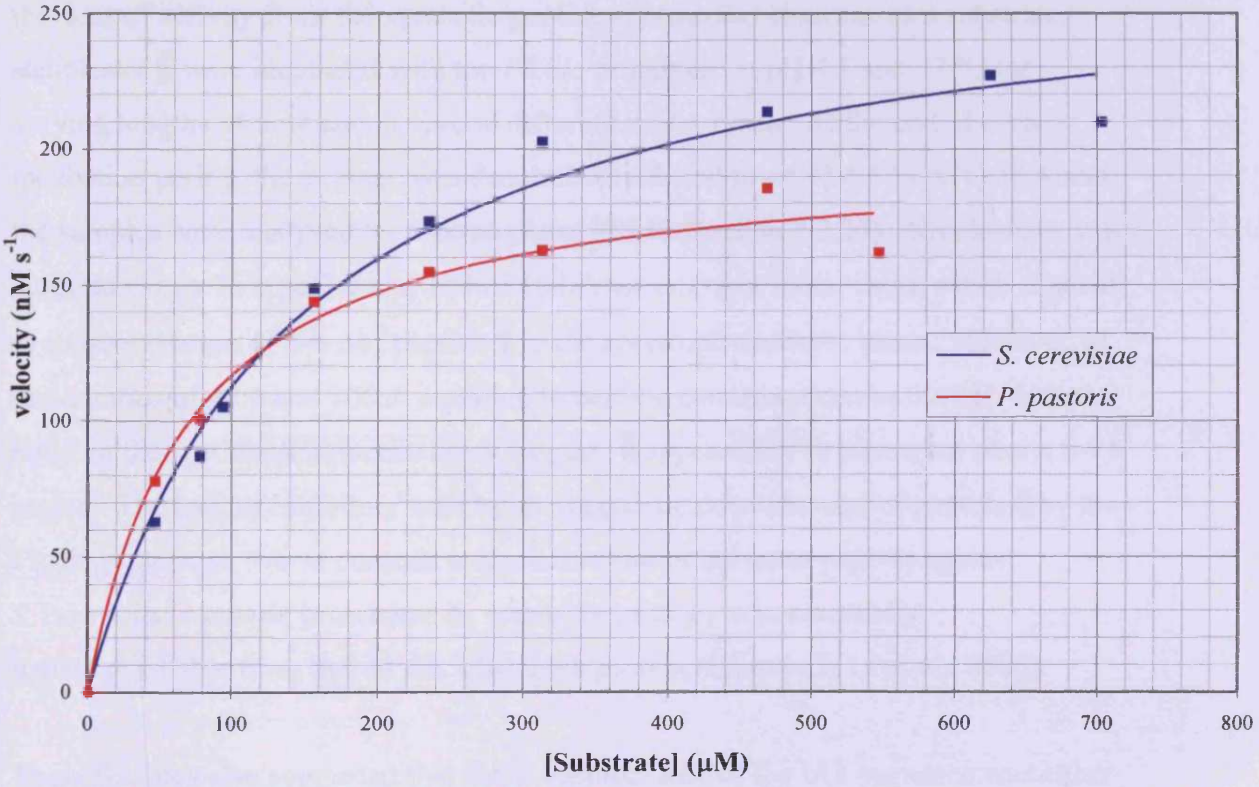


Figure 8.2 Cleavage of the chromogenic substrate (KPIEF*F′RL) by the vacuolar proteinases of *S. cerevisiae* (blue) and *P. pastoris* (red) at pH 4.7, 37 °C. Initial rates of substrate cleavage were monitored as a change in absorbance at 300 nm, as described in Section 2.2.29. The change in A_{300} upon substrate hydrolysis is 1600 $M^{-1}cm^{-1}$. The curves were fitted with Graphpad Prism® software to apply a non-linear regression analysis to the data, using the Michaelis-Menten equation $v = V_{max}[S]/(K_m+[S])$.

8.2.2 Paradigms lost: the interactions of *S. cerevisiae* IA3 with *P. pastoris* proteinase

Recombinant wild-type *S. cerevisiae* IA3 (C-terminally His-tagged, as described in Chapter 3) interacted weakly with the *P. pastoris* enzyme at pH 4.7 (see **1** in Table 8.2), but the synthetic peptide consisting of residues 2-34 (**6** in Table 8.2) showed no inhibition of the *Pichia* proteinase whatsoever. To investigate whether this loss of activity from the synthetic peptide represented cleavage as a substrate, samples of **6** were incubated with the *Pichia* proteinase at pH 4.7 and 37 °C for varying lengths of time and at several different molar ratios. At the end of each incubation period, the enzyme was denatured by the addition of 4.5 % (v/v) TFA and the samples were analysed by reverse-phase FPLC (Section 2.2.27). No cleavage was detected over a 16 h period at a 1000:1 (inhibitor:enzyme) molar ratio, which equated to a concentration of 5-6 μM peptide **6** in the spectrophotometric assay. However, at molar ratios of 40:1 and 100:1, equating to peptide concentrations in the 250-600 nM range in the spectrophotometric assay, peptide **6** appeared to be destroyed over a 6-9 h period. The lack of inhibitory activity and apparent slow cleavage of peptide **6** by the *Pichia* proteinase was in contrast to the behaviour of the same peptide against *S. cerevisiae* vacuolar proteinase A, where the potency was essentially indistinguishable from that of the full-length protein (Chapter 5; Li *et al.*, 2000).

These findings also suggested that the C-terminal half of the IA3 sequence was either influencing the ability of the N-terminal region to interact with the *Pichia* proteinase, or even that the activity against this new enzyme resided elsewhere in the polypeptide. Consequently, synthetic peptides representing residues 35-68 (both with and without the $LE(H)_6$ tag of the full-length recombinant protein form of *S. cerevisiae* IA3) were tested against the *Pichia* proteinase. Neither peptide produced any inhibition at a concentration of 2 μM (**77** and **78** in Table 8.2), confirming that the C-terminal half of the molecule was not exclusively responsible for the weak activity against the *Pichia* enzyme observed in the full-length protein form (**1** in Table 8.2). Therefore, residues from both N- and C-terminal regions were required.

The separate synthetic peptides that together spanned the entire sequence of IA3, i.e. 2-34 and 35-68, were tested against the *P. pastoris* proteinase in combination at a 4:1 molar ratio of C-terminal peptide (35-68) : N-terminal peptide (2-34). The N-terminal peptide was present at a concentration 4-times that of the K_i value derived for the full-length protein (**1**). Both His-tagged and tagless forms were investigated. No inhibition was observed in either of these situations (i.e. **6** + **77** and **6** + **78**, see Table 8.2). This suggested that the mere presence of both parts of the IA3 sequence was not sufficient: they had to be connected for activity against the *Pichia* proteinase to be observed.

The importance of the spatial relationship of the two domains was investigated with the domain-swap protein 3AI, where residues 35-68 of IA3 were transposed to be positioned in front of residues 1-34 (Section 3.5). Perhaps unsurprisingly, this protein was still a very potent inhibitor of *S. cerevisiae* vacuolar proteinase A, with a K_i value at pH 4.7 at the limit of accurate determination (0.1 nM). Against *P. pastoris* proteinase, however, this transposition compromised the activity (compare **79** with **1** in Table 8.2), although a weak interaction was still quantifiable (compared to the total absence of inhibition by the two unconnected peptides representing the same overall sequence).

To gain more insight into the role of the tail (residues 35-68), the two recombinant forms of wild-type *S. cerevisiae* IA3 with partial C-terminal truncations were tested against the *Pichia* enzyme (**4** and **5** in Table 8.2). The two truncation points, after Tyr57 and Ala45 respectively, divided the tail region into three roughly equal segments. The interaction of these mutants with the *Pichia* proteinase was comparable to that observed for 3AI (**79**). Furthermore, the similarity in potency between the two truncation mutants suggested that, although removal of residues 57-68 was detrimental to potency, the subsequent removal of a further 12 residues to Ala45 was not.

Nevertheless, the fact that even a partial truncation of the C-terminus was disadvantageous strongly suggested that the inhibitory activity against *P. pastoris* proteinase resulted from a synergistic action from both the N- and C-terminal halves of the molecule. This was an apparent contrast of the findings against *S. cerevisiae*

vacuolar proteinase A. As a result, there was no conclusive evidence from these findings that the inhibitor acted against the *Pichia* enzyme in a manner similar to that observed against the natural target.

An X-ray crystallographic study of a complex of IA3 with the *P. pastoris* proteinase offered an attractive method of studying the origin of the inhibitory action. However, the relatively poor K_i value (55 ± 11 nM) suggested that the enzyme-inhibitor complex might not be sufficiently stable for crystallographic study. Therefore, the nature of the inhibition and the origin of the potency were investigated in a less direct manner by comparing the effect of mutations in selected regions of the sequence.

8.3 The effect of mutations in the centrepiece residues of *S. cerevisiae* IA3

The centrepiece residues (18 and 22) form one of the regions critical to the activity of IA3 against the natural target. Therefore, the effect of mutations to this centrepiece region on the interaction with the *P. pastoris* proteinase was investigated. Replacement of Lys18 with the more strongly basic amino acid Arg (**31** in Table 8.3) produced a 10-fold improvement in potency. The low nanomolar K_i value represented potent inhibition of the new target enzyme by this protein containing a single, seemingly conservative, residue change. A more modest improvement also resulted from replacement of Asp22 with a longer Glu residue (**32** in Table 8.3). However, simultaneous introduction of both mutations in the double mutant K18R/D22E (**33** in Table 8.3) did not improve upon the potency of the K18R single mutant (**31**).

Other centrepiece mutations were not so beneficial. For instance, transposition of the centrepiece residues in the mutant K18D/D22K (**36** in Table 8.3), and three other mutants in which the acidic/basic nature of the residues at positions 18 and 22 was reversed (**37**, **40**, **41** in Table 8.3), all showed varying degrees of reduced activity, relative to the wild-type (**1**).

However, the mutant K18V/D22K (**38** in Table 8.3) displayed a level of potency against the *Pichia* proteinase that was comparable or even modestly superior to that of the wild-type IA3 (**1**). The K_i value was approximately an order of magnitude better than the other mutants containing a basic residue at position 22 described above. This revealed that such a mutation was not detrimental, indeed could be beneficial, so long as an appropriate residue was introduced simultaneously at position 18. The mutant K18M/D22K reinforced this finding (**39** in Table 8.3), with a low nanomolar K_i value representing an improvement of approximately 20-fold upon the wild-type protein and with marginally superior potency to either of the K18R-containing mutants (**31** and **33** in Table 8.3).

Therefore, these effects, both advantageous and detrimental, confirmed the importance of the centrepiece region for the inhibitory potency of *S. cerevisiae* IA3 against the *Pichia* enzyme. This represented the first suggestion of a similar mode of inhibitory action against both enzymes.

8.4 Inhibitory potency of *S. castellii* IA3 against the *Pichia* proteinase

The most potent *S. cerevisiae* IA3 mutant studied in Section 8.3 (**39** in Table 8.3) contained the centrepiece residues Met18 and Lys22, which are those found naturally in *S. castellii* IA3. Therefore, C-terminally His-tagged recombinant *S. castellii* IA3 was tested against the *Pichia* enzyme in order to assess how well this centrepiece performed within its natural sequence context.

Notably, recombinant *S. castellii* IA3 was less potent than the *S. cerevisiae* IA3 mutant containing the same centrepiece (compare **42** in Table 8.4 with **39** in Table 8.3), although its potency was modestly superior to wild-type *S. cerevisiae* IA3 (**1** in Table 8.3). Any suggestion that a *S. cerevisiae* IA3 sequence context outside of the centrepiece might be superior to that of *S. castellii* IA3 was dispelled, however, by the discovery that the synthetic peptide representing residues 2-34 of *S. castellii* IA3 (**44**) was of equal potency to **39**. As such, the potency of *S. castellii* IA3 had been

improved 4-fold by the removal of residues 35-81 and the His-tag. This clearly established that the inhibitory activity of *S. castellii* IA3 towards the *Pichia* enzyme is located in the N-terminal 34 residues of the molecule, and reprised the findings against *S. cerevisiae* vacuolar proteinase A (Chapter 6). However, the result also represented a remarkable departure from the apparent requirement of *S. cerevisiae* IA3 for an intact C-terminal half in order to achieve inhibition of the *P. pastoris* proteinase. Indeed the implication of this result was that the *S. castellii* IA3 C-terminal tail was *limiting* the potency of the inhibitory sequence (residues 2-34).

Therefore, although producing opposite effects, neither tail species was "inert" when connected to its own respective inhibitory sequence. To investigate the effect of each tail on the opposite inhibitory sequence, the N*cere*-C*cast* and N*cast*-C*cere* chimaeric proteins (**45** and **47**, respectively, in Table 8.4) were tested against *P. pastoris* vacuolar proteinase. Attachment of the *S. castellii* tail (residues 35-81) to the *S. cerevisiae* inhibitory sequence (residues 1-34) in **45** had a small (2-fold) detrimental effect (compare with **1** in Table 8.3), although the observed potency was still significantly better than the 2-34 inhibitory sequence on its own (peptide **6** in Table 8.2). Therefore, for the *S. cerevisiae* inhibitory sequence, even the attachment of an *S. castellii* IA3 tail was preferable to no tail at all.

The reciprocal chimaera, **47**, where the *S. cerevisiae* IA3 tail (residues 35-68) was substituted for residues 35-81 of *S. castellii* IA3, resulted in a 5-fold improvement in the K_i value against the *Pichia* enzyme (compare **47** with **42** in Table 8.4). Thus, the potency of **47** was equivalent to the tailless synthetic peptide **44** (Table 8.4), suggesting that the *S. cerevisiae* IA3 tail allowed the *S. castellii* inhibitory sequence to perform to its tailless "potential".

The relatively detrimental nature of the *S. castellii* IA3 tail was further indicated by a comparison of **48** and **49** (Table 8.4). In the presence of the natural *S. castellii* IA3 tail, this single M18V (**48**) mutation in the inhibitory sequence effectively destroyed the inhibitory activity of the molecule. However, when the tail was replaced with that from *S. cerevisiae* IA3 in **49**, the K_i value was restored to within 3-fold of wild-type *S. castellii* IA3. The potency of **49** was very close to that of **38** (Table 8.3), the

S. cerevisiae IA3 mutant with the same centrepiece. Seemingly, this supported the view that, in the absence of the *S. castellii* IA3 tail, either context for the inhibitory species might be approximately equally favourable for a given centrepiece combination.

To extend these studies, the centrepiece residues (Lys18 and Asp22) of wild-type *S. cerevisiae* IA3 were introduced into the *S. castellii* IA3 sequence. In full-length recombinant protein form (**66** in Table 8.4), the effect of this substitution was minimal. Strikingly, the 2-32 peptide form of the same sequence (**67** in Table 8.4) was essentially devoid of inhibitory activity. That is, by exchanging the centrepiece residues of *S. castellii* IA3 for those of *S. cerevisiae* IA3, or *vice versa*, seemingly one sequence would exhibit the properties of the other.

Transposition of the centrepiece residues of *S. cerevisiae* IA3 had already been shown to compromise inhibitory potency (**36** in Table 8.3). The transposition of the centrepiece residues of *S. castellii* IA3 (M18K/K22M) also effectively destroyed all inhibitory potency in both recombinant protein and tailless synthetic peptide forms (**63** and **64**, respectively, in Table 8.4). It was noted, however, that the centrepiece thus produced - Lys18 and Met22 – was modestly *beneficial* when introduced into an *S. cerevisiae* IA3 sequence context (**20** in Table 8.4). Generally, though, the data suggested that a Lys at 18 was not an optimal residue choice, and could severely impair the potency of an *S. castellii* IA3 molecule in particular. This conclusion was exemplified by comparison of two mutants – M18K/K22L and K22L (**65** and **68**, respectively in Table 8.4). Whilst the double mutant **65** lost all activity, **68** was the most potent inhibitor yet discovered from either species against *P. pastoris* proteinase.

Notably, a 2-32 synthetic peptide form of the K22L mutant (**69** in Table 8.4) had essentially identical potency to the recombinant protein form. The implication was that the replacement of Lys22 with Leu removed the detrimental effect of the *S. castellii* IA3 tail or, at the very least, the advantageous effect of tail removal was no longer apparent. Such a finding was reminiscent of the work described in Chapter 7 in relation to the pH switch phenomenon at pH 3.1 against *S. cerevisiae* vacuolar proteinase A.

Taken in totality, these data emphasise the importance of the centrepiece residues for the inhibitory potency of both species of IA3, and suggest that *P. pastoris* proteinase is inhibited by these polypeptides in a similar manner to that observed for *S. cerevisiae* vacuolar proteinase A. Indeed, the only significant difference was the apparent requirement for an intact tail, of either species, on *S. cerevisiae* IA3. This conclusion was further supported by the discovery that a 2-32 synthetic peptide (**80** in Table 8.4) corresponding in sequence to the most potent *S. cerevisiae* IA3 protein (**39** in Table 8.3) described above did not demonstrate any potency against the *Pichia* enzyme. This was despite the presence of the *S. castellii* IA3 centrepiece residues. Therefore, regardless of the importance of the binding contributions made by the centrepiece, and the frequent similarity in performance of equivalent centrepiece combinations between the two species, the superior potency of the peptide form of *S. castellii* IA3 against the *Pichia* enzyme must also be dependent on other features.

8.5 The effect of interchanging residues outside the centrepiece between IA3 species

Other than the centrepiece residues 18 and 22, the major contributors to binding identified from the crystal structure of the *S. cerevisiae* IA3 - *S. cerevisiae* vacuolar proteinase A complex (Li *et al.*, 2000) are the hydrophobic interactions made by the front and back pins, and Leu19 (Section 1.8.5). As described in Section 6.4, the front pins of both species are strikingly similar, and Leu19 is conserved. However, the back pin region is substantially different. All four positions (boxed below) predicted by the crystal structure to oriented towards the floor of the active site cleft were less bulky in *S. castellii* IA3:

residue number	25	26	27	28	29	30	31
S. cerevisiae IA3:	V	V	S	D	A	F	K
S. castellii IA3:	A	A	S	E	G	M	K

The simultaneous introduction of the *S. castellii* IA3 back pin residues at positions 26, 29 and 30 into *S. cerevisiae* IA3 produced a moderate improvement in potency over

the wild-type molecule against the *Pichia* proteinase (compare **52** and **1** in Table 8.5). Indeed, the observed K_i value was similar to that of recombinant wild-type *S. castellii* IA3 (**42** in Table 8.5).

The mutant **53** (Table 8.5), which had incorporated fortuitously a fourth mutation (Q20E) towards the *S. castellii* IA3 sequence during the production of **52** (see Section 3.2.3), was more potent. The low nanomolar K_i value represented an 8-fold improvement on wild-type *S. cerevisiae* IA3, similar to the *S. cerevisiae* IA3 mutant containing the *S. castellii* IA3 centrepiece (**39** in Table 8.3). These results, therefore, suggested a significant contribution from the back pin region to the superior potency demonstrated by *S. castellii* IA3 against the *Pichia* proteinase. This was in addition to the role of the centrepiece identified in Section 8.4.

The chimaeric recombinant protein **56** (Table 8.5) examined the effect of a simultaneous introduction of both the *S. castellii* IA3 centrepiece and back pin regions into *S. cerevisiae* IA3. This chimaera was more potent than the protein (**52**) in which only the back pin residues were exchanged. However, it did not improve further the K_i value observed upon the introduction of the *S. castellii* IA3 centrepiece residues alone (**39**). This suggests that the dominant factor may be the presence of the "correct" centrepiece residues.

This was examined further by the introduction of the *S. cerevisiae* IA3 back pin residues into the *S. castellii* IA3 sequence in the synthetic peptides **57** - **60** (Table 8.5). Incorporation of *S. cerevisiae* IA3 residues at either position 25 (**59**) or positions 29 and 30 (**57**) had a marginal effect at most (compare with **44** in Table 8.5), and the simultaneous introduction of all three back pin mutations in **58** resulted in a K_i value that was identical to that of the wild-type peptide (**44** in Table 8.5). Substitution of a bulkier Ile residue at position 25 in **60** did not affect activity either, showing that the *S. castellii* inhibitory sequence could accommodate numerous changes to this region with almost no effect on potency.

On this basis, the ability of *S. castellii* IA3 to tolerate more extensive changes to this region of its sequence was examined in the chimaeric synthetic peptide **61** (Table 8.5). All residues from 24-34 of *S. castellii* IA3 were replaced with their

equivalent residues from *S. cerevisiae* IA3. Only a modest reduction (4-fold) in potency was observed against the *Pichia* enzyme. As such, **61** was the first peptide length inhibitor harbouring a significant portion (almost one third) of contiguous *S. cerevisiae* IA3 sequence to show inhibition of the *P. pastoris* proteinase.

8.6 The effect of hydrophobic mutations in the centrepiece of *S. cerevisiae* IA3

An inspection of the data revealed that the most potent inhibitor of the *Pichia* enzyme yet described, from either yeast species, was the recombinant *S. castellii* mutant IA3, K22L (**68** in Table 8.4). Therefore, because the introduction of the wild-type *S. castellii* centrepiece into *S. cerevisiae* IA3 had yielded a potency improvement of 20-fold (see **39** Section 8.3), the centrepiece from **68** was also tested in an *S. cerevisiae* IA3 context. This *S. cerevisiae* IA3 mutant, K18M/D22L (**24** in Table 8.6), was shown previously to be particularly potent against *S. cerevisiae* vacuolar proteinase A, with a K_i value below the detection limit (0.1 nM) at pH 3.1 as well as pH 4.7 (see Section 5.5). Against the *Pichia* enzyme, the potency of this form of IA3 was even more remarkable. The subnanomolar K_i value measured represented a 100-fold improvement upon that of the wild-type sequence (**1** in Table 8.6). Again, therefore, a potent centrepiece from *S. castellii* IA3 was even better in an *S. cerevisiae* IA3 context, just as was described above when the M18/K22 centrepiece arrangement was "transplanted" into *S. cerevisiae* IA3 in place of the wild-type K18/D22 combination (**39** and **1**, Table 8.3)

Several other *S. cerevisiae* IA3 mutants containing hydrophobic mutations to the centrepiece residues were tested against the *Pichia* enzyme also. Introduction of the mutations K18I/L19M (**23** in Table 8.6) had a much less profound effect than **24**, with the K_i value only improving 3-4 fold over the wild-type. However, an extra hydrophobic mutation at position 22 in the triple mutant K18I/L19M/D22I (**26** in Table 8.6) produced a 25-fold improvement upon the double mutant **23**, resulting in a second inhibitor of the *Pichia* enzyme that had a subnanomolar K_i value.

Against *S. cerevisiae* vacuolar proteinase A, a plethora of results (see Section 5.4) had identified Leu19 as one of the critical hydrophobic pins. The importance of this residue for the inhibition of *P. pastoris* proteinase was also confirmed by the K18I/D22I mutant *S. cerevisiae* IA3 (**25** in Table 8.6). This mutant was of superior potency to **26** (i.e. the same residues at positions 18 and 22 but lacking the L19M mutation), with a K_i value against the *Pichia* enzyme below the limit of accurate measurement.

The potencies of **24**, **25** and **26** against the *Pichia* proteinase were substantially better than any of the *S. castellii* IA3-derived inhibitory sequences that were active in both full-length protein and tailless, synthetic peptide forms. Therefore, C-terminal truncations of the triple mutant (**26**) were produced to consider whether these mutations had introduced sufficient improvements in binding to enable an *S. cerevisiae* IA3-derived inhibitory sequence to inhibit the *Pichia* enzyme in the absence of a C-terminal tail. The truncation points selected were identical to those described in Section 8.2.2, which had compromised the activity of wild-type *S. cerevisiae* IA3 against the *Pichia* enzyme.

Both truncated forms (**27** and **28** in Table 8.6) of **26** demonstrated a modestly improved potency versus the *Pichia* proteinase, suggesting that the tail region was not so critical to the activity of these mutants. This conclusion was supported by the results with a 2-32 synthetic peptide (**29** in Table 8.6) that represented the same inhibitory sequence. Although the potency was reduced 10-fold with respect to the full-length form (**26**), it nevertheless retained a K_i value in the very low nM range. A 2-32 synthetic peptide version of the K18M/D22L mutant (**30** in Table 8.6) also produced a low nM K_i value against the *Pichia* enzyme. As such, **29** and **30** were the first synthetic, tailless, *S. cerevisiae* IA3-based inhibitory sequences, of either 2-32 or 2-34 length, that were capable of inhibition of *P. pastoris* proteinase. Thus, the paradigms, lost in Section 8.2.2, were (at least partially) regained.

8.7 The effect of IA3 C-terminal tail length and His-tag location on potency

A notable contrast in the inhibition of the *Pichia* proteinase had been observed between the behaviour of *S. cerevisiae* IA3 sequences and those based on *S. castellii* IA3. Even the *S. cerevisiae* IA3 inhibitory sequences capable of potent inhibition of the *Pichia* enzyme as tailless peptides (**29** and **30** in Table 8.6) were *more* active in full-length form. Conversely, removal of the tail from *S. castellii* IA3-based inhibitors was almost universally beneficial. The possibility that the production of the *S. cerevisiae* IA3-based peptides **29** and **30** in 2-32, rather than 2-34, form was responsible for this reduction (see Section 5.3) was considered, especially because the losses were similar to those witnessed upon this truncation to wild-type *S. cerevisiae* IA3 against the *S. cerevisiae* vacuolar proteinase A (compare **7** and **6** in Table 5.1). However, both the recombinant protein and 2-32 synthetic peptide forms of the *S. castellii* IA3 mutant K22L (**68** and **69**, respectively, in Table 8.4), which contained the same centrepiece as **30**, displayed almost identical K_i values against the *Pichia* proteinase. The reduction in length to 2-32 rather than 2-34 was therefore considered unlikely to be the *sole* cause of the diminished potency between **24** and **30**, or **26** and **29**.

The role and effect of the tail region in the two IA3 species still appeared quite distinct, therefore. This conclusion was examined further by the testing of a chimaeric form of the K18I/L19M/D22I triple mutant (**46** in Table 8.6), where the *S. cerevisiae* IA3 tail (residues 35-68) was replaced with that of *S. castellii* IA3 (35-81). Despite the attachment of this *S. castellii* IA3 tail, so detrimental to the potency of *S. castellii* IA3-derived inhibitory sequences, the K_i value recorded for this chimaera was only marginally worse than the form with an *S. cerevisiae* IA3 tail (**26** in Table 8.6), and was modestly superior to the tailless peptide (compare **46** and **29** in Table 8.6).

As described in Section 8.4, the disadvantageous nature of a tail to the *S. castellii* inhibitory sequence could, apparently, be alleviated either by replacement of Lys22 with Leu (**68** in Table 8.4) or the attachment of an *S. cerevisiae* IA3 tail in its place

(for example, the N*cast*-C*cere* chimaera **47** in Tables 8.4, and 8.7). However, because the *S. cerevisiae* IA3 tail is 13 residues shorter than that of *S. castellii* IA3 (Figure 6.2), **47** had introduced alterations to both sequence identity *and* tail length. A form of *S. castellii* IA3 truncated to the same length as the *S. cerevisiae* molecule (and, therefore, **47**) had already been produced for the studies with *S. cerevisiae* vacuolar proteinase A (**50** in Tables 8.7 and 6.2). This provided a convenient tool to investigate whether the difference in tail length was responsible for the superior potency of the N*cast*-C*cere* chimaera (**47**) over wild-type *S. castellii* IA3 (**42** in Table 8.7). However, for reasons already described in Section 7.4, it had been necessary to produce this protein in N-terminally His-tagged ($NT(H)_6$) form rather than with the usual C-terminal His-tag. Therefore, the $NT(H)_6$ form of the N*cast*-C*cere* chimaera (**51** in Table 8.7) was also tested against the *Pichia* proteinase, in order to afford a direct comparison.

The potencies of **50** and **51** were very similar, with both proteins displaying subnanomolar K_i values against the *Pichia* proteinase. The Q68-truncated *S. castellii* IA3 (**50**) was >20-fold more potent than full-length, wild-type, C-terminally His-tagged *S. castellii* IA3 (**42** in Table 8.7). However, the fact that the N*cast*-C*cere* chimaera was also of approximately 10-fold superior potency in $NT(H)_6$ form (compare **47** and **51** in Table 8.7) suggested that this improvement probably was related to the His-tag relocation rather than the C-terminal truncation. Certainly, more severe truncations in either *S. castellii* IA3 (**70** and **71** in Table 8.7) or N*cast*-C*cere* (**72** and **73**) forms did not improve potencies any further.

Relocation of the His-tag to the N-terminus of full length, wild-type *S. castellii* IA3 increased the potency of the molecule 150-fold with respect to the C-terminally His-tagged form (compare **43** and **42** in Table 8.7). Moreover, the K_i value, which was at the limit of measurement, was approximately 10-fold better than any of the truncation mutants. This suggested that, if anything, the truncations were actually mildly disadvantageous.

To investigate further the effect of His-tag relocation on inhibitory potency, the $NT(H)_6$ form of *S. cerevisiae* IA3 was tested against the *Pichia* proteinase also. A similar improvement was observed, with the $NT(H)_6$ form of *S. cerevisiae* IA3 (**2** in

Table 8.7) of 100-fold higher potency than the C-terminally His-tagged counterpart (**1** in Table 8.7).

These results indicated that relocation of the His-tag to the N-terminus of wild-type IA3 molecules of either species greatly improved the potency of the molecules against the *P. pastoris* proteinase. The fact that it also improved the potency of the N*cast*-C*cere* chimaera (**51** versus **47** in Table 8.7) was a contrast to the effect of relocation on the same inhibitor against the *S. cerevisiae* proteinase A, where the $NT(H)_6$ tag had been modestly detrimental (Section 6.3). This improved potency against the *Pichia* proteinase, however, could represent either the removal of a detrimental effect of the C-terminal His-tag or an active benefit of an N-terminal His-tag. To examine these two possibilities, the N-terminal His-tags were removed from wild-type *S. cerevisiae* IA3 (**3**), the N*cast*-C*cere* chimaera (**74**) and wild-type *S. castellii* IA3 (**75**). All three tagless proteins were substantially less potent than their $NT(H)_6$ tagged equivalents (**2**, **51** and **43**, respectively), with K_i values that were moderately inferior to the equivalent C-terminally His-tagged form. Furthermore, the tagless form of *S. castellii* IA3 truncated after K45 (**76** in Table 8.7) recorded an identical K_i value to the full-length *S. castellii* IA3, and remained substantially less potent than the 2-34 synthetic peptide form (**44** in Table 8.5). Thus, it would appear that the $NT(H)_6$ tag was of active benefit to the inhibitory sequences.

To test the beneficial effect further, an N-terminally His-tagged form of the *S. cerevisiae* IA3 mutant K18D/D22K was produced. As described in Section 8.3, in C-terminally His-tagged form (**36** in Table 8.3) this protein had demonstrated very little activity against the *Pichia* enzyme. However, with an $NT(H)_6$ tag (**81** in Table 8.7) the potency improved >10-fold, resulting in an inhibitor of superior potency to C-terminally His-tagged wild-type *S. cerevisiae* IA3 (**1** in Table 8.7). Once again, removal of the N-terminal His-tag returned the K_i of this mutant (**82** in Table 8.7) to a value worse than the $CT(H)_6$ form. Thus, the universal improvement in potency brought about by the attachment of an $NT(H)_6$ tag even extended to the conversion of a protein of low activity (**82**) (and that was probably slowly cleaved, see Sections 5.6, and 8.2.2) into an effective inhibitor (**81**) of the *Pichia* proteinase.

8.8 Discussion

After *S. cerevisiae* proteinase A, the vacuolar aspartic proteinase from *P. pastoris* is only the second enzyme susceptible to inhibition by wild-type *S. cerevisiae* IA3, albeit at a level of potency that was several orders of magnitude poorer. Against *S. cerevisiae* proteinase A, where the potency of wild-type *S. cerevisiae* IA3 is at or beyond the limit of detection by our apparatus, studies of the beneficial effects of mutations had been largely limited to peptides in which potency had been attenuated by truncation. The reduced potency against the *Pichia* proteinase provided a first opportunity to judge more accurately the effects, both positive and negative, of various mutations to the potency of the *S. cerevisiae* IA3 sequence in the full-length recombinant form. Additionally, by comparing the activity of mutants against two enzymes, influences on specificity, as well as potency, could be assessed. Also in this regard, *S. castellii* IA3 provided a second working inhibitory "template" with which to inspect the effect of various mutations and to compare the roles and contributions of individual residues.

Regrettably, however, one potentially invaluable source of information remained elusive throughout the studies. This was the sequence of any putative IA3 from *P. pastoris* itself. This may reflect an absence of such a protein from the *P. pastoris* genome, although it seems unlikely that this yeast does not regulate the activity of its vacuolar aspartic proteinase in a similar manner to the various *Saccharomyces* species identified in Section 6.1 (Figure 6.1B). Alternatively, it could be that the gene has not yet been identified during the ongoing *P. pastoris* genome sequencing project. A more sobering option in the context of the aims of this chapter, however, is that the *Pichia* IA3 sequence is sufficiently distinct from our starting templates of *S. cerevisiae* IA3 and *S. castellii* IA3 as to be apparently unrecognisable by BLAST searches. Neither BLAST searches using the *S. cerevisiae* and *S. castellii* IA3 sequences, nor searches using shared characteristics of the two, identified a potential IA3 analogue in *P. pastoris*. However, the fact that several very potent inhibitors of the *Pichia* enzyme were produced, from both *S. cerevisiae* and *S. castellii* IA3-derived sequences, displays once again the apparent diversity of possible IA3 sequences that can be functional inhibitors, as long as the correct

combination of residues is employed. This conclusion was reinforced by the range of different methods discovered for imparting such improvements in potency against the *Pichia* proteinase. Several features, depicted schematically in Figure 8.3, emerged consistently:

- Wild-type *S. castellii* IA3 in C-terminally His-tagged form (**cyan** square) is marginally superior to *S. cerevisiae* IA3 (**gold** square), although both are considerably poorer inhibitors of the *Pichia* enzyme than of *S. cerevisiae* proteinase A (4 nM and <0.1 nM, respectively).
- **Removing** the C-terminal tail (**yellow** arrows) is detrimental in **all** cases for *S. cerevisiae* IA3, but *beneficial* to *S. castellii* IA3, **except** when the centrepiece of *S. cerevisiae* IA3 (M18K/K22D, **brown** arrow and boxed in **gold**) is introduced. Then, as in *S. cerevisiae* IA3, tail removal is detrimental.
- **Exchanging** the C-terminal tails (**pink** arrows) showed that the *S. castellii* IA3 tail (boxed in **cyan**) was mildly detrimental to *S. cerevisiae* IA3. However, the *S. cerevisiae* IA3 tail (boxed in **gold**) was *beneficial* to the *S. castellii* inhibitory sequence, even greatly improving the potency of an otherwise very poor mutant (*S. castellii* M18V).
- **Interchanging** the back pins between IA3 species (**dark green** arrows,) improved *S. cerevisiae* IA3 moderately (boxed in **cyan**), but had little effect on *S. castellii* IA3 (boxed in **gold**)
- **Interchanging** the centrepieces between IA3 species (**brown** arrows and mutation boxed in colour corresponding to the origin of the centrepiece) also effectively interchanged the observed levels of potency. Thus, very low nanomolar potency was achieved by *S. cerevisiae* IA3 containing the *S. castellii* IA3 centrepiece (K18M/D22K).
- Other centrepiece mutations (**red** arrows) were similarly advantageous (e.g. *S. cerevisiae* mutants containing K18R), but could also be detrimental to potency.

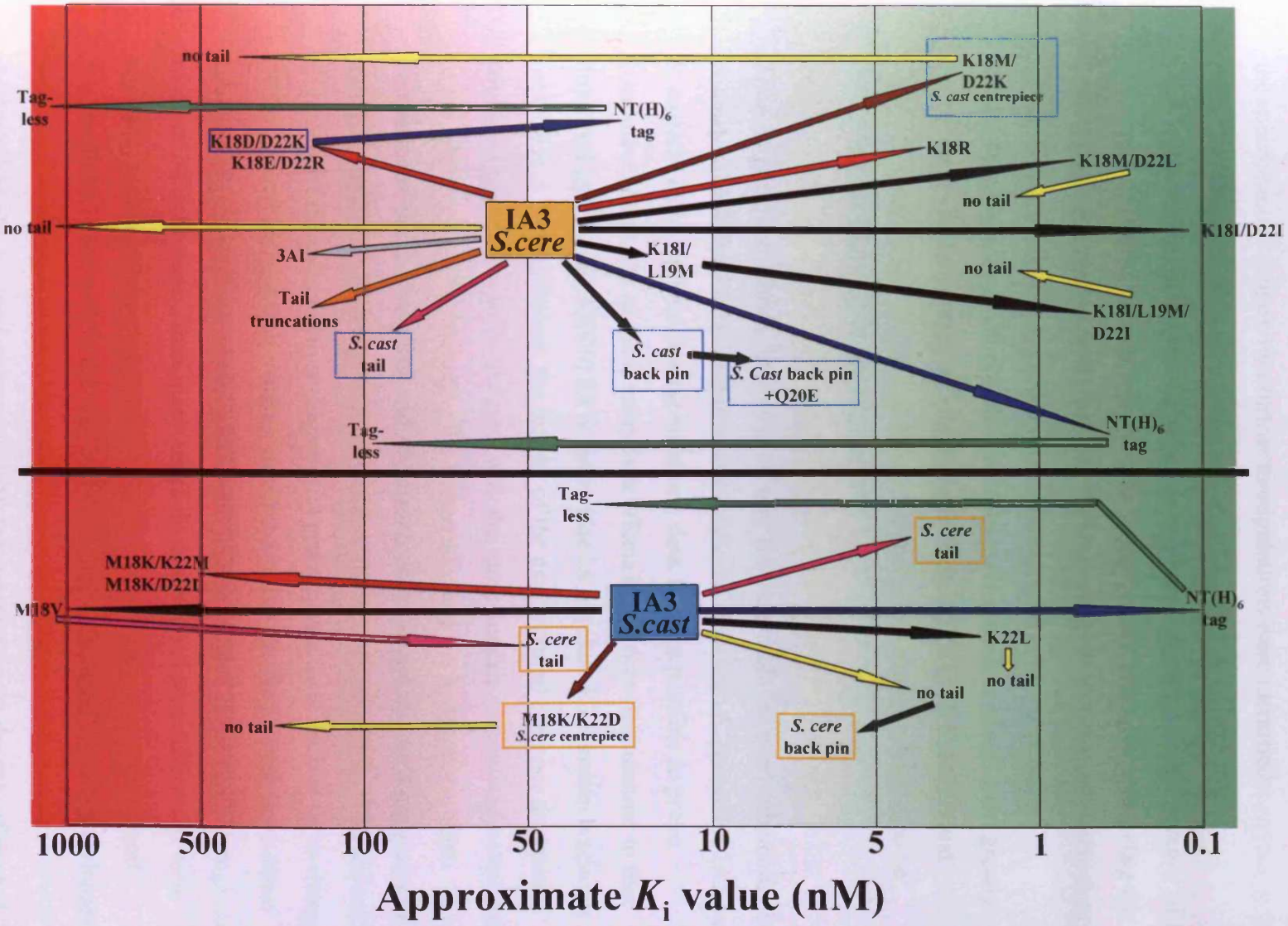


Figure 8.3 Graphical summary of the inhibitory data collected for the inhibition of the *P. pastoris* proteinase by derivatives of *S. cerevisiae* IA3 (gold square) and *S. castellii* IA3 (cyan square). The arrows are colour-coded. **Red**= centrepiece; **Black**= hydrophobic; **Blue** = $NT(H)_6$ tag attachment; **Green** = His-tag removal; **yellow**= tail removal (i.e. synthetic peptide); **pink** = tail exchange between IA3 species; **orange**= tail truncation; **grey**= domain swap; **brown**= centrepiece interchanged between IA3 species; **dark green**= back pin interchanged between IA3 species. Where a mutation introduced an element of the other "parent" IA3, it is boxed in **cyan** (for *S. castellii*) or **gold** (for *S. cerevisiae*). The diagram is not exhaustive, but highlights some of the more significant results.

However, in order to improve the K_i values by a further order of magnitude (i.e. into the subnanomolar range), two further manipulations were identified:

- **Relocation** of the His-tag (**blue** arrows) greatly improved the potency of IA3 proteins from both *S. castellii* and *S. cerevisiae*. However, **removing** the His-tag (**green** arrows) was detrimental, resulting in K_i values worse than the C-terminally His-tagged form
- **Hydrophobic** mutations to the centrepiece (**black** arrows) could greatly improve the potency of both *S. cerevisiae* IA3 (e.g. K18I/D22I) and *S. castellii* IA3 (K22L), although inappropriate changes could also be detrimental (e.g. *S. castellii* M18V).

Thus, by several methods, inhibitors of very low nanomolar or even subnanomolar potency against the *Pichia* enzyme were produced from both *S. cerevisiae* IA3 and *S. castellii* IA3. Without precise structural data, it is not possible to prove conclusively that any of these exert their effects in an equivalent manner to that described by Li *et al.*, (2000) for *S. cerevisiae* IA3 against *S. cerevisiae* vacuolar proteinase A. Nevertheless, the totality of the experimental evidence amassed strongly suggests this to be the case, with the same features present and contributing to the inhibitory action, including the centrepiece and the hydrophobic pins. Furthermore, the *S. cerevisiae* and *P. pastoris* enzymes are also strikingly similar in terms of sequence. This is particularly true of the active site region. The sequence alignment of the two enzymes (Figure 8.1) predicts that there are just three changes (boxed in red) amongst the residues lining the active site cleft which have direct contact with the inhibitory helix of a bound IA3 molecule. However, although the mode of inhibition contained many parallels, some significant differences were apparent, both between the two inhibitory "templates" (i.e. *S. cerevisiae* and *S. castellii* IA3) and also in the way by which they act against the two proteinases.

As described above, one fundamental difference centres upon the role of the tail region of IA3 (yellow arrows in Figure 8.3). Against *S. cerevisiae* vacuolar proteinase A, the activity of IA3 from both *S. cerevisiae* and *S. castellii* is located solely in the first 34 residues (Chapters 5 and 6). In contrast, however, wild-type

S. cerevisiae IA3 has an absolute requirement for a tail in order to inhibit *P. pastoris* proteinase (compare the arrows in Figure 8.3). Perhaps even more striking was that *S. castellii* IA3 still did *not* need a tail. Indeed, as was noted against *S. cerevisiae* vacuolar proteinase A also, the tail of *S. castellii* IA3 actually impairs the potency of the inhibitory sequence (compare **42** and **44** in Table 8.2). Furthermore, the truncation experiments and interspecies chimaeras (Section 8.7) again point to a role for residues 35-45 of *S. castellii* IA3, but not *S. cerevisiae* IA3, in this detrimental effect (see Section 7.5.2 in reference to the inhibition of the *S. cerevisiae* proteinase A).

Thus, two main conclusions in relation to the distinct roles of the two types of tail region of the two inhibitory sequences have been reached:

(i) the tail region of *S. cerevisiae* IA3 makes vital contributions to the activity of its inhibitory sequence against the *Pichia* enzyme. Against *S. cerevisiae* vacuolar proteinase A, a similar role for the C-terminal tail was witnessed only once, with the discovery that the K18D/D22K *S. cerevisiae* IA3 mutant was only potent in the presence of its C-terminal tail (compare **62** and **36** in Table 7.1).

(ii) The inhibitory sequence of *S. castellii* IA3 (residues 1-34) might have an inherently better binding energy, so negating the requirement for the contributions from the tail region.

These two points are supported by the weight of kinetic data. Even when the inhibitory potency is significantly improved by mutation, the tail region of *S. cerevisiae* IA3 still has a critical role. For example, two mutant *S. cerevisiae* IA3 sequences (**29** and **30** in Table 8.6) inhibit the *P. pastoris* proteinase in the absence of the tail region, although they are both approximately 10-fold less potent than in full-length recombinant form. Both possess hydrophobic replacements of the centrepiece residues at positions 18 and 22. Thus, it can be postulated that their ability to inhibit in 2-32 form arises from extra binding energy that is contributed by significant gains in the hydrophobic interactions between the inhibitor and the enzyme. These are sufficient to overcome the loss of the contributions made by the *S. cerevisiae* IA3 tail in the full-length forms.

No single factor appears to be responsible for the apparently superior binding of the wild-type *S. castellii* inhibitory sequence for the *Pichia* enzyme. However, a series of contributions are identifiable. In particular, as is also suggested by the benefit of the double hydrophobic mutations into *S. cerevisiae* IA3 (above), the centrepiece appears to be a critical region. For instance, the introduction of the *S. castellii* IA3 centrepiece (Met18 and Lys22) into *S. cerevisiae* IA3 did improve the K_i value of the protein form (**39** in Table 8.3) to be equivalent to that of the tailless *S. castellii* synthetic peptide (**44** in Table 8.4). However, this beneficial mutation was still insufficient in itself to enable the tailless synthetic peptide of the same inhibitory sequence to display significant levels of activity against *Pichia* proteinase (**80** in Table 8.4).

Perhaps to a lesser degree than the centrepiece, the back pin residues of *S. castellii* IA3 also appear to contribute small advantages over their *S. cerevisiae* IA3 counterparts, with a transfer of the *S. castellii* IA3 residues at positions 26, 29 and 30 into *S. cerevisiae* IA3 resulting in marginally superior inhibition of the *Pichia* proteinase (compare **52** with **1**, Table 8.5, or see dark green arrow in Figure 8.3). This might again reflect the alteration of one of the residues of the polyproline loop (Section 1.8.5) because Val295 of the *S. cerevisiae* enzyme is replaced with a larger Ile in the *Pichia* proteinase (Figure 8.1). This more closely resembles the Met of the *S. castellii* proteinase at this position, which may explain the preference for smaller residues in this region of the inhibitory helix (see Figure 6.2 and the associated discussion in Section 6.5). Once again, the change of Val26→Ala between *S. cerevisiae* and *S. castellii* IA3 also appears to confer a modest advantage (by comparison of **58** with **61**), probably because of a reduced interference with Tyr189 of the enzyme (see Figure 6.2 and Section 6.5).

However, the simultaneous introduction of both the *S. castellii* IA3 centrepiece and back pin residues into *S. cerevisiae* IA3 (**56**) did not improve upon the potency of the centrepiece-only mutant (**39**). This suggests that the differences between the two sequences in the back pin region are of secondary importance to the centrepiece. Likewise, comparison of **56** and **47** (Table 8.5) suggests that the first half of the *S. castellii* IA3 inhibitory sequence (residues 1-17) confers little overall advantage

over the equivalent *S. cerevisiae* IA3 region. Thus, the single change in the front pin region (Ile11→Met) does not account for the superior binding energy of the *S. castellii* IA3 inhibitory sequence either.

Therefore, the individual benefits of the *S. castellii* IA3 sequence are only modest, but occur at several positions and, cumulatively, amount to a significant effect. Also, it is conceivable that other regions or features might now make more significant positive (or negative) contributions. Although not previously considered to be significant (Section 5.3), differences between residues on the hydrophilic face of the inhibitory helix may also have a role. These more subtle effects were not manifest against *S. cerevisiae* proteinase A, probably because they were largely obscured by the major contributions made elsewhere in the sequences, for instance at the centrepiece and back pin. However, with these "big players" making smaller contributions against the new target enzyme, perhaps the beneficial donations of other residues are now more detectable in the kinetic assays. With no less than 19 differences between *S. castellii* and *S. cerevisiae* IA3 in the first 34 residues, it seems conceivable that some interactions of hitherto unidentified significance may also play roles.

For example, the presence of a Glu residue at position 20 of *S. castellii* IA3 offers the prospect of superior (*i*, *i*+4) interactions with Lys16 than those achievable by the Gln20 of *S. cerevisiae* IA3. Certainly, the introduction of a Q20E mutation into *S. cerevisiae* IA3 produced an improvement in potency (compare **52** and **53** in Table 8.5) that was perhaps larger than might have been expected from a residue alteration on the hydrophilic face of the helix (see Section 5.3).

Identifying the validity of such a suggestion, and disentangling the web of interactions is complicated when based purely on interpretation of K_i values, because these represent merely the accretion of all the positive and negative interactions along the sequence. In this regard, a crystal structure of a complex of at least one inhibitor with the *Pichia* proteinase would be invaluable. Regrettably, however, there are currently insufficient stocks of this proteinase to permit such a study and production of new batches of enzyme would require adjustment and optimisation of an industrial scale fermentation process. Despite this, the *Pichia* enzyme has served to enhance our

understanding of the complex interplay between a proteinase and an IA3-based inhibitor in a way that the *S. cerevisiae* vacuolar proteinase work alone could not.

With regard to the major objectives of these studies, the production of inhibitors of *P. pastoris* proteinase with potency equivalent to that observed for wild-type *S. cerevisiae* and *S. castellii* IA3 against the *S. cerevisiae* proteinase A was a success. However, all potent inhibitors of the *Pichia* vacuolar proteinase retained their potency against the *S. cerevisiae* enzyme as well. Thus, although several mutations (such as K18R, K18I/D22I and K18M/D22L in an *S. cerevisiae* IA3 context, for example) had introduced much improved activity against the *Pichia* proteinase, this reflected a *broadened* rather than an *altered* specificity. No inhibitor was truly re-targeted *exclusively* towards *Pichia* proteinase, and the level of specificity demonstrated by wild-type *S. cerevisiae* IA3 for *S. cerevisiae* vacuolar proteinase A was not recreated.

However, using wild-type *S. cerevisiae* IA3 as the reference point, some mutants were *partially* re-directed towards the *Pichia* enzyme. For example, both of the *S. cerevisiae* IA3 mutants K18V/D22K and K18M/D22K (**38** and **39**, respectively in Table 8.3) were less potent than the wild-type (**1**) against the *S. cerevisiae* proteinase, but more potent than wild-type against the *Pichia* enzyme. However, the chimaeric protein **56** (Table 8.5) came closest to achieving re-targeting, with almost identical K_i values measured against both enzymes. As such, this value was >20-fold worse than that of wild-type *S. cerevisiae* IA3 against *S. cerevisiae* vacuolar proteinase A, but >18-times *better* than wild-type against the *Pichia* enzyme.

In contrast, some of the mutations to *S. cerevisiae* IA3 achieved selectivity in the opposite direction, with all activity against the *Pichia* proteinase removed whilst retaining subnanomolar or very low nanomolar potencies against *S. cerevisiae* vacuolar proteinase A (for example, see **36**, **37**, **40**, **41** in Table 8.3). Presumably, the detrimental effects arising from these mutations are not sufficiently compensated for by the remaining binding contributions when the inhibitor is confronted with the active site of a proteinase which is not the normal target of the "parent" IA3.

Although potent inhibitors of the *Pichia* proteinase have been produced, true re-targeting of a specific IA3-like inhibitor against the *Pichia* proteinase using rational

re-design will require more information. In particular, precise structural details of the differences between the active sites of the two enzymes would be extremely valuable, so that differential selectivity could be engineered. However, regions outside the confines of the active site also appear to influence selectivity. The fortuitous discovery of the positive effect of an N-terminal His tag on the potency of IA3 molecules against the *Pichia* proteinase also has potential uses in re-targeting. Although His-tag relocation to the N-terminus had a modest effect on *S. castellii* IA3 against *S. cerevisiae* vacuolar proteinase A, the inhibitory potency against the *Pichia* proteinase was enhanced by approximately two orders of magnitude. Furthermore, against the *Pichia* enzyme the same phenomenon was displayed upon attachment to *S. cerevisiae* IA3, too. Thus, the N-terminally His-tagged forms of both *S. castellii* IA3 and *S. cerevisiae* IA3 displayed K_i values in the subnanomolar range.

The origins of this dramatic enhancement could be explained in a number of ways. It may, for example, reflect an increased stabilisation of the N-terminal region of the inhibitory helix, perhaps by dissipation of the helix dipole moment, leading to an increase in potency. However, the >100-fold improvement in the observed K_i value for both IA3 species would seem to be too large to be explained in this manner. Neither His nor Met, or even Gln, the three residues of the tag nearest to the IA3 sequence, are particularly favoured N-terminal helix-capping residues on their own (Fersht, 1998).

Molecular modelling studies conducted in conjunction with our collaborator Dr. Daniel Bur (Actelion Ltd., Allschwil, Switzerland) offer another explanation for the effects of the NT$(H)_6$ tag. Comparison of the X-ray crystal structure of *S. cerevisiae* vacuolar proteinase A and the model of the *P. pastoris* enzyme reveals a differentially-charged loop on the surface of the two proteinases relatively near to the entrance of the active site cleft, and within reach of a His-tag attached to the N-terminus of the inhibitory helix (see Figure 8.4):

	159				164	
S. cerevisiae proteinase:	**S**	**K**	**D**	**T**	**E**	**N**	**net charge = -1**
P. pastoris proteinase:	**D**	**K**	**D**	**E**	**S**	**D**	**net charge = -3**

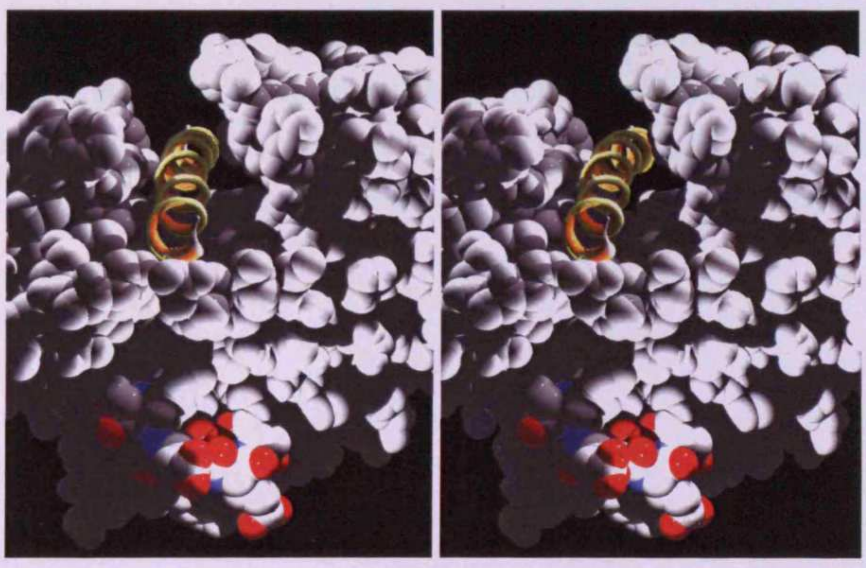


Figure 8.4 Stereo image of the location of the surface loop (residues 159-164) of *S. cerevisiae* proteinase A. The proteinase is displayed in spacefill, with all residues coloured grey apart from residues 159-164. The inhibitory α-helix of *S. cerevisiae* IA3 (viewed from its N-terminal end) is represented in gold.

Thus, if the $NT(H)_6$ tag extending from the N-terminus of the IA3 polypeptide establishes additional contacts with this region of the *Pichia* proteinase, this could make a significant contribution to the increased potency of binding. Any benefits to the inhibition of the less-charged *S. cerevisiae* proteinase A are less prominent. This possible explanation can be tested in future experiments using additional IA3 mutants or synthetic peptides with varying lengths of N-terminal His-tags. In addition, it would also be attractive, although considerably more complicated, to exchange this loop (159-164) between the *Pichia* and *S. cerevisiae* proteinases in order to study whether the increased beneficial effect of an $NT(H)_6$ tag on an IA3-type inhibitor is also transplanted to the other enzyme.

In conclusion, therefore, as well as producing several subnanomolar and very low nanomolar inhibitors of *P. pastoris* based upon either *S. cerevisiae* or *S. castellii* IA3 sequences, and beginning to identify methods of introducing a novel differential selectivity into IA3 for the first time, a fortuitous and fascinating phenomenon to improve binding to the *P. pastoris* proteinase was also discovered in these studies.

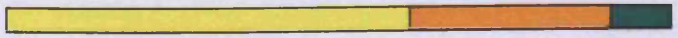
No	Name		K_i (nM)
1	WT *S. cerevisiae* IA3		55 ± 11
6	WT *S. cerevisiae* 2-34		NI at 2 μM
77	WT *S. cerevisiae* 35-68		NI at 2 μM
78	WT *S. cerevisiae* 35-68 $CT(H)_6$		NI at 2 μM
6+77	WT *S. cerevisiae* 2-34 + 35-68		NI at 200 nM (peptide 6) + 800 nM (peptide 77)
6+78	WT *S. cerevisiae* 2-34 + 35-68 $CT(H)_6$		NI at 200 nM (peptide 6) + 800 nM (peptide 78)
79	3AI (WT *S. cerevisiae* 35-68-1-34)		200 ± 20
4	WT *S. cerevisiae* IA3 Tyr57		225 ± 25
5	WT *S. cerevisiae* IA3 Ala45		200 ± 20

Table 8.2 Location of the inhibitory activity in *S. cerevisiae* IA3 against the *P. pastoris* proteinase, and the role of the C-terminal half of the polypeptide. The N-terminal 34 residues of IA3 are depicted schematically as , whilst residues 35-68 and the His-tag are shown as and , respectively. The schematic also reflects the partial tail truncations in **4** (after Tyr57) and **5** (after Ala45). NI = no inhibition. Assays were conducted at 37 °C, pH 4.7.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			P. pastoris	*S. cerevisiae*
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	55 ± 11	<0.1
31	K18R	MNTDQ QKVSE IFQSS KERLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	6 ± 2	<0.1
32	D22E	MNTDQ QKVSE IFQSS KEKLQ GEAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	20 ± 2	<0.1
33	K18R/D22E	MNTDQ QKVSE IFQSS KERLQ GEAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	5 ± 0.3	<0.1
36	K18D/D22K	MNTDQ QKVSE IFQSS KEDLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	300 ± 50	0.3 ± 0.1
37	K18E/D22K	MNTDQ QKVSE IFQSS KEELQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	IC50 ~ 730	1.3 ± 0.2
40	K18D/D22R	MNTDQ QKVSE IFQSS KEDLQ GRAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	IC50 > 500	5 ± 0.8
41	K18E/D22R	MNTDQ QKVSE IFQSS KEELQ GRAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	160 ± 15	0.7 ± 0.3
38	K18V/D22K	MNTDQ QKVSE IFQSS KEVLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	35 ± 2	2 ± 0.4
39	K18M/D22K	MNTDQ QKVSE IFQSS KEMLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	3 ± 0.9	0.2 ± 0.1

Table 8.3 The effect of mutation of centrepiece residues of recombinant *S. cerevisiae* IA3 on inhibition of the vacuolar proteinase from *P. pastoris* at pH 4.7. Sequence derived from *S. cerevisiae* IA3 is coloured yellow, with mutations highlighted in red, and residues corresponding to those from *S. castellii* IA3 in blue. Residues 1-34 are given in full, with the C-terminal tail region represented as ~~~~~~~~. Data determined at pH 4.7 against *S. cerevisiae* vacuolar proteinase A (from previous chapters) are included for comparison.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			P. pastoris	*S. cerevisiae*
42	WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	15 ± 3	4 ± 0.5
44	WT *S. castellii* IA3 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	4 ± 0.5	0.4 ± 0.1
45	N*cere*-C*cast* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~~~~~LE(H)$_6$	100 ± 20	<0.1
47	N*cast*-C*cere* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	3 ± 0.5	0.1 ± 0.1
48	*S. castellii* M18V IA3	MSDKN ANVSE MFQQA KEVLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	IC50 >1000	125 ± 20
49	N*cast* M18V-C*cere* IA3	MSDKN ANVSE MFQQA KEVLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	45 ± 4	25 ± 2
66	*S. castellii* M18K/K22D IA3	MSDKN ANVSE MFQQA KEKLE GDANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	IC50 ≈ 60	3 ± 0.8
67	*S. castellii* M18K/K22D 2-32	SDKN ANVSE ZFQQA KEKLE GDANA ASEGZ KN	280 ± 30	10 ± 1.5
63	*S. castellii* M18K/K22M IA3	MSDKN ANVSE MFQQA KEKLE GMANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	IC50 >500	370 ± 50
64	*S. castellii* M18K/K22Z 2-32	SDKN ANVSE ZFQQA KEKLE GZANA ASEGZ KN	IC50 >>500	350 ± 100
20	*S. cerevisiae* D22M IA3	MNTDQ QKVSE IFQSS KEKLQ GMAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	15 ± 3	<0.1
65	*S. castellii* M18K/K22L IA3	MSDKN ANVSE MFQQA KEKLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	IC50 >800	1000 ± 200
68	*S. castellii* K22L IA3	MSDKN ANVSE MFQQA KEMLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	2 ± 0.4	1.1 ± 0.2
69	*S. castellii* K22L 2-32	SDKN ANVSE ZFQQA KEZLE GLANA ASEGZ KN	3 ± 0.4	2.5 ± 0.5
80	*S. cerevisiae* K18Z/D22K 2-32	MNTDQ QKVSE IFQSS KEZLQ GKAKV VSDAF KK	IC50 >320	15 ± 3

Table 8.4 The effect of varying the tail and centrepiece regions of IA3 from *S. cerevisiae* and *S. castellii* on the inhibition of *P. pastoris* proteinase at pH 4.7. Sequence derived from *S. cerevisiae* IA3 is yellow, with *S. castellii* IA3 sequence in blue. Other mutations are in red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag and only residues 1-34 are depicted. The C-terminal tail region is represented as ~~~~~~~~, with the extra length of the *S. castellii* IA3 tail reflected in this. Z = norleucine. Data determined at pH 4.7 against *S. cerevisiae* vacuolar proteinase A (from previous chapters) are included for comparison.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			P. pastoris	*S. cerevisiae*
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	55 ± 11	<0.1
6	WT *S. cerevisiae* IA3 2-34	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKZA	NI at 2 µM	<0.1
42	WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	15 ± 3	4 ± 0.5
44	WT *S. castellii* IA3 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	4 ± 0.5	0.4 ± 0.1
47	N*cast*-C*cere* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	3 ± 0.5	0.1 ± 0.1
52	*S. cerevisiae* V26A/A29G/F30M	MNTDQ QKVSE IFQSS KEKLQ GDAKV ASDGM KKMA~~~~~~~~LE(H)$_6$	20 ± 2	<0.1
53	*S. cerevisiae* Q20E/V26A/A29G/F30M	MNTDQ QKVSE IFQSS KEKLE GDAKV ASDGM KKMA~~~~~~~~LE(H)$_6$	7 ± 0.9	<0.1
56	N*cast Sac*I/*Nhe*I IA3	MNTDQ QKVSE IFQSS KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	3 ± 0.2	2 ± 0.3
39	*S. cerevisiae* K18M/D22K	MNTDQ QKVSE IFQSS KEMLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	3 ± 0.9	0.2 ± 0.1
59	*S. castellii* A25V 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEGZ KNZA	6 ± 0.4	1.2 ± 0.3
57	*S. castellii* G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEAF KNZA	9 ± 0.9	3 ± 0.4
58	*S. castellii* A25V/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEAF KNZA	4 ± 0.4	1.1 ± 0.1
60	*S. castellii* A25I/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANI ASEAF KNZA	3 ± 0.2	1.4 ± 0.2
61	SEX 4	SDKN ANVSE ZFQQA KEZLE GKAKV VSDAF KKMA	15 ± 5	0.8 ± 0.1

Table 8.5 The effect of mutating the back pin residues on potency against the *Pichia* enzyme at pH 4.7. Sequence derived from *S. cerevisiae* IA3 is in yellow with *S. castellii* IA3 sequence in blue and, where mutations or chimaeras directly exchanged residues between species, the mutation is represented in the appropriate colour. Other mutations are in red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag and only residues 1-34 are depicted. The C-terminal tail region is represented as ~~~~~~~~. Z= norleucine. Data determined at pH 4.7 against *S. cerevisiae* vacuolar proteinase A (from previous chapters) are included for comparison.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			P. pastoris	*S. cerevisiae*
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	55 ± 11	<0.1
24	K18M/D22L IA3	MNTDQ QKVSE IFQSS KEMLQ GLAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.5 ± 0.5	<0.1*
23	K18I/L19M IA3	MNTDQ QKVSE IFQSS KEIMQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	15 ± 3	<0.5 (at pH 3.1)*
26	K18I/L19M/D22I IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.6 ± 0.15	<0.1
25	K18I/D22I IA3	MNTDQ QKVSE IFQSS KEILQ GIAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	<0.1
27	K18I/L19M/D22I Y57 IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~LE(H)$_6$[57]	0.2 ± 0.2	<0.1
28	K18I/L19M/D22I A45 IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~LE(H)$_6$[45]	0.2 ± 0.2	<0.1
29	K18I/L19Z/D22I 2-32	NTDQ QKVSE IFQSS KEIZQ GIAKV VSDAF KK	6 ± 0.5	<0.1
30	K18Z/D22L 2-32	NTDQ QKVSE IFQSS KEZLQ GLAKV VSDAF KK	3 ± 0.5	<0.1
46	NcereTM-Ccast IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~~~~~~~~LE(H)$_6$	1.5 ± 0.3	<0.1

Table 8.6 The effect of hydrophobic mutations to the centrepiece of *S. cerevisiae* IA3. All data were determined at pH 4.7. Sequence derived from *S. cerevisiae* IA3 is yellow with sequence derived from *S. castellii* IA3 in blue. Other mutations are in red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag and only residues 1-34 are depicted. The C-terminal tail region is represented as ~~~~~~~~. Z = norleucine. Data determined at pH 4.7 against *S. cerevisiae* vacuolar proteinase A (from previous chapters) are included for comparison. * = from Phylip *et al.* (2001)

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			P. pastoris	*S. cerevisiae*
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	55 ± 11	<0.1
42	WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	15 ± 3	4 ± 0.5
47	N*cast*-C*cere* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	3 ± 0.5	0.1 ± 0.1
50	NT(H)$_6$ *S. castellii* Q68 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~[68]	0.7 ± 0.1	0.8 ± 0.1
51	NT(H)$_6$ N*cast*-C*cere* IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	0.4 ± 0.1	0.7 ± 0.2
70	NT(H)$_6$ *S. castellii* Y57 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	1.1 ± 0.1	1.8 ± 0.3
71	NT(H)$_6$ *S. castellii* K45 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	0.9 ± 0.1	0.9 ± 0.4
72	NT(H)$_6$ N*cast*-C*cere* Y57 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	2 ± 0.2	3 ± 0.3
73	NT(H)$_6$ N*cast*-C*cere* A45 IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	2 ± 0.2	2 ± 0.6
43	NT(H)$_6$ WT *S. castellii* IA3	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	0.1 ± 0.1	0.3 ± 0.2
2	NT(H)$_6$ WT *S. cerevisiae* IA3	MK(H)$_7$MQ NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	0.6 ± 0.2	<0.1
3	Tagless WT *S. cerevisiae* IA3	*QNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	100 ± 10	0.1± 0.1
74	Tagless N*cast*-C*cere* IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	15 ± 1	1.5 ± 0.2
75	Tagless WT *S. castellii* IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	30 ± 4	3.9 ± 0.3
76	Tagless *S. castellii* K45 IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	30 ± 10	10 ± 1.5
81	NT(H)$_6$ *S. cerevisiae* K18D/D22K IA3	MK(H)$_7$MQ NTDQ QKVSE IFQSS KEDLQ GDAKV VSDAF KKMA~~~~~~~~	25 ± 2	<0.1
82	Tagless *S. cerevisiae* K18D/D22K IA3	*QNTDQ QKVSE IFQSS KEDLQ GDAKV VSDAF KKMA~~~~~~~~	>1000	0.6 ± 0.2

Table 8.7 The effect of His-tag location, tail length and tail identity on inhibitory potency of *S. cerevisiae* and *S. castellii* IA3 against the *P. pastoris* proteinase. Sequence derived from *S. cerevisiae* IA3 is coloured in yellow, *S. castellii* IA3 sequence is in blue. Mutations are coloured red. Recombinant proteins were produced with either a C-terminal LE(H)$_6$ or N-terminal MK(H)$_6$HMQ tag. Only residues of the inhibitory sequence (1-34) are shown, with the C-terminal tail region depicted as~~~~~~~~, relative to its length in that molecule. Selected proteins had their His-tag removed, a process involving the cyclisation of the exposed N-terminal Gln to pyroglutamate (*Q). Data determined at pH 4.7 against *S. cerevisiae* vacuolar proteinase A (from previous chapters) are included for comparison.

Chapter 9

IA3-based inhibitors of the probable vacuolar aspartic proteinase from *Aspergillus fumigatus*

9.1 Introduction

The successful production of mutant IA3 molecules with subnanomolar potencies against the vacuolar proteinase from *P. pastoris* (Chapter 8) inspired attempts to inhibit more distantly related, but pharmaceutically-relevant enzyme targets.

The classification of fungi appears to be a complex and somewhat controversial debate. The numerous phyla are classified primarily by their reproductive structures, however, the Kingdom of Fungi can also be divided broadly into two categories - the yeasts and the moulds. The evolutionary distance between *S. cerevisiae*, *S. castellii* and *P. pastoris* is relatively small, with all three contained within the same sub-phylum of yeasts (Figure 9.1). To broaden the scope of the challenge, therefore, efforts concentrated on the identification of a suitable target from a mould. Reichard *et al.*, (2000) reported the cloning and targeted deletion of an aspartic proteinase from the human fungal pathogen *Aspergillus fumigatus*. Despite the increased evolutionary distance between the organisms, a relatively high sequence identity with the vacuolar aspartic proteinases from *S. cerevisiae* and *P. pastoris* was maintained. This chapter recounts the efforts to produce highly potent IA3-based inhibitors of this orthologous proteinase from a medically-important mould.

9.2 The (probable) vacuolar aspartic proteinase from *A. fumigatus*

Early descriptions of the pathogenic saprophytic fungus *A. fumigatus* included those of Bennet (1842), who reported an aspergilloma, and Fresenius who in 1863 described an airway infection in poultry (Latgé, 1999). The most common agent of aspergillosis in both Man and animals, *A. fumigatus,* is an almost ubiquitous pathogen in the environment. As a result, humans and animals are prone to constant inhalation of conidia (asexual spores), although generally these are cleared quickly by the innate immune mechanisms of the host. However, with the increases in immuno-compromised human patients in recent years, for instance due to the prevalence of HIV/AIDS, the frequency of aspergillosis had increased some 14-fold in the period

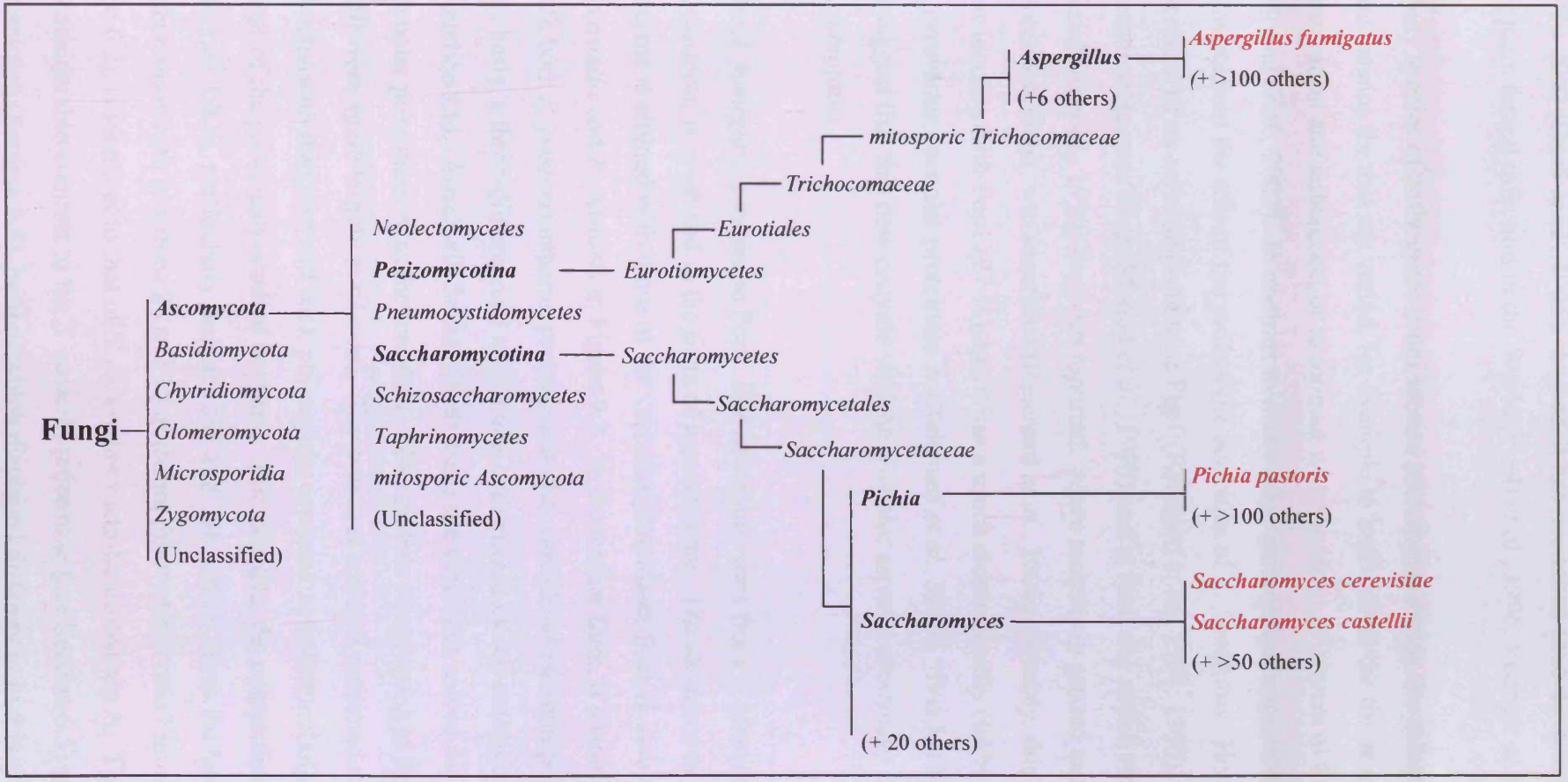


Figure 9.1 Taxonomic tree representing the distance between *A. fumigatus* and the other fungal species that feature significantly in this thesis (in red). (N.B. Only the most relevant levels of classification have been expanded at each rank).

1980-1992 (Groll *et al.*, 1996). Indeed, it has overtaken candidiasis as the most frequent fungal infection in the World (Groll *et al.*, 1996, Vogeser *et al.*, 1999).

Many species of pathogenic fungi secrete proteinases during the infection process. The reasons for this are varied, for example to break down the tissue and help with penetration and adherence, or to interact with the immune system of the infected host (Monod *et al.*, 2002). In contrast to many other pathogenic fungi, relatively little is known about the role of the proteolytic activities of *A. fumigatus*. However, the secretion of an aspartic proteinase Pep1 (Reichard *et al.*, 1994; 1995), a metalloproteinase Mep (Monod *et al.*, 1993) and at least one serine proteinase Alp1 (Reichard *et al.*, 1990) has been reported. More recently, a second aspartic proteinase, Pep2, was identified (Reichard *et al.*, 2000). Notably, this enzyme has a low identity with Pep1 (27 %), but it has a much closer identity (64 %) to *S. cerevisiae* vacuolar proteinase A (Reichard *et al.*, 2000). This led the investigators to suggest that this new enzyme was the vacuolar aspartic proteinase from *A. fumigatus*.

The *A. fumigatus* proteinase Pep2, like its counterparts from *S. cerevisiae* and *P. pastoris*, is produced in the form of a proenzyme. The sequence of the mature enzyme is aligned with those of the vacuolar proteinases from *S. cerevisiae*, *S. castellii* and *P. pastoris* in Figure 9.2. In the mature form, it shares 70 % identity with both *P. pastoris* aspartic proteinase and *S. cerevisiae* vacuolar proteinase A. On this basis, a three-dimensional model was constructed by our colleague Dr. D. Bur (Actelion Ltd., Allschwil, Switzerland) using the structures solved for *S. cerevisiae* vacuolar proteinase A as the template. The residues contributing to the active site cleft were again largely conserved. The principal residues predicted to interact directly with the front and back pin regions appeared to be particularly consistent with those of the previously-studied enzymes. For example, the polyproline loop (Section 1.8.5), particularly residues 295 and 298, which forms the basis of the interactions with positions 26 and 30 of the inhibitory helix (see Figures 1.14 and 6.2), is identical to that of *S. cerevisiae* vacuolar proteinase A. This in itself was a considerable contrast to the *S. castellii* proteinase (see Section 6.5) and the *Pichia* proteinase (Section 8.8), both of which displayed differences in this region.

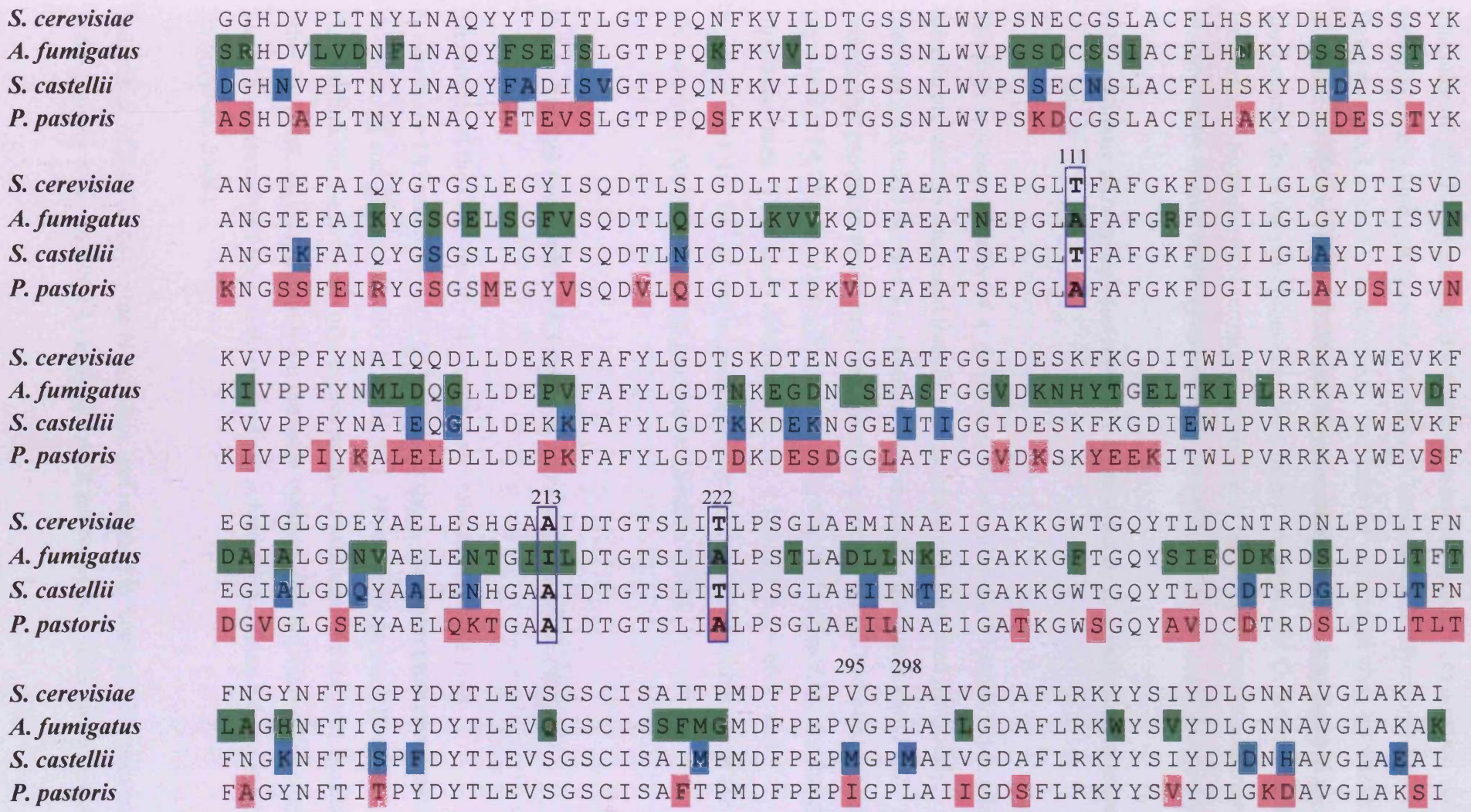


Figure 9.2 Sequence alignment of the *A. fumigatus proteinase* Pep2 with the vacuolar proteinases of *S. cerevisiae*, *S. castellii* and *P. pastoris*. Residues of the *S. cerevisiae* proteinase (and, by inference, the other enzymes) that potentially contact the inhibitory helix of IA3 directly, and that are different in the *A. fumigatus* proteinase, are boxed in blue.

The plasmid pPEP2, containing the *PEP2* gene in the *P. pastoris* expression vector pPICZaA was a kind gift from Dr. M. Monod (Laboratoire de Mycologie Service de Dermatologie, Lausanne, Switzerland). Induction of expression was performed as described by Monod *et al.*, (2002) and the recombinant enzyme was purified to homogeneity from the conditioned medium by my colleague Dr. D. Wyatt (Cardiff University). Subsequent N-terminal sequencing of the purified protein confirmed the identity as the mature *PEP2* gene product.

9.2.1 Kinetic parameters for substrate hydrolysis by *A. fumigatus* proteinase Pep2

The high sequence identity of *A. fumigatus*, *S. cerevisiae* and *P. pastoris* vacuolar aspartic proteinases suggested that the chromogenic substrate used previously (Section 2.2.29 and Section 8.2.1) might also be suitable for assays with Pep2. Indeed, this proved to be the case and kinetic parameters for the hydrolysis of this peptide at pH 4.7 by the *Aspergillus* enzyme were derived as described for the other fungal enzymes. The values obtained were K_m = 190 μM, k_{cat} = 66 s^{-1} and $k_{cat}/K_m = 3.5 \times 10^5$ M^{-1} s^{-1} and as such, are directly comparable with those of the *Pichia* and *S. cerevisiae* aspartic proteinases (Section 8.2.1).

9.3 Initial tests with IA3 from *S. cerevisiae* and *S. castellii*

The ability of the C-terminally His-tagged recombinant forms of both *S. cerevisiae* IA3 and *S. castellii* IA3 to inhibit the *A. fumigatus* proteinase was examined (**1** and **42**, respectively in Table 9.1). Neither protein was effective. Furthermore, on the evidence of previous analyses with related inhibitors against other enzymes, such weak inhibition probably indicates that the *Aspergillus* enzyme would also cleave both polypeptides slowly as substrates (Sections 5.6 and 8.2.2; Phylip *et al.*, 2001).

Relocation of the His-tag to the N-terminus had resulted in substantial improvements to the potency of *S. castellii* IA3 against both *S. cerevisiae* vacuolar proteinase A

(10-fold, Section 6.2) and, particularly, against the *Pichia* proteinase (160-fold, Section 8.7) where it had also produced a similar improvement to *S. cerevisiae* IA3. However, against the *A. fumigatus* proteinase, His-tag relocation did not improve either *S. castellii* or *S. cerevisiae* IA3 as inhibitors (**43** and **2**, respectively, in Table 9.1). Removal of the N-terminal His-tag was also without benefit (**75** and **3** in Table 9.1). In these recombinant forms, therefore, the wild-type sequences of the two IA3 species did not significantly inhibit the new enzyme.

The C-terminally truncated variants of *S. castellii* IA3 (**50**, **70**, **71** and **76**) and the chimaeras (**45**, **47**, **51**, **72**, **73**, **74**) were also found to be ineffective as inhibitors of the *Aspergillus* enzyme (Table 9.1). Similarly, further shortening at the C-terminal end to produce the variant consisting of residues 2-34 from *S. cerevisiae* IA3 (**6**) resulted in a peptide that did not inhibit the Pep2 enzyme.

Remarkably, however, the peptide corresponding to residues 2-34 of wild-type *S. castellii* IA3 (**44** in Table 9.1) was 35-fold better than the C-terminally His-tagged full-length recombinant form (**42**) and >100-fold more potent than the tagless form (**75**), which was the nearest full-length equivalent. Mass spectrometric analysis confirmed that peptide **44** remained intact following a 16 h incubation at 37 °C, pH 6.8 at a 20:1 (inhibitor:enzyme) molar ratio. The implications of these results were two-fold. Firstly, it demonstrated the presence of effective inhibitory activity against the *Aspergillus* enzyme within residues 2-34 of *S. castellii* IA3 that was absent from the *S. cerevisiae* IA3 counterpart. Secondly, in recombinant protein form, the C-terminal IA3 tail region of *either species* obscured this activity (compare **44** with **42** and **47**).

9.4 Comparing the roles of the centrepiece residues

One of the regions identified as critical to the superior potency of *S. castellii* IA3 against the *Pichia* proteinase was the centrepiece residues. Variants with this M/K (18/22) centrepiece changed to K/M or K/L, or the K/D arrangement of *S. cerevisiae* IA3, were examined therefore, in peptide and full-length forms (**63** – **67**

in Table 9.2). None was effective as an inhibitor. Clearly, then, the centrepiece is also of key importance important for the potency of *S. castellii* IA3 against the *A. fumigatus* proteinase. However, introduction of the M/K centrepiece of *S. castellii* IA3 into the *S. cerevisiae* polypeptide did not generate any inhibitory activity against the *Aspergillus* proteinase in either peptide or full-length protein forms (**39**, **80** in Table 9.2), and nor did any of the other centrepiece variants of the *S. cerevisiae* IA3 sequence depicted in Table 9.2 (**20**, **31** - **33**, **36** - **38**, **40**, **41**). This suggested that the *S. castellii* IA3 centrepiece combination was only of benefit if other regions of the sequence were also of sufficient compatability with the *Aspergillus* proteinase active site. That is, the introduction of the *S. castellii* IA3 centrepiece was not sufficient to overcome other disadvantages elsewhere in the *S. cerevisiae* inhibitory sequence.

9.5 The role of the back pin region of IA3

As one of the regions of most substantial difference between the *S. castellii* and *S. cerevisiae* inhibitory sequences that had also been shown previously to make significant binding contributions, the back pin residues were investigated next. Introduction of the back pin residues (positions 26, 29 and 30) of *S. castellii* IA3 into the *S. cerevisiae* IA3 sequence, and in the fortuitous Q20E mutant (that was described in Section 6.4), had no effect (**52** and **53** in Table 9.3). In the chimaeric protein **56** (Table 9.3) the centrepiece and back pin regions of *S. castellii* IA3 were both introduced into *S. cerevisiae* IA3 simultaneously. However, this did not result in a substantial improvement to the potency of *S. cerevisiae* IA3 either (compare with **1** in Table 9.1). Furthermore, the chimaeric synthetic 2-34 peptide **61**, where residues 24-34 from *S. cerevisiae* IA3 were introduced into the *S. castellii* inhibitory sequence, showed a >10-fold reduction in potency. That is, 2-34 *S. castellii* IA3 effectively lost all activity upon the introduction of the entire *S. cerevisiae* IA3 back pin region, despite the maintenance of the M18/K22 centrepiece arrangement (compare **61** and **44** in Table 9.3).

In an attempt to identify whether individual residues or the overall nature of the *S. cerevisiae* back pin region were responsible for the loss of potency in chimaeric peptide **61**, the series of peptides **57** - **60** (Table 9.3) was examined. Replacing Gly29 and Met30 with their corresponding residues from *S. cerevisiae* IA3 (Ala and Phe, respectively) produced an almost 10-fold drop in potency (compare **57** with **44** in Table 9.3), reflecting the situation with peptide **61**. A similar effect was noted with the A25V mutation (**59** in Table 9.3). Individually, therefore, the components of the *S. cerevisiae* back pin at positions 25, 29 and 30 were extremely disadvantageous. Remarkably, however, the introduction of all three of these mutations simultaneously in peptide **58** (Table 9.3) returned a K_i value approaching that of the wild-type peptide (**44**), and replacement of Val25 by Ile (peptide **60** in Table 9.3) resulted in a K_i value matching that of the wild-type peptide **44**. Thus, as long as the Met/Nle at 18 with Lys at 22 combination is maintained as the centrepiece, the *S. castellii* inhibitory sequence is almost equally effective with either its own wild-type residues at positions 25, 29 and 30 or those of *S. cerevisiae* IA3. This finding appeared consistent with the fact that the predicted helix-contacting residues constituting the polyproline loop region of the *A. fumigatus* proteinase were identical to those of *S. cerevisiae* proteinase A (see Section 9.2).

9.6 The effect of hydrophobic mutations to the IA3 sequence

Most of the binding contributions identified from the crystal structure of *S. cerevisiae* IA3 complexed with *S. cerevisiae* vacuolar proteinase A arise from hydrophobic interactions (Li *et al.*, 2000; Phylip *et al.*, 2001; Section 1.8.5). The replacement of the centrepiece with residues to increase the hydrophobic binding energy had improved the potency of both species of IA3 against the *P. pastoris* proteinase (Section 8.6). With a similar objective, the C-terminally His-tagged *S. castellii* IA3 mutant K22L (**68** in Table 9.4) was tested against the *Aspergillus* proteinase. In synthetic peptide form, replacement of the Z/K (18/22) centrepiece of the wild-type *S. castellii* IA3 peptide (**44** in Table 9.1) by this Z/L arrangement proved detrimental to inhibition, with a loss of potency of approximately 10-fold (**69** in Table 9.4). Nevertheless, the introduction of the equivalent centrepiece

arrangement (M/L) into the full-length recombinant *S. castellii* IA3 protein form resulted in an 8-fold *improvement* in the apparent inhibition (compare **68** with **42** in Table 9.4). As such, protein **68** was the first recombinant protein with an observed K_i value <100 nM. Furthermore, it also appeared to be approximately 2-fold tighter than the corresponding peptide (compare **68** and **69** in Table 9.4), suggesting that this K22L mutation had counteracted the detrimental effect of the *S. castellii* IA3 tail in some way.

The introduction of this double hydrophobic centrepiece of M18/L22 into *S. cerevisiae* IA3 had an arguably more profound effect. For the first time in a *S. cerevisiae* IA3 context, a significant level of inhibitory activity was observed. Both recombinant protein and 2-32 synthetic peptide forms displayed K_i values <50 nM (**24** and **30** in Table 9.4). Positioning Ile residues at 18 and 22 in *S. cerevisiae* IA3 (**25** in Table 9.4) generated an even more potent inhibitor, with the first K_i value below 10 nM recorded for any IA3 derivative from either inhibitory species against the *A. fumigatus* enzyme. Even the additional replacement of Leu19 by Met (**26** and **46** in Table 9.4) did not significantly alter the potency of the I18/I22-containing sequence. Indeed, removal of all residues beyond 32 again resulted in only a mild loss of potency for the resulting peptide form of this mutant (**29** in Table 9.4). Therefore, as long as the correct combination of residues is found, potent inhibition of the *Aspergillus* enzyme can be achieved by polypeptides based on the *S. cerevisiae* IA3 sequence as well as that of *S. castellii* IA3.

9.7 Discussion

From a consideration of the sequence, structural modelling and kinetic data, the Pep2 proteinase from *A. fumigatus* shows the hallmarks of being from the same stable of enzymes as the vacuolar aspartic proteinases of *S. cerevisiae* and *P. pastoris*. Although phylogenetically *A. fumigatus* is more distantly related to *S. cerevisiae* than is *P. pastoris*, the actual shared identity of the mature forms of the proteinases remains high (see Figure 9.1). The values (approximately 70%, see Section 9.2) are only marginally lower than the identity between *S. cerevisiae* proteinase A and the

Pichia enzyme (approximately 75%), which was still inhibited by both wild-type *S. cerevisiae* IA3 and *S. castellii* IA3 to varying degrees (Chapter 8). Yet the potency of both inhibitors in recombinant form was reduced a further 10-40 fold against the *Aspergillus* enzyme, such that no significant inhibition by either was noted. This appeared indicative, once again, of the exquisite specificity of the interaction between wild-type IA3 and its target species.

Despite these results, the insights gained from the studies against the *S. cerevisiae* and *P. pastoris* proteinases were brought successfully to bear on the task of identifying potent inhibitors of the *Aspergillus* enzyme. Several of the key features, such as the centrepiece, the back hydrophobic pin and the C-terminal tail were once again shown to be critical. This also suggests that activity against this enzyme is mediated in an approximately equivalent manner to the established paradigm of *S. cerevisiae* IA3 inhibiting the *S. cerevisiae* proteinase A.

The detrimental effect of the *S. castellii* IA3 C-terminal tail (i.e. residues post-34), first witnessed against *S. cerevisiae* proteinase A and again with the *Pichia* enzyme, is even more pronounced against the *A. fumigatus* proteinase. Remarkably, the absence of the C-terminal tail from the 2-34 synthetic peptide form converted the inactive (and almost certainly slowly cleaved) *S. castellii* IA3 into an effective inhibitor of the new target. In contrast, against the *P. pastoris* and *S. cerevisiae* aspartic proteinases at pH 4.7, this influence of the tail was apparently restricted to the *S. castellii* IA3 tail acting on its own inhibitory sequence (Chapters 6 and 8). However, against the *A. fumigatus* enzyme, the interference by the tail is irrespective of tail type and the catastrophic loss of potency upon attachment of a tail is also witnessed in the N*cast*-C*cere* chimaeric IA3 (compare **44** with **75** and **74** in Table 9.1). Furthermore, the loss of inhibitory power remains upon partial truncation of either tail as far as residue 45 (**76**).

Although collectively these results point to residues 35-45 in the C-terminal tail (of either IA3 species studied) shrouding the action of the inhibitory sequence of *S. castellii* IA3, at least one result indicates that the interplay between the two regions of IA3 has a further complexity. The retention of the tail sequence does NOT appear to interfere with the inhibitory action of the *S. castellii* IA3 mutant K22L (compare **68**

and **69** in Table 9.4). This is a clear contrast. Also, it raises the questions of the significance or possible relationship of a Lys at 22 to the detrimental effect originating in the C-terminal tail. However, the discovery of a possible interplay between Lys22 and the tail region is not unique to the *S. castellii* IA3 inhibition of *A. fumigatus* proteinase. This "neutralisation" of a detrimental effect of the *S. castellii* IA3 tail region in specific situations was also noted in Chapters 7 and 8. On those occasions, too, an apparent inter-relationship between Lys22 and the C-terminal tail could be removed by replacing Lys22 with a hydrophobic Leu.

The differing levels of activity displayed by IA3 from *S. castellii* and *S. cerevisiae* represents a further step in the progression from *S. cerevisiae* proteinase A, via the *Pichia* enzyme, to the new target. Against the *P. pastoris* proteinase, both wild-type IA3 forms had functioned to varying degrees: *S. cerevisiae* IA3 in full-length protein form only; *S. castellii* IA3 as either a protein or (more potently) a peptide (2-34). In contrast against the *A. fumigatus* proteinase, only the peptide form of *S. castellii* IA3 remains functional, but even the binding of this 2-34 region now exhibits greatly increased sensitivity to the introduction of sub-optimal mutations or detrimental elements, such as the C-terminal tail. However, because removal of the tail sequence does *not* improve the potency of *S. cerevisiae* IA3 against the *A. fumigatus* proteinase (**6** in Table 9.1), a situation also consistent with the findings against the *P. pastoris* proteinase, the *S. castellii* inhibitory sequence appears to be the inherently superior of the two for the inhibition of this new enzyme.

As described previously, the two IA3 species have different residues constituting the back hydrophobic pin. These are all smaller or less bulky in *S. castellii* IA3. Against the *Pichia* enzyme, the simultaneous introduction of the *S. castellii* IA3 pin residues at 26, 29 and 30 into *S. cerevisiae* IA3 had improved inhibition to a level comparable with recombinant *S. castellii* IA3. However, against the *A. fumigatus* proteinase this does not help. Conversely, the *S. castellii* IA3 sequence can accommodate the *S. cerevisiae* IA3 residues at 25, 29 and 30 with little effect on potency (compare **58** with **44**, Table 9.3), but *only* if all three residues are inserted simultaneously (compare **58** with **57** and **59**). Closer inspection of the mutants suggests that the (*i*, *i*+4) pairing at positions 25 and 29 on adjacent turns of the inhibitory helix is the most critical consideration in this region of *S. castellii* IA3. *Either* the *S. cerevisiae* IA3 pair of

V25/A29 *or* the natural *S. castellii* IA3 pair (A25/G29) satisfy the requirements, but not a hybrid of the two. Therefore, it would appear that in the back pin region the *combination* of residues, rather than the identity of individual constituent side chains, is the key to inhibition of the *Aspergillus* proteinase. This probably reflects the clustering of the side chains, from both the enzyme and the inhibitor, within the confines of the active site cleft. Thus, some positional adjustments can be accommodated and may even generate more productive contacts, but others reduce functional interactions or result in disadvantageous conflicts, clashes or distortions to such a degree that the potency is much reduced. This will be considered further in the General Discussion (Chapter 11).

Another distinction between the behaviour of the two inhibitors against the *Pichia* and *Aspergillus* proteinases is apparent in the role of the centrepiece region. Although it retains activity in synthetic peptide form, the inhibitory activity of the *S. castellii* IA3 sequence also appears to be "in the balance" in the centrepiece region. None of the mutations introduced into the centrepiece of *S. castellii* IA3 are beneficial and only one, the K22L mutation discussed earlier (itself detrimental by comparison of **69** with **44** in Table 9.4), is tolerated even remotely. The *S. castellii* IA3 sequence is also no longer capable of accommodating the *S. cerevisiae* centrepiece residues of K18/D22. Possibly then, this centrepiece combination is far less appropriate for the inhibition of the *A. fumigatus* proteinase than for either the *P. pastoris* or *S. cerevisiae* enzymes. This seems likely as a factor, not least because *S. cerevisiae* IA3 fails to inhibit the new target at all. However, it could also reflect an inability of the rest of the *S. castellii* inhibitory sequence to compensate for this detrimental change, suggesting that the remainder of the *S. castellii* sequence is also considerably less well suited to the inhibition of the *Aspergillus* proteinase.

Differences are also noted in the effect of mutations to the *S. cerevisiae* IA3 centrepiece. Devices that enhance the activity of *S. cerevisiae* IA3 against the *P. pastoris* proteinase, such as K18R, or the introduction of alternative single hydrophobic mutations into the centrepiece (K18V/D22K, D22M) are ineffectual with the *Aspergillus* enzyme. Moreover, whereas the introduction of the *S. castellii* IA3 centrepiece had resurrected activity from the impotent *S. cerevisiae* IA3 sequence against the *P. pastoris* proteinase, this K18M/D22K mutation no longer resolves the

issue against the *Aspergillus* enzyme (**39** and **80** in Table 9.2). Even in combination with the simultaneous introduction of the entire *S. castellii* IA3 back pin region (**56** in Table 9.3), the *S. castellii* IA3 centrepiece is insufficient to overcome the disadvantages of the *S. cerevisiae* IA3 sequence which presumably remain in the first 17 residues and/or the tail region. Furthermore, these negative effects must be more significant than against the *Pichia* enzyme, where **56** was a potent inhibitor (3 ± 0.2 nM). As with the *P. pastoris* work, identifying contributions to the apparent superiority of the *S. castellii* inhibitory sequence (for the inhibition of the *A. fumigatus* proteinase) from outside of the centrepiece and hydrophobic pin regions is difficult without fresh structural data. In contrast to the *P. pastoris* proteinase, however, it was possible for my colleague Dr. Wyatt to produce crystallographic quantities of the *Aspergillus* enzyme within our laboratory. Therefore, a complex of peptide **44** with the *A. fumigatus* proteinase has been prepared, and supplied to our collaborating crystallographic laboratory (Dr Alex Wlodawer, National Cancer Institute, Frederick, MD, USA). At the time of writing, no crystallographic data have yet been obtained.

However, the total removal of charged residues from the *S. cerevisiae* IA3 centrepiece, and replacement with hydrophobic residues at both positions 18 and 22 simultaneously, does provide an avenue for the generation of low nanomolar, *S. cerevisiae* IA3-derived, inhibitors of the *Aspergillus* proteinase (see **24**, **25**, **26** and **46** in Table 9.4). These function in either the presence or absence of the C-terminal halves (compare **24** with **30**, and **26** with **29**). Thus, the introduction of a totally hydrophobic centrepiece is sufficient to offset the other disadvantages in the *S. cerevisiae* inhibitory sequence. Nevertheless, whilst this strategy creates effective inhibitors of the *Aspergillus* enzyme, the hydrophobic centrepiece also achieves a broadening rather than re-targeting of specificity. This is apparent because these hydrophobic centrepiece mutants are also amongst the most potent inhibitors of the *S. cerevisiae* and *P. pastoris* vacuolar aspartic proteinases as well. However, the broadening is now extended to embrace a different evolutionary subphylum of fungi. Thus the success of this work indicates that, ultimately, the creation of IA3-based inhibitors for other pharmaceutically-relevant fungal proteinases is also a realistic and achievable proposition.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			A. fumigatus	*S. cerevisiae*
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	630 ± 110	<0.1
42	WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	600 ± 50	4 ± 0.5
43	NT(H)$_6$ WT *S. castellii* IA3	MK(H)$_6$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	170 ± 25	0.3 ± 0.2
2	NT(H)$_6$ WT *S. cerevisiae* IA3	MK(H)$_6$MQ NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	530 ± 100	<0.1
75	Tagless WT *S. castellii* IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	1900 ± 200	4 ± 0.3
3	Tagless WT *S. cerevisiae* IA3	*QNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	>1000	0.1± 0.1
50	NT(H)$_6$ *S. castellii* Q68	MK(H)$_6$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~[68]	800 ± 150	0.8 ± 0.1
70	NT(H)$_6$ *S. castellii* Y57	MK(H)$_6$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	270 ± 25	2± 0.3
71	NT(H)$_6$ *S. castellii* K45	MK(H)$_6$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	150 ± 25	0.9 ± 0.4
76	Tagless *S. castellii* K45 IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	320 ± 60	10 ± 1.5
45	N*cere*-C*cast* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~~~~~LE(H)$_6$	1200 ± 100	<0.1
47	N*cast*-C*cere* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	340 ± 80	0.1 ± 0.1
51	NT(H)$_6$ N*cast*-C*cere* IA3	MK(H)$_6$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	680 ± 100	0.7 ± 0.2
72	NT(H)$_6$ N*cast*-C*cere* Y57	MK(H)$_6$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	1800 ± 250	3 ± 0.3
73	NT(H)$_6$ N*cast*-C*cere* A45	MK(H)$_6$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	1750 ± 200	2 ± 0.6
74	Tagless N*cast*-C*cere* IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	550 ± 80	1.5 ± 0.2
6	WT *S. cerevisiae* 2-34	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKZA	1500 ± 300	<0.1
44	WT *S. castellii* 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	17 ± 2	0.4 ± 0.1

Table 9.1 The effect of tail length, tail identity and His-tag location on the inhibitory potency of *S. cerevisiae* and *S. castellii* IA3 against the *A. fumigatus* proteinase at pH 4.7. Sequence derived from *S. cerevisiae* IA3 is coloured in yellow, *S. castellii* IA3 sequence is in blue. Recombinant proteins were produced with either a C-terminal LE(H)$_6$ or N-terminal MK(H)$_6$HMQ tag. Only residues of the inhibitory sequence (1-34) are shown, with the C-terminal tail region depicted as~~~~~~~~, relative to its length in that molecule. Selected proteins had their His-tag removed, a process involving the cyclisation of the exposed N-terminal Gln to pyroglutamate (*Q). Z= norleucine. Data determined at pH 4.7 against *S. cerevisiae* vacuolar proteinase A (from previous chapters) are included for comparison.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			A. fumigatus	*S. cerevisiae*
63	*S. castellii* M18K/K22M IA3	MSDKN ANVSE MFQQA KEKLE GMANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	> 5 μM	370 ± 50
64	*S. castellii* M18K/K22Z 2-32	SDKN ANVSE ZFQQA KEKLE GZANA ASEGZ KN	> 5 μM	350 ± 100
65	*S. castellii* M18K/K22L IA3	MSDKN ANVSE MFQQA KEKLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	> 5 μM	1000 ± 200
66	*S. castellii* M18K/K22D IA3	MSDKN ANVSE MFQQA KEKLE GDANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	> 5 μM	3 ± 0.8
67	*S. castellii* M18K/K22D 2-32	SDKN ANVSE ZFQQA KEKLE GDANA ASEGZ KN	> 5 μM	10 ± 1.5
39	*S. cerevisiae* K18M/D22K IA3	MNTDQ QKVSE IFQSS KEMLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	> 5 μM	0.2 ± 0.1
80	*S. cerevisiae* K18Z/D22K 2-32	MNTDQ QKVSE IFQSS KEZLQ GKAKV VSDAF KK	> 5 μM	16 ± 3
20	*S. cerevisiae* D22M IA3	MNTDQ QKVSE IFQSS KEKLQ GMAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	720 ± 250	<0.1
31	*S. cerevisiae* K18R IA3	MNTDQ QKVSE IFQSS KERLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	620 ± 60	<0.1
32	*S. cerevisiae* D22E IA3	MNTDQ QKVSE IFQSS KEKLQ GEAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	NI at 10 μM	<0.1
33	*S. cerevisiae* K18R/D22E IA3	MNTDQ QKVSE IFQSS KERLQ GEAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	350 ± 30	<0.1
36	*S. cerevisiae* K18D/D22K IA3	MNTDQ QKVSE IFQSS KEDLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	NI at 10 μM	0.3 ± 0.1
37	*S. cerevisiae* K18E/D22K IA3	MNTDQ QKVSE IFQSS KEELQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	NI at 12 μM	1.3 ± 0.2
38	*S. cerevisiae* K18V/D22K IA3	MNTDQ QKVSE IFQSS KEVLQ GKAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	NI at 9 μM	2 ± 0.4
40	*S. cerevisiae* K18D/D22R IA3	MNTDQ QKVSE IFQSS KEDLQ GRAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	NI at 9 μM	5 ± 0.8
41	*S. cerevisiae* K18E/D22R IA3	MNTDQ QKVSE IFQSS KEELQ GRAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	~ 5 μM	0.7 ± 0.3

Table 9.2 The effect of mutations to the centrepiece residues of *S. castellii* IA3 and *S. cerevisiae* IA3. All data were collected at pH 4.7. Sequence derived from *S. cerevisiae* IA3 is coloured in yellow, *S. castellii* IA3 sequence is in blue. Mutations are coloured red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag. Only residues of the inhibitory sequence (1-34) are shown, with the C-terminal tail region depicted as~~~~~~~~, relative to its length. Z = norleucine. Data determined at pH 4.7 against *S. cerevisiae* vacuolar proteinase A (from previous chapters) are included for comparison.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			A. fumigatus	*S. cerevisiae*
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	630 ± 110	<0.1
44	WT *S. castellii* 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	17 ± 2	0.4 ± 0.1
52	*S. cerevisiae* V26A/A29G/F30M IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV ASDGM KKMA~~~~~~~~LE(H)$_6$	NI at 9 μM	<0.1
53	*S. cerevisiae* Q20E/V26A/A29G/F30M IA3	MNTDQ QKVSE IFQSS KEKLE GDAKV ASDGM KKMA~~~~~~~~LE(H)$_6$	IC50 > 10 μM	<0.1
56	N*cast Sac*I/*Nhe*I C*cere* IA3	MNTDQ QKVSE IFQSS KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	280 ± 50	2 ± 0.3
61	SEX 4	SDKN ANVSE ZFQQA KEZLE GKAKV VSDAF KKZA	230 ± 50	0.8 ± 0.1
57	*S. castellii* G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEAF KNZA	120 ± 20	3 ± 0.4
59	*S. castellii* A25V 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEGZ KNZA	130 ± 8	1.2 ± 0.3
58	*S. castellii* A25V/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEAF KNZA	25 ± 3	1.1 ± 0.1
60	*S. castellii* A25I/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANI ASEAF KNZA	15 ± 3	1.4 ± 0.2

Table 9.3 The effect of mutating the back pin residues. Sequence derived from *S. cerevisiae* IA3 is yellow, *S. castellii* IA3 sequence is blue. Where mutations or chimaeras directly interchanged equivalent residues between species, the mutation is represented in the appropriate colour. Other mutations are in red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag and only residues 1-34 are depicted. The C-terminal tail region is represented as ~~~~~~~~, with the extra length of the *S. castellii* IA3 tail reflected in this. Z= norleucine. Data determined at pH 4.7 against *S. cerevisiae* vacuolar proteinase A (from previous chapters) are included for comparison.

No	Name	sequence 1 5 10 15 20 25 30	K_i (nM)	
			A. fumigatus	*S. cerevisiae*
1	WT *S. cerevisiae* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	630 ± 110	<0.1
42	WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	600 ± 50	4 ± 0.5
44	WT *S. castellii* 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	17 ± 2	0.4 ± 0.1
68	*S. castellii* K22L IA3	MSDKN ANVSE MFQQA KEMLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	75 ± 15	1.1 ± 0.2
69	*S. castellii* K22L 2-32	SDKN ANVSE ZFQQA KEZLE GLANA ASEGZ KN	140 ± 10	2.5 ± 0.5
24	*S. cerevisiae* K18M/D22L IA3	MNTDQ QKVSE IFQSS KEMLQ GLAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	35 ± 5	<0.1
30	*S. cerevisiae* K18Z/D22L 2-32	NTDQ QKVSE IFQSS KEZLQ GLAKV VSDAF KK	40 ± 5	<0.1
25	*S. cerevisiae* K18I/D22I IA3	MNTDQ QKVSE IFQSS KEILQ GIAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	9 ± 1	<0.1
26	*S. cerevisiae* K18I/L19M/D22I IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	13 ± 2	<0.1
46	N*cere*TM-C*cast* IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~~~~~~~~LE(H)$_6$	12 ± 3	<0.1
29	*S. cerevisiae* K18I/L19Z/D22I 2-32	NTDQ QKVSE IFQSS KEIZQ GIAKV VSDAF KK	22 ± 4	<0.1

Table 9.4 Hydrophobic mutations to the centrepieces of *S. castellii* IA3 and *S. cerevisiae* IA3. Sequence derived from *S. cerevisiae* IA3 is yellow, *S. castellii* IA3 sequence is blue. Where mutations or chimaeras directly interchanged equivalent residues between species, the mutation is represented in the appropriate colour. Other mutations are in red. Recombinant proteins were produced with a C-terminal LE(H)$_6$ tag and only residues 1-34 are depicted. The C-terminal tail region is represented as ~~~~~~~~, with the extra length of the *S. castellii* IA3 tail reflected in this. Z = norleucine.

Chapter 10

The future? A human enzyme target and investigating phage display-based directed evolution as a fresh approach for inhibitor discovery

10.1 Introduction

In the work described in the last two chapters, modified IA3 molecules were produced with novel inhibitory activities against fungal vacuolar proteinases of increasing evolutionary distance from the natural target enzymes. A more ambitious final challenge was to screen the existing repertoire of variant IA3 molecules for activity against an enzyme from a species of a different evolutionary kingdom entirely. Based on the similarity to fungal vacuolar aspartic proteinases, the human lysosomal enzyme cathepsin D was identified as an appropriate new target.

Principally, the choices of IA3 mutations described in this thesis were informed by either the X-ray crystallographic structures of *S. cerevisiae* IA3 variants complexed with *S. cerevisiae* vacuolar proteinase A (Li *et al.*, 2000; Phylip *et al.*, 2001), or the models of the *Aspergillus* and *Pichia* enzymes, constructed by our colleague Dr. D. Bur. However, with each step away from the natural target, the range of practical solutions identified became more restricted, which highlighted the increased difficulty when confronted with an absence of genuine structural data. Without fresh information from the X-ray crystallographic studies of the new inhibitors complexed with the *P. pastoris* and *A. fumigatus* aspartic proteinases, therefore, strategies based on rational design were becoming exhausted, so demonstrating the limitations of this approach.

Popular alternatives to rational design in the pharmaceutical and biotechnological industries are combinatorial screening methods (Section 1.9). With such approaches, no specific knowledge of the role of individual residues is necessary, depending instead on the ability to produce and screen an extensive repertoire of closely-related variant sequences. Therefore, this chapter also describes the initial feasibility testing of the biologically-based combinatorial approach of bacteriophage display as an alternative strategy for the identification of new IA3-derived inhibitors.

10.2 Similarity of cathepsin D to fungal vacuolar aspartic proteinases and relevance of cathepsin D as a target

Aguilar *et al.*, (1997) compared the lysosomal enzyme human cathepsin D with their elucidated structure of vacuolar aspartic proteinase A from *S. cerevisiae*. Yeasts lack lysosomes, but employ their vacuole to perform similar functions, as well as other tasks such as nutrient storage and osmotic balance (Thumm, 2000). The equivalence of location of proteinase A in *S. cerevisiae* and cathepsin D in humans is reflected in the sequence similarity of the two enzymes, which share 46% identity (and this figure rises to 63% if conservative replacements are included). Furthermore, the two enzymes are also structurally very similar. Aguilar *et al.*, (1997) commented that the proteinase A structure (PDB code 2JXR) was within 0.88 Å (rms deviation at the αC atoms) of the crystal structure of cathepsin D (PDB code 1LYA) reported by Baldwin *et al.* (1993), and that most differences were localised to surface loop regions. Thus, although the sequence identity between *S. cerevisiae* proteinase A and human cathepsin D was much lower than the identity shared by the fungal vacuolar enzymes considered in this thesis, the presence of crystal structures of both cathepsin D and *S. cerevisiae* vacuolar proteinase A in the Protein Data Bank simplified their comparison.

Cathepsin D was considered a particularly attractive, pharmaceutically-relevant target because it has been implicated in many of the most-researched medical conditions of recent times. For example, roles have been reported in tumorigenesis (Cooper, 2002) and regulation of angiogenesis (Perchick & Jabbour, 2003; Piwnica *et al.*, 2004), breast cancer metastasis (Majer *et al.*, 1997; Dash *et al.*, 2003), Alzheimer's disease (Majer *et al.*, 1997) and apoptosis (Kagedal *et al.*, 2001; Li *et al.*, 2004; Erdal *et al.*, 2005). Furthermore, Brown *et al.* (2004) noted that cathepsin D expression was upregulated in scrapie-infected mouse brains and, some time ago, a critical role in development was proposed (Saftig *et al.*, 1995). There are also reports (Igarashi *et al.*, 2004) of a role in the skin condition psoriasis.

However, despite involvement in such a diversity of medical conditions, relatively few new inhibitors against cathepsin D have been reported in recent years

(Kick *et al.*, 1997; Dumas *et al.*, 1999; Bi *et al.*, 2000; McConnell *et al.*, 2003). In fact, it is probable that this paucity of inhibitors of cathepsin D is *because* its many and varied roles render it a particularly complicated therapeutic target (Dash *et al.*, 2003). Previously, wild-type IA3 from *S. cerevisiae* showed no inhibitory activity against cathepsin D in either naturally-occurring (Dreyer *et al.*, 1985) or 2-34 synthetic peptide forms (Phylip *et al.*, 2001). However, one of the most attractive elements of IA3 as an inhibitor is the exquisite specificity for the natural target, making its redirection against such a potentially problematic but pharmaceutically attractive target as cathepsin D a fascinating challenge.

As detailed in Chapter 1, very few other naturally-occurring, gene-encoded inhibitors of aspartic proteinases have been described (Dunn, 2002), although at least four have been found to act against cathepsin D. Perhaps the most relevant to the work described here is the API-8 inhibitor from *Solanum tuberosum* (potato) and its homologue from tomato (Section 1.7.2), which inhibit both cathepsin D and *S. cerevisiae* proteinase A (Cater *et al.*, 2002), but no other aspartic proteinases tested. Whilst the sequence of API-8 is quite different to IA3, this apparent similarity in terms of the behaviour of an inhibitor encouraged attempts to identify IA3 sequences capable of inhibiting human cathepsin D.

10.3 Determination of the kinetic parameters for substrate hydrolysis by cathepsin D

Human cathepsin D was a kind gift from Dr. S. Gulnik (NCI, Frederick, MD, USA). To simplify the comparison with the other enzymes assessed in this thesis, the same chromogenic substrate (sequence KPIEF*F′ RL, Sections 2.2.29, 8.2.1 and 9.2.1) was used. As it was of interest to examine the susceptibility of cathepsin D to inhibition at pH 3.1 and pH 4.7, it was first necessary to establish the characteristics of the enzyme at the both pH values.

The enzyme activity was lower at pH 4.7, meaning that the concentration of cathepsin D needed in the assays at pH 4.7 was approximately 2-fold higher than at

pH 3.1, as determined by active site titration with a known concentration of isovaleryl-peptstatin (Section 2.2.31). The reduced activity was reflected in the differences in k_{cat} at the two pH values:

	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} M^{-1})
pH 3.1	170	130	7.5×10^5
pH 4.7	230	65	3×10^5

10.4 Testing IA3-based inhibitors against cathepsin D

10.4.1 pH 3.1 data

Without exception, at pH 3.1 none of the sequences derived from *S. cerevisiae* IA3 (listed in Table 10.1) showed any inhibitory activity whatsoever against cathepsin D at the maximum concentration that could be included feasibly in a spectrophotometric assay (<5% of total assay volume). The highest concentrations used were within the range of 400 - 12,000 nM. Likewise, the majority of sequences based on *S. castellii* IA3 displayed no inhibitory activity at pH 3.1 (Table 10.2). Nevertheless, a few of these *S. castellii* IA3-derived sequences (marked with ✓ in Table 10.2) did show an interaction, which was reflected as very weak concentration-dependent inhibition. However, the K_i value was never <1000 nM. The fact that some of the *S. castellii* IA3 molecules had an effect at all was a significant contrast to the *S. cerevisiae* IA3-derived polypeptides. To illustrate the difference, Figure 10.1 shows the dose-response curves for the synthetic peptide representing residues 2-34 of wild-type *S. castellii* IA3 (**44** blue curve) and, by way of comparison, the *S. cerevisiae* IA3 mutant K18M/D22K containing the same centrepiece (**39** red line). However, mass spectrometric analysis following a 16 h incubation of peptide **44** with cathepsin D at pH 3.1 in a 300:1 molar ratio confirmed that this peptide was cleaved by the enzyme. The observed mass peaks at 2431.1 Da and 1956.8 Da correlated exactly with the predicted masses of residues $F^{12} \rightarrow A^{34}$ and $K^{16} \rightarrow A^{34}$, respectively. Thus, it would appear that, despite the very weak apparent inhibitory activity, peptide **44** is slowly digested at the Z^{11}-F^{12} and A^{15}-K^{16} bonds by cathepsin D.

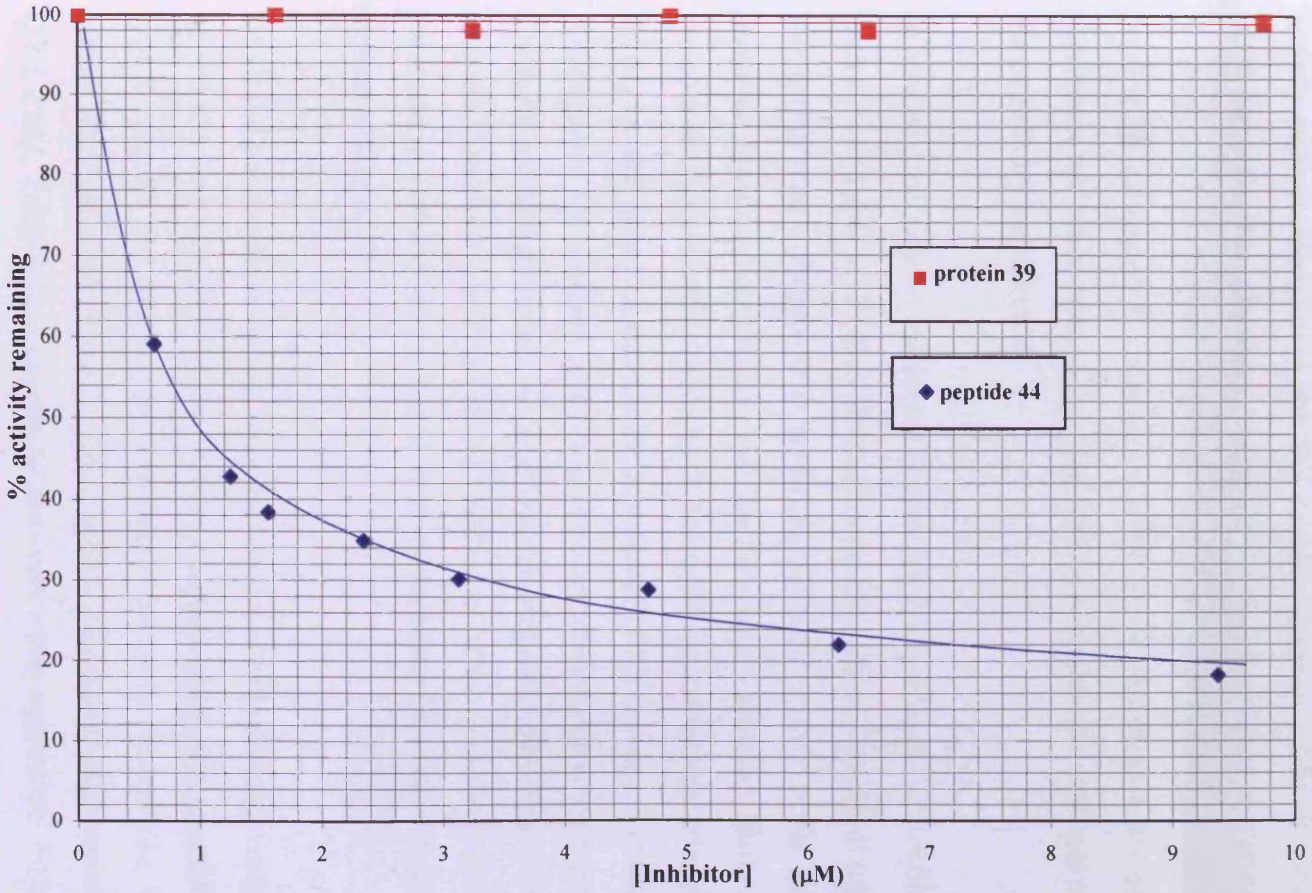


Figure 10.1 Concentration-dependent inhibition of human cathepsin D at pH 3.1. Activity was calculated as a percentage of the rate of substrate hydrolysis observed in the absence of inhibitor. Peptide **44** (blue): the peptide representing residues 2-34 of wild-type *S. castellii* IA3; protein **39** (red): the *S. cerevisiae* IA3 mutant protein containing the same centrepiece residues (K18M/D22K) as **44**.

10.4.2 pH 4.7 data

Against *S. cerevisiae* vacuolar proteinase A, IA3 potency was found to be universally superior at pH 4.7 (Chapters 5, 6 and 7). Therefore, to assess any pH-dependency of activity against cathepsin D, the IA3 variants listed in Table 10.1 were tested at pH 4.7 as well. No inhibitory interaction with cathepsin D was apparent from any *S. cerevisiae* IA3-derived inhibitory sequences, with the exception of the synthetic peptide **30** (estimated K_i value of 1.5 μM). However, the potency of this dose-dependent inhibition was so low that it most probably also reflects slow cleavage of the peptide by the enzyme.

In contrast, the majority of tested sequences derived from *S. castellii* IA3 (marked with a ✓ in Table 10.2) demonstrated a concentration-dependent inhibitory interaction with cathepsin D at the higher pH value of 4.7. In the cases where the observed K_i values were <1000 nM, the numerical values are given in Table 10.2. In the course of experiments against all three of the enzymes described previously in this thesis, i.e. the vacuolar aspartic proteinases of *S. cerevisiae*, *P. pastoris* and *A. fumigatus*, such values would not have been noteworthy, except to attract a comment on their probable "second" nature as slowly cleaved substrates of the enzyme. However, the fact that ***any*** inhibition of the hydrolysis of the chromogenic substrate was observed contrasted with all the other IA3-based molecules tested against cathepsin D.

Of all the polypeptides tested, only one produced effective inhibition. A K_i value of 15 ± 3 nM was recorded for the synthetic peptide representing residues 2-34 of wild-type *S. castellii* IA3 (peptide **44**, highlighted in red in Table 10.2). This K_i value was almost identical to the potency of this peptide against the *Aspergillus* proteinase (17 ± 2 nM, Table 9.1), and only 4-fold poorer than against the *Pichia* enzyme (4 ± 0.5 nM, Table 8.4).

To establish whether this K_i value reflected genuine inhibitory action, a sample of peptide **44** was incubated with cathepsin D at pH 4.7 and 37 °C at a molar ratio of 35:1 (inhibitor:enzyme). This equated to approximately 80 nM peptide **44** in a

spectrophotometric assay. After 4 min (the typical duration of a spectrophotometric assay), any residual activity of the enzyme was stopped by the addition of 50 μM isovaleryl-pepstatin and the sample was analysed by MALDI-TOF mass spectrometry. The single mass peak observed (3489.5 Da) correlated with the theoretical mass of the intact peptide (3487.8 Da). This indicated that, over the course of the assay period, the observed inhibitory constant reflected genuine inhibition of the enzyme. However, MALDI-TOF mass spectrometry of a second sample of **44** incubated with cathepsin D, but this time for 18 h at a molar ratio of 10:1 (inhibitor:enzyme), revealed distinct cleavage products. As at pH 3.1 (Section 10.4.1), peptide **44** appeared to be cleaved at the Z^{11}-F^{12} and the A^{15}-K^{16} bonds (observed masses of 2430.15 Da and 1955.9 Da, respectively). Therefore, the complex of cathepsin D with peptide **44** was not sufficiently stable to prevent slow cleavage of the peptide at the lower ratios of inhibitor:enzyme and over the extended time period typically required for crystallisation and X-ray structural analysis. Nevertheless, the apparent inhibition by peptide **44** over the timecourse of the kinetic assay demonstrates that a fungal IA3-type molecule is capable of occupying the active site of human cathepsin D in an inhibitory manner, despite the considerable evolutionary distance between the species of the target enzyme and the inhibitor. However, without any structural data relating to the interaction of this solitary "lead" compound with the enzyme, it was not immediately obvious how to improve the potency further. This is perhaps a reflection of the limitations of rational design approaches. Investigations into an alternative strategy for the discovery of new IA3-based inhibitors were thus initiated. These centred on directed evolution.

10.5 A fresh approach to new inhibitor discovery: bacteriophage display

The theoretical sequence space encompassed by a polypeptide of the length of the inhibitory sequence region of IA3 (33 residues) is bewilderingly large – certainly too large for all 20^{33} (approximately 8.6×10^{42}) possible combinations to be represented in a combinatorial library! In real terms, however, by a combination of random chance and co-evolution with their natural target enzymes, Nature's design of IA3

sequences has itself closed down much of this space. Nevertheless, the chances of improved sequence discovery using combinatorial techniques lie in the production and screening of as many functionally divergent forms of the "template" as possible. As a result, neither of the test sample production systems employed during our rational design procedures, i.e. solid-phase peptide synthesis and pET vector-based recombinant protein expression in *E. coli*, were suitable for a combinatorial approach. This was because the production of more than a handful of molecules by either system was prohibitively expensive in terms of either cost (solid-phase synthesis) or time (recombinant production).

Biologically-based combinatorial techniques, such as bacteriophage display, provide an attractive alternative because library populations of 10^{10} individuals can be produced simply and cheaply. Hypermutagenic PCR, using the IA3 gene (or an assortment of IA3 genes encoding some of the more potent inhibitory sequences) as a template, could generate a vast range of diverging mutant DNA sequences. These could be cloned into a coat protein gene of an appropriate bacteriophage. Each bacteriophage particle in the "library", therefore, would display a unique variant IA3 peptide sequence within the coat (although the degeneracy of the genetic code and the hypermutagenic rate of the PCR, together with random chance, suggest that many sequences would be represented more than once). Theoretically, by the application of a suitable selection pressure, such as an ability to bind to an immobilised target aspartic proteinase, bacteriophage displaying variant sequences with superior activity would be isolated from the library. Sequencing the appropriate region of the genome(s) of these phage would identify the promising new peptide sequence(s), enabling their production as synthetic peptides or recombinant proteins and quantification of their inhibitory potencies. Furthermore, their DNA sequences would provide templates for new rounds of hypermutagenesis. Using increasingly stringent selection conditions, it was hoped that improvements in both potency and specificity could be "evolved", leading to the discovery of not only potent inhibitors of new target enzymes but also to an improved understanding of the origins of the activity of IA3.

The most widely exploited bacteriophage for directed evolution is the filamentous family, Ff. However, the relatively new T7Select™ system marketed by Novagen offered a number of potential advantages for the work described here (listed below).

(i) T7, an icosahedral bacteriophage, is more robust than the filamentous forms. In particular, the manufacturer of theT7Select™ system suggests that this bacteriophage demonstrates a wider pH tolerance, which is of considerable advantage because the aspartic proteinase targets to be used in the biopanning steps are only active at acidic pH values.

(ii) Unlike filamentous bacteriophage systems, the polypeptides displayed on the surface of T7 do not need to be secreted through the cell membrane (Russel, 1991).

(iii) The fusion is at the solvent-exposed C-terminus of the T7 major coat protein. Thus, although the N-terminus of the displayed peptide sequence is "tethered", the C-terminus is not. In the case of IA3, it was hoped this would both aid accessibility to the enzyme active site and permit formation of the inhibitory α-helix. Certainly, the domain-swap protein, 3AI (**79** in Table 8.2, and see Section 3.5) demonstrated that an N-terminally tethered inhibitory sequence had a potency equivalent to that of the unrestricted sequence (against *S. cerevisiae* proteinase A, at least).

(iv) Multivalent display of longer (>7 residue) sequences is much simpler using the T7Select™ system. For polypeptides <50 residues in length, the T7Select™ vector 415-1b allows for display of the sequence in all 415 copies of the bacteriophage capsid coat protein which, theoretically, increases the sensitivity of detection.

(v) Growth and replication times of T7 bacteriophage are much faster, reducing the time to perform multiple rounds of enrichment.

For these reasons, the T7Select™ system was chosen for these studies. As an initial "proof-of-concept", attempts were made to optimise a system that would exclusively select T7 bacteriophage displaying residues 2-34 of wild-type *S. cerevisiae* IA3 from within a mixed population of bacteriophage displaying an alternative peptide sequence.

10.5.1 Production of T7Select™ bacteriophage displaying either residues 2-34 of wild-type *S. cerevisiae* IA3 or S•Tag™

Using the pET22b-IA3 plasmid vector (Section 2.1.1) as template, the DNA encoding residues 2-34 of wild-type *S. cerevisiae* IA3 was amplified by PCR (Figure 10.2). The forward primer (ϕF, see Appendix II.G) attached the restriction enzyme site (*Eco*RI) required for cloning into the T7Select™ vector, together with three Gly codons (to encode a flexible linker between the phage coat protein and the IA3 sequence) and a single base that was required to maintain the reading frame. The reverse primer (ϕR, see Appendix II.G) introduced a stop codon immediately after codon 34 of the IA3 sequence, followed by the *Hin*dIII restriction site also required for the cloning process. Double-stranded DNA sequencing (Section 2.2.11) using the vector-specific M13 forward and reverse primers confirmed that the PCR had produced the desired sequence. The PCR product was prepared for ligation into the T7Select™ cloning kit (Novagen) vector DNA, which is supplied pre-cut into two "arms", by treatment with *Eco*RI and *Hin*dIII (Section 2.2.2). The ligation, together with packaging *in vitro* of the ligation reaction mixture, was performed as described in Section 2.2.12. Double stranded DNA sequencing, using the vector-specific PhageSeqF and PhageSeqR primers (see Appendix II.E) that anneal to either side of the T7Select™ vector cloning region, confirmed the incorporation of the sequence encoding residues 2-34 of wild-type *S. cerevisiae* IA3, preceded by a $(Gly)_3$ linker, into the phage coat protein 10B. The correct reading frame was also maintained, and the coat protein gene was now terminated by the stop codon introduced after codon 34 of the IA3 sequence. These T7Select™ phage, displaying 415 copies of residues 2-34 of *S. cerevisiae* IA3 on their capsid, were referred to as "*ϕphage*".

The T7Select™ cloning kit is also supplied with a ligation-ready positive control insert DNA fragment, encoding the 15 amino acid peptide tag known as S•Tag™. Thus, S•Tag™-displaying T7Select™ phage were produced in a similar fashion and used as the control "strain" to assess non-specific binding. These were referred to as "*+phage*". Again, double-stranded DNA sequencing confirmed the success of the ligation and packaging. To investigate any effects of displaying residues 2-34 of IA3

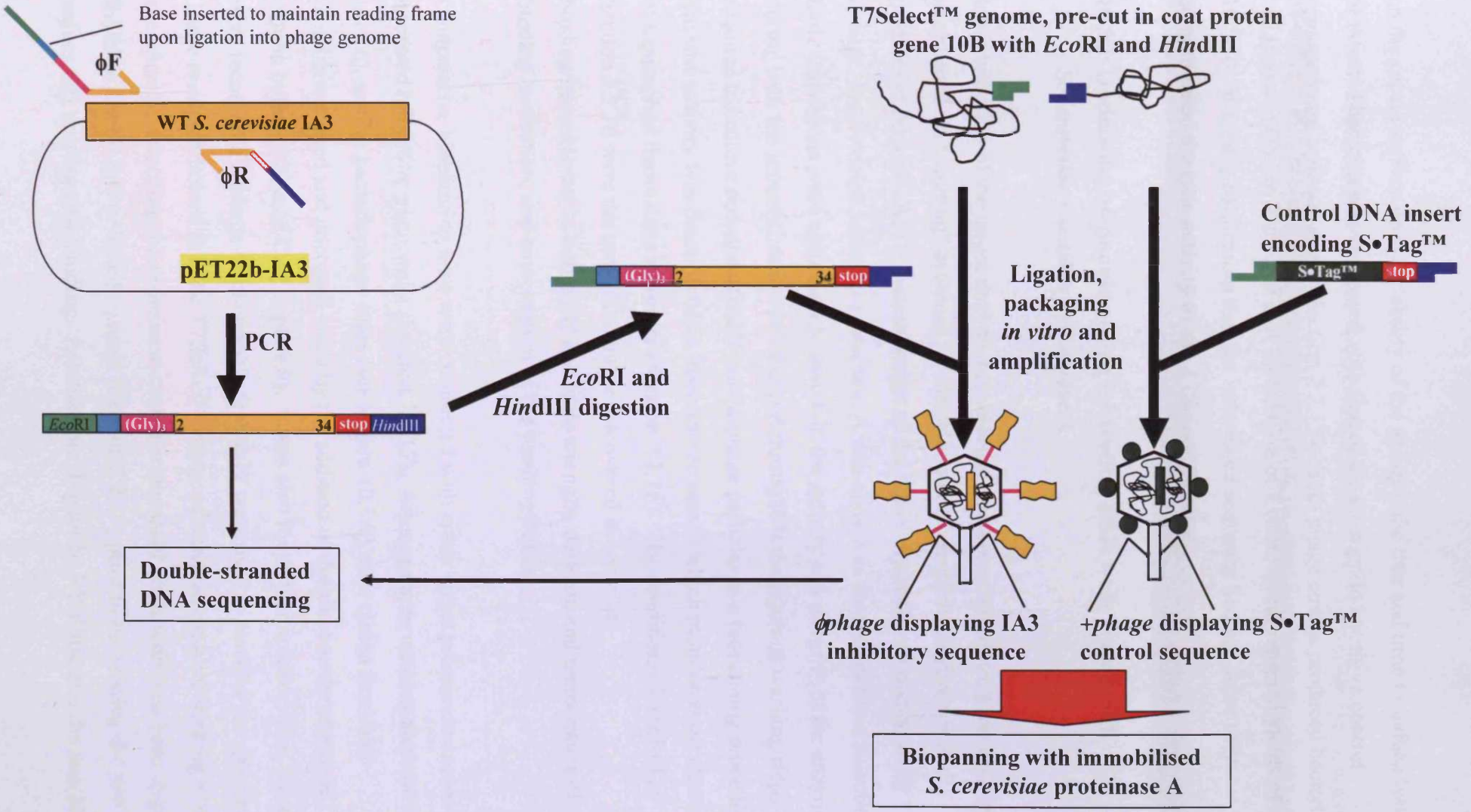


Figure 10.2 Schematic representation of the production of T7Select™ bacteriophage displaying either residues 2-34 of wild-type *S. cerevisiae* IA3 ("*ϕphage*") or S•Tag™ ("*+phage*") on their capsid.

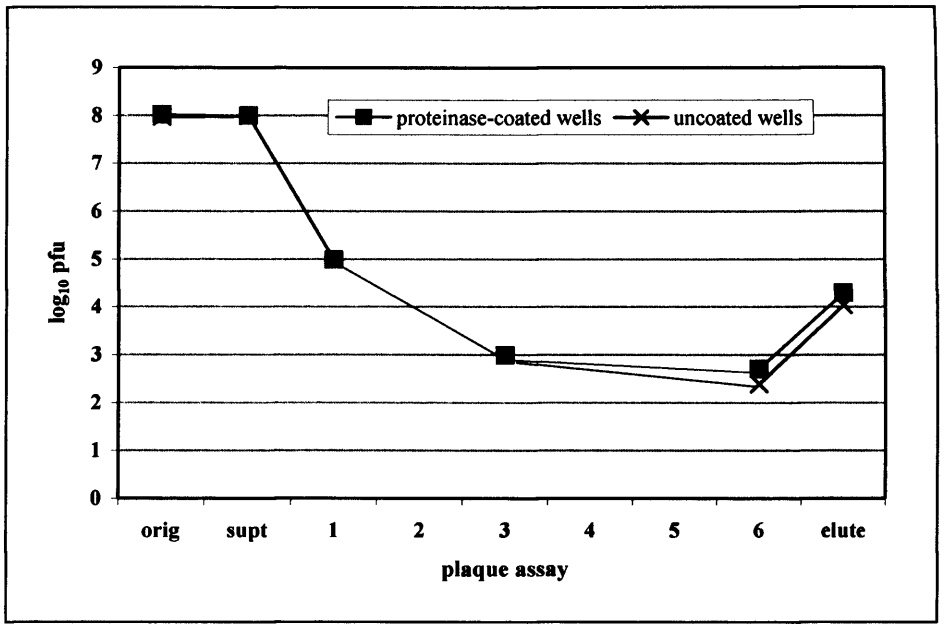
on the capsid surface on the viability of the *ϕphage*, the titre and time to induce lysis of infected bacteria was compared with that of the kit-supplied positive control *+phage* using a plaque assay (Section 2.2.13). Both phage strains produced bacterial cell lysis at 37 °C within 1.5-3 h of inoculation of a culture, and returned yields of 10^{10}-10^{11} pfu ml^{-1}, confirming that the introduced sequence had not detectably compromised the lytic activity of the T7Select™ phage.

10.5.2 Optimising biopanning using mictrotitre plate wells coated with *S. cerevisiae* vacuolar proteinase A

Immobilisation of the target enzyme (or selection tool) on a microtitre plate is a well-established "biopanning" protocol for the selection of phage displaying specific peptides of interest. As the natural target of the inhibitory sequence displayed by *ϕphage*, *S. cerevisiae* vacuolar proteinase A was chosen as the immobilised selection tool. Conditions were optimised to maintain the activity and integrity of the enzyme during both the immobilisation process and throughout the rigorous washing steps required to remove non-specifically bound phage particles in a biopanning procedure. Enzyme activity was determined in fluorometric assays which monitored the cleavage of a quenched fluorescent substrate (Section 2.2.16). The conditions described in Section 2.2.16 were the most favourable discovered in terms of binding/immobilisation buffer pH and ionic strength, duration and temperature of binding incubation, and composition of the washing buffer.

Comparative biopanning tests were conducted with either target proteinase-coated or uncoated microtitre plate wells (Section 2.2.17). After multiple washing steps (at least 6), similar bacteriophage titres (see Figure 10.3A) were eluted from both proteinase coated and uncoated wells by the addition of the alkaline denaturation elution buffer (100 mM CAPS, pH 9.9). It was also found that near-identical values were recorded if *+phage* displaying the S•Tag™ were used in place of the *ϕphage*. These results indicated that the T7Select™ phage of either strain were sticking non-specifically. Blocking the proteinase-coated or uncoated wells with lima bean trypsin inhibitor, rabbit IgG or killed *+phage* (Section 2.2.17) prior to biopanning did not reduce this non-specific binding. For example, Figure 10.3 B illustrates the data for

A



B

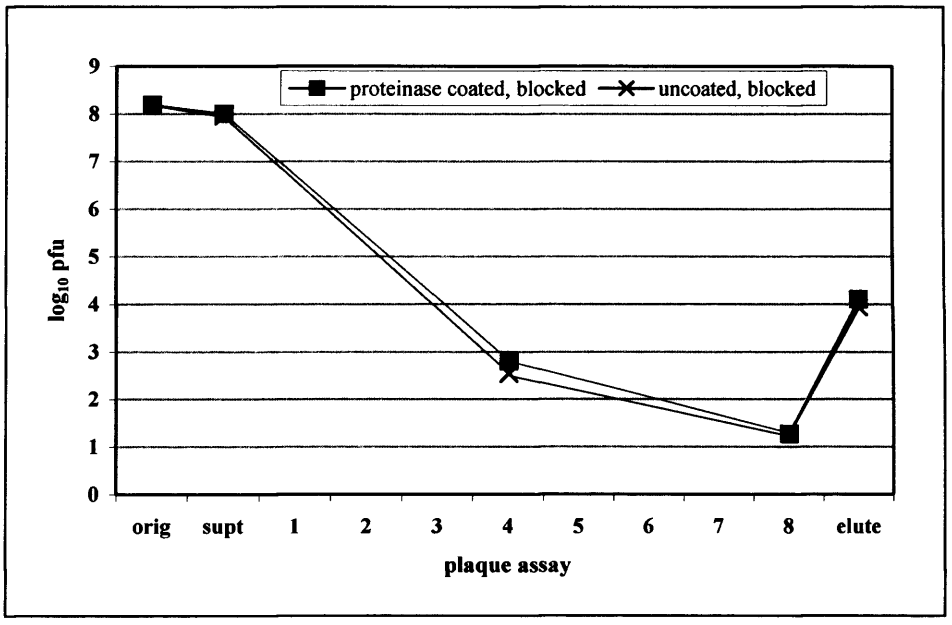


Figure 10.3 Binding of *ϕphage* to microtitre plate wells. A. Wells were either coated (red) or uncoated (blue) with *S. cerevisiae* proteinase A (Section 2.2.16), before the addition of a titred concentration of *ϕphage* (orig). The titre of phage aspirated from the well after this "panning" step was measured by plaque assay (supt). Selected subsequent washes (1-6) were similarly titred for phage. Following introduction of elution solution (in this instance 100 mM CAPS pH 9.9), the titre of phage thus eluted was determined (elute). **B. Blocking microtitre plate wells with killed *+phage*.** The experiment was identical to that of **A**, except that wells were blocked with killed control phage (*+phage*) before the addition of *ϕphage* to be "panned". In this example, the number of wash steps was increased to 8.

blocking with killed +*phage*. Attempts to elute *ϕphage* specifically from proteinase-coated wells in a more selective manner using an excess of a free competitive inhibitor, either isovaleryl pepstatin (Section 1.7.1) or the synthetic peptide of the same sequence as that displayed on *ϕphage* (peptide [redacted]), did not circumvent the non-specific binding problem either. Therefore, an alternative method of biopanning was sought.

10.5.3 Biopanning using *S. cerevisiae* vacuolar proteinase A immobilised on ConA-Sepharose

Concanavalin A (ConA) is a lectin which binds reversibly to molecules containing α-D-mannopyranosyl or α-D-glucopyranosyl groups, such as those found on the glycoprotein *S. cerevisiae* vacuolar proteinase A. ConA coupled to Sepharose beads is used widely for purification of glycoproteins, although no reports of the use of this matrix for bacteriophage biopanning were found in literature searches.

Prior to undertaking biopanning, preliminary experiments established conditions under which the *S. cerevisiae* proteinase A was bound by ConA-Sepharose. After incubation (pH 4.1 for 5 min at 4 °C) with an excess of the matrix followed by centrifugation, all enzyme activity was removed from the supernatant, as determined by a normal spectrophotometric assay (Section 2.2.29). All of the activity was recovered from the matrix, back into the supernatant, upon the addition of an excess of α-D-methylmannoside (0.5 M) to compete the bound enzyme from the ConA-Sepharose. Also, proteinase A was shown (by analytical FPLC, Section 2.2.27) to hydrolyse the usual chromogenic substrate (KPIEF*F'RL, Section 2.2.29, Section 9.2.1) at similar rates when free in solution and when bound to ConA-Sepharose. Furthermore, no digestion of free ConA was detected in a 16 h incubation at pH 6.25 and 25 °C with *S. cerevisiae* proteinase A at a 40:1 molar ratio (Figure 10.4A). Therefore, *S. cerevisiae* proteinase A formed a functional, stable, reversible, non-covalent association with the ConA-Sepharose, and was not able to catalyse its own release from the immobilisation matrix. These are essential requirements for a successful biopanning procedure.

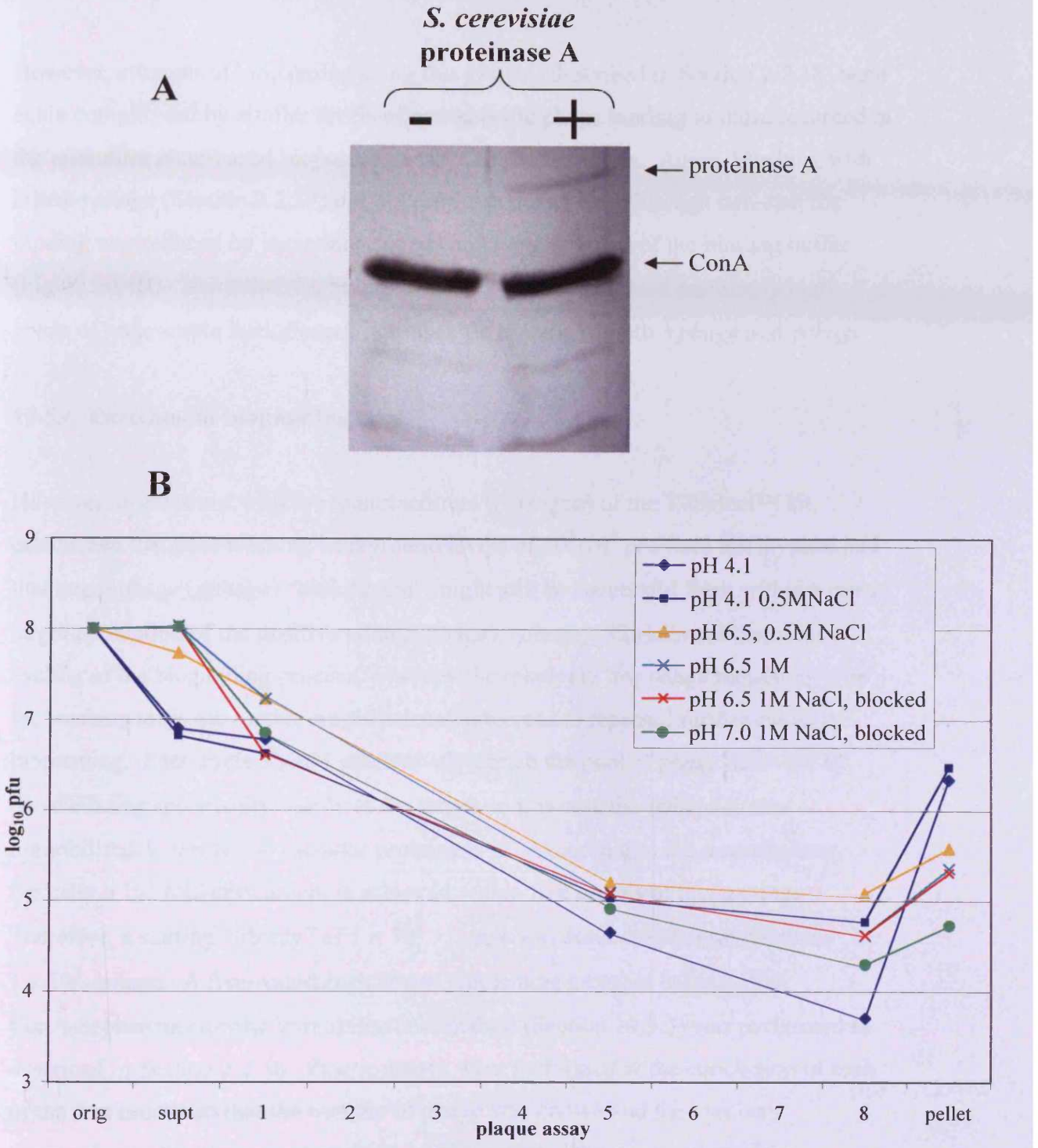


Figure 10.4 A Resistance of concanavalin A (ConA) to digestion by *S. cerevisiae* proteinase A. ConA was incubated for 16 h at 25 °C in 0.1 M MES buffer, pH 6.25, 1M NaCl, with (+) or without (-) *S. cerevisiae* vacuolar proteinase A, prior to analysis by SDS-PAGE (20% w/v acrylamide). The molar ratio of ConA:proteinase A = 40:1. **B. The effect of pH and salt on the binding of *φphage* to proteinase A in the ConA-Sepharose-based biopanning system.** A known titre of *φphage* (orig) was incubated with *S. cerevisiae* proteinase A (Section 2.2.18). ConA-Sepharose, equilibrated with the binding buffers indicated, was incubated with this phage-proteinase A mixture (Section 2.2.18). ConA-Sepharose was separated by centrifugation and the residual phage titre in the supernatant was determined (supt). The phage titres of selected subsequent wash steps (1-8) were measured. The phage titre remaining bound to the ConA-Sepharose-proteinase A matrix after the washes (pellet) was assayed. Blocking, when performed (red and green lines), was by incubation of ConA-Sepharose with killed control phage (+*phage*) before the panning step.

However, attempts at biopanning using this system (described in Section 2.2.18) were again complicated by similar levels of non-specific phage binding to those recorded in the microtitre plate-based biopanning assays described above. Again, blocking with killed *+phage* (Section 2.2.17) did not diminish this effect, although non-specific binding was reduced by increasing the pH and ionic strength of the binding buffer (Figure 10.4B). Nevertheless, both biopanning methods yielded seemingly high levels of undesirable background, non-specific binding of both *+phage* and *ϕphage.*

10.5.4 Enrichment biopanning trials

However, discussions with the manufacturers (Novagen) of the T7Select™ kit, established that post-washing background levels of 10^2-10^5 pfu were not atypical and that target phage (*ϕphage*) "enrichment" might still be successful from within a much larger population of the positive control phage (*+phage*). Enrichment describes cycling of the biopanning process, whereby the relatively few phage remaining after the washing steps are eluted, amplified and subjected to repeated further cycles of biopanning. Each cycle should successively enrich the pool of phage in favour of those binding specifically/tightly to the target, in this case the *ϕphage* to the immobilised *S. cerevisiae* vacuolar proteinase A. According to the manufacturer, typically a 10^6-fold enrichment is achieved within five rounds of biopanning. Therefore, a starting "library" of 1×10^8 *+phage* was seeded with approximately 1×10^2 *ϕphage*. A five-round enrichment biopanning protocol based on the ConA-Sepharose enzyme immobilisation method (Section 10.5.3) was performed as described in Section 2.2.18. Plaque assays were performed at the conclusion of each of the five rounds so that the number of phage was known and the numbers progressing to the next round could be kept constant. The progression of the enrichment was assessed by PCR screen (Section 2.2.5) using the PhageSeqF and PhageSeqR primers (Appendix II.E) that anneal upstream and downstream of the cloning region of the T7Select™ bacteriophage genome. The *+phage*, which contain the smaller S•Tag™ insert, could thus be distinguished from the *ϕphage* on the basis of PCR fragment size, as determined by submarine agarose (2 % w/v) gel electrophoresis (Section 2.2.3). In the first three rounds, only control phage (*+phage*) were apparent (Figure 10.5A). However, at the end of round four, 2 out of 12 (16%)

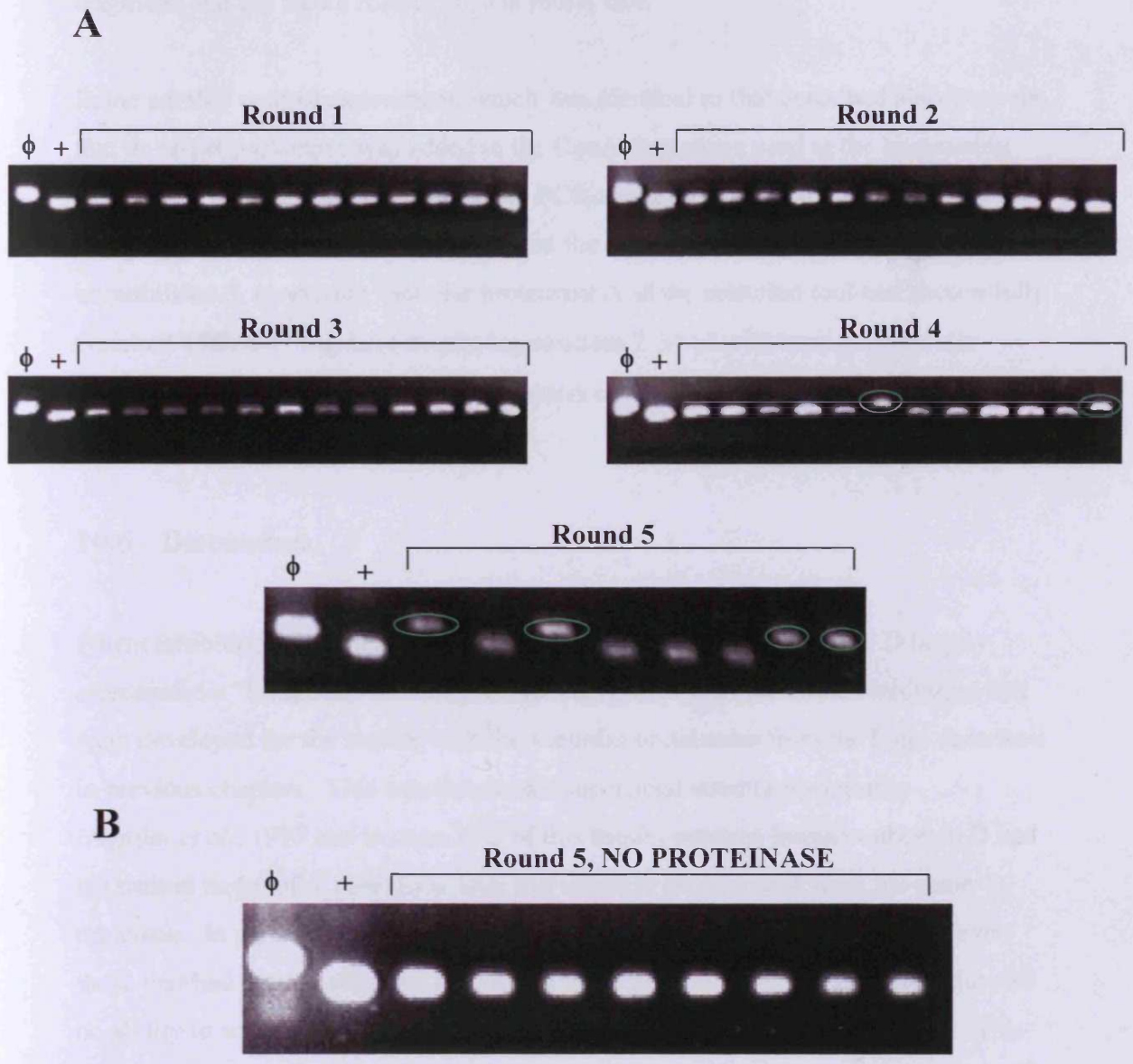


Figure10.5 Analysis of enrichment biopanning for *ϕphage*. A population of 10^8 *+phage* (displaying S•Tag™) was mixed with approximately 10^2 *ϕphage* (displaying residues 2-34 of *S. cerevisiae* IA3). This "library" was subjected to 5 rounds of enrichment biopanning, using ConA-Sepharose as the matrix (Section 2.2.18), in the presence (panel **A**) or absence (panel **B**) of the selection tool, *S. cerevisiae* proteinase A. After each round of enrichment, phage titre was assayed and 8 or 12 of the resulting plaques were screened by PCR. The PCR products were analysed by submarine agarose (2% w/v) gel electrophoresis (Section 2.2.3). PCR products positive for the "ϕ" insert (and therefore displaying the IA3 sequence) are circled in green. "ϕ" = marker for PCR product produced from *ϕphage* containing the IA3 insert; + = marker for product of PCR with *+phage* control containing S•Tag™.

screened plaques were positive for the *ϕphage* displaying the 2-34 IA3 peptide fragment, and the figure rose to 50% in round five.

In the parallel control experiment, which was identical to that described above except that no target proteinase was added to the ConA-Sepharose used in the biopanning steps, only *+phage* were revealed in the PCR screen at the conclusion of round 5 (See Figure 10.5B). Therefore, it appeared that the 5 rounds of biopanning using immobilised *S. cerevisiae* vacuolar proteinase A as the selection tool had successfully enriched T7Select™ *ϕphage* displaying residues 2-34 of wild-type *S. cerevisiae* vacuolar proteinase A from a 10^6-fold excess of control-insert *+phage*.

10.6 Discussion

Potent inhibition of the human lysosomal aspartic proteinase cathepsin D largely represented a "bridge too far" for the existing repertoire of IA3-based inhibitors that were developed for the studies with the vacuolar proteinases from the fungi described in previous chapters. This was despite the superficial structural similarity (Aguilar *et al.*, 1997 and Section 10.2 of this thesis) between human cathepsin D and the natural target of *S. cerevisiae* IA3, the vacuolar proteinase A from the same organism. In particular, the majority of *S. cerevisiae* IA3-derived sequences, even those that had proven effective against the more intractable fungal enzymes, showed no ability to so much as reduce the rate of hydrolysis of the reporter substrate. This was true even when the inhibitors were present at concentrations approximately $1/6^{th}$ that of the substrate, where substrate competition might have been observed. Therefore, although prolonged incubation with cathepsin D leads to its cleavage (Phylip *et al.*, 2001), it would appear that *S. cerevisiae* IA3 does not bind particularly well as a substrate either, in addition to not being an inhibitor. However, it perhaps re-enforces the original proposition of Phylip *et al.*, (2001) that the *S. cerevisiae* IA3 sequence is a highly specific inhibitor, tuned for an exclusive interaction with its own natural target.

In contrast, although generally not effective as inhibitors in the context of these studies, most of the *S. castellii* IA3-derived sequences tested do at least demonstrate a concentration-dependent interaction with cathepsin D at pH 4.7. Most notably, the synthetic peptide representing residues 2-34 of wild-type *S. castellii* IA3 (**44** in Table 10.2) is an effective inhibitor (K_i = 15 ± 3 nM). Although prolonged incubation with a high concentration (10:1 molar ratio for 18 h) of cathepsin D did result in a slow hydrolysis, notably this cleavage did not occur at locations equivalent to those described previously for the attack of cathepsin D on wild-type IA3 from *S. cerevisiae*, i.e. between residues 10 and 11, and 29 and 30 (Phylip *et al.*, 2001). Perhaps significantly, the residue pairs constituting these two peptide bonds are different in *S. castellii* IA3, being Glu-Met (instead of Glu-Ile) and Gly-Met (instead of Ala-Phe), respectively. The *S. castellii* IA3 combinations at these two locations may be refractory to cleavage. However, the fact that *S. castellii* IA3 is still cleaved, albeit slowly and at different positions, indicates that removal of one or more cleavage locations merely directs digestion by the non-target proteinase to the next accessible peptide bond meeting the sub-site preferences of the enzyme. This is reminiscent of the observations made with regard to *S. cerevisiae* IA3 by Phylip *et al.* (2001).

Although only one IA3 sequence displays effective inhibitory activity, the very fact that one *does* is highly significant. This marked contrast between one sequence and all the others is remarkable and informative in itself, and merits further comment.

Conclusions must, of necessity, be predicated on the assumption that the 2-34 *S. castellii* IA3 peptide (**44**) inhibits human cathepsin D by the same mechanism proposed for the fungal enzymes described previously. Certainly, the potency and the location of the activity within residues 2-34 of the IA3 sequence are highly suggestive that the mechanism may well be the same. With this caveat, therefore, several significant conclusions may be inferred.

(i) No *S. cerevisiae* IA3-derived sequences show detectable activity against cathepsin D, whereas numerous *S. castellii* IA3 sequences do interact. Furthermore, the introduction of the *S. castellii* IA3 centrepiece, back pin or even both sets of residues at once is insufficient to bestow activity on

S. cerevisiae IA3. This suggests that for future investigations, *S. castellii* IA3 may form a more suitable starting template.

(ii) One notable facet to the activity of the *S. castellii* IA3 is that the most potent form is the wild-type inhibitory sequence. None of the mutations, with modifications to either the centrepiece or the back pin regions, improved the inhibitory potency of **44**. In particular, the replacement of Lys22 with Leu was significantly disadvantageous. This is a contrast with the role of this position in *S. castellii* IA3 against the fungal enzymes tested, where this replacement had appeared to help overcome the apparently detrimental effect of the *S. castellii* IA3 tail.

(iii) The attachment of a tail from either *S. castellii* or *S. cerevisiae* IA3 destroys the potency of the interaction of the *S. castellii* 2-34 inhibitory sequence with the enzyme. This is a situation also seen with the *A. fumigatus* proteinase (Chapter 9). However, the finding with cathepsin D further adds to the body of evidence that the C-terminal tail is not a completely neutral, inert or ineffectual partner in the inhibitory operation of IA3.

(iv) His-tag removal may be moderately beneficial to the *S. castellii* IA3 recombinant proteins, although the potencies remain so poor as to make interpretation difficult. Nevertheless the improvement in activity, particularly noticeable with the N*cast*-C*cere* chimaera (**74**) and the Lys45 truncated form of *S. castellii* IA3 (**76**), represent a significant contrast to the situation with the fungal enzymes where the absence of a His-tag had a universally detrimental effect on inhibition. This again points to the His-tag having a modulating role in the interaction of an IA3 with a target enzyme, although the severity of this effect varies widely from one enzyme to another.

Despite these conclusions and even with the knowledge derived from several detailed structures available for cathepsin D, significant further progression of this project by conventional rational design approaches would prove difficult. Most probably, it will require detailed structural information of an IA3-based inhibitor (such as peptide **44**) bound in the cathepsin D active site cleft. This would reveal the precise mode of inhibition and may offer insight for future improvements to the inhibitory sequence.

However, without any of the available sequences generating potencies in the very low nanomolar or subnanomolar K_i range, it is not feasible at this time to produce sufficiently stable complexes of the human cathepsin D with an IA3-derived inhibitor to permit structural studies by X-ray crystallography. This highlights the limitations of rational design. In contrast, approaches based on directed evolution offer the opportunity to circumvent this "Catch22" requirement for fresh structural data and may, ultimately, provide improved, retargeted inhibitors that facilitate the formation of stable complexes amenable to structural analysis.

The preliminary experiments using bacteriophage display were a qualified success and gave an insight into the potential of this technology. The two methods of biopanning, which use different methods of target enzyme immobilisation, are not just applicable to *S. cerevisiae* proteinase A but can be readily extended to other enzymes in the aspartic proteinase family. Indeed, although not described in detail in this thesis, the conditions for immobilisation of target enzymes other than *S. cerevisiae* proteinase A, namely the vacuolar proteinase from *P. pastoris* and human pepsin, have also been optimised for the microtitre plate panning system. This is in preparation for biopanning of the planned library of bacteriophage displaying variant sequences derived from residues 2-34 of IA3.

However, one problem in particular may hamper future efforts based on the current system. High levels of non-specific binding were observed with both biopanning methods, even after various blocking methods and more stringent washing procedures were applied. Eluting bacteriophage specifically using an excess of a free, highly selective, competing inhibitor does not eliminate the problem either. These non-specific binding problems, although not widely reported in the literature, appear from informal contacts with the T7Select™ manufacturer to be inherent in phage-display systems. Therefore, the key event will be to reduce the non-specific binding, possibly by the introduction of one or more panning steps that remove those bacteriophage displaying sequences that are particularly prone to non-specific binding. Preliminary experiments indicate that as a corollary this may significantly reduce the size of the library. Therefore, improved methods of reducing non-specific binding *within* the biopanning protocol will also be needed. In this regard, trialling one of the more well-established systems based on filamentous bacteriophage could

be advantageous. However, for the specific case of panning for aspartic proteinase inhibitors, there may be several potential pitfalls to these systems as well, as noted in Section 10.5.

If the balance of advantages and disadvantages indicates that the T7Select™ system has the greater potential, an alternative possibility may be to reduce the number of copies of the inhibitory sequence displayed on each phage particle from 415 to, for instance, 10. Whilst this may reduce the sensitivity of the biopanning procedure, it is also possible that tethering 415 identical copies of the same 33 amino acid sequence, which has the ability to form amphipathic α-helices, to the outside of a phage particle may be too high a density and thus could interfere with effective biopanning.

Similarly, reducing any steric hindrances to enzyme-inhibitor interaction posed by the proximity of the phage particle may also prove beneficial. In this regard, one attractive possibility for tethering the displayed sequence further from the surface of the bacteriophage might be to use an extended linker region containing two proline residues separated by glycines and alanines. This may create a flexible "angle-poise" swinging-arm, like that found in the lipoyl domains of multienzyme complexes such as pyruvate dehydrogenase (Milne *et al.*, 2002), although in our context perhaps a fishing rod with bait would be a more appropriate analogy (!). However, flexibility must be accompanied by resistance to proteolysis; otherwise a longer linker may be susceptible to cleavage by the immobilised enzyme that serves as the selection tool.

Despite these concerns, the potential of the combinatorial phage display approach for the continuation of this research has been demonstrated. The T7Select™ phage displaying residues 2-34 of wild-type *S. cerevisiae* IA3 can be enriched by approximately 10^6-fold within five rounds of biopanning from within a background of bacteriophage expressing S•Tag™. This suggests strongly that the system is sound and that, with the refinements suggested above, production and biopanning of a preliminary T7Select™ bacteriophage library can commence. Until fresh structural information becomes available from the X-ray crystallographic preparations described in preceding chapters, bacteriophage display represents a promising route of progression towards the production of IA3-based inhibitors of new aspartic proteinase

targets, such as human cathepsin D. It may also, paradoxically, provide the avenue towards the eventual generation of inhibitor-protein complexes for crystallography, which would in turn inform the rational design approach.

No	name	sequence	pH 3.1	pH 4.7
1	CT$(H)_6$ WT IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
2	NT$(H)_6$ WT IA3	MK$(H)_3$MQ NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	×	×
3	Tagless WT IA3	*QNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	×	×
20	D22M	MNTDQ QKVSE IFQSS KEKLQ GMAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
25	K18I/D22I	MNTDQ QKVSE IFQSS KEILQ GIAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
26	K18I/L19M/D22I	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
29	K18I/L19Z/D22I 2-32 IA3	NTDQ QKVSE IFQSS KEIZQ GIAKV VSDAF KK	×	×
30	K18Z/D22L 2-32 IA3	NTDQ QKVSE IFQSS KEZLQ GLAKV VSDAF KK	×	✓
	TJW14 (E17I/K18J/L19Z/D22I) 2-32	NTDQ QKVSE IFQSS KIJZQ GIAKV VSDAF KK	×	×
31	K18R	MNTDQ QKVSE IFQSS KERLQ GDAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
32	D22E	MNTDQ QKVSE IFQSS KEKLQ GEAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
33	K18R/D22E	MNTDQ QKVSE IFQSS KERLQ GEAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
36	K18D/D22K	MNTDQ QKVSE IFQSS KEDLQ GKAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
37	K18E/D22K	MNTDQ QKVSE IFQSS KEELQ GKAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
38	K18V/D22K	MNTDQ QKVSE IFQSS KEVLQ GKAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
39	K18M/D22K	MNTDQ QKVSE IFQSS KEMLQ GKAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
40	K18D/D22R	MNTDQ QKVSE IFQSS KEDLQ GRAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
41	K18E/D22R	MNTDQ QKVSE IFQSS KEELQ GRAKV VSDAF KKMA~~~~~~~~LE$(H)_6$	×	×
45	N*cere*-C*cast* IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~~~~~LE$(H)_6$	×	×
46	N*cere*TM-C*cast* IA3	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~~~~~~~~LE$(H)_6$	×	×
52	V26A/A29G/F30M IA3	MNTDQ QKVSE IFQSS KEKLQ GDAKV ASDGM KKMA~~~~~~~~LE$(H)_6$	×	×
53	Q20E/V26A/A29G/F30M IA3	MNTDQ QKVSE IFQSS KEKLE GDAKV ASDGM KKMA~~~~~~~~LE$(H)_6$	×	×
56	N*cast Sac*I/*Nhe*I IA3	MNTDQ QKVSE IFQSS KEMLE GKANA ASEGM KNMA~~~~~~~~LE$(H)_6$	×	×
62	*S. cerevisiae* K18D/D22K 2-32	NTDQ QKVSE IFQSS KEDLQ GKAKV VSDAF KK	×	×
80	*S. cerevisiae* K18Z/D22K 2-32	NTDQ QKVSE IFQSS KEZLQ GKAKV VSDAF KK	×	×

Table 10.1 Lack of effect of *S.cerevisiae* IA3-derived sequences (yellow) against cathepsin D. Where mutations introduced sequence derived from *S. castellii* IA3, the residues are in blue. Other mutations are in red. × = no apparent interaction whatsoever (reflected by the red line on Figure 10.1, which is that of **39** at pH 3.1); ✓ = weak dose-dependent response (more closely resembling the blue line on Figure 10.1).

No	name	sequence	pH 3.1	pH 4.7
42	CT(H)$_6$ WT *S. castellii* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	×	✓ 800 ± 100
43	NT(H)$_6$ WT *S. castellii* IA3	MK(H)$_6$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	×	✓ < 1000
44	WT *S. castellii* IA3 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	✓	✓ 15 ± 3
47	N*cast*-C*cere* IA3	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	×	✓ 850 ± 350
48	*S. castellii* M18V IA3	MSDKN ANVSE MFQQA KEVLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	×	×
49	N*cast* M18V-C*cere* IA3	MSDKN ANVSE MFQQA KEVLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	×	×
57	*S. castellii* G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANA ASEAF KNZA	✓	✓ 130 ± 20
58	*S. castellii* A25V/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEAF KNZA	×	✓ 360 ± 70
59	*S. castellii* A25V 2-34	SDKN ANVSE ZFQQA KEZLE GKANV ASEGZ KNZA	×	✓ 270 ± 50
60	*S. castellii* A25I/G29A/M30F 2-34	SDKN ANVSE ZFQQA KEZLE GKANI ASEAF KNZA	×	✓ 250 ± 40
61	SEX 4	SDKN ANVSE ZFQQA KEZLE GKAKV VSDAF KKZA	×	✓
63	*S. castellii* M18K/K22M IA3	MSDKN ANVSE MFQQA KEKLE GMANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	×	×
64	*S. castellii* M18K/K22Z 2-32	SDKN ANVSE ZFQQA KEKLE GZANA ASEGZ KN	×	✓
65	*S. castellii* M18K/K22L IA3	MSDKN ANVSE MFQQA KEKLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	×	×
66	*S. castellii* M18K/K22D IA3	MSDKN ANVSE MFQQA KEKLE GDANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	×	×
67	*S. castellii* M18K/K22D 2-32	SDKN ANVSE ZFQQA KEKLE GDANA ASEGZ KN	×	×
68	*S. castellii* K22L IA3	MSDKN ANVSE MFQQA KEMLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	×	✓
69	*S. castellii* K22L 2-32	SDKN ANVSE ZFQQA KEZLE GLANA ASEGZ KN	✓	✓
50	NT(H)$_6$ *S. castellii* Q68 IA3	MK(H)$_6$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~[68]	×	✓
70	NT(H)$_6$ *S. castellii* Y57 IA3	MK(H)$_6$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	×	✓
71	NT(H)$_6$ *S. castellii* K45 IA3	MK(H)$_6$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	×	✓ 750 ± 100
51	NT(H)$_6$ N*cast*-C*cere* IA3	MK(H)$_6$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	✓	✓ 650 ± 70
72	NT(H)$_6$ N*cast*-C*cere* Y57 IA3	MK(H)$_6$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	×	✓ 850 ± 350
73	NT(H)$_6$ N*cast*-C*cere* A45 IA3	MK(H)$_6$MQSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	×	✓
74	Tagless N*cast*-C*cere* IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	✓	✓ 280 ± 25
75	Tagless WT *S. castellii* IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	×	✓ 530 ± 50
76	Tagless *S. castellii* K45 IA3	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	✓	✓ 300 ± 40

Table 10.2 Effect of *S. castellii* IA3-derived sequences (blue) against cathepsin D. Where mutations introduced sequence derived from *S. cerevisiae* IA3, the residues are in yellow. Other mutations are in red. × = no apparent interaction whatsoever (resembling the red line on Figure 10.1); ✓ = dose-dependent response more closely resembling peptide **44** (the blue line on Figure 10.1). Where this concentration-dependent inhibition produced a K_i<1000 nM, the value is included in the table.

Chapter 11

General Discussion

Investigations of the sequence elements that contribute to the remarkable potency, specificity and mode of inhibition demonstrated by IA3 have generated a wealth of new information, which has advanced our understanding of the intricate interplay between inhibitor and enzyme. Although, most of the results presented in this thesis have been discussed in the relevant experimental chapter, certain overarching themes were identified to unify this work. This general discussion will examine these themes, and place them in a broader context.

Furthermore, although the aims of the project (Section 1.9) were all met, to use a baseball analogy, the "bases are still loaded" in this project. That is to say, further large gains in information are anticipated from this work, such as the X-ray crystallography of key inhibitory complexes and NMR studies of IA3 forms in solution. Therefore, this chapter will also discuss the potential avenues of research that could maximise the future yield of results.

11.1 Front and back hydrophobic pins

The residues of *S. cerevisiae* proteinase A that directly contact the hydrophobic front pin region of *S. cerevisiae* IA3 in the complex (Li *et al.*, 2000, and see Figure 1.12) are almost entirely conserved amongst the enzymes studied here. Even the more distantly related *A. fumigatus* proteinase and cathepsin D contain only conservative substitutions:

	Residue					
	9	**10**	**12**	**276**	**278**	**283**
S. cerevisiae	Tyr	Leu	Ala	Leu	Val	Ile
S. castellii	Tyr	Leu	Ala	Leu	Val	Ile
P. pastoris	Tyr	Leu	Ala	Leu	Val	Ile
A. fumigatus	**Phe**	Leu	Ala	Leu	Val	Ile
Cathepsin D	Tyr	**Met**	Ala	Leu	Val	**Leu**

This suggests that whilst the front pin may contribute significantly to the potency of the interaction with all these enzymes, the potential gains in specificity from this region may be small. A principal focus of this thesis, however, was on the

re-direction of IA3 activity towards new targets. Therefore, most investigations of the hydrophobic interactions between the inhibitory sequence and the various enzymes concentrated on the back pin and its interaction with the considerably more variable corresponding regions of the target enzymes, such as the polyproline loop (see earlier end of chapter discussions).

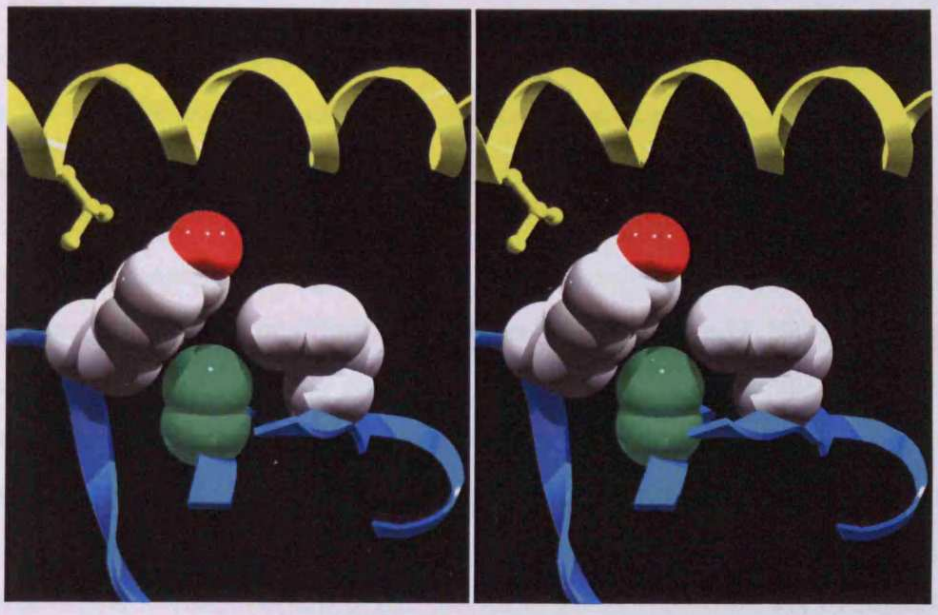
These studies substantiate that the back pin is indeed integral to both the potency and specificity of the IA3 inhibitory activity. Notably, however, against *S. cerevisiae* vacuolar proteinase A, the benefit of the *S. cerevisiae* IA3 back pin over that of *S. castellii* IA3 was remarkably small. Indeed, one of the *S. castellii* IA3 residues – Ala26 (instead of Val in *S. cerevisiae* IA3) – was even modestly beneficial (Chapter 6). Generally, though, the main requirement in this back pin region against *S. cerevisiae* proteinase A was for *compatibility* between residues on adjacent turns of the inhibitory helix, i.e. in (*i*, *i*+4) proximity. For optimum activity, the residues constituting the 25/29 and 26/30 "pairs" had to be either *S. cerevisiae* IA3-derived or from *S. castellii* IA3, but <u>not</u> a combination of the two. However, the 25/29 and 26/30 pairs themselves could have distinct origins. For example, a molecule containing 25/29 (*S. cerevisiae* IA3) and 26/30 (*S. castellii* IA3) would be predicted to perform at least as well as the wild type form, whilst an IA3 with 25/**30** (*S. cerevisiae*) and 26/**29** (*S. castellii*), or vice versa, would not.

In contrast, against the proteinase from *P. pastoris*, the *S. castellii* IA3 back pin residues were superior to those of *S. cerevisiae IA3*. The introduction of the *S. castellii* IA3 back pin into *S. cerevisiae* IA3 improved potency, whereas the placement of *S. cerevisiae* IA3 back pin residues into *S. castellii* IA3 gave a poorer inhibitor (Chapter 8). This possibly reflected subtle but important differences in the active site cleft of the *P. pastoris* proteinase, such as the residues constituting the polyproline loop (Sections 1.8.5 and 8.8).

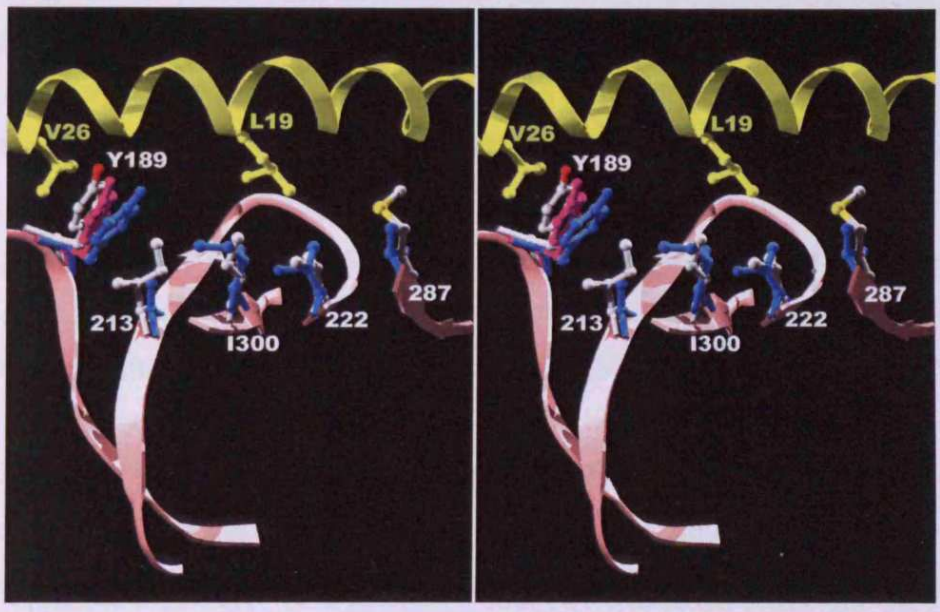
However, adjustments to the back pin region affected inhibition of the *A. fumigatus* proteinase and cathepsin D more profoundly. Against the *Aspergillus* enzyme, the compatibility of residues on adjacent turns of the inhibitory helix was again vitally important. Introduction of the *S. cerevisiae* IA3 residues at either position 25 or 29/30 into the *S. castellii* IA3 inhibitory sequence was extremely detrimental, but the

potency of the mutant containing all three changes simultaneously was barely distinguishable from the wild type inhibitory sequence. The same could not be said of the inhibition of cathepsin D. This enzyme was highly sensitive to *any* change to the sequence of *S. castellii* IA3. This may reflect the more substantial differences between cathepsin D and the other enzymes with regard to the residues in contact with the back pin of IA3. However, against both the *Aspergillus* proteinase and cathepsin D, the most critical position in the back pin region of IA3 was again the residue at 26. An Ala at position 26 was much more advantageous against these two enzymes than *S. cerevisiae* proteinase A (see above). In Chapter 6 (Section 6.5) it was proposed that the replacement of Val26 with Ala in *S. cerevisiae* IA3 reduced a mild steric clash between the residue at this position of the inhibitory α-helix and Tyr189 of the enzyme. Tyr189, which is conserved amongst all five enzymes considered in detail in this thesis, is "held" in position in *S. cerevisiae* proteinase A by Ala213 (green) and Ile300 (Figure 11.1A). These three residues make a particularly snug fit. Whilst Ile300 is also conserved amongst the five proteinases, the "sandwiched" residue at position 213 is not. In many other aspartic proteinases, including the *Aspergillus* enzyme and cathepsin D, the Ala213 in the vacuolar proteinases of *S. cerevisiae*, *S. castellii* and *P. pastoris* is replaced by a much larger Ile. Inspection of the cathepsin D structure (Baldwin *et al.*, 1993) reveals that, rather like an over-filled sandwich, Ile213 pushes Tyr189 back out into the active site cleft. This relative displacement of Tyr189 in cathepsin D is illustrated in Figure 11.1B, which shows with the relevant residues of *S. cerevisiae* proteinase A in both IA3-bound (proteinase sidechains in blue, inhibitory helix in gold) and IA3-unbound (purple) forms superimposed on the structure of cathepsin D (CPK sidechains). Thus, it can be seen that in cathepsin D (and probably the *Aspergillus* enzyme that also contains Ile213), Tyr189 is displaced beyond the "default" location adopted by the corresponding residue in the IA3-unbound form of *S. cerevisiae* proteinase A (purple). Certainly, it is a considerable distance from the location adopted in *S. cerevisiae* proteinase A upon IA3 binding (blue).

A



B



C

Residue	*S. cerevisiae*	*S. castellii*	*P. pastoris*	*A. fumigatus*	Cathepsin D
189	Tyr	Tyr	Tyr	Tyr	Tyr
213	Ala	Ala	Ala	**Ile**	**Ile**
222	Thr	Thr	**Ala**	**Ala**	**Val**
287	Thr	**Met**	Thr	**Met**	**Met**
300	Ile	Ile	Ile	Ile	Ile

Figure 11.1 The proposed "domino effect" in *S. cerevisiae* proteinase A and cathepsin D. **A.** Stereo representation of Val26 of IA3 (yellow) and (left to right) Tyr189 (CPK), Ala213 (green) and Ile300 (CPK) of *S. cerevisiae* proteinase A (PDB code 1DPJ). **B.** Stereo view of a superimposition of selected residues from *S cerevisiae* proteinase A in both IA3-bound (blue, PDB code 1DPJ) and IA3-unbound (pink, PDB code 2JXR) conformations onto the structure of cathepsin D (CPK, PDB code 1LYA). **C.** Differences between the proteinases studied in this thesis at the residues involved in the domino effect.

Therefore, this hypothesis may explain the strong preference exhibited by the *Aspergillus* enzyme and cathepsin D for Ala over Val at position 26 of *S. castellii* IA3. The isopropyl sidechain of Val26 of IA3 may push Tyr189 of *S. cerevisiae* proteinase A out of its uninhibited position (compare the positions of the blue and purple tyrosines in Figure 11.1B). This is only possible because of the relatively small residue at position 213 (Ala, in blue in Figure 11.1B) in *S. cerevisiae* proteinase A onto which Tyr189 can collapse. However, in cathepsin D, and probably the *Aspergillus* proteinase, Tyr189 is held further out into the active site cleft as a result of the "overstuffed sandwich" containing Ile213 instead of Ala. This reduction in the size of the binding pocket would lead to a more severe clash with a Val at position 26 of IA3. In cathepsin D and the *Aspergillus* enzyme, Tyr189 cannot reduce this conflict by collapsing to the position it adopts in the *S. cerevisiae* proteinase A-IA3 complex (in blue), because the space is filled with Ile213 braced by Ile300. Thus, an Ala at 26, which is smaller than the Val found in wild-type *S. cerevisiae* IA3, would therefore reduce this clash with Tyr189 and be more favoured.

An inspection of Figure 11.1B reveals that any displacement of Tyr189 as a result of clashes with residue 26 of IA3 could have repercussions further along the active site cleft, and consequently might affect more central binding pockets, such as the S2 pocket utilised by Leu19 of IA3 (see also Figure 1.13). In the *Aspergillus* enzyme and cathepsin D, where Tyr189 has no space to move into unless both Ile213 and Ile300 also move in sympathy (Figure 11.1B), the implications could be more severe. In these enzymes, Ile213/Ile300 movement could interfere with position 222 and, possibly, 287, in a "domino effect". Notably, the residues at 222 and 287 of the *Aspergillus* enzyme and cathepsin D also differ from those in the other enzymes (see Figure 11.1C). In the *Aspergillus* proteinase, Ala222, replaces the Thr found in *S. cerevisiae* proteinase A. This smaller sidechain may provide a little more space for the "domino collapse" of Tyr189, Ile213 and Ile300, and therefore could explain why a select few *S. cerevisiae* IA3-derived inhibitors containing Val26 remained effective inhibitors of this enzyme. However, Val222 of cathepsin D would not be expected to be so accommodating of the demands indirectly placed upon it by the choice of IA3 residue 26. This may partially explain the more stringent requirement for Ala26 for the inhibition of cathepsin D. It may be intriguing to test the hypothesis by exchanging residues 213, 222 and 287 of the "domino cascade" so that one proteinase

is made to resemble another at these key positions, and then re-investigate with mutant IA3 forms to study any change in residue requirements.

11.2 Centrepiece residues

The identities of the "centrepiece" residues (positions 18 and 22) were consistently critical to the inhibitory potency of all forms of IA3. This is perhaps not surprising considering that these residues are in close proximity to, and interact with, the catalytic machinery of *S. cerevisiae* proteinase A. The evidence in this thesis indicates that similar interactions exist in the active sites of the other enzymes that are inhibited by forms of IA3.

The centrepiece is the only hydrophilic part of the otherwise hydrophobic face of the amphipathic *S. cerevisiae* IA3 inhibitory helix. Lys18 and Asp22 engage in a complex network of electrostatic handshakes with each other, the two catalytic aspartates, the catalytic water molecule and Tyr75 on the enzyme flap, all of which help to protect the inhibitor from cleavage (Section 1.8.6 and Figure 1.15). However, in view of the discovery that appropriate hydrophobic substitutions in this region yield significant improvements in inhibitory potency, it is interesting to consider why Nature has selected a pair of charged rather than hydrophobic residues for these positions in *S. cerevisiae* IA3.

The results presented in this thesis reveal that the undoubted benefits to potency resulting from such hydrophobic replacements are not without other costs. For instance, one of the most distinguishing features of the inhibitory activity of IA3 is the exclusivity of its inhibitory interaction with its natural target, which is in stark contrast to other naturally-occurring aspartic proteinase inhibitors like the pepstatins (Section 1.7.1). This would suggest that the selectivity exhibited by IA3 is probably by design rather than a fortunate or fortuitous coincidence. However, the hydrophobic centrepiece mutants, such as K18I/D22I, that are potent inhibitors of the natural target proteinase, are also very potent inhibitors of the *P. pastoris* and *A. fumigatus* enzymes. Thus, although hydrophobic centrepiece residues increase the

potency of *S. cerevisiae* IA3, this improvement comes at the cost of greatly reduced specificity. Moreover, in view of the already picomolar potency of the wild type sequence, any further improvement in potency might be of questionable advantage. Indeed, there may even be detrimental physiological consequences if the inhibition is *too* tight.

It was also noted that all *S. cerevisiae* IA3 sequences containing multiple hydrophobic centrepiece mutations accumulated poorly as both recombinant proteins in *E. coli* (Chapter 3) and as chemically synthesised peptides (Dr. J. Fox, Alta Bioscience Ltd., personal communication). This was a significant contrast to the other IA3 mutants, all of which accumulated well in both recombinant and synthetic peptide forms. It is possible that the sequence of IA3 has also been selected by yeast as an optimum for expression and yield, and that an increase in the hydrophobicity of IA3 could reduce yields of IA3 *in vivo* as well. Possibly related to this, Gustchina et al. (2002) reported difficulties in crystallising a complex of *S. cerevisiae* proteinase A with a mutant form of IA3 containing two hydrophobic substitutions (K18M/D22L). The inhibitor was absent from the solved structure, suggesting that it might have been cleaved by the proteinase.

The introduction of hydrophobic residues into the centrepiece of *S. cerevisiae* IA3 therefore appears to represent a "poisoned chalice", where the modest improvements in potency are more than outweighed by the loss of both the specificity for, and resistance to eventual cleavage by, its natural target proteinase. Although hydrophobic substitutions in the centrepiece of IA3 are attractive in certain circumstances within the narrow confines of our experimental systems *in vitro*, it seems that Nature has selected charged residues for the centrepiece of *S. cerevisiae* IA3 with a wider view.

It is less immediately apparent, however, why/how Nature has specifically selected Lys at 18 and Asp at 22 for *S. cerevisiae* IA3 rather than, for example, any of the other multiple combinations (Chapter 5) that were similarly effective against *S. cerevisiae* proteinase A at approximately "physiological" pH values. On the one hand, perhaps the *S. cerevisiae* IA3 sequence results from the fact that Nature does not necessarily select the *best* individual, but only one that is *good enough* to meet the

selection pressure. That is to say, perhaps Lys18 and Asp22 are good enough for the purpose, without necessarily being the best. Of course, in such a situation it must also be carefully considered what Nature was selecting for. A closer inspection of our data indicates that although several of the alternative centrepiece combinations tested were either more potent (for example K18R, K18R/D22E) *or* more specific (e.g. K18E/D22K, K18D/D22K), none were improved in *both* aspects. Therefore, *in vivo*, perhaps Lys18 and Asp22 strike just the right balance of potency without compromising on specificity for *S. cerevisiae* IA3.

In this regard, however, the *S. castellii* IA3 sequence is curious because it appears to be more intrinsically promiscuous than *S. cerevisiae* IA3. Once again, the centrepiece residues (this time Met18 and Lys22) play a critical role in the inhibitory activities displayed by *S. castellii* IA3. Furthermore, interchanging the centrepieces between the two IA3 species also transferred many of the inhibitory characteristics. However, significant distinctions also exist between the behaviour of the two centrepieces when contained within their respective wild-type sequences. For example, the introduction of the *S. castellii* centrepiece residues (Met18/Lys22) into *S. cerevisiae* IA3 results in a greatly accentuated pH-dependency of the inhibitory interaction with *S. cerevisiae* proteinase A (Section 5.6). Although this so-called "pH switch" phenomenon, which is reminiscent of pH-dependent activitation of zymogens, is also present in full-length recombinant *S. castellii* IA3, the origin in the two IA3 species appears to be distinct. The sole responsible factor in *S. cerevisiae* IA3 was the presence of a basic residue at 22 (Chapters 5 and 7). However, whilst this residue is a critical factor in the pH switch in *S. castellii* IA3 as well, the His-tag and possibly the C-terminal tail (i.e. residues after 34) of the molecule also contribute to the effect (Chapter 7). Other evidence, such as the preliminary NMR experiments and the discovery that *S. castellii* IA3 is more potent without its C-terminal half (discussed in Section 11.3), also leads to the conclusion that the mode of inhibition employed by *S. castellii* IA3, although closely related, may be influenced by more factors than are proposed for *S. cerevisiae* IA3.

Finally, it is perhaps not intuitively obvious why *S. castellii* IA3, but not *S. cerevisiae* IA3, might have evolved a severe pH dependency to activity, as well as an apparently broadened specificity. It is tempting to suggest that this might reflect a

difference in the environment in which *S. cerevisiae* (brewers' and bakers' yeast) and *S. castellii* (a soil-dwelling yeast) live. However, the pH at which the catastrophic loss of activity was observed – pH 3.1 – is well below the normal "physiological" pH of the vacuole of fungi (typically pH 5 - 6.5, Preston *et al.*, 1989; Harrison *et al.*, 2002). Thus, in all probability, the observed pH switch is merely an intriguing artefact of our experimental system. Similarly, as discussed in Section 11.3 (next), there may be alternative explanations for the seemingly broader specificity of the inhibitory sequence of *S. castellii* IA3.

11.3 The C-terminal tail: a role in specificity?

Despite extensive investigation, the role of the residues after position 34, i.e. the "tail", of both *S. cerevisiae* and *S. castellii* IA3 remains something of an enigma. Against its natural target, the tail of *S. cerevisiae* IA3 has little effect on inhibition. Similarly, this tail region has minimal influence on the inhibition of the *Aspergillus* enzyme or cathepsin D. However, in other situations the *S. cerevisiae* IA3 tail *can* contribute to the potency of its inhibitory sequence (residues 2-34). This was particularly noticeable against the vacuolar enzyme from *P. pastoris*, where the presence of a tail was critical to activity of certain *S. cerevisiae* IA3-derived inhibitory sequences, including wild-type *S. cerevisiae* IA3. In all cases its presence was beneficial.

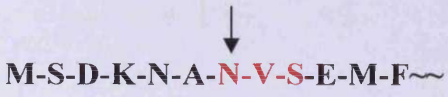
However, whilst the tail was at worst inconsequential and at best vital to the activity of *S. cerevisiae* IA3, the *opposite* was true of *S. castellii* IA3. The *S. castellii* IA3 tail was universally detrimental, and against both the *A. fumigatus* enzyme and cathepsin D it destroyed the effective inhibitory activity of the wild-type *S. castellii* inhibitory sequence altogether.

The fact that the C-terminal tails of these two IA3s behave quite distinctly raises broader questions regarding the role of the tail region of IA3 *in vivo*. Why would *S. castellii* IA3 maintain a detrimental tail sequence, and why would *S. cerevisiae* IA3 maintain a tail that appears unnecessary against its natural target? Natural selection

might be expected to remove wasteful, unnecessary or deleterious elements of a given system, so it seems unlikely that the tails of *S. cerevisiae* and *S. castellii* IA3 remain without good purpose. Several possible explanations can be proposed. Perhaps the most obvious is that the inhibitory sequences (34 residues) may be too small in themselves for efficient accumulation/survival in yeast. Certainly, a recombinant form of *S. cerevisiae* IA3 truncated after residue 34 could not be produced in *E. coli* (Dr. L. Phylip, personal communication). Therefore, the tail may be maintained as an aid to clearance from the ribosome. However, such a purpose would presumably have a low sequence-dependency and, although the sequences of the tails of the IA3 forms found thus far are considerably more divergent than the first 34 residues (see Figure 6.1B), there are nevertheless also distinct regions of identity. This suggests that the tail serves a more active purpose than merely providing ribosome clearance, especially since this would still not explain why Nature would select a tail sequence for *S. castellii* IA3 that appears detrimental to potency. As an alternative explanation, the tails may serve to protect the inhibitory sequences from cleavage by non-target proteinases, although it is not immediately clear how they may do this considering the apparent lack of structure in the IA3 molecule from *S. cerevisiae* at least (Green *et al.*, 2004).

However, when seeking roles for the C-terminal tail, an important point to remember is that the inhibitors tested in this thesis were NOT naturally-occurring products from the relevant yeast, but either recombinant forms produced in *E. coli* or chemically synthesised peptides. Whereas the characteristics of naturally-occurring, recombinant and synthetic forms of *S. cerevisiae* IA3 *in vitro* have been shown to be essentially identical in terms of inhibitory potency against *S. cerevisiae* proteinase A in this thesis (Chapter 5) and elsewhere (Phylip *et al.*, 2001), the same cannot be said of *S. castellii* IA3. Whilst the efforts have not been detailed in this thesis (see Section 6.2), a potent inhibitory activity *was* partially purified from cultured *S. castellii* using principles based on those described for the purification of IA3 from *S. cerevisiae* (Saheki *et al.*, 1974; Núñez de Castro & Holzer, 1976 ; Biedermann *et al.*, 1980; Meußdoerffer, 1980). The preparation was not sufficiently homogeneous for detailed characterisation to establish that the activity originated from naturally occurring *S. castellii* IA3; however the parallels of the purification processes, the similarities of the two yeasts and the size of the semi-purified protein as

estimated by SDS-PAGE might suggest this to be the case. Assuming that the inhibitory activity was a natural form of *S. castellii* IA3, the limited data collected indicated that the activity more closely resembles that of the synthetic peptide (namely, potent concentration-dependent inhibition of all the enzymes tested in this thesis) than the recombinant protein form (i.e. only active against the *S. cerevisiae* and *P. pastoris* proteinases). The possibility exists, therefore, that *in vivo S. castellii* IA3 may be subject to post-translational modification. For example, perhaps *in vivo* the tail region is released from residues 1-34 of *S. castellii* IA3 following cleavage by an endopeptidase. Such systems are widely reported for short antimicrobial peptides. Glycosylation is another post-translational modification, common in eukaryotes but absent from prokaryotes. Interestingly, in this context *S. castellii* IA3 sequence contains a glycosylation site at residue 7, close to the N-terminus:



Therefore, if the nascent polypeptide should encounter the glycosylation machinery within the *S. castellii* cell, naturally-occurring *S. castellii* IA3 may be a glycoprotein, whilst the recombinant and synthetic forms are certainly not. Notably, assuming the modes of inhibition of *S. castellii* IA3 and *S. cerevisiae* IA3 are equivalent, this glycosylation site is in one of the few locations where sterically such a modification might be tolerated. Residue 7, a Lys in *S. cerevisiae* IA3 (Figure 11.2) is on the hydrophilic face of the helix and exposed to the solvent. Thus, if a bulky oligosaccharide unit typical of yeasts is linked to Asn7 of *S. castellii* IA3, it is in the best location to avoid interference with the enzyme.

However, with regard to the C-terminal tail region, glycosylation could conceivably affect the behaviour of *S. castellii* IA3 in other ways. For example, an oligosaccharide could prevent inappropriate intramolecular (or intermolecular) interactions between the inhibitory sequence and another region of *S. castellii* IA3, such as the tail. One hypothetical scenario is illustrated schematically in Figure 11.3 as an example. Therefore, the universally detrimental effect of residues 35-45 of the tail (Chapters 7, 8 and 9) of recombinant *S. castellii* IA3 may, in fact, be an experimental artefact resulting from the use of recombinant, and therefore unglycosylated, rather than naturally-occurring IA3. Such hypotheses illustrate how

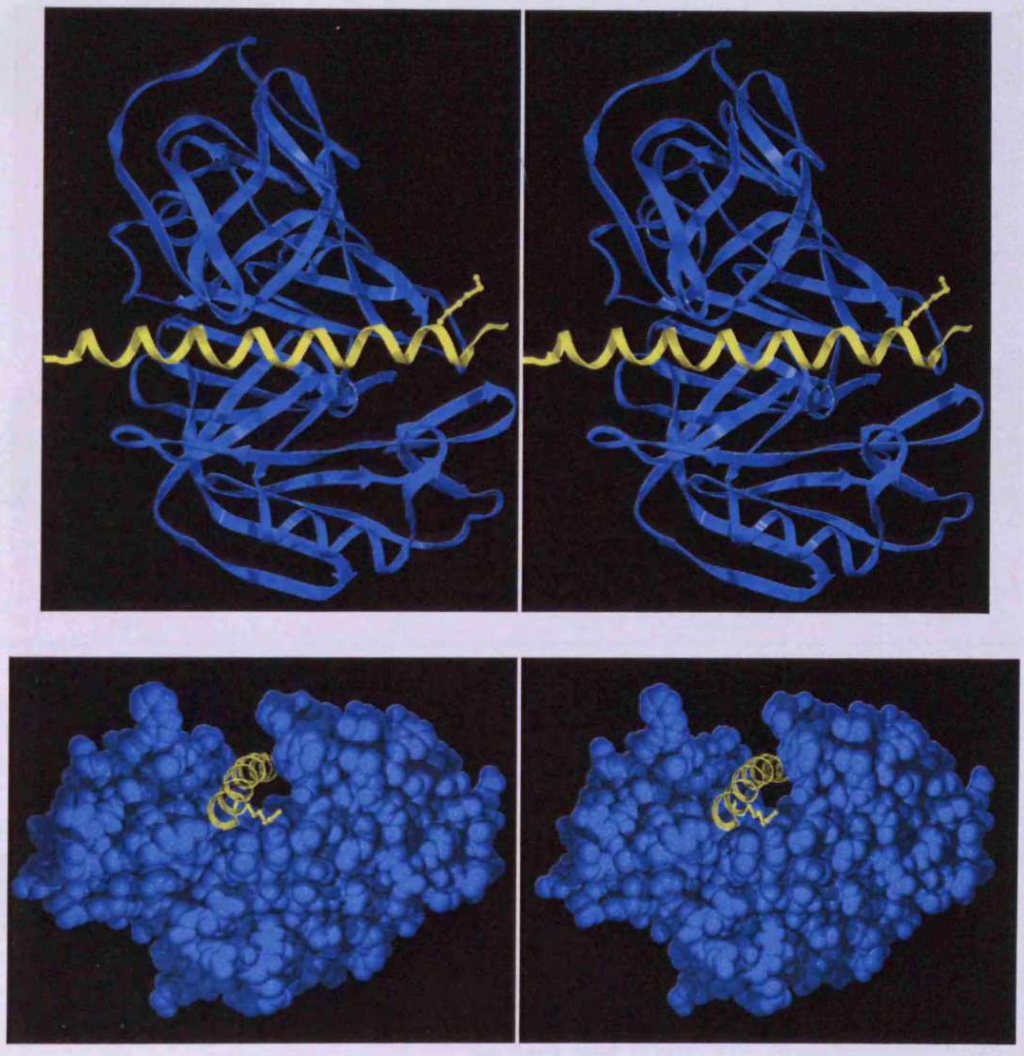


Figure 11.2 Stereo views illustrating the location of Lys7 (ball and stick) of *S. cerevisiae* IA3 (gold) in complex with *S. cerevisiae* proteinase A (blue). In *S. castellii* IA3, residue 7 is an Asn, and represents a putative glycosylation site. (PDB code 1DPJ).

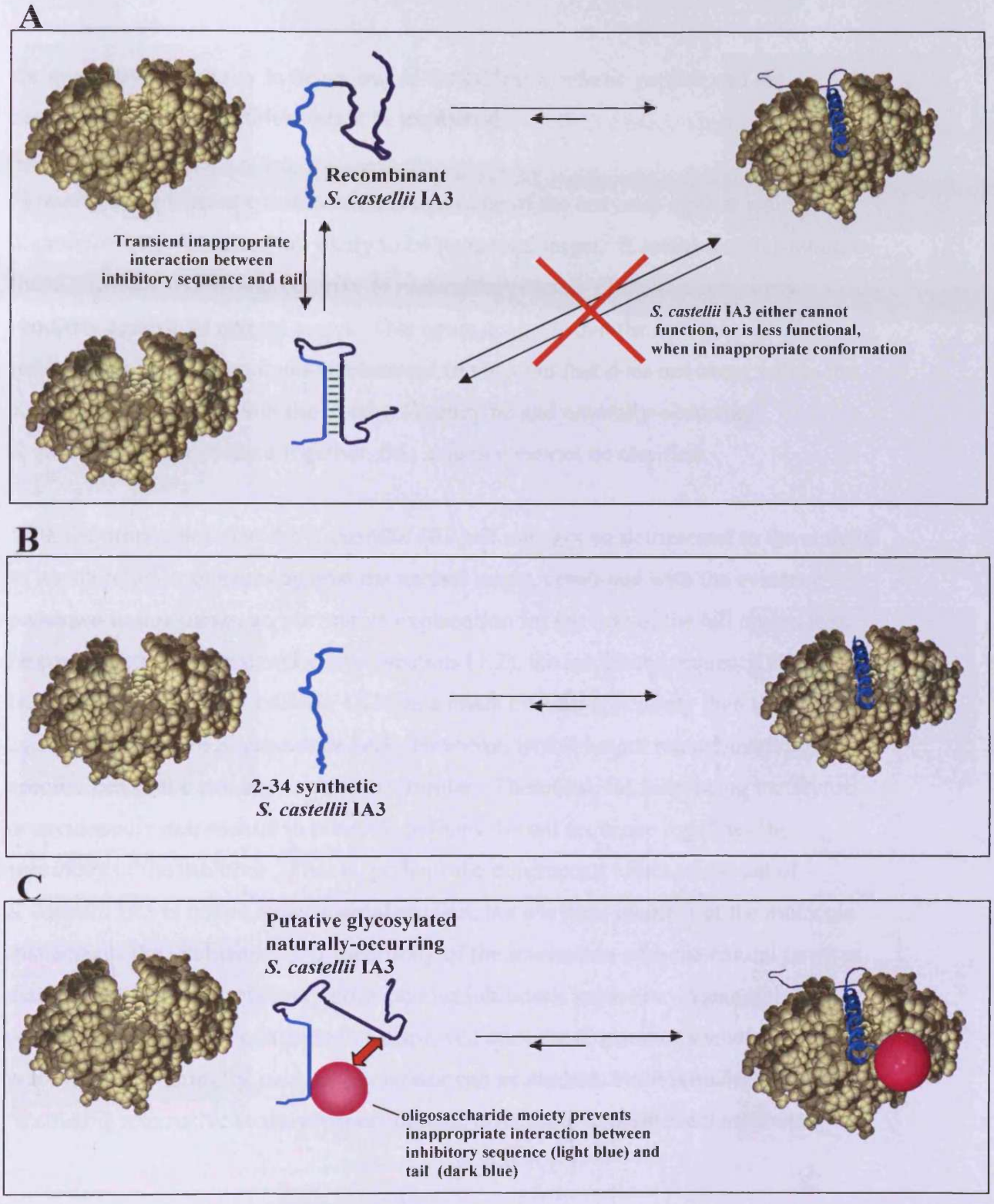


Figure 11.3 Possible explanation for the observed discrepancy between the inhibitory behaviour of *S. castellii* IA3 in the 3 forms studied: recombinant, tailless synthetic peptide and the qualitatively-assessed naturally-occurring inhibitor. **A.** The C-terminal tail (residues 35-81, dark blue) of *S. castellii* IA3 can form transient reversible interactions with the inhibitory sequence (1-34, light blue), leading to a temporary conformation that is incompatible with inhibitory binding to the enzyme. **B.** The absence of a tail means that this does not occur in synthetic peptide, which displays superior potency. **C.** If naturally-occurring IA3 in *S. castellii* IA3 is glycosylated at Asn7, the oligosaccharide moiety (purple sphere) may obstruct an interaction between the C-terminal tail and the inhibitory sequence, whilst still permitting inhibitory binding to the target enzyme.

the qualitative similarity in behaviour of the tailless synthetic peptide and the naturally-occurring inhibitor might be explained.

However, a significant consideration is that none of the enzymes against which *S. castellii* IA3 was tested are likely to be its natural target. It seems counter-intuitive that *S. castellii* IA3 would contrive to *reduce* the potency of its own inhibitory sequence against its natural target. This again suggests that the detrimental effect may reflect an artificial phenomenon, observed *in vitro* but that does not occur within the *S. castellii* cell. Until both the *S. castellii* enzyme and naturally-occurring *S. castellii* IA3 are studied together, this situation cannot be clarified.

With the proposition that the *S. castellii* IA3 tail may not be detrimental to the activity of its inhibitory sequences against the natural target, combined with the evidence presented in this thesis, an alternative explanation for the role of the tail region may be considered. As discussed above (Section 11.2), the inhibitory sequence alone (residues up to 34) of *S. castellii* IA3 has a much broader specificity than the equivalent region of *S. cerevisiae* IA3. However, in full-length recombinant form, the specificities of the two are remarkably similar. Therefore, far from being ineffectual or accidentally detrimental to potency, perhaps the tail sequence regulates the *specificity* of the inhibitor. That is, perhaps the detrimental effect of the tail of *S. castellii* IA3 is not an experimental artefact, but a critical element of the molecule that ensures the exclusivity and specificity of the interaction with the natural target is maintained, despite a relatively promiscuous inhibitory sequence. Again, this possibility can only be confirmed or disproved once the *S. castellii* vacuolar proteinase and naturally-occurring inhibitor can be studied, but it remains a fascinating alternative to the prospect that the effects are experimental artefacts.

11.4 Which technique for future progress: rational design or directed evolution?

Through a combination of rational design, common sense and good fortune, forms of both *S. cerevisiae* and *S. castellii* IA3 were produced during these studies that

displayed new and improved inhibitory activities. However, as discussed in Chapters 9 and 10, the requirement for more structural information, and the limitations of our understanding of the relationship between the sequence, structure and function of IA3, may retard future progress by rational methods. In the course of this work, the models of the *P. pastoris* and *A. fumigatus* proteinases have proved invaluable tools to inform the design of IA3 sequences and to gain an insight into the relationship between structure and activity. However, their constructor Dr. D. Bur (Actelion Ltd., Allschwil, Switzerland) would be the first to excuse them as "just models"; there is no substitute for "real" (crystallographic or NMR) structural data in the rational design of inhibitors. In relation to this project, this is especially true for *S. castellii* IA3, for which we have not yet obtained a structure in either free solution or in an inhibitory complex with any enzyme.

However, of more use than "just" the structure of an enzyme or an inhibitor, the elucidation of the structures of some of the most promising new IA3 "leads" in complex with the enzyme targets described in this thesis will be invaluable for rational design approaches. These will identify both the manner by which the inhibitors are mediating their effect and possible avenues to improve the interaction further. The benefit of detailed information regarding a specific inhibitory complex, rather than just the enzyme or inhibitor alone, is exemplified by the cathepsin D work (Chapter 10). Despite having access to detailed crystallographic structures of cathepsin D and of *S. cerevisiae* IA3 complexed with *S. cerevisiae* proteinase A, it is difficult to apply the available information to the design of inhibitors of cathepsin D based on *S. cerevisiae* IA3, or to improve the potency displayed by the wild-type inhibitory sequence of *S. castellii* IA3. Therefore, until the various crystallisation projects already in progress bear fruit, the implementation and optimisation of the directed molecular evolution framework utilising phage display (Chapter 10) is an attractive alternative to rational design.

That is not to say, however, that the knowledge gained thus far by rational means will be rendered any less significant by the adoption of the complementary method of directed evolution. Indeed, it is likely that the information gained by rational approaches will prove invaluable to the design of an effective directed evolutionary approach, and *vice versa*.

As we cannot hope to represent even an infinitesimal fraction of the 20^{34} possible IA3-derived inhibitory sequences (Section 10.5), there will be little point placing value on the absolute *quantity* of individuals within the library. Therefore, the "traditional" directed evolutionary technique of fully random whole-gene mutagenesis is unlikely to represent the most effective method of directed evolution. Instead, the knowledge gained from the "rational" investigations described in this thesis could be used to design "smart" libraries (Chica *et al.*, 2005), where random mutagenesis is targeted to specific regions, such as the centrepiece or hydrophobic pin residues of the IA3 inhibitory sequence, that have been identified as more critical to activity. This will improve the *quality* of the library.

The application of specific experimental knowledge to the design of the library in this way will help to avoid other limitations of more traditional techniques. For instance, each application of fully random whole-gene mutagenesis by error-prone PCR normally only results in individual point mutations to a handful of codons spread throughout the DNA sequence, often in areas not affecting the function or activity. Furthermore, because single base changes typically only access 6 of the 19 possible amino acid substitutions at that location (Bloom *et al.*, 2005), the potential diversity is reduced still further and a large percentage of the mutations generate either only conservative amino acid changes or, due to the degeneracy of the genetic code, no change at all. In contrast, by specifically targeting a few carefully selected codons, the mutation rate in these nucleated regions increases, meaning that larger jumps can be made in one round of error-prone PCR. Thus, libraries of variant IA3 inhibitory sequences could be greatly reduced in size whilst simultaneously increasing the comprehensiveness, relevance and functional diversity of mutations represented. Such libraries should contain a much higher frequency of IA3-derived sequences with desired functions (Roth, 2005) which, in turn, greatly increases the chances of detecting such individual(s) with the desired characteristics. These functionally diverse libraries could then be screened against a wide range of enzymes, to detect "lead" individuals displaying a variety of novel specificities. Screening libraries in this manner should require much less time and effort than the individual re-design of IA3 against each new target. By virtue of the genotype-phenotype link provided by phage display, these sequences could be identified, characterised and used as

templates for both fresh rounds of mutagenesis and also further optimisation by rational design.

Directed evolution also offers a simpler alternative for the generation of specificity or even exclusivity of an inhibitory interaction. Despite their taxonomic divergence, the substrate preferences of all the enzymes studied in this thesis are very similar, and the residues constituting the conventionally-recognised substrate/inhibitor binding pockets (Section 1.6) lining the active site clefts remain largely conserved. This makes rational design of specificity, so that an inhibitor will discriminate between closely-related enzymes, quite difficult. In contrast, a library can be screened not only for individuals displaying potent activity towards the new target enzyme, but these "positives" can be re-screened for reduced (or absent) activity towards other enzymes. By multiple iterations of the mutation and selection process, using the most promising candidate sequences as the parental templates for the production of the next library generation, together with successively increasing stringencies of the respective selection pressures, it should be possible to "evolve" exclusive and potent inhibitors against new enzyme targets.

Since the start of this project, there have been many advances in the field of directed evolution, not least in the generation of more representative genetic diversity within the library by error prone PCR systems. The traditional methods (for example, Cadwell & Joyce, 1992) contained high levels of polymerase-induced bias towards either A/T or G/C mutations. Modern approaches, such as "sequence saturation mutagenesis" described by Wong *et al.* (2004) which uses universal bases like inosine, can circumvent these biases and thus enable a much wider spectrum of mutations to be encountered in a single round. There are also commercially available kits (such as GeneMorph®II, marketed by Stratagene) that reduce bias in random mutagenesis PCRs by using two different DNA polymerases with opposite biases to counteract each other.

Thus, far from being mutually exclusive processes, the future progression of this project may lie in combining rational design and directed evolution. Such a hybrid or "semi-rational" approach (Chica *et al.*, 2005) should help to avoid the limitations of either individual process.

11.5 New targets, feasibility of IA3 as a drug and new directions

Another of the most important future elements of research into IA3 will be the identification/selection of suitable new enzyme targets. With the application of the "semi rational" approach, enzymes of commercial interest will become increasingly feasible targets.

At first glance, IA3 is an unlikely contender as a drug. It is a small, unstructured protein that is highly susceptible to cleavage by non-target enzymes. Furthermore, as a protein that is non-human (let alone a non-self molecule) it might be immunogenic, with associated risks of side-effects.

However, whilst delivering IA3-derived proteins to intracellular enzymes *in vivo* may prove difficult, *extracellular* aspartic proteinases, such as those secreted by many animal and plant pathogens as an integral part of their infection process (Section 1.1), are more attractive targets. For example, oral thrush could be treated by a sucking lozenge containing a cocktail of IA3-derived inhibitors directed against the aspartic proteinases secreted by *Candida albicans* (the SAPs, Section 1.1.5). Furthermore, the inhibitors in the lozenge could be safely ingested because gastric enzymes, including pepsin and gastricsin, would not be inhibited. Indeed, the IA3-derived molecules would be digested by these non-target enzymes in the stomach and rendered harmless. Thus, the perceived weakness of susceptibility to cleavage by non-target proteinases could circumvent two of the most common problems associated with drugs and non-hydrolysable compounds, namely undesirable/coincidental cross reactivity and slow drug clearance.

The lack of structure, small size and simple nature of IA3 imparts other, more industrially- or biotechnologically-relevant advantages. Production of eukaryotic proteins in prokaryotes is often complicated by the functional requirement for post-translational modifications. Furthermore, overexpression of large eukaryotic proteins in *E. coli* often leads to insoluble and mis-folded products. In contrast, IA3 (at least

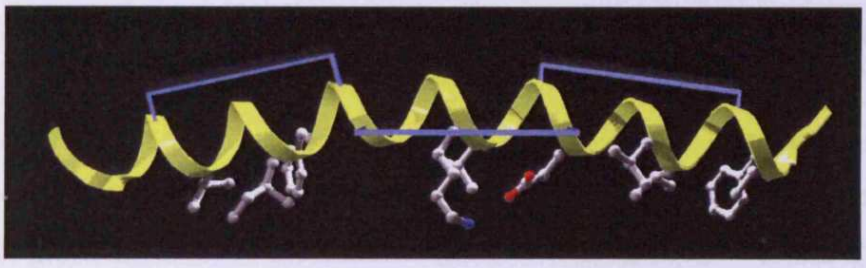
from *S. cerevisiae*) does not appear to require post-translational modification, it is unfolded in free solution, and generally produces high yields in *E. coli*. Moreover, biosynthetic production avoids many of the complications associated with the chemical synthesis of non-peptide small molecules, and the heat-stable nature of IA3 may aid a large-scale purification process.

For several reasons, in recent years therapeutic drug designers have increasingly adopted naturally-occurring molecules as "lead compounds". Dekker *et al.* (2005) describe natural products as "biologically validated starting points in structural space". By extension, even if IA3-like sequences do not themselves reach clinical viability, the information gained by their study may bear much fruit for inhibitor design. For example, it is noticeable that the small-molecule inhibitors, which normally only interact with the most central binding pockets of the much more extensive enzyme active site (Sections 1.6 and 1.7.1), are often prone to undesirable cross-reactivity with other aspartic proteinases. A prime example of this would be the pepstatins (Section 1.7.1), where sensitivity to these molecules is so widespread that it is used to define the aspartic proteinase family! In contrast, the interaction of *S. cerevisiae* IA3 with its target proteinase is remarkably specific and, from the work described in this thesis, it is concluded that much of the specificity of IA3 originates directly or indirectly from interactions at more distal locations, such as those between the back pin region and the polyproline loop.

Thus, in the future, either IA3-derived molecules, or molecules designed using information gained from studies with IA3, may yield drugs of high potency but with lower cross reactivities or side effects.

A greater understanding of the relationship between amino acid sequence, structure and function in IA3 will also be useful in this regard. One of the most fascinating and paradoxical elements of the behaviour of IA3 is its transition from an apparently unstructured and inactive random coil in free solution into a near-perfect amphipathic α-helix of subnanomolar inhibitory potency that is tailored precisely to the active site of its target enzyme. Thus, predicting the effect of mutations to the sequence is complicated because changes may affect helix-forming propensity as well as the

interaction with the binding sites within the enzyme active site. Potentially, both these consequences may affect the observed K_i value, especially since folding is a binding event and *vice versa*. Whilst the structural transition probably plays a critical role in the activity of IA3, at first consideration this event appears to be an entropically unfavourable element to the inhibitory interaction. Both entropic and enthalpic elements contribute to the free energy of binding of an inhibitor to an enzyme ($\Delta G = \Delta H - T\Delta S$). A simplistic proposition, therefore, might be that the interaction of IA3 with its enzyme is enthalpically driven and that the $T\Delta S$ term may be far from optimised. If this is the case, one possible way to improve potency might be to improve the helix-forming propensity of IA3, or even to constrain IA3 in a permanent helical formation, for example by the introduction of covalent links between residues on the non-critical solvent exposed face of the molecule:



However, it must be remembered that, *in vivo*, positive evolutionary pressures have selected for the disordered-to-ordered transition observed in IA3, and thus it must bring with it some advantage. Such beneficial considerations may also be applicable to the design of IA3 as a drug. IA3 appears to fold into its stable, active structure only when afforded the opportunity to do so by its own folding template, namely the active site of the target enzyme. In recent years, many papers have reported the widespread roles of unstructured proteins in mediating precisely-tailored protein-protein interactions (Section 1.9). Indeed, current estimates suggest that approximately one third of all eukaryotic proteins are either partially (>30 contiguous residues) or completely unfolded (Ward *et al.*, 2004; Schreiber & Serrano, 2005; Fink, 2005). Like IA3, a number of these proteins only assume regular structure upon binding their target and, in a further parallel, this structure may not encompass the whole protein but only that part involved with the interaction (Dyson & Wright, 2002). Perhaps, therefore, being pre-folded would sacrifice the exquisite specificity of the interaction of IA3 with its target, or even negate the interaction altogether.

It is also possible that, *over all*, the binding of IA3 to the enzyme does not convey an entropic penalty. For example, unstructured IA3 in free solution, and the unliganded active site cleft of the enzyme, would presumably exhibit extensive hydration layers. Many of these water molecules might be released upon inhibitor binding to the enzyme. This might offset partially the concomitant loss of conformational entropy in both IA3 and the enzyme, associated with the formation of the inhibitory complex, and so result in a net increase in entropy.

A further possibility is that IA3 is, in fact, not fully unstructured in free solution. Other disordered proteins, or disordered regions within proteins, often contain sub-regions with residual but uncoalesced structure connected by flexible linkers (Tompa, 2002; Fuxreiter *et al.*, 2004; Fink, 2005). In free solution, such proteins interconvert among a range of conformers and maintain most of the configurational adaptability associated with an unstructured protein. The data from Green *et al.* (2004) on *S. cerevisiae* IA3 (Section 1.8.4) and the unpublished data (Dr. D. Wyatt) with *S. castellii* IA3 (Section 7.5) are consistent with this situation. Therefore, if the unbound forms of IA3 *do* exhibit such regions of interchanging partial order, or at least not total disorder, the entropy change on binding to the target enzyme would be less than if IA3 was a truly random coil, and thus more readily offset by water release on binding (see above) together with the enthalpic binding terms.

It can be seen, therefore, that a deeper understanding of factors contributing to the free energy of binding of IA3 to its target would provide a fresh perspective on inhibitor design. Techniques such as isothermal titration calorimetry (Holdgate & Ward, 2005) could disentangle the effects of mutations in terms of both their entropic and enthalpic consequences for IA3. Ultimately, this may afford more sensitive inhibitor design than merely studying the effect of mutations on the K_i value, which is a reflection of the free energy as a whole. Breaking down the free energy of the inhibitory interaction into its component terms will enable the design of inhibitors that are both enthalpically and entropically optimised. Furthermore, IA3 and its interactions with the target enzyme active site would provide a paradigmatic "model system" for the

study of this fascinating breed of proteins that transit from disorder to order in the act of performing their function.

11.6 Concluding remarks

Rather like a male spider entering a female's web with amorous intentions, IA3 walks a precisely engineered tightrope between subnanomolar inhibitory potency and destruction as a substrate, in order to retain its remarkable level of specificity. Mutations that tilt the balance in favour of superior inhibition also risk compromising the exclusivity of the interaction. From the work of this thesis, the centrepiece in particular appears to be a key element of IA3. Notably, however, the wild type centrepiece of IA3 (from either species examined) has not evolved to make a major contribution to binding energy (and, therefore, to inhibitory *potency*). Instead, the residues in this region appear to have been selected for their contribution to *specificity*. This may explain why the introduction of appropriate hydrophobic residues into the centrepiece improved inhibitory potency against the vacuolar proteinases of all three fungi studied in this thesis. Not only did these introductions remove the primary specificity-determining feature of the IA3 sequence, but they also increased the hydrophobic interactions in the single "untapped" region of the *S. cerevisiae* IA3 inhibitory sequence.

However, although the identities of some residues, such as the centrepiece, are more critical than others, to consider IA3 as merely an α-helix with a few select "pins" dictating specificity and potency is a gross oversimplification of the situation. The suitability of each residue is not just dictated by its potential to interact with the enzyme; the compatibility of each residue with its near-neighbours on the inhibitory helix itself must also be considered. This was particularly noticeable, for example, in the back pin region (Section 11.1) and with certain centrepiece mutants such as *S. cerevisiae* IA3 D22K and *S. castellii* IA3 M18K/K22L. Individually, or sometimes in combination with an appropriate residue at the (*i*, *i*+4) position on the inhibitory helix, all of these centrepiece mutations were present in other mutant IA3 forms of roughly equivalent or even superior potency to the wild-type sequence when pitted

against *S. cerevisiae* proteinase A. Thus, each of these residues is compatible with the enzyme active site. Yet inhibitory activity was lost from these two centrepiece mutants as a result of incompatibilities with other residues *within* the IA3 sequence itself.

The discovery of a second, quite distinct form of IA3 in *S. castellii* revealed a whole new corridor of doors to be opened. The contrasts in the behaviour, specificities and potencies of *S. castellii* IA3 reinforces the observation that it is not just the centrepiece and hydrophobic pins that hold the key to inhibitory activity. Other regions, previously considered ineffectual or unimportant, such as the C-terminal tail and even other residues within the inhibitory sequence, also play roles. For instance, the introduction of the established "key elements" of *S. castellii* IA3 were not sufficient to bestow activity against either the *Aspergillus* enzyme or cathepsin D on *S. cerevisiae* IA3. Thus, the accumulated variations between the *S. cerevisiae* and *S. castellii* inhibitory sequences at locations other than the centrepiece and back pin regions add up to significant effects.

As a result, the task of re-directing IA3 towards new, more distant targets in a truly exclusive manner is more complicated than was perhaps first envisaged. Despite the many successes of this project, the work has also highlighted the complex and integrated nature of the interactions both within IA3 and between IA3 and the enzyme. This is epitomised by the fact that the exquisite precision of the inhibitory interaction between IA3 and its target enzyme does not prevent the inhibitory sequence from being remarkably adaptable to change. This, in turn, has led to the following overriding conclusion of the project.

Almost any mutation, even individually destructive ones, may be accommodated within an IA3 sequence, even within regions conventionally recognised to be critical to the inhibitory activity against that enzyme, as long as appropriate alterations are also made in the surrounding IA3 sequence. These combinations not only enable inhibitory potency to be regained, but also lead, directly or indirectly, to changes in the specificity of that molecule.

In many ways, the adaptive nature of IA3, and our limited understanding of the relationship between sequence, structure and function, is exemplified by the limited success of our extensive BLAST searches for other IA3 sequences in Nature. Despite the fact that others must surely exist within the available genome sequence records, we can still only detect IA3 orthologues in a handful of closely-related yeast species. This points to the divergence of sequence whilst, presumably, it succeeds in maintaining functionality. It would seem, therefore, that Nature discovered the unusually accommodating disposition of IA3 some time before us.

Acknowledgements

I would like to thank my supervisors, Dr. Colin Berry and Professor John Kay, for the benefit of their collective wisdom and for providing me with the invaluable opportunity to work in their laboratories. I thank the Biotechnology and Biological Sciences Research Council, Cardiff University and my supervisors for funding my time here. I am indebted to Dr. Lowri Phylip, whose work I inherited and who helped me so much in the short time that our paths overlapped. The help, advice and expertise of Dr. Daniel Bur (Actelion Ltd., Allschwill, Switzerland) have helped my work immeasurably. Heart-felt thanks also to Bill Edwards and John Robertson for their efforts to minimise the disruption to my research during each of the 10 or so laboratory room changes that were necessary during the last three years.

On a lighter note, my lab mates Dr. "Straightleg" Dave Wyatt and Gareth "G-funk" Wyn Jones have been infinite sources of information and advice. More importantly, they were an integral part of our "Grand Slam" winning side of 2003/4 which claimed both the Departmental pub quiz and the football tournament (*in absentia*). A good lesson learned there, I feel: never let football get in the way of a good night out – the team (or at least Jon Urch) is better off without us anyway. The only thing I won't miss about working with you two is your "duets" of *Wild Horses*.

Although it pains me to do so, I reluctantly acknowledge the existence of John Curry, Mike Allen and my bald patch. I'm still better than either of you at Halo, though. 

Most sincere gratitude, however, is reserved for my parents, my girlfriend Ele "Tim, what the @&#$ is PCR?" Ragg, and all of my friends from the "normal" World. Thanks, all of you, for being so understanding and for forgiving my almost total disappearance from public for the last 3 years. As you no doubt surmised, "I've been in the lab a fair bit recently".

References

Abdelmeguid, S.S. (1993). Inhibitors of aspartyl proteinases. Med. Res. Rev., **13**: 731-778.

Abuerrei, G.M. & Peanasky, R.J. (1974). Pepsin inhibitors from *Ascaris lumbricoides* - isolation, purification, and some properties. J. Biol. Chem., **249**: 1558-1565.

Aguilar, C.F., Cronin, N.B., Badasso, M., Dreyer, T., Newman, M.P., Cooper, J.B., Hoover, D.J., Wood, S.P., Johnson, M.S. & Blundell, T.L. (1997). The three-dimensional structure at 2.4Å resolution of glycosylated proteinase A from the lyzosome-like vacuole of *Saccharomyces cerevisiae*. J. Mol. Biol., **267**: 899 - 915.

Ammerer, G., Hunter, C.P., Rothman, J.H., Saari, G.C., Walls, L.A. & Stevens, T.H. (1986). PEP4 gene of *Saccharomyces cerevisiae* encodes proteinase A, a vacuolar enzyme required for processing of vacuolar precursors. Mol. Cell. Biol., **6**: 2490-2499.

An, C.I., Fukusaki, E.I. & Kobayashi, A. (2002). Aspartic proteinases are expressed in pitchers of the carnivorous plant *Nepenthes alata Blanco*. Planta, **214**: 661-667.

Andreeva, N.S. & Rumsh, L.D. (2001). Analysis of crystal structures of aspartic proteinases: on the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. Protein Sci., **10**: 2439-2450.

Athauda, S.B., Takahashi, T., Inoue, H., Ichinose, M. & Takahashi, K. (1991). Proteolytic activity and cleavage specificity of cathepsin E at the physiological pH as examined towards the B chain of oxidized insulin. FEBS Lett., **292**: 53-56.

Atlas, S.A., Christofalo, P., Hesson, T., Sealy, J.E. & Fritz, L.C. (1985). Immunological evidence that inactive renin is prorenin. Biochem. Biophys. Res. Commun., **132**: 1038-1045.

Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder, R.C., Cachau, R.E., Collins, J., Silva, A.M. & Erickson, J.W. (1993). Crystal-structures of native and inhibited forms of human cathepsin-D - implications for lysosomal targetting and drug design. Proc. Natl. Acad. Sci. USA, **90**: 6796-6800.

Barrett A.J. & Rawlings N.D. (1995). Families of aspartic peptidases, and those of unknown catalytic mechanism. Meth. Enzymol., **248**: 105- 120.

Beher, D., Fricker, M., Nadin, A., Clarke, E., Wrigley, J.D.J., Li, Y.M., Culvenor, J.G., Masters, C.L., Harrison, T. & Shearman, M.S. (2003). *In vitro* characterization of the presenillin-dependent γ-secretase complex using a novel affinity ligand. Biochemistry, **42**: 8133-8142.

Bennet, J.H. (1842). On the parasitic vegetable structures found growing in living animals. Trans. Roy. Soc. Edin., **15**: 277-279.

Bennett, K., Levine, T., Ellis, J.S., Peanasky, R.J., Samloff, I.M., Kay, J. & Chain, B.M. (1992). Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. Eur. J. Immunol., **22**: 1519-1524.

Bi, X., Haque, T.S., Zhou, J., Skillman, A.G., Lin, B., Lee, C.E., Kuntz, I.D., Ellman, J.A. & Lynch, G. (2000). Novel cathepsin D inhibitors block the formation of hyperphosphorylated tau fragments in hippocampus. J. Neurochem., **74**: 1469-1477.

Biedermann, K., Montali, U., Martin, B., Svendsen, I. & Ottesen, M. (1980). The amino acid sequence of proteinase A inhibitor 3 from baker's yeast. Carlsberg Res. Commun., **45**: 225-235.

Birnboim, H.C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res., **7 (6)**: 1513-1523.

Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D. & Arnold, F.H. (2005). Evolving strategies for enzyme engineering. Curr. Opin. Struct. Biol., **15**: 447-452.

Brown, A.R., Webb, J., Rebus, S., Williams, A. & Fazakerley, J.K. (2004). Identification of up-regulated genes by array analysis in scrapie-infected mouse brains. Neuropath. Appl. Neurobiol., **30**: 555-567.

Brunelle, F., Cloutier, C. & Michaud, D. (2004). Colorado potato beetles compensate for tomato cathepsin D inhibitor expressed in transgenic potato. Arch. Insect Biochem. Physiol., **55**: 103-113.

Bruun, A.W., Svendsen, I., Sorensen, S.O., Kielland-Brandt, M.C. & Winther, J.R. (1998). A high-affinity inhibitor of yeast carboxypeptidase Y is encoded by TFS1 and shows homology to a family of lipid binding proteins. Biochemistry, **37**: 3351-3357.

Bursavich, M.G. & Rich, D.H. (2002). Designing non-peptide peptidomimetics in the 21st century: inhibitors targeting conformational ensembles. J. Medic. Chem., **45**: 541-558.

Cadwell, R.C. & Joyce, G.F. (1992). Randomization of genes by PCR mutagenesis. PCR Methods Appl., **2**: 28-33.

Casso, D.J., Tanda, S., Biehs, B., Martoglio, B. & Kornberg, T.B. (2005). *Drosophila* signal peptide peptidase is an essential protease for larval development. Genetics, **170**: 139-148.

Castanheira, P., Samyn, B., Sergeant, K., Clemente, J.C., Dunn, B.M., Pires, E., van Beeumen & Faro, C. (2005). Activation, proteolytic processing, and peptide specificity of recombinant cardosin A. J. Biol. Chem., **280**: 13047-13054.

Cater, S.A., Lees, W.E., Hill, J., Brzin, J., Kay. J & Phylip, L.H. (2002). Aspartic proteinase inhibitors from tomato and potato are more potent against yeast proteinase A than cathepsin D. Biochim. Biophys. Acta, **1596**: 76-82.

Chain, B.M., Free, P., Medd, P., Swetman, C., Tabor, A.B. & Terrazzini, N. (2005). The expression and function of cathepsin E in dendritic cells. J. Immunol., **174**: 1791-1800.

Chen, K.C.S. & Tang.J. (1972). Amino acid sequence around the epoxide-reactive residues in pepsin. J. Biol. Chem., **247**: 2566-2574.

Chen, X., Pfeil, J.E. & Gal, S. (2002). The three typical aspartic proteinase genes of *Arabidopsis thaliana* are differentially expressed. Eur. J. Biochem., **269**: 4675-4684.

Chica, R.A., Doucet, N. & Pelletier, J.N. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr. Opin. Biotechnol., **16**: 278-384.

Chowdhury, S.F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Menard, R., Lefebvre, J., Konishi, Y., Cygler, M., Sulea, T. & Purisma, E.O. (2002). Design of noncovalent inhibitors of human cathepsin L from the 96-residue proregion to optimized tripeptides. J. Medic. Chem., **45**: 5321-5329.

Christeller, J.T., Farley, P.C., Ramsay, R.J., Sullivan, P.A. & Laing, W.A. (1998). Purification, characterization and cloning of an aspartic proteinase inhibitor from squash phloem exudate. Eur. J. Biochem., **254**: 160-167.

Chuman, Y., Bergman, A., Ueno, T., Saito, S., Sakaguchi, K., Alaiya, A.A., Franzen, B., Bergman, T., Arnott, D., Auer, G., Appella, E., Jornvall, H. & Linder, S. (1999). Napsin A, a member of the aspartic protease family, is abundantly expressed in normal lung and kidney tissue and is expressed in lung adenocarcinomas. FEBS Lett., **462**: 129-134.

Clark, S.J., Templeton, M.D. & Sullivan, P.A. (1997). A secreted aspartic proteinase from *Glomerella cingulata*: purification of the enzyme and molecular cloning of the cDNA. Microbiology, **143**: 1395-1403.

Coates, L., Erskine, P.T., Crump, M.P., Wood, S.P. & Cooper, J.B. (2002). Five atomic resolution structures of endothiapepsin inhibitor complexes: implications for the aspartic proteinase mechanism. J. Mol. Biol., **318**: 1405-1415.

Cooper, J.B. (2002). Aspartic proteinases in disease: a structural perspective. Curr. Drug Targets, **3**: 155-173.

Dalboge, H., Bayne, S. & Pedersen, J. (1990). *In vivo* processing of N-terminal methionine in *Escherichia coli*. FEBS Lett., **266**: 1-3.

Dame, J.B., Reddy, G.R., Yowell, C.A., Dunn, B.M., Kay, J. & Berry, C. (1994). Sequence, expression and modelled structure of an aspartic proteinase from the human malarial parasite *Plasmodium falciparum*. Mol. Biochem. Parasitol., **64**: 177-190.

Dash, C., Phadtare, S., Deshpande, V. & Rao, M. (2001). Structural and mechanistic insight into the inhibition of aspartic proteases by a slow-tight binding inhibitor from an extremophilic *Bacillus sp.*: correlation of the kinetic parameters with the inhibitor induced conformational changes. Biochemistry, **40**: 11525-11532.

Dash, C., Ahmad, A., Nath, D. & Rao, M. (2001). Novel bifunctional inhibitor of xylanase and aspartic protease: implications for inhibition of fungal growth. Antimicrobial Agents Chemotherapy, **45**: 2008-2017.

Dash, C., Kulkarni, A., Dunn, B. & Rao, M. (2003). Aspartic peptidase inhibitors: implications in drug development. Crit. Rev. Biochem. Mol. Biol., **38**: 89-119.

de Carvalho, M.H., Pham-Thi, A.T., Gareil, M., D'Arcy-Lameta, A. & Fodil, Y.Z. (2004). Isolation and characterization of an aspartic proteinase gene from cowpea (*Vigna unguiculata L. Walp.*). J. Plant Physiol., **161**: 971-976.

De Maere, V., Vercauteren, I., Gevaert, K., Vercruysse, J. & Claereboujt, E. (2005). An aspartyl protease inhibitor of *Osteragia astertagi*: molecular cloning, analysis of stage and tissue specific expression and vaccine trial. Mol. Biochem. Parasitol., **141**: 81-88.

Dekker, F.J., Koch, M.A. & Waldmann, H. (2005). Protein structure similarly clustering (PSSC) and natural product structure as inspiration sources for drug development and chemical genomics. Curr. Opin. Chem. Biol., **9**: 232-239.

Delaney, A., Williamson, A., Brand, A., Ashcom, J., Varghese, G., Goud, G.N. & Hawdon, J.M. (2005). Cloning and characterisation of an aspartyl protease inhibitor (API-1) from *Ancylostoma* hookworms. Int. J. Parasitol., **35**: 303-313.

Dixon, M. (1953). The determination of enzyme inhibitor constants. Biochem. J., **55**: 170-171.

Dreyer, T., Biedermann, K. & Ottesen, M. (1983). Yeast proteinase in beer. Carlsberg Res. Commun., **48**: 249-253.

Dreyer, T., Valler, M.J., Kay, J., Charlton, P. & Dunn, B.M. (1985). The selectivity of action of the aspartic-proteinase inhibitor IA3 from yeast (*Saccharomyces cerevisiae*). Biochem. J., **231**: 777-779.

Dreyer, T. (1989). Substrate specificity of proteinase yscA from *Saccharomyces cerevisiae*. Carlsberg Res. Commun., **54**: 85-97.

Duffy, M.S., MacAfee, N., Burt, M.D.B. & Appleton, J.A. (2002). An aspartyl protease inhibitor orthologue expressed by *Parelaphostrongylus tenuis* is immunogenic in an atypical host. Clin. Diagnost. Lab. Immunol., **9**: 763-770.

Dumas, J., Brittelli, D., Chen, J.S., Dixon, B., Hatoum-Mokdad, H., Konig, G., Subley, R., Witowsky, J. & Wong, S. (1999). Synthesis and structure activity relationships of novel small molecule cathepsin D inhibitors. Bioorg. Medicinal Chem. Lett., **9**: 2531-2536.

Dunn, B.M. (1997). Splitting image. Nat. Struct. Biol., **4**: 969-972.

Dunn, B.M. & Hung, S.H. (2000). The two sides of enzyme-substrate specificity: lessons from the aspartic proteinases. Biochim. Biophys. Acta, **1477**: 231-240.

Dunn, B.M. (2002). Structure and mechanism of the pepsin-like family of aspartic peptidases. Chem. Rev., **102**: 4431-4458.

Dunn, B.M., Goodenow, M.M., Gustchina, A. & Wlodawer, A. (2002). Retroviral proteases. Genome Biol., **3**: Reviews/3006.

Dyson, H.J. & Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol., **12**: 54-60.

Erdal, H., Berndtsson, M., Castro, J., Brunk, U., Shoshan, M.C. & Linder, S. (2005). Induction of lysosomal membrane permeabilization by compounds that activate p53-independent apoptosis. Proc. Natl. Acad. Sci. USA, **102**: 192-197.

Ersmark, K., Feierberg, I., Bjelic, S., Hamelink, E., Hackett, F., Blackman, M.J., Hulton, J., Samuelsson, B., Aqvist, J. & Hallberg, A. (2004). Potent inhibitors of the *Plasmodium falciparum* enzymes plasmepsin I and II devoid of cathepsin D inhibitory activity. J. Medicinal Chem., **47**: 110-122.

Esler, W.P., Felix, A.M., Stimson, E.R., Lachenmann, M.J., Ghilardi, J.R., Lu, Y.A., Vinters, H.V., Mantyh, P.W., Lee, J.P. & Maggio, J.E. (2000). Activation barriers to structural transition determine deposition rates of Alzheimer's disease Aβ amyloid. J. Struct. Biol., **130**: 174-183.

Farley, P.C., Christeller, J.T., Sullivan, M.E., Sullivan, P.A. & Laing, W.A. (2002). Analysis of the interaction between the aspartic peptidase inhibitor SQAPI and aspartic peptidases using surface plasmon resonance. J. Mol. Recognit., **15**: 135-144.

Faro, C., Ramalho-Santos, M., Vieira, M., Mendes, A., Simoes, I., Andrade, R., Verissimo, P., Lin, X., Tang, J. & Pires, E. (1999). Cloning and characterization of cDNA encoding cardosin A, an RGD-containing plant aspartic proteinase. J. Biol. Chem., **274**: 28724-28729.

Farzan, M., Schnitzler, C.E., Vasilieva, N., Leung, D. & Choe H. (2000). BACE2, a beta-secretase homolog, cleaves at the beta site and within the amyloid-beta region of the amyloid-beta precursor protein. Proc. Natl. Acad. Sci. USA, **97**: 9712-9717.

Fersht, A.R. (1998). Structure and mechanism in protein science. W.H. Freeman: New York.

Fink, A.L. (2005). Natively unfolded proteins. Curr. Opin. Struct. Biol., **15**: 35-41.

Foltmann, B. & Szecsi, P.B. (1998). Chymosin. The Handbook of Proteolytic Enzymes. Barrett, A.J., Rawlings, N.D. & Woessner, J.F. (eds), Academic Press, London.

Fuchs, S., Germain, S., Phillipe, J., Corvol, P. & Pinet, F. (2002). Expression of renin in large arteries outside the kidney revealed by human renin promoter/LacZ transgenic mouse. Amer. J. Path., **161**: 717-725.

Fuxreiter, M., Simon, I., Friedrich, P. & Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol., **338**: 1015-1026.

Galleschi, L., Bottari, A., Capocchi, A., Repiccioli, R. & Saviozzi, F. (1997). Partial characterization of a pepsin inhibitor from soft wheat bran. Sciences Aliments, **17**: 173-182.

Gleeson, M.A. & Howard, B.D. (1996). Genes which influence *Pichia* proteolytic activity, and uses therefor. Patent: US 5541112., :

Green, T.B., Ganesh, O., Perry, K., Smith, L., Phylip, L.H., Logan, T.M., Hagen, S.J., Dunn, B.M. & Edison, A.S. (2004). IA(3), an aspartic proteinase inhibitor from *Saccharomyces cerevisiae*, is intrinsically unstructured in solution. Biochemistry, **43**: 4071-4081.

Gribouval, O., Gonzales, M., Neuhaus, T., Aziza, J., Bieth, E., Laurent, N., Bouton, J.M., Feuillet, F., Makni, S., Ben Amar, H., Laube, G., Delezoide, A.L., Bouvier, R., Dijoud, F., Ollagon-Roman, E., Roume, J., Joubert, M., Antignac, C. & Gubler, M.C. (2005). Mutations in genes in the renin-angiotensin system are associated with autosomal recessive renal tubular dysgenesis. Nat. genetics, **37**: 964-968.

Groll, A.H., Shah, P.M., Mentzel, C., Schneider, M., Just-Nuebling, G. & Huebner, K. (1996). Trends in the postmortem epidemiology of invasive fungal infections at a university hospital. J. Infect., **33**: 23-32.

Gustchina, A., Li, M., Phylip, L.H., Lees, W.E., Kay, J. & Wlodawer, A. (2002). An unusual orientation for Tyr75 in the active site of the aspartic proteinase from *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun., **295**: 1020-1026.

Gustchina, A., Li, M., Wunschmann, S., Chapman, M.D., Pomes, A. & Wlodawer, A. (2005). Crystal structure of cockroach allergen Bia g2, an unusual zinc binding aspartic protease with a novel mode of self-inhibition. J. Mol. Biol., **348**: 433-444.

Hagen, L.S., Jacquemond, M., Lepingle, A., Lot, H. & Tepfer, M. (1993). Nucleotide sequence and genomic organization of cacao swollen shoot virus. Virology., **196**: 619-628.

Hanahan, H. (1983). Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol., **166 (4)**: 557-580

Harrison, T.S., Chen, J., Simons, E. & Levitz, S.M. (2002). Determination of the pH of the *Cryptococcus neoformans* vacuole. Medical Mycology, **40**: 329-332.

Harrop, S.A., Prociv, P. & Brindley, P.J. (1996). Acasp, a gene encoding a cathepsin D-like aspartic protease from the hookworm *Ancylostoma caninum*. Biochem. Biophys. Res. Commun., **227**: 294-302.

Hartley, B.S. (1960). Proteolytic enzymes. Annu. Rev. Biochem., **29**: 45-72.

Hay, J., Grieco, F., Druka, A., Pinner, M., Lee, S.C. & Hull. R. (1994). Detection of rice *Tungro bacilliform* virus gene products *in vivo*. Virology, **205**: 430-437.

Heimgartner, U., Pietrzak, M., Geersten, R., Brodelius, P., Figuerdo, A.C.D. & Pais, M.S.S. (1990). Purification and partial characterization of milk clotting proteases from flowers of *Cynara cardulculus*. Phytochemistry, **29**: 1405-1410.

Henderson, P.J.F. (1972). A linear equation that describes the steady-state kinetics of enzymes and subcellular particles interacting with tightly bound inhibitors. Biochem. J., **127**: 321-333.

Hill, J. & Phylip, L.H. (1997). Bacterial aspartic proteinases. FEBS Lett., **409**: 357-360.

Hill, J. & Phylip, L.H. (1998). Aspartic proteinases and inhibitors in plant pathogenesis. Aspartic proteinases: ed. James, Plenum Press: New York: 441-444.

Hiratake, J. (2005). Enzyme inhibitors as chemical tools to study enzyme catalysis: rational design, synthesis, and applications. Chem. Record, **5**: 209-228.

Hirel, P.H., Schmitter, J.M., Dessen, P., Fayat, G. & Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. USA, **86**: 8247-8251.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene, **77**: 51-59.

Holdgate, G.A. & Ward, W.H.J. (2005). Measurements of binding thermodynamics in drug discovery. Drug Discovery Today, **10**: 1543-1550.

Holzer, H. & Saheki, T. (1976). Mechanism and regulation of proteolytic processes in yeast. J. Japanese Biochem. Soc., **48**: 962-970.

Horovitz, A., Matthews, J.M. & Fersht, A.R. (1992). α-Helix stability stability in proteins: factors that influence stability at an internal position. J. Mol. Biol., **227**: 560-568.

Hussain, I., Powell, D.J., Howlett, D.R., Chapman, G.A., Gilmour, L., Murdock, P.R., Tew, D.G., Meek, T.D., Chapman, C., Schneider, K., Ratcliffe, S.J., Tattersall, D., Testa, T.T., Southan, C., Ryan, D.M., Simmons, D.L., Walsh, F.S., Dingwall, C. & Christie, G. (2000). ASP1 (BACE2) cleaves the amyloid precursor protein at the beta-secretase site. Mol. Cell. Neurosci., **16**: 609-619.

Hyland, L.J., Tomasek, T.A., Roberts, G.D., Carr, S.A., Magaard, V.W., Bryan, H.L., Fakhoury, S.A., Moore, M.L., Minnich, M.D., Culp, J.S., Desjarlais, R.L. & Meek, T.D. (1991). Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic charcterisation of reaction intermediates by O-18 isotope exchange. Biochemistry, **30**: 8441-8453.

Igarashi, S., Tazikawa, T., Tazikawa, T., Yasuda, Y., Uchiwa, H., Hayashi, S., Brysk, H., Robinson, J.M., Yamamoto, K., Brysk, M.M. & Horikoshi, T. (2004). Cathepsin D, but not cathepsin E, degrades desmosomes during epidermal desquamation. Brit. J. Dermatol., **151**: 355-361.

Ilgen, C., Cereghino, J.L. & Cregg, J.M. (2004). *Pichia pastoris*. Chapter 7 in Production of recombinant proteins: microbial and eukaryotic expression systems. Gellison, G. (ed.), Wiley-VCH Verlag, Weinheim, Germany.: 143-162.

Iqbalsyah, T.M. & Doig, A.J. (2005). Anticooperativity in a Glu-Lys-Glu salt bridge triplet in an isolated α-helical peptide. Biochemistry, **44**: 10449-10456.

James, M.N., Sielecki, A.R., Hayakawa, K. & Gelb, M.H. (1992). Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. Biochemistry., **31**: 3872-3886.

James. M.N.G. & Sielecki, A. (1983). Structure and refinement of penicillopepsin at 1.8Å resolution. J. Mol. Biol., **163**: 299-361.

Jean, L., Grosclaude, J., Labbe, M., Tomley, F. & Pery, P. (2000). Differential localisation of an *Eimeria tenella* aspartyl proteinase during the infection process. Int. J. Parasitol., **30**: 1099-1107.

Jolodar, A. & Miller, D.J. (1997). Preliminary characterisation of an *Onchocerca volvulus* aspartic protease. Int. J. Parasitol., **27**: 1087-1090.

Kagedal, K., Johansson, U. & Ollinger, K. (2001). The lysosomal protease cathepsin D mediates apoptosis induced by oxidative stress. FASEB J., **15**: U376-U391.

Kageyama, T. (1998). Molecular cloning, expresssion and characterization of an *Ascaris* inhibitor for pepsin and cathepsin E. Eur. J. Biochem., **253**: 804-809.

Kageyama, T., Ichinose, M., Tsukada-Kato, S., Omata, M., Narita, Y., Moriyama, A. & Yonezawa, S. (2000). Molecular cloning of neonate/infant-specific pepsinogens from rat stomach mucosa and their expressional change during development. Biochem. Biophys. Res. Commun., **267**: 806-812.

Kageyama, T. (2002). Pepsinogens, progastricsins, and prochymosins: structure, function, evolution and development. Cellular Mol. Life Sci., **59**: 288-306.

Kay, J. (1982). Proteolysis – a degrading business but food for thought. Biochem. Soc. Trans., **10**: 277-280.

Kay, J. (1985). Aspartic proteinases and their inhibitors. Aspartic proteinases and their inhibitors. V. Kostka. Berlin, Walter de Gruyter & Co.: 1-5.

Kay, J. & Tatnell, P.J. (1998). Cathepsin E. The Handbook of Proteolytic Enzymes, Barrett, A.J., Rawlings, N.D. & Woessner, J.F. (eds): Academic Press, London.

Keilova, H. & Tomasek, V. (1976). Isolation and properties of cathepsin D inhibitor from potatoes. Coll. Czech. Chem. Commun., **41**: 489-497.

Kesavulu, M.M., Gowda, A.S.P., Ramya, T.N.C., Surolia, N. & Suguna, K. (2005). Plasmepsin inhibitors: design, synthesis, inhibitory studies and crystal structure analysis. J. Peptide Res., **66**: 211-219.

Khan, A.R., Cherney, M.M., Tarasova, N.I. & James, M.N.G. (1997). Structural characterization of activation 'intermediate 2' on the pathway to human gastricsin. Nat. Struct. Biol., **4**: 1010-1015.

Khan, A.R. & James, M.N.G. (1998). Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Protein Sci., **7**: 815-836.

Khan, A.R., Khazanovich-Bernstein, N., Bergmann, E.M. & James, M.N.G. (1999). Structural aspects of activation pathways of aspartic protease zymogens and viral 3C protease precursors. Proc. Natl. Acad. Sci. USA, **96**: 10968-10975.

Khazanovich-Bernstein, N., Cherney, M.M., Loetscher, H., Ridley, R.G. & James, M.N. (1999). Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. Nat. Struct. Biol., **6**: 32-37.

Kick, E.K., Roe, D.C., Skillman, A.G., Liu, G.C., Ewing, T.J.A., Sun, Y.X., Kuntz, I.D. & Ellman, J.A. (1997). Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D. Chem. Biol., **4**: 297-307.

Kim, Y.T., Downs, D., Wu, S., Dashti, A., Pan, Y., Zhai, P., Wang, X., Zhang, X.C. & Lin X. (2002). Enzymic properties of recombinant BACE2. Eur. J. Biochem., **269**: 5668-5677.

Kitajima, S. & Sato, F. (1999). Plant pathogenesis-related proteins: Molecular mechanisms of gene expression and protein function. J. Biochem. (Tokyo), **125**: 1-8.

Klemba, M. & Goldberg, D.E. (2005). Characterization of plasmepsin V, a membrane-bound aspartic protease homolog in the endoplasmic reticulum of *Plasmomodium falciparum*. Mol. Biochem. Parasitol., **143**: 183-191.

Klionsky, D.J., Banta, L.M. & Emr, S.D. (1988). Intracellular sorting and processing of a yeast vacuolar hydrolase: proteinase A propeptide contains vacuolar targetting information. Mol. Cell. Biol., **8**: 2105-2116.

Klionsky, D.J., Herman, P.K. & Emr, S.D. (1990). The fungal vacuole: composition, function, and biogenesis. Microbiol. Rev., **54**: 266-292

Koelsch, G., Mares, M., Metcalf, P. & Fusek, M. (1994). Multiple functions of pro-parts of aspartic proteinase zymogens. FEBS Lett., **343**: 6-10.

Kondo, H., Shibano, Y., Amachi, T., Cronin, N., Oda, K. & Dunn, B.M. (1998). Substrate specificities and kinetic properties of proteinase A from the yeast *Saccharomyces cerevisiae* and the development of a novel substrate. J. Biochem. (Tokyo), **124**: 141-147.

Labrou, N.E. & Rigden, D.J. (2004). The structure-function relationship in the clostripain family of peptidases. Eur. J. Biochem., **271**: 983-992.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, **227**: 680-685.

Lai, M.T., Chen, E., Crouthamel, M.C., DiMuzio-Mower, J., Xu, M., Huang, O., Price, E., Register, R.B., Shi, X.P., Donoviel, D.B., Bernstein, A., Hazuda, D., Gardell, S.J. & Li, Y.M. (2003). Presenillin-1 and presenellin-2 exhibit distinct yet overlapping γ-secretase activities. J. Biol. Chem., **278**: 22475-22481

Latgé, J.P. (1999). *Aspergillus fumigatus* and aspergillosis. Clin. Microbiol. Rev., **12**: 310-350.

Laurent, F., Bourdieu, C., Kaga, M., Chilmonczyk, S., Zgrzebski, G., Yvore, P. & Pery P. (1993). Cloning and characterization of an *Eimeria acervulina* sporozoite gene homologous to aspartyl proteinases. Mol. Biochem. Parasitol., **62**: 303-312.

Lees, W.E., Kalinka, S., Meech, J., Capper, S.J., Cook, N.D. & Kay J. (1990). Generation of human endothelin by cathepsin E. FEBS Lett., **273**: 99-102.

Lenarcic, B., Ritonja, A., Strukelj, B., Turk, B. & Turk, V. (1997). Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*, is structurally related to thyroglobulin type-1 domain. J. Biol. Chem., **272**: 13899-13903.

Lenarčič, B. & Turk, V. (1999). Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. J. Biol. Chem., **274**: 563-566.

Leung, D., Abbenante, G. & Fairlie, D.P. (2000). Protease inhibitors: current status and future prospects. J. Medicinal Chem., **43**: 305-341.

Li, M., Phylip, L.H., Lees, W.E., Winther, J.R., Dunn, B.M., Wlodawer, A., Kay, J. & Gustchina, A. (2000). The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix. Nat. Struct. Biol., **7 (2)**: 113-117.

Li, X.P., Rayford, H., Shu, R.J., Zhuang, J.J. & Uhal, B.D. (2004). Essential role for cathepsin D in bleomycin-induced apoptosis of alveolar epithelial cells. Amer. J. Physiol., **287**: L46-L51.

Lin Cereghino, J. & Cregg, J.M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol. Rev., **24**: 45-66.

Lowther, W.T., Majer, P. & Dunn, B.M. (1995). Engineering the substrate specificity of rhizopuspepsin: the role of Asp 77 of fungal aspartic proteinases in facilitating the cleavage of oligopeptide substrates with lysine in P1. Protein Sci., **4**: 689-702.

Luo, Y., Bolon, B., Kahn, S., Bennett, B.D., Babu-Khan, S., Denis, P., Fan, W., Kha, H., Zhang, J., Gong, Y., Martin, L., Louis, J.C., Yan, Q., Richards, W.G., Citron, M. & Vassar R. (2001). Mice deficient in BACE1, the Alzheimer's beta-secretase, have normal phenotype and abolished beta-amyloid generation. Nat. Neurosci., **4**: 231-232.

Macriniszyn, N., Hartsuck, J.A. & Tang, J. (1976). Mode of inhibition of acid proteases by pepstatin. J. Biol. Chem., **251**: 7088-7094.

Majer, P., Collins, J.R., Gulnik, S.V. & Erickson, J.W. (1997). Structure-based subsite specificity mapping of human cathepsin D using statine-based inhibitors. Prot. Sci., **6**: 1458-1466.

Martzen, M.R., McMullen, B.A., Smith, N.E., Fujikawa, K. & Peanasky, R.J. (1990). Primary structure of the major pepsin inhibitor from the intestinal parasitic nematode *Ascaris suum*. Biochemistry, **29**: 7366-7372.

McClelland, M. & Nelson, M. (1992). Effect of site-specific methylation on DNA modification methyltransferases and restriction endonucleases. Nucleic Acids Res., **20 Suppl**: 2145-2157.

McConnell, R.M., Godwin, W.E., Stefan, A., Newton, C., Myers, N. & Hatfield, S.E. (2003). Synthesis and cathepsin D inhibition of peptide-hydroxyethyl amine isosteres with cyclic tertiary amines. Lett. Peptide Sci., **10**: 69-78.

Meußdoerffer, F., Tortora, P. & Holzer, H. (1980). Purification and properties of proteinase A from yeast. J. Biol. Chem., **255**: 12087-12093.

Milne, J.L.S., Shi, D., Sunshine, J.S., Domingo, G.J., Wu, X., Brooks, B.R., Perham, R.N., Henderson, R. & Subramanian, S. (2002). Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine. EMBO J., **21**: 5587-5598.

Miroux, B. & Walker, J.E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol., **260**: 289-298.

Mitchell, W.M. & Harrington, W.F. (1968). Purification and properties of clostridiopeptidase B (Clostripain). J. Biol. Chem., **243**: 4683-4692.

Monod, M., Paris, S., Sanglard, D., Jaton-Ogay, K., Ave, P. & Latgé, J.P. (1993). Isolation and characterization of a secreted metalloprotease of *Aspergillus fumigatus*. Infect. Immun., **61**: 4099-4104.

Monod, M., Capoccia, S., Lechenne, B., Zaugge, C., Holdum, M. & Jousson, O. (2002). Secreted proteases from pathogenic fungi. Int. J. Med. Microbiol., **292**: 405-419.

Morrison, C.J., Hurst, S.F., Bragg, S.L., Kuykendall, R.J., Diaz, H., McLaughlin, D.W. & Reiss, E. (1993). Purification and characterization of the extracellular aspartyl proteinase of *Candida albicans*: removal of extraneous proteins and cell wall mannoprotein and evidence for lack of glycosylation. J. Gen. Microbiol., **139**: 1177-1786.

Movahedi, S., Norey, C.G., Kay, J. & Heale, J.B. (1991). Infection and pathogenesis of cash crops by *Botrytis cinerea*: primary role of an aspartic proteinase. Adv. Exp. Med. Biol., **306**: 213-216.

Murakami, S., Kondo, Y., Nakano, T. & Sato, F. (2000). Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. FEBS Lett., **468**: 15-18.

Nakanishi, H., Amano, T., Sastradipura, D.F., Yoshimine, Y., Tsukuba, T., Tanabe, K., Hirosu, I., Ohono, T. & Yamamoto, K. (1997). Increased expression of cathepsins E and D in neurons of the aged rat brain and their colocalization with lipofuscin and carboxy-terminal fragments of Alzheimer amyloid precursor protein. J. Neuro. Chem., **68**: 739-749.

Ng, K.K.S., Petersen, J.F.W., Cherney, M.M., Garen, C., Zalatoris, J.J., Rao-Naik, C., Dunn, B.M., Martzen, M.R., Peanasky, R.J. & James, M.N.G. (2000). Structural basis for the inhibition of porcine pepsin by *Ascaris* pepsin inhibitor-3. Nature Struct. Biol., **7**: 653-657.

Núñez de Castro, I. & Holzer, H. (1976). Studies on the proteinase A inhibitor I3(A) from yeast. Hoppe-Seyler's Z. Physiol. Chem., **357**: 727-734.

Oefner, C., Binggeli, A., Breu, V., Bur, D., Clozel, J.P., D'Arcy, A., Dorn, A., Fischli, W., Gruninger, F., Guller, R., Hirth, G., Marki, H., Mathews, S., M ller, M., Ridley, R.G., Stadler, H., Vieira, E., Wilhelm, M., Winkler, F. & Wostl, W. (1999). Renin inhibition by substituted piperidines: a novel paradigm for the inhibition of monomeric aspartic proteinases? Chem. Biol., **6**: 127-131.

Ord, T., Kolmer, M., Villems, R. & Saarma, M. (1990). Structure of the human genomic region homologous to the bovine prochymosin-encoding gene. Gene, **91**: 241-246.

Parris, K.D., Hoover, D.J., Damon, D.B. & Davies, D.R. (1992). Synthesis and crystallographic analysis of two rhizopuspepsin inhibitor complexes. Biochemistry., **31**: 8125-8141.

Patel, S., Vuillard, L., Cleasby, A., Murray, C.W. & Yon, J. (2004). Apo and inhibitor complex structures of BACE (β-secretase). J. Mol. Biol., **343**: 407-416.

Pauling, L. (1946). Chem. Eng. News, **24**: 1375

Pearl, L. & Blundell, T. (1984). The active-site of aspartic proteinases. FEBS Letts., **174**: 96-101.

Pedersen, J., Lauritzen, C., Madsen, M.T. & Dahl, S.W. (1999). Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases. Protein Expr. Purif., **15**: 389-400.

Perchick, G.B. & Jabbour, H.N. (2003). Cyclooxygenase-2 overexpression inhibits cathepsin D-mediated cleavage of plasminogen to the potent antiangiogenic factor angiostatin. Endocrinol., **144**: 5322-5328.

Phylip, L.H., Lees, W.E., Brownsey, B.G., Bur, D., Dunn, B.M., Winther, J.R., Gustchina, A., Li, M., Copeland, T., Wlodawer, A. & Kay, J. (2001). The potency and specificity of the interaction between the IA3 inhibitor and its target aspartic proteinase from *Saccharomyces cerevisiae*. J. Biol. Chem., **276**: 2023-2030.

Piwnica, D., Touraine, P., Struman, I., Tabruyn, S., Bolbach, G., Clapp, C., Martial, J.A., Kelly, P.A. & Goffin, V. (2004). Cathepsin D processes human prolactin into multiple 16k-like N-terminal fragments: study of their antiangiogenic properties and physiological relevance. Mol. Endocrinol., **18**: 2522-2542.

Porath, J., Carlsson, J., Olsson, I. & Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. Nature, **258**: 598-599.

Powell, D.J., Bur, D., Wlodawer, A., Gustchina, A., Payne, S.L., Dunn, B.M. & Kay, J. (1996). Expression, characterisation and mutagenesis of the aspartic proteinase from equine infectious anaemia virus. Eur. J. Biochem., **241**: 664-674.

Preston, R.A., Murphy, R.F. & Jones, E.W. (1989). Assay of the vacuolar pH in yeast and identification of acidification-defective mutants. Proc. Natl. Acad. Sci. USA, **86**: 7027-7031.

Rahuel, J., Rasetti, V., Maibaum, J., Rueger, H., Goschke, R., Cohen, N.C., Stutz, S., Cumin, F., Fuhrer, W., Wood, J. & Grutter, M.G. (2000). Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of renin. Chem. Biol., **7**: 493-504.

Ramalho-Santos, M., Verissimo, P., Faro, C. & Pires, E. (1996). Aspartic proteinases from the flowers of the cardoon *Cynara cardunculus L.* Biochim. Biophys. Acta, **1297**: 83-89.

Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.a., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. & Wongstaal, F. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature, **313**: 277-284.

Rawlings, N.D., Tolle, D.P. & Barrett, A.J. (2004). Evolutionary families of peptidase inhibitors. Biochem. J., **378**: 705-716.

Rawlings, N.D., Tolle, D.P. & Barrett, A.J. (2004). MEROPS: the peptide database. Nucleic Acids Res., **32**: D160-D164.

Reichard, U., Buttner, S., Eiffert, H., Staib, F. & Ruchel, R. (1990). Purification and characterization of an extracellular serine proteinase from *Aspergillus fumigatus* and its detection in tissue. J. Med. Microbiol., **33**: 243-251.

Reichard, U., Eiffert, H. & Ruchel, R. (1994). Purification and characterization of an extracellular aspartic proteinase from *Aspergillus fumigatus*. J. Med. Vet. Mycology, **32**: 427-436.

Reichard, U., Monod, M. & Ruechel, R. (1995). Molecular cloning and sequencing of the gene encoding an extracellular aspartic proteinase from *Aspergillus fumigatus*. FEMS Microbiol. Lett., **130**: 69-74.

Reichard, U., Cole, G.T., Ruchel, R. & Monod, M. (2000). Molecular cloning and targeted deletion of PEP2 which encodes a novel aspartic proteinase from *Aspergillus fumigatus*. Int. J. Med. Microbiol., **290**: 85-96.

Rich, D.H., Bursavich, M.G. & Estiarte, M.A. (2002). Discovery of nonpeptide, peptidomimetic peptidase inhibitors that target alternate enzyme active site conformations. Biopolymers, **66**: 115-125.

Richards, A.D., Phylip, L.H., Farmerie, W.G., Scarborough, P.E., Alvarez, A., Dunn, B.M., Hirel, P.H., Konvalinka, J., Strop, P., Pavlickova, L., Kostka, V. & Kay, J. (1990). Sensitive, soluble chromogenic substrates for HIV-1 proteinase. J. Biol. Chem., **265**: 7733-7736.

Richardson, J.S., (1981). The anatomy and taxonomy of protein structure. Advan. Protein Chem., **34**: 167-339

Richter, C., Tanaka, T. & Yada, R.Y. (1998). Mechanism of activation of the gastric aspartic proteinases: pepsinogen, progastricsin and prochymosin. Biochem. J., **335**: 481-490.

Ritonia, A., Krizaj, I., Mesko, P., Kopitar, M., Lucovnik, P., Strukelj, B., Pungercar, J., Buttle, D.J., Barrett, A.J., & Turk, V. (1990). The amino acid sequence of a novel inhibitor of cathepsin-D from potato. FEBS Lett., **267**: 13-15.

Roberds, S.L. and others. (2001). BACE knockout mice are healthy despite lacking the primary β-secretase activity in brain: implications for Alzheimer's disease therapeutics. Human Mol. Genetics, **10**: 1317-1324.

Rodrigo, I., Vera, P. & Conejero, V. (1989). Degradation of tomato pathogenesis-related proteinase by an endogenous 37-kDa aspartyl endoproteinase. Eur. J. Biochem., **184**: 663-669.

Rodrigo, I., Vera, P., Valoon, L.C. & Conejero, V. (1991). Degradation of tobacco pathogenesis-related proteins – evidence for conserved mechanisms of degradation of pathogenesis-related proteins in plants. Plant Physiol., **95**: 616-622.

Rodriguez, E.J., Angeles, T.S. & Meek, T.D. (1993). Use of N-15 kinetic isotope effects to elucidate details of the chemical mechanism of human immunodeficiency virus-1 protease. Biochemistry, **32**: 12380-12385.

Rokas, A., Williams, B.L., King, N. & Carroll, S.B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenetics. Nature, **425**: 798-802.

Rost, B. & Sander, C. (1993). Prediction of protein secondary structure at better than 70-percent accuracy. J. Mol. Biol., **232**: 584-599.

Roszkowska-Jakimiec, W. & Lesniewska, J. (2004). The activity and location of cathepsin D inhibitor in seeds of common vetch (*Vicia sativa L.*). Rocz. Akad. Med. Bialynmst., **49 (Suppl 1)**: 231-233.

Roth, H.J. (2005). There is no such thing as 'diversity'! Curr. Opin. Chem. Biol., **9**: 293-295.

Rothnie, H.M., Chapdelaine, Y. & Hohn, T. (1994). Pararetroviruses and retroviruses: a comparative review of viral structure and gene expression strategies. Adv. Virus Res., **44**: 1-67.

Rupp, S., Hirsch, H.H. & Wolf, D.H. (1991). Biogenesis of the yeast vacuole (lysosome). Active site mutation in the vacuolar aspartate proteinase yscA blocks maturation of vacuolar proteinases. FEBS Lett., **293**: 62-66.

Rusconi, F. & Belghazi, M. (2002). Desktop prediction/analysis of mass spectrometric data in proteomic projects by using massXpert. Bioinformatics, **18**: 644-645.

Russel, M. (1991). Filamentous phage assembly. Mol. Microbiol., **5**: 1607-1613.

Saftig, P., Hetman, M., Schmahl, W., Weber, K., Heine, L., Mossmann, H., Koster, A., Hess, B., Evers, M, von Figura, K. & Peters, C. (1995). Mice deficient for the lysosomal proteinase cathepsin D exhibit progressive atrophy of the intestinal-mucosa and profound destruction of lymphoid-cells. EMBO J., **14**: 3599-3608.

Saheki, T., Matsuda, Y. & Holzer, H. (1974). Purification and characterization of macromolecular inhibitors of proteinase A from yeast. Eur. J. Biochem., **47**: 325-332.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, **74 (12)**: 5463-5467.

Sarkkinen, P., Kalkkinen, N., Tilgmann, C., Siuro, J., Kervinen J. & Mikola, L. (1992). Aspartic proteinase from barley grains is related to mammalian lysosomal cathepsin D. Planta, **186**: 317-323.

Saura, C.A., Choi, S.Y., Beglopoulos, V., Malkani, S., Zhang, D., Rao, B.S.S., Chatterji, S., Kelleher III, R.J., Kandel, E.R., Duff, K., Kirkwood, A. & Shen, J. (2004). Loss of presenilin function causes impairments of memory and synaptic plasticity followed by age-dependent neurodegeneration. Neuron, **42**: 23-36.

Schaller, A. & Ryan CA. (1996). Molecular cloning of a tomato leaf cDNA encoding an aspartic protease, a systemic wound response protein. Plant Mol. Biol., **31**: 1073-1077.

Schaller, M., Schafer, W., Korting, H.C. & Hube, B. (1998). Differential expression of secreted aspartyl proteinases in a model of human oral candidosis and in patient samples from the oral cavity. Mol. Microbiol., **29**: 605-615.

Schaller, M., Hube, B., Ollert, M.W., Schafer, W., Borg-von Zepelin, M., homa-Greber, E. & Korting, H.C. (1999). *In vivo* expression and localization of *Candida albicans* secreted aspartyl proteinases during oral candidiasis in HIV-infected patients. J. Invest. Dermatol., **112**: 383-386.

Schauer-Vukasinovic, V., Bur, D., Kling, D., Gruninger, F. & Giller, T. (1999). Human napsin A: expression, immunochemical detection, and tissue localization. FEBS Lett., **462**: 1-2.

Schechter, I. & Berger, A. (1967). On the size of the active site in Proteases. I. Papain. Biochem. Biophys. Res. Commun., **27 (2)**: 157-162.

Schramm, V.L. (2005). Enzymatic transition states: thermodynamics, dynamics and analogue design. Arch. Biochem. Biophys., **433**: 13-26.

Schreiber, G. & Serrano, L. (2005). Folding and binding: an extended family business. Curr. Opin. Struct. Biol., **15**: 1-3.

Schu, P. & Wolf, D.H. (1991). The proteinase yscA-inhibitor, I(A)3, gene. Studies of cytoplasmic proteinase inhibitor deficiency on yeast physiology. FEBS Lett., **283**: 78-84.

Schu, P., Suarez-Rendueles, P. & Wolf, D.H. (1991). The proteinase yscB inhibitor (PBI2) gene of yeast and studies on the function of its protein product. Eur. J. Biochem., **197**: 1-7.

Schwartz, D.E., Tizard, R. & Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. Cell, **32**: 853-869.

Shaw, R.J., McNeill, M.M., Maass, D.R., Hein, W.R., Barber, T.K., Wheeler, M., Morris, C.A. & Schoemaker, C.B. (2003). Identification and characterisation of an aspartyl protease inhibitor homologue as a major allergen of *Trichostrongylus colubriformis*. Int. J. Parasitol., **33**: 1233-1243.

Shintani, T., Kobayashi, M. & Ichishima, E. (1996). Characterization of the S1 subsite specificity of aspergillopepsin I by site-directed mutagenesis. J. Biochem. (Tokyo), **120**: 974-981.

Sielecki, A.R., Fujinaga, M., Read, R.J. & James, M.N. (1991). Refined structure of porcine pepsinogen at 1.8 A resolution. J. Mol. Biol., **219**: 671-692.

Silva, A.M., Lee, A.Y., Gulnik, S.V., Maier, P., Collins, J., Bhat, T.N., Collins, P.J., Cachau, R.E., Luker, K.E., Gluzman, I.Y., Francis, S.E., Oksman, A., Goldberg, D.E. & Erickson, J.W. (1996). Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA, **93**: 10034-10039.

Simoes, I. & Faro, C. (2004). Structure and function of plant aspartic proteinases. Eur. J. Biochem., **271**: 2067-2075.

Sinha, S. & Lieberburg, I. (1999). Cellular mechanisms of beta-amyloid production and secretion. Proc. Natl. Acad. Sci. USA, **96**: 11049-11053.

Smyth, T.P. (2004). Substrate variants versus transition state analogues as noncovalent, reversible inhibitors. Bioorg. Medicinal Chem., **12**: 4081-4088.

Sorensen, S.O., van den Hazel, H.B., Kielland-Brandt, M.C. & Winther, J.R., , (1994). pH-dependent processing of yeast procarboxypeptidase Y by proteinase A *in vivo* and *in vitro*. Eur. J. Biochem., **220**: 19-27.

Strisovsky, K., Tessmer, U., Langner, J., Konvalinka, J. & Krausslich, H.G. (2000). Systematic mutational analysis of the acitve-site threonine of HIV-1 proteinase: rethinking the "fireman's grip" hypothesis. Protein Sci., **9**: 1631-1641.

Strukelj, B., Lenarcic, B., Gruden, K., Pungercar, J., Rogelj, B., Turk, V., Bosch, D. & Jongsma, M.A. (2000). Equistatin, a protease inhibitor from the sea anemone *Actinia equina*, is composed of three strutural and functional domains. Biochem. Biophys. Res. Commun., **269**: 732-736.

Studier, F.W. & Moffatt, B.A. (1986). Use of bacteriophage-T7 RNA-polymerase to direct selective high-level expression of cloned genes. J. Mol Biol., **189 (1)**: 113-130.

Sugana, K., Padlan, E.A., Smith, C.W., Carlson, W.D. & Davis, D.R. (1987). Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinesis* – implications for a mechanism of action. Proc. Natl. Acad. Sci. USA, **84**: 7009-7013.

Suzuki, A., Shijubo, N., Yamada, G., Ichimiya, S., Satoh, M., Abe, S. & Sato, N. (2005). Napsin A is useful to distinguish primary lung adenocarcinomas of other organs. Pathology Res. Practice, **201**: 579-586.

Suzuki, F., Murakami, K., Nakamura, Y. & Inagami, T. (1998). Renin. The Handbook of Proteolytic Enzymes, Barrett, A.J., Rawlings, N.D. & Woessner, J.F. (eds): Academic Press, London.

Takahashi, S., Takahashi, K., Kaneko, T., Ogasawara, H., Shindo, S. & Kobayashi, M. (1999). Human renin binding protein is the enzyme N-acetyl-glucosamine 2-epimerase. J. Biochem. (Tokyo), **125**: 348-353.

Tanaka, T. & Yada, R.Y. (2001). N-terminal portion acts as an initiator of the activation of pepsin at neutral pH. Protein Engineering, **14**: 669-674.

Tang, J., James, M.N.G. & Hsu, I.N. (1978). Structural evidence for gene duplication in the evolution of the acid proteases. Nature, **271**: 618-621.

Tang, J. & Wong, R.N.S. (1987). Evolution in the stucture and function of aspartic proteases. J. Cell. Biochem., **33**: 53-63.

Tang, J. & Koelsch, G. (1995). A possible function of the flaps of aspartic proteases: the capture of substrate side chains determines the specificity of cleavage positions. Protein Peptide Lett., **2**: 257-266.

Tang. J. (1998). Pepsin A. The Handbook of Proteolytic Enzymes., Barrett, A.J., Rawlings, N.D. & Woessner, J.F. (eds): Academic Press, London.

Tatnell, P.J., Powell, D.J., Hill, J., Smith, T.S., Tew, D.G. & Kay J. (1998). Napsins: new human aspartic proteinases. Distinction between two closely related genes. FEBS Lett., **441**: 43-48.

Teichert, U., Mechler, B., Muller, H. & Wolf, D.H. (1989). Lysosomal (vacuolar) proteinases of yeast are essential catalysts for protein degradation, differentiation, and cell survival. J. Biol. Chem., **264**: 16037-16045.

ten Have, A., Dekkers, E., Kay, J., Phylip, L.H. & van Kan, J.A.L. (2004). An aspartic proteinase gene family in the filamentous fungus *Botrytis cinerea* contains members with novel features. Microbiology, **150**: 2475-2489.

Terauchi, K., Asakura, T., Nishizawa, N.K., Matsumoto, I. & Abe, K. (2004). Characterization of the genes for two soybean aspartic proteinases and analysis of their different tissue dependent expression. Planta, **218**: 947-957.

Thum, M. (2000). Structure and function of the yeast vacuole and its role in autophagy. Microscopy Research and Technique, **51**: 563-572.

Tompa, P. (2002). Intrinsically unstructured proteins. Trends Biochem. Sci., **27**: 527-533.

Torruella, M.M., Gordon, K. & Hohn, T. (1989). Cauliflower mosaic virus produces an aspartic proteinase to cleave its polyproteins. EMBO J., **8**: 2819-2825.

Tsukuba, T., Okamoto, K., Okamoto, Y., Yanagawa, M., Kohmura, K., Yasuda, Y., Uchi, H., Nakahara, T., Furue, M., Nakayama, K., Kadowaki, T., Yamamoto, K. & Nakayama, K.I. (2003). Association of cathepsin E deficiency with development of atopic dermatitis. J. Biochem. (Tokyo), **134**: 893-902.

Tyas, L. (1997). Plasmepsins I and II, malarial aspartic proteinases. PhD thesis, University of Wales.

Tyndall, J.D.A., Nall, T. & Fairlie, D.P. (2005). Proteases universally recognize beta strands in their active sites. Chem. Rev., **105**: 973-999.

Valler, M.J. & Kay, J. (1985). The interaction of aspartic proteinases with naturally occurring inhibitors from *Actinomycetes* and *Ascaris lumbricoides*. J. Enzyme Inhibition., **1**: 77-82.

van den Hazel, H.B., Kielland-Brandt, M.C. & Winther, J.R. (1993). The propeptide is required for *in vivo* formation of stable active yeast proteinase A and can function even when not covalently linked to the mature region. J. Biol. Chem., **268**: 18002-18007.

van den Hazel, H.B., Kielland-Brandt, M.C. & Winther, J.R. (1995). Random substitution of large parts of the propeptide of yeast proteinase A. J. Biol. Chem., **270**: 8602-8609.

van den Hazel, H.B., Kielland-Brandt, M.C. & Winther, J.R. (1996). Review: biosynthesis and function of yeast vacuolar proteases. Yeast, **12**: 1-16.

van den Hazel, H.B., Wolff, A.M., Kielland-Brandt, M.C. & Winther, J.R. (1997). Mechanism and ion-dependence of *in vitro* autoactivation of yeast proteinase A: possible implications for compartmentalized activation *in vivo*. Biochem. J., **326**: 339-344.

van Es, J.H., van Gijn, M.E., Riccio, O., van den Born, M., Voojis, M., Begthel, H., Cozijnsen, M., Robine, S., Winton, D.J., Radtke, F. & Clevers, H. (2005). Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas not goblet cells. Nature, **435**: 959-963.

Veerapandian, B., Cooper, J.B., Sali, A., Blundell, T., Rosati, R.L., Dominy, B.W., Damon, D.B. & Hoover, D.J. (1992). Direct observation by X-ray analysis of the tetrahedral 'intermediate' of aspartic proteinases. Protein Sci., **1**: 322-328.

Verissimo, P., Faro, C., Moir, A.J., Lin, Y., Tang, J. & Pires, E. (1996). Purification, characterization and partial amino acid sequencing of two new aspartic proteinases from fresh flowers of *Cynara cardunculus L.* Eur. J. Biochem., **235**: 762-768.

Vogeser, M., Haas, A., Ruckdeschel, G. & Wanders, A. (1999). A four-year review of fatal aspergillosis. Eur. J. Clin. Microbiol. Infect. Dis., **18**: 42-45.

Wang, P.H., Do, Y.S., Macaulay, L., Shinagawa, T., Anderson, P.W., Baxter, J.D. & Hsueh, W.A. (1991). Identification of renal cathepsin B as a human prorenin-processing enzyme. J. Biol. Chem., **266**: 12633-12638.

Wang, W. & Kollman, P.A. (2001). Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. Proc. Natl. Acad. Sci. USA, **98**: 14937-14942.

Wang, Z.Y., He, G.Q., Liu, Z.S., Ruan, H., Chen, Q.H. & Xiong, H.P. (2005). Purification of yeast proteinase A from fresh beer and its specificity on foam proteins. Int. J. Food Sci. Technol., **40**: 1-6.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. & Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three Kingdoms of Life. J. Mol. Biol., **337**: 635-645.

Weihofen, A., Lemberg, M.K., Friedmann, E., Rueeger, H., Schmitz, A., Paganetti, P., Rovelli, G. & Martoglio, B. (2003). Targeting presenilin-type aspartic protease signal peptidase with γ-secretase inhibitors. J. Biol. Chem., **278**: 16528-16533.

Werner, R., Guitton, M.C. & Muhlbach, H.P. (1993). Nucleotide sequence of a cathepsin D inhibitor protein from tomato. Plant Physiol., **103**: 1473-1473.

White, P.C., Cordeiro, M.C., Arnold, D., Brodelius, P.E. & Kay, J. (1999). Processing, Activity, and inhibition of recombinant cyprosin, an aspartic proteinase from cardoon (*Cynara cardunculus*). J. Biol. Chem., **274**: 16685 - 16693.

Williamson, A.L., Brindley, P.J., Abbenante, G., Prociv, P., Berry, C., Girdwood, K., Pritchard, D.I., Fairlie, D.P., Hotez, P.J., Dalton, J.P. & Loukas, A. (2002). Cleavage of hemoglobin by hookworm cathepsin D aspartic proteases and its potential contribution to host specificity. FASEB J., **16**: 1458-1460.

Winther, J.R., Phylip, L.H. & Kay, J. (2004). Saccharopepsin. Handbook of Proteolytic Enzymes, 2nd ed, Elsevier.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J. & Kent, S.B.H. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science, **245**: 616-621.

Wlodawer, A., Gustchina, A., Reshetnikova, L., Lubkowski, J., Zdanov, A., Hui, K.Y., Angleton, E.L., Farmerie, W.G., Goodenow, M.M., Bhatt, D., Zhang, L. & Dunn, B. (1995). Structure of an inhibitor complex of the proteinase from feline immunodeficiency virus. Nat. Struct. Biol., **2**: 480-488.

Wlodawer, A. & Gustchina, A. (2000). Structural and biochemical studies of retroviral proteases. Biochim. Biophys. Acta, **1477**: 16-34.

Wolfe, M.S., Xia, W., Ostaszewski, B.L., Diehl, T.S., Kimberly, W.T. & Selkoe, D.J. (1999). Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity. Nature, **398**: 513-517.

Wolfenden, R. (1999). Conformational aspects of inhibitor design: enzyme-substrate interactions in the transition state. Bioorg. Medicinal Chem., **7**: 647-652.

Wong, J.Y., Harrop, S.A., Day, S.R. & Brindley, P.J. (1997). Schistosomes express two forms of cathepsin D. Biochim. Biophys. Acta, **1338**: 156-160.

Wong, T.S., Encell, L.P., Levinson, W.E., Crist, M.J., Loomis, A.K., Licato, L.L., Aransdorf, J.J., Sica, N., Pienkos, P.T. & Monticello, D.J. (2004). Growth factor engineering by degenerate homoduplex gene family recombination. Nat. Biotechnol., **20**: 1246-1250.

Yokoi, S., Shigyo, T. & Tamaki, T. (1996). A fluorimetric assay for proteinase A in beer and its application for investigation of enzymatic effects on foam stability. J. Inst. Brewing, **102**: 33-37.

Zhao, B., Winborne, E., Minnich, M.D., Culp, J.S., Debouck, C., & Abdel-Meguid, S.S. (1993). Three-dimensional structure of a simian immunodeficiency virus protease/inhibitor complex. Implications for the design of human immunodeficiency virus type 1 and 2 protease inhibitors. Biochemistry, **32**: 13054-13060.

Zhou, M.Y. & Gomez-Sanchez, C.E. (2000). Universal TA cloning. Curr. Issues. Mol. Biol., **2 (1)**: 1-7.

Appendix I

Composition and structure of wild-type IA3 from *S. cerevisiae* and *S. castellii*

Gene sequences

S. cerevisiae IA3

```
1   ATGAATACAG ACCAACAAAA
21  AGTGAGCGAA ATATTTCAGA
41  GCTCAAAGGA AAAATTGCAG
61  GGCGATGCAA AGGTAGTGAG
81  TGACGCTTTT AAGAAAATGG
101 CTAGTCAAGA CAAGGACGGC
121 AAGACTACCG ATGCTGATGA
141 AAGTGAAAAA CACAACTATC
161 AAGAGCAATA CAACAAGCTC
181 AAAGGGGCGG GGCATAAGAA
201 GGAG
```

S. castellii IA3

```
1   ATGAGTGATA AAAACGCTAA
21  CGTCTCGGAA ATGTTTCAAC
41  AGGCAAAGGA AATGCTTGAA
61  GGTAAAGCTA ATGCAGCTAG
81  TGAAGGAATG AAAAACATGG
101 CCTCCCCTGA AAGTAAGGAA
121 AAAATGAAGA CAAAAGGCCA
141 GGATATGGAA AAGAAAGCTC
161 AAGATTCGTA TAATAAACTA
181 AGAGGTGCTG GGAAGCAAGA
201 ACAAACAAAG GAAAGACAAG
221 ATAAAGATGA ATTGATGGAT
241 CAT
```

Primary structure

S. cerevisiae IA3

```
1   MNTDQQKVSE IFQSSKEKLQ
21  GDAKVVSDAF KKMASQDKDG
41  KTTDADESEK HNYQEQYNKL
61  KGAGHKKE
```

S. castellii IA3

```
1   MSDKNANVSE MFQQAKEMLE
21  GKANAASEGM KNMASPESKE
41  KMKTKGQDME KKAQDSYNKL
61  RGAGKQEQTK ERQDKDELMD
81  H
```

Amino acid composition

	S. cerevisiae			*S. castellii*		
	IA3	**1-34**	**35-68**	**IA3**	**1-34**	**35-81**
A	5	3	2	8	6	2
C	0	0	0	0	0	0
D	7	3	4	6	1	5
E	6	2	4	10	4	6
F	2	2	0	1	1	0
G	4	1	3	5	2	3
H	2	0	2	1	0	1
I	1	1	0	0	0	0
K	13	6	7	14	4	10
L	2	1	1	3	1	2
M	2	2	0	8	5	3
N	3	1	2	5	4	1
P	0	0	0	1	0	1
Q	7	4	3	7	2	5
R	0	0	0	2	0	2
S	6	4	2	6	3	3
T	3	1	2	2	0	2
V	3	3	0	1	1	0
W	0	0	0	0	0	0
Y	2	0	0	1	0	1
hydrophobic (%)	30.9	38.2	23.5	34.6	47.1	25.5
charged (%)	41.1	32.4	50.0	40.7	26.5	51.1
estimated net charge at						
pH 7.0	+0.2	+1.0	-0.8	+0.1	-1.0	+1.1
pH 4.7	+4.6	+2.0	+2.6	+4.7	+0.3	+4.4
pH 3.1	+13.6	+5.4	+8.2	+15.6	+3.7	+11.9

Notes:

1. Overall composition for both IA3 species is also sub-partitioned into the N-terminal (1-34) and C-terminal (35-68 or 35-81) regions.

2. Hydrophobic residues were taken to be: A, C, F, G, I, L, M, P, V, W, Y

3. Estimated net charge calculated assuming the following pK values for side chains:

Arginine	12.5
Aspartic acid	3.9
Glutamic acid	4.3
Histidine	6.0
Lysine	10.5

Appendix II

Oligonucleotides

A. Oligonucleotides for site directed mutagenesis by Quikchange™ PCR (Section 3.2.1)

Mutant No.	Name	F or R	sequence	Template DNA
4	*S. cerevisiae* Y57 truncation	F	CAGGAGCAATACCTCGAGCTCAAAGGGGCG	WT *S. cerevisiae* IA3 (1)-pET22b
		R	CGCCCCTTTGAGCTCGAGGTATTGCTCCTG	
5	*S. cerevisiae* A45 truncation	F	GACTACCGATGCTCTCGAGAGTGAAAAACACAAC	WT *S. cerevisiae* IA3(1) -pET22b
		R	GTTGTGTTTTTCACTCTCGAGAGCATCGGTAGTC	
26	*S. cerevisiae* K18I/L19M/D22I	F	GGAAAAGTTGCAGGGCATCGCAAAGGTAGTGAGTG	K18I/L19M (23)-pET22b
		R	CACTCACTACCTTTGCGATGCCCTGCAACTTTTCC	
27	*S. cerevisiae* K18I/L19M/D22I Y57	F	As for 4	K18I/L19M/D22I (26)-pET22b
		R		
28	*S. cerevisiae* K18I/L19M/D22I A45	F	As for 5	K18I/L19M/D22I (26)-pET22b
		R		
31	*S. cerevisiae* K18R	F	CAGAGCTCAAAGGAAAGGTTGCAGGGCGATGCAAAGG	WT *S. cerevisiae* IA3 (1)-pET22b
		R	CCTTTGCATCGCCCTGCAACCTTTCCTTTGAGCTCTG	
32	*S. cerevisiae* D22E	F	GGAAAAGTTGCAGGGCGAGGCAAAGGTAGTGAGTG	WT *S. cerevisiae* IA3 (1)-pET22b
		R	CACTCACTACCTTTGCCTCGCCCTGCAACTTTTCC	
36	*S. cerevisiae* K18D/D22K	F	CAGAGCTCAAAGGAAGACTTGCAGGGCAAGGCAAAGGTAGTGAG	WT *S. cerevisiae* IA3(1) -pET22b
		R	CTCACTACCTTTGCCTTGCCCTGCAAGTCTTCCTTTGAGCTCTG	
37	*S. cerevisiae* K18E/D22K	F	CAGAGCTCAAAGGAAGAATTGCAGGGCAAGGCAAAGG	K18D/D22K (36)-pET22b
		R	CCTTTGCCTTGCCCTGCAATTCTTCCTTTGAGCTCTG	
38	*S. cerevisiae* K18V/D22K	F	CAGAGCTCAAAGGAAGTGTTGCAGGGCAAGGC	K18D/D22K (36)-pET22b
		R	GCCTTGCCCTGCAACACTTCCTTTGAGCTCTG	
39	*S. cerevisiae* K18M/D22K	F	CAGAGCTCAAAGGAAATGTTGCAGGGCAAGGCAAAGGTAGTGAG	WT *S. cerevisiae* IA3 (1)-pET22b
		R	CTCACTACCTTTGCCTTGCCCTGCAACATTTCCTTTGAGCTCTG	
40	*S. cerevisiae* K18D/D22R	F	GGAAGACTTGCAGGGCAGGGCAAAGGTAGTGAGTG	K18D/D22K (36)-pET22b
		R	CACTCACTACCTTTGCCCTGCCCTGCAAGTCTTCC	

B. Oligonucleotides for overlap-extension PCR

With the exception of the domain swap protein 3AI (Section 3.5) overlap-extension PCR schemes involving mutation of *S. cerevisiae* IA3 used the **T7F** and **R** primers (Appendix II.E) as the gene-flanking primers, F1 and R2 (Figure 3.2). For *S. castellii* IA3-derived mutants, F1 and R2 were *S. castellii* F and R, respectively (Appendix II.E). Thus, all 4 primers for the production of 3AI are detailed below whilst only the "central" primers F2 and R1 (Figure 3.2) are shown for the others.

Mutant No.	Name	F or R	sequence	Template DNA
20	*S. cerevisiae* D22M	F2	GAAAAATTGCAGGGCATGGCAAAGGTAGTGAGTGACGC	WT *S. cerevisiae* IA3 (1)-pET22b
		R1	TGCCATGCCCTGCAATTTTTCCTTTGAGCTCTGAAATATTTCGC	
25	*S. cerevisiae* K18I/D22I	F2	GAAATCTTGCAGGGCATCGCAAAGGTAGTGAGTGACGC	WT *S. cerevisiae* IA3 (1)-pET22b
		R1	TGCGATGCCCTGCAAGATTTCCTTTGAGCTCTGAAATATTTCGC	
33	*S. cerevisiae* K18R/D22E	F2	GAAAGGTTGCAGGGCGAGGCAAAGGTAGTGAGTGACGC	WT *S. cerevisiae* IA3 (1)-pET22b
		R1	TGCCTCGCCCTGCAACCTTTCCTTTGAGCTCTGAAATATTTCGC	
34	*S. cerevisiae* K18D	F2	GAAGATTTGCAGGGCGATGCAAAGGTAGTGAGTGACGC	WT *S. cerevisiae* IA3 (1)-pET22b
		R1	TGCATCGCCCTGCAAATCTTCCTTTGAGCTCTGAAATATTTCGC	
35	*S. cerevisiae* D22K	F2	GAAAAATTGCAGGGCAAAGCAAAGGTAGTGAGTGACGC	WT *S. cerevisiae* IA3 (1)-pET22b
		R1	TGCTTTGCCCTGCAATTTTTCCTTTGAGCTCTGAAATATTTCGC	
41	*S. cerevisiae* K18E/D22R	F2	GAAGAGTTGCAGGGCAGGGCAAAGGTAGTGAGTGACGC	WT *S. cerevisiae* IA3 (1)-pET22b
		R1	TGCCCTGCCCTGCAACTCTTCCTTTGAGCTCTGAAATATTTCGC	
63	*S. castellii* M18K/K22M	F2	GAAAAGCTTGAAGGTATGGCTAATGCAGCTAGTGAAGGAATG	WT *S. castellii* IA3 (42)-pET22b
		R1	AGCCATACCTTCAAGCTTTTCCTTTGCCTGTTGAAACATTTC	
65	*S. castellii* M18K/K22L	F2	GAAAAGCTTGAAGGTCTAGCTAATGCAGCTAGTGAAGGAATG	WT *S. castellii* IA3 (42)-pET22b
		R1	AGCTAGACCTTCAAGCTTTTCCTTTGCCTGTTGAAACATTTC	
66	*S. castellii* M18K/K22D	F2	GAAAAGCTTGAAGGTGACGCTAATGCAGCTAGTGAAGGAATG	WT *S. castellii* IA3 (42)-pET22b
		R1	AGCGTCACCTTCAAGCTTTTCCTTTGCCTGTTGAAACATTTC	
68	*S. castellii* K22L	F2	GAAATGCTTGAAGGTCTAGCTAATGCAGCTAGTGAAGGAATG	WT *S. castellii* IA3 (42)-pET22b
		R1	AGCTAGACCTTCAAGCATTTCCTTTGCCTGTTGAAACATTTC	
79	3AI	F1	ATAAGAAGGAGATGAATACAGACCAACAAAAAGTG	WT *S. cerevisiae* IA3 (1)-pET22b
		R1	CTCGAGCTAAGCCATTTTCTTAAAAGCGTCAC	
		F2	CATATGCATCATCATCATCATCATAGTCAAGACAAGGACGGCAAG	
		R2	GTCTGTATTCATCTCCTTCTTATGCCCCGCCCC	

C. Oligonucleotides for the construction and use of AntiSacJ35-pET22b (Section 3.2.3)

Two site-directed mutations were required to convert *S. cerevisiae* (wild type) IA3 (**1**)-pET22b into AntiSacJ35-pET22b (Section 3.2.3a and Figure 3.4). Both the removal of a *Sac*I site ("antiSac") and introduction of an *Nhe*I site at codon34/35 ("J35") were performed by Quikchange™ site-directed mutagenesis. This process was also performed on a version of the pET22b vector containing the triple mutant *S. cerevisiae* IA3 K18I/L19M/D22I (**26**), resulting in second version of the vector termed "AntiSac**TM**J35-pET22b".

Name	F or R	Sequence	Template DNA
AntiSac	**F**	CGGGGCATAAGAAGGAACTCGAGCACCACCACC	WT *S. cerevisiae* IA3 (**1**)-pET22b
	R	GGTGGTGGTGCTCGAGTTCCTTCTTATGCCCCG	
J35	**F**	GCTTTTAAGAAAATGGCTAGCCAAGACAAGGACGGCAAGAC	AntiSac IA3-pET22b
	R	GTCTTGCCGTCCTTGTCTTGGCTAGCCATTTTCTTAAAAGC	(i.e. product of the antiSac SDM PCR)

The mutants listed below were constructed by annealing the chemically synthesised oligonucleotides described and ligating into AntiSacJ35-pET22b digested with *Sac*I and *Nhe*I (Section 3.2.3b).

Mutant No.	Name	Oligonucleotide name	Sequence
52	***S. cerevisiae*** **V26A/A29G/F30M**	**AGM TOP**	CAAAGGAAAAATTGCAGGGCGATGCAAAGGTAGCTAGTGACGGTATGAAGAAAATGG
		AGM BOTTOM	CTAGCCATTTTCTTCATACCGTCACTAGCTACCTTTGCATCGCCCTGCAATTTTTCCTTTGAGCT
53	***S. cerevisiae*** **Q20E/V26A/A29G/F30M**	**EAGM TOP**	CAAAGGAAAAATTGGAGGGCGATGCAAAGGTAGCTAGTGACGGTATGAAGAAAATGG
		EAGM BOTTOM	CTAGCCATTTTCTTCATACCGTCACTAGCTACCTTTGCATCGCCCTGCAATTTTTCCTTTGAGCT
56	**N*cast Sac*I/*Nhe*I**	**TOP**	CAAAGGAAATGCTTGAAGGTAAAGCTAATGCAGCTAGTGAAGGAATGAAAAACATGG
		BOTTOM	CTAGCCATGTTTTTCATTCCTTCACTAGCTGCATTAGCTTTACCTTCAAGCATTTCCTTTGAGCT

D. *S. cerevisiae-S. castellii* IA3 gene chimaeras

The following oligonucleotides were used in the formation of the chimaeric forms of IA3, using either the AntiSacJ35-pET22b vector (**A**) or the AntiSacTMJ35-pET22b (**B**) vector, as appropriate. The general principle of this scheme is described in Section 3.4.

Mutant No.	Name	F or R	Sequence	Template	Cloning vector	cut with
45	N*cere*-C*cast*	F	CTAGCTAGCCCTGAAAGTAAGGAAAAAATGAAGAC	*S. castellii* IA3 (42)-pET22b	A	*Nhe*I & *Xho*I
		R	CCGCTCGAGATGATCCATCAATTCATCTTTATCTTG			
46	N*cere*TM-C*cast*	F	As for 45	*S. castellii* IA3 (42)-pET22b	B	*Nhe*I & *Xho*I
		R	As for 45			
47	N*cast*-C*cere*	F	GGAATTCCATATGAGTGATAAAAACGCTAACGTC	*S. castellii* IA3 (42)-pET22b	A	*Nde*I & *Nhe*I
		R	CTAGCTAGCCATGTTTTTCATTCCTTCACTAGC			
49	N*cast* M18V-C*cere*	F	As for 47	*S. castellii* M18V IA3 (48)-pET22b	A	*Nde*I & *Nhe*I
			As for 47			

E. Sequencing primers and gene-specific primers

Name	F or R	sequence
T7	F	TAATACGACTCACTATAGGG
	R	GCTAGTTATTGCTCAGGCG
M13	F	GTAAAACGACGGCCAG
	R	CAGGAAACAGCTATGAC
PhageSeq	F	GACCAGATTATCGCTAAGTAC
	R	AACCCCTCAAGACCCGTTTA
S. castellii	F	GGAATTCCATATGAGTGATAAAAACGCTAACGTC
	R	CCGCTCGAGATGATCCATCAATTCATCTTTATCTTG

F. Oligonucleotides to convert pET22b into an N-terminal His-tag vector, and PCR primers to construct IA3 genes for use in this vector

The region between *Xba*I and *Nde*I in the multiple cloning site of pET22b was replaced as described in Section 4.3.1 by a cassette produced from the following oligonucleotides annealed together (see also Figure 4.4):

Name	sequence
newpetTOP	CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATAAGCTTATGAAACACCACCACCACCACCACCA
newpetBOT	TATGGTGGTGGTGGTGGTGGTGTTTCATAAGCTTATCTCCTTCTTAAAGTTAAACAAAATTATTT

The following oligonucleotides were used to produce IA3 constructs compatible with this vector, as described in Section 4.3.2:

Mutant No.	Name	F or R	sequence	Template DNA
2	NT(H)$_6$ *S. cerevisiae* IA3	F	GGAATTCCATATGCAGAATACAGACCAACAAAAAGTGAG	WT *S. cerevisiae* IA3 (1)-pET22b
		R	CCGCTCGAGCGGCTATCTATCTACTCCTTCTTATGCCCCGCC	
43	NT(H)$_6$ *S. castellii* IA3	F	GGAATTCCATATGCAGAGTGATAAAAACGCTAACGTCT	WT *S. castellii* IA3 (42)-pET22b
		R	CCGCTCGAGCGGCTATCTATCTAATGATCCATCAATTCATCTTTATC	
50	NT(H)$_6$ N*cast*-C*cere* IA3	F	As F for 43	N*cast*-C*cere* IA3 (47)-pET22b
		R	As R for 2	
51	NT(H)$_6$ *S. castellii* Q68 IA3	F	As F for 43	WT *S. castellii* IA3 (42)-pET22b
		R	CCGCTCGAGCGGCTATCTATCTATTGTTCTTGCTTCCCAGCACC	
70	NT(H)$_6$ *S. castellii* Y57 IA3	F	As F for 43	WT *S. castellii* IA3 (42)-pET22b
		R	CCGCTCGAGCGGCTATCTATCTAATACGAATCTTGAGCTTTCTTTTC	
71	NT(H)$_6$ *S. castellii* K45 IA3	F	As F for 43	WT *S. castellii* IA3 (42)-pET22b
		R	CCGCTCGAGCGGCTATCTATCTATTTTGTCTTCATTTTTTCCTTACTTTC	
72	NT(H)$_6$ N*cast*-C*cere* Y57 IA3	F	As F for 43	N*cast*-C*cere* IA3 (47)-pET22b
		R	CCGCTCGAGCGGCTATCTATCTAGTATTGCTCCTGATAGTTGTG	
73	NT(H)$_6$ N*cast*-C*cere* K45 IA3	F	As F for 43	N*cast*-C*cere* IA3 (47)-pET22b
		R	CCGCTCGAGCGGCTATCTATCTAAGCATCGGTAGTCTTGCCGTC	
81	NT(H)$_6$ *S. cerevisiae* K18D/D22K IA3	F	As F for 2	K18D/D22K IA3 (36)-pET22b
		R	As R for 2	

G. Oligonucleotides to amplify the ϕ insert for phage display by PCR (Section 10.5.1)

Name	sequence
ϕF	ATATTCGAATTCCGGAGGTGGAAATACAGACCAACAAAAAGTGAG
ϕR	GCGCTTAAGCTTCTCGAGCTAAGCCATTTTCTTAAAAGCGTCA

H. Oligonucleotides for the construction of other mutants

The oligonucleotides required for the construction of the mutants listed below can be found in Phylip *et al.* (2001).

11	***S. cerevisiae* $(Gly)_9$ IA3**
16	***S. cerevisiae* L19M IA3**
21	***S. cerevisiae* D22L IA3**
22	***S. cerevisiae* K16M/K18M IA3**
23	***S. cerevisiae* K18I/L19M IA3**
24	***S. cerevisiae* K18M/D22L IA3**

Appendix III

Inhibitory data summary

A complete list of the K_i values and inhibitors described in this thesis is provided below. Numbering and colouring is consistent with the rest of the thesis i.e. *S. cerevisiae* IA3 (yellow), *S. castellii* IA3-derived sequence (cyan), and other mutations (red). All inhibitory data was collected spectrophotometrically as described in Section 2.2.30, at 37 °C. Residues 1-34 are detailed, the tail region (after residue 34) represented as ~~~~~~~~~ according to its length. For complete sequences of wild-type *S. cerevisiae* and *S. castellii* IA3, see Appendix I. Recombinant proteins were produced with either a C-terminal $LE(H)_6$ tag or an N-terminal $MK(H)_6HMQ$ tag. The N-terminal His-tag could be removed by exopeptidase treatment (Section 4.3.3), exposing an N-terminal pyroglutamate residue (*Q).

Synthetic peptides used the artificial amino acid norleucine (Z) in place of methionine. J = phenylglycine.

× = no apparent interaction with cathepsin D (resembling the red line on Figure 10.1); ✓ = dose-dependent response more closely resembling blue line on Figure 10.1. Where this concentration-dependent inhibition produced a K_i<1000 nM, the value is included in the table. * = from Phylip *et al.*, 2001.

No	Sequence 1 10 20 30	*S. cerevisiae* pH 4.7	*S. cerevisiae* pH 3.1	*P.pastoris* pH 4.7	*A. fumigatus* pH 4.7	Cathepsin D pH 4.7	Cathepsin D pH 3.1
1	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	1.1 ± 0.4 *	55 ± 11	630 ± 110	×	×
2	MK(H)$_7$MQ NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	<0.1	0.6 ± 0.2	0.6 ± 0.2	530 ± 100	×	×
3	*Q NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~	0.1 ± 0.1	0.7 ± 0.3	100 ± 10	>1000	×	×
4	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~LE(H)$_6$[57]	<0.1	1.2 ± 0.4	225 ± 25			
5	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~LE(H)$_6$[45]	<0.1	2 ± 0.4	200 ± 20			
6	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKZA	<0.1 *	3 ± 0.6 *	NI at 2 µM	1500 ± 300		
7	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KK	2 ± 0.4	15 ± 4 *				
8	NTDQ QKVSE IFQAS KEKLQ GDAKV VTDAF KK	5 ± 0.4	45 ± 4				
9	NAGQ AKVSE IFQAS KEKLE GDAKV VADAF KK	35 ± 3					
10	NTDQ QPVSE IFQSS KEKLQ GDAKV VSDAF KK	7 ± 1					
11	MGGGG GGGGE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	1 ± 0.7 *	40 ± 10 *				
12	KVSE IFQSS KEKLQ GDAKV VSDAF KKZA	30 ± 8 *	85 ± 20 *				
13	NTDQ QKVSE IFQSS KEKLQ GDAKV PSDAF KK	125 ± 20					
14	NTDQ QKVSE IFQSS KEKLQ GDAKV V	NI *					
15	NTDQ QPVSE IFQSS KEKLQ GDAKV PSDAF KK	NI					
16	MNTDQ QKVSE IFQSS KEKMQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	nd	6 ± 1.2 *				
17	NTDQ QKVSE IFQSS KEKZQ GDAKV VSDAF KK	6 ± 1					
18	NTDQ QKVSE IFQSS KEKCQ GDAKV VSDAF KK	9 ± 1					
19	NTDQ QKVSE IFQSS KEKJQ GDAKV VSDAF KK	700 ± 120 *	NI *				
20	MNTDQ QKVSE IFQSS KEKLQ GJAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	2 ± 0.3	15 ± 3	720 ± 250	×	×
21	MNTDQ QKVSE IFQSS KEKLQ GJAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.6 ± 0.3 *	8 ± 0.2 *				
22	MNTDQ QKVSE IFQSS KEMLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1 *	<0.2 *				
23	MNTDQ QKVSE IFQSS KEIMQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1 *	0.5 ± 0.3 *	15 ± 3			
24	MNTDQ QKVSE IFQSS KEMLQ GJAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1 *	<0.1 *	0.5 ± 0.5	35 ± 5		
25	MNTDQ QKVSE IFQSS KEILQ GIAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	<0.1	< 0.1	9 ± 1	×	×
26	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	2 ± 0.3	0.6 ± 0.15	13 ± 2	×	×
27	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~~~~LE(H)$_6$[57]	<0.1	2 ± 0.3	0.2 ± 0.2			
28	MNTDQ QKVSE IFQSS KEIMQ GIAKV VSDAF KKMA~~LE(H)$_6$[45]	<0.1	2 ± 0.3	0.2 ± 0.2			
29	NTDQ QKVSE IFQSS KEIZQ GIAKV VSDAF KK	0.6 ± 0.1	5 ± 1	6 ± 0.5	22 ± 4	×	×
30	NTDQ QKVSE IFQSS KEZLQ GIAKV VSDAF KK	0.5 ± 0.1	1.5 ± 0.3	3 ± 0.5	40 ± 5	✓	×

No	Sequence (1 10 20 30)	*S. cerevisiae* pH 4.7	*S. cerevisiae* pH 3.1	*P.pastoris* pH 4.7	*A. fumigatus* pH 4.7	Cathepsin D pH 4.7	Cathepsin D pH 3.1
31	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	0.5 ± 0.1	6 ± 2	620 ± 60	×	×
32	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	3 ± 0.3	20 ± 2	NI at 10 µM	×	×
33	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	<0.1	1.1 ± 0.3	5 ± 0.3	350 ± 30	×	×
34	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.3 ± 0.1	6 ± 0.8				
35	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	50 ± 3	NI				
36	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.3 ± 0.1	120 ± 20	300 ± 50	NI at 10 µM	×	×
37	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	1.3 ± 0.2	610 ± 120	IC50 ~ 730	NI at 12 µM	×	×
38	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	2 ± 0.4	1000 ± 200	35 ± 2	NI at 9 µM	×	×
39	MNTDQ QKVSE IFQSS KEMLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.2 ± 0.1	85 ± 15	3 ± 0.9	> 5 µM	×	×
40	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	5 ± 0.8	800 ± 300	IC50 > 500	NI at 9 µM	×	×
41	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~LE(H)$_6$	0.7 ± 0.3	280 ± 30	160 ± 15	~ 5 µM	×	×
42	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	4 ± 0.5	350 ± 70	15 ± 3	600 ± 50	✓ 800 ± 100	×
43	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	0.3 ± 0.2	125 ± 30	0.1 ± 0.1	170 ± 25	✓ < 1000	×
44	SDKN ANVSE ZFQQA KEZLE GKANA ASEGZ KNZA	0.4 ± 0.1	8 ± 1	4 ± 0.5	17 ± 2	✓ 15 ± 3	✓
45	MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA~~~~~~~~~~~~LE(H)$_6$	<0.1	4 ± 0.6	100 ± 20	1200 ± 100	×	×
46	MNTDQ QKVSE IFQSS KEMLQ GDAKV VSDAF KKMA~~~~~~~~~~~~LE(H)$_6$	<0.1	3 ± 0.4	1.5 ± 0.3	12 ± 3	×	×
47	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	0.1 ± 0.1	180 ± 40	3 ± 0.5	340 ± 80	✓ 850 ± 350	×
48	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	125 ± 20		IC50 >1000		×	×
49	MSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	25 ± 2		45 ± 4		×	×
50	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~[68]	0.8 ± 0.1	910 ± 80	0.7 ± 0.1	800 ± 150	✓	×
51	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	0.7 ± 0.2	190 ± 80	0.4 ± 0.1	680 ± 100	✓ 650 ± 70	✓
52	MNTDQ QKVSE IFQSS KEKLQ GDAKV ASDGM KKMA~~~~~~~~LE(H)$_6$	<0.1	30 ± 4	20 ± 2	NI at 9 µM	×	×
53	MNTDQ QKVSE IFQSS KEKLE GDAKV ASDGM KKMA~~~~~~~~LE(H)$_6$	<0.1	15 ± 3	7 ± 0.9	IC50 > 10 µM	×	×
54	NTDQ QKVSE IFQSS KEKLQ GDAKV ASDAF	6 ± 1 *	10 ± 2 *				
55	NTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF	5 ± 1 *	25 ± 8 *				
56	MNTDQ QKVSE IFQSS KEMLE GKANA ASEGM KNMA~~~~~~~~LE(H)$_6$	2 ± 0.3	500 ± 100	3 ± 0.2	280 ± 50	×	×
57	SDKN ANVSE ZFQQA KEZLE GKANA ASEAF KNZA	3 ± 0.4	65 ± 8	9 ± 0.9	120 ± 20	✓ 130 ± 20	✓
58	SDKN ANVSE ZFQQA KEZLE GKANV ASEAF KNZA	1.1 ± 0.1	22 ± 1.1	4 ± 0.4	25 ± 3	✓ 360 ± 70	×
59	SDKN ANVSE ZFQQA KEZLE GKANV ASEGZ KNZA	1.2 ± 0.3	26 ± 1	6 ± 0.4	130 ± 8	✓ 270 ± 50	×
60	SDKN ANVSE ZFQQA KEZLE GKANI ASEAF KNZA	1.4 ± 0.2	26 ± 2	3 ± 0.2	15 ± 3	✓ 250 ± 40	×

III.iii

No	Sequence 1 10 20 30	*S. cerevisiae* pH 4.7	*S. cerevisiae* pH 3.1	*P.pastoris* pH 4.7	*A. fumigatus* pH 4.7	Cathepsin D pH 4.7	Cathepsin D pH 3.1
61	SDKN ANVSE ZFQQA KEZLE GZAKV VSDAF KKZA	0.8 ± 0.1	25 ± 4	15 ± 5	230 ± 50	✓	×
62	NTDQ QKVSE IFQSS KEDLQ GDAKV VSDAF KK	80 ± 7	>2000			×	×
63	MSDKN ANVSE MFQQA KEKLE GMANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	370 ± 50	NI	IC50 >500	> 5 µM	×	×
64	SDKN ANVSE ZFQQA KEKLE GZANA ASEGZ KN	350 ± 100	550 ± 50	IC50 >500	> 5 µM	✓	×
65	MSDKN ANVSE MFQQA KEKLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	1000 ± 200	NI	IC50 >800	> 5 µM	×	×
66	MSDKN ANVSE MFQQA KEKLE GDANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	3 ± 0.8	50 ± 15	IC50 ≈ 60	> 5 µM	×	×
67	SDKN ANVSE ZFQQA KEKLE GDANA ASEGZ KN	10 ± 1.5	35 ± 6	280 ± 30	> 5 µM	×	×
68	MSDKN ANVSE MFQQA KEMLE GLANA ASEGM KNMA~~~~~~~~~~~~LE(H)$_6$	1.1 ± 0.2	13 ± 3	2 ± 0.4	75 ± 15	✓	×
69	SDKN ANVSE ZFQQA KEZLE GLANA ASEGZ KN	2.5 ± 0.5	3 ± 0.8	3 ± 0.4	140 ± 10	✓	✓
70	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~[57]	2 ± 0.3	750 ± 70	1.1 ± 0.1	270 ± 25	✓	×
71	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	0.9 ± 0.4	680 ± 70	0.9 ± 0.1	150 ± 25	✓ 750 ± 100	×
72	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~[57]	3 ± 0.3	460 ± 50	2 ± 0.2	1800 ± 250	✓ 850 ± 350	×
73	MK(H)$_7$MQ SDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	2 ± 0.6	570 ± 50	2 ± 0.2	1750 ± 200	✓	×
74	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~	1.5 ± 0.2	240 ± 20	15 ± 1	550 ± 80	✓ 280 ± 25	✓
75	*QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~~~~~~~~~	4 ± 0.3	150 ± 20	30 ± 4	1900 ± 200	✓ 530 ± 50	×
76	°QSDKN ANVSE MFQQA KEMLE GKANA ASEGM KNMA~~~~[45]	10 ± 1.5	190 ± 30	30 ± 10	320 ± 60	✓ 300 ± 40	✓
77	Synthetic peptide of residues 35-68 of *S. cerevisiae* IA3	NI		NI at 2 µM			
78	Synthetic peptide of residues 35-68 of *S. cerevisiae* IA3 + LE(H)$_6$	NI		NI at 2 µM			
79	M(H)$_6$~~~~~~~~~MNTDQ QKVSE IFQSS KEKLQ GDAKV VSDAF KKMA (see Section 3.5)	0.1 ± 0.1		200 ± 20			
80	NTDQ QKVSE IFQSS KEZLQ GZAKV VSDAF KK	15 ± 3		IC50 >320	> 5 µM	×	×
81	MK(H)$_7$MQ NTDQ QKVSE IFQSS KEDLQ GDAKV VSDAF KKMA~~~~~~~~	<0.1		25 ± 2			
82	*QNTDQ QKVSE IFQSS KEDLQ GNAKV VSDAF KKMA~~~~~~~~	0.6 ± 0.2		>1000			
TJW3	NTDQ QKVSE IFQSS KLKLQ GDAKV VSDAF KK	13 ± 2					
TJW5	NTDQ QKVSE IFQSS KEJJQ GDAKV VSDAF KK	NI					
TJW6	NTDQ QKVSE IFQSS KIJJQ GDAKV VSDAF KK	80 ± 10					
TJW7	NTDQ QKVSE IFQSS KIJZQ GDAKV VSDAF KK	35 ± 5					
TJW8	NTDQ QKVSE IFQSS KLJJQ GDAKV VSDAF KK	625 ± 100					
TJW12	NTDQ QKVSE IFQSS KEJJQ GIAKV VSDAF KK	NI					
TJW13	NTDQ QKVSE IFQSS KIJJQ GIAKV VSDAF KK	NI					
TJW14	NTDQ QKVSE IFQSS KIJZQ GIAKV VSDAF KK	1.7 ± 0.4	8 ± 1			×	×
TJW15	NTDQ QKVSE IFQSS KLJJQ GIAKV VSDAF KK	NI					

PRIFYSGOL CAERDYDD CARDIFF UNIVERSITY