

**Zymogenaktivierung in humanpathogenen Pilzen**

**Charakterisierung des Substratspektrums regulatorischer Kex2-ähnlicher Proteasen**

**Dissertation zur Erlangung des akademischen Grades des  
Doktors der Naturwissenschaften (Dr. rer. nat.)**

**eingereicht im Fachbereich Biologie, Chemie, Pharmazie  
der Freien Universität Berlin**

**vorgelegt von**

**Oliver Bader  
aus Hamburg**

**April 2008**

- 
1. Gutachter: Prof. Dr. Bernhard Hube
  2. Gutachter: Prof. Dr. Rupert Mutzel

Disputation am 27.6.2008

---

<b>1 EINLEITUNG.....</b>	<b>7</b>
1.1 CANDIDA ALBICANS.....	7
1.2 CANDIDA GLABRATA.....	8
1.3 DIE PROTEASE KEX2.....	8
1.4 FUNKTION EUKARYONTISCHER PROHORMONKONVERTASEN.....	10
1.5 BEDEUTUNG VON KEX2 BEI PATHOGENESE- UND VIRULENZMECHANISMEN IN CANDIDA ALBICANS UND ANDEREN PATHOGENEN PILZEN.....	12
1.6 BIOCHEMISCHE UND GENETISCHE METHODEN ZUR ANALYSE DER SUBSTRATSPEZIFITÄT VON PROTEASEN	13
1.7 SUBSTRATSPEZIFITÄT DER PROTEASE KEX2.....	14
1.7.1 UNTERSUCHUNGEN MIT SYSTEMATISCHEN SUBSTRATEN	15
1.7.2 UNTERSUCHUNGEN MITTELS RÖNTGENBEUGUNG AN KEX2 UND FURIN	17
1.7.3 UNTERSUCHUNGEN MITTELS MUTAGENESE VON KEX2	20
1.8 IN SILICO METHODEN DER VORHERSAGE VON PROTEASESCHNITTSTELLEN.....	20
1.9 STEREOCHEMISCHE VORAUSSETZUNGEN IM SUBSTRAT FÜR DIE PROTEOLYSE NATIVER PROTEINE.....	21
1.10 FUNKTION UND STRUKTUR DES KEX2-ENTHALTENDEN ZELLKOMPARTIMENTS.....	21
<b>2 ZIEL DIESER ARBEIT.....</b>	<b>23</b>
<b>3 MATERIALIEN UND METHODEN.....</b>	<b>24</b>
3.1 ORGANISMEN UND NÄHRMEDIEN.....	24
3.1.1 BAKTERIEN	24
3.1.1.1 Stämme.....	24
3.1.1.2 Medien.....	24
3.1.2 PILZE	25
3.1.2.1 Stämme.....	25
3.1.2.2 Medien.....	25
3.2 MOLEKULARBIOLOGISCHE METHODEN.....	25
3.2.1 ISOLATION VON NUKLEINSÄUREN	25
3.2.1.1 Isolation genomischer DNA aus Pilzen.....	25
3.2.1.2 Isolation von Gesamt-RNA aus Pilzen.....	26
3.2.1.3 Isolation von Plasmid-DNA aus Bakterien.....	26
3.2.2 MANIPULATION VON NUKLEINSÄUREN	27
3.2.2.1 Reverse Transkription von mRNA zu cDNA.....	27
3.2.2.2 Sequenzierungen.....	27
3.2.2.3 Klonierung von PCR-Produkten in TOPO®-Vektoren.....	27
3.2.2.4 Vektorklonierung.....	27
3.2.2.5 Herstellung und Transformation dauerkompetenter Escherichia coli Zellen.....	28
3.2.2.6 Kolonienhybridisierung.....	28
3.2.3 VERWENDETE OLIGONUKLEOTIDE	30
3.3 PROTEINBIOCHEMISCHE METHODEN.....	33
3.3.1 SDS-PAGE UND WESTERNBLOT	33
3.3.1.1 Herstellung von Polyacrylamidgelen.....	33
3.3.1.2 Coomassiefärbung.....	33
3.3.1.3 Silberfärbung.....	33
3.3.1.4 Westernblot.....	34
3.3.1.5 Größen-Ausschluss-Chromatographie.....	35
3.3.2 ANIENEN-AUSTAUSCH-CHROMATOGRAPHIE	35
3.3.3 HETEROLOGE EXPRESSION VON PROTEINEN IN E. COLI	35
3.3.3.1 Expression.....	35

3.3.3.2	Aufreinigung von löslichen Proteinen mit His-tag.....	36
3.3.3.3	Aufreinigung von löslichen Proteinen mit GST-tag.....	36
3.3.3.4	Renaturierung von unlöslichen Proteinen aus Inklusionskörpern.....	36
	Aufreinigung von Protein aus Inklusionskörpern.....	36
	Renaturierung mittels Harnstoff.....	36
	Renaturierung mittels TritonX-100 und $\beta$ -Cyclodextrin.....	37
3.3.4	TRANSFORMATION VON PILZEN MIT REKOMBINANTER DNA.....	37
3.3.5	EXPRESSION REKOMBINANTER PROTEINE IN PICHIA PASTORIS.....	37
3.3.5.1	Herstellung und Transformation der Expressionskonstrukte.....	37
3.3.5.2	Testexpression.....	38
3.3.6	MESSUNG DER PHOSPHATASEAKTIVITÄT.....	38
3.3.7	MESSUNG DER KEX2-AKTIVITÄT.....	39
3.3.8	PROTEOLYTISCHER VERDAU MIT KEX2-HOMOLOGEN PROTEASEN.....	39
<b>3.4</b>	<b>STATISTISCHE VERFAHREN UND DATENBANKEN.....</b>	<b>39</b>
3.4.1	MATTHEWS-KORRELATION.....	39
3.4.2	RECEIVER OPERATING CHARACTERISTIC KURVE.....	40
3.4.3	SEPARIERENDE HYPEREBENEN/SUPPORT VEKTOREN.....	40
3.4.4	VERWENDETE SOFTWARE UND GENOMDATENBANKEN.....	41
<b>4</b>	<b>ERGEBNISSE.....</b>	<b>42</b>
<b>4.1</b>	<b>BEKANNTE SUBSTRATE KEX2-HOMOLOGER PROTEASEN IN PILZEN.....</b>	<b>42</b>
<b>4.2</b>	<b>SLoPF-DB: DATENBANK FÜR SUBZELLULÄRE LOKALISATION VON PROTEINEN AUS FUNGL.....</b>	<b>45</b>
4.2.1	VORHERSAGE DER LOKALISATION VON PROTEINEN.....	45
4.2.1.1	Erstellung eines Validierungs- und Testsets.....	47
4.2.1.2	Zuordnung aufgrund von Lokalisationssequenzen.....	48
4.2.1.3	Zuordnung aufgrund der Aminosäurezusammensetzung.....	49
4.2.1.4	Zuordnung aufgrund von Homologien zu Proteinen bekannter Lokalisation.....	51
4.2.2	GRUPPIERUNG ORTHOLOGER PROTEINE.....	52
4.2.3	VALIDIERUNG DER LOKALISATIONSVORHERSAGE IN SLoPF-DB.....	55
<b>4.3</b>	<b>VERTEILUNG VON POTENTIELLEN KEX2-SCHNITTSTELLEN IM SEKRETORISCHEN PROTEOM.....</b>	<b>57</b>
4.3.1	VORHERGESAGTE SUBSTRATE VON KEX2 BEI C. ALBICANS, C. GLABRATA UND S. CEREVISIAE.....	59
<b>4.4</b>	<b>IN VITRO ÜBERPRÜFUNG EINIGER PUTATIVER SUBSTRATE.....</b>	<b>71</b>
4.4.1	KLONIERUNG EINES KEX2-HOMOLOGEN GENS AUS CRYPTOCOCCUS NEOFORMANS.....	71
4.4.2	HETEROLOGE EXPRESSION UND AUFREINIGUNG KEX2-ÄHNLICHER PROTEASEN.....	72
4.4.2.1	Heterologe Expression in Pichia pastoris.....	72
	Candida glabrata Kex2.....	73
	Candida albicans Kex2.....	74
	Cryptococcus neoformans Kex2.....	75
4.4.2.2	Heterologe Expression von CaKex2 in Candida albicans.....	76
4.4.2.3	Aufreinigung heterolog exprimierter Kex2 Proteasen aus S. cerevisiae, P. pastoris, C. glabrata und C. albicans.....	77
4.4.3	KEX2 IST BEI UNPHYSIOLOGISCH HOHEN TEMPERATUREN AM AKTIVSTEN.....	78
4.4.4	EXPRESSION VON SUBSTRATPROTEINEN IN E. COLI.....	79
4.4.5	DURCH RENATURIERUNG LASSEN SICH NATIV GEFALTETE, ENZYMATISCH AKTIVE PROTEINE HERSTELLEN.....	81
4.4.6	IN VITRO PROZESSIERUNG DER PUTATIVEN KEX2-SUBSTRATE.....	82
4.4.6.1	Die $\alpha$ -Pheromone von S. cerevisiae, C. glabrata und C. albicans werden in ähnlicher Weise prozessiert.....	82
4.4.6.2	CaEce1 ist ein Polypeptidvorläufer.....	83
4.4.6.3	CA0365 wird von den Proteasen unterschiedlich prozessiert.....	84
4.4.6.4	CA0365 und Ece1 könnten Strukturmoleküle ähnlich UmRep1 sein.....	85
4.4.6.5	Proteine mit Ops4-ähnlicher Struktur werden an mehreren Stellen prozessiert.....	87
4.4.6.6	Pir-Proteine werden von Kex2 prozessiert.....	89
4.4.6.7	Proteine der Sun/Scw4 Familie, nicht aber renaturiertes Tos1, werden an einer konservierten histidinreichen Struktur prozessiert.....	91
4.4.6.8	Manche Proteine der Pry-Proteinfamilie besitzen glykosylierte Propeptide.....	92

4.4.6.9 Auch Proteine aus dem ER tragen prozessierbare Schnittstellen.....	94
4.4.6.10 Einige Proteine mit potentiellen Kex2 Schnittstellen werden nicht prozessiert.....	97
CA0375/Crh11.....	97
CAGL0J11770g/Plb1.....	97
CA5147/Pho114.....	98
4.4.7 UNTERSUCHUNG VON PROTEINEN OHNE HOMOLOGE BEKANNTER STRUKTUR	99
4.4.8 IN PROZESSIERTEN SCHNITTSTELLEN SIND NEGATIVE LADUNGEN AN DEN POSITIONEN P1'-P4' ÜBERREPRÄSENTIERT	100
<b>4.5 VERGLEICH DER STRUKTUR VON KEXINEN AUS PILZEN UND SÄUGERN.....</b>	<b>103</b>
4.5.1 IDENTIFIZIERUNG VON KEX2-HOMOLOGEN GENEN IN GENOMDATEN VERSCHIEDENER PILZE	103
4.5.2 IN SILICO ANALYSE DER P'-SUBSTRATBINDUNGSSTELLEN VON PROHORMONKONVERTASEN	104
4.5.2.1 In silico Analyse der S1'-S4' Substratbindungstaschen von Kex2 und Furin im Vergleich zu Kumamolisin.....	105
4.5.2.2 Besetzung der an der Substraterkennung beteiligten Positionen bei Kex2 im Vergleich zu anderen Prohormonkonvertasen.....	111
<b>4.6 STRUKTURELLE DETERMINANTEN DER PROTEASESUSZEPTIBILITÄT IM SUBSTRAT.....</b>	<b>114</b>
4.6.1 DIE AUSGEBILDETE STRUKTUR DER SCHNITTSTELLE BEEINFLUSST DIE PROZESSIERUNG	114
4.6.2 BERECHNUNG DER VERBESSERUNG VON VORHERSAGEALGORITHMEN DURCH BERÜCKSICHTIGUNG STRUKTURELLER DETERMINANTEN	116
<b><u>5 DISKUSSION.....</u></b>	<b><u>124</u></b>
<b>5.1 ENTWICKLUNG EINER NEUEN METHODE ZUR VORHERSAGE SUBZELLULÄRER PROTEINLOKALISATION....</b>	<b>125</b>
<b>5.2 SCHNITTSTELLEN IM GESAMT- UND SEKRETORISCHEN PROTEOM.....</b>	<b>127</b>
<b>5.3 SUBSTRATERKENNUNG AN DER P1'-P4'-REGION.....</b>	<b>128</b>
<b>5.4 DIE BIOCHEMISCHEN DATEN VON S. CEREVISIAE KEX2 SIND NUR BEDINGT AUF DIE KEX2-PROTEINE ANDERER PILZE ÜBERTRAGBAR.....</b>	<b>129</b>
<b>5.5 DIE DREIDIMENSIONALE STRUKTUR DES SUBSTRATS BEEINFLUSST SEINE PROZESSIERBARKEIT.....</b>	<b>131</b>
5.5.1 SCHNITTSTELLEN PROZESSIERTER PROTEINE SIND EXPONIERT	133
5.5.2 UNPROZESSIERTE MOTIVE SIND FÜR DIE PROTEASEN NICHT ERREICHBAR	134
5.5.3 COMPUTERMODELL ZUR VERBESSERTEN VORHERSAGE	136
<b>5.6 KORRELATION VON PHÄNOTYPEN MIT SUBSTRATEN.....</b>	<b>138</b>
5.6.1 ZELLWAND ASSOZIIERTE PHÄNOTYPEN DER KEX2-DELETIONSMUTANTEN	139
5.6.2 MANCHE PROTEINE DER PRY-PROTEINFAMILIE BESITZEN GLYKOSYLIERTE POTENTIELLE PROPEPTIDE	141
5.6.3 CAEce1 UND CA0365 SIND POLYPEPTIDVORLÄUFER MIT ÄHNLICHKEIT ZU UmREP1	142
5.6.4 MATINGASSOZIIERTE PHÄNOTYPEN DER KEX2-DELETIONSMUTANTEN	143
5.6.5 AUCH PROTEINE AUS DEM ER TRAGEN PROZESSIERBARE SCHNITTSTELLEN	144
<b>5.7 RELEVANZ DER IN VITRO DATEN FÜR DIE IN VIVO SITUATION.....</b>	<b>145</b>
<b><u>6 ZUSAMMENFASSUNG.....</u></b>	<b><u>147</u></b>
<b><u>7 SUMMARY.....</u></b>	<b><u>149</u></b>
<b><u>8 DANKSAGUNG.....</u></b>	<b><u>151</u></b>
<b><u>9 LITERATUR.....</u></b>	<b><u>152</u></b>
<b><u>10 PUBLIKATIONSLISTE.....</u></b>	<b><u>161</u></b>

---

<b>11 ANHANG.....</b>	<b>162</b>
<b>11.1 GENE BANK EINTRAG FÜR CRYPTOCOCCUS NEOFORMANS KEX2 .....</b>	<b>162</b>
<b>11.2 BEWERTUNG VON KR-MOTIVEN.....</b>	<b>164</b>

## 1 Einleitung

Weltweit sind etwa 300.000 Pilzarten bekannt. Neben biotechnologisch relevanten Arten, wie etwa *Saccharomyces cerevisiae*, *Pichia pastoris* oder verschiedene Schimmelpilze, die zur industriellen Naturstoffproduktion eingesetzt werden, stellen ca. 150 Arten Krankheitserreger des Menschen dar [1]. Ein großer Anteil der systemischen humanen Pilzinfektionen in Deutschland (ca. 50-60%) lässt sich auf Stämme der Gattung *Candida* zurückführen [2]. Von den rund 200 bekannten *Candida*-Arten sind wiederum nur wenige medizinisch relevant: über 50% aller diagnostizierten *Candida*-Infektionen werden von *C. albicans* verursacht. Daneben sind *C. glabrata*, *C. parapsilosis*, *C. tropicalis*, *C. krusei* und *C. dubliniensis* für *Candida*-Infektionen beim Menschen verantwortlich [2-4].

### 1.1 *Candida albicans*

*Candida*-Arten gehören als *Ascomycota* (Schlauchpilze) in die Klasse der *Saccharomycetales*. *C. albicans* ist durch seine heterotrophe, fakultativ anaerobe Lebensweise, die Nutzung verschiedener Kohlenstoff- und Stickstoffverbindungen zur Energiegewinnung und die Fähigkeit bei 37°C wachsen zu können gut an den menschlichen Körper als Habitat angepasst.

*C. albicans* ist ein pleiomorpher Pilz: in seiner einzelligen Wachstumsform bildet *C. albicans* ellipsoide Hefezellen. Diese können sich durch Sprossung vermehren, wobei kleine, vegetative Tochterzellen aus der Mutterzelle herauswachsen und sich als neue Hefezellen abnabeln. Auf externe Reize hin kann *C. albicans* seine Gestalt verändern: *in vitro* lässt sich durch Erhöhung der Temperatur auf 37-41°C und einem pH-Wert von 6,5-7,0 oder durch Anreicherung des Kulturmediums mit Blutserum aus der ellipsoiden Hefeform eine filamentöse Hyphenform erzeugen. Ebenso wird die Hyphenbildung in nährstoffarmem Medium induziert [5]. Dieser morphologische Übergang geht mit einem starken Umbau der Zellwand und der Expression von dafür spezifischen Genen einher [6].

*C. albicans* wird als obligat diploider Organismus betrachtet, der keinen sexuellen Lebenszyklus besitzt. Erst 1999 zeigten Genomsequenzierungen, dass *C. albicans* einen „mating type“ ähnlichen Locus (MTL) besitzt [7]. Weitere Studien zeigten, dass „Mating“, also Paarung, bei *C. albicans* tatsächlich möglich ist, jedoch natürlicherweise nur selten vorkommt [8-10].

Bei einigen *C. albicans* Stämmen kann zusätzlich zu dem von äußeren Bedingungen induziertem Wechsel zwischen Hefe und filamentöser Form, ein spontaner und reversibler Wechsel zwischen der normalen ellipsoiden Hefezellform und größeren, verlängerten Zellen auftreten. Dieses Phänomen wird aufgrund des Aussehens der Kolonien, welche von den beiden Zellformen gebildet werden, als „white-opaque switching“ bezeichnet [11]. Die Fähigkeit zur Paarung

korreliert bei *C. albicans* mit dem der „opaque“-Form, welches die paarungs-kompetente Zellform darstellt. Die meisten *C. albicans*-Stämme, einschließlich SC5314, sind heterozygot für den MTL, was letztlich zur Repression des switching von „white“- zu „opaque“-Zellen führt. Stämme, die homozygot für *MTLa* oder *MTL $\alpha$*  geworden sind, können hingegen in die „opaque“-Phase übergehen und als a- oder  $\alpha$ - Zellen mit „opaque“-Zellen des anderen Paarungstyps verschmelzen [8, 12].

## 1.2 *Candida glabrata*

Der humanpathogene Pilz *Candida glabrata* ist ebenfalls ein weit verbreiteter Ascomycet [13], der allerdings phylogenetisch näher an *S. cerevisiae* als an *C. albicans* steht. Im Gegensatz zu *C. albicans* fehlt ihm die Fähigkeit zur Ausbildung von Hyphen unter *in vivo* Bedingungen und er besitzt ein haploides Genom. Bevorzugte Habitate sind neben faulenden Früchten auch der menschliche Gastrointestinaltrakt und die Mucosa [14] [15].

Diese Spezies ist nach *C. albicans* der zweithäufigste Pilz, der aus Blutkulturen isoliert werden kann [2, 16, 17]. Eine Schwächung des Immunsystems, wie sie beispielsweise bei AIDS aber auch bei Leukämie und Neutropenie auftritt oder medikamentös bei Organtransplantationen induziert wird, ist wie bei *C. albicans* häufig Voraussetzung für die Entwicklung einer systemischen *C. glabrata* Infektion [18]. Das Erregerreservoir bilden dabei hauptsächlich endogene Stämme der Normalflora des Gastrointestinaltrakts. Ein sexueller Zyklus ist bei *C. glabrata*, trotz eines vorhandenen MTL [19], noch nicht gezeigt worden.

*C. glabrata* zeigt eine geringere *in vivo* Empfindlichkeit gegenüber Azolderivaten als andere *Candida* Arten, insbesondere den lange Zeit gebräuchlichsten Antimykotika Itraconazol und Fluconazol, was die Therapie erschwert [2, 16]. Hinzu kommt die Fähigkeit zur schnellen Entwicklung sekundärer Resistenzen. Bis zu 18% der klinischen *C. glabrata* Isolate aus Blutkultur zeigen eine Resistenz gegenüber Fluconazol [2, 16, 20].

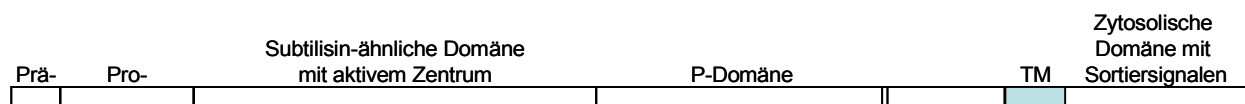
## 1.3 Die Protease Kex2

Die Flexibilität, äußere Gestalt und Interaktion mit der Umwelt der Pilzzelle hängt direkt von einer großen Bandbreite sekretorischer Proteine ab. Diese werden teilweise von der regulatorischen Protease Kex2 prozessiert. Kex2 ist eine Endoprotease aus der Familie der Subtilisin-ähnlichen Serinproteasen. Da jedoch einige deutliche genetische Unterschiede zu anderen Proteasen dieser Familie bestanden, wurde es das namensgebende Protein für die Untergruppe der „Kexine“. Es sind mittlerweile 28 verschiedene Familien von Serinproteasen in der MEROPS Datenbank für Proteasen klassifiziert worden [21]. Ihnen allen ist gemein, dass sich im katalytischen Zentrum eine „Triade“ aus den drei konservierten Aminosäuren Serin

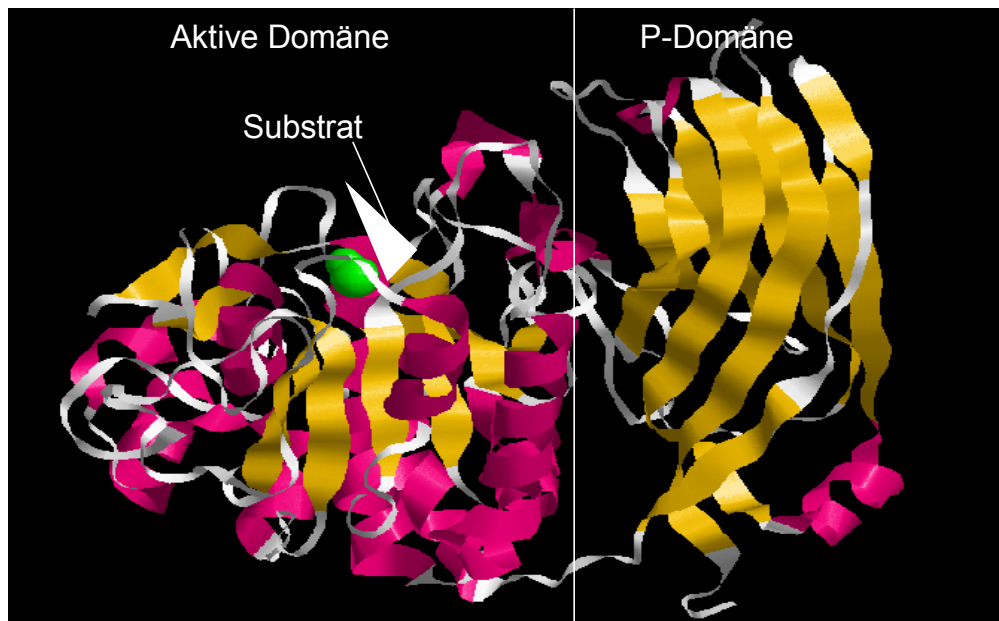


(nukleophil), Aspartat (elektrophil) und Histidin (nukleophil) befinden. Die Bezeichnung „Serinproteasen“ stammt daher, dass das katalytische Serin ungewöhnlich reaktiv ist, was anhand von Inhibitionsstudien bei Chymotrypsin gezeigt wurde [22]. Das Serin im aktiven Zentrum lässt sich durch organische Difluorophosphate (DFP) inhibieren, welche dort kovalent gebunden werden. Die starke Reaktivität des Serins im aktiven Zentrum des Enzyms wird daran deutlich, dass keines der weiteren Serine im Protein DFP bindet [23]. Auf dieses aktive Serin wird während der Proteolyse der N-terminale Teil des zu spaltenden Peptids transient übertragen. Kex2 besitzt eine Substratspezifität für dibasische Motive, die meistens aus einem Lysin und einem Arginin bestehen [24]. Die Proteolyse erfolgt direkt hinter dem Lysin-Arginin-Motiv (KR-Motiv). Das Arginin ist essentiell, das Lysin kann durch andere Aminosäuren ersetzt werden, dabei nimmt allerdings die Effizienz der Proteolyse stark ab [24]. Eine Schnittstelle aus zwei Argininen wird nur noch  $\frac{1}{12}$ -mal so effektiv hydrolysiert wie das KR-Motiv (Abschnitt 1.7). Bei Ersatz des Lysins durch andere Aminosäuren als Arginin erfolgt fast keine Hydrolyse mehr ([24] siehe Abschnitt 1.7).

Kex2 ist ein Transmembranprotein der Klasse I. D. h. es besitzt ein Signalpeptid, welches den N-Terminus in das ER-Lumen dirigiert, eine Transmembrandomäne und einen im Zytosol gelegenen C-Terminus. Die Synthese erfolgt als Prä-Pro-Protein in das Endoplasmatische Retikulum (ER) hinein. Die Abspaltung des Prä-Peptids erfolgt co-translational durch den Signalpeptidasekomplex, die des Pro-Peptides autokatalytisch [25, 26].



**Abbildung 1: Schematischer Aufbau der Protease Kex2 bei Pilzen.** Das Kex2-Protein besteht aus einer Golgi-luminalen und einer zytosolischen Domäne [27]. Die katalytische Aktivität befindet sich auf der N-terminalen, luminalen Seite gefolgt von einer strukturellen Domäne unbekannter Funktion, der P-Domäne. Proteolytisch abgespalten werden das Signalpeptid („Prä-“) und die Pro-Domäne.



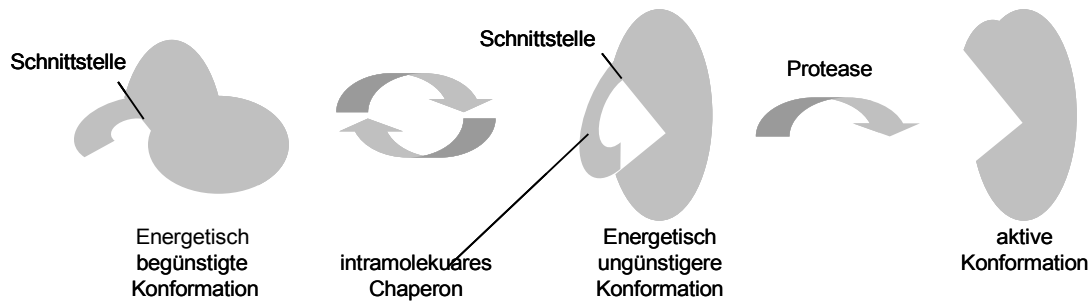
**Abbildung 2: Die Golgi-luminale Domänen von Kex2 aus *S. cerevisiae* (nach [28]).** Die luminale Domäne von Kex2 besteht aus zwei Subdomänen: der katalytischen, Subtilisin-ähnlichen Domäne und der strukturellen P-Domäne. Die P-Domäne besteht hauptsächlich aus einer  $\beta$ -Faltblatt Struktur (gelb), die die katalytische Domäne mit der Transmembrandomäne verbindet. Die katalytische Domäne setzt sich aus gleichen Teilen  $\alpha$ -Helix (rot) und  $\beta$ -Faltblatt Strukturen (gelb) zusammen. Das aktive Serin (grün) befindet sich in der Mitte der Substratbindungstasche (weißes Dreieck).

#### 1.4 Funktion eukaryontischer Prohormonkonvertasen

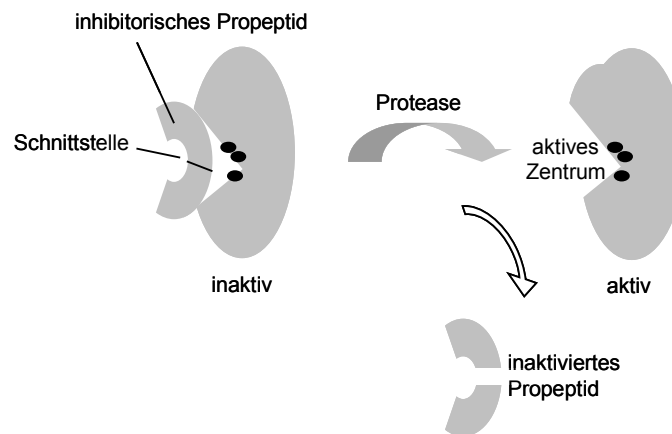
Bei eukaryontischen Organismen erfüllen Prohormonkonvertasen eine Vielzahl von Aufgaben, wie beispielsweise die Prozessierung von Peptidhormonvorläufern (Abbildung 3), die Aktivierung hydrolytischer Enzyme (Abbildung 4). Bei Proteasen, wie beispielsweise Kex2 selbst, haben die abgespaltenen Propeptide weiterhin eine inhibitorische Funktion, die durch einen zweiten Schnitt aufgehoben wird (Abbildung 5). Eine weitere Wirkungsweise von regulatorischen Proteasen ist das Trennen der Proteinen von Membranverankerungen (Abbildung 6), was zur Relokalisation führt. Höhere Organismen, wie z. B. Säugetiere, besitzen im Gegensatz zu Pilzen und anderen Einzellern eine große Familie dieser Prohormonkonvertasen [29].



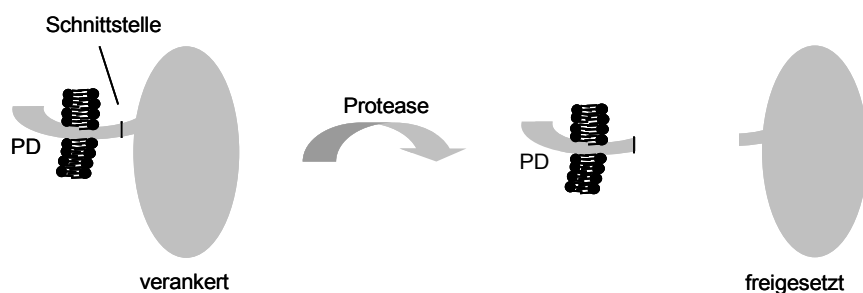
**Abbildung 3: Prozessierung des Pheromons MFa1 bei *S. cerevisiae*.** Das Pheromon MFa1 von *S. cerevisiae* wird durch Kex2 aus einem Vorläuferprotein hergestellt. Die durch die Proteolyse entstandenen Fragmente werden durch die Exoproteasen Kex1 und Ste13 an den Termini weiterprozessiert, wodurch biologisch aktives Pheromon in Form eines Peptidgemisches entsteht [30].



**Abbildung 4: Faltung von Proteinen durch intramolekulare Chaperone.** Viele Proteasen benötigen zur korrekten Faltung ein intramolekulares Chaperon in Form eines Propeptides, welches später durch Proteolyse abgespalten wird [31].



**Abbildung 5: Autoinhibition durch Propeptide.** Propeptide können als Inhibitoren wirken und dadurch einen Mechanismus für die proteolytische Aktivierung darstellen.



**Abbildung 6: „Shedding“ von Membrandomänen.** In einer Membran verankerte Proteine können durch Proteolyse freigesetzt werden und somit zu ihrem Wirkungsort gelangen.

## 1.5 Bedeutung von Kex2 bei Pathogenese- und Virulenzmechanismen in *Candida albicans* und anderen pathogenen Pilzen

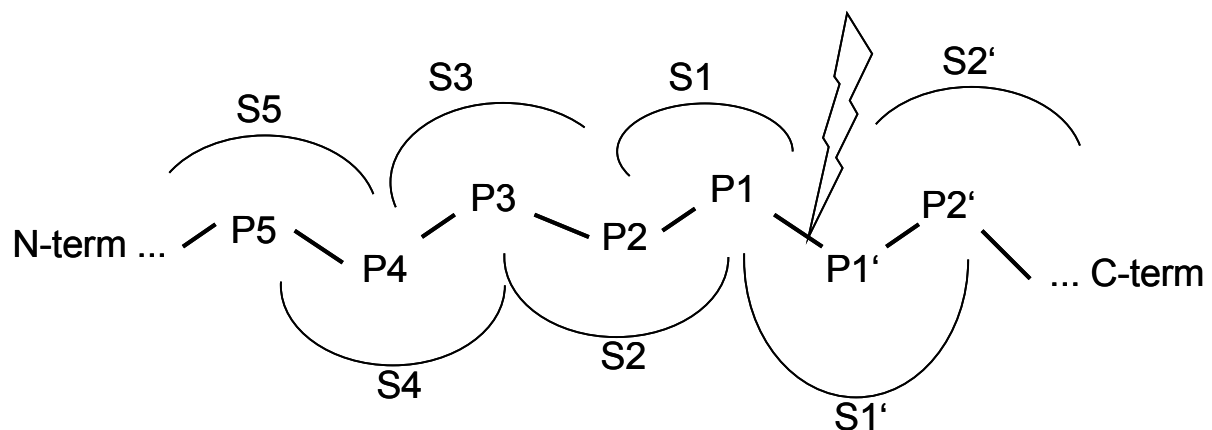
Deletionsmutanten von ScKEX2 und ScKEX2-ähnlichen Genen bei verschiedenen Pilzen zeigen ein breites Spektrum an phänotypischen Veränderungen gegenüber Wildtyp Stämmen. Eine direkte Folge des Kex2-Verlustes ist die Sekretion und Anreicherung unprozessierter Proteinvorläufer im Kulturmedium. Die Folgen einer ausbleibenden Prozessierung lassen sich in zwei Fällen direkt beobachten: die Fehlende Prozessierung des  $\alpha$ -Pheromons führt zum Ausbleiben der sexuellen Konjugation der Paarungstypen und die ausbleibende Prozessierung des viral kodierten Killertoxins bei *S. cerevisiae* führt zu einer Resistenz gegenüber diesem Toxin. Andere Phänotypen sind pleiotroperer Natur: alle bisher hergestellten KEX2-Mutanten zeigen morphologische Defekte der Zellwand, insbesondere in der Ausbildung von Myzelstrukturen und bei der Separation von Tochter- und Mutterzelle nach der Zellteilung. Die Summe dieser Effekte und die potentiell fehlende Prozessierung von Virulenzproteinen, wie bspw. den sekretorischen Aspartatproteasen bei *C. albicans*, führen zu einer Reduktion der Virulenz.

Organismus/Gen	Phänotypen
<i>Saccharomyces cerevisiae</i> KEX2	Fehlende Prozessierung des $\alpha$ -Pheromons und des Killertoxins [27]
<i>Candida albicans</i> KEX2	Aomale Prozessierung sekretorischer Proteine [32] Defekte im Myzelwachstum [33] Reduzierte Virulenz [33]
<i>Candida glabrata</i> KEX2	Defekt in der Separation von Mutter und Tochterzellen [34] Hypersensitivität gegenüber der Zellwand oder Plasmamembran angreifenden Antimykotika [35] Defekte in der Zellwandintegrität [35] Reduzierte Virulenz im <i>C. elegans</i> Modell [36]
<i>Yarrowia lipolytica</i> XRP6	Defekt in der Separation von Mutter und Tochterzellen [37] Defekte im Myzelwachstum [37] Verlangsamtes Wachstum [37]
<i>Aspergillus oryzae</i> KEXB	Verkleinerte Kolonien und verkürzte Hyphen [38] Zellwanddefekte [38]
<i>Aspergillus niger</i> KEXB	Langsameres Wachstum [39] Verkürzte Hyphen [39] Sekretion falsch prozessierter Glukoamylase und unprozessierter Fusionsproteine [39]
<i>Pichia pastoris</i> KEX2	Fehlende Prozessierung der pic9- $\alpha$ -Faktor Leadersequenz bei der Produktion rekombinanter Proteine in einer KEX2/YPS1 Doppelmutante [40]
<i>Schizosaccharomyces pombe</i> KRP1	Letal [41]
<i>Trichoderma reesii</i> KEX2	Akkumulation unprozessierter Xylanasevorläufer im Kulturüberstand [42]

**Tabelle 1: Phänotypen von Kex2-Deletionsmutanten bei Pilzen.** Die verschiedenen, in der Literatur beschriebenen Deletionsmutanten von KEX2-homologen Genen zeigen verschiedene auffällige Phänotypen.

## 1.6 Biochemische und genetische Methoden zur Analyse der Substratspezifität von Proteasen

Die Nomenklatur von Proteaseschnittstellen basiert auf dem Modell von Schechter und Berger [43]. In diesem Modell wird angenommen, dass das katalytische Zentrum beiderseits von spezifischen Subsites flankiert vorliegt. In diesen Subsites können sich die Seitenketten der Aminosäuren des Substrates anlagern. Es wurde festgelegt, dass die Substratseitenketten P (Peptid) und die mit diesen in Interaktion stehenden Enzymseitenketten als S (Subsite) bezeichnet werden. Die Subsites werden von der Spaltstelle in Richtung N-Terminus des Substrates („non-prime-side“) mit S1, S2, S3, ..., Sn und in Richtung C-Terminus des Substrates („prime-side“) mit S1', S2', S3', ..., Sn' bezeichnet. Die Seitenketten des Substrates werden analog dazu entsprechend der Anlagerung in den Subsites mit P1, P2, ..., Pn bzw. P1', P2', ..., Pn' bezeichnet (Abbildung 7).



**Abbildung 7: Nomenklatur von Proteaseschnittstellen nach Schechter und Berger.** Die Positionen im Substrat werden, ausgehend von der hydrolysierten Bindung zwischen P1 und P1', mit P1-Pn in Richtung des N-Terminus und mit P1'-Pn' in Richtung des C-Terminus bezeichnet. Die Substratbindungstaschen des Enzyms werden entsprechend mit S1-Sn und S1'-Sn' bezeichnet [43].

Es wurden viele verschiedene Methoden der Analyse der Substratspezifität von Proteasen entwickelt und in der Literatur beschrieben. An dieser Stelle sollen die Methoden vorgestellt werden, die entweder in dieser Arbeit oder in der Literatur zur Analyse der Substratspezifität von Kex2 oder Kex2-ähnlichen Proteasen verwendet wurden.

Die einfachste Methode zur Überprüfung, ob eine Protease ein bestimmtes Substrat schneidet, ist, beide Komponenten unter geeigneten Bedingungen *in vitro* zusammenzugeben und die entstehenden Produkte mittels SDS-PAGE aufzutrennen.

Voraussetzung hierfür ist jedoch, dass sowohl Substrat als auch Protease in ausreichender Menge als gereinigte Proteine zur Verfügung stehen. Dies wird vor allem dann problematisch, wenn eine größere Anzahl von systematischen Substraten untersucht werden soll, da die Expression

und Aufreinigung von Substratproteinen technisch nicht einfach und zeitaufwendig ist. Zudem führt der Austausch einzelner Aminosäuren teilweise dazu, dass ein Protein nicht mehr exprimiert wird, nicht die richtige Konformation hat, oder sich nicht mehr ohne größeren Aufwand aufreinigen lässt. Hier kann man auf synthetische Peptide als Substrate zurückgreifen, die auf der FRET-Technologie („Fluorescence Resonance Energy Transfer“) basieren [44]. Diese Substratpeptide werden bei der Synthese beispielsweise terminal jeweils mit EDANS- und Dabcyl-Gruppen modifiziert. Das Absorptionsspektrum der Dabcyl-Gruppe überlagert das Emmissionsspektrum der EDANS-Gruppe und somit ist keine Fluoreszenz messbar, solange sich die beiden Gruppen in räumlicher Nähe befinden. Wird das Peptid gespalten, so wird die Entfernung der beiden Gruppen zu groß und die Fluoreszenz der EDANS-Gruppe dadurch messbar.

Eine ähnliche Methode, die ohne synthetische Peptide auskommt, wurde von Patel *et al.* [45] beschrieben. Hier wird die Proteaseschnittstelle zwischen ein 6xHIS-Affinitätstag und das green fluorescent protein (GFP) gesetzt. Das Substratprotein wird an Nickel-Agarose gebunden: wenn die GFP-Domäne durch Zugabe der Protease abgespalten wird, kann die Fluoreszenz nach Zentrifugation im Reaktionsüberstand quantifiziert werden. Ebenso wie die Analyse der Reaktionsprodukte in der SDS-PAGE, ist es für diese Methode notwendig, sowohl Protease als auch Substrat heterolog zu exprimieren und aufzureinigen.

## 1.7 Substratspezifität der Protease Kex2

In der Literatur beschriebene Untersuchungen zur Substratspezifität von *S. cerevisiae* Kex2 gliedern sich in drei prinzipiell unterschiedliche Ansätze. Zum einen wurde die Aktivität von Wildtyp Kex2 gegenüber systematischen Mutationen von Substraten untersucht (Abschnitt 1.7.1). Zum anderen wurden durch gezielte Mutationen in dem Proteasegen die an der Substraterkennung beteiligten Aminosäuren gefunden und untersucht (Abschnitt 1.7.3). Schließlich ließ auch die Kristallstruktur der Protease, z. B. im Komplex mit einem substratähnlichen Inhibitor, Schlussfolgerungen über die an der Substratbindung beteiligten Aminosäuren zu (Abschnitt 1.7.2). Erst die feste räumliche Einbindung des Substrats in die Substratbindungstasche ermöglicht die Katalyse durch das Enzym. Dazu sind starke Wechselwirkungen zwischen Enzym und Substrat vonnöten, die im Fall der Kexine hauptsächlich über elektrostatische Kontakte vermittelt werden. Aber auch negative Selektion gegenüber bestimmten, die räumliche Anordnung der Substratbindungsregion perturbierenden Aminosäuren, findet statt.

### 1.7.1 Untersuchungen mit systematischen Substraten

Systematische Untersuchungen mit Substraten an *S. cerevisiae* Kex2 wurden erstmals 1997 von Rockwell [46] durchgeführt. Hier wurde die katalytische Aktivität ( $K_{\text{cat}}/K_{\text{M}}$ ) von Kex2 gegenüber verschiedenen fluorogenen Peptiden mit Variationen an P1 gemessen und die relative katalytische Aktivität in Bezug auf ein Wildtyp-Peptid, welches von der Sequenz des  $\alpha$ -Pheromons von *S. cerevisiae* abgeleitet wurde, errechnet. Dabei konnten die Autoren zeigen, dass *S. cerevisiae* Kex2 ausschließlich C-terminal nach Arginin (Position P1) schneidet, nicht jedoch nach entweder ungeladenen Derivaten mit ähnlicher Wasserstoffbrücken-Bildung (Citrullin) oder ähnlicher Ladung (Lysin, Ornithin) (Tabelle 2, „P1“). Des Weiteren zeigten die Autoren, dass *S. cerevisiae* Kex2 positiv geladene Aminosäuren an P2 bevorzugt und an P3 gegen Aspartat diskriminiert.

Die Substratspezifität von *S. cerevisiae* Kex2 an Position P2 wurde von der gleichen Gruppe ein Jahr später genauer charakterisiert [24]. Dazu bedienten sie sich eines genetischen Systems, bei dem die Effizienz des Matings mit verschiedenen, systematischen Substitution an der P2-Position des  $\alpha$ -Pheromons gemessen wurde. Hier zeigte sich, dass alle Substitutionen für Lysin an P2 die Effizienz des Matings reduzierten (Tabelle 2, „P2“). Die besten Werte erzielte Arginin (50%), aber auch andere Aminosäuren wie Threonin (13%) oder Prolin (9%) ließen noch eine Prozessierung zu. Diese Spezifitäten an P1 und P2 wurden mittels sGASP bestätigt [47] (siehe Abschnitt 1.6).

Die Spezifität an den Positionen P3 und P4 wurde wiederum ein Jahr später mittels systematischen Substitutionen an Schnittstellen einer synthetischen Fusion von  $\beta$ -Galaktosidase aus *Escherichia coli*, einem Linker mit der Schnittstelle und einem 34 Aminosäuren langen Abschnitt aus dem humanen parathyroiden Hormon untersucht. Hier konnte die Prozessierung in der Linkerdomäne per HPLC quantifiziert werden [48]. Es zeigte sich für P3 (Tabelle 2, „P3“) lediglich die schon beobachtete negative Selektion gegen Aspartat, für P4 konnte jedoch ein starker positiver Effekt bei der Substitution durch positiv geladene Aminosäuren beobachtet werden (Tabelle 2, „P4“). Eine Untersuchung der Selektivität an Position P6 [49] mit fluorogenen Substratpeptiden, welche für eine Untersuchung der humanen Protease Furin erzeugt wurden, zeigte, dass die Substitution an P6 für Kex2 keine relevante Veränderung der Substrataffinität oder Katalyse zur Folge hatte.

P1 Rockwell et al., 1997 <sup>1</sup> [46]			P2 Bevan et al., 1998 [24]		
Name	untersuchte Sequenz	Relative Aktivität [%]	Name	untersuchte Sequenz	rel. Mating Effizienz [%]
	<b>IQ-Substrate</b>		Lys	...SLD <b>K</b> R/EAEA...	100
I-1	RJβY <b>K</b> R/EAEABR	100	Arg	...SLD <b>R</b> R/EAEA...	51
I-12	RJβY <b>K</b> K/EAEABR	1,4	Thr	...SLD <b>T</b> R/EAEA...	13
I-13	RJβY <b>K</b> O/EAEABR	0,88	Pro	...SLD <b>P</b> R/EAEA...	9,5
I-14	RJβY <b>R</b> K/EAEABR	0,019	Glu	...SLD <b>E</b> R/EAEA...	8,8
	<b>AMC-Substrate</b>		Ile	...SLD <b>I</b> R/EAEA...	5,4
A-1	AcPβY <b>K</b> R/Mca	100	Ser	...SLD <b>S</b> R/EAEA...	4,3
A-10	AcPβY <b>K</b> K/Mca	0,44	Ala	...SLD <b>A</b> R/EAEA...	3,2
A-11	AcPβY <b>K</b> Ç/Mca	0,015	Asn	...SLD <b>N</b> R/EAEA...	2,9
			Val	...SLD <b>V</b> R/EAEA...	1,4
			Cys	...SLD <b>C</b> R/EAEA...	0,85
			Asp	...SLD <b>D</b> R/EAEA...	0,27
			Gln	...SLD <b>Q</b> R/EAEA...	0,24
			Gly	...SLD <b>G</b> R/EAEA...	0,23
			His	...SLD <b>H</b> R/EAEA...	0,072
			Met	...SLD <b>M</b> R/EAEA...	0,068
			Leu	...SLD <b>L</b> R/EAEA...	0,043
			Tyr	...SLD <b>Y</b> R/EAEA...	0,017
			Phe	...SLD <b>F</b> R/EAEA...	0,00068
			Trp	...SLD <b>W</b> R/EAEA...	0,00009

P3 Suzuki et al., 2000 [48]			P4 Suzuki et al., 2000 [48]		
Name	untersuchte Sequenz	Relative Aktivität [%]	Name	untersuchte Sequenz	Relative Aktivität [%]
Gly	...PSV <b>G</b> KR/SVS...	100	Gly	...PS <b>G</b> HKR/SVS...	6
Ala	...PSV <b>A</b> KR/SVS...	92	Ala	...PS <b>A</b> HKR/SVS...	45
Val	...PSV <b>V</b> KR/SVS...	130	<b>Val</b> <sup>2</sup>	...PS <b>V</b> HKR/SVS...	100
Ile	...PSV <b>I</b> KR/SVS...	79	Ile	...PS <b>I</b> HKR/SVS...	51
Leu	...PSV <b>L</b> KR/SVS...	128	Leu	...PS <b>L</b> HKR/SVS...	16
Phe	...PSV <b>F</b> KR/SVS...	141	Phe	...PS <b>F</b> HKR/SVS...	108
Pro	...PSV <b>P</b> KR/SVS...	0	Pro	...PS <b>P</b> HKR/SVS...	36
Trp	...PSV <b>W</b> KR/SVS...	126	Trp	...PS <b>W</b> HKR/SVS...	58
Met	...PSV <b>M</b> KR/SVS...	N/A	Met	...PS <b>M</b> HKR/SVS...	95
Cys	...PSV <b>C</b> KR/SVS...	N/A	Cys	...PS <b>C</b> HKR/SVS...	9
Ser	...PSV <b>S</b> KR/SVS...	92	Ser	...PS <b>S</b> HKR/SVS...	31
Thr	...PSV <b>T</b> KR/SVS...	110	Thr	...PS <b>T</b> HKR/SVS...	30
Asn	...PSV <b>N</b> KR/SVS...	98	Asn	...PS <b>N</b> HKR/SVS...	88
Gln	...PSV <b>Q</b> KR/SVS...	102	Gln	...PS <b>Q</b> HKR/SVS...	74
Tyr	...PSV <b>Y</b> KR/SVS...	121	Tyr	...PS <b>Y</b> HKR/SVS...	85
His	...PSV <b>H</b> KR/SVS...	143	His	...PS <b>H</b> HKR/SVS...	131
<b>Lys</b> <sup>3</sup>	...PSV <b>K</b> KR/SVS...	100	Lys	...PS <b>K</b> HKR/SVS...	128
Arg	...PSV <b>R</b> KR/SVS...	128	Arg	...PS <b>R</b> HKR/SVS...	179
Asp	...PSV <b>D</b> KR/SVS...	38	Asp	...PS <b>D</b> HKR/SVS...	0
Glu	...PSV <b>E</b> KR/SVS...	89	Glu	...PS <b>E</b> HKR/SVS...	0

**Tabelle 2: Biochemische Aktivität von *S. cerevisiae* Kex2.** Zusammenfassung der in der Literatur beschriebenen systematischen Untersuchungen der katalytischen Aktivität von *S. cerevisiae* Kex2 gegenüber verschiedenen Peptiden.

<sup>1</sup> Von den Autoren verwendete Abkürzungen für ungewöhnliche Aminosäuren und fluorogene Gruppen bzw. Quencher: Ac: Acetyl, Mca: C-terminales Methylcoumarinamid, Ç: Citrullin, β: Norleucin, O: Ornitin, J: Lys(DABCYL)-Konjugat, B: Glu(EDANS)-Konjugat.

<sup>2</sup>Als Referenzwert auf 100% gesetzt.

<sup>3</sup>Als Referenzwert auf 100% gesetzt.



Die Protease Kex2 aus *S. cerevisiae* schneidet also ausschließlich nach Arginin an P1, hat eine hohe Präferenz für die positiv geladenen Aminosäuren Lysin und Arginin an der Position P2, diskriminiert an P3 gegen negative Ladungen und bevorzugt an P4 positiv geladene oder große aliphatische Aminosäuren. Die Schnittstelle in Substraten von Kex2 wird in der Literatur im Allgemeinen trotzdem als [K/R]R bzw. XX[K/R]R angegeben.

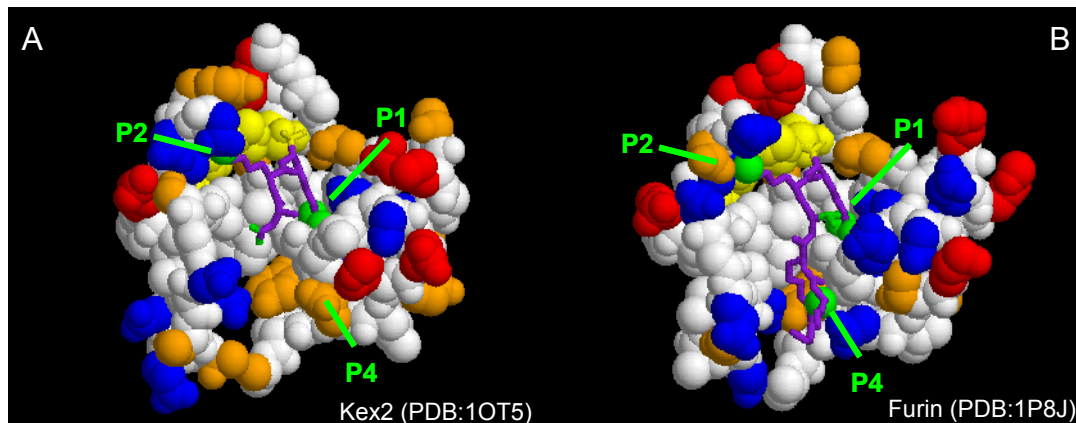
Im Gegensatz dazu hat das homologe Protein Furin aus höheren Organismen eine davon sehr unterschiedliche Substratspezifität. Furin zeigt an P2 keine stringente Präferenz, dafür aber eine hohe Selektivität für Lysin und Arginin an P4. Die Schnittstelle für Furinsubstrate wird daher in der Literatur im Allgemeinen als [K/R]XXR angegeben. Allerdings existiert in höheren Säugern nicht nur eine Kex2-ähnliche Protease (Furin), sondern eine Familie Kex2-ähnlicher Proteasen mit mindestens sieben Mitgliedern, welche wahrscheinlich mit einer breiten Substratspezifität für eine vielfältige Prozessierung, in Abhängigkeit von Expressionszeitpunkt und subzellulärer Lokalisation, von Proteinvorläufern sorgt [29]. Im Folgenden soll daher betrachtet werden, in welchen strukturellen Merkmalen die Spezifität der Proteasen begründet liegt.

### 1.7.2 Untersuchungen mittels Röntgenbeugung an Kex2 und Furin

Nach mehreren Ansätzen, das reaktive Zentrum von Prohormonkonvertasen durch Sequenzvergleiche zu modellieren [50-52], wurden die dreidimensionalen Strukturen der Golgi-luminalen Proteasedomänen von *S. cerevisiae* Kex2 und *M. musculus* Furin in den Jahren 2003 bis 2005 mehrfach experimentell bestimmt [28, 53-55]. Diese Analysen boten erstmals die Möglichkeit, direkt aus der Struktur der Proteasen Aussagen über die globale Struktur des Proteins (siehe Abschnitt 1.3), über Glykosylierungen und über als Kofaktoren gebundene Kalzium- und Kaliumionen zu treffen.

In allen vier Fällen wurden die Proteasen mit kovalent an das aktive Serin gebundenen Substrat-ähnlichen Inhibitoren, entsprechend den Positionen P4-P1, kristallisiert. Dadurch war es möglich, die an der Substraterkennung beteiligten Aminosäuren der Substratbindungstaschen S1-S4 direkt durch ihre Interaktion mit dem Peptid-analogen Inhibitor zu bestimmen (Abbildung 8). Aus biochemischen Untersuchungen war bekannt, dass beide Proteasen ausschließlich nach der Aminosäure Arginin an Position P1 schneiden (Abschnitt 1.7.1). Tatsächlich besteht die Tasche S1 sowohl in Kex2 als auch in Furin aus einem lang gestreckten, von hydrophoben Seitenketten flankierten und von negativ geladenen Aminosäuren terminierten (D276 und D325 in Kex2, D257 und D306 in Furin) Tunnel, so dass nur ein Arginin hier mit seiner terminalen positiven Ladung richtig fixiert werden kann. Eine rigide Fixierung ist wiederum die Voraussetzung für die Hydrolyse der P1-P1' Bindung. Des Weiteren interagiert die P1-Seitenkette mit dem einen

Rückgrat Carbonyl (A311 in Kex2, A292 in Furin) über eine Wasserstoffbrückenbindung. Die S1-Tasche wird durch ein Kalziumion über D277 in Kex2, D258 in Furin, stabilisiert, was aufgrund der hohen negativen elektrostatischen Ladungen in diesem Bereich für die Stabilität der Tasche essentiell scheint [55]. Dieses findet auch darin Ausdruck, dass Kalziumionen essentiell für die Katalysereaktion *in vitro* sind [56].



**Abbildung 8:** Räumliche Darstellung der Substratbindungstaschen P4-P1 von *S. cerevisiae* Kex2 (A) und *M. musculus* Furin (B). Die beiden Enzyme ScKex2 und MmFurin besitzen fast identisch aufgebaute Substratbinderegionen. Lediglich die Verteilung der negativ geladenen Aminosäuren (blau) unterscheidet sich und reflektiert die unterschiedliche Präferenz von positiven Ladungen an P2 (Kex2) bzw. P4 (Furin). Weiterhin hervorgehoben sind die Aminosäuren der katalytischen Triade (gelb), die kovalent gebundenen Substratanaloga (violett und grün) und positiv geladene Aminosäuren der Enzyme (rot).

An der Position P2 unterscheiden sich die Spezifitäten der beiden Proteasen stark. Während Kex2 an dieser Stelle positive Ladungen bevorzugt, besitzt Furin hier keine ausgeprägte Präferenz. Ein Vergleich der Strukturen an der S2 Tasche zeigt, dass auch hier in beiden Proteinen eine starke lokale negative Ladung herrscht, vermittelt durch die Aminosäuren D176, D210 und D211 in Kex2, sowie D154, D191 und N192 in Furin. Der entscheidende Unterschied an dieser Stelle ist eine Insertion in Furin und anderen Prohormonkonvertasen gegenüber Kex2: diese Insertion einer kleinen Schleife zwischen den Positionen 185 bis 192 (entsprechend 204 bis 211 in Kex2) bewirkt, dass das Aspartat D191 (D210 in Kex2) von der P2 Aminosäure des Substrats weg orientiert ist. Daraus resultiert eine wesentlich geringere Affinität zu positiv geladene P2-Gruppen in Furin gegenüber Kex2.

Aus den Strukturen der Proteine ist ersichtlich, dass die Seitenkette des Substrats an der Position P3 aus der Substraterkennungsregion herausragt und somit eigentlich nicht zur Erkennung beitragen sollte. Die in kinetischen Untersuchungen jedoch beobachtete Diskriminierung gegen negative Ladungen an P3 lässt sich durch die elektrostatische Abstoßung einer potentiellen Aspartatseitenkette an dieser Position in relativer Nähe (ca. 4 Å) zu dem Aspartat D276 erklären [28].

Ähnlich wie an der Position P2, ist auch an der Position P4 die Substratspezifität der beiden Proteasen unterschiedlich. Kex2 generiert seine Spezifität hauptsächlich über die S1- und S2-Taschen, Furin dagegen hauptsächlich über S1 und S4. Die S4 Bindungstasche von Kex2 selektiert sowohl in Richtung positiv geladener als auch großer aliphatischer Aminosäuren. Dieser Unterschied rührt hauptsächlich daher, dass in der S4-Tasche von Furin die P4-Aminosäure eine andere Konformation annimmt, als in Kex2. Bei Furin werden sowohl die terminale positive Ladung als auch die Stickstoffatome der Guanidiniumgruppe eines Arginins an P4 ähnlich wie an S1 erkannt. Dies wird bei Furin dadurch möglich, dass hier die Bindung in eine hydrophobe Umgebung wie bei Kex2 durch V231 (I250 bei Kex2) versperrt ist und die P4 Aminosäure so mit der ausgeprägten P4 Bindungstasche wechselwirken muss. Im Gegensatz dazu hat Kex2 hier die Möglichkeit, sowohl die geladene P4 Bindungstasche als auch eine hydrophobe Region zur Bindung der P4 Aminosäure zu nutzen. Verursacht wird dies durch die Aminosäure D254, welche in Kex2 nach außen gewandt ist, die äquivalente Aminosäure V235 in Furin dagegen nach innen, wodurch der Zugang zur hydrophoben Tasche für das Substrat versperrt ist [54].

Die in den Strukturanalysen verwendeten Inhibitoren verfügten weder über die Position P5 und Position P6 noch über die Positionen P1' und Folgende, daher gibt es über Bindungen zwischen Substrat und Protease an diesen Positionen nur Vermutungen. Durch Vergleich der Strukturen von Kex2 und der bakteriellen Subtilase Thermitase [28] konnte eine potentielle P1' Bindungsregion für Kex2 ausgeschlossen werden. Dies stimmt mit biochemischen Experimenten überein, nach denen Kex2 keine Selektivität an dieser Stelle aufweist, jedoch hier gegen große Aminosäuren diskriminiert [57].

Ergänzend dazu konnten Cai *et al.* [58] durch Modellierung elektrostatischer Interaktionen mit verschiedenen Varianten des Kexin-Inhibitors Eglin C verschiedene mögliche Bindungsregionen für die Positionen P2' bis P4' identifizieren und diese Aussagen mit Inhibitionsstudien untermauern. Danach ist die S2'-Tasche durch schwache hydrophobe Interaktionen mit den Aminosäuren P347 und Y348 (W328 und Y329 in Furin) gekennzeichnet. Einen größeren Einfluss hat das Arginin an Position 318 (E299 in Furin), welches eine Abstoßung positiv geladener Aminosäuren im Substrat an dieser Position verursacht. Die Bindung des Inhibitors an der Position P3' erfolgt bei Kex2 durch Ausbildung von Wasserstoffbrücken an S380 und Y212. Ein Austausch des Arginins im Inhibitor an dieser Stelle durch die negativ geladene Aminosäure Aspartat führt bei Kex2 zu einem Verlust der Inhibition, nicht aber bei Furin, da hier R193 und R197 (T und R in Kex2) kompensieren können. Die Position P4' ist im Furin-Eglin C-Komplex vom Protein abgewandt, bei Kex2 gibt es jedoch eine elektrostatische Wechselwirkung mit

R318, was wiederum zu einer Abstoßung von positiven Ladungen im Inhibitor an dieser Position führt. Arbeiten mit einer randomisierten Bank von Eglin C Mutationen führte zu vergleichbaren Ergebnissen [59].

### 1.7.3 Untersuchungen mittels Mutagenese von Kex2

Ein weiterer Ansatz, die Substratspezifität eines Enzyms zu verstehen, ist es, durch gezielte Mutagenese Codons für Aminosäuren des aktiven Zentrums auszutauschen. Dies wurde in *S. cerevisiae* *KEX2* und in *M. musculus* in den Genen für Furin und der verwandten Protease PC2 unternommen.

Brenner *et al.* [60] waren die ersten, die durch Mutagenese des Asparagins an Position 314 (entsprechend N295 bei Furin) zu Aspartat zeigen konnten, dass die Mutation N314D zu einer reduzierten Katalyserate führte. Zeitgleich mit den verschiedenen Strukturaufklärungen wurden mittels Mutagenese auch die Unterschiede in der P4- und P6-Spezifität von Kex2 und Furin untersucht [61]. Die in *S. cerevisiae* Kex2 vorgenommenen Substitutionen wurden hier so ausgewählt, dass das aktive Zentrum bezüglich der Substratspezifität an den Positionen P4 und P6 an das von Furin angenähert wird. Tatsächlich führt die Kombination der Mutationen T252D und Q283E (entsprechend D233 und E264 in Furin) zu einer furinähnlicheren Substratspezifität: beide Substitutionen führten in Kex2 zu einer erhöhten Selektivität gegenüber argininenthaltenden Substraten an den Positionen P4, P5 und P6. In einer anderen Studie konnten durch Mutagenese der Aminosäuren D176, D210 und D211 Varianten von Kex2 erzeugt werden, die eine erhöhte Aktivität gegenüber Methionin an P2 zeigten [62]. Untersuchungen mit Aminosäuresubstitutionen im Substratbindezentrum der Protease PC2 legen des Weiteren nahe, dass nicht nur spezifische Interaktionen zwischen Enzym und Substrat an einzelnen Positionen für die Selektivität der Protease ausschlaggebend sind, sondern der Gesamtbau des Substratbindezenters für diesen Prozess von großer Wichtigkeit ist [63].

## 1.8 *In silico* Methoden der Vorhersage von Proteaseschnittstellen

Etablierte Methoden, speziell zur Vorhersage von Proteaseschnittstellen in Proteinsequenzdaten, existieren nicht. Man bedient sich bisher allgemeinen Methoden zur Motivsuche und der Statistik. Das gebräuchlichste Verfahren ist die Verwendung einer positionsspezifischen Matrix, in der die Häufigkeitsverteilung der 20 Aminosäuren in jeder Position relativ zur Schnittstelle aufgeführt ist. Hieraus lässt sich für ein zu untersuchendes Motiv berechnen, wie gut es mit der Verteilung in der Matrix übereinstimmt. Eine solche Matrix wurde von Rholam *et al.* [57] für Furinsubstrate aufgestellt, um die generelle Spezifität der Protease abzuleiten. Modernere

Verfahren setzen Methoden der Künstlichen Intelligenz, beispielsweise „Artifizielle Neuronale Netze (ANN)“ [64], oder stochastische Methoden wie z. B. „Self Organizing Maps (SOM)“ oder „Separierende Hyperebenen“ („support Vektoren“) ein.

Zu den untersuchten Proteasen gehören Caspasen und Granzyme B [65, 66], virale Proteasen [67, 68, 69] und die Familie der Prohormonkonvertasen bei Säugern [64].

### **1.9 Stereochemische Voraussetzungen im Substrat für die Proteolyse nativer Proteine**

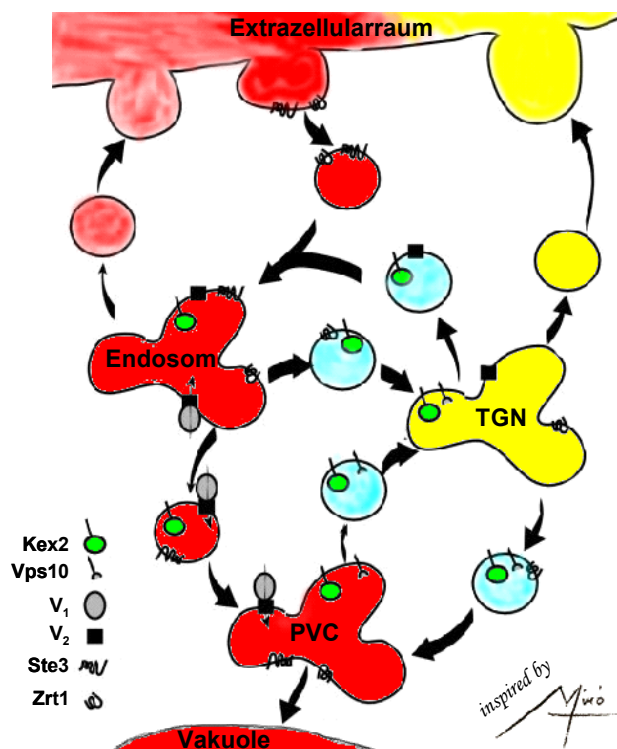
Bei der Vorhersage von Proteaseschnittstellen in Proteinen spielt nicht nur die Spezifität der Protease eine Rolle, sondern auch die ausgeprägte Tertiärstruktur des Substrats [70]. So muss eine Schnittstelle an einer Position im Substrat gelegen sein, die für die Protease zugänglich ist um tatsächlich prozessiert werden zu können. Es gibt verschiedene Modelle, wie dieses geschehen kann. Zunächst kann eine Schnittstelle an einer exponierten Stelle, beispielsweise einer Verbindung zweier von einander unabhängig gefalteten Domänen oder in einem exponierten Teil des Proteins gelegen sein. Des Weiteren gibt es Teile von Proteinen, die sich leichter entfalten als andere. Beispielsweise können Teile von Helices wesentlich leichter entfaltet werden als Teile von  $\beta$ -Faltblättern [71]. Auch kommen in Proteinen Bereiche vor, die über keine ausgeprägte Sekundärstruktur verfügen, und somit im nativen Zustand über einen hohen Grad an intrinsischer Unordnung verfügen [72]. Diese Bereiche fallen in Kristallstrukturanalysen dadurch auf, dass die Atome an diesen Stellen über hohe Temperaturwerte verfügen und ihre Position daher im Hintergrundrauschen verschwindet [73]. Eine andere Untersuchung kommt zu dem Schluss, dass die Oberflächenlokalisierung eine weit wichtigere Determinante als die Flexibilität ist [74]. Im Fall der degradativen Proteolyse und der Prozessierung von Polypeptiden kommt ein weiterer Faktor hinzu: die Prozessierung an einer Stelle führt zur Freisetzung weiterer, bisher verborgener Schnittstellen, so dass das Protein in viele Einzelfragmente zerfällt.

### **1.10 Funktion und Struktur des Kex2-enthaltenden Zellkompartiments**

Das Kex2 Protein ist in einem Kompartiment des späten Trans-Golgi-Netzwerkes (TGN) lokalisiert [75]. Es wird teilweise auch in angrenzenden Kompartimenten, wie z. B. sekretorischen Vesikeln, dem prä-vakuolären Kompartiment und späten Endosomen gefunden [76], nie allerdings in den früheren Teilen des sekretorischen Pfads (Abbildung 9). Die Sortierung von Membranproteinen in dieses Kompartiment erfolgt durch im Zytoplasma gelegene Motive, im Fall von Kex2 sind diese auf der C-terminalen Domäne gelegen [77]. Das Kex2-enthaltende TGN-Kompartiment beherbergt auch zwei weitere Membranproteine des

proteolytischen Aktivierungsprozesses, Kex1 und Ste13 [78]. Kex1 ist eine Carboxypeptidase, welche terminale Arginine und Lysine entfernt [79] und für die Generierung von aktivem  $\alpha$ -Pheromon wichtig ist, ebenso wie die Dipeptidylaminopeptidase Ste13 [80].

Es gibt keine Studien darüber, ob alle sekretorischen Proteine dieses Kompartiment durchlaufen, oder ob noch weitere parallele Wege existieren. Aufgrund der Tatsache, dass sich Kex2 auch durch das prä-vakuoläre Kompartiment und Endosomen bewegt, kommen daher auch vakuoläre Proteine und endophagozytierte Protein als theoretische Ziele der proteolytischen Aktivität von Kex2 in Frage (Abbildung 9).



**Abbildung 9: Sortierung von Kex2 im späten Trans-Golgi-Netzwerk (nach Sipos *et al.*, 2004 [81]).** Das Kex2-Protein befindet sich sowohl in einem späten Kompartiment des Trans-Golgi-Netzwerks (TGN) im sekretorischen Weg (gelb) als auch in endosomalen Kompartimenten (rot) und dem prä-vakuolären Kompartiment (PVC).

## 2 Ziel dieser Arbeit

Die Protease Kex2 ist eine im Trans-Golgi-Netzwerk von Pilzen lokalisierte Endoprotease, welche im Transit befindliche Proteine an spezifischen Schnittstellen schneidet und auf diese Weise aktiviert. Trotz der biochemisch gut charakterisierten Aktivität von Kex2 aus *S. cerevisiae* sind über die zwei bekannten Substrate Killertoxin und  $\alpha$ -Kreuzungspheromon hinaus nur wenige Proteine bekannt, für die eine Prozessierung durch Kex2 erfolgt. Trotzdem ist es in der Literatur üblich, Proteine mit einem in relativer Nähe zum N-Terminus gelegenen dibasischen „KR“-Motiv als putative Kex2-Substrate zu annotieren. Es gibt jedoch Beispiele dafür, dass Kex2 auch C-terminale oder endogene Prozessierungsstellen erkennt und schneidet, so dass generell davon ausgegangen werden muss, dass Kex2-Prozessierungen nicht nur auf N-terminale Bereiche beschränkt sind. Gleichzeitig gibt es aber auch Beispiele für unprozessiert bleibende „KR“-Motive. Daraus folgt, dass die in der Literatur übliche Beschreibung der Schnittstelle als ein in Nähe des N-Terminus gelegenes „KR“-Motiv weder ausreichend noch spezifisch ist.

In dieser Arbeit sollten daher Substratproteine von Kex2-Proteasen in Pilzen im Allgemeinen, sowie in *Candida albicans* und *C. glabrata* im Speziellen, identifiziert werden. Aufgrund der erweiterten experimentellen Basis sollte anschließend ein erweitertes Modell zur Vorhersage von Kex2-Substraten in Genomdaten entwickelt werden.

Im experimentellen Teil dieser Arbeit sollten ausgewählte putative Kex2-Substrate heterolog exprimiert und *in vitro* mit rekombinanten Kex2-Enzymen auf ihre Suszeptibilität gegenüber spezifischer Proteolyse hin überprüft werden.

Im theoretischen Teil dieser Arbeit sollte versucht werden, aus den gewonnenen Daten Kriterien für die verbesserte Vorhersage von Kex2-Schnittstellen in Proteinsequenzen abzuleiten und anhand dieser eine eingeschränkte Liste mit möglichen Kex2-Substraten zu erstellen.

### 3 Materialien und Methoden

#### 3.1 Organismen und Nährmedien

Alle gängigen Verbrauchskemikalien wurden, soweit nicht anders angegeben, von der Firma Roth, Feinchemikalien von der Firma Merck, Enzyme von der Firma Fermentas Life Science und Oligonukleotide von der Firma TibMolbiol, bezogen.

##### 3.1.1 Bakterien

###### 3.1.1.1 Stämme

Name	Genotyp
DH5- $\alpha$	F-, phi80dlacZ $\Delta$ M15, $\Delta$ (lacZYA-argF), U169, deoR, recA1, hadR17 (rk-,mk+), gal-, phoA, supE44, $\lambda$ -, thi-1, gyrA96m relA1 (Stratagene)
Rosetta blue	F- ompT hsdS <sub>B</sub> (r <sub>B</sub> <sup>-</sup> m <sub>B</sub> <sup>-</sup> ) dcm pRARE <sup>2</sup> (Cm <sup>R</sup> ) (Novagen)

Tabelle 3: Verwendete Bakterienstämme.

###### 3.1.1.2 Medien

Medium	Zusammensetzung
LB-Medium „Luria-Bertani“	1% w/v Hefeextrakt (Becton Dickinson) 1% w/v Trypton (Becton Dickinson) 0,5% w/v NaCl pH 7,5 + 2% w/v Agar (Becton Dickinson) bei Festmedien + ggf. 50 $\mu$ g/ml Ampicillin (Sigma) zur Selektion + ggf. 50 $\mu$ g/ml Chloramphenicol (Sigma) zur Selektion
Autoinduzierendes LB/Laktose-Medium	1% w/v Hefeextrakt (Becton Dickinson) 1% w/v Trypton (Becton Dickinson) 0,5% w/v NaCl pH 7,5 1% w/v Laktose 0,04 % w/v Glukose

Tabelle 4: Medien zur Anzucht von Bakterien.



### 3.1.2 Pilze

#### 3.1.2.1 Stämme

Name	Organismus	Genotyp/Beschreibung	Referenz
GS 115	<i>Pichia pastoris</i>	His-	Invitrogen manual
KM 71	<i>Pichia pastoris</i>	His-	Invitrogen manual
	<i>Pichia pastoris</i>	Expression ssScKex2	G. Boileau
PpCg12	<i>Pichia pastoris</i>	Expression ssCgKex2	Diese Arbeit
CAI4	<i>Candida albicans</i>	Ura-	
CaActKEX2-4	<i>Candida albicans</i>	Expression ssCaKex2	Diese Arbeit
SC5314	<i>Candida albicans</i>	Wt	

Tabelle 5: Verwendete Pilzstämme.

#### 3.1.2.2 Medien

Medium	Zusammensetzung
YPG Medium "yeast, peptone, glucose" Vollmedium zur Anzucht von Pilzen	1% w/v Hefeextrakt (BD) 1% w/v Pepton (BD) 2% w/v Glukose +2% w/v Agar (BD) bei Festmedien
SD Selektionsmedium „synthetic defined“ Selektionsmedium zur Anzucht auxotropher Mutanten	0,67% (w/v) Yeast-Nitrogen Base®, frei von Aminosäuren (BD) 1% (w/v) Glukose 2% (w/v) Bactoagar® bei Festmedien (BD)
Sabouraud Medium Unselektives Medium zur Anzucht von Pilzen	1% (w/v) Casein Hydrolysat (BD) 4% (w/v) Glukose 1,5% (w/v) Agar (BD)
BMGY Medium "buffered minimal glycerol+yeast extract" Vollmedium zur Anzucht von <i>P. pastoris</i>	1% (w/v) Hefeextrakt 2% (w/v) Pepton 100 mM K <sub>2</sub> PO <sub>4</sub> , pH 6,0 1% (v/v) Glycerol
MD Medium „minimal dextrose“ Minimalmedium zur Selektion von <i>P. pastoris</i> Transformanden	1,34% (w/v) Yeast Nitrogen Base (BD) 2% (w/v) Glukose 2% Bactoagar® bei Festmedien (BD)
BMM Medium „buffered minimal methanol“ Minimalmedium zur heterologen Proteinexpression in <i>P. pastoris</i>	1,34% (w/v) Yeast Nitrogen Base (BD) 0,5% (v/v) Methanol 100 mM K <sub>2</sub> PO <sub>4</sub> , pH 6,0

Tabelle 6: Medien zur Anzucht von Pilzen.

## 3.2 Molekularbiologische Methoden

### 3.2.1 Isolation von Nukleinsäuren

#### 3.2.1.1 Isolation genomischer DNA aus Pilzen

Für die Isolation genomischer DNA aus Pilzen im analytischen Maßstab [präparativer Maßstab] werden die Zellen über Nacht in 3 ml [50 ml] YPG Medium bei 37°C unter Schütteln angezogen.

Die Zellen werden pelletiert, in 200 µl ml [2 ml] Lysislösung I (0,9 M Sorbitol; 0,1 M EDTA; 14 mM β-Mercaptoethanol; 0,2% Lytikase) resuspendiert und bei 37°C für 1 Stunde inkubiert.

Danach werden die Zellen 5 min bei 3000xg pelletiert und der Überstand verworfen. Das Pellet wird in 250 µl [2,5 ml] Lysislösung II (15 ml 0,5 M EDTA; 6 ml 2 M Tris; 6 ml 10% SDS; 120 ml ddH<sub>2</sub>O) vorsichtig resuspendiert und eine Stunde bei 68°C inkubiert. Zur Fällung der Proteine werden 400 µl [4 ml] 5 M Kaliumacetatlösung (pH 4,2) zugegeben und vorsichtig gemischt. Die Proteine werden 10 min bei 13000xg pelletiert und aus dem Überstand die DNA mit Ethanol gefällt.

Bei der Isolation im analytischen Maßstab wird die DNA in 50 µl 1x TE (10 mM Tris-HCl; 1 mM EDTA; pH 8,0) mit 10 µg/ml RNase A aufgenommen und ist nach kurzer Inkubation bei Raumtemperatur bereit für Restriktionsanalysen.

### **3.2.1.2 Isolation von Gesamt-RNA aus Pilzen**

Zu den gefrorenen Zellen (ca. 10<sup>8</sup> Zellen) wird 1 ml Tri-Reagenz (PeqLab) gegeben und die Zellen 10 min mit Glasperlen aufgeschlossen. Nach 5 min Inkubation bei Raumtemperatur werden 200 µl Chloroform zugegeben, gemischt und weitere 10 min inkubiert. Zur Phasentrennung wird 5 min bei 13000xg zentrifugiert und aus der oberen, wässrigen Phase mit 500 µl Isopropanol die RNA ausgefällt. Die pelletierte Gesamt-RNA wird mit 70% Ethanol gewaschen und 10 min bei 50°C in 10 µl H<sub>2</sub>O gelöst. Von 0,5 µl der RNA-Lösung wird im Photometer bei einer Verdünnung von 1:100 die Konzentration bestimmt und 1 µg RNA in einem 1%igen Agarosegel auf Reinheit und Degradation überprüft.

### **3.2.1.3 Isolation von Plasmid-DNA aus Bakterien**

Zur Isolation von Plasmid-DNA aus *Escherichia coli* im analytischen Maßstab [präparativer Maßstab] werden die Zellen aus einer 3 ml Über-Nacht-Kultur [50 ml] in LB Medium mit Ampicillin verwendet.

Die Zellen werden bei 2700xg für 10 min bei 4°C zentrifugiert, in 250 µl [5 ml] Lysislösung (50 mM Glukose; 25 mM Tris-HCl; 10 mM EDTA; 0,3 mg/ml Lysozym; pH 8,0) resuspendiert und 10 min bei Raumtemperatur inkubiert. Zur Lyse der Zellen werden 250 µl [5 ml] frisch hergestellter Denaturierungslösung (0,2 N NaOH; 1% SDS) zugegeben, vorsichtig gemischt und 10 min bei Raumtemperatur inkubiert. Danach werden zur Fällung der Proteine und der an Proteine gebundenen genomischen DNA 250 µl [5 ml] 5 M Kaliumacetat zugesetzt, erneut gemischt und für weitere 10 min bei Raumtemperatur inkubiert. Anschließend wird bei 13000xg und 4°C für 20 min zentrifugiert. Aus dem Überstand wird die DNA mit 1 Volumen Isopropanol und 1/10 Volumen 5 M Na<sub>2</sub>Acetat (pH 4,3) gefällt. Das Präzipitat wird mit 70% Ethanol gewaschen, anschließend getrocknet und in 1 ml ddH<sub>2</sub>O gelöst.

### 3.2.2 Manipulation von Nukleinsäuren

#### 3.2.2.1 Reverse Transkription von mRNA zu cDNA

Um isolierte mRNA für Expressionsanalysen zu cDNA umzuschreiben, muss zunächst sicher gestellt werden, dass der Hintergrund an schon vorhandener DNA so klein wie möglich gehalten wird. Daher werden 0,5-1 µg Gesamt-RNA zunächst mit DNase I (Promega) behandelt. Dazu wird die RNA mit 1 µl (40 u) RNase-Inhibitor (RNaseOut, Invitrogen), 1 µl 10x DNase-Puffer und 1 µl (1 u) DNase versetzt und mit ddH<sub>2</sub>O auf 10 µl aufgefüllt. Der Ansatz wird 15 min bei Raumtemperatur inkubiert. Zur Inaktivierung der DNase wird 1 µl EDTA (0,25 mM; pH 8,0) zugegeben und 15 min bei 65°C inkubiert. Der Ansatz wird dann auf Eis gekühlt und 9 µl RT-PCR Gemisch (1 µl Superscript; 4 µl 5x Puffer; 1 µl Oligo(dT)<sub>23</sub>-Primer; 1 µl dNTP-Mix; 2 µl DTT) zugegeben und im Thermocycler 60 min bei 42°C und anschließend 10 min bei 70°C inkubiert.

#### 3.2.2.2 Sequenzierungen

Alle Sequenzierungen wurden mit dem ABI BigDye System durchgeführt. Dazu wurde in einem Gesamtvolumen von 10 µl jeweils 1 µg der zu sequenzierenden DNA, mit 10 pmol Sequenzierprimer und 2 µl BigDye Mix vermischt, mit dem Standard-Sequenzierungsprotokoll im Thermocycler behandelt. Die Proben wurden in einem 3100-Avant Genetic Analyser (ABI Prism) analysiert.

#### 3.2.2.3 Klonierung von PCR-Produkten in TOPO<sup>®</sup>-Vektoren

PCR-Produkte, die nur für eine Sequenzierung kloniert werden sollten, wurden mit dem TOPO-TA System (Invitrogen) kloniert. Dazu wurden sie zunächst mit Taq-Polymerase (Invitrogen) von genomischer DNA oder cDNA amplifiziert und dann nach Herstellerangaben in den Vektor ligiert.

PCR-Produkte, die für die Expression in *E. coli* bestimmt waren, wurden in den Vektor pET-D100 (Invitrogen) kloniert. Dazu wurden sie zunächst mit Pfu-Polymerase (Fermentas Life Science) von genomischer DNA amplifiziert und danach entsprechend den Herstellerangaben in den Vektor ligiert.

#### 3.2.2.4 Vektorklonierung

PCR-Produkte, die in einen spezifischen Vektor kloniert werden sollten, wurden nach folgendem Schema behandelt [82]. Um zu verhindern, dass residuale Taq-Polymerase während des

folgenden Restriktionsverdau die überstehenden Enden wieder auffüllt, wurde die Taq-Polymerase durch Zugabe von Proteinase K zum PCR-Produkt und Inkubation für 1 h bei 37°C proteolytisch abgebaut. Die Proteinase K wurde anschließend durch 2-malige Phenol-Chloroform-Extraktion entfernt und die DNA anschließend gefällt. Wurde als Template-DNA in der PCR ein Plasmid verwendet, so wurde dieses durch Zugabe von *DpnI* im folgenden Restriktionsverdau abgebaut. *DpnI* schneidet methylierte DNA an der Sequenz GATC, also ausschließlich Plasmid-DNA aber kein PCR-Produkt. Anschließend wurde das verdaute PCR-Produkt durch Auftrennung im Agarosegel gereinigt. Der Zielvektor wurde mit den gleichen Enzymen, die für die Restriktion der PCR-Produkte verwendet wurden, geöffnet und mittels Agarosegelelektrophorese aufgereinigt. Die Ligation von PCR-Produkt und Vektor erfolgte in einem Gesamtvolumen von 20 µl mit T4-Ligase. Dabei wurden pro Vektor-Insert-Kombination verschiedene Mengenverhältnisse verwendet. 5 µl des Ligationsansatzes wurden in ein Aliquot dauerkompetenter DH5- $\alpha$  *E. coli* Zellen transformiert (Abschnitt 3.2.2.5).

### 3.2.2.5 Herstellung und Transformation dauerkompetenter *Escherichia coli* Zellen

Die Herstellung dauerkompetenter *E. coli* Zellen erfolgt in Anlehnung an die Methode von Hanahan [83]. Die Zellen werden in Gegenwart von Antibiotika entsprechend der Resistenzen des Stammes (keine bei DH5- $\alpha$ , 50 mM Chloramphenicol bei Rosetta) in einer 3 ml Über-Nacht-Kultur angezogen. Daraus wird eine 200 ml Kultur angeimpft und bis zu einer optischen Dichte ( $OD_{600}$ ) von 0,4-0,6 bei 37°C unter Schütteln angezogen. Nach Erreichen der gewünschten Zelldichte werden die Zellen geerntet und in 40 ml Puffer 1 (30 mM Kaliumacetat; 50 mM  $MnCl_2$ ; 100 mM  $RbCl$ ; 10 mM  $CaCl_2$ ; 15% Glycerol; pH 5,8) gewaschen. Anschließend werden die Zellen vorsichtig in 8 ml Puffer 2 (10 mM MOPS; 75 mM  $CaCl_2$ ; 10 mM  $RbCl$ ; 15% Glycerol; pH 7,0) resuspendiert und in Fraktionen à 80 µl in flüssigem Stickstoff schockgefroren.

Zur Transformation werden die Zellen auf Eis aufgetaut, mit Plasmid-DNA bzw. einem Aliquot eines Ligationsansatzes versetzt und nach 5 min Inkubation auf Eis 30 sec einem Hitzeschock bei 42°C ausgesetzt. Die Zellen werden dann mit 0,5 ml LB Medium versetzt, 45-60 min bei 37°C geschüttelt und anschließend auf LB-Platten mit Antibiotikum zur Selektion ausplattiert.

### 3.2.2.6 Koloniehybridisierung

Die Koloniehybridisierung wird beim Durchmustern großer Mengen von *E. coli* Klonen verwendet. Die Einzelkolonien werden dazu entweder jeweils auf zwei auf LB/Amp-Platten liegenden Nylonmembranen (Amersham) ausgestrichen und für mindestens 3 h bei 37°C

wachsen gelassen, oder es werden Replikate der Platten durch Abklatsch angefertigt. Danach werden die Zellen auf der ersten Membran bzw. der Replik in drei Schritten aufgeschlossen: 15 min Denaturierung (0,5 N NaOH; 1,5 M NaCl), 15 min Neutralisierung (1 M Tris-HCl; 1,5 M NaCl; pH 7,5), 10 min Äquilibrierung (2x SSC: 0,3 M NaCl; 0,03 M Na<sub>3</sub>Citrat). Die andere Membran dient später als Quelle für lebende Zellen. Die erste Membran wird kurz getrocknet und zur Fixierung der DNA im UV-Crosslinker (Stratagene) mit 1200 J/cm<sup>2</sup> bestrahlt. Die verbliebenen Zelltrümmer werden mit einem Tuch abgewischt. Die Detektion der gesuchten DNA wurde mit dem Dig-Easy-System (Roche) durchgeführt. Hierfür werden mittels PCR spezifische, mit Digoxigenin markierte DNA-Sonden hergestellt und in Hybridisierungspuffer (DIG-EasyHyb+2% w/v Blocking Reagent) aufgenommen. Die Membran wird zunächst mindestens 2 h in Hybridisierungspuffer (ohne Sonde) vorinkubiert und dann über Nacht mit der Sonde bei 42°C hybridisiert. Um die überschüssige Sonde zu entfernen, wird die Membran am Folgetag 2x 5 min bei Raumtemperatur in W1 (2x SSC; 0,1% w/v SDS) und 2x 15 min bei der Hybridisierungstemperatur in W2 (0,5x SSC; 0,1% w/v SDS) gewaschen. Zur Blockierung der unspezifischen Proteinbindungsstellen wird die Membran 30 min in Blockierungslösung (1% w/v Blocking Reagent; 100 mM Maleinsäure; 150 mM NaCl; pH 7,5) inkubiert, dann der DIG-Antikörper in einer Verdünnung von 1:7500 zugegeben und weitere 30 min geschüttelt. Der überschüssige Antikörper wird durch dreimaliges, 15 minütiges Waschen in Waschpuffer (0,3% v/v Tween80; 100 mM Maleinsäure; 150 mM NaCl; pH 7,5) entfernt. Die Membran wird mindestens 5 min in Äquilibrierungspuffer (100 mM NaCl; 100 mM Tris-HCl; pH 9,0) gewaschen und zur Detektion wird entweder das Substrat NBT/BCIP (Boehringer) für die an den Antikörper gebundene alkalische Phosphatase zugegeben, welches nach Reaktion ein dunkles Präzipitat auf der Membran bildet, oder das Chemilumineszenz CDP-Star verwendet und entsprechend ein Film mit dem Blot exponiert.

### 3.2.3 Verwendete Oligonukleotide

Primer	Sequenz
<i>Candida glabrata</i>	
CgKEX2ssATG	CGGATCCGTATGAAGTGGTGGACAAG
CgKEX2Hisrev	CCCGCGGCCGCTTAATGATGATGATGATGATGATGATCAAATGAACACCAGTTGG
<i>Candida albicans</i>	
CaKEX2ssATG	CGGATCCGCTATGTTGCCAATAAAAATTAC
CaKEX2revHIS0	CCCGCGGCCGCTTAATGATGATGATGATGATGATGGTCTTCTTCACCTTCTTGTTTAT
CaKEX2revHIS1	CCCGCGGCCGCTTAATGATGATGATGATGATGATGAGAATTGGCCTTTTTTGTGGCC
CaKEX2revHIS2	CCCGCGGCCGCTTAATGATGATGATGATGATGATGTCTGTATTATCAGAATTGGC
CaKEX2revHIS3	CCCGCGGCCGCTTAATGATGATGATGATGATGATGTGACATAAGGTAAAATCCTGT
CaKEX2_ACTfwd	CTATGTTGCCAATAAAAATTACTAATA
CaKEX2_ACTrev	GGGCTAGCTTACAGAATTGGCCTTTTTTGTGGCC
<i>Cryptococcus neoformans</i>	
CnKEX2ssATG	CGGATCCGGCAAAATGCGCACCTTATTATCC
CnKEX2-Hisrev	CCCGCGGCCGCTTAATGATGATGATGATGATGATGCCAGCCAGGGAAACCACTATCTGC
CnKEX2revHIS2	CCCGCGGCCGCTTAATGATGATGATGATGATGATGGCCGACGAAGGGTTCTTCAGC
CnKEX2revHIS3	CCCGCGGCCGCTTAATGATGATGATGATGATGATGCCATGTAGAGGATGAAGCGAG

**Tabelle 7: Zur Konstruktion von *KEX2*-Expressionsvektoren verwendete Oligonukleotide.**

Insert	Sequenz
EGFPf	GGGATTCTCTAAAGGTGAAGAATTA
EGFPPr	GCGATGCGGGCGGCTCTATTTGTACAATTCATCC
CaCcw14f	CCGGATCCAATTATTATTCCAATCAACAGG
CaCcw14r	CCGGATCCTGATTCATTAGCATTTTTAGAA
CA0365	CCGGATCCGCTATTCCACAAGAATCAACCA
CA0365r	CCGGATCCTAAGGTTCAAAATGTTAAGAGC

**Tabelle 8: Zur Konstruktion von GST-GFP Fusionsproteinen verwendete Oligonukleotide.**

Primer	Sequenz
<b>pIC3.5-spezifische Vektorprimer</b>	
AOX 3'	GCAAATGGCATTCTGACATCC
AOX 5'	GACTGGTTCCAATTGACAAGC
<b><i>Candida glabrata</i></b>	
OB D13-1	TATGAGTGTAGCTCACTGGG
OB D13-2	GAACGCAATATACTTAAGTAGC
KEX2-5'	ACGACAATTAGGTCATGAACAAG
KEX2-3'	CTCTACCTCTATAAGTAGAATC
KAP2 ob1	AGGATGCCCCGTACCAGCACC
KAP2 ob2	TGCCATCATAGCAGCCACC
OBUp1	GTAGTGGTTATCTAGACCCC
OBUp2	GAGGAGGTGCGGTACGGGC
OBdown2	TGCTATGATAGGCATGGAGG
<b><i>Candida albicans</i></b>	
Ca Kex2 seq3f	CCTGTCCCCATTGAAATGG
Ca Kex2 seq3r	GTGGTAGATTTGGTGATTCTTG
CaKex2 rev 2	CTTAAGAATTGGCCTTTTTGTTGCC
CaKex2 seq-midf	GGGATTAGAATATTATCTGGCC
CaKex2rev4	CTATCCTGTATTATCAGAATTGGCC
<b><i>Cryptococcus neoformans</i></b>	
Cn seq2f	CCAATATCCGATGCCGACGAAGC
Cn seq2r	GAAGTGACCATGCGACGAAAC
Cneo Kex2 seq3f	GTTATGGAAAGCTTGACGCAGG
cneo seq1f	AGCATTGCCTTCAAGTCTTGGC
cneo seq1r	CCTCAATGCAGTCCTCTCATCC
CneofKEX2f	CGCCATTTGTTTTATTTTCTTTA
CneofKEX2r	TTGCTGGATCGCTCTACACTCTTC
Cnfwd	ATGGGCACCTTATTATCCCTTTGG

Tabelle 9: Zur Sequenzierung von *KEX2*-Genen und Konstrukten verwendete Oligonukleotide.

Topo-Substrat	Forwardprimer	Rückwärtsprimer
1 CA1855	CA1855F	CA1855R
2 CA1856	CA1856F	CA1856R
3 CA1857	CA1857F	CA1857R
4 CA1858	CA1858F	CA1858R
5 CA1859	CA1859F	CA1859R
6 CA1860	CA1860F	CA1860R
7 CA1861	CA1861F	CA1861R
8 CA1862	CA1862F	CA1862R
9 CA1863	CA1863F	CA1863R
10 CA1864	CA1864F	CA1864R
11 CA1865	CA1865F	CA1865R
12 CA1866	CA1866F	CA1866R
13 CA1867	CA1867F	CA1867R
14 CA1868	CA1868F	CA1868R
15 CA1869	CA1869F	CA1869R
16 CA1870	CA1870F	CA1870R
17 CA1871	CA1871F	CA1871R
18 CA1872	CA1872F	CA1872R
19 CA1873	CA1873F	CA1873R
20 CA1874	CA1874F	CA1874R
21 CA1875	CA1875F	CA1875R
22 CA1876	CA1876F	CA1876R
23 CA1877	CA1877F	CA1877R
24 CA1878	CA1878F	CA1878R
25 CA1879	CA1879F	CA1879R
26 CA1880	CA1880F	CA1880R
27 CA1881	CA1881F	CA1881R
28 CA1882	CA1882F	CA1882R
29 CA1883	CA1883F	CA1883R
30 CA1884	CA1884F	CA1884R
31 CA1885	CA1885F	CA1885R
32 CA1886	CA1886F	CA1886R
33 CA1887	CA1887F	CA1887R
34 CA1888	CA1888F	CA1888R
35 CA1889	CA1889F	CA1889R
36 CA1890	CA1890F	CA1890R
37 CA1891	CA1891F	CA1891R
38 CA1892	CA1892F	CA1892R
39 CA1893	CA1893F	CA1893R
40 CA1894	CA1894F	CA1894R
41 CA1895	CA1895F	CA1895R
42 CA1896	CA1896F	CA1896R
43 YFL187W	YFL187WF	YFL187WR

Tabelle 10: Zur Konstruktion von Substrat-Expressionsvektoren verwendete Oligonukleotide.



### 3.3 Proteinbiochemische Methoden

#### 3.3.1 SDS-PAGE und Westernblot

##### 3.3.1.1 Herstellung von Polyacrylamidgelen

Zur Auftrennung von Proteinen mittels Polyacrylamidgelelektrophorese (PAGE) wurde die Apparatur Mini Protean III (BioRad) verwendet. Die Gelelektrophorese wurde bei 160 V durchgeführt, bis die Laufmittelfront der Proteine ca. 1 cm von der Unterseite des Gels entfernt war. Zur Auftrennung von Peptiden und kleineren Proteinen (<20 kDa) wurden Tris-Tricingele verwendet. Die Zusammensetzungen der Puffer und Gele sind in Tabelle 11 und Tabelle 12 angegeben.

##### 3.3.1.2 Coomassiefärbung

Zur Färbung mit Coomassie Reagenz wurden die Gele für 2 h in Coomassiefärbelösung (0,1% Coomassie Brilliant Blau; 20% v/v Methanol; 0,5% v/v Eisessig) unter Schütteln inkubiert und der Hintergrund mit Entfärberlösung (50% v/v H<sub>2</sub>O; 40% v/v Methanol; 10% v/v Eisessig) nach Sicht entfärbt.

##### 3.3.1.3 Silberfärbung

Für die Silberfärbung wurden die Polyacrylamidgele zunächst 20 min in Fixierer (50% v/v Methanol; 10% v/v Essigsäure; 0,05% v/v Formaldehyd) inkubiert. Reste des Fixierers wurden durch zweimaliges Waschen in 50% v/v Ethanol entfernt und das Gel 1 min Lösung A (0,0002% v/v Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> x 5H<sub>2</sub>O) inkubiert. Nach zweimaligem Waschen in ddH<sub>2</sub>O wurde das Gel 10-15 min in Lösung B (0,16% w/v AgNO<sub>3</sub>, 0,72% v/v Formaldehyd) inkubiert. Ungebundenes Silber wurde durch zweimaliges Waschen mit ddH<sub>2</sub>O entfernt, das Gel in einem frischen Gefäß in Lösung C (4% w/v Na<sub>2</sub>CO<sub>3</sub>, 0,00005% Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> x 5H<sub>2</sub>O; 0,4% Formaldehyd) nach Sicht entwickelt und schließlich in Fixierer gestoppt.

	Tris-Glycin Gele		Tris-Tricin Gele	
<b>Sammelgele</b>	Sammelgelpuffer	1,26 ml	Puffer mit SDS	1,89 ml
	10% SDS	50 µl		
	10% APS	50 µl	10% APS	75 µl
	TEMED	5 µl	TEMED	75 µl
	ddH <sub>2</sub> O	2,8 ml	ddH <sub>2</sub> O	7,5 ml
	30% Acrylamid	840 µl	Acrylamid	4,2 ml
<b>Trenngele</b>				
	Trenngelpuffer	3,8 ml	Puffer mit SDS	5,7 ml
	10% SDS	150 µl	Glycerol	225 µl
	10% APS	150 µl	10% APS	225 µl
	TEMED	6 µl	TEMED	9 µl
7,5%	ddH <sub>2</sub> O	7,15 ml	ddH <sub>2</sub> O	10,73 ml
	Acrylamid	3,75 ml	Acrylamid	5,625 ml
10,0%	ddH <sub>2</sub> O	5,9 ml	ddH <sub>2</sub> O	8,85 ml
	Acrylamid	5 ml	Acrylamid	7,5 ml
12,0%	ddH <sub>2</sub> O	4,9 ml	ddH <sub>2</sub> O	7,35 ml
	Acrylamid	6 ml	Acrylamid	9 ml
15,0%	ddH <sub>2</sub> O	3,4 ml	ddH <sub>2</sub> O	5,1 ml
	Acrylamid	7,5 ml	Acrylamid	11,25 ml

Tabelle 11: Zusammensetzung von Acrylamidgelen.

Puffer	Zusammensetzung
Laufpuffer nach Lämmli (5-fach)	0,125 M Tris-Base 0,96 M Glycin 0,5% (w/v) SDS
Probenpuffer (4-fach)	62,5 mM Tris-HCl 20% (v/v) Glycerol 2% (w/v) SDS 0,1 M DTT 0,01% (w/v) Bromphenolblau
Tris-Glycin Sammelgelpuffer	0,5 M Tris-HCl, pH 6,8
Tris-Glycin Trenngelpuffer	1,5 M Tris-HCl, pH 8,8
Tris-Tricin Gelpuffer	3 M Tris-HCl, pH 8,45 0,3% SDS
Kathodenpuffer	0,2 M Tris-Base 0,2 M Tricin
Anodenpuffer	0,4 M Tris-Base, pH 8,8

Tabelle 12: Zusammensetzung der Puffer für PAGE.

### 3.3.1.4 Westernblot

Zum Transfer von Proteinen aus Polyacrylamidgelen auf PVDF Membranen wurde die Trans-Blot Apparatur (BioRad) verwendet. Der Transfer der in der SDS-PAGE aufgetrennten Proteine auf eine PVDF-Membran der Firma Millipore (Immobilon P bei Proteinen größer als 20 kDa, Immobilon P<sup>SO</sup> bei Proteinen/Peptiden unter 20 kDa) erfolgte in Transferpuffer nach dem Semi-Dry-Verfahren. Das Gel wurde entweder 2 h bei 400 mA/60-70 V, oder ü. N. bei 100 mA/60-70 V geblottet. Zur Sättigung unspezifischer Bindungsstellen wurde die Membran 2 h bei RT in Blocklösung (10% w/v Milchpulver in PBS (0,8% w/v NaCl; 0,02% w/v KCl; 0,144% w/v Na<sub>2</sub>HPO<sub>4</sub>; 0,024% w/v KH<sub>2</sub>PO<sub>4</sub>; pH 7,4); 0,5% v/v Tween20) geschwenkt. Anschließend wurde der Primärantikörper (Qiagen) in Blocklösung verdünnt zugegeben (Verdünnung 1:10000 für

Anti-His) und ü. N. bei 4°C inkubiert. Die Membran wurde 3x 15 min mit Waschlösung (PBS+0,5% v/v Tween20) gewaschen und der Peroxidase-gekoppelte Zweitantikörper (Ziege-anti-Maus-Peroxidase) in Blocklösung 1:10000 verdünnt zugegeben. Die Inkubation erfolgte für 1 h bei RT. Die Membran wurde erneut gewaschen und das Signal mit Luminol (Pierce) detektiert.

### **3.3.1.5 Größen-Ausschluss-Chromatographie**

Zum Säubern von Proteinlösungen von hochmolekularen Aggregaten wurden die Proteinlösungen durch einen Sterilfilter mit einer Porengröße von 0,22 µm filtriert.

Zum Konzentrieren und Trennen von Proteinen anderer Größen wurden die Größen-Ausschluss-Systeme Centricon 20 (Millipore) für Größenordnungen ab mehreren ml und die Systeme Microcon YM10/YM30/YM50 (Amicon) für Größenordnungen von wenigen µl bis wenigen ml verwendet.

### **3.3.2 Anionen-Austausch-Chromatographie**

Für die Reinigung von Proteinen mittels Anionen-Austausch-Chromatographie wurden HiTrap ANX FF Säulen (Amersham Biosciences) verwendet. Die sterilfiltrierten Kulturüberstände wurden zunächst aufkonzentriert (Abschnitt 3.3.1.5) und in Bindepuffer (20 mM Bis-Tris; 10 mM NaCl; pH 4,5) aufgenommen. Die Proteine aus der Lösung wurden auf die Säule gebunden und die Säule anschließend mit 20 ml Bindepuffer gewaschen. Zur Elution der Proteine wurde die Säule mit 3 ml Elutionspuffer (20 mM Bis-Tris; 100 mM NaCl; pH 5,5) gespült.

### **3.3.3 Heterologe Expression von Proteinen in *E. coli***

#### **3.3.3.1 Expression**

Die Expression rekombinanter Proteine in *E. coli* erfolgte durch Induktion mit Laktose in Über-Nacht-Kultur (LB-Medium 0,04% Glukose; 0,2% Laktose [84, 85]) bei RT, 25 oder 30°C und 200 rpm im Schüttler. Ein Zellpellet aus 50 ml Kulturvolumen wurde in 5 ml Lysislösung (Bugbuster (Novagen) mit 20-50 mM Imidazol, 1% v/v Protease Inhibitor Cocktail III (Calbiochem), 10 u Benzonase (Novagen) und eine kleine Spatelspitze Lysozym (Sigma)) 20-30 min bei RT unter Bewegung lysiert. Durch Zentrifugation wurden lösliche Bestandteile von unlöslichen getrennt. Lysat und Pellet wurden dann, wie in den folgenden Abschnitten beschrieben, weiterbehandelt.

### **3.3.3.2 Aufreinigung von löslichen Proteinen mit His-tag**

Proteine mit 6xHis-tag wurden mittels Affinitätschromatographie isoliert. Dazu wurde Ni-NTA-Agarose (Novagen) verwendet: das Zelllysat wurde auf die Ni-NTA-Agarose gebunden, intensiv mit Waschpuffer (40 mM Imidazol; 0,5 M NaCl; 20 mM Tris-HCl; pH 8,0) gewaschen und mit Elutionspuffer (350 mM Imidazol; 0,001% v/v Triton X-100; 0,5 M NaCl; 20 mM Tris-HCl; pH 8,0) in 1 ml-Schritten eluiert.

### **3.3.3.3 Aufreinigung von löslichen Proteinen mit GST-tag**

Zur Aufreinigung mit Glutathion-S-Transferase fusionierter Proteine wurde Glutathion-Sepharose (Novagen) verwendet. Nach Bindung der Proteine aus dem Lysat für ca. 20 min wurde die Sepharose intensiv mit GST-Waschpuffer (4,3 mM Na<sub>2</sub>HPO<sub>4</sub>; 1,5 mM KH<sub>2</sub>PO<sub>4</sub>; 0,14 mM NaCl; 0,27 mM KCl; pH 7,2) gewaschen und anschließend in 1 ml-Schritten mit Elutionspuffer (Waschpuffer mit 10 mM reduziertem Glutathion) eluiert.

### **3.3.3.4 Renaturierung von unlöslichen Proteinen aus Inklusionskörpern**

#### **Aufreinigung von Protein aus Inklusionskörpern**

Ein Zellpellet aus einer 50 ml ü. N. Laktose-Expressionskultur wurde, wie oben beschrieben, lysiert und erneut zentrifugiert. Der Überstand wurde verworfen und das verbleibende Pellet in 5 ml Harnstoff-Puffer (8 M Urea; 50 mM Imidazol; 1% v/v Proteaseinhibitor Cocktail II (Calbiochem) in Ni-NTA-Lysispuffer (s.o.)) gelöst, ggf. ü. N. bei 60°C. Das gelöste Protein wurde anschließend durch Zentrifugation von unlöslichen Bestandteilen getrennt, auf 1 ml Ni-NTA-Agarose (Bettvolumen) gegeben und die Säule/das Gefäß für 10 min sanft geschüttelt. Optional kann die Agarose mit Isopropanol gewaschen werden, um Verunreinigungen wie Endotoxine zu entfernen [86]. Die Agarose wurde dann mindestens zwei Mal mit Harnstoff-Puffer gewaschen, um Reste anderer Lösungsmittel und unspezifisch gebundenes Protein zu entfernen. Sie war danach für die Renaturierung der gebundenen Proteine bereit.

#### **Renaturierung mittels Harnstoff**

Die Renaturierung mit Harnstoff beruht auf der sequentiellen Entfernung des Harnstoffs als chaotropes Reagenz aus den Proteinen. Das kann entweder durch Pufferwechsel oder durch Verdünnen geschehen. Bei an Ni-NTA-Agarose gebundenen Proteinen bot sich der Pufferwechsel an. Das geschah durch Wechsel mit Harnstoffpuffern immer geringerer Konzentration (4 M; 3 M; 2 M; 1 M; 0 M Harnstoff in Ni-NTA-Lysispuffer ohne Inhibitoren und

Imidazol) und jeweils ca. 12 h Inkubation bei RT unter Bewegung. Essentiell war hier ein ausreichendes abschließendes Waschen mit normalem Ni-NTA-Waschpuffer, um verbleibenden Harnstoff auszuwaschen. Nach der Faltung konnte die Säule wie oben eluiert werden.

### **Renaturierung mittels TritonX-100 und $\beta$ -Cyclodextrin**

Bei Renaturierung mit TritonX-100 und  $\beta$ -Cyclodextrin [87] wurde der Harnstoff zunächst durch intensives Waschen der an die Ni-NTA-Agarose gebundenen Proteine mit Puffer A (0,1 M NaCl; 20 mM Tris; 0,1% Triton X-100; pH 8,0) entfernt und durch Triton X-100 ersetzt. Überschüssiges Triton wurde durch Waschen mit 20 ml Puffer B (0,1 M NaCl; 20 mM Tris; 5 mM  $\beta$ -Cyclodextrin; pH 8,0) entfernt. Die Renaturierung erfolgte ü. N. in 10 ml Puffer B unter Bewegung. Die Säule wurde am Folgetag mit 20 ml Ni-NTA-Lysispuffer gewaschen und, wie oben beschrieben, eluiert.

### **3.3.4 Transformation von Pilzen mit rekombinanter DNA**

Pilzzellen von *P. pastoris* oder *C. albicans* wurden nach einem aus Lin-Cereghino *et al.* [88] abgewandelten Elektroporationsprotokoll transformiert: eine 50 ml Zellkultur wurde bis zu einer OD<sub>600</sub> von 0,8 angezogen, die Zellen durch Zentrifugation (500xg, RT) geerntet, in 9 ml eiskaltem TEDS-Puffer (10 mM Tris-Base; pH 8,3; 3% v/v Polyethylenglycol; 5% v/v Dimethylsulfoxid; 1 M Sorbitol) und 1 ml 1 M Dithiothreitol resuspendiert und 5 min unter leichtem Schütteln bei RT inkubiert. Die Zellen wurden danach erneut pelletiert (500xg, RT), in 1 ml TEDS ohne Dithiothreitol resuspendiert und in Aliquots à 50  $\mu$ l in flüssigem Stickstoff schockgefroren. Zur Elektroporation wurden die Zellen auf Eis aufgetaut, mit 1  $\mu$ g Plasmid-DNA in max. 20  $\mu$ l Lösung in einer Elektroporationsküvette (d= 2 mm) vermischt und 10 min auf Eis inkubiert. Die Küvette wurde in den Elektroporator transferiert und der Puls (450 V; 150  $\mu$ F) sofort gestartet. Nach dem Puls wurde die Küvette wieder auf Eis gestellt, die Zellen später auf 3 SD-Agarplatten ausplattiert und diese bei 30°C inkubiert. Die ersten Kolonien erschienen nach 2-3 Tagen.

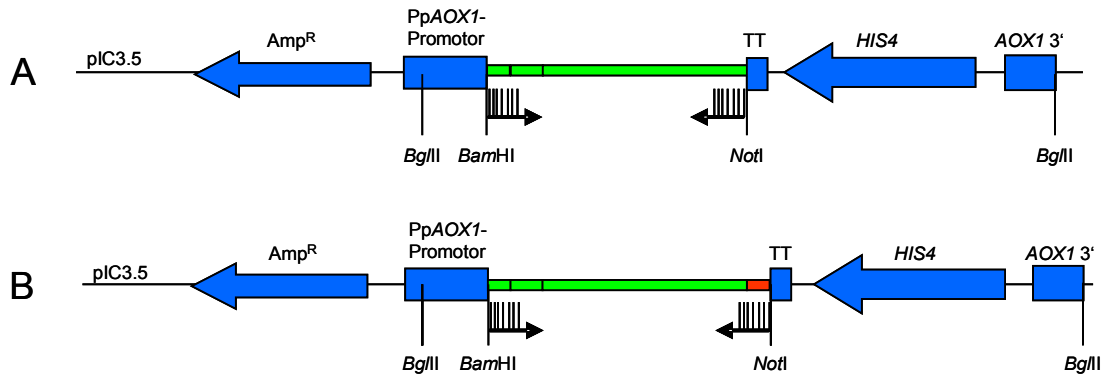
### **3.3.5 Expression rekombinanter Proteine in *Pichia pastoris***

#### **3.3.5.1 Herstellung und Transformation der Expressionskonstrukte**

Zur heterologen Expression von rekombinanten Kex2 Proteasen wurde das Plasmid pIC3.5 aus dem „*Pichia* Expression kit“ (Invitrogen) verwendet. Zur Herstellung der spezifischen Konstrukte wurde das Zielgen mittels PCR amplifiziert und über *Bam*HI und *Not*I

Restriktionsschnittstellen in das Plasmid pIC3.5 eingesetzt. Es wurden sowohl Konstrukte mit C-terminalem 6xHis-tag als auch ohne erstellt (Abbildung 10).

Zellen des Stammes GS115 wurden mittels Elektroporation mit durch *Bgl*II linearisierter Plasmid-DNA transformiert und auf Minimalmedium selektiert (siehe Abschnitt 3.3.4).



**Abbildung 10: Schematische Darstellung der verwendeten pIC3.5-basierten Expressionskonstrukte.** Vektorrückgrad (blau) mit Resistenzgen (Amp<sup>R</sup>) zur Vermehrung in *E. coli*, Selektionsmarker (HIS4) für *P. pastoris* und die von den zur Linearisierung verwendeten *Bgl*II Restriktionsschnittstellen flankierten AOX1-homologen Bereiche. Das zu exprimierende Gen (grün) wurde entweder ohne (A) oder mit (B) im rückwärtsgewandten Primer enthaltenen C-terminalen 6xHis-tag (rot) fusioniert.

### 3.3.5.2 Testexpression

Zur Identifizierung von Transformanten, welche das gewünschte Protein sezernieren, wurden die einzelnen Klone in 3 ml BMM Medium über Nacht bei 30°C angezogen und am Folgetag auf Proteaseaktivität (siehe Abschnitt 3.3.7) im Kulturüberstand überprüft.

Die Transformanten mit der höchsten Aktivität unter diesen Bedingungen wurden für die Herstellung größerer Proteinmengen verwendet.

### 3.3.6 Messung der Phosphataseaktivität

Die Aktivität von Phosphatasen wurde mittels der Umsetzung von p-Nitrophenylphosphat (pNPP) zu p-Nitrophenol bestimmt. Die Konzentration des Endproduktes dieser Reaktion kann photometrisch durch Messung der Absorption bei 405 nm gemessen werden. Da die Absorptionsstärke von p-Nitrophenol vom pH-Wert des Milieus abhängt und im sauren Bereich am niedrigsten ist, muss die Lösung vor der Messung auf einen alkalischen pH-Wert gebracht werden und für jeden verwendeten pH-Wert eine eigener Referenzwert verwendet werden.

Für jeden Messwert werden 50 µl 0,1 M KAc-Puffer vorgelegt, jeweils mit und ohne Enzym. Die Reaktion wird durch Zugabe von 50 µl des Substrats (5 mg/ml) gestartet und bei Raumtemperatur inkubiert. Die Reaktionen werden nach 1-3 h durch Zugabe von 200 µl 0,2 M NaOH gestoppt und die Absorption bei 405 nm gemessen.

### 3.3.7 Messung der Kex2-Aktivität

Zur Messung der Aktivität von Kex2 und anderen Proteasen mit ähnlicher Spezifität wurde das Substrat Benzyloxycarbonyl-Tyr-Lys-Arg-para-nitroanilid (Z-Tyr-Lys-Arg-pNA, Bachem, Weilheim) verwendet. Umsetzung des Substrates kann durch Messung der Absorption bei 405 nm verfolgt werden. Zur Messung wurde ein Spektrophotometer (Tecan) eingesetzt. Die Reaktion erfolgte in einem Gesamtvolumen von 100 µl mit variablen Enzymmengen und einer Standardmenge Substrat (1 µl einer Stocklösung von 50 mM Substrat/H<sub>2</sub>O). Substrat und Enzym wurden jeweils getrennt in Reaktionspuffer (50 mM Tris-HCl; 0,2 M KCl; 1 mM CaCl<sub>2</sub>; 0,001% Triton X-100) vorbereitet und die Reaktion durch Vermischen der beiden Komponenten gestartet. Die Messung von kinetischen Daten wurde unmittelbar nach dem Start der Reaktion begonnen.

### 3.3.8 Proteolytischer Verdau mit Kex2-homologen Proteasen

Der proteolytische Verdau von heterolog exprimierten Substraten erfolgte in einem Gesamtvolumen von 50 oder 100 µl in Reaktionspuffer (50 mM Tris-HCl; 0,2 M KCl; 1 mM CaCl<sub>2</sub>; 0,001% Triton X-100) bei 37°C. Standardmäßig wurden 3 µl einer zuvor auf ihre Aktivität eingestellte Protease verwendet. Das Volumen des eingesetzten Substrats war abhängig von dessen Konzentration und den Färbereigenschaften des Proteins in der SDS-PAGE.

## 3.4 Statistische Verfahren und Datenbanken

### 3.4.1 Matthews-Korrelation

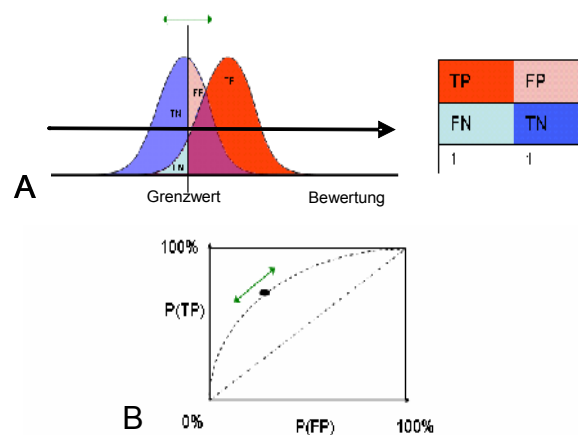
Der Matthews-Korrelations-Koeffizient ist ein Maß der Vorhersagequalität, der sich aus dem Verhältnis von richtigen und falschen positiven und negativen Beispielen errechnet. Ein perfekter Vorhersagealgorithmus hätte einen Wert von 1 (nur richtige), ein Algorithmus mit einem Wert von 0 entspräche einer Zufallsverteilung. Ein Wert von -1 entspräche „nur falsch“.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

**Formel 1: Matthews-Korrelations-Koeffizient.**

### 3.4.2 Receiver Operating Characteristic Kurve

Die Receiver Operating Characteristic (ROC)-Kurve ist eine Methode zur Bewertung und Optimierung von Analysestrategien. Die ROC-Kurve stellt die Abhängigkeit der Effizienz von der Fehlerrate dar und ermöglicht so die Grenzwertoptimierung (Abbildung 11). Es werden für jeden möglichen Grenzwert (0..max, in Intervallen von 0,01) die Spezifität (Richtignegativrate,  $TN/(TN+FP)$ ) und Sensitivität (Richtigpositivrate,  $TP/(TP+FN)$ ) ermittelt (Abbildung 11A) und die Fläche unter der resultierenden ROC Kurve (AUC, „area under curve“) mittels Trapezoidapproximation ermittelt [89] (Abbildung 11B).



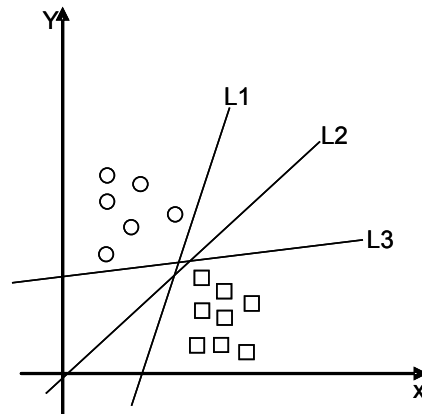
Bildquelle: de.wikipedia.org/wiki/Receiver\_Operating\_Characteristic

**Abbildung 11: Beispieldiagramm ROC-Kurven.** A: Durch Verschieben des Grenzwertes auf einer Bewertungsskala lassen sich Richtigpositiv- und Richtignegativ-Rate optimieren. B: Das Integral der Auftragung von Richtigpositiv- und Falschpositiv-Rate (AUC) bei allen Grenzwerten im Intervall ist ein Maß für die Gesamteffizienz eines Algorithmus.

### 3.4.3 Separierende Hyperebenen/Support Vektoren

Die Support-Vector-Maschine (SVM) ist ein Lernalgorithmus zur Klassifizierung von Objekten. Die SVM bestimmt anhand einer Menge von Trainingsbeispielen eine Hyperebene, welche beide Klassen so voneinander trennt, dass der kleinste Abstand zur Hyperebene für die Beispiele beider Klassen maximiert wird, um eine möglichst gute Generalisierbarkeit des Klassifikators zu garantieren (Abbildung 12). Dieses Verfahren lässt sich unter Verwendung nicht-linearer Separatoren auch zur Klassifikation von Mehr-Klassen-Problemen anwenden. Hierzu wurde das Programm libSVM-prob verwendet [90].





**Abbildung 12: Eine separierende Hyperebene trennt zwei Klassen von Daten.** Alle drei Ebenen L1, L2 und L3 trennen die Datensätze „rund“ und „eckig“ anhand der Messparameter x und y voneinander, aber nur L2 verläuft optimal, mit größtem Abstand zu beiden Populationen.

### 3.4.4 Verwendete Software und Genomdatenbanken

Name	Version	Referenz
BioPerl	1.5	<a href="http://www.bioperl.org">www.bioperl.org</a>
TMHMM	1.0	<a href="http://cbs.dtu.dk/services/TMHMM">cbs.dtu.dk/services/TMHMM</a> [91]
TargetP	1.0	<a href="http://cbs.dtu.dk/services/TargetP">cbs.dtu.dk/services/TargetP</a> [92]
SignalP	2.0	<a href="http://cbs.dtu.dk/services/SignalP">cbs.dtu.dk/services/SignalP</a> [93]
Disopred	2.0	<a href="http://bioinf.cs.ucl.ac.uk/disopred">bioinf.cs.ucl.ac.uk/disopred</a> [94]
libSVM-prob	1.0	<a href="http://www.csie.ntu.edu.tw/~cjlin/libsvm/">www.csie.ntu.edu.tw/~cjlin/libsvm/</a> [90]
Mass	1.0	<a href="http://bioinfo3d.cs.tau.ac.il/MASS/">bioinfo3d.cs.tau.ac.il/MASS/</a> [95]
Rvp-net	1.0	<a href="http://gibk26.bse.kyutech.ac.jp/jouhou/shandar/netasa/rvp-net/">gibk26.bse.kyutech.ac.jp/jouhou/shandar/netasa/rvp-net/</a> [96]
RasMol		<a href="http://www.openrasmol.org/">www.openrasmol.org/</a> [97]

**Tabelle 13: Verwendete Software.**

Spezies	Webadresse
<i>Aspergillus fumigatus</i>	<a href="http://www.tigr.org/tdb/e2k1/afu1/">www.tigr.org/tdb/e2k1/afu1/</a>
<i>Aspergillus niger</i>	<a href="http://www.broad.mit.edu/annotation/genome/aspergillus_group/">www.broad.mit.edu/annotation/genome/aspergillus_group/</a>
<i>Ashbya gossypii</i>	<a href="http://agd.unibas.ch/">agd.unibas.ch/</a>
<i>Candida albicans</i>	<a href="http://candidagenome.org/">candidagenome.org/</a>
<i>Candida glabrata</i>	<a href="http://natchaug.labri.u-bordeaux.fr/Genolevures/">natchaug.labri.u-bordeaux.fr/Genolevures/</a>
<i>Candida famata</i>	<a href="http://natchaug.labri.u-bordeaux.fr/Genolevures/">natchaug.labri.u-bordeaux.fr/Genolevures/</a>
<i>Cryptococcus neoformans</i>	<a href="http://www.tigr.org/tdb/e2k1/cna1/">www.tigr.org/tdb/e2k1/cna1/</a>
<i>Fusarium graminearum</i>	<a href="http://www.broad.mit.edu/annotation/genome/fusarium_graminearum/">www.broad.mit.edu/annotation/genome/fusarium_graminearum/</a>
<i>Kluveromyces lactis</i>	<a href="http://natchaug.labri.u-bordeaux.fr/Genolevures/">natchaug.labri.u-bordeaux.fr/Genolevures/</a>
<i>Magnaporthe grisea</i>	<a href="http://www.broad.mit.edu/annotation/fungi/magnaporthe/">www.broad.mit.edu/annotation/fungi/magnaporthe/</a>
<i>Neurospora crassa</i>	<a href="http://www.broad.mit.edu/annotation/fungi/neurospora/">www.broad.mit.edu/annotation/fungi/neurospora/</a>
<i>Saccharomyces cerevisiae</i>	<a href="http://yeastgenome.org/">yeastgenome.org/</a>
<i>Schizosaccharomyces pombe</i>	<a href="http://www.sanger.ac.uk/Projects/S_pombe/">www.sanger.ac.uk/Projects/S_pombe/</a>
<i>Stagonospora nodorum</i>	<a href="http://www.broad.mit.edu/annotation/genome/stagonospora_nodorum/">www.broad.mit.edu/annotation/genome/stagonospora_nodorum/</a>
<i>Trichoderma reesei</i>	<a href="http://genome.jgi-psf.org/Trire2/Trire2.home.html">genome.jgi-psf.org/Trire2/Trire2.home.html</a>
<i>Ustilago maydis</i>	<a href="http://www.broad.mit.edu/annotation/fungi/ustilago_maydis/">www.broad.mit.edu/annotation/fungi/ustilago_maydis/</a>
<i>Yarrowia lipolytica</i>	<a href="http://natchaug.labri.u-bordeaux.fr/Genolevures/">natchaug.labri.u-bordeaux.fr/Genolevures/</a>

**Tabelle 14: Verwendete Genomdaten.** Für die Analysen wurden von den angegebenen Internetseiten die aus den Sequenzierdaten vorhergesagten und translatierten Proteinsequenzen im Fasta-Format bezogen. Dies geschah im Zeitraum Januar 2004 bis März 2005. Daher sind in den Analysen weder neuere Daten dieser Organismen noch die Daten von später veröffentlichten Genomsequenzen berücksichtigt.

## 4 Ergebnisse

### 4.1 Bekannte Substrate Kex2-homologer Proteasen in Pilzen

In der Literatur ist nur für relativ wenige Proteine experimentell gezeigt worden, dass sie von Kex2 prozessiert werden. Dabei handelt es sich vor allem um Proteine mit multiplen Schnittstellen, etwa von Kreuzungspheromonen und Killertoxinen oder dem Rep1 Protein aus *Ustilago maydis*. Darüber hinaus sind einige wenige Enzyme bekannt, die Substrate von Kex2 sind. Dazu gehören die Exoglukanase Exg1 aus *S. cerevisiae*, die Xylanasen Xyn1 und Xyn2 aus *Trichoderma reesei* und eine Chloroperoxidase aus *Aspergillus niger*. Des Weiteren gibt es Strukturproteine der Zellwand aus *S. cerevisiae*, die Pir-Proteine, die zu den Substraten von Kex2 gehören.

Spezies	Protein	Pos.	Ref.	Accession#	Sequenz der Schnittstelle
<i>A. niger</i>	Clp Chloroperoxidase	322	[98]	P04963	Napyyaqdptlgpndkr eaapaattsmavf
<i>C. albicans</i>	MF $\alpha$	81	[10]	CA2412	livngtltstsgannekr eakskggfrltnf
<i>C. albicans</i>	MF $\alpha$	102	[10]	CA2412	kkgfrltnfgyfepgkr danadagfrltnf
<i>C. albicans</i>	MF $\alpha$	123	[10]	CA2412	dagfrltnfgyfepgkr danaeagfrltnf
<i>C. albicans</i>	MF $\alpha$	143	[10]	CA2412	dagfrltnfgyfepgkr danaeagfrltnf
<i>C. albicans</i>	Exoglukanase 1	38	[99]	P29717	-visnfpkpngnlkfkkr ggghnvawdydnn
<i>C. immitis</i>	Sowgp	61	[100]		satpshydeygykmrkr gatshkehsycdt
<i>K. lactis</i>	RF2 Killertoxin	29	[101]	S07915	fvqglehthrrgslvkr avcydtdqvpjni
<i>K. lactis</i>	RF2 Killertoxin	894	[101]	S07915	tslavkeynnekvlykr gdipppgsnnrli
<i>S. cerevisiae</i>	Cis3 (Pir4)	64	[102]	NP_012377	ttfgiavqaitsskkr dvisqigdgqvqa
<i>S. cerevisiae</i>	Pir2	72	[103]	S33651	tssassaattasskkr aasqigdgqvqaa
<i>S. cerevisiae</i>	K1 Killertoxin	149	[104]		nrpvsqsdpngatvarr distvadgdipld
<i>S. cerevisiae</i>	K1 Killertoxin	233	[104]		gtyttfttphipavakr yvypmcehgikas
<i>S. cerevisiae</i>	K1 Killertoxin	44	[104]		ndvagpaetapvsllpr eapwydkiewvkd
<i>S. cerevisiae</i>	K1 Killertoxin	188	[105]		ehgisilpanasqyvkr sdaehttsfvvt
<i>S. cerevisiae</i>	K2 Killertoxin	221	[106]	S12933	vvhlgvyydnstgivr dsrpsmtswtvlh
<i>S. cerevisiae</i>	K28 Killertoxin	149	[107]	S51548	ankvisktenaeniqr slipgllsmdyns
<i>S. cerevisiae</i>	K28 Killertoxin	49	[107]	S51548	yargmptserqqgleer dfsaatcvlmgae
<i>S. cerevisiae</i>	K28 Killertoxin	245	[107]	S51548	fgfaphsngnstglqkr asiddavwlqsay
<i>S. cerevisiae</i>	<i>Rhizopus</i> Lipase RNL	231	[108]	AAC60540	yyvestgipfqrtvhkr divphvppqsfgf
<i>S. cerevisiae</i>	<i>Rhizopus</i> Lipase RNL	95	[109]	P21811	tewyeshggnltsigkr ddnlvvgmtldlp
<i>S. cerevisiae</i>	MF $\alpha$	104	[110]	P01149	eawhlqlkpgqpmykr eaeaeawhlqlk
<i>S. cerevisiae</i>	MF $\alpha$	125	[110]	P01149	eawhlqlkpgqpmykr eadaeawhlqlk
<i>S. cerevisiae</i>	MF $\alpha$	145	[110]	P01149	eawhlqlkpgqpmykr eadaeawhlqlk
<i>S. cerevisiae</i>	MF $\alpha$	85	[110]	P01149	tiasiaakeegvsldkr eaeawhlqlkpg
<i>T. reesei</i>	Xylanase I	51	[42]	XYN1_TRIRE	rgmydfvlgahndhrrr asinydqnyqtgg
<i>T. reesei</i>	Xylanase II	32	[42]	XYN2_TRIRE	ascrpaaevesvavekr qtiqpgtgynngy
<i>U. maydis</i>	KP6 Killertoxin	26	[111]	A34778	-----lpnglspr nnafcagfglsck
<i>U. maydis</i>	KP6 Killertoxin	139	[111]	A34778	elhrddppptasdigkr gkrprpvmcqvcd
<i>U. maydis</i>	KP6 Killertoxin	107	[111]	A34778	hcsslennmscrslskr tiqdsatdtdvdlg
<i>U. maydis</i>	Rep1, hydrophobin	59	[112]	S71753	klvdasdltlgvdilkr tdysackkyvssy

Tabelle 15: Schnittstellen bekannter Substrate von Kex2-homologen Proteasen aus Pilzen.

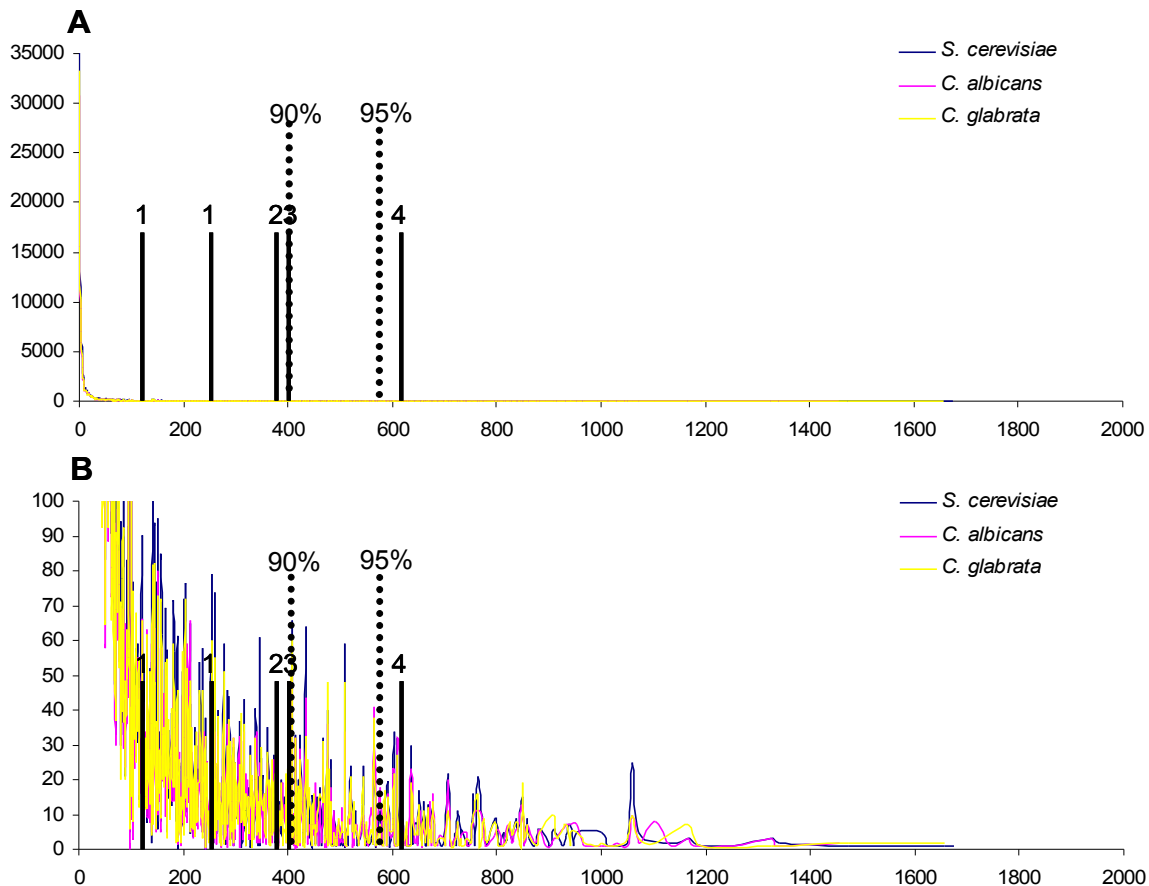
Da diese bekannten Substrate in ihrer Anzahl zu gering sind um eine statistisch verlässliche Beschreibung der Schnittstelle über die P1- und P2-Position hinaus zu generieren, wurde aus den in Tabelle 2 gezeigten Daten der biochemischen Aktivität von *S. cerevisiae* Kex2 gegenüber fluorogenen Peptiden eine positionsspezifische Matrix (PSSM) für die Positionen P4-P1 berechnet. Dazu wurden die gegebenen Werte (Spalte „relativ“, Tabelle 16) der Positionen P4 und P3 in Prozent vom Maximalwert (Arg an P4, His an P3) umgerechnet. Mit Hilfe dieser Matrix können nun Peptidsequenzen mit Arginin an P1 hinsichtlich ihrer biochemischen Suszeptibilität gegenüber Kex2 klassifiziert werden.

	P4		P3		P2	P1
	relativ	%Max	relativ	%Max	%Max	%Max
Lys	128	0,72	100	0,70	1,000	0,014
Arg	179	1,00	128	0,90	0,510	1,000
Thr	30	0,17	110	0,77	0,130	0,0088
Pro	36	0,20	0	0,01	0,095	0,00019
Glu	0	0,01	89	0,62	0,088	0,00
Ile	51	0,28	79	0,55	0,054	0,00
Ser	31	0,17	92	0,64	0,043	0,00
Ala	45	0,25	92	0,64	0,032	0,00
Asn	88	0,49	98	0,69	0,029	0,00
Val	100	0,56	130	0,91	0,014	0,00
Cys	9	0,05	101	0,71	0,0085	0,00
Asp	0	0,01	38	0,27	0,0027	0,00
Gln	74	0,41	102	0,71	0,0024	0,00
Gly	6	0,03	100	0,70	0,0023	0,00
His	131	0,73	143	1,00	0,00072	0,00
Met	95	0,53	101	0,71	0,00068	0,00
Leu	16	0,09	128	0,90	0,00043	0,00
Tyr	85	0,47	121	0,85	0,00017	0,00
Phe	108	0,60	141	0,99	0,0000068	0,00
Trp	58	0,32	126	0,88	0,0000009	0,00

**Tabelle 16: Matrix „biochemisch“.** Positionsspezifische Matrix für die Substratspezifität von Kex2 in *S. cerevisiae*, generiert aus den in Tabelle 2 gezeigten Daten.

Die Durchmusterung von Genomdaten macht relativ schnell klar, dass es eine Vielzahl solcher Motive gibt. Alleine das Motiv „KR“ kommt im Genom von *S. cerevisiae* 11499 mal, bei *C. albicans* 9639 mal und bei *C. glabrata* 9876 mal vor. Die Verteilung der durch die Matrix bewerteten Motive (Abbildung 13) zeigt, dass dieses aufgrund der verschiedenen Möglichkeiten an der Position P2 noch weit häufiger vorkommt (20475 mal bei *C. albicans*). Die Motive bekannter Substrate (Abbildung 13, 1-4) werden, wie erwartet, höher bewertet als der Durchschnitt. Tatsächlich liegen die bekannten Substrate von *S. cerevisiae* und *C. albicans*, mit Ausnahme des  $\alpha$ -Pheromons aus *C. albicans*, im Bereich der oberen 10%.

Die Tatsache, dass es in diesem Bereich etwa 3000 weitere Motive gibt, macht gleichzeitig deutlich, dass Motive, welche Kex2-Schnittstellen ähneln, nicht exklusiv in einer kleinen Gruppe von Substratproteinen vorkommen, sondern statistisch im Genom verteilt sind. Es müssen daher zur Vorhersage von Substratproteinen weitere Kriterien als nur die Bewertung der biochemischen Aktivität/Affinität an den Bindungsstellen S4-S1, herangezogen werden.



**Abbildung 13: Verteilung der durch die PSSM bewerteten Motive bei *C. albicans*.** A: Etwa 90% der 20475 durch die PSSM bewertbaren Motive liegen in einem Bereich unterhalb der Bewertung von bekannten Kex2-Schnittstellen (2,3,4). B: Selbe Daten wie (A) mit vergrößerter Y-Achse. 1: CaMF $\alpha$  (122, 264); 2: ScMF $\alpha$  (392); 3: ScPir2/ScPir4 (401), 4: CaExg1 (621).

## 4.2 SLoPF-DB: Datenbank für Subzelluläre Lokalisation von Proteinen aus Fungi

Um zwischen tatsächlich geschnittenen und unprozessiert bleibenden Motiven unterscheiden zu können, bieten sich zunächst zwei Parameter an:

- Erstens verhindert die Kompartimentierung der Zelle, dass Proteine unterschiedlicher subzellulärer Lokalisation direkt miteinander interagieren können, so dass Motive in einem Kompartiment ohne den entsprechenden Interaktionspartner ohne Funktion sein können.
- Zweitens kann der Grad der Konservierung eines Motivs zwischen orthologen Proteinen verschiedener Organismen Aufschluss über die funktionelle Relevanz eines Motivs geben.

Dazu sollte zunächst eine Methode bereitgestellt werden, die diese beiden Parameter vereint. Im Folgenden wird ein Algorithmus beschrieben, der für eine Gruppe zueinander orthologer Proteine aus verschiedenen Pilzen aus den einzelnen vorhergesagten Lokalisationen eine Wahrscheinlichkeit für die subzelluläre Lokalisation berechnet. Die so erzeugten Daten wurden in eine interaktive „Datenbank für Subzelluläre Lokalisation von Proteinen aus Fungi“ (SLoPF) integriert, die es ermöglicht, Proteinmotive, sortiert nach subzellulären Kompartimenten, auf Grad der Konservierung und lokale Struktur motive hin zu untersuchen.

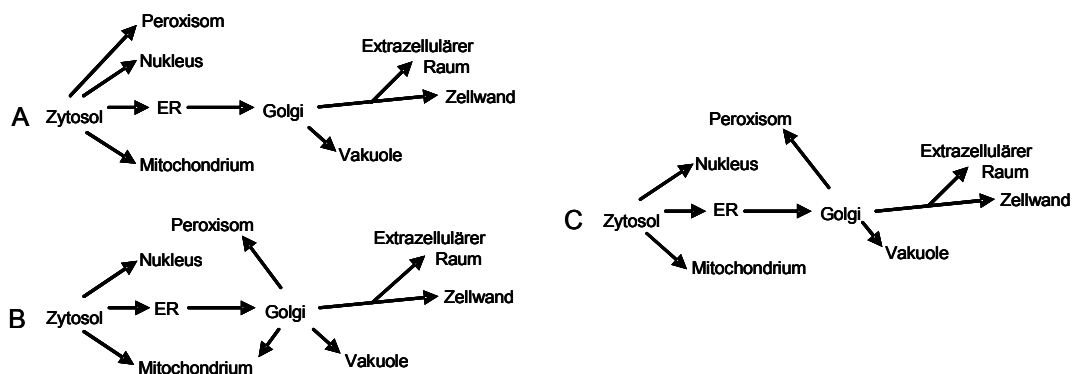
### 4.2.1 Vorhersage der Lokalisation von Proteinen

In biologischen Systemen gibt es aufgrund der Kompartimentierung von Zellen eine Vielzahl möglicher, von einander distinkter Lokalisationen von Proteinen. Bereits co-translational während der Proteinbiosynthese entscheidet sich für viele Proteine durch Bindung des „signal recognition particle“ (SRP) an das Signalpeptid oder eine Signal-Anker-Sequenz, ob das Protein in eines der Subkompartimente des sekretorischen Wegs eintritt oder ob es im Zytosol verbleibt. Über diesen Mechanismus werden auch integrale Membranproteine, mit Ausnahme derer, die über einen sog. „C-tail-anchor“ in der Membran verankert werden, in der Zelle verteilt. Darüber hinaus enthalten Proteine weitere Lokalisationssequenzen, die post-translational zur Translokation in andere Kompartimente der Zelle führen können, beispielsweise für den Zellkern (Kernlokalisationssequenz, „NLS“), Peroxisomen („peroxisomal targeting sequence“ (PTS)) oder Mitochondrien („mitochondrial targeting peptide“ (mTP)). Des Weiteren werden viele Proteine an einem bestimmten Ort der Zelle durch Protein-Protein Interaktion gehalten, z. B. ER-residente Proteine durch Bindung durch den HDEL-Rezeptor.

Da die bekannten Lokalisationssignale jedoch in der Regel keine einfache Konsensussequenz aufweisen, ist eine genaue Klassifizierung anhand dieser nicht immer möglich. Trotzdem besitzt diese Art der Lokalisationsvorhersage eine sehr hohe positive Vorhersagequalität, d. h. Proteine

in denen die entsprechenden Motive gefunden werden, können mit hoher Wahrscheinlichkeit zugeordnet werden. Allerdings ist das Fehlen eines solchen Motivs kein klarer Beweis für die Lokalisation an anderer Stelle, so z. B. bei Proteinen, welche nicht über das klassische Signalpeptid in das ER gelangen [113]. Daher wird an dieser Stelle eine Kombination aus verschiedenen Methoden verwendet, deren Zusammenspiel zu einer erhöhten Qualität der Aussage führen sollte. Die Ergebnisse dieser Methoden werden wiederum mit der in Abschnitt 4.2.2 vorgestellten Gruppierung von Sequenzen gekoppelt, so dass eine Lokalisationsvorhersage im günstigsten Fall auf mehreren orthologen Proteinsequenzen beruht. Ausnahmen bilden hier abgeleitete Genprodukte, die spezifisch für eine der betrachteten Spezies sind („orphans“) oder Genprodukte, deren Sequenzhomologien nicht ausreichen, um diese mit der verwendeten Methode zu finden.

Das Sortierschema, welchem SLoPF-DB folgt, ähnelt im Wesentlichen einem angenäherten Zwei-Schritt Sortierungsprozess, wie er auch in der Zelle funktioniert. Es wird zunächst eine grobe Klassifizierung in „nicht-sekretorisch“/„sekretorischer Weg“/„mitochondrial“ vorgenommen, wobei in die Klasse „sekretorischer Weg“ auch alle nicht-mitochondrialen Transmembranproteine und Proteine der Mikrokörperkompartimente fallen. Die Klassen „nicht-sekretorisch“ und „sekretorischer Weg“ werden dann weiter aufgeteilt auf „zytosolisch“/„nukleär“ bzw. „ER“/„Golgi“/„Vakuole“/„Peroxisom“/„Extrazellulärraum“. Eine weitergehende Unterteilung des mitochondrialen Proteoms wurde nicht vorgenommen, weil dies im Rahmen dieser Arbeit nicht von Bedeutung war.



**Abbildung 14: Subzelluläre Lokalisationsvorhersage mit SLoPF-DB.** SLoPF-DB imitiert den Sortierprozess der Zelle, vereinfacht aber die Sortierung von Membranproteinen der Mikrokörper und des Mitochondriums. Sortierung von (A) löslichen und (B) Membranproteinen innerhalb der Zelle und (C) Sortierung beider Proteinklassen gleichzeitig in SLoPF-DB.

#### 4.2.1.1 Erstellung eines Validierungs- und Testsets

Die hier durchgeführten Vorhersagen von subzellulären Lokalisationen wurden mittels Proteinen bekannter Lokalisation von *S. cerevisiae* validiert. Diese wurden aus der GO-Annotationsdatenbank<sup>4</sup> für das *S. cerevisiae* Genom entnommen, welche die aus der Literatur entnommene Beschreibung der einzelnen Proteine mit den entsprechenden Verweisen („GO-terms“) enthält [114]. Alle Einträge, die selbst auf elektronischer Annotation beruhen (Evidence Codes „IEA: inferred from electronic annotation“ und „RCA: inferred from Reviewed Computational Analysis“) oder die für Gene, welche keine Proteine codieren (tRNAs, etc.) wurden vorher aus der Datenbank entfernt. Außerdem wurden die Daten der Arbeit von Sickmann *et al.* [115] über das mitochondriale Proteom vernachlässigt, da die Annotationen jener Arbeit nicht nach luminalen und peripher assoziierten (d. h. zytosolischen) Proteinen unterscheiden.

Verschiedene GO-terms sind bezüglich der Lage der Proteine nicht eindeutig. Beispielsweise wird der GO-term „vesicle coat“ auf die Lokalisation „ER-membrane“ zurückgeführt. Dies ist zwar richtig, allerdings sind die entsprechenden, vesikelumgebenden COP-Proteine auf der luminalen Seite des Vesikels lokalisiert und daher für diesen Algorithmus als „zytosolisch“ einzustufen. Die so vorgenommenen Ausnahmen sind in Tabelle 17 aufgeführt. Des Weiteren wurden nur solche Einträge berücksichtigt, denen genau eine Lokalisation zugeordnet wurde. Insgesamt enthält das auf diese Weise erzeugte Validierungsset eindeutige Lokalisierungen für 4940 Proteine aus *S. cerevisiae*.

GO-term	beschreibt	wird zu	weil
GO:0030479	actin cortical patch	GO:0005829	ist assoziiert mit Plasmamembran
GO:0030008	TRAPP complex	GO:0005829	ist assoziiert mit Golgi-Apparat
GO:0030897	HOPS complex	GO:0005829	ist assoziiert mit Vakuolenmembran
GO:0030120	vesicle coat	GO:0005829	COPI, COPII and Clathrin-coat
GO:0005885	ARP3 complex	GO:0005829	ist assoziiert mit Plasmamembran
GO:0017119	golgi transport complex	GO:0005829	ist assoziiert mit Golgi-Apparat
GO:0045009	chitosome	GO:0005829	ist assoziiert mit Plasmamembran
GO:0000221	vacuolar ATPase complex	GO:0005829	auf zytosolischer Seite der Membran
GO:0000938	GARP complex	GO:0005829	auf zytosolischer Seite der Membran
GO:0000306	extrinsic to vacuolar membrane	GO:0005829	auf zytosolischer Seite der Membran
GO:0042406	extrinsic to ER membrane	GO:0005829	auf zytosolischer Seite der Membran
GO:0005780	extrinsic to peroxisomal membrane	GO:0005829	auf zytosolischer Seite der Membran
GO:0031234	plasmamembrane, cytosolic side	GO:0005829	auf zytosolischer Seite der Membran
GO:0031232	plasmamembrane, extracellular side	GO:0005615	auf extrazellulärer Seite der Membran
GO:0005680	anaphase promoting complex	löschen	zu diesem Zeitpunkt des Zellzyklus existiert kein Zellkern
GO:0005935	bud site	löschen	beschreibt sowohl intra- als auch extrazelluläre Proteine
GO:0000943	transposable elements	löschen	keine spezifische Lokalisation
GO:0005619	spore wall	löschen	im Zytosol lokalisierte Zellwandproteine

**Tabelle 17: Ausnahmen im Validierungsset für die Lokalisationsvorhersage.** Nicht alle GO-Termini sind korrekt an die Kompartimentierung der Zellen angelehnt. Das betrifft insbesondere im zytosolisch lokalisierten Vesikelmantel und an der zytosolischen Seite von ER- und Golgi-Membran lokalisierte Proteinkomplexe. Die Bedeutungen dieser GO-Termini wurden entweder manuell geändert oder bei Mehrfachbedeutungen gelöscht.

<sup>4</sup> [ftp://genome-ftp.stanford.edu/pub/yeast/literature\\_curation/gene\\_association.sgd.gz](ftp://genome-ftp.stanford.edu/pub/yeast/literature_curation/gene_association.sgd.gz)

#### 4.2.1.2 Zuordnung aufgrund von Lokalisationssequenzen

Die am häufigsten verwendete Methode um Lokalisationen vorherzusagen, ist die Suche nach Peptidsequenzen, die von der Zelle benutzt werden, um Proteine in bestimmte Kompartimente zu transportieren oder sie dort festzuhalten, so genannten „Lokalisationssequenzen“. Die in dieser Arbeit benutzten Motive sind in Tabelle 18 zusammengestellt.

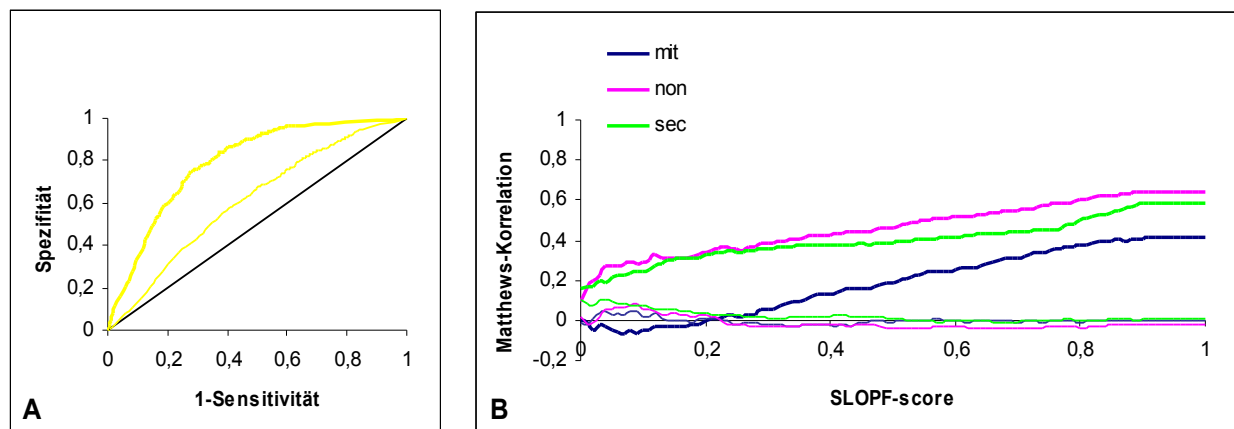
Lokalisation	Name des Motivs	Motiv <sup>5</sup> /Programm	Referenz
Sekretorischer Weg	Signalpeptid	SignalP 2.0	[116]
Mitochondrium	mitochondriales targeting-Peptid (mTP)	TargetP 1.0	[92]
Peroxisom, Lumen	PTS1, peroxisomal targeting sequence 1	[SA][KR][LM]\$	[92]
Peroxisom, Lumen	PTS2, peroxisomal targeting sequence 2	^R[LI]{0..5}HL	[92]
ER, Lumen	ER-Retentionsmotiv für lösliche Proteine	[HDY]DEL\$	
Zellkern	Kernlokalisationssequenz (NLS)	predictNLS 1.0	[117]
ER, Membran	ER-Retentionsmotiv für integrale Membranproteine	KK..\$	[118]
Sekretorischer Weg/ Membranfraktion	Kombination von Signalpeptid und Transmembrandomänen	Phobius 1.0	[119]
Plasmamembran, zytosolische Seite	CaaX-Box	C[DENQ][LIVM].\$	Prosite PS00294
Zellwand oder Plasmamembran	Glykosylphosphatidylinositol-Anker (GPI)	[NSGDAC][GASVIETKDLF] [GASV]{4,19} [FILMVAGPSTCYWN]{10}. {1,3}\$	Modifiziert nach [120]
Zellwand oder Plasmamembran	Glykosylphosphatidylinositol-Anker (GPI)	DGPI 1.0	[121]
Membranfraktion	C-tail-Anker	TMHMM	[122]

**Tabelle 18: Zur Lokalisationsvorhersage verwendete Motive.**

Bei Anwendung dieser Methode auf das Testset zeigte sich, dass die Lokalisationsvorhersage aufgrund von Lokalisationssequenzen eine sehr hohe Effizienz bezüglich der Erkennung von Proteinen des sekretorischen Weges aufweist (Abbildung 15). Für mitochondriale Proteine ist diese Methode weniger spezifisch, weil viele der Proteine aufgrund eines von TargetP nicht erkannten mitochondrialen Signalpeptids als zytoplasmatisch eingestuft werden (nicht gezeigt). Auch verfügen die Proteine, welche im mitochondrialen Genom kodiert sind, nicht über erkennbare Lokalisationssequenzen und müssten hier als Sonderfall behandelt werden.

<sup>5</sup> Die Notierung von Sequenzmotiven erfolgt hier nach den Regeln für „regular expressions“ der Programmiersprache Perl: „^“: Sequenzanfang, „\$“: Sequenzende, „[xx]“: X oder Y, „.“: beliebige Aminosäure, „{x..y}“: vorhergehende Aminosäure mindestens x-, maximal y-mal.





**Abbildung 15: Messung der Qualität der Lokalisationsvorhersage mittels Lokalisationssequenzen.** A: Die Klassifikation des Validierungssets mittels Lokalisationssequenzen erfolgt mit einer hohen Gesamttrefferquote (fette Linie) im Vergleich zu einer Zufallsverteilung (feine Linie). B: Die Trefferwahrscheinlichkeiten der einzelnen Kompartimente (fette Linien) sind dabei für die Lokalisationen „sekretorisch“ (grün) und „nicht-sekretorisch“ (pink) wesentlich höher als für „mitochondrial“ (blau).

#### 4.2.1.3 Zuordnung aufgrund der Aminosäurezusammensetzung

Aufgrund der verschiedenen physikochemischen Bedingungen wie beispielsweise dem pH-Wert oder der Ionenstärke in den verschiedenen Kompartimenten der Zelle unterscheiden sich Proteine, die dort lokalisiert sind, in der Aminosäurezusammensetzung der Oberfläche [123]. Durch statistische Analyse dieser Unterschiede in der Aminosäurezusammensetzung des Gesamtproteins lässt sich die subzelluläre Lokalisation vorhersagen [124, 125]. Für den hier verfolgten Ansatz waren die in der Literatur beschriebenen Verfahren allerdings nicht geeignet, da sie zwar Zuordnungen zu den einzelnen Lokalisationen zulassen (Ja/Nein-Aussagen), diese aber nicht mit Werten für die Wahrscheinlichkeit bzw. Qualität der Aussagen versehen. Ein Programm, das erstmals eine solche Wahrscheinlichkeit der Zugehörigkeit zu einer Klasse berechnet und welches auch hier verwendet wurde, ist libsvm-prob [90].

Die Trainingsdaten für den hier erzeugten Klassifikator wurden aus der Datenbank DBSubLoc<sup>6</sup> Version 1.0 [126], erzeugt. Dazu wurden die einzelnen Einträge nach ihren Klassifizierungen mit Geneontology<sup>7</sup>-Kategorien („GO“) sortiert, wobei integrale Membranproteine aus dem Datensatz eliminiert wurden, da sie aus Domänen unterschiedlicher Lokalisation aufgebaut sind und sich somit nicht eindeutig zuordnen lassen. Die schematische Reihenfolge der Sortierung für Membranproteine wurde bereits beschrieben (4.2.1). Die Anzahl der in den Trainingsklassen enthaltenen Einzelsequenzen sind in Tabelle 19 aufgeführt.

Ein Hauptproblem bei der Erzeugung von Modellen mit Supportvektoren stellen hier die unterschiedlich großen Datenmengen für die einzelnen Lokalisationen dar, was bei der

<sup>6</sup> [www.bioinfo.tsinghua.edu.cn/~guotao/data/dbsubloc2.fungi.dat.gz](http://www.bioinfo.tsinghua.edu.cn/~guotao/data/dbsubloc2.fungi.dat.gz)

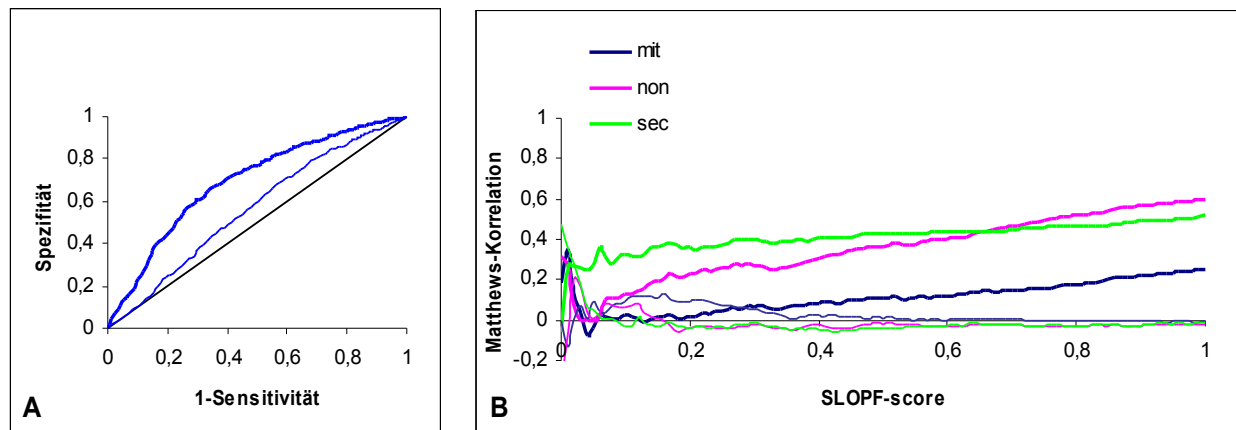
<sup>7</sup> [www.geneontology.org](http://www.geneontology.org)

Kreuzvalidierung zu einer überproportionalen Gewichtung der größeren Klasse führen kann. Daher wurden drei verschiedene Modelle erzeugt („Primär“, „Intra“ und „Extra“), bei denen Anzahl der Einzelsequenzen in den einzelnen Klassen größtmöglich gleich ist (Tabelle 19). Die drei Modelle entsprechen dabei, wie oben dargestellt, in etwa dem biologischen Sortierungsweg der Zelle, in dem zunächst eine Unterscheidung von Proteinen des sekretorischen Weges, der Mitochondrien und dem Rest vorgenommen wird. Diese werden anschließend weiter unterteilt, entweder in die Klassen „Peroxisom“, „Vakuole“, „extrazellulär“, „Golgi-Apparat“ oder „Endoplasmatisches Retikulum“ für Proteine des sekretorischen Weges, und „zytosolisch“ oder „nukleär“ für Proteine die weder das Mitochondrion noch den sekretorischen Weg nehmen. Obwohl lösliche peroxisomale Proteine nicht über den sekretorischen Weg sortiert werden, sondern direkt aus dem Zytosol in das Peroxisom gelangen [127, 128], wurden sie hier aus technischen Gründen in die Gruppe der sekretorischen Proteine integriert. Dieses Vorgehen hat den Vorteil, dass die relativ kleine Gruppe der peroxisomalen Proteine nicht gegenüber den im Verhältnis dazu sehr großen Gruppen der zytosolischen und nukleären Proteine statistisch unberücksichtigt bleibt. Da außerdem peroxisomale Membranproteine zumeist über den sekretorischen Weg in das Peroxisom gelangen [127, 128], ermöglicht dies eine gemeinsame Betrachtung von löslichen Proteinen und integralen Membranproteinen.

Lokalisation	GO-term	Anzahl Sequenzen	Genauigkeit des Klassifikators		
			primär	Sekundär	
Zytosol	GO:0005829	395	77,6%	Int	94,4%
Nukleus	GO:0005634	1109		ra	
Mitochondrium	GO:0005739	331		(Keine weitere Klassifikation mitochondrialer Proteine)	
ER	GO:0005783	27		Ex	76,3%
Golgi	GO:0005794	9		tra	
Vakuole	GO:0005773	8			
Extrazellularraum	GO:0005615	92			
Peroxisom	GO:0005777	18			

**Tabelle 19: Supportvektor-Klassifikatoren zur Vorhersage der subzellulären Lokalisation.** Die Genauigkeit bei der fünffachen Kreuzvalidierung ergibt Klassifikationsgenauigkeiten um 77% für die Klassifikatoren „Primär“ und „Extra“, und sogar annähernd 95% für den Klassifikator „Intra“.

Für das Training wurde hier eine radiale Kernelfunktion mit den Parametern  $C=500$  und  $\gamma=8$  verwendet. Die Ergebnisse der fünffachen Kreuzvalidierung für die drei verschiedenen Klassifikatoren sind in Tabelle 19 dargestellt. Den größten Anteil an Fehlklassifikation der primären Ebene stellt die Verwechslung von mitochondrialen und nukleären Proteinen dar, was auf die Redundanz des transkriptionellen Apparates zurückzuführen ist (nicht gezeigt).

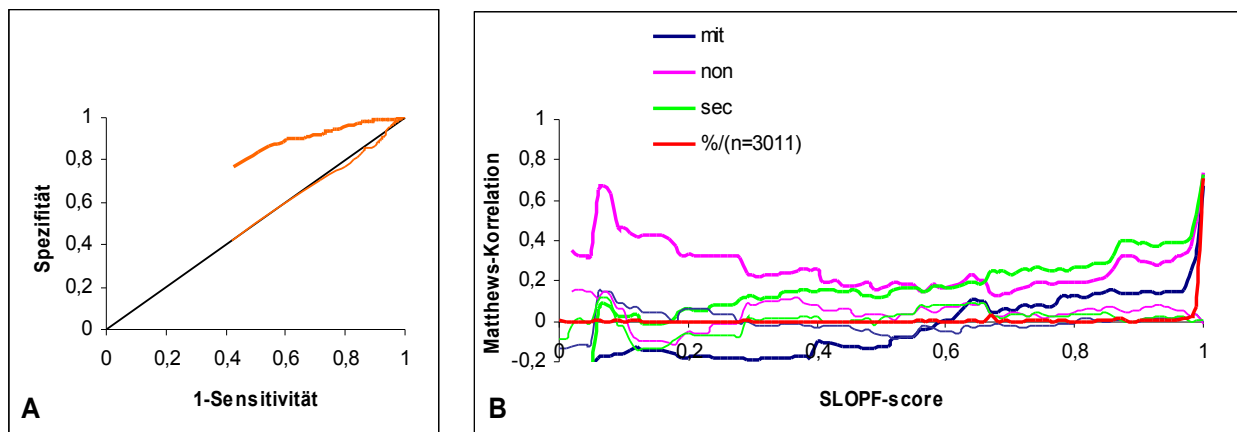


**Abbildung 16: Messung der Qualität der Vorhersage mittels Aminosäurezusammensetzung.** A: Die Klassifikation des Validierungssets mittels Lokalisationssequenzen in die drei Kategorien „sekretorisch“, „nicht-sekretorisch“ und „mitochondrial“ erfolgt im Vergleich zu einer Zufallsverteilung (feine Linie) mit einer niedrigeren Gesamttrefferquote (fette Linie) als die Vorhersagen mittels Lokalisationssequenzen. B: Wiederum sind die Trefferwahrscheinlichkeiten der einzelnen Kompartimente (fette Linien) für die Lokalisationen „sekretorisch“ (grün) und „nicht-sekretorisch“ (pink) wesentlich höher als für „mitochondrial“ (blau).

#### 4.2.1.4 Zuordnung aufgrund von Homologien zu Proteinen bekannter Lokalisation

Eine dritte Möglichkeit die Lokalisation von Proteinen vorherzusagen, besteht darin, die in der Literatur vorhandenen Annotationen dieses und/oder sequenzhomologer Proteine zu analysieren. Ein Webservice, der dieses ermöglicht ist der „Proteome Analyst“ [129]<sup>8</sup>. Dort werden Wahrscheinlichkeiten für die Lokalisationskategorien errechnet, indem für in den Annotationen der Proteine gefundene, vorher festgelegte, Vokabeln Punkte vergeben werden und daraus eine Wahrscheinlichkeit für eine bestimmte Lokalisation errechnet wird. In SLoPF-DB ergibt sich die Wahrscheinlichkeit für „sekretorischer Weg“ aus der Summe der Wahrscheinlichkeitswerte des Proteome Analyst für die Teilkompartimente. Analog die Wahrscheinlichkeit für die Fraktion der Proteine, welche weder in den sekretorischen Weg noch in das Mitochondrium transportiert werden (Summe von „zytoplasmatisch“ und „nukleär“). Problematisch bei dieser Methode ist, dass sie nicht für alle Proteine Vorhersagen liefern kann. Liegen keine Annotationen für dieses Protein (oder für solche mit genügend hoher Sequenzähnlichkeit) vor, so gibt das Programm keine Werte zurück. Somit sind aus dem Validierungsset nur 3001 Proteine (60,5%) mit dieser Methode klassifizierbar, dafür allerdings mit einem sehr hohen positiv prädiktiven Wert.

<sup>8</sup> [www.cs.ualberta.ca/~bioinfo/PA/Sub/](http://www.cs.ualberta.ca/~bioinfo/PA/Sub/)



**Abbildung 17: Messung der Qualität der Vorhersage mittels Proteome Analyst.** A: Der Proteome Analyst weist zu 98% nur Werte oberhalb von 0,99 zu. Daher lässt sich keine vollständige ROC-Kurve erstellen. Die Effizienz liegt ähnlich hoch wie die der Vorhersage aufgrund von Lokalisationssequenzen. B: Die vom Proteome Analyst vorhergesagten Werte unterhalb einer Wahrscheinlichkeit von 95% sind insignifikant, allerdings sind bei diesem Wert bereits mehr als 99% der Sequenzen klassifiziert. Wie bei der Vorhersage mittels Lokalisationssequenzen oder Aminosäurezusammensetzung, ist auch hier die Klassifikation von mitochondrialen Proteinen (blau) am schwierigsten. Dicke Linien: Klassifikator, dünne Linien: Werte einer Zufallsverteilung.

#### 4.2.2 Gruppierung orthologer Proteine

Ein Teil der Problematik bei statistischen Verfahren, wie sie hier angewandt werden, sind Ausreißersequenzen, die leicht vom Idealmodell abweichen und auf diese Weise „gerade eben“ falsch klassifiziert werden. Fehler durch solche Ausreißer können durch Mittelwertbildung vermindert werden, wie es beispielsweise bei physikalischen Messverfahren durch mehrere Messwiederholungen üblich ist. Da eine schlichte „Messwiederholung“ der Lokalisationsvorhersage hier wieder zu einem identischen Ergebnis führt, ist dies nicht praktikabel. Allerdings besteht die Möglichkeit, dass orthologe Sequenzen eines Ausreißers richtig klassifiziert werden und daher als „quasi“-Messwiederholung herangezogen werden könnten.

Zu diesem Zweck wurden die proteinkodierenden Sequenzen der 17 zum Zeitpunkt dieser Arbeit vollständig sequenzierten und veröffentlichten Pilzgenome mittels des Inparanoid Algorithmus [130] in Orthologgruppen aufgeteilt (Tabelle 20) und untersucht, ob die Bildung von Mittelwerten der Lokalisationswahrscheinlichkeiten innerhalb einer Gruppe einen positiven Einfluss auf die Lokalisationsvorhersage hat (Abbildung 18).

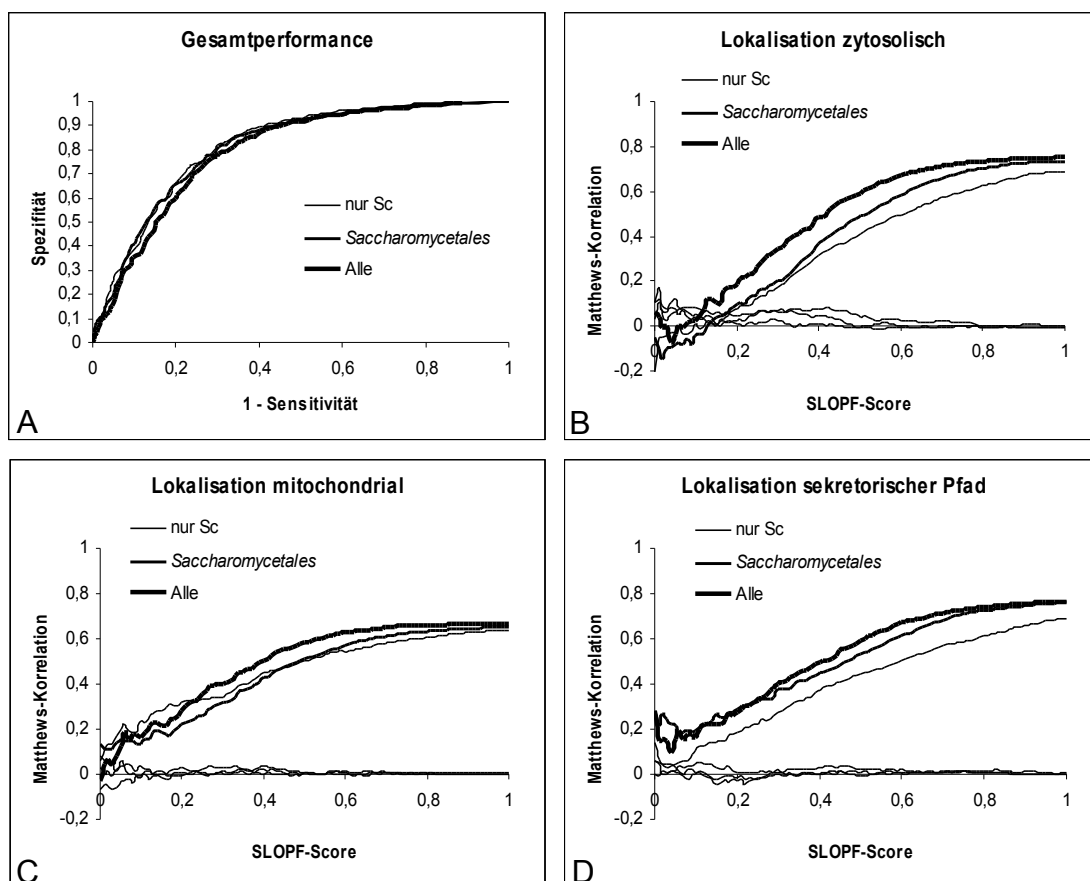
	1 <i>S. cerevisiae</i>	2 <i>Saccharomycetales</i>	3 Alle
unterer cutoff [bits]	-	50	50
oberer cutoff [bits]	-	350	500
min. Sequenzüberlappung	-	30%	30%
max. Clustergröße	-	28	42
Sequenzen in Datenbank	6703	41672	138596
Orthologgruppen in Datenbank	6703	15342	53021
Anteil Sequenzen in Orthologgruppen	0%	36%	38%
<b>Spezies</b>			
<i>Ashbya gossipii</i>		X	X
<i>Aspergillus fumigatus</i>			X
<i>Aspergillus nidulans</i>			X
<i>Candida albicans</i>		X	X
<i>Candida famata (Debaryomyces hansenii)</i>		X	X
<i>Candida glabrata</i>		X	X
<i>Cryptococcus neoformans</i>			X
<i>Fusarium graminearum</i>			X
<i>Kluyveromyces lactis</i>		X	X
<i>Magnaporthe grisea</i>			X
<i>Neurospora crassa</i>			X
<i>Saccharomyces cerevisiae</i>	X	X	X
<i>Schizosaccharomyces pombe</i>			X
<i>Stagonospora nodorum</i>			X
<i>Trichoderma reesei</i>			X
<i>Ustilago maydis</i>			X
<i>Yarrowia lipolytica</i>		X	X

**Tabelle 20: Übersicht orthologe Gruppierungen.** Die drei verschiedenen Genomgruppierungen haben in etwa gleiche Anteile an Sequenzen in Orthologgruppen. Er ist in Gruppe 3 (enthält alle 17 Genome) höher als in Gruppe 2 (enthält nur Genome der *Saccharomycetales*-Gruppe).

Der Inparanoid-Algorithmus ist BLAST-basiert und findet zueinander orthologe Sequenzen zwischen zwei Genomen, aber nicht paraloge Proteine innerhalb nur eines Genoms. Auch lassen sich mit ihm nur zwei Genome vergleichen, nicht aber mehrere. Daher wurden alle Genome paarweise verglichen und im Anschluss daran die Treffer gruppiert. Variable Parameter sind hier die minimale Länge der homologen Region („% overlap“), welcher optimal als 30% gefunden wurden (Daten nicht gezeigt), und der als „bits“ angegebene, minimale BLAST-score (unterer cutoff), welcher optimal bei 50 bits lag (Daten nicht gezeigt). Die jeweils zueinander orthologen Sequenzen wurden als bester bi-direktionaler Treffer oberhalb des minimalen BLAST-scores definiert, bei weiteren Treffern handelte es sich um zueinander paraloge. Diese wurden gemeinsam gruppiert. Um eine zu große Größe der Gruppen zu vermeiden, wurde als optimale maximale Gruppengröße eine Anzahl von  $4 \times n_{\text{Genome}}$  festgelegt (Daten nicht gezeigt). Wird sie überschritten, so wird der minimale BLAST-score (unterer cutoff) für die Gruppenzugehörigkeit so lange angehoben, bis die maximale Zahl an Sequenzen nicht mehr überschritten wird. Dies

passiert allerdings nur bis zu einer Obergrenze (oberer cutoff), oberhalb derer alle Treffer als signifikant angesehen werden.

Die Performance der kombinierten Lokalisationsvorhersage aus Lokalisationssequenzen, Aminosäurezusammensetzung und Textanalysen (siehe vorhergehende Abschnitte) wurde unter Verwendung der verschiedenen Genomgruppierungen gemessen. Die Messung mittels ROC-Analyse zeigt zunächst keinen signifikanten Anstieg der Klassifikationsgenauigkeit (Abbildung 18A). Die einzelnen Messungen der Performance für die drei verschiedenen Lokalisationen mittels Matthews-Korrelation (Abbildung 18B, C und D) zeigen hingegen, dass die Effizienz proportional zur Anzahl der verwendeten Genome steigt. Wiederum ist der Effekt bei der Klassifikation in die Klassen „sekretorisch“ und „nicht-sekretorisch“ am stärksten, während er für die Gruppe „mitochondrial“ nur moderat ansteigt.



**Abbildung 18: Einfluss der Anzahl von betrachteten Proteomen auf die Vorhersagequalität.** A: Die Messung der Gesamtpformance mittels ROC-Analyse zeigt keinen signifikanten Anstieg der Vorhersagequalität. B, C, D: Dagegen zeigt eine Analyse mittels Matthews-Korrelation eine zur Anzahl der verwendeten Genome proportionale Steigerung für alle drei Unterkompartimente. Untere Linien in A, B und C: Werte einer Zufallsverteilung. Sc: *S. cerevisiae* Datensatz.

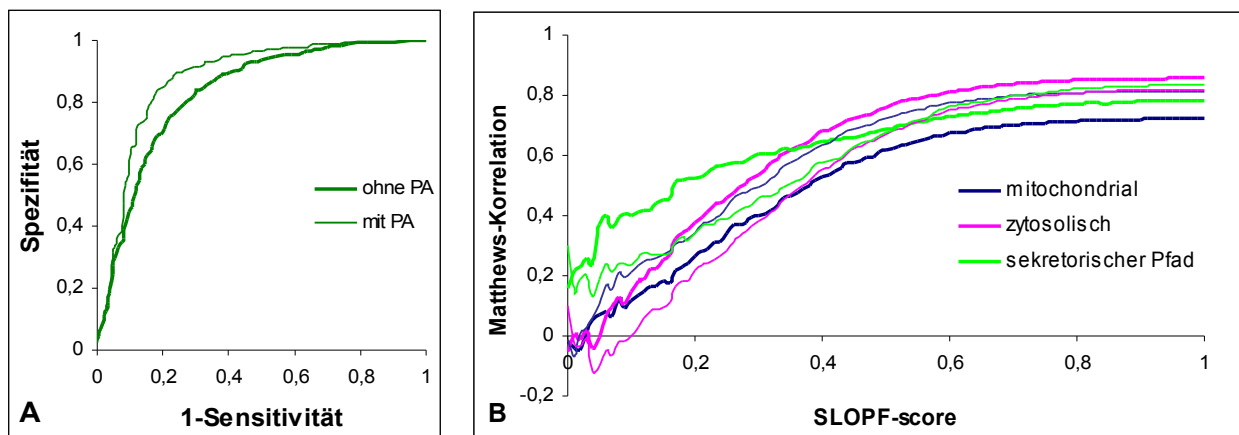
### 4.2.3 Validierung der Lokalisationsvorhersage in SLoPF-DB

Im Rahmen der Erstellung des Validierungssets wurden verschiedene Sequenzen aus der Klassifikation manuell herausgefiltert (Abschnitt 4.2.1.1), da sie beispielsweise als „ER“ eingestuft waren weil sie peripher mit dem ER assoziiert waren, sich somit aber nicht im Lumen des Kompartiments befanden. Da sich die Lokalisationsvorhersage des Proteome Analyst (Abschnitt 4.2.1.4) auf die gleichen Grunddaten wie die GO-Annotation in SGD stützt und somit die gleichen Fehler auftreten konnten, musste überprüft werden wie sich die Verwendung des Proteome Analyst auf die Vorhersage auswirkt.

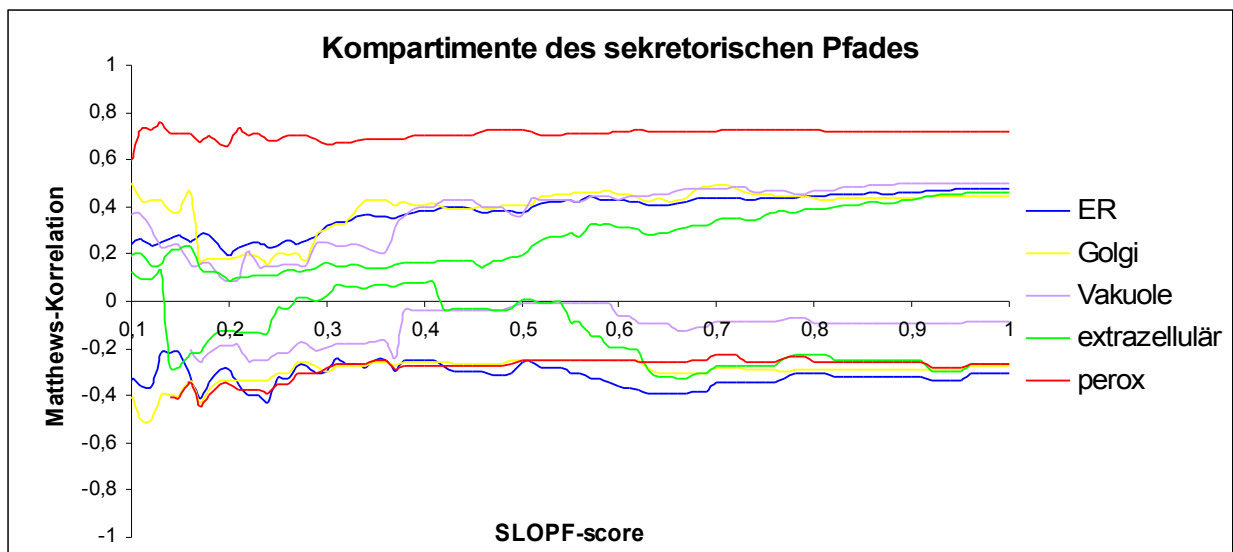
Tatsächlich konnte beobachtet werden, dass bei Verwendung der Proteome Analyst Daten die Gesamtperformance signifikant ansteigt (Abbildung 19A), sich aber spezifisch die Vorhersage von sekretorischen Proteinen verschlechtert (Abbildung 19B). Dies war vor allem auf eine wesentlich höhere Falsch-Positiv-Rate durch eben jene peripher assoziierten Protein zurückzuführen (Daten nicht gezeigt).

Weiterhin war von Interesse, wie die Vorhersage bezüglich der Unterkompartimente des sekretorischen Weges gelang, da Kex2 erst im Golgi Apparat aktiv ist und somit ER-residente Proteine in der Theorie nicht das primäre Ziel darstellen.

Eine Analyse des Validierungssets ohne Verwendung des Proteome Analyst mit der Orthologengruppe 3 zeigte eine nur moderate Vorhersagequalität für die Unterkompartimente (Abbildung 20). Im Folgenden wurden daher diese Parameter verwendet, aber die Unterlokalisierung im sekretorischen Weg nicht für die Vorhersage berücksichtigt.



**Abbildung 19: Analyse der Gesamtperformance von SLOPF-DB mit und ohne Proteome Analyst.** A: Die Gesamtperformance der Lokalisationsvorhersage in die drei Gruppen „sekretorischer Pfad“, „zytosolisch“ und „mitochondrial“ ist mit Berücksichtigung der Textanalyse (feine Linie) durch Proteome Analyst (PA) höher als ohne (fette Linie). B: Die Klassifikation von mitochondrialen Sequenzen (blau) profitiert stark von PA, während die Klassifikation sekretorischer Sequenzen (grün) darunter leidet.

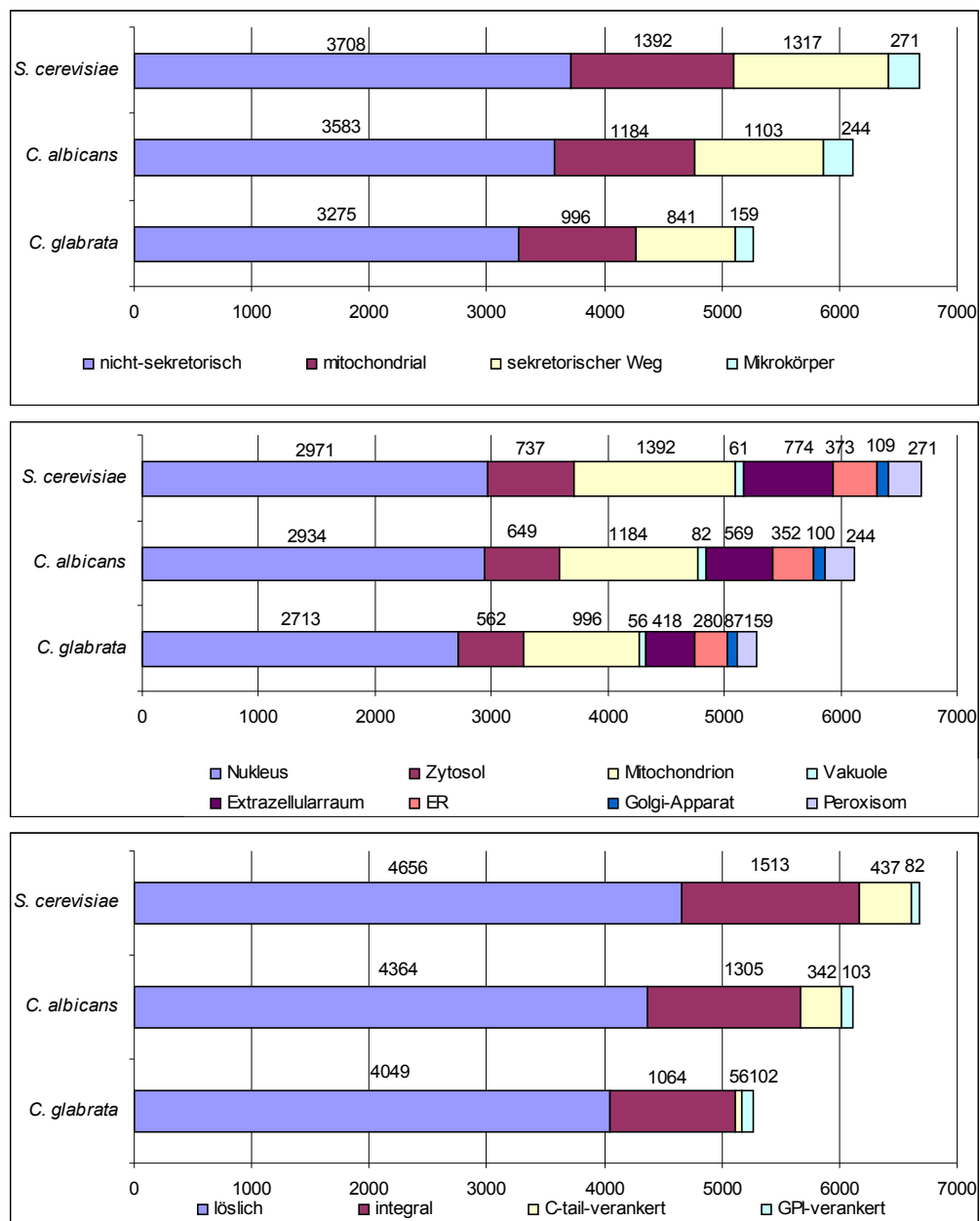


**Abbildung 20: Analyse der Performance von SLOPF-DB bezüglich der Unterkompartimente des sekretorischen Weges.** Für die Analyse wurde die Orthologengruppe 3 verwendet und die Daten des Proteome Analyst vernachlässigt. Die Klassifikationsrate innerhalb der Subkompartimente, mit Ausnahme der Mikrokörper (rot), ist nur moderat. Obere Linien: Klassifikator, untere Linien: Werte einer Zufallsverteilung.



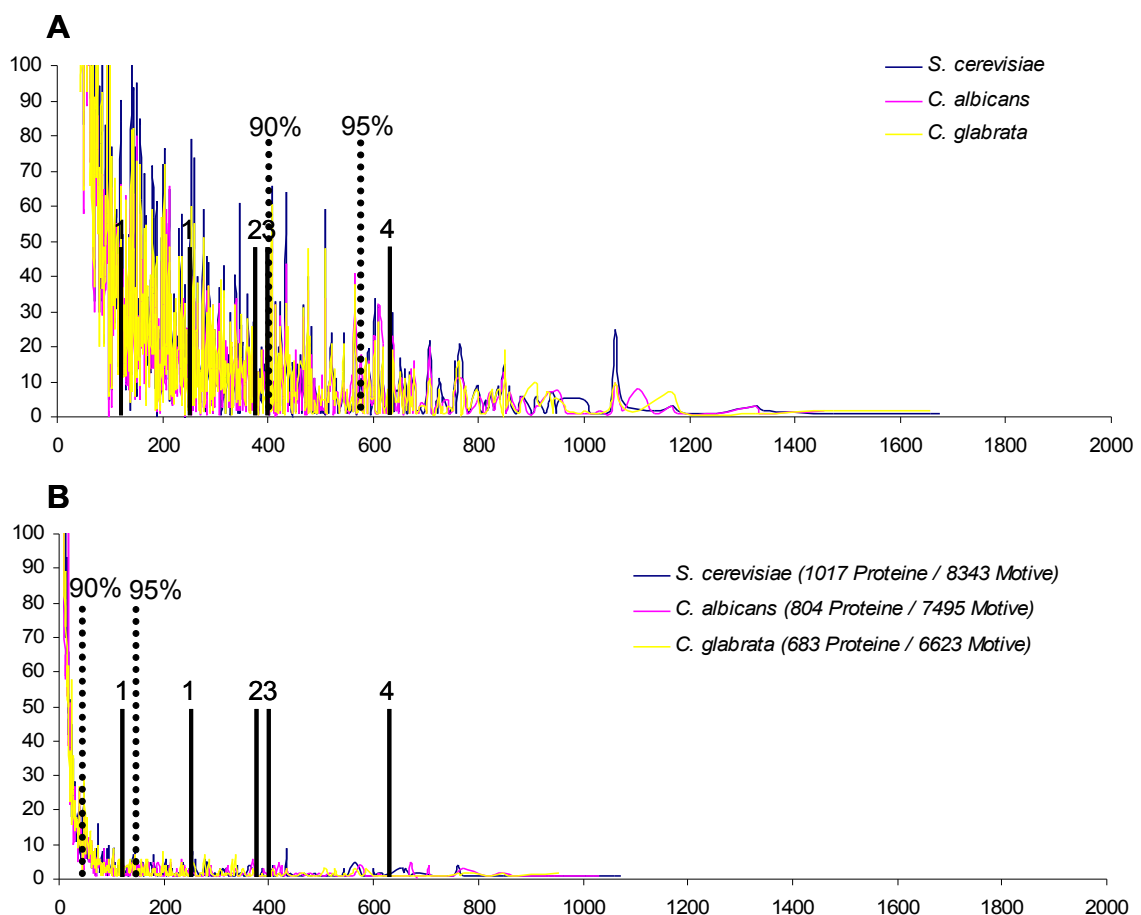
### 4.3 Verteilung von potentiellen Kex2-Schnittstellen im sekretorischen Proteom

Stellvertretend für die 17 in der Datenbank vertretenen Genome ist hier die Verteilung von Proteinen auf die einzelnen Kompartimente der beiden humanpathogenen *Candida*-Spezies *C. albicans* und *C. glabrata* und von *S. cerevisiae* dargestellt (Abbildung 21).



**Abbildung 21: Verteilung der Subzellulären Lokalisation.** Wie erwartet, werden die meisten Proteine als zytoplasmatisch eingestuft. Die Proteine des sekretorischen Wegs und der Mitochondrien machen etwa 1/3 des Gesamtproteoms aus (oben). Innerhalb der Kompartimente des sekretorischen Wegs wird fast die Hälfte der Proteine als extrazellulär vorhergesagt, etwas weniger als ER-resident. Der Rest verteilt sich zu gleichen Teilen auf die Vakuole und das Golgi-Kompartiment (mittig). Fast 3/4 aller Proteine sind nicht mit Membranen assoziiert (unten).

Wie erwartet, werden die meisten Proteine als zytoplasmatisch eingestuft, während die Proteine des sekretorischen Weges und der Mitochondrien zusammen nur etwa 1/3 des Gesamtproteoms ausmachen. Innerhalb der Kompartimente des sekretorischen Weges wird fast die Hälfte der Proteine als extrazellulär vorhergesagt, etwas weniger als ER-resident. Der Rest verteilt sich zu gleichen Teilen auf die Vakuole und das Golgi-Kompartiment.



**Abbildung 22: Verteilung der PSSM-Werte in sekretorischen Proteinen.** Im Gegensatz zum Gesamtproteom (A) werden die Schnittstellen von bekannten Kex2-Substratproteinen (1-4) im sekretorischen Subproteom (B) relativ höher bewertet. Auf diese Weise können zwar ca. 2/3 der Motive ausgeschlossen werden, da sie nicht mit der Protease in Kontakt kommen. Es zeigt sich aber auch, dass diese Methode alleine nicht ausreicht, um die Anzahl auf eine handhabbare Menge zu reduzieren. 1: CaMF $\alpha$  (Bewertung durch PSSM: 122, 264); 2: ScMF $\alpha$  (392); 3: ScPir2/ScPir4 (401), 4: CaExg1 (621).

Die Suche nach Kex2-Schnittstellenmotiven im so vorhergesagten Subproteom des sekretorischen Weges inklusive der entsprechenden Membranen zeigt, dass auf diese Weise nur 2/3 der Motive ausgeschlossen werden können. Daraus lässt sich folgern, dass diese Motive in Proteinen des sekretorischen Weges mit etwa der doppelten Frequenz auftreten, wie im restlichen Proteom. Die Bewertung bekannter Schnittstellen (Abbildung 22, 1-4) ist im sekretorischen Subproteom relativ zur Gesamtzahl der gefundenen Proteine höher als im Gesamtproteom. Dies zeigt, dass die hier entwickelte PSSM im Zusammenspiel mit der

Lokalisationsvorhersage erfolgreich Schnittstellen von Kex2 aus Genomdaten anreichern kann, die so gefilterten Daten aber manuell weiter analysiert werden müssen.

#### 4.3.1 Vorhergesagte Substrate von Kex2 bei *C. albicans*, *C. glabrata* und *S. cerevisiae*

Die oberen 10% der im vorhergehenden Abschnitt beschriebenen, aus den Genomdaten von *C. albicans*, *C. glabrata* und *S. cerevisiae* herausgefilterten Proteine mit Motiven, die Ähnlichkeit zu Kex2-Schnittstellen aufwiesen, wurden manuell auf Fehler in der Lokalisationsvorhersage überprüft. Dabei wurde ein Großteil der Membranproteine wegen unklarer Topologien aus der Analyse ausgeschlossen. Es wurden bei *C. albicans* 214 Motive aus 132 Proteinen (163/116 bei *C. glabrata* und 177/139 bei *S. cerevisiae*) weiterverwendet und in funktionelle Gruppen bzw. Gruppen von ähnlich aufgebauten Strukturen eingeteilt und diese auf konservierte Motive hin untersucht. Die Gruppen sind in den Tabelle 21 bis Tabelle 35 aufgelistet, die restlichen Motive finden sich in Anhang 11.2.

Um experimentell zu überprüfen, welche von diesen putativen Proteine von Kex2-Proteasen tatsächlich hydrolysiert werden können, wurden aus der Gesamtmenge 44 Proteine für die heterologe Expression in *E. coli* herausgesucht (Abschnitt 4.4.4) und die erfolgreich aufgereinigten Proteine *in vitro* auf die Funktionalität der Schnittstellen hin überprüft (Abschnitt 4.4.6).

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. glabrata</i>	CAGL0H03135g	85	LFVNTTLYNQATKGEKLSDFTKR/DANPDAAEAEA	402	21/(no TMs)159
<i>C. glabrata</i>	CAGL0H03135g	110	NPDAEAEAWHWVKIRKGGQLFR/ SADASPEAEA	39	21/(no TMs)159
<i>C. glabrata</i>	CAGL0H03135g	135	DASPEAEAWHWVRLRKGQGLFR/ SADASPEAEA	39	21/(no TMs)159
<i>S. cerevisiae</i>	YGL089C/MF2	100	KREAVADAWHWLNLPGQPMYKR/EANADAWHWL	392	20/(no TMs)120
<i>S. cerevisiae</i>	YPL187W/MF1	84	LLFINTTIIASIAAKEEGVSLDKR/EAEAWHWLQL	21	20/(no TMs)165
<i>S. cerevisiae</i>	YPL187W/MF1	103	LDKREAEAWHWLQLKPGQPMYKR/EAEAEAWHWL	392	20/(no TMs)165
<i>S. cerevisiae</i>	YPL187W/MF1	124	KREAEAEAWHWLQLKPGQPMYKR/EADAEAWHWL	392	20/(no TMs)165
<i>S. cerevisiae</i>	YPL187W/MF1	145	KREADAEAWHWLQLKPGQPMYKR/EADAEAWHWL	392	20/(no TMs)165
<i>C. albicans</i>	CA2412/MF $\alpha$ 1	80	DNKPVILIVNGTTLTSGANNEKR/EAKSKGGFRL	264	19/(no TMs)143
<i>C. albicans</i>	CA2412/MF $\alpha$ 1	101	KREAKSKGGFRLTNFGYFEPGKR/DANADAGFRL	122	19/(no TMs)143
<i>C. albicans</i>	CA2412/MF $\alpha$ 1	122	KRDANADAGFRLTNFGYFEPGKR/DANAEAGFRL	122	19/(no TMs)143

Tabelle 21: In  $\alpha$ -Kreuzungspheromonen vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA2825/Hwp1	38	*VPQVDGQGETEEALIQKR/SYDYYQPCD	173	20/(no TMs)(604-GPI)o634
<i>C. albicans</i>	CA2830/Rbt1	59	PVCTDLGLYHDSSISLSGSKNKR/EAEIANKDGT	432	18/(no TMs)(684-GPI)o714
<i>C. albicans</i>	CA2830/Rbt1	73	SLSGSKNKREAEIANKDGTIEKR/TFGSAGVNAG	151	18/(no TMs)(684-GPI)o714
<i>C. albicans</i>	CA2830/Rbt1	225	NAIGTFDASALFNQVKAFAKPR/ELDFEELSN	111	18/(no TMs)(684-GPI)o714
<i>C. albicans</i>	CA2830/Rbt1	243	FPAKRELDFEELSDNGVTHSKR/TLGLLLGLLK	407	18/(no TMs)(684-GPI)o714
<i>S. cerevisiae</i>	YKR102W/Flo10	131	VPVKRGVKLCSNDNTLSSKTEKR/ENDDCDQGA	91	25/(no TMs)1169
<i>C. albicans</i>	CA5699/Als7	1268	ESSSGTILSEESSDSIPTTFSTR/YWSPSGMSSR	43	19/(no TMs)(1980-GPI)o2000
<i>C. albicans</i>	CA5699/Als7	1405	ESSIGTILSEESSDSIPTTFSTR/YWSPSGMSSR	43	19/(no TMs)(1980-GPI)o2000
<i>C. albicans</i>	CA5699/Als7	1537	ESSSGTILSEENSDSIPTTFSTR/YVSVSLTVGE	43	19/(no TMs)(1980-GPI)o2000
<i>C. albicans</i>	CA5699/Als7	1915	KSVQAESVQTDSTTYSEMMSKR/NKNSGFGTSS	94	19/(no TMs)(1980-GPI)o2000
<i>C. albicans</i>	CA5713/Als6	918	LSESTDTTSSIGTDSSTVTFCCR/DNGDGCIVTG	189	19/(no TMs)(1345-GPI)o1366
<i>C. glabrata</i>	CAGL0E00275g/Flo10	398	IPMPVTTTSKIMSSRFEPFGPR/IGTWLLEDKP	36	20/(no TMs)880
<i>C. glabrata</i>	CAGL0E00275g/Flo10	449	RSHLAQFSVNLPSNDTQRPYSKR/NSSFFPSNLS	262	20/(no TMs)880
<i>C. glabrata</i>	CAGL0M00132g/Flo10	324	VGIPCSSSQSIYGSSSSSSSSKR/LDCPSGLFFD	94	24/(no TMs)(892-GPI)o922
<i>C. glabrata</i>	CAGL0M00132g/Flo10	872	AIFRSATSRHDYTTISAFPTPR/LSLQITTLVQ	40	24/(no TMs)(892-GPI)o922

Tabelle 22: In Adhäsinen vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. glabrata</i>	CAGL0I00484g/Exg1	38	*VIAPKNKDTSLHFVNEKR/YYDYDSKAIG	264	20/(no TMs)443
<i>S. cerevisiae</i>	YLR300W/Exg1	39	*VPARDPSSIQFVHEENKRR/YYDYDHGSLG	256	20/(no TMs)448
<i>C. albicans</i>	CA0822/Exg1	37	*ISNPFKPNGNLKFKR/GGGHNVAWDY	621	22/(no TMs)438
<i>A. gossipii</i>	AAR146W	44	*LPALTPNSIQVVTNTR/YFDYENKTMR	328	27/(no TMs)442
<i>C. albicans</i>	CA3867/Phr2	73	QNNLDSNESFVDPLANPEHCKR/DIPYLEAVDT	451	23/(no TMs)(510-GPI)544
<i>C. albicans</i>	CA4857/Phr1	83	GYPDAPNRKYNDPLADADACKR/DVKYFKESNT	154	25/(no TMs)(515-GPI)548
<i>C. albicans</i>	CA5572/Bgl22	59	SSINHKNQPLSLRNNPQLIQKR/SRHAESPLFI	173	21/(no TMs)924
<i>Y. lipolytica</i>	YALI0F05390g	85	ANVPVDEYHYTAWLGKEAEAKR/LTDHWNTWIT	135	16/(no TMs)421
<i>C. glabrata</i>	CAGL0D02530g/Egt2	190	GIYQTNNTAWFTLDIPISQFKR/PIRIQSTPSV	353	23/(no TMs)(656-GPI)o676

Tabelle 23: In sekretorischen Glukanasen vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA2660/Sap1	49	VTLLDFVIKTPVFNATGQEGKVKR / QALPVTLNNE	571	24/(no TMs)391
<i>C. albicans</i>	CA3138/Sap2	55	FSVVKTPKAFPVTNGQEGKTSKR / QAVPVTLNNE	94	19/(no TMs)398
<i>C. albicans</i>	CA6065/Sap3	57	VIKTHKNVTGPQGEINTNVNKKR / QTVPVKLINE	388	19/(no TMs)398
<i>C. albicans</i>	CA2055/Sap4	22	*VKRSTGFVTLDFNVKR / STGFVTLDFN	N/V	19/(no TMs)417
<i>C. albicans</i>	CA2055/Sap4	35	*VKRSTGFVTLDFNVKR / SLVDPKDPTV	388	19/(no TMs)417
<i>C. albicans</i>	CA2055/Sap4	74	LFLDIEPTEIPVDDTGRNDVGKR / GPVAVKLDNE	341	19/(no TMs)417
<i>C. albicans</i>	CA2499/Sap5	35	*GFVTLDFNVKR / SLVDPDDPTV	388	24/(no TMs)418
<i>C. albicans</i>	CA2499/Sap5	75	FLEFTPSEFPVDETRGDVDVKR / GPVAVTLHNE	131	24/(no TMs)418
<i>C. albicans</i>	CA0968/Sap6	35	*GFVTLDFNVKR / SLVDPDDPTV	388	24/(no TMs)418
<i>C. albicans</i>	CA0968/Sap6	75	FLDLDPKTIIPVDDTGRNDVGDKR / GPVAVKLDNE	131	24/(no TMs)418
<i>C. albicans</i>	CA1266/Sap8	32	*EDIDKR / AEKVSLDFT	65	26/(no TMs)405
<i>C. albicans</i>	CA1266/Sap8	74	AHQHHQSQQQQQQQQPAQKR / GTVQTSLINE	154	26/(no TMs)405
<i>C. albicans</i>	CA4700/Sap9	49	VRRGESKDDLSPEDDSNRPFVKR / DGSLDMTLTN	475	18/(no TMs)(520-GPI)544
<i>C. albicans</i>	CA4863/Sap10	37	*KLDFNKVSTPSKYTKR / DALPMLIND	315	21/(no TMs)(428-GPI)453
<i>C. glabrata</i>	CAGL0E01419g	49	YCKNSDNKSRNLDALGSNRFAKR / ASSVASASPS	334	?/(noTMs)(?-GPI)519
<i>C. glabrata</i>	CAGL0E01727g	30	*DFERQTAQDVALAKR / HTSNGVANAM	50	15/(no TMs)(519-GPI)539
<i>C. glabrata</i>	CAGL0E01749g	33	*KLDFEKVPEQNLLNKR / DDKGLQLELE	54	16/(no TMs)482
<i>C. glabrata</i>	CAGL0E01771g	33	*DFEKTPKISDELTKR / ADNDSFIAPL	60	18/(no TMs)(501-GPI)519
<i>C. glabrata</i>	CAGL0E01793g	32	*KLDFTRVPSASVLEKR / ADDLSPVPLR	48	16/(no TMs)(480-GPI)516
<i>C. glabrata</i>	CAGL0E01815g	33	*KLDFEKSPKASSDLVKKR / DDEYVNVPLK	71	16/(no TMs)(503-GPI)519
<i>C. glabrata</i>	CAGL0E01837g	32	*KLDFTKTPGSDLAQR / DVVDPEAAQL	50	17/(no TMs)(498-GPI)521
<i>C. glabrata</i>	CAGL0E01859g	29	*ALDFEKSHGDDLVKR / DVQDVKLINS	71	14/(no TMs)(484-GPI)505
<i>C. glabrata</i>	CAGL0E01881g	30	TAAAAGYLRLPFTKITDNLAKR / ADDDYLNQVL	50	(no TMs)(469-GPI)508
<i>C. glabrata</i>	CAGL0J02288g	37	VVNCRLRTIEKRIVSSHASLSKR / SAVDLQFRRF	50	(no TMs)(508-GPI)541
<i>C. glabrata</i>	CAGL0J02288g	46	EKRIVSSHASLSKR.SAVDLQFRR / FNNLYYESVL	180	(no TMs)(508-GPI)541
<i>C. glabrata</i>	CAGL0M04191g	44	ADLSYVKLDFDKYYGETFETAKR / GRSQADIRVN	94	19/(no TMs)(580-GPI)601
<i>C. glabrata</i>	CAGL0M04191g	56	YYGETFETAKRGRSQADIRVNKR / ANGYEEVQIT	336	19/(no TMs)(580-GPI)601
<i>S. cerevisiae</i>	YLR120C/Yps1	29	*IPAANKR / DDDSNSKFVK	150	22/(no TMs)(542-GPI)569
<i>S. cerevisiae</i>	YDR144C/Yps2	64	KKYGSSFENALDDTKGRTRLMTR / DDDYELVELT	7	23/(no TMs)(575-GPI)596
<i>S. cerevisiae</i>	YLR121C/Yps3	46	DGKYVKIPFTKKKNGDNGELSKR / SNGHEKFVLA	50	21/(no TMs)(483-GPI)508
<i>S. cerevisiae</i>	YGL259W/Yps5	51	FPVQKFADI INIGTQDVSTVFVKR / NEVLNTTVIN	483	22/(no TMs)165
<i>S. cerevisiae</i>	YIR039C/Yps6	51	FPVQKLADI INICTQDVSTVFVKR / NEVLNTTVIN	483	22/(no TMs)537

Tabelle 24: In sekretorischen Aspartatproteasen vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA2633/Kex2	99	DKSHPHNSFLGNHNSNEYNLMKR/QLGHEQDYDE	55	21/o774-794i936
<i>C. albicans</i>	CA2633/Kex2	128	DYDELISHVESIHLLPMKKLSKR/IPVPIEMEDV	50	21/o774-794i936
<i>C. albicans</i>	CA2633/Kex2	526	YSDVIQVNQTTTTPPEQKAPSKR/DSPQKIHSS	111	21/o774-794i936
<i>C. albicans</i>	CA5814/subtilisin	67	NDSYFVSLKASESIDHFMEYDKR/YPKHLQVRDF	110	i21-40o421
<i>C. albicans</i>	CA5814/subtilisin	104	IAIGDFKGFSGRFSKDIIDRLKR/CPLVQEITED	784	i21-40o421
<i>C. albicans</i>	CA5814/subtilisin	140	NALEFEIQEDAPRHLARI SRRRR/MKPNKPYPYM	266	i21-40o421
<i>C. albicans</i>	CA5815/subtilisin	116	KAYDAVFDQDSAPRHLARI SRRKR/MKPIKKYSYI	522	22/(no TMs)379
<i>C. glabrata</i>	CAGL0J00583g/Dap2	516	ALLSYDGPDPVYQKIIELDTDKR/DEHIHNSIG	40	i24-45o817
<i>C. glabrata</i>	CAGL0J07546g/Kex2	89	YVFSVAKSENEDDIDSGNSLRKR/AIENEDHILS	47	24/o671-693i789
<i>C. glabrata</i>	CAGL0J07546g/Kex2	112	AIENEDHILSSEFLTSPNKLQKR/MPVPAPPLDS	55	24/o671-693i789
<i>C. glabrata</i>	CAGL0L13266g/Prb1	45	DINTLEQTSESEQEYDNFFKQSKR/PGNVKAPLVA	228	20/(no TMs)579
<i>C. glabrata</i>	CAGL0L13266g/Prb1	114	VAKLQLRGSISLAPNDFPFHHTR/DKEFSTEEELG	82	20/(no TMs)579
<i>C. glabrata</i>	CAGL0M02211g/Pep4	68	VQQLGHKYVSYQEKANPGTVLPR/DHLFYPDQVH	43	23/(no TMs)415
<i>S. cerevisiae</i>	YEL060C/Prb1	86	HGKFHKKGRKGQDKESPEFNGKR/ASGSHGSAHE	298	19/(no TMs)635
<i>S. cerevisiae</i>	YEL060C/Prb1	188	TAQFNPD AISKIIPNRYIIVFKR/GAPQEEIDFH	483	19/(no TMs)635
<i>S. cerevisiae</i>	YNL238W/Kex2	79	EHDVRLPNHYVFSKELLKLGKR/SSEELQGDN	54	24/o676-699i814
<i>S. cerevisiae</i>	YNL238W/Kex2	108	QGDNDHILSVHDLFPRNDFKRLPVPAPMDS	77	24/o676-699i814
<i>S. cerevisiae</i>	YNL238W/Kex2	503	TEETLESVITISEKSLQDANFKR/IEHVTVTVDI	422	24/o676-699i814
<i>S. cerevisiae</i>	YOR219C/Ste13	162	SKLLPTKITRPKTSAGDSSLGKR/SFSIENVLNG	54	i120-139o931

Tabelle 25: Vorhergesagte Schnittstellen in anderen Proteasen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA0177	123	DEGSDCDDDDCKKKKKAHRYTKR/CGGGDDDDCE	315	19/(no TMs)275
<i>C. albicans</i>	CA0856	190	SGSDCEDDECKRKKKVRHNYKR/GGYGSERRQS	363	o28-45i57-78o356
<i>C. albicans</i>	CA0856	198	ECKRKKKVRHNYKRGGYGSERR/QSDCESDCER	46	o28-45i57-78o356
<i>C. albicans</i>	CA0856	211	KRGYGSERRQSDCESDCERSR/CNYPFCELYE	50	o28-45i57-78o356
<i>C. albicans</i>	CA1334	116	YRNSEEECSDEDDHKKKKR/PHRHGGKSDD	376	19/(no TMs)201
<i>C. albicans</i>	CA1334	135	KKKRPHRHGGKSDDDDDKKWKR/GGDYSDNDN	552	19/(no TMs)201
<i>C. albicans</i>	CA1844	123	DEGSDCDDDDCKKKKKAHRYAKR/CGGGDDDDCE	262	19/(no TMs)275
<i>C. albicans</i>	CA2912	84	DDGQLERNDKKLDCKNSRVS/PAPSPSAIAV	31	19/(no TMs)247
<i>C. albicans</i>	CA2912	119	GNEECDCCDEHRKKGSKYKR/GEVENPRETR	303	19/(no TMs)247
<i>C. albicans</i>	CA2912	129	DEHRKKGSKYKRGEVENPRETR/DCDFCTIEKS	70	19/(no TMs)247
<i>C. albicans</i>	CA3030	70	IQVVTVESASALSTDTATSTLTR/NDNKKEATPV	17	19/(no TMs)(328-GPI)346
<i>C. albicans</i>	CA3614	121	DNDECEDEDCDEDKKKGKQYKR/GEVEEPCETS	303	19/(no TMs)251
<i>C. albicans</i>	CA3720	128	DDDDNDCDDCKKKKKVYFAKR/GDDDDDKCD	334	25/(no TMs)274
<i>C. glabrata</i>	CAGL0I06160g	60	AKPEYASFGI AVKPISSVAKR/AVVSQIGDGQ	312	19/(no TMs)233
<i>C. glabrata</i>	CAGL0I06182g	61	TDYASTFGI AVQPIATPASVAKR/AVSQIGDGQV	312	19/(no TMs)340
<i>C. glabrata</i>	CAGL0I06204g	61	TDYASTFGI AVQPIATPASVAKR/AVSQIGDGQV	312	19/(no TMs)349
<i>C. glabrata</i>	CAGL0M08492g	62	TDYSSSTFGI AVQPIATSSVAKR/AVSQIGDGQV	312	20/(no TMs)335
<i>S. cerevisiae</i>	YJL158C/Cis3	63	GAAEYTTTFGI AVQAITSSAKR/DVISQIGDGQ	401	22/(no TMs)(209-GPI)227
<i>S. cerevisiae</i>	YJL160C	61	TEYRTSFLAVIPFTVTESKVKR/NVISQINDGQ	571	24/(no TMs)180
<i>S. cerevisiae</i>	YKL163W/Pir3	66	SFGIAIEAVATSASSVASSAKR/AASQIGDGQV	401	19/(no TMs)325
<i>S. cerevisiae</i>	YKL164C/Pir1	62	DYSSSTFGI AVEPIAT TASSAKR/AAASQIGDG	401	19/(no TMs)341

Tabelle 26: In Pir-Proteinen vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA1402/Ece1	30	* IHHAPEFNMKR / DVAPAAPAAP	303	19/(no TMs)271
<i>C. albicans</i>	CA1402/Ece1	60	AAPADQAPTVPAPQEFNTAITKR / SIIGIIMGIL	187	19/(no TMs)271
<i>C. albicans</i>	CA1402/Ece1	92	LGNIPQVIQIIMSIVKAFKGNKR / EDIDSVVAGI	18	19/(no TMs)271
<i>C. albicans</i>	CA1402/Ece1	125	IADMPFVVRVAVDTAMTSVASTKR / DGANDDVANA	114	19/(no TMs)271
<i>C. albicans</i>	CA1402/Ece1	159	VRLPEIVARVATGVQQSIENAKR / DGVPDVGLNL	273	19/(no TMs)271
<i>C. albicans</i>	CA1402/Ece1	193	ANAPRLISNVFDGVSETVQQAKR / DGLEDFLDEL	228	19/(no TMs)271
<i>C. albicans</i>	CA1402/Ece1	227	QRLPQLITRSAESALKDSQPVKR / DAGSVALSNL	158	19/(no TMs)271
<i>C. albicans</i>	CA3122	41	* IFDLESKQGLQDEETVNNDKR / EPVNLLYLDR	115	19/(no TMs)(?-GPI)410
<i>C. albicans</i>	CA3122	68	LLYLDRFKMGVSDAEAKGNKFKR / DPKNVIDPAS	621	19/(no TMs)(?-GPI)410
<i>C. albicans</i>	CA3122	92	PKNVIDPASLKEGSAEEEEQDKR / EPKNLFLNQA	169	19/(no TMs)(?-GPI)410
<i>C. albicans</i>	CA3122	116	PKNLFLNQAALHEGLKDEETKSKR / EAKNLPNLEA	401	19/(no TMs)(?-GPI)410
<i>C. albicans</i>	CA3122	161	KDAKNLIDLVALKQSLEKEAAKR / DAKNIPDLEA	139	19/(no TMs)(?-GPI)410
<i>C. albicans</i>	CA3122	186	KNIPDLEALKTGIEEEEGQVAKR / DAKNVINLSN	312	19/(no TMs)(?-GPI)410
<i>C. albicans</i>	CA3122	204	QVAKRDAKNVINLSNFIETPSKR / EGKNLFDLTK	111	19/(no TMs)(?-GPI)410
<i>C. albicans</i>	CA3122	225	KREGKNLFDLTKFQQSGQPIKRR / DQKILKQEK	146	19/(no TMs)(?-GPI)410
<i>C. albicans</i>	CA3295	318	GGILGGSSGYTGSTSKVCTLKKR / SLEKKEIRKL	47	23/(no TMs)(?-GPI)1249
<i>C. albicans</i>	CA3295	354	SVNKVIARRQQAIAEKKQANAKR / ESISFVSFEA	273	23/(no TMs)(?-GPI)1249
<i>C. albicans</i>	CA3295	401	VTAVSSVLLTVAALTNANEIDKR / SFFGDLFSGL	65	23/(no TMs)(?-GPI)1249
<i>C. albicans</i>	CA3295	551	APSSSEAPSSSSSEALSSSSTTKR / PTAAAKGFFG	114	23/(no TMs)(?-GPI)1249
<i>S. cerevisiae</i>	YFL051C	33	* EACLPTNKR / EDGMNINFYE	102	24/(no TMs)160
<i>S. cerevisiae</i>	YFL051C	99	LSIYSPPESTPTCVTYAVLKR / DEDGYDPCGP	439	24/(no TMs)160
<i>S. cerevisiae</i>	YFL051C	116	YAVLKRDEDEGYDPCGPLYETKRR / DTEYCDPNTA	88	24/(no TMs)160
<i>C. albicans</i>	CA0365	31	* IPQESTTEVDKR / LDADVAAQLA	131	19/(no TMs)178
<i>C. albicans</i>	CA0365	62	LALNILNLIQLGIGANVNVTAQR / DESAVTIEQS	94	19/(no TMs)178
<i>C. albicans</i>	CA0365	74	IGANVNVTAQRDESAVTIEQSKR / LDADVAAQLA	228	19/(no TMs)178
<i>C. albicans</i>	CA0365	105	LAGNILNLIQLGAGAEISASTKR / DESAVDTVEE	114	19/(no TMs)178
<i>C. albicans</i>	CA0365	118	GAEISASTKRDESAVDTVEESKR / LDADVAAQLA	5	19/(no TMs)178
<i>C. albicans</i>	CA0365	149	IAANILNLIQLGIGADVNVSAKR / EDTVAAQIGA	94	19/(no TMs)178

Tabelle 27: In Polypeptid-Vorläuferartigen Proteinen vorhergesagte Schnittstellen (ohne  $\alpha$ -Kreuzungspheromone).

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA0229	193	TNDDSQLSSFSRQILEAHNIKR / ASHGVNPLTW	234	24/(no TMs)317
<i>C. albicans</i>	CA5344	134	DATPTADVEFAEELKHEHNVKR / ALHGVPAISW	388	20/(no TMs)271
<i>S. cerevisiae</i>	YJL079C/Pry1	173	SSSDSDLSDFASSVLAHEHNKRR / ALHKDTPALS	256	20/(no TMs)299
<i>S. cerevisiae</i>	YKR013W/Pry2	203	SSTQSSSDFSTSMVNEHNTKR / ALHKDTGSLT	328	19/(no TMs)329
<i>C. glabrata</i>	CAGL0F05137g	76	KKTKTQSQKTKPTPSTDLDKR / AQKKRNLSE	21	18/(no TMs)227
<i>C. glabrata</i>	CAGL0F05137g	80	KSQKTKPTPSTDLDKRAQKRR / SNLSEWQQKM	256	18/(no TMs)227
<i>C. glabrata</i>	CAGL0F05137g	98	QKKRNLSEWQQKMLDQHNKRR / ELHKDTPDSL	214	18/(no TMs)227
<i>C. glabrata</i>	CAGL0G07667g	132	PAQDSNLSDFAKSMLEHNIKR / ALHQDTNPLT	234	18/(no TMs)258

Tabelle 28: In Pry-Proteinen vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA2974	54	LTVRESNLVNSALANLQHYNAKR/DLMSQEEIIK	273	21/(no TMs)328
<i>C. albicans</i>	CA2974	65	ALANLQHYNAKRDLMSQEEIIKR/ENQIVTDVLT	134	21/(no TMs)328
<i>C. albicans</i>	CA2974	180	NLIGNALGLNANTVSSKMASAKR/EIMVESAPEP	94	21/(no TMs)328
<i>C. albicans</i>	CA3713/Ops4	33	*ANIVSEQTMVKK/EDVNAIVELI	420	21/(no TMs)403
<i>C. albicans</i>	CA3713/Ops4	53	VKREDVNAIVELINEIKKHINQKR/DLAEGEDLLE	303	21/(no TMs)403
<i>C. albicans</i>	CA3713/Ops4	67	EIKKHINQKRDLAEGEDLLELQRR/ADSVIGELVS	28	21/(no TMs)403
<i>C. albicans</i>	CA3713/Ops4	191	SSSSSSAAPAAAAAPATNGASKR/EIMEAAEYLS	139	21/(no TMs)403
<i>C. famata</i>	DEHA0C08173g	33	*AGTPTTTTSLTKR/DAENIGKAIA	60	21/(no TMs)329
<i>C. famata</i>	DEHA0C08173g	52	LTKRDAENIGKAIALLNDYNAKR/ETSPSYALAE	273	21/(no TMs)329
<i>C. famata</i>	DEHA0C08173g	63	AIALLNDYNAKRETSPSYALAER/DYPIVTTVLG	4	21/(no TMs)329
<i>C. famata</i>	DEHA0C08173g	170	GSYIGDLFDKIKDKISNIGSSKR/ESLEERALTT	94	21/(no TMs)329
<i>C. famata</i>	DEHA0E15136g	33	*AAVVAENSLVTR/GDANEILLEIL	9	21/(no TMs)367
<i>C. famata</i>	DEHA0E15136g	53	VTRGDANEILLEILSSLKQVNEKR/DLVSESEFSE	264	21/(no TMs)367
<i>C. famata</i>	DEHA0E15136g	67	SLKQVNEKRDLVSESESEFELSKR/ADGLVTELVA	50	21/(no TMs)367
<i>C. famata</i>	DEHA0E15136g	218	SSGSDVDISGIPSPSEYNSGAKR/DFIDGFLDK	16	21/(no TMs)367
<i>C. famata</i>	DEHA0F12496g	73	IHSLVSSLNQYNAAHKVPGYTKR/EDILVREAHE	315	23/(no TMs)310
<i>C. famata</i>	DEHA0F12496g	187	LPGLLRIGKELLKQNGIDIFSKR/SIDTTELETA	334	23/(no TMs)310
<i>C. famata</i>	DEHA0F12496g	208	KRSIDTTELETAAVFEETSIDKR/ESQLLNELFA	65	23/(no TMs)310
<i>C. albicans</i>	CA1873	39	*EQQQQESQFASKALTKR/EEQDIQELVQ	60	21/(no TMs)351
<i>C. albicans</i>	CA1873	58	LTKREEQDIQELVQHINNYKTRR/DAIDEEIMKR	246	21/(no TMs)351
<i>C. albicans</i>	CA1873	68	ELVQHINNYKTRRDAIDEEIMKR/DYAIIVTDVLA	173	21/(no TMs)351
<i>C. albicans</i>	CA1873	195	QLITRDETDIDALAPYVVTMEKR/LDLDGVDNL	286	21/(no TMs)351
<i>C. albicans</i>	CA4679/PGA17	322	FGSVPTSINYSALSQISGALRKR/EYNDAVEAAL	47	20/(no TMs)(535-GPI)557
<i>C. albicans</i>	CA4679/PGA17	338	SGALRKREYNDAVEAALREIQKR/EEGIDDVEIA	173	20/(no TMs)(535-GPI)557
<i>C. albicans</i>	CA4679/PGA17	354	LREIQKREEGIDDVEIALRKMKR/DNIEDLLTTI	445	20/(no TMs)(535-GPI)557
<i>C. albicans</i>	CA4679/PGA17	457	SRAFNDDELKAVLWNDITSIFKR/DMAFRDEIVK	241	20/(no TMs)(535-GPI)557
<i>C. albicans</i>	CA4679/PGA17	468	VLWNDITSIFKRDMAFRDEIVKR/SNGTITSLPV	222	20/(no TMs)(535-GPI)557

Tabelle 29: In OPS4-ähnlichen Proteinen vorhergesagte Schnittstellen.



Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA0825/Plb2	37	*PTNLYTPGYVQCPEGKLTR/SSLDGINSNE	73	18/(no TMs)609
<i>C. albicans</i>	CA0825/Plb2	299	LRSFVDTKYIGTRLDGAPVSKR/CVNGFDNAGF	312	18/(no TMs)609
<i>C. albicans</i>	CA1975/Plb1	193	GIEHPKQVEEWNVGNQVASKR/NANFNVSLTD	139	18/(no TMs)605
<i>C. albicans</i>	CA1975/Plb11	552	VGCAIIRREQERQIEQTEQCKR/CFENYCWDGT	253	18/(no TMs)605
<i>C. albicans</i>	CA2223/Plb5	138	SKKTFIRTASELSQQEKDYIHKR/QETTNKNLID	244	20/(no TMs)(720-GPI)o754
<i>C. albicans</i>	CA3182/Lip4	408	ALTWIIINRFNGKQTVSGCQHVKR/TSNFEYPNIP	578	15/(no TMs)459
<i>C. glabrata</i>	CAGL0J11748g/Plb2	201	YDKDFSRTLARWTAIGSQVQGKR/DAGFNVTITD	250	20/(no TMs)(667-GPI)o695
<i>C. glabrata</i>	CAGL0J11748g/Plb2	53	NANFIRNAADGLSPAKEWLKCR/DPIITRDALQT	47	20/(no TMs)(667-GPI)o695
<i>C. glabrata</i>	CAGL0J11770g/Plb1	73	SEIDWLKRDVNTREALESFLLKRV/VTSNFTSNSS	470	23/(no TMs)(633-GPI)o659
<i>C. glabrata</i>	CAGL0J11770g/Plb1	59	NDINLLRNATGLSQSEIDWLKCR/DVNTREALES	47	23/(no TMs)(633-GPI)o659
<i>C. glabrata</i>	CAGL0J11770g/Plb1	64	LRNATGLSQSEIDWLKRDVNTREALESFLLKRV	43	23/(no TMs)(633-GPI)o659
<i>S. cerevisiae</i>	YCR068W/Atg15	139	TNPFDFEFELRRMPLLMKRMKCR/DPEFIESYIY	24	i12-34o520
<i>S. cerevisiae</i>	YMR006C/Plb2	61	DDTSLVRNASGLSTAETDWLKCR/DAYTKEALHS	47	20/(no TMs)(680-GPI)o706
<i>S. cerevisiae</i>	YMR008C/Plb1	59	DDINLVREASGLSDNETEWLKCR/DAYTKEALHS	47	23/(no TMs)(637-GPI)o664
<i>S. cerevisiae</i>	YNL012W/Spo1	182	SEEEIVTNAKKEYNDNDLKMKCR/QGGSLITSSS	277	18/o374-396i631
<i>S. cerevisiae</i>	YNL012W/Spo1	217	YEQIEEIMNSIEEIPEDYMITKR/NLNPLARLKK	187	18/o374-396i631

Tabelle 30: In lipolytischen Enzymen vorhergesagte Schnittstellen.

	Spezies	Name	Pos.	Sequenz	PSSM	Topologie
S u n - F a m i l i e	<i>T. reesei</i>	Tr46210	39	*PSHHHHHAHRHAKKHAAARVEKR/APDVVTEVV	302	17/(no TMs)436
	<i>S. nodorum</i>	SNU10592.1	33	*RHAHQVVKR/DARPDVVY	444	25/(no TMs)467
	<i>A. gossipii</i>	ADR322W	27	*PTHKRRRHR/DSQGLTMT	444	18/(no TMs)508
	<i>A. gossipii</i>	AER372C	40	*EDCSTSAGHQHQR/AIAYEYVYE	357	28/(no TMs)436
	<i>A. fumigatus</i>	Afu7g05450	31	*PHHQHRRHQHQR/EDVVESAT	357	20/(no TMs)414
	<i>A. nidulans</i>	AN6697.2	31	*QHGHGHHHQHQR/ENVTVDGPT	407	19/(no TMs)410
	<i>C. albicans</i>	CA0883	45	*DNKNIKR/EDCDKTSFHGHKHKR/AVAYDYAYV	627	23/(no TMs)418
	<i>C. albicans</i>	CA5232	30	*PLQHQHHHHEHARR/AEVTKVVYV	207	16/(no TMs)372
	<i>C. glabrata</i>	CAGL0G08668g	44	*PQFDRYKEHSKDDVSHQHQHQR/AVTVEYVYA	214	23/(no TMs)(449-GPI)469
	<i>C. glabrata</i>	CAGL0J09922g	42	*AMPDLKDCCTTTAQQGNHQHQR/AVAVEYVYE	357	23/(no TMs)427
	<i>C. glabrata</i>	CAGL0L03289g	29	*PAVQHQDKHAHEKR/DLVVVTEYV	394	16/(no TMs)366
	<i>C. glabrata</i>	CAGL0L05434g	30	*PSPKHDEDRHAIKR/DVDVVTVVE	119	16/(no TMs)(326-GPI)346
	<i>D. hansenii</i>	DEHAOC15092g	31	*PLBGKHHQHQHQR/DSVVKQVTQ	357	18/(no TMs)384
	<i>D. hansenii</i>	DEHAOF22242g	42	*PAPAQDEDCATTQVHAHHKHKR/ENVVYDYAYV	627	21/(no TMs)404
	<i>F. graminearum</i>	FG01351.1	40	*HHNHQHIHARRHAGSKVEKR/SPDVITEYV	207	20/(no TMs)443
	<i>K. lactis</i>	KLLA0E14982g	33	*PAPGSKIHRRKEKR/AAVTVTQYY	389	20/(no TMs)398
	<i>K. lactis</i>	KLLA0F04433g	43	*PHAAKR/DEDCSTTVHAHHKHKR/AVAVEYVYQ	627	21/(no TMs)413
	<i>M. griseae</i>	MG00505.4	37	*HANHHRHLHAKRDVAKR/SPEVVVYVE	312	21/(no TMs)416
	<i>N. crassa</i>	NCU02668.1	39	*PHNHGHHHAKKHAHAGIDKR/ADGTVVVTE	65	19/(no TMs)445
	<i>S. pombe</i>	SPAC1002.13c	30	*PFRHPHLLNKR/DVSVVTSKV	54	19/(no TMs)417
	<i>S. pombe</i>	SPBC2G2.17c	?	*PLGNAPYHHHHHAGL.NASNITVGV	N/V	20/(no TMs)319
	<i>Y. lipolytica</i>	YALI0E33539g	42	*PATPDADCNEEKEVR/DFHAHQHQR/AVQVEYVYV	357	18/(no TMs)(377-GPI)414
	<i>S. cerevisiae</i>	YIL123W	34	*PHVDVHQEDAHQHQR/AVAYKYVYE	357	20/(no TMs)475
	<i>S. cerevisiae</i>	YJL116C	?	*PAPAPADSHHEDHHKDEKP.AVTVTQYI	N/V	19/(no TMs)(318-GPI)337
<i>S. cerevisiae</i>	YKR042W	37	*DFKPQFTLLSSVTKKKKKVVR/PHNFQCIHS	3	17/(no TMs)365	
<i>S. cerevisiae</i>	YNL066W	44	*PYAADIDTGTCTTAHGSQHQHQR/AVAVTYVYE	357	23/(no TMs)420	
S c w 4 / 1 0 - ä h n l i c h	<i>A. fumigatus</i>	AFU6g12380	25	*HPHVGHR/NAEHDVKPK	13	18/(no TMs)369
	<i>A. gossipii</i>	AGL354C	28	*EVAHEHDDR/AVAVVTNYR	87	19/(no TMs)451
	<i>C. albicans</i>	CA6263	31	*APLAHQHHQHKKEKR/AVHVTTTN	5	17/(no TMs)???
	<i>C. glabrata</i>	CAGL0G00308g	28	*PHAAHTHKEKR/DVVTKTMA	389	18/(no TMs)374
	<i>C. glabrata</i>	CAGL0M13805g	30	*PANVHEHKNKR/DVVTKTVMH	432	19/(no TMs)371
	<i>D. hansenii</i>	DEHAOC13937g	?	*PLQQHQHHEHKMEKK/DVKVVTQTS	0	18/(no TMs)372
	<i>K. lactis</i>	KLLA0A03201g	31	*AAAVHKHAAEKKR/DVTVVVTFN	125	18/(no TMs)378
	<i>K. lactis</i>	KLLA0C14047g	28	*PIHGKDNKR/DLVVVTNRV	6	19/(no TMs)378
	<i>Y. lipolytica</i>	YALI0D20680g	29	*PIVGHHAHQHQR/EAGIETVVFV	357	17/(no TMs)420
	<i>S. cerevisiae</i>	YGR279C	29	*PANHEKDKR/AVVTTTVQK	169	20/(no TMs)386
	<i>S. cerevisiae</i>	YMR305C	28	*PAVRHKHEKR/DVVTATVHA	394	19/(no TMs)389

T o s s i - ä h n i c h	<i>A. gossipii</i>	ADR056W	116/125	GSDGGSEQKP .GHKHNSGHKHPGHDHDKR / AVHYVHVTS	0/171	20/(no TMs)501
	<i>C. albicans</i>	CA2303	105/123	YYPSSSGNSKR / QIDDQDCNVKVVHKKHR / ATEVVQVTQ	273/627	22/(no TMs)468
	<i>C. glabrata</i>	CAGL0M05599	100/120	YPASSNSKR / AVAEEEDCDSVVAKPVHKKHR / DVAVEVIEV	273/627	18/(no TMs)446
	<i>D. hansenii</i>	DEHAOD16599	94/111	YYPSSSGSNKKR / EDEEECTTKHVHKKHR / ATEVVQVTQ	256/627	21/(no TMs)435
	<i>F. graminearum</i>	FG00876.1	106/130	SKNKR / DTESAPSPHLHGRRHLHEQRKKR / AEWVTATID	432/376	20/(no TMs)440
	<i>K. lactis</i>	KLLA0B09746	103/126	VSFKKR / DQVAEDAPCTTVSTAKKHHEHR / EAAVQWVEV	313/8	19/(no TMs)457
	<i>S. pombe</i>	SPBP23A10.11	111/125	YYPDTSNALSSLRKR / SNNQHMKRHPHHR / DDVIDSTLT	47/1074	23/(no TMs)507
	<i>Y. lipolytica</i>	YAL10A17919g	115/141	AKR / DAVEEEDKDCGETAKREFHAHHAHR / GVVTEIVQV	273/217	31/(no TMs)467
	<i>S. cerevisiae</i>	YBR162C	109/134	LKKR / SEKQSIESCKEGEAVVSRHKHQHR / DVAVEYVQV	47/357	24/(no TMs)455

Tabelle 31: In dem neuartigen „His-Box“-Motiv vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA0894/Orm1	61	TPTPIAEKSLSTPIPTPSPSHRR / GSSIGSSTSP	75	o142-162i219-235o241-259i268
<i>C. albicans</i>	CA0894/Orm1	105	SANALTPPSHHQPAHSITRQR / SSSLIQHLEP	315	o142-162i219-235o241-259i268
<i>C. albicans</i>	CA1387/Erg11	156	PNSRLMEQKKFAKFAKALTTDSFKR / YVPKIREEIL	146	o13-36i48-65o528
<i>C. albicans</i>	CA1387/Erg11	266	RRDAAQKKISATYMKEIKSRER / GDIDPNRDLI	47	o13-36i48-65o528
<i>C. albicans</i>	CA1387/Erg11	264	YWRDAAQKKISATYMKEIKSRER / ERGDIDPNRD	204	o13-36i48-65o528
<i>C. albicans</i>	CA1676/Erg9	419	LPENKFTKIVASRESIDLVSQRR / IEQENFNQCNV	176	o430-447i448
<i>C. albicans</i>	CA1676/Erg9	31	HPTELKAVIQLFGFRQPLHPGKR / DVNDKELVRC	122	o430-447i448
<i>C. albicans</i>	CA1872	358	KREKNSVIATTVSGDSPTALKR / LSGGSNSSGS	196	o895-919i943-968o972
<i>C. albicans</i>	CA1872	159	TSTGSTHPIPPAASVLSSSSQRR / ATIEDFADTE	53	o895-919i943-968o972
<i>C. albicans</i>	CA1872	116	ERTRTKKSSSIGSAELSPRSR / MMNFSLRPKF	55	o895-919i943-968o972
<i>C. albicans</i>	CA1872	337	TQIAQQYYGDARKHIEVVDPTKR / EKNSNSVIAT	134	o895-919i943-968o972
<i>C. albicans</i>	CA2007/Gpi8	143	VTVENFMRLLLTDKWDSDQPRSKR / LLSDENSNIF	557	21/o360-379i383
<i>C. albicans</i>	CA2786/Kre5	1390	DAKMIDLNNPLTRENKLDAAKR / LIPEWIEYEQ	139	23/(no TMs)1447
<i>C. glabrata</i>	CAGL0H06919g/Ero1	428	NTFDKLSVESVEYINNFEKLYKRR / EGGDSYTNFF	125	19/(no TMs)627
<i>C. glabrata</i>	CAGL0I07777g/Mpd2	164	GSRSVNPLQGFITMVVNKDVMKR / QLMDIIEGRP	346	16/(no TMs)240
<i>C. glabrata</i>	CAGL0I07777g/Mpd2	122	KPPTLDANTTVDTPTLFERVKKR / AIFGENTWTI	793	16/(no TMs)240
<i>C. glabrata</i>	CAGL0L10670g/Rot1	207	STGSHPGSSSESESQSDDTSKR / SLLKKRSLRE	94	21/(no TMs)(241-GPI)o263
<i>C. glabrata</i>	CAGL0M10571g/Are2	172	KFAGPTLEASIKSKEKELRKLRR / EQKRQLDSGE	288	o200-221i242-262o274-293i305-324o380-397i418-442o462-482i544-563o569-587i599-617o619
<i>C. glabrata</i>	CAGL0M13453g/Gpi16	319	GICKFEMEIAIVDMVNTPQSLRR / GTNPPIRPLS	68	22/o582-602i646
<i>S. cerevisiae</i>	YAL058W/Cne1	461	GPTERKFMNSRLGNLQTTFHNER / ESNPFDRII	39	30/(no TMs)502
<i>S. cerevisiae</i>	YDR331W/Gpi8	140	VTVENFIRLLTDRWTEHDHPKSKR / LLTDENSNIF	401	23/o378-397i411
<i>S. cerevisiae</i>	YHR101C/Big1	103	GLRKLDFLEFETEFVSLQRYIRR / SSTAIFKFKV	114	20/o276-295i335
<i>S. cerevisiae</i>	YIL005W/Eps1	138	SGYIKSFTETPRTKESLIAFARR / ESMDPNNLDT	170	28/o652-672i701
<i>S. cerevisiae</i>	YOL031C/Sil1	310	LISKILKADMYENDDNLILFKR / NAENWSNLQ	77	19/(no TMs)421
<i>S. cerevisiae</i>	YHR188C/Gpi16	276	HYEMDARIEMVTHVPSALARGER / PIPKPLDGNT	54	20/o548-568i610
<i>C. glabrata</i>	CAGL0M01298g/Gpi8	128	VTVENFIRLLTDRWTEHDHPKSKR / LQTDENSNIF	401	22/o359-377i390
<i>C. albicans</i>	CA4418/Erg5	510	SEEIKVFATIFPKDDLILEWKKR / DPLKSL*	167	(no TMs)517
<i>S. cerevisiae</i>	YHR079C/Ire1	34	*PLSSRSTRR / QIVEDEVAST	48	25/o534-555i897-914o1115
<i>S. cerevisiae</i>	YHR079C/Ire1	98	SSYPTPNLLNTADNRRANKKGRR / AANSISVPYL	224	25/o534-555i897-914o1115

<i>C. albicans</i>	CA0915/Kar2	308	KQFKKKHNIDITANAKAISLKR/EAEKAKRTLS	564	35/(no TMs)687
<i>S. cerevisiae</i>	YJL034W/Kar2	309	KAFKKKHGIDVSDNNKALAKLKR/EAEKAKRALS	564	i12-29o682
<i>C. glabrata</i>	CAGL0D02948g/Kar2	295	KTFKKKHGIDVSDNEKALAKLKR/EAEKAKRALS	564	29/(no TMs)667
<i>S. cerevisiae</i>	YJL159W/Hsp150	71	VQPISTTSSASSAATTASSKAKR/AASQIGDGQV	401	24/(no TMs)(280-GPI)o310
<i>S. cerevisiae</i>	YLR207W/Hrd3	269	ESGLLGKGLNSVPSSTVVRKRTRR/PDIGSPFIAQ	87	21/o768-789i833
<i>S. cerevisiae</i>	YLR207W/Hrd3	266	SDFESGLLGKGLNSVPSSTVVRKR/TTRPDIGSPF	292	21/o768-789i833
<i>C. glabrata</i>	CAGL0F06369g/Hsp70	734	IGLNVTEKYNSILSKMSFSSIRR/SSEENIKTLA	41	16/(no TMs)889
<i>C. glabrata</i>	CAGL0F06369g/Hsp70	209	SVKDEAVLVSDGLSVAIDYALKR/PDLEINVPQY	196	16/(no TMs)889
<i>C. albicans</i>	CA1507/Hsp30	13	*SAAVSTLSDIIKR/NDAVNVNPPN	134	o38-59i71-89o117-139i159-179o185-203i215-233 o253-275i342

Tabelle 32: In ER-residenten Proteinen vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA0046	395	SDALQYILTDPTWSPGLATFFKR/MDDAGAYQRL	517	24/(no TMs)415
<i>C. albicans</i>	CA0397/Fre31	27	*PRKYKR/YGQLKSFYAC	533	21/o158-181i318-340o346-368i375-393o399-417i710
<i>C. albicans</i>	CA3160/Zrt2	15	*NSDSISQVFEYLNKR/DECPTDNDYN	54	o33-52i64-85o105-127i212-235o275-295i307-326o346-366i370
<i>C. albicans</i>	CA3460/Cfl1	50	IVMSTCTTFIGKYGTCTSTGKR/STNWNVCYCKT	103	19/o212-233i288-306o326-347i368-387o402-424i431-449o455-473i760
<i>C. glabrata</i>	CAGL0C03333g/Fre2	24	*FQKR/DVWEVNAVMS	N/V	20/o186-206i257-276o296-316i337-358o370-393i400-420o426-445i727
<i>C. glabrata</i>	CAGL0E01353g/Zrt2	9	*DIVQLLFKR/DDDDGPPAC	77	o34-53i65-86o106-126i277-297o338-358i370-389o409-429i433
<i>S. cerevisiae</i>	YKL220C/Fre2	55	LVTNACTRIFQKVTWEYTSKSKR/SSPVCSYEPA	401	21/o165-184i236-256o276-297i318-339o351-370i377-395o401-419i711
<i>S. cerevisiae</i>	YNR060W/Fre4	33	*PPSKTSLINTHERR/SIYSCYVGLR	201	19/o157-180i229-249o269-290i311-332o344-364i371-388o394-412i719
<i>S. cerevisiae</i>	YOR384W/Fre5	26	*ASTKKR/TQWDQIAIDA	88	20/o162-184i205-225o237-254i266-288o350-368i375-392o398-417i694

Tabelle 33: In Proteinen des Ionenstoffwechsels vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA0932/Ept1	104	FFYAFGLFMYQTFDGCDCGHARR/TGQSGPLGEL	207	i50-69o138-158i179-200o220-240i274-292o298-322i334-353o359-379i402
<i>C. albicans</i>	CA1039/(Flavin carrier)	278	TGGSPTTYFISSTKQVLVQKHKR/SLDYVAGVAK	627	24/o170-190i202-220o327-351i372-394o406-426i466-485o491-510i522-541o553-579i808
<i>C. albicans</i>	CA5614	56	VTVSTYFVNSYTCNQPHHNHSTR/PSHYLPINGT	52	i20-42o270-293i337
<i>C. glabrata</i>	CAGL0G02959g/Nnf2	86	FFLSKKPSPDIQVVSNSAKNKKR/YVTPPTLYKL	256	i27-51o114-134i146-173o179-199i239-258o918
<i>C. glabrata</i>	CAGL0G08338g/YLR241w	176	DNAVTKAGISLLKRYIQTSNSKR/DHNSGSSNPE	273	o32-57i120-139o191-212i449-471o499-523i544-567o587-614i635-656o662-682i703-721o727-745i794
<i>C. glabrata</i>	CAGL0I02508g/Ctr2	115	EAVVTSTSATNGSLNSPSPITKR/YAVSRSLWYG	60	i64-82o131-156i175
<i>C. glabrata</i>	CAGL0M12320g/YAL053w	284	VVVSNDILSISVQRRANLLKR/GISVASSSDS	70	20/o167-191i203-220o341-367i396-416o428-448i486-506o512-531i543-564o576-602i769
<i>S. cerevisiae</i>	YBR042C/Cst26	96	VRITTENSSVPKGTFFLDLKKR/ILSHLKSNSV	376	o13-37i58-77o120-138i377-396o397
<i>S. cerevisiae</i>	YGL126W/Scs3	220	NTFIHQMHRRQGPLTTSACRR/FGGHWAGGHD	129	i12-31o51-66i87-105o236-254i326-349o355-371i380
<i>S. cerevisiae</i>	YIL037C/Prm2	400	NLSTLYPIPYQLTKLINQKIFKR/EPEKIDDKKV	241	o20-43i318-336o421-440i631-654o656
<i>S. cerevisiae</i>	YOR365C	313	ATSGSPTVVYRNKEVLSISVQKR/SLNSKIISAS	346	25/o203-222i234-259o374-396i417-436o456-478i517-537o543-561i573-595o607-634i703
<i>C. albicans</i>	CA5160/Pho89	291	STDNIPPIPEGQTLTKDYKGRR/YDEAGNLVVL	224	o7-24i44-64o85-104i146-166o186-205i225-244o476-495i566-590o600
<i>S. cerevisiae</i>	YLR083C/Emp70	249	YHDRGEGNYRVVGVIVEPVSIRK/SSPGTCETTG	81	23/o305-327i371-395o401-426i438-462o474-496i521-548o560-583i595-615o627-657i667

Tabelle 34: In luminalen Schleifen verschiedener Transmembranproteine vorhergesagte Schnittstellen.

Spezies	Name	Pos.	Sequenz	PSSM	Topologie
<i>C. albicans</i>	CA0854/Mnn1	800	CSGYMWCAYDKIGGKTKSGNTR/LEGKIINFDE	56	i14-32o828
<i>C. albicans</i>	CA4176/lfx2	193	SEIMKPIGLKNNCEWLQKEANKR/TRFPDFCKDL	150	i7-26o808
<i>C. albicans</i>	CA4295/Mnn7	188	KDTSIPDLENYLTKANDKYFKR/QEYWRVLLKD	405	i12-29o756
<i>C. albicans</i>	CA4295/Mnn7	642	HPKIDPLKNSKEADAEFGVYSRR/NMGLPDKVEK	133	i12-29o756
<i>C. glabrata</i>	CAGL0F09053g/Mnl1	608	KVSKYRQNDAGFSFSNIMSSFKR/ISEPERIFNT	146	17/(no TMs)809
<i>C. glabrata</i>	CAGL0F09053g/Mnl1	781	STLFKEVVSFEELYEDPEILHKKR/DILRAFVGNK	78	17/(no TMs)809
<i>C. glabrata</i>	CAGL0G04279g/Mnt3	578	DIESEVGIRKRSPLSTDVNVVRR/DEQQFASKEP	226	i27-44o660
<i>C. glabrata</i>	CAGL0G04279g/Mnt3	642	CTFVDRIGDADSKLVRFDKKTTR/FVNHLAQIWR	483	i27-44o660
<i>C. glabrata</i>	CAGL0G04279g/Mnt3	566	FYECESDKLQEI DIESEVGIRKR/SLPSTDVNVV	146	i27-44o660
<i>C. glabrata</i>	CAGL0H09130g/Mnn4	243	DRKKYSQASWYKMERKKADFHR/NKISDYCEPF	266	i7-25o1015
<i>C. glabrata</i>	CAGL0H09240g/Mnt3	93	INTRIMKLEEKDDPKKAGKTR/TESEEDKILL	48	33/(no TMs)642
<i>C. glabrata</i>	CAGL0H09240g/Mnt3	505	DYVIEPEDAATIGKLEYSEKDRR/YSICAIQMAH	86	33/(no TMs)642
<i>C. glabrata</i>	CAGL0H09240g/Mnt3	41	*QRRSTR/YEEDLLFSSG	72	35/(no TMs)625
<i>C. glabrata</i>	CAGL0H09240g/Mnt3	92	FFFSLENMLPSWMLSTHTAREKR/VLLNWDKSSG	540	35/(no TMs)625
<i>C. glabrata</i>	CAGL0I04532g/Mnn2	320	FWRRTMPAYEIANLPYNQNK/VRNAFDDITP	246	i12-30o612
<i>C. glabrata</i>	CAGL0L12804g/Mnn9	166	GFIIIPKSASGDAALKQLEAAVKK/AQTDKKNRNF	198	(no TMs)393
<i>C. glabrata</i>	CAGL0L12804g/Mnn9	207	DSKGFDFKLQEKERHALAVQKERR/SAMALARNEL	198	(no TMs)393
<i>C. glabrata</i>	CAGL0B04565g/Ktr2	159	GLIPREHWSYPEWIDQEKARLTR/EEMKKQKIIY	101	37/(no TMs)401
<i>C. glabrata</i>	CAGL0M05841g/Ktr2	44	DIDRNYFERSVAVKTKYDGVRRR/EYFVPLKLAT	149	(no TMs)411
<i>S. cerevisiae</i>	YGL156W/Ams1	74	HVKVWWYQVSFERGSSPVS PDKR/PSWKSIIERD	47	(no TMs)1083
<i>S. cerevisiae</i>	YGL156W/Ams1	547	GNVI PKLQVGI TVDEFYDDILKR/TNQGHDLPW	219	(no TMs)1083
<i>S. cerevisiae</i>	YIL014W/Mnt3	61	KLNNHASERTAVESSAFNWIEKR/QHQVRSENLM	151	31/(no TMs)630
<i>S. cerevisiae</i>	YIL014W/Mnt3	387	NEQTFQYCIEMLNVEPSAQERR/FIGSRLVFDS	112	31/(no TMs)630
<i>S. cerevisiae</i>	YIL014W/Mnt3	557	EDDFGREPEYFKSRYGDISKLKR/IYDASLNVEG	564	31/(no TMs)630
<i>S. cerevisiae</i>	YJR075W/Hoc1	292	KPGHPVLREIISRIETTLQKRR/DDQLNVNLRN	214	29/(no TMs)396
<i>S. cerevisiae</i>	YKL201C/Mnn4	1039	DKDPIIVYEDYAYAKLLEERKRR/EKKKKEEEEK	266	i28-45o1178
<i>S. cerevisiae</i>	YNR059W/Mnt4	279	PQELWFLDVKMDLNDEYLARFKR/FSNKWLAITF	862	31/(no TMs)580
<i>S. cerevisiae</i>	YHR204W/Mnl1	511	DVIFSTEAHPMWLPQEVRSNYKR/NAKFNNSVYS	363	21/(no TMs)796
<i>S. cerevisiae</i>	YHR204W/Mnl1	764	ILPRHGSVQIKKHSPLVTSNGTR/EEDEFKMDGI	38	21/(no TMs)796
<i>S. cerevisiae</i>	YHR204W/Mnl1	69	DNYLEHGYPYDEVKPI SCVPKRR/NFEDPTDQGT	104	21/(no TMs)796

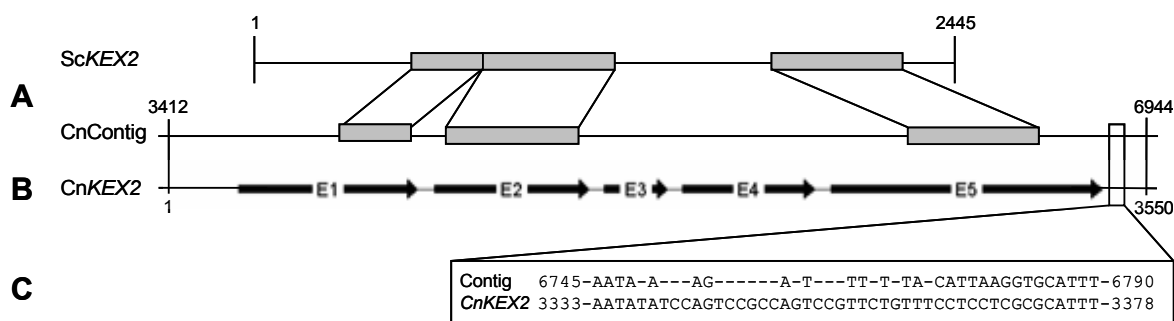
Tabelle 35: In Proteinen des Mannosylierungsapparats vorhergesagte Schnittstellen.

## 4.4 *In vitro* Überprüfung einiger putativer Substrate

### 4.4.1 Klonierung eines *KEX2*-homologen Gens aus *Cryptococcus neoformans*

*Cryptococcus neoformans* ist ein wichtiger humanpathogener Pilz, der einen Großteil seiner Virulenz über das  $\alpha$ -Pheromon reguliert. Daher war die Betrachtung des Kex2-Proteins dieser Spezies von großem Interesse. Da die Gensequenz jedoch aus dem Genomprojekt nur teilweise ersichtlich war, wurde die genomische experimentell bestimmt. Weiterhin besitzt die genomische DNA von *Cr. neoformans*, im Gegensatz zu *Candida* Spezies, häufig Introns. Daher wurde ebenfalls die cDNA Sequenz von Cn*KEX2* ermittelt.

Der *Cr. neoformans* Stamm ATCC34544 wurde für zwei Tage in Sabouraud-Medium bei 30°C kultiviert und aus den pelletierten Zellen die genomische DNA isoliert. Gesamt RNA wurde aus Zellen einer gleichen Kultur isoliert und zur cDNA-Synthese eingesetzt. Aus der Nukleotidsequenz des Genomprojektes wurde ein Primerpaar (Cn<sub>fwd</sub> und Cn<sub>rev</sub>) zur Amplifikation der genomischen DNA abgeleitet, welche sich flankierend zur kodierenden Sequenz des Gens befanden und unter Verwendung einer korrekturlesenden DNA-Polymerase die genomische DNA amplifiziert ( $T_A = 58^\circ\text{C}$ ,  $t_E = 3'30''$ , 33 Zyklen). Da sich aus der cDNA mit diesen Primern kein Produkt amplifizieren ließ, wurde ein weiteres Primerpaar (Cneo\_seq1f, Cneo\_seq1r) innerhalb der kodierenden Sequenz, aber außerhalb der putativen Introns abgeleitet und mit diesem auch die cDNA amplifiziert (PCR-Bedingungen wie oben). Beide PCR-Fragmente wurden in den Vektor pCR2.1 kloniert und mit verschiedenen Oligonukleotiden sequenziert (Tabelle 9). Die Sequenz von Cn*KEX2* wurde in GeneBank unter der Nummer AY184224 hinterlegt (siehe Anhang 11.1).



**Abbildung 23: Klonierung und Sequenzierung des *KEX2*-Gens aus *Cryptococcus neoformans*.** A: Mittels BLAST-Analyse identifizierte, zu *ScKEX2* homologe Bereiche (graue Balken) im Genom von *Cr. neoformans*. B: Intron-Exon Struktur des Cn*KEX2*-Gens ermittelt aus dem Vergleich von genomischer zu cDNA Sequenz. C: Die einzige starke Abweichung zwischen Contig- und der hier ermittelten Sequenz war ein kurzer Bereich im 3'-untranslatierten Bereich.

## 4.4.2 Heterologe Expression und Aufreinigung Kex2-ähnlicher Proteasen

### 4.4.2.1 Heterologe Expression in *Pichia pastoris*

Die Expression von *S. cerevisiae* Kex2 ist in der Literatur sowohl in *S. cerevisiae* [26] als auch *P. pastoris* [131] beschrieben. Bei den Kex2-orthologen Proteasen handelt es sich um Transmembranproteine. Zur Erzeugung von löslichen, sekretorischen Proteinen kann der nur jeweils für den luminalen Teil des Proteins kodierende vordere Teil des Gens kloniert werden, wie in den in der Literatur beschriebenen Vektoren zur Expression von *S. cerevisiae* Kex2 (Abbildung 24). In dieser Arbeit wurden verschiedenen Plasmidvektoren konstruiert, die zur heterologen Expression der löslichen Formen von Kex2 aus *C. albicans*, *C. glabrata* und *Cr. neoformans* verwendet werden sollten. Dieses erfolgte durch Amplifikation des Fragments aus genomischer DNA (*C. albicans* und *C. glabrata*) oder cDNA (*Cr. neoformans*) mit spezifischen Oligonukleotiden, wie unter Abschnitt 3.3.5.1 beschrieben.

Da nicht bei allen *KEX2*-homologen Genen die exakte, für die Transmembrandomäne kodierende Region bestimmt werden konnte, und mit den zunächst hergestellten Konstrukten keine Expression der Protein gefunden wurde, wurden weitere luminaire Bereiche verschiedener Längen mit und ohne fusioniertem C-terminalem 6xHis-tag zur Expression getestet. Zur Herstellung von Kex2-homologen Proteasen aus verschiedenen Pilzen wurde das *Pichia pastoris*-Expressionssystem der Firma Invitrogen verwendet. Die Gene der Proteasen kodieren bereits auch für ein Signalpeptid, daher wurde als Vektor das Plasmid pIC3.5 ausgewählt und das native Signalpeptid der Proteasen verwendet.

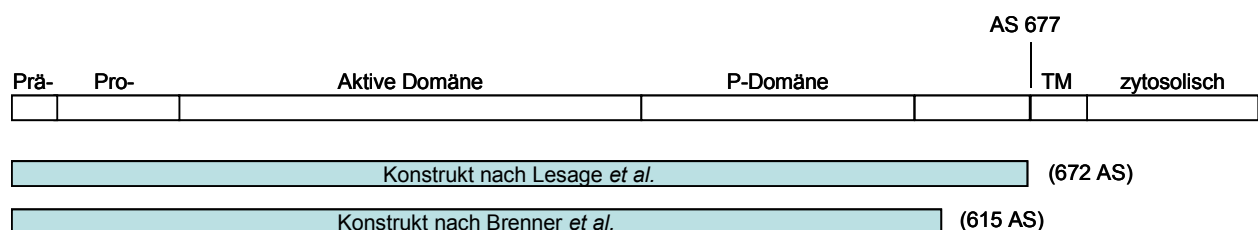
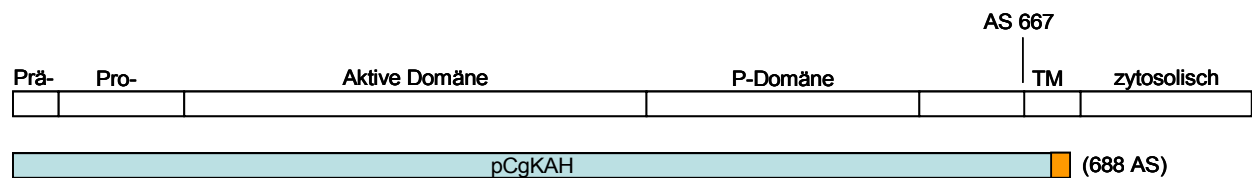


Abbildung 24: In der Literatur beschriebene Vektorkonstrukte zu Expression von *S. cerevisiae* Kex2.

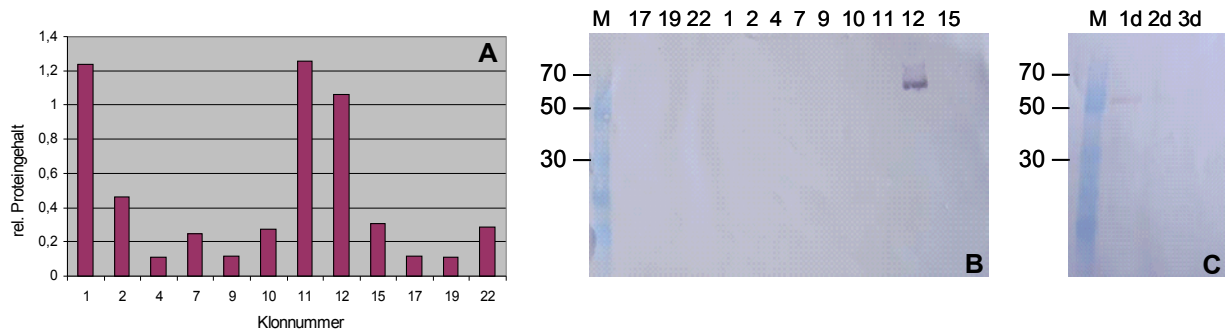


***Candida glabrata* Kex2**

Durch die Transformation des Stammes GS115 mit dem Plasmid pCgKAH (Abbildung 25) wurden 22 Transformanten erzeugt. Von 12 dieser Transformanten wurde zunächst in den Kulturüberständen einer Testexpression der relative Proteingehalt bestimmt (Abbildung 26A) und die Klone anschließend im Westernblot auf eine Bande der richtigen Größe hin überprüft. Eine solche Bande wurde lediglich bei Klon #12 gefunden (Abbildung 26B) und dieser im Folgenden für die Expression von His-markiertem *C. glabrata* Kex2 (CgKex2) verwendet. Die maximale Anreicherung von CgKex2 im Kulturüberstand konnte bereits 24 h nach Induktion beobachtet werden. Kulturüberstände untransformierter Zellen zeigten keine Kex2-ähnliche Aktivität (Abbildung 26C).



**Abbildung 25:** Vektorkonstrukts zur heterologen Expression von *C. glabrata* Kex2 in *P. pastoris*.



**Abbildung 26:** *C. glabrata* Kex2 Expression. A: Bestimmung des Gesamproteingehalts im Kulturüberstand der Testexpression. B: Nachweis eines His-markierten Proteins im Kulturüberstand von Klon #12. C: Die Expression von CgKex2 erreicht bereits nach 24 h ihr Maximum.

### *Candida albicans* Kex2

Zur Expression von *C. albicans* Kex2 (CaKex2) wurden sukzessive vier verschieden lange Vektoren (pCaKAH0, pCaKAH1, pCaKAH2 und pCaKAH3) konstruiert, wobei die Region nahe der Transmembranregion mehrmals verkürzt wurde (Abbildung 27). Insgesamt wurden 440 Transformanten (pCaKAH0: 15, pCaKAH1: 291, pCaKAH2: 21, pCaKAH3:53) auf Sekretion eines His-markierten Proteins bzw. eine Kex2-Aktivität gegen das synthetische Substrat KVKR-pNPP im Kulturüberstand getestet. Es konnte nur bei zwei Transformanten von pCaKAH3 ein schwaches Signal im Dotblot nachgewiesen werden (Abbildung 28, Punkte B9 und B10). Es zeigte sich aber in einer Westernblot Analyse, dass dieses Protein lediglich eine Größe von ~20 kDa besaß (Abbildung 28, Spuren B9 und B10). Es handelte sich daher entweder nicht um das richtige Protein, oder nur um ein Abbauprodukt. Es wurde demzufolge ein anderer Ansatz gewählt um dieses Protein herzustellen (siehe Abschnitt 4.4.2.2).

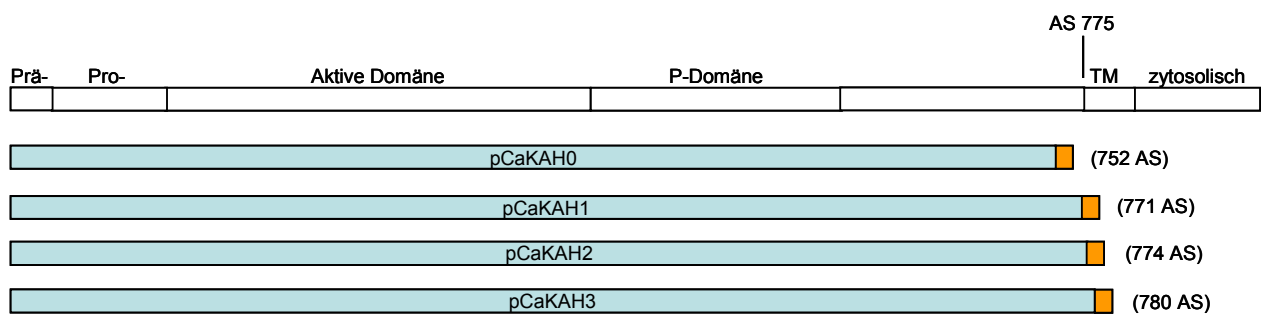


Abbildung 27: Vektorkonstrukte zur Expression von CaKex2 in *P. pastoris*.

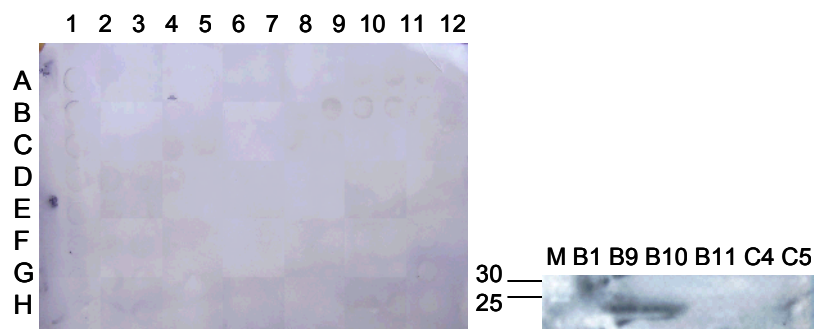


Abbildung 28: *C. albicans* Kex2-Expression in *P. pastoris*. Dotblotanalyse von Transformanten auf HIS-markiertes Protein im Kulturüberstand (links). Westernblot Analyse ausgewählter Transformanten auf HIS-markiertes Protein der richtigen Größe (rechts). Bei den zwei Transformanten mit dem stärksten Signalen im Dotblot lässt sich eine ~20 kDa große Bande beobachten.

### *Cryptococcus neoformans* Kex2

Zur Expression von *Cr. neoformans* Kex2 (CnKex2) wurden sukzessive drei verschieden lange Vektoren konstruiert (pCnKAH1, pCnKAH2 und pCnKAH3; Abbildung 29). Insgesamt wurden 32 Transformanden (pCnKAH1: 10, pCnKAH2: 10, pCnKAH3: 12) auf Sekretion eines His-markierten Proteins oder einer Kex2-Aktivität gegen das synthetische Substrat KVKR-pNPP im Kulturüberstand getestet. Dabei konnte kein Transformand mit der gewünschten Aktivität gefunden werden. Aufgrund der bereits mit der Expression von CaKex2 in *P. pastoris* gewonnenen Erfahrungen (siehe Vorhergehender Abschnitt) wurden diese Versuche nicht weiterverfolgt. Experimente, Kex2-Enzyme aus *Aspergillus fumigatus* oder *Histoplasma capsulatum* zu exprimieren, verliefen ebenso erfolglos (Daten nicht gezeigt).

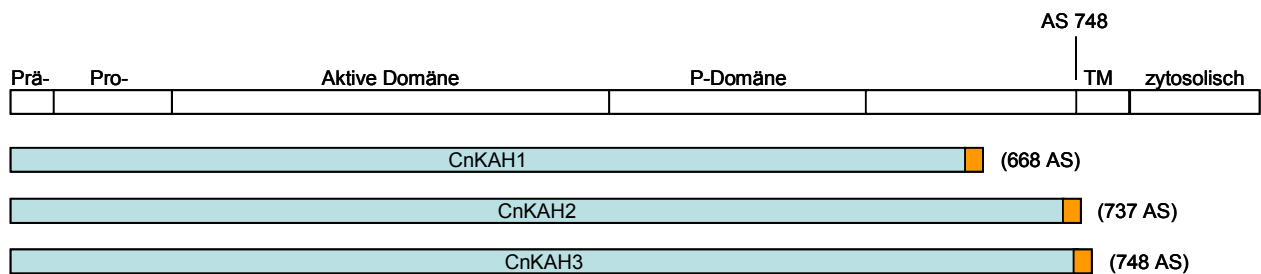


Abbildung 29: Vektorkonstrukte zur Expression von *Cryptococcus neoformans* Kex2.

#### 4.4.2.2 Heterologe Expression von CaKex2 in *Candida albicans*

Da das Kex2-Protein aus *C. albicans* in *Pichia pastoris* bei der Expression bereits degradierte, wurde für weitere Versuche mit diesem Protein als Expressionssystem *C. albicans* selbst gewählt. Dazu standen zwei verschiedene Vektoren mit unterschiedlichen Promotoren zur Verfügung: zum einen das Plasmid pCip10 [132] mit dem konstitutiven ACT1-Promotor, zum anderen das Plasmid pBI [133] mit dem unter Glukosemangel-Bedingungen aktiven Promotor des *PCK1* Gens aus *C. albicans* (Abbildung 30). Mit beiden Plasmiden wurden Expressionsvektoren konstruiert (pBI-CaKex2 und pCip10-CaKex2). In keiner von ca. 50 getesteten pBI-CaKex2-0 Transformanden konnte eine Kex2-ähnliche Aktivität im Kulturüberstand nachgewiesen werden (nicht gezeigt).

Dagegen zeigten die Kulturüberstände aller acht mit dem Plasmid pCip10-CaKex2 erzeugten CAI-4 Transformanden Kex2-Aktivität gegenüber dem synthetischen Substrat KVKR-pNA in einer Testexpression. Diese war bei Klon #7 nach 24 h Kulturdauer am stärksten (Abbildung 31A). Kulturüberstände untransformierter Zellen (Wildtypstamm SC5314) zeigten keine Kex2-ähnliche Aktivität gegenüber KVKR-pNA (Abbildung 31 B).

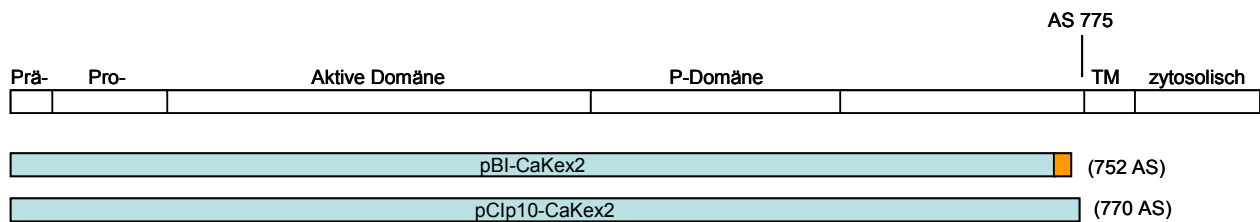


Abbildung 30: Vektorkonstrukte zur Expression von CaKex2 in *C. albicans*.

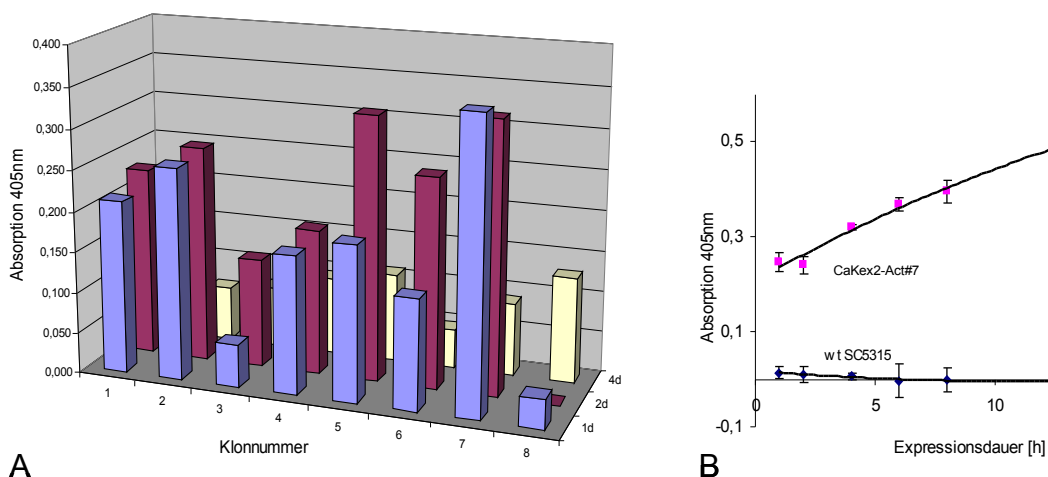
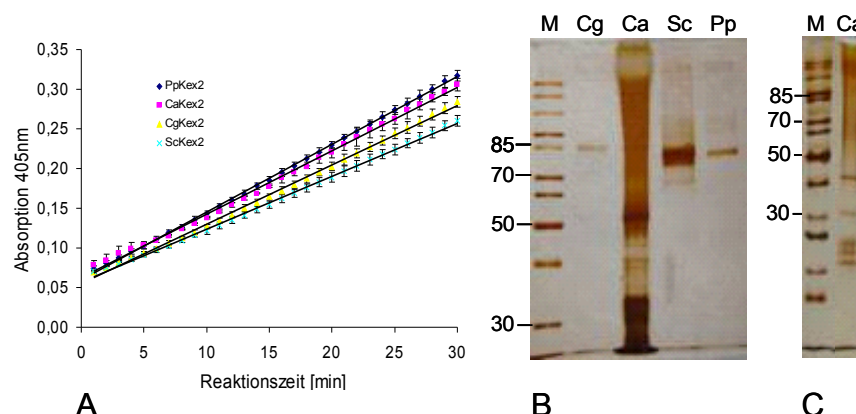


Abbildung 31: Expression von CaKex2. A: CaKex2-Aktivität in Kulturüberständen von Testexpressionen der acht Transformanden nach 1, 2 und 4 Tagen. B: Im Vergleich zu Klon #7 zeigten Kulturüberstände des Wildtypstamms SC5314 diese Aktivität nicht.

#### 4.4.2.3 Aufreinigung heterolog exprimierter Kex2 Proteasen aus *S. cerevisiae*, *P. pastoris*, *C. glabrata* und *C. albicans*

Zur Aufreinigung von sekretorischen Kex2-Proteasen standen vier Stämme zur Verfügung: die beiden in dieser Arbeit hergestellten Stämme Pp-CgKex2#12 und Ca-CaKex2#7 und die zwei Stämme Pp-PpKex2 [134] und Pp-ScKex2 [131]. Die Proteine CgKex2 und PpKex2 konnten über das C-terminale 6xHIS-tag mittels Ni-NTA-Affinitätschromatographie aufgereinigt werden. Das Protein ScKex2 wurde in großer Menge sezerniert, so dass es bereits im Kulturüberstand in hohen Konzentrationen und fast in Reine vorlag.

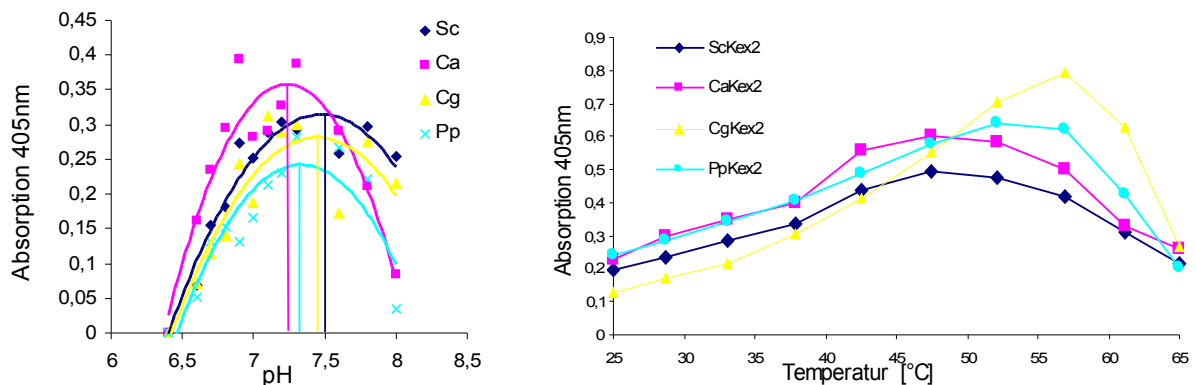
Die Aufreinigung des CaKex2 Proteins war dagegen problematisch. Da kein 6xHIS-tag vorhanden war, konnte das Enzym nicht über Ni-NTA-Agarose aufgereinigt werden. Im Gegensatz zu *P. pastoris* sezerniert *C. albicans* eine Vielzahl von Proteinen in den Kulturüberstand. Auch enthält das Kulturmedium YPD hohe Mengen niedermolekularer Stoffe, die bei der Konzentration der Proteine durch Ausschlusschromatographie ebenfalls angereichert werden und in größeren Mengen Signale bei späteren Versuchen mit Kex2-Substraten überdecken könnten (Abbildung 32A, Spur 3). Daher wurde das Kulturmedium vor der Inokulation über eine 30 kDa Ausschlusschromatographie von Kontaminationen >30 kDa größtenteils befreit. Durch eine an die Expression anschließende 50 kDa Ausschluss-Chromatographie konnte eine Enzym-Präparation hergestellt werden, die soweit von Kontaminationen befreit war, dass sie in Enzymreaktionen einsetzbar war (Abbildung 32B, Spur 2). Die Enzympräparationen wurden so verdünnt, dass sie alle die gleiche Aktivität gegenüber dem synthetischen Substrat KVKR-pNPP bei 37°C besaßen (Abbildung 32C).



**Abbildung 32: Kex2 Enzympräparationen.** A: Aktivität der eingestellten Proteasen gegenüber dem chromogenen Substrat KVKR-pNPP. B: Kex2 Enzympräparationen in der Silberfärbung. Die Präparation des *C. albicans* Enzyms CaKex2 ist stark verunreinigt. C: *C. albicans* CaKex2 Präparation aus vorweg gereinigtem Medium.

#### 4.4.3 Kex2 ist bei unphysiologisch hohen Temperaturen am aktivsten

Für ScKex2 liegen die optimalen Reaktionsbedingungen bei pH 7,2 [135]. Versuche mit ScKex2 wurden in der Literatur bei 25°C durchgeführt, was allerdings nicht der Umgebungstemperatur der humanpathogenen Pilze *C. albicans* und *C. glabrata* unter Infektionsbedingungen entspricht. Daher wurde überprüft, ob die Proteasen unter gleichen Versuchsbedingungen ähnliche Aktivitäten aufweisen. Getestet wurden pH-Wert und Reaktionstemperatur. Die Aktivität gegenüber dem Substrat KVKR-pNA war bei allen vier Proteasen bei pH 7,3-7,5 (Abbildung 33A) am stärksten und damit ähnlich wie für ScKex2 bereits beschrieben [135]. Auch das Temperaturoptimum war bei allen vier Protease vergleichbar, allerdings lag es nicht, wie angenommen, in einem Bereich um 30°C, welches einer idealen Umgebungstemperatur unter Kulturbedingungen entspräche, sondern bei unphysiologisch hohen Temperaturen um 50°C (Abbildung 33B). Daher konnte in den Versuchen für alle Proteasen die gleiche Reaktionstemperatur verwendet werden.



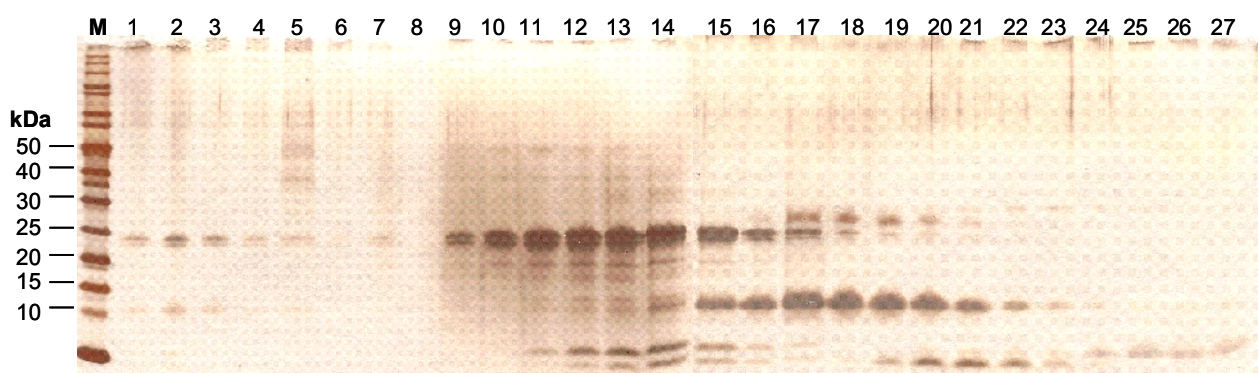
**Abbildung 33: pH- und Temperaturabhängigkeit der enzymatischen Aktivität von Kex2 aus den untersuchten Pilzen.** Die mittels Messung des Umsatzes von KVKR-pNA gemessene Kex2-Aktivität war bei allen Kex2 Enzymen bei einem pH-Wert von 7,3-7,5 und Temperaturen im Bereich von 45°C (*S. cerevisiae* und *C. albicans* Kex2) bis 55°C (*C. glabrata* und *P. pastoris*) am höchsten.

#### 4.4.4 Expression von Substratproteinen in *E. coli*

Zur heterologen Expression in *E. coli* wurden repräsentative Proteine aus der Gesamtmenge der vorhergesagten Substratproteine ausgewählt und die für den reifen Teil des Proteins (ohne Signalpeptid- und GPI-Ankersequenzen) kodierenden Sequenzen in den Expressionsvektor TOPO pET-D100 kloniert (Übersicht Tabelle 36). Es wurden Expressionskonstrukte für 28 Proteine aus *C. albicans*, 14 aus *C. glabrata* und zwei aus *S. cerevisiae* hergestellt.

In anschließenden Expressionsversuchen konnten bei 28 der insgesamt 44 Expressionsvektoren (63%) sehr schwache bis sehr starke Signale im Westernblot detektiert werden. Von diesen 28 Proteinen wurden 16 (57%) in einer löslichen Form exprimiert, die restlichen 12 (43%) in einer unlöslichen Form als Inklusionskörper. In löslicher Form exprimierte Proteine wurden direkt über Ni-NTA-Affinitäts-Chromatographie aufgereinigt. Die Aufreinigung von Proteinen aus Inklusionskörpern wurde durch Renaturierung aus mit 8 M Harnstoff denaturierten Lösungen ebenfalls auf Ni-NTA-Agarose erreicht.

Teilweise wurden die Proteine in *E. coli* bereits abgebaut, insbesondere solche, die Hormonvorläuferproteinen ähnlich sind (CA0365 und CaEce1). Diese konnten mittels Ni-NTA-Chromatographie nicht separiert werden, da sie teilweise das N-terminale 6xHis-tag enthielten. Daher wurden diese Proteine mittels Gelfiltration abgetrennt (Abbildung 34).



**Abbildung 34: Gelfiltration zur Aufreinigung von Ca0365.** Einige Proteine konnten mittels Ni-NTA-Chromatographie nicht genügend aufgereinigt werden, da die Abbauprodukte ebenfalls das N-terminale 6xHis-tag enthielten. Diese wurden mittels Gelfiltration entfernt. Im abgebildeten Beispiel für das Protein CA0365 wurden die Fraktionen 9-13 vereinigt und erneut mittels Gelfiltration gereinigt, was zu einer reinen Präparation führte.

Nr. <sup>9</sup>	Organismus	Name	kDa	Expressionsbedingungen	Renaturierungsbedingungen
01	<i>C. albicans</i>	CA0365	21	30°C, Laktose	löslich
02	<i>C. albicans</i>	CA1402/Ece1	31	30°C, Laktose	löslich
04	<i>C. albicans</i>	CA1873/OPS4-like 1	39	30°C, Laktose	löslich
05	<i>C. albicans</i>	CA2412/MF $\alpha$	18	30°C, Laktose	löslich
06	<i>C. albicans</i>	CA2974/OPS4-like 2	37	30°C, Laktose	löslich
07	<i>C. albicans</i>	CA3713/OPS4	44	nicht exprimiert	-
08	<i>C. albicans</i>	CA3943	48	30°C, Laktose	IK, nicht renaturierbar
09	<i>C. albicans</i>	CA5814/Subtilisin	51	nicht exprimiert	-
10	<i>C. albicans</i>	CA2303/Tos1	51	25°C, Laktose	IK, schlecht renaturierbar
11	<i>C. albicans</i>	CA0883/Sun41	45	30°C, IPTG	IK, Cyclodextrin
12	<i>C. albicans</i>	CA2942/Ccw14	33	30°C, IPTG	IK, Cyclodextrin
13	<i>C. albicans</i>	CA5147/Pho114	54	30°C, IPTG	IK, Cyclodextrin
14	<i>C. albicans</i>	CA1583/Rot1	28	nicht exprimiert	-
15	<i>C. albicans</i>	CA0212	29	nicht exprimiert	-
16	<i>C. albicans</i>	CA2470/Sap9	58	degradiert	-
17	<i>S. cerevisiae</i>	YDR144c/Yps3	63	nicht exprimiert	-
18	<i>C. albicans</i>	CA3867/Phr1	56	nicht exprimiert	-
19	<i>S. cerevisiae</i>	YMR307c/Gas1	56	nicht exprimiert	-
20	<i>C. albicans</i>	CA4679/Pga17	58	25°C, Laktose	teilweise löslich
21	<i>C. albicans</i>	CA0104/Rbt4	39	25°C, Laktose	IK, Cyclodextrin
22	<i>C. albicans</i>	CA0811/Pir-like	39	30°C, Laktose	IK, nicht renaturierbar
23	<i>C. albicans</i>	CA0375/Crh1	21	30°C, Laktose	löslich
24	<i>C. albicans</i>	CA2825/Hwp1	65	nicht exprimiert	-
25	<i>C. albicans</i>	CA6006	18	nicht exprimiert	-
26	<i>C. albicans</i>	CA1621/IFF6	103	nicht exprimiert	-
27	<i>C. albicans</i>	CA1394	49	Glukose, 30°C, IPTG	löslich
28	<i>C. albicans</i>	CA0416	25	30°C, Laktose	wenig IK, nicht renaturierbar
29	<i>C. albicans</i>	CA0195/Kar2	76	nicht exprimiert	-
30	<i>C. glabrata</i>	CAGL0H03135/MF $\alpha$	19	30°C, Laktose	löslich
31	<i>C. glabrata</i>	CAGL0J11770/Plb	73	30°C, Laktose	IK, Cyclodextrin
32	<i>C. glabrata</i>	CAGL0M13805/Scw4	40	30°C, Laktose	löslich
33	<i>C. glabrata</i>	CAGL0L05434/Sun4	38	30°C, Laktose	löslich
34	<i>C. glabrata</i>	CAGL0H08910g	41	30°C, Laktose	IK, Cyclodextrin
35	<i>C. glabrata</i>	CAGL0I04092g	65	nicht exprimiert	-
36	<i>C. glabrata</i>	CAGL0D02530/Egt1	70	30°C, Laktose	IK, Cyclodextrin
37	<i>C. glabrata</i>	CAGL0L02607	68	nicht exprimiert	-
38	<i>C. glabrata</i>	CAGL0M08492/Pir1	36	30°C, Laktose	löslich
39	<i>C. glabrata</i>	CAGL0A02277	31	30°C, Laktose	IK, Cyclodextrin
40	<i>C. glabrata</i>	CAGL0F05137/Pry1	27	30°C, Laktose	IK, Cyclodextrin
41	<i>C. glabrata</i>	CAGL0G07667/Pry2	29	30°C, Laktose	IK, schlecht renaturierbar
42	<i>C. glabrata</i>	CAGL0L10802/Subtilisin	59	nicht exprimiert	-
43	<i>S. cerevisiae</i>	YPL187W/MF $\alpha$	20	30°C, Laktose	löslich
44	<i>C. glabrata</i>	CAGL0D02948/Kar2	74	30°C, Laktose	löslich
45	<i>S. cerevisiae</i>	YJL034W/Kar2 <sup>10</sup>	72	30°C, Laktose	löslich

**Tabelle 36: Übersicht exprimierter Substratproteine und Expressionsbedingungen.** Substratproteine wurden in *E. coli* bei 30°C oder 37°C durch Zugabe von IPTG oder Laktose exprimiert. Proteine, die in Form von Inklusionskörpern (IK) vorlagen, wurden ggf. durch Cyclodextrin renaturiert.

<sup>9</sup>Substrat #3 entfällt.

<sup>10</sup>Plasmid #45 wurde freundlicher Weise von Jeff Brodsky zur Verfügung gestellt.

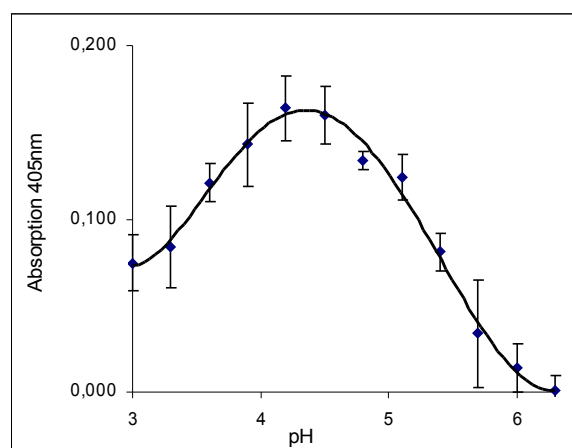


#### 4.4.5 Durch Renaturierung lassen sich nativ gefaltete, enzymatisch aktive Proteine herstellen

Die Renaturierung von durch Harnstoff denaturierten Proteinen ist ein statistischer Prozess, der nicht unbedingt zu einem korrekt gefalteten Protein führen muss. Unter den hier bearbeiteten Proteinen befand sich nur eines, welches starke Homologien zu bekannten Enzymen mit einer nachweisbaren Aktivität besaß und somit als Kontrolle für eine korrekte Faltung in die aktive Form dienen konnte. Dabei handelte es sich um Substrat #13 (CA5147, CaPho114), welches eine 42%ige Sequenzidentität zu *S. cerevisiae* Pho11, einer im sauren pH aktiven Phosphatase, aufwies.

In der Tat konnte gezeigt werden, dass auch CaPho114 diese Aktivität aufweist: Präparationen des renaturierten Proteins waren in der Lage, das Substrat *para*-Nitrophenolphosphat im sauren pH-Bereich zu dem photometrisch nachweisbaren Stoff *para*-Nitrophenol umzusetzen. Die maximale Aktivität lag dabei im pH-Bereich von 4,3-4,5 (Abbildung 35).

Daraus ließ sich folgern, dass die Renaturierung von denaturierten Proteinen, zumindest zum Teil, zu nativ gefalteten und aktiven Proteinen führte. Des Weiteren konnte festgestellt werden, dass für die Aktivität dieses Proteins offensichtlich keine Prozessierung durch Kex2 nötig war, obwohl das Protein eine potentielle Schnittstelle besitzt.

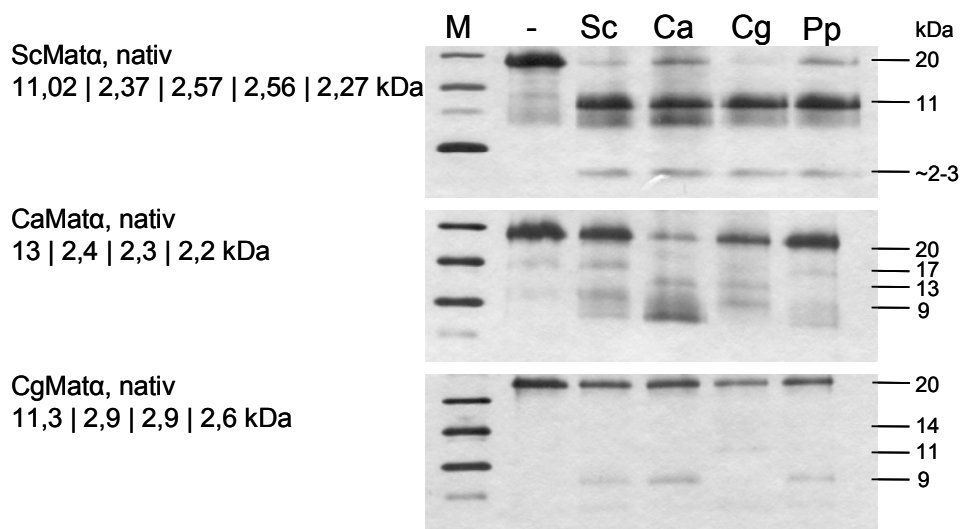


**Abbildung 35: Aktivitätstest der sauren Phosphatase CaPho114 (Substrat#13) aus *C. albicans*.** CaPho114 zeigt maximale Aktivität gegenüber dem Substrat pNPP im Bereich von pH= 4,4-4,5.

#### 4.4.6 *In vitro* Prozessierung der putativen Kex2-Substrate

##### 4.4.6.1 Die $\alpha$ -Pheromone von *S. cerevisiae*, *C. glabrata* und *C. albicans* werden in ähnlicher Weise prozessiert

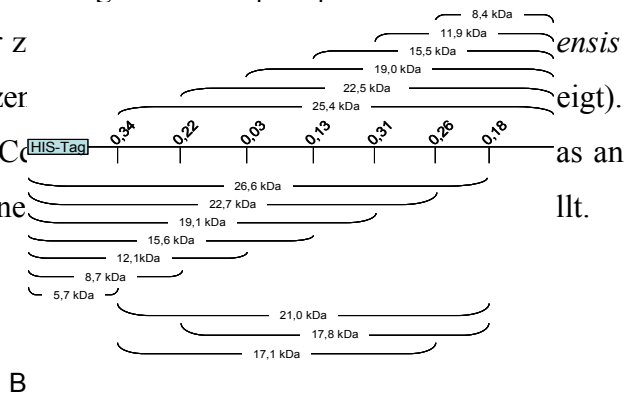
Neben der Prozessierung von Killertoxinvorläuferproteinen gehört die Prozessierung der  $\alpha$ -Pheromone zu den bestuntersuchten biologischen Prozessen, in die Kex2-Enzyme involviert sind. Daher wurde zunächst die Prozessierung der rekombinanten  $\alpha$ -Pheromone durch die verschiedenen Proteasen untersucht. Zur Verfügung standen die Vorläuferproteine für die  $\alpha$ -Pheromone aus *S. cerevisiae* (Substrat #44), *C. albicans* (Substrat #5) und *C. glabrata* (Substrat #30). Zunächst konnte beobachtet werden, dass alle Proteine von allen Proteasen geschnitten werden und dass die der Vorhersage entsprechenden Zwischenprodukte in Erscheinung treten. Des Weiteren war zu beobachten, dass die Proteasen das jeweils aus ihrem Hintergrund stammende Pheromonvorläuferprotein am besten prozessieren.



**Abbildung 36: Prozessierung der  $\alpha$ -Pheromone von *S. cerevisiae*, *C. glabrata* und *C. albicans*.** Alle rekombinanten Pheromonvorläuferproteine werden von allen Kex2-Enzymen geschnitten, allerdings mit unterschiedlicher Effizienz: das aus dem gleichen Organismus, wie das Substrat, stammende Enzym schneidet immer am effektivsten.

4.4.6.2 CaEce1 ist ein Polypeptidvorläufer

Das bei der Induktion von Hyphenwachstum bei *C. albicans* am stärksten hochregulierteste Gen ist *CaECE1*. Es wurde bereits 1993 beschrieben und war das erste mit der Ura-Blaster-Methode deletierte Gen [136]. Trotz der starken Expression konnte für die *Caece1*-Mutante jedoch kein Phänotyp beobachtet werden. Bei der Betrachtung der von *CaECE1* kodierten Proteinsequenz fällt auf, dass sie acht degenerierte Peptidwiederholungen enthält [136]. *CaEce1* besitzt keine Sequenzhomologien zu anderen Proteinen außer zu *CdEce1* (Abbildung 37). Andere *Candida* Spezies besitzen das zusätzliche KR-Motiv an Position 253 bei *CdEce1* an der entsprechenden Position in *CaEce1* vorhandene



```

1 (SP) -----IHHAPE-----EFN-MKR/
2----DVAPAAPAAPADQAP-----TVPAPQEFNTAITKR/
3----SIIGIIMGILGNIP--QVIQIIMSIVKAFKGNKR/
4----EDIDSVVAGI IADMP--FVVRVAVDTAMTSVASTKR/
5----DGANDDVANAVVRLPEI VARVATGVQQSIENAKR/
6----DGVDPVGLNLVANAPRLISNVFDGVSETVQQAKR/
7----DGLDFLDELLQRLPQLITR SAESALKDSQPVK R/
8----DAGSVALSNLIKKS IETVGIENAAQIV-S---ER/
9----DISSLIEEYFGKA*
    
```

A Abbildung nach Birse *et al.*, 1993 [136]

```

CdEce1 MKFSKIACAAAFVLSSQAAIIHHAPEFNMKRD IATADPVV---AEPDVSAANVDTSIAKRSIIGILTAILNNVPQIINVITTI I
CaEce1 MKFSKIACATVFALSSQAAIIHHAPEFNMKRDVAPAAPAAPADQAPTVPAPQEFNTAITKRSIIGIIMGILGNIPQVIQIIMSIV
*****:.*.*****:.*.* *.. * *... :.:*:*:*****: .**.*:**:.*.:*:*

CdEce1 KSITGNKREDIDSVVSGVIADMPFVVRVAVDTALTSVASTKRDGTNDVDVTAIVRLPEI VAPVATGVQQTVENAKRDGVEDIGLNI
CaEce1 KAFKGNKREDIDSVVAGI IADMPFVVRVAVDTAMTSVASTKRDGANDDVANAVVRLPEI VARVATGVQQSIENAKRDGVPDVG LNL
*.:.*****:.*:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:

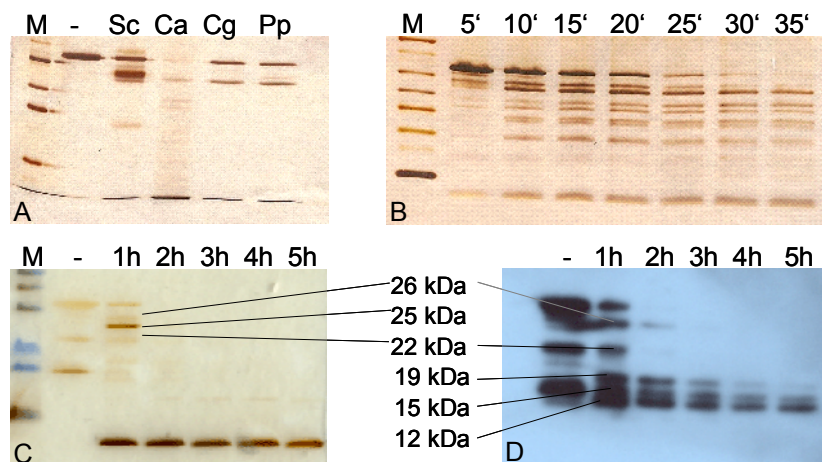
CdEce1 VANSPLRVSDVIGGVLETVKQAKRDNAEDILTKVLQELPDIISKVSNSTLKNSPFFKRDANTVTLSKLIKKS IETIGVENAAKMV
CaEce1 VANAPRLISNVFDGVSETVQQAKRDGLEDFLDELLQRLPQLITR SAESALKDSQPVKRDAGSVALSNLIKKS IETVGIENAAQIV
***:**:.*:.* **:*:*****: **:* ::*: **:*:.*: :.*:*:* * **:*:.*:**:*****:.*:*****:.*

CdEce1 SKRDISSLIAEYFEEA
CaEce1 SERDISSLIEEYFGKA
*:*:*****:*****.*
    
```

C

Abbildung 37: Aufbau des Proteins *CaEce1*. A: *CaECE1* kodiert für ein Protein mit sieben ca. 34 Aminosäuren langen, degenerierten Wiederholungen, welche von potentiellen Kex2 Schnittstellen begrenzt werden. B: Schematische Darstellung des rekombinanten *CaEce1* Proteins und der theoretischen Fragmentlängen nach Prozessierung an den potentiellen Kex2-Schnittstellen. C: *CaEce1* weist nur ein einziges orthologes auf, welches aus *C. dubliniensis* stammt (*CdEce1*).

In wiederholten Experimenten konnte gezeigt werden, dass CaEce1 tatsächlich von Kex2 Proteasen gespalten wurde. Dabei war, wie bei den Pheromonen (Abschnitt 4.4.6.1), die Protease des Herkunftsorganismus (hier CaKex2) am aktivsten (Abbildung 38A). Ein Zeitreihenexperiment zeigte, dass das Protein bevorzugt an den Enden prozessiert wurde, da anfänglich genau diese Intermediate entstanden, nicht aber solche, die auf eine Prozessierung in der Mitte schließen ließen (Abbildung 38B). Diese Zuordnung der N-terminalen Fragmente konnte im Westernblot bestätigt werden (Abbildung 38C).

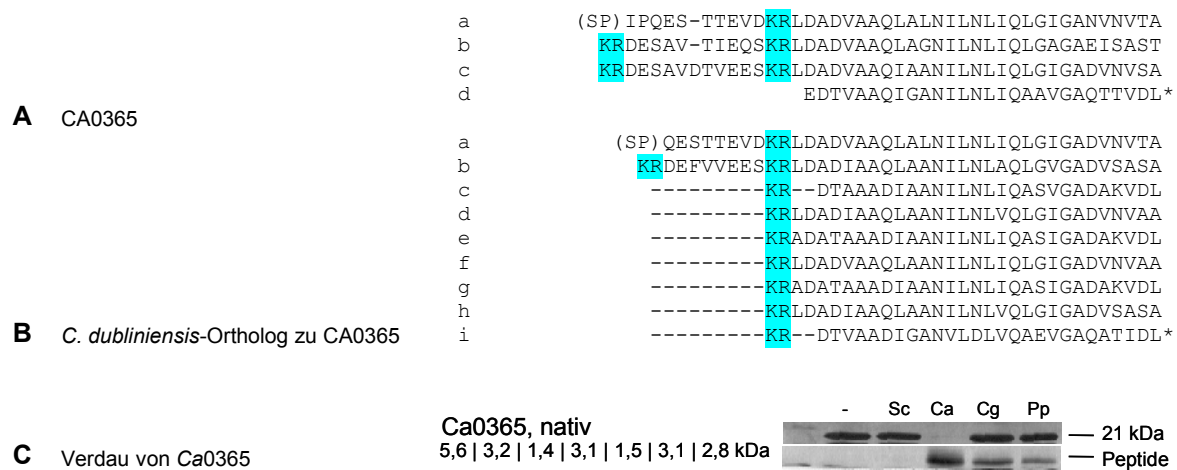


**Abbildung 38: Prozessierung von CaEce1.** A: CaEce1 wurde von allen vier Proteasen geschnitten, von CaKex2 am effizientesten. B: Ein Zeitreihenexperiment zeigte, dass genau die Fragmente entstanden, die bei einer Prozessierung von den Enden her erwartet wurden. C, D: Im Westernblot konnten die Fragmente mit intaktem N-terminalen 6xHis-tag den entsprechenden Banden in der SDS-PAGE zugeordnet werden.

#### 4.4.6.3 CA0365 wird von den Proteasen unterschiedlich prozessiert

Auch das von ORF CA0365 kodierte Protein zeigt, wie CaEce1, die Struktur eines Prohormonvorläufers. Vier konservierte Wiederholungen werden von potentiellen Kex2-Schnittstellen flankiert, die ersten drei Wiederholungen beinhalten dabei eine weitere Schnittstelle. Die vollständige Prozessierung würde so zu drei kurzen (~1,5 kDa) und vier längeren (~3 kDa) Peptiden führen (Abbildung 39A). Der Verdau mit den vier Kex2 Proteasen führte tatsächlich, ohne nennenswerte Bildung von Intermediaten, zu peptidgroßen Abbauprodukten (Abbildung 39C). Wiederum wurde das Polypeptid am effizientesten von der Protease des Herkunftsorganismus (CaKex2) prozessiert. Im Gegensatz zu den vorigen Proteinen wurde CA0365 jedoch gar nicht von ScKex2 prozessiert. Dies weist darauf hin, dass es im Einzelfall zwischen den Proteasen gravierende Unterschiede in der Substraterkennung geben muss.

Wie CaEce1 besitzt auch CA0365 keine Sequenzhomologien zu anderen Proteinen außer zu genau einem Ortholog bei *C. dubliniensis*. Dieses kodiert allerdings für insgesamt neun Wiederholungen, wovon fünf denen von CA0365 entsprechen. Die anderen vier sind leicht unterschiedlich, in sich aber wieder konserviert (Abbildung 39B).



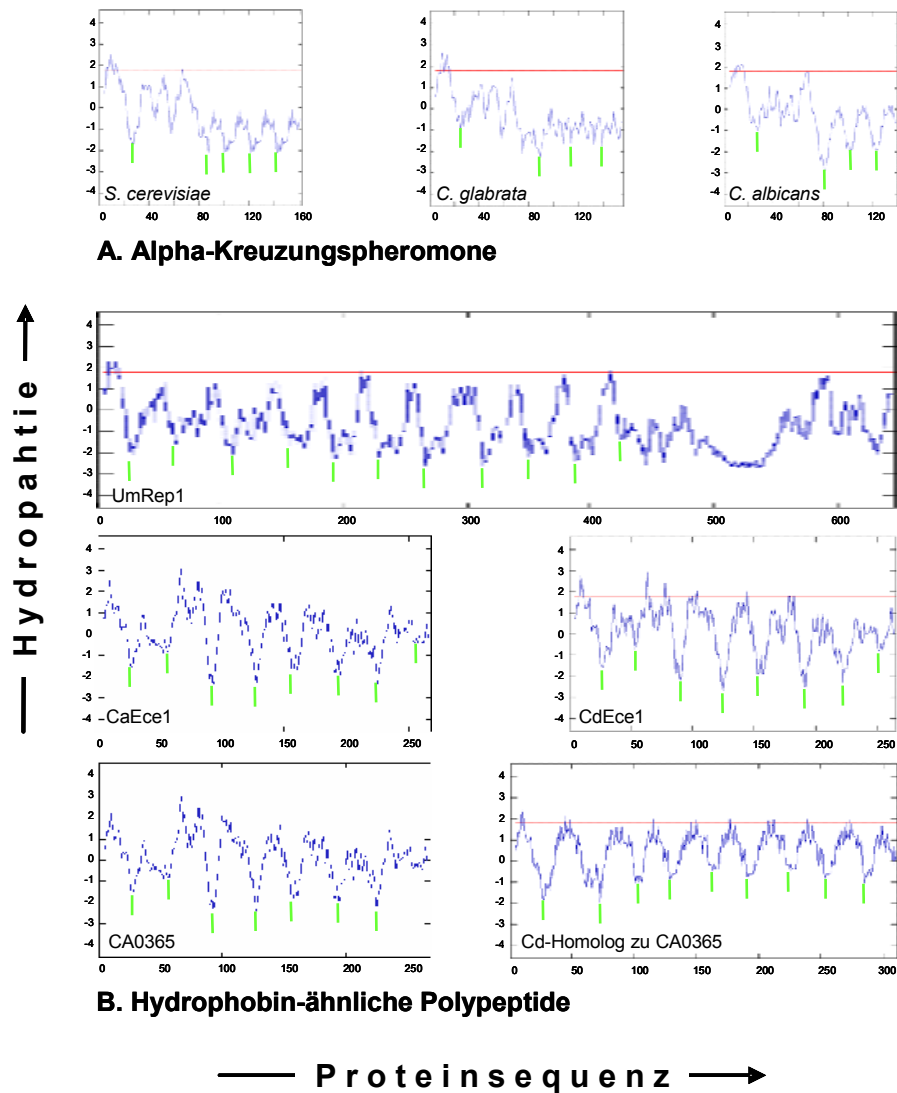
**Abbildung 39: Aufbau und Prozessierung von CA0365.** A: Polypeptidstruktur des Proteins: vier konservierte Wiederholungen werden von potentiellen Kex2-Schnittstellen flankiert. B: Das orthologe Protein zu CA0365 aus *C. dubliniensis* kodiert für neun durch Kex2-Schnittstellen begrenzte konservierte Peptide. C: CA0365 wurde von CaKex2 am schnellsten zu Peptiden prozessiert, von ScKex2 dagegen nicht.

#### 4.4.6.4 CA0365 und Ece1 könnten Strukturmoleküle ähnlich UmRep1 sein

CA0365, CaEce1 und die beiden entsprechenden Orthologen Proteine aus *C. dubliniensis* weisen keine längeren Sequenzhomologien zu weiteren bekannten Proteinen auf. Das einzige bekannte Protein mit einem von Kex2 prozessierten ähnlichen Aufbau ist das Hydrophobin UmRep1 aus *Ustilago maydis*. UmRep1 kodiert für zehn konservierte Wiederholungen (Abbildung 40) und eine längere C-terminale Domäne [112]. Es handelt sich hierbei um sekretorische Peptide, die in Lufthyphen in der Zellwand verankert werden und der Oberfläche hydrophobe Eigenschaften vermitteln. Eine Analyse des Hydropathieverlaufs von UmRep1, den Ece1-Proteinen, CA0365 und den ebenfalls aus mehreren Wiederholungen kurzer Peptide bestehenden  $\alpha$ -Kreuzungspheromonen zeigt, dass (Abbildung 41).

N-term	(SP)-
a	KR----PHYEY-KSYNAGHNVESVVENKLVDSADLTGVDIL
b	KRTDYSACKKYVSSYNAGYNVYSINENKLIDLSDATVKLSLLTNGRPVKTWNKA
c	PR----YCLEYIKSYNGGHNVIIDNENKLIDLSGLNLGIDIL
d	KR----KAPEF-SYYPNGHNVKSTTENKLVDSADLLANVNILS
e	KR----TGSQY-SYNGGHNVESTNENKLVDSADLLANVDILK
f	KR----GDKY-SYNGGYNVESDTENKLIIDLSGLLANIDILK
g	KR----GDKY-TYNGGHNIESTNENKLIDLSDLTANVGVLDVHVNIIGLLD
h	KR----TDKY-SYDGGHNIELTNEKLIIDLSADLLANVDVLK
i	KR----TDKF-SYNGGHNVESDTQNKLIIDVSHLTAIVNALS
j	KR----TDKY-SYNGGHNVESTNENKLIDLSADLTALVNVLS
C-term	KR-(längeres C-terminales Fragment)

**Abbildung 40: Aufbau des Hydrophobins Rep1 aus *Ustilago maydis*.** UmRep1 kodiert für ein Protein aus zehn wiederholten, konservierten Peptiden, die durch Kex2-Schnittstellen getrennt sind, und ein ca. 200 AS langes C-terminales Fragment.

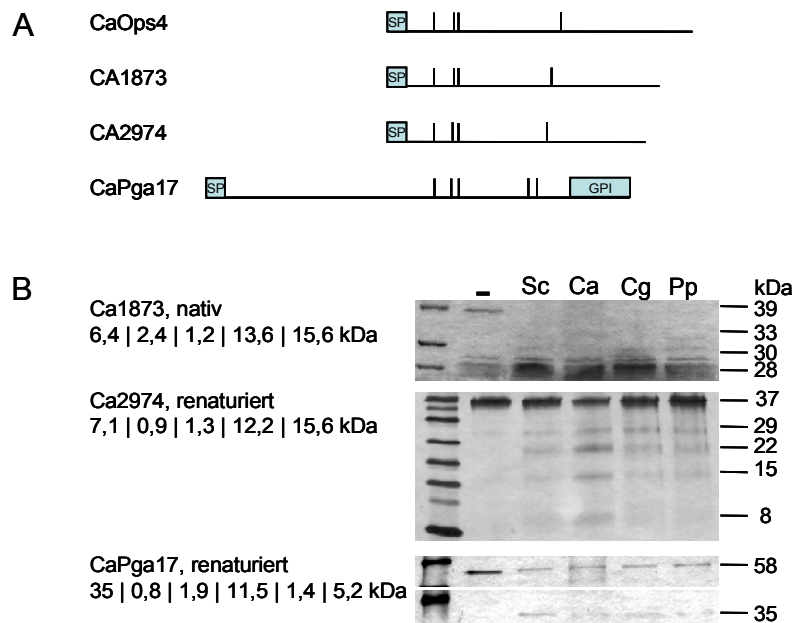


**Abbildung 41: Hydropathieanalyse von Polypeptidvorläufern.** Die Hydropathieanalyse identifiziert hydrophobe (Maxima) und hydrophile (Minima) Bereiche in Proteinsequenzen. Die Bereiche zwischen den stark geladenen Schnittstellen (grün) von Kex2 und Signalpeptidase (die jeweils erste Schnittstelle) sind in Pheromonvorläufern (A) weniger hydrophob als in Rep1, Ece1 und CA0365 (B). Keine der Domänen überschreitet die minimale Hydrophobizität (rote Linie) für eine Transmembrandomäne.

#### 4.4.6.5 Proteine mit Ops4-ähnlicher Struktur werden an mehreren Stellen prozessiert

Eine Gruppe von Proteinen, welche eine ähnliche Verteilung von potentiellen Schnittstellen trägt, besteht aus CaOps4, den zwei zu Ops4 homologen Proteinen CA1873 und CA2974, und dem GPI-verankerten Protein Pga17. Diese Proteine tragen N-terminal eine Gruppe aus zwei oder drei potentiellen Schnittstellen und eine weitere Schnittstelle in der Mitte. Bei Pga17 ist die gesamte Anordnung zum C-Terminus hin verschoben (Abbildung 42).

Für die Proteine CA1873, CA2974 und Pga17 konnte eine Prozessierung an diesen Schnittstellen gezeigt werden. Ops4 konnte in *E. coli* nicht exprimiert werden. Bei Pga17 konnte lediglich der die Prozessierung in der Gruppe der ersten drei Schnittstellen festgestellt werden.



**Abbildung 42: Aufbau und Prozessierung von *C. albicans* Proteinen mit Ops4-ähnlicher Struktur.** A: Proteine dieser Gruppe besitzen drei N-terminale und zwei um ca. 12 kDa weiter zum C-Terminus gelegene Schnittstellen. Bei Pga17 ist diese Anordnung insgesamt zum C-Terminus verschoben, zusätzlich besitzt es eine GPI-Verankerungssequenz. B: Die Ops4-ähnlichen Proteine werden von Kex2 in die erwarteten Fragmente zerschnitten.

Ein Vergleich der Proteinsequenzen zeigt nicht nur eine ähnliche Anordnung der Kex2-Schnittstellen, sondern auch weitere starke Sequenzähnlichkeiten. Der zwischen beiden Gruppen von Kex2-Schnittstellen gelegene mittlere Teil der Proteine trägt potentielle N-Glykosylierungsstellen, während der C-terminale Teil mögliche O-Glykosylierungsstellen aufweist. Pga17 besitzt ein C-terminales GPI-Verankerungssignal [120]. Bei den drei anderen Proteinen ist der C-Terminus auffällig positiv geladenen und endet in einem konservierten Tyrosin (Abbildung 43).

```

CaOPS4      MKFSOATILAI FASSALVSAAP-----ANIVSEQTMVKREDVNAIVELINEIKHINQKRD
Ca1873      MKLSTTTALAVLATATAINAAPQQQQQESQFASKALTKREEQDIQELVQHINNYKTRRD
Ca2479      MKLDFATATLVLAAISSVAAAP-----AADAQVKVTQLTVRESNLVNSALANLQHYNAKRD
CaPGA17     MKFSTVFTAIFALGTA VSAQOV (.....) NYSALSQISGALRREYNDAVEAALREIQKREEG

CaOPS4      LAEGEDLLELQRRADSVIGELVSALYNSGVIGLVWDKLTDPSSISLNSNIKSAVQGA I
Ca1873      AID-EEIM---KRDYAIVTDVLAAIQSQ L APKILDYLVSETFKPIVVVTIATMKSGI
Ca2479      LMSQEEII---KRENQIVTDVLTAIKNTNLSPGI IKYLITDPTLSKISVDVIVGAIKGT
CaPGA17     IDDVEIALRKMKRDNIEDLLTTIFASVARSNLLTTIQYLVTDQRFESTVVELLQGVFEN

CaOPS4      VQG G ALIQAVWNSGLLDVFNKLI DTLRQALLDVGKALFN SAANLISIWLGSSSSSS
Ca1873      ISLQALLDALVSSNLAVNLVNDLISDCSLYVELFNAAKSVIN---DLASKVKGLISKGVS
Ca2479      I LTTLLKSL DSGLAVDLIKDLINDCAFYAYIYKAILDKLTSLPNLIGNALGLNANTVS
CaPGA17     IGSTLTGILD TDWSALQPLVSSLLNSGLLTD FISR AFNDELKAVLWNDITSIFKDMAF

CaOPS4      AAPAAAA PATNGASKREIMEAAEYLSERDLASILSWIVQTIKDTGIVQSLVNQVINNPD
Ca1873      QLITRDETDIDALAPYVVTME-----KRLDL DGVVDNLLDSLYKSGLATSVVKDILTNSD
Ca2479      SKMASAKREIMVESAPEPIY-----ARDGQDVLTSLMESLKSSGLANQVVEALVIDDQ
CaPGA17     RDEIVKRS GTITISLPVDFITGVATETSALDGADG-----

CaOPS4      TVIS---FLTSALKNGLVIVEDLYDWAKQSGLWDQALVYIQNNAGSWIKALAGLFGNAL S
Ca1873      YIPFATDLIKAMIANNLIDLGNIVDAVKQSGLV TQLFQKLVN-----FG---T
Ca2479      FYKWGGDLIKELIDEKAITLGLIDALADSGLIPSLFQAF LN-----FG---T
CaPGA17     -----

CaOPS4      NGTITASDINNAGSSSKPTGTTTASTATAAPKAATPAAASA AVASKAATS GSSSSGSSNN
Ca1873      VQTV AETAFAAYAGECQSGAISGGSGSGSGSGSGSGSTITTTGGSSSGSSGSSSHG
Ca2479      LKSVIVTALAAAFGKQCN-----ATPTSSLKTTPTGTATVSIPTATGTS GPVTC-----
CaPGA17     -----TLSSLDVTA FINTVSHSFITSNASS-----

CaOPS4      ADLNALINKYGGGSGSTSTPTVDT SGLSSDVNTLVNAAGQAASSL-----KKRKLY*
Ca1873      SSAGPC-----KKRRRRRRRSNY*
Ca2479      -----KKKRRRNY*
CaPGA17     -----AVITIQSDNAGSSYGPGFYSTIFAVFGLFAMMI*

```

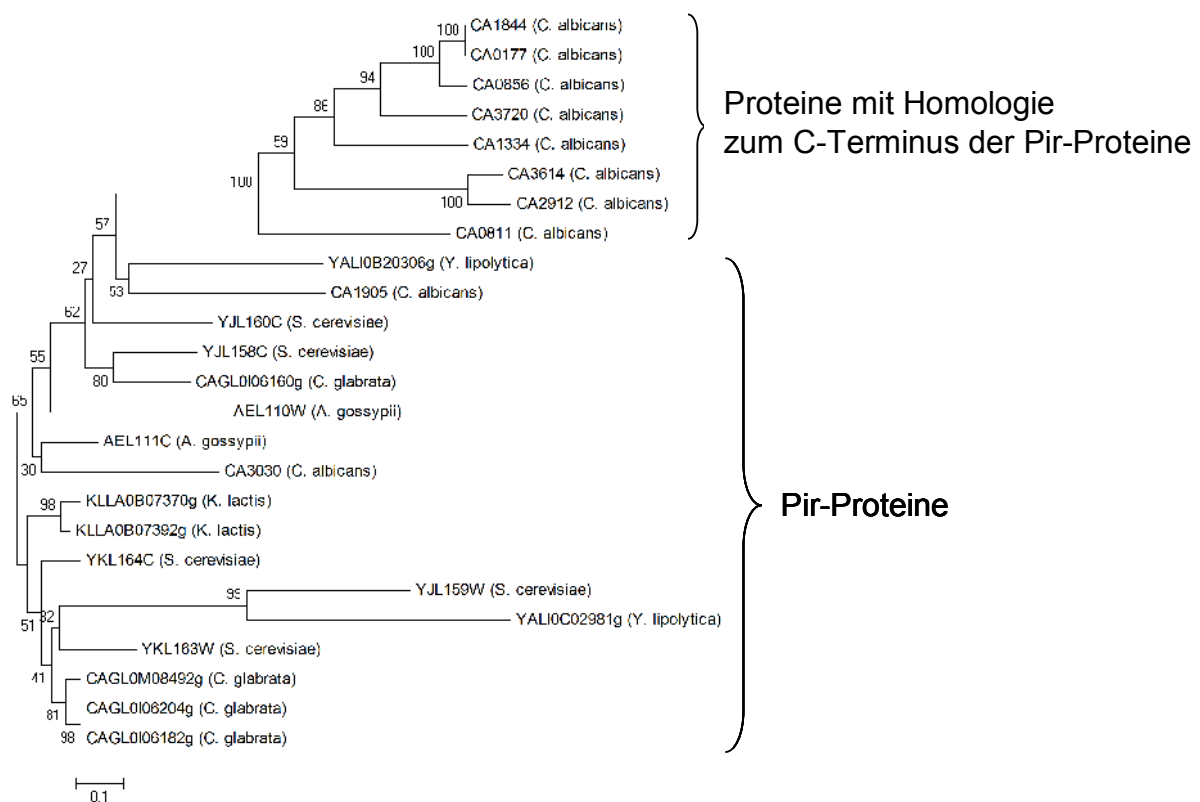
**Abbildung 43: Aufbau Ops4-homologer Proteine.** Sequenzalignment der Ops4-homologen Proteine CA2479, CA1873 und dem C-terminalen Teil von Pga17 mit Ops4. Blau: Kex2-Schnittstellen, rot: potentielle N-Glykosylierungsstellen, gelb: S/T-reiche Regionen (potentielle O-Glykosylierung), grün: potentielle C-terminale Lokalisationssignale, unterstrichen: Signalpeptid.



#### 4.4.6.6 Pir-Proteine werden von Kex2 prozessiert

Von Proteinen der Pir-Familie (Proteins with internal repeats, Abbildung 45) ist bekannt, dass sie in *kex2*-Deletionsmutanten anomal prozessiert werden [102, 103]. Zusätzlich zu den Pir-Proteinen gibt es bei *C. albicans* eine Gruppe zum C-Terminus der Proteine der Pir-Familie ähnlicher Proteine (Abbildung 44). Diese verfügen zwar nicht über die charakteristisch kurzen internen Wiederholungen, aber über eine im Vergleich zu den Pir-Proteinen weiter C-terminal gelegene potentielle Kex2-Schnittstelle. Da bisher nicht gezeigt wurde, ob Pir-Proteine von Kex2 oder einer anderen, aber selbst Kex2-abhängigen, Protease geschnitten werden, wurden zwei Pir-homologe Proteine in das Spektrum der zu untersuchenden Substrate aufgenommen (CA0811, Substrat #22, und CAGL0M08492, Substrat #38). Eines davon (CAGL0M08492) konnte teilweise aufgereinigt und für Versuche verwendet werden (Abbildung 46).

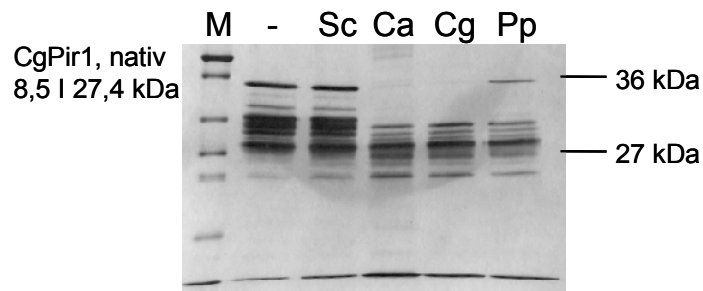
Während sich das von CA0811 kodierte Protein nicht in löslicher Form herstellen ließ, konnte CgPir1 in ausreichend sauberer Form aufgereinigt werden. Der Verdau mit den verschiedenen Kex2-Enzymen zeigte, dass es von allen geschnitten wurde (Abbildung 46). ScKex2 zeigte dabei eine wesentlich geringere Effizienz als die anderen Enzyme.



**Abbildung 44: Phylogenie der Pir-Proteine.** Innerhalb der Gruppe von *S. cerevisiae*, *C. glabrata*, *K. lactis* und *A. gossypii* besitzen die Organismen jeweils mehrere Pir-Protein. *C. albicans* besitzt nur ein Pir-Protein (Pir1, CA3030), dafür acht weitere, die hohe Sequenzähnlichkeiten zum C-Terminus der Pir-Proteine aufweisen.

Name	Sequenzmotiv mit Schnittstelle
CA1844 ( <i>C. albicans</i> )	--DDEGS--DCDDDDCKK-KKKAHR-----YAKR CGGGDD---
CA0177 ( <i>C. albicans</i> )	--DDEGS--DCDDDDCKK-KKKAHR-----YTKR CGGGDD---
CA0856 ( <i>C. albicans</i> )	SSDDEGSGSDCEDDECKKRRKKVHR-----NYKRGGYGSR--
CA3720 ( <i>C. albicans</i> )	--DDNC-----DDCKKKKKK-----VYFAKRGDDDD--
CA1334 ( <i>C. albicans</i> )	-----DDHHKKKKRPHRHGGKSDDDDDKKWKRGGDYSDND
CA3614 ( <i>C. albicans</i> )	YWDGGDDNDECEDEDCDEDEKKGHK-----QYKRGEVEEPC--
CA2912 ( <i>C. albicans</i> )	AVVGG--NEECEDEDCDEHRKKGSK-----QYKRGEVENPR--
CA0811 ( <i>C. albicans</i> )	SSALFNKGWSVV-----SSNGKRLLSLNGSTR
YALI0B20306g ( <i>Y. lipolytica</i> )	FGIAIQPITGAVQAG-----AEATAVPTAHKRDEHDD
YJL160C ( <i>S. cerevisiae</i> )	FGLAVIPFTVTESK-----VKRNVISQINDG
YJL158C ( <i>S. cerevisiae</i> )	FGIAVQA--ITSSK-----AKRDVISQIGDG
CAGL0I06160g ( <i>C. glabrata</i> )	FGIAVKP--ISSV-----AKRAVVSQIGDG
AEL110W ( <i>A. gossypii</i> )	FGIAVVPLTP-----RVFDGSKKRDLSAVEQIIDG
CA3030 ( <i>C. albicans</i> )	FGIQVVTVESASALSTDTATSTLT RNDNKKEATPVAQITDGGVQHQTGG
AEL111C ( <i>A. gossypii</i> )	FGIAVVPLATEKAS-----KAKRAVPVVSQIGDG
KLLA0B07370g ( <i>K. lactis</i> )	FGIAVIPITTSVSTKTSTTATS-----ATFTSTPKAKRDVAAISQIGDG
KLLA0B07392g ( <i>K. lactis</i> )	FGIAVIPITTSVSTVSTGTATSTSA--KTTATTTSKAKRDVAAISQIGDG
YKL164C ( <i>S. cerevisiae</i> )	FGIAVEPIATASSKA-----KRAAISQIGDG
YJL159W ( <i>S. cerevisiae</i> )	FGIAVQPIST-TSSASSAAT-----TASSKAKRAASQIGDG
YALI0C02981g ( <i>Y. lipolytica</i> )	SQINDGQIQAPKSTAAPAPA-----PKPETAAPVVSQINDG
YKL163W ( <i>S. cerevisiae</i> )	FGIAIEAVATSASSVA-----SSKAKRAASQIGDG
CAGL0M08492g ( <i>C. glabrata</i> )	FGIAVQPIATSS-----SVAKRAVSQIGDG
CAGL0I06204g ( <i>C. glabrata</i> )	FGIAVQPIATPA-----SVAKRAVSQIGDG
CAGL0I06182g ( <i>C. glabrata</i> )	FGIAVQPIATPA-----SVAKRAVSQIGDG

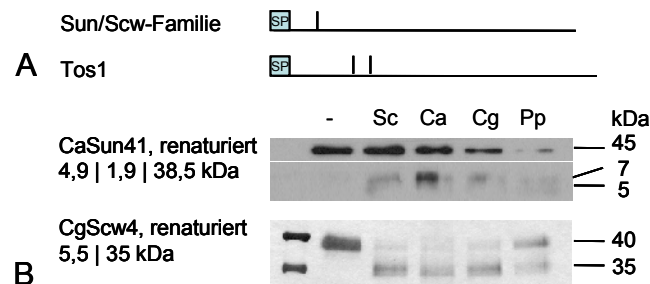
**Abbildung 45: Kex2-Schnittstellen in Pir- und Pir-ähnlichen Proteinen.** Sowohl die Pir-Proteine als auch die Gruppe der Pir-ähnlichen Proteine aus *C. albicans* besitzen potentielle Kex2-Schnittstellen. Diese Motive sind in allen Pir-homologen Proteinen konserviert.



**Abbildung 46: Prozessierung von CgPir1.** CgPir1 wird von allen Kex2 Enzymen geschnitten. ScKex2 zeigt dabei die niedrigste Effizienz.

#### 4.4.6.7 Proteine der Sun/Scw4 Familie, nicht aber renaturiertes Tos1, werden an einer konservierten histidinreichen Struktur prozessiert

Eine besonders auffällige Struktur mit potentiellen Kex2-Schnittstellen ist bei Proteinen der Sun/Scw-Familie und bei ScTos1 und seinen orthologen Proteinen vorhanden. Sie besteht aus einem KR-Motiv, welches N-terminal von einer Reihe positiv geladener Aminosäuren, meist Histidin, flankiert wird. Bei Tos1 befindet sich das Motiv, nicht wie bei Sun/Scw direkt am N-Terminus, sondern ca. 120 AS weit in das Protein hinein. Allerdings befindet sich direkt vor dem Motiv eine weitere potentielle Schnittstelle, so dass bei einer Prozessierung an dieser vorderen Schnittstelle die histidinreiche Struktur wieder am N-Terminus des Proteins gelegen wäre (Abbildung 47A). In einem Westernblot konnten für CaSun41 die N-terminal abgespaltenen Produkte von 5 und 7 kDa nachgewiesen werden. Ebenso konnte in einer SDS-PAGE eine Prozessierung bei CgScw4 gezeigt werden. Eine Prozessierung bei renaturiertem CaTos1 wurde hingegen nicht beobachtet (nicht gezeigt).



**Abbildung 47: Struktur und Prozessierung von Proteinen mit His-Box-Struktur.** A: Lokalisation der potentiellen Schnittstellen in Proteinen mit HIS-Box. B: CaSun41 und CgScw4 werden von Kex2 geschnitten.

#### 4.4.6.8 Manche Proteine der Pry-Proteinfamilie besitzen glykosylierte Propeptide

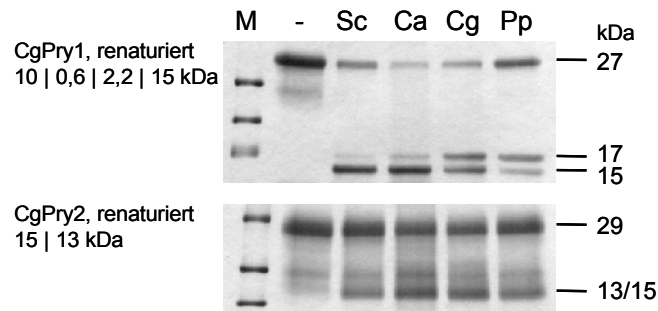
Die Familie der Pry-Proteine (Familie der PR-1-homologen Proteine) besteht aus 2-5 Mitgliedern pro untersuchter Spezies. Sie kommt in allen hier untersuchten Spezies vor. Einige weisen N-terminal vor der konservierten Domäne potentielle Kex2-Schnittstellen auf (Tabelle 37). Die Funktion von Pry-homologen Proteinen bei Pilzen ist nicht bekannt. Es gibt aber strukturell homologe Proteine anderer Organismen mit bekannter Funktion, beispielsweise das in Wespengift vorkommende Protein V5/Tpx1 (siehe potentielle Substratbindungstasche in Abbildung 48B), für das gezeigt worden ist, dass es sich um eine Protease handelt. Tpx1 schneidet, wie auch Kex2, hinter Arginin.

PR1-homologe Proteine zeichnen sich durch einen hoch konservierten Bereich aus, der homolog zu der aktiven Domäne von Tpx1 ist und entweder N- oder C-terminal von einer S/T-reichen Region flankiert wird. Bei wenigen findet sich auch ein GPI-Verankerungsmotiv. Da es sich auch bei den Pry-Proteinen um Proteasen handeln kann, lag die Vermutung nahe, dass sie über abspaltbare Propeptide verfügen. Tatsächlich gibt es am Anfang der konservierten Domäne ein konserviertes Arginin, welches in den Pry-Proteinen der Pilze einer Kex2-Schnittstelle ähnelt (Abbildung 48A, Region 3). Andere Proteine der Familie tragen potentielle Schnittstellen am N-Terminus und CgPry1 kann kurz vor der konservierten Domäne prozessiert werden. Dies würde zum Verlust von ~60 Aminosäuren N-terminal zur konservierten Domäne führen, welche möglicherweise für eine katalytische Funktion des Proteins nicht benötigt werden, da sie bei anderen Proteinen der Familie vollständig fehlen (z. B. ScPry3).

Spezies und Sequenzname	N-terminal (Region 1)	S/T-reich	(Region 2)	(Region 3)	Topologie
<i>S. cerevisiae</i> YJL078C/Pry3			*QTTFFPFESDVLNEHNKFRALHVDTPAPLTSWSDTLA		19/(no TMs)(853-GPI)o881
<i>S. cerevisiae</i> YJL079C/Pry1	*PAVVVTEHAHEAAVVTVQGVVYVENGQTRTYETLAP	-(85)-	TSQAAATSSASSSDSDLSDFASSVLAHEHNKFRALHKDTPALSWSDTLA		20/(no TMs)299
<i>S. cerevisiae</i> YKR013W/Pry2	*APVAVTVTQHVHQAAATVVVQGVVYVENGQTLTTFITKG	-(117)-	TTATTTQSTASSTQSSSDFSTSMVNEHNKFRALHKDTGSLTWSDTLA		19/(no TMs)329
<i>C. albicans</i> CA0229	*KLVYEYETKYVTVVEIVTVSGETTYTERLETNGPTST	-(103)-	PQPQENSGTNDSSQLSS-FSRQILEAHNKRASH-GVNPLTWSNELY		24/(no TMs)317
<i>C. albicans</i> CA3335	*RVVKVEYLHLVVEVEPSTITE	-----	SCLSDGIDGYSGSFYLNNTFAELMLKEHNKFRKLH-QSCPLKWSSELF		21/(no TMs)202
<i>C. albicans</i> CA0104/Rbt4	*EIAYVTQTRGVTVGETATVATVTVTGATVTGGDQGDQ	-(135)-	TASGSSSGSNDGSGVKDTQFAQQILDHANNKFRARH-GVPDLTWDATGY		20/(no TMs)358
<i>C. albicans</i> CA1909/Pry1	*APAIITVWQTVTDAQVAAGPTAAAPAANANANVQQAAA	-(201)-	SSSGDIYAAISQCDGIDASFASEILDHANNKFRAQH-KVGDLSWDVDTY		21/(no TMs)410
<i>C. albicans</i> CA5344/Pry2	*ATITKFFTASTQTLFVTQTSQTVVATKSFVETIYSAPP	-(47)-	ISSSTASESDATPTADVEFAEELKEHNKFRALH-GVPALSWSNKLA		20/(no TMs)271
<i>C. glabrata</i> CAGLOG07667g	*APAVVTVTEVAHEHDVKTVRGVVYVQGGETKTSYTTLD	-(47)-	AAASPSPSPPAQDNLSDFAKSMLEHNKFRALHQDTNPLTWSDELA		18/(no TMs)258
<i>C. glabrata</i> CAGL0F05137g	*APAVVTVTKEAPPARVTVQAVNVNEDGGRQSSLSTLEK	-(13)-	TPSTDLDKFRQKFRSNLSEWQQMLDQHNKFRALHKTDSLWVNDNLA		18/(no TMs)227
<i>C. famata</i> DEHA0C19184g	*ENLGAFKIDFKVLRGNDKFRDLFLSDEPYIMKFRDSSEM	-(48)-	STSTQASPSSTSTMIASFGFEEDILNSHNEKFRAIH-GVQSLEWNETLV		19/(no TMs)276
<i>C. famata</i> DEHA0B01232g	*APALITRHTAPAVTHWVITYTNTVVKVVTAQDTENNA	-(103)-	TSGTSTSTSTSGSTASSGFEADILNAHNEKFRALH-GIQSLAWNDDTLA		22/(no TMs)330
<i>K. lactis</i> KLLA0C01496g	*QNTVTLDLTLDATSTRLSKPVATSTDSGYSATSSSTSVP	-(222)-	ASTTADSAATTTAEAISDADRESVLNAHNEYFRARHQSTNPLVWNEELA		19/(no TMs)456
<i>K. lactis</i> KLLA0D02442g	*RIYFLTYIYVLIFFAFKISWFKASIYITRGLISFCLR	-(150)-	PATTTAATSSSSSDSNLSSFASMLDEHNAKFRALHKDTPTLSWSELA		i12-39o45-63i84-102o368
<i>Y. lipolytica</i> YALI0D08140g	*APIDCDDNKLVARVTVFAHETTYVYANQDAAPAPAA	-(110)-	PSPSQAEAPAKSSNLDSSWSQSLDTONAKRAEH-GVGAFAWNETLA		18/(no TMs)316
<i>A. gossypii</i> AAL179W	*HQHDRKKPHLVVYGDVYVSGTDTFTKYYSKMTDKLPA	-(146)-	EPTGSPQRSRPPAGGQNGASFEELILRAHNSRRQLHKDTKPLTWSEELA		20/(no TMs)360
<i>A. gossypii</i> AAL178W	*RTVTVSVTATRVTIVISAAAVTTLRRTVFTTAYRT	-----	QTTPLTETPLSRPTAVNNTFASAVLDLHNDYRRRHHAVP-LRWNSTLY		19/(no TMs)205
<i>A. fumigatus</i> Afu1g02040	*SNTAGTTV	-----	IVTVPTIHPASYSLEIFKDTVLSTSNAYRRRHH-NASHLTWNETLT		26/(no TMs)(226-GPI)o262
<i>A. fumigatus</i> Afu1g12350	*KPVDKRSYVTDWTVIVTETITLTPYQPPPTSTPAYSPI	-(85)-	TSTSTTAPASVPTAAGSNAYQSAVLYNHNIHRSNH-SASSMTWNTTLE		18/(no TMs)320
<i>M. grisea</i> MG07807.4	*MLPAILLASIGTTSAPQETRT	-----	ITANPASPTDEPSYTRSDDFQAAMLTTHNRFRSQH-SAPDLEWDDTLA		17/(no TMs)(193-GPI)o215
<i>M. grisea</i> MG03085.4	*MRASLCLVASGAILALAGPLEIRGIHTEFVTDIVYVTV	-(52)-	NIKKQSSNSNSKASKGTNAFQNTAVYHHNIHRENN-SASALEWDSTYA		18/(no TMs)279
<i>F. graminearum</i> FG02744.1	*ADVVT	-----	ITAPVSIPSNEPEWKDKDTFTSAVLNSTNFYRHH-NATDLKWNKTLE		19/(no TMs)196
<i>F. graminearum</i> FG09548.1	*SPIGNVLNERAMKTEWVYEVTVVTEGAEPKKAHTA	-(107)-	PKVTVEASIDTNSLSLDGAYDTVMLAYHNIHRSNH-SAPALEWDDTLA		18/(no TMs)348
<i>T. reesei</i> orf39324	*VPLEKRVMETDWTVDIVTVTVTPDNASPSATPTAAEA	-(11)-	*LSASPEPSATPEPAPQPAPQPHNVHRANH-SSPALEWDDTLA		17/(no TMs)245
<i>N. crassa</i> NCU02470.1	*SPILQDRRVHFTTKVVEEWVTVTGTAGNVPFATVNAFH	-(86)-	MQEHAATTKVDSPPAASGSDYASTLLYHHNIHRSNH-SAGAMTWGETYA		18/(no TMs)332
<i>U. maydis</i> UM01204.1	*ASQPVIERNISIASDQNDLQPRDTSSCKASARIRNHQH	-(56)-	TRSSAESASTGTADSNLSNFETTMLEIHNMDFRAKH-SASPLTWDTTLA		23/(no TMs)279
<i>V. vulgaris</i> 1qnx/antigen5	*NNYCKIKCLKGGVHTACKYG	-----	SLKPNCGNKVVVSYGLTKQEKQDILKEHNDFRQKIANMKNLVWDELA		18/(no TMs)348
<i>Solanum lycopersicum</i> 1cfe			*QNSPDYLAHVNDARAAQV-GVGPMWDANLA		18/(no TMs)348
<i>H. sapiens</i> 1smb/Golgi- Pr1			*MGKSASKQFHNEVLKAHNEYRQKH-GVPPLKCLKNLN		(no TMs)

**Tabelle 37: Potentielle Kex2-Schnittstellen in Pry-homologen Genen.** In Region 1, am N-terminalen Ende der Proteine, befinden sich bei einigen Proteine potentielle Schnittstellen. Am Beginn der konservierten Region (Region 3) befindet sich ein konserviertes Arginin, welches bei den Proteinen der Saccharomyces Gruppe einer Kex2-Schnittstelle ähnelt. CAGL0F05137g besitzt zusätzlich eine potentielle Schnittstelle kurz vor der konservierten Region (Region 2).

Aus dieser Familie wurden die drei Proteine CaRbt4, CgPry1 und CgPry3 erfolgreich exprimiert und untersucht. CaRbt4 wurde dabei nicht von Kex2-Enzymen geschnitten (nicht gezeigt). Bei CgPry1 wurden zwei Fragmente von jeweils ~17 und ~18 kDa beobachtet. Bei CgPry2 wurde nur eine Produktbande von ca. ~13-15 kDa beobachtet, welche sich jedoch nicht eindeutig einem spezifischen Produkt zuordnen ließ.



**Abbildung 48: Prozessierung und Struktur von Proteinen der PR-1-Proteinfamilie.** Die Proteine CgPry1 und CgPry2 werden von den Kex2-Proteasen geschnitten.

#### 4.4.6.9 Auch Proteine aus dem ER tragen prozessierbare Schnittstellen

Das Protein Kex2 ist in einem späten Kompartiment des Golgi lokalisiert, und wird auch dort erst aktiviert. Trotzdem konnte in Zusammenarbeit mit H. Mora, Universität Aberdeen, UK, [137] gezeigt werden, dass auch ER-lokalisierte Proteine durch Kex2 *in vivo* prozessiert werden. In diesem Fall handelt es sich um das Membranprotein CaMns1, eine  $\alpha$ -1,2-Mannosidase welche an der Glykosylierung von Proteinen im ER beteiligt ist. Mittels Kombination von Proteaseinhibitoren wurde zunächst eine Prozessierung durch eine Serinprotease nachgewiesen, später die Prozessierung durch rekombinantes Kex2 (Abbildung 49A). Durch N-terminale Sequenzierung (Tabelle 38, identifizierte Sequenz fett gedruckt) konnte hier gezeigt werden, dass der Schnitt hinter der Membrandomäne an der unkonventionellen Schnittstelle SRAR<sub>152</sub>/DWIK stattfindet. Die Schnittstelle befindet sich zwischen zwei konservierten Teilen des Proteins (Tabelle 38) und durch diese Prozessierung wird das Membranprotein in ein lösliches konvertiert und der Re-Import in das Zytosol eingeleitet. Für *C. glabrata* Mns1 konnten Mora *et al.* eine entsprechende Prozessierung zeigen [137].

Eine entsprechende Schnittstelle ist zwischen den Orthologen konserviert (Tabelle 38) und die Raumstruktur von ScMns1 (Abbildung 49B) zeigt, dass das Arginin an P1 exponiert an der Oberfläche des Proteins liegt. Obwohl es sich hier innerhalb einer Helixstruktur befindet, ist die Seitenkette vom Protein weg gerichtet und somit für die Protease erreichbar. In den Spezies der *Saccharomyces* Gruppe befindet sich eine unkonservierte Insertion von ca. 9-12 AS um die Schnittstelle herum, was nahe legt, dass dieser Abschnitt eine Schleife an der Oberfläche des

Proteins bildet und damit suszeptibel gegenüber Proteolyse ist. Dies bestätigt sich auch in der Raumstruktur von ScMns1 (Abbildung 49B).

Name	Sequenzmotiv
<i>S. cerevisiae</i> YJR131W (PDB:16GI)	DSVDTLMLMYNSS--TLYKSEFEAEIQ <b>R</b> SEHWINDVLD <del>FD</del> IDAENVVVFETTIRM
<i>C. glabrata</i> CAGL0M00528	DSLDTLMVMYNS--ANYKDEF <b>R</b> EHINAATEWIDRELNFYD <del>NS</del> VSLFETTIRL
<i>A. gossipii</i> AER165W	DALDTMLMQAHTKDVRRHLRL <b>R</b> REVAR <b>R</b> CEAWIRDEL <del>LD</del> YDMNTEVSVFETTIRM
<i>K. lactis</i> KLLA0B04356g	DSVDTMMLMHNKTKDDTHR <b>K</b> RFAEEINIAESWIKNTLDY <del>NV</del> DSEVNLFFETTIRM
<i>C. albicans</i> CA1037	DSLDTLMIMDCPE-----EVSRA <b>R</b> DWIKNDLDYTFDYNVNTFETTIRM
<i>C. famata</i> DEHA0D10153	DSLDTLMIMDCPE-----QLARA <b>R</b> KFKVDDLDYHFDY <del>NV</del> NVVFETTIRM
<i>Y. lipolytica</i> YALI0D17424	DSLDSLMLMGLEK-----ELAEAR <b>R</b> YWVDKDL <del>DY</del> NI <del>DY</del> EVNTFETTIRM
<i>U. maydis</i> UM02227	DTLDTLILMGLKD-----EYERAR <b>R</b> EWVAEELNWN <del>V</del> AGRMNVFETTIRT
<i>S. nodorum</i> SNU03652	DALDTLMMLNLT-----ELKHAR <b>R</b> EWVSTLDY <del>NK</del> DQDVNTFETTIRM
<i>A. nidulans</i> AN5748	DALDTLMIMNLT-----RVQHAR <b>R</b> TWIHNSLQYDQD <del>H</del> VSTFETTIRM
<i>N. crassa</i> NCU02778	DSLDTLMIMNLT-----RVTHAR <b>R</b> EWISKSLTWDQDQ <del>D</del> VNTFETTIRM
<i>A. fumigatus</i> Afu6g06790	DALDTLMIMNLT-----RVQHAR <b>R</b> NWIHNSLQY <del>N</del> QDHDVNTFETTIRM
<i>F. graminearum</i> FG00612	DSLDTMILMNLTS-----RLQDAR <b>R</b> QWISESLTWDQDQ <del>D</del> VNTFETTIRM
<i>T. reesei</i> TR40031	DSLDTMMLNQTT-----RLQHAR <b>R</b> EWISTSLTWDQDQ <del>D</del> VNTFETTIRM
<i>Cr. neoformans</i> cn05461	DSIDSLIMDLIP-----EYQRA <b>R</b> DWVRDHLN <del>FD</del> KAQFNTFETTIRL

Tabelle 38: Kex2-Schnittstellen in der  $\alpha$ -Mannosyltransferase Mns1. Blau: Potentielle Kex2-Schnittstellen, fett: die von Mora *et al.* [137] identifizierte Schnittstelle in CA1037.

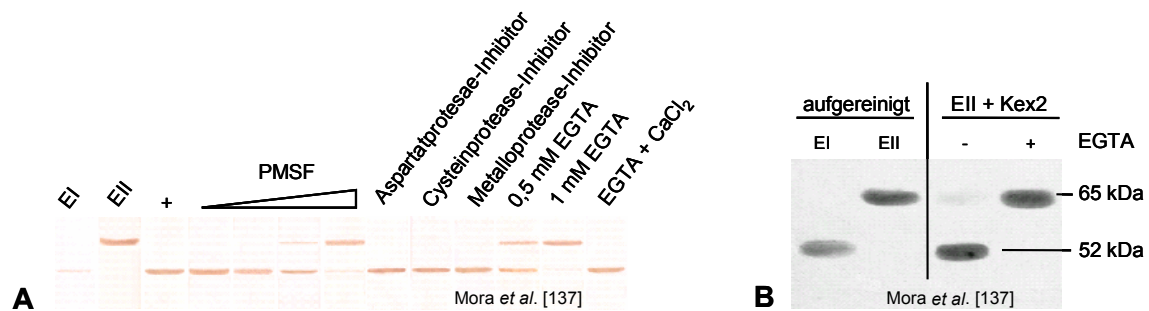


Abbildung 49: Prozessierung und Struktur der  $\alpha$ -Mannosyltransferase Mns1. Durch Einsatz verschiedener Proteaseinhibitoren konnte gezeigt werden dass die Prozessierung von EII zu EI durch eine Protease mit einem Inhibitionsprofil wie Kex2 geschieht (A). Die Prozessierung von EII zu EI kann durch Kex2 erfolgen und wird durch Zugabe von EDTA gehemmt (B) [137].

Das Mns1-Protein ist nicht das einzige ER-lokalisierte Protein, welches potentielle Kex2-Schnittstellen besitzt: in den untersuchten Organismen wurde die Schnittstelle mit der höchsten Bewertung durch die PSSM in dem Chaperon Kar2 gefunden. Kar2 (auch „BiP“) ist ein Hitzeschockprotein, welches über das Ire1-Hac1-System die „unfolded protein response“-Signalkaskade auslöst. Das Motiv ist hoch konserviert (Tabelle 39), was zunächst eine Lokalisation in einer strukturell unzugänglichen Region des Proteins nahe legt.

Aus dieser Gruppe konnte ScKar2 erfolgreich exprimiert werden, die orthologen Proteine aus *C. albicans* und *C. glabrata* hingegen nicht. Der *in vitro* Verdau zeigt, dass ScKar2 tatsächlich von Kex2 prozessiert werden kann (Abbildung 50).

Name	Sequenzmotiv
<i>S. cerevisiae</i> Kar2	DVSDNNKALAKL KREA EKAKR ALSSQMSTR IEID
<i>B. taurus</i> Hsc70 (1YUW)	DI SEN KRAV RRL RTACERAKR TLSSSTQAS IEID
<i>C. albicans</i> CA0915	DITANAKAISKL KREA EKAKR TLSSQMSTR VEID
<i>C. glabrata</i> CAGL0D02948	DVSDNEKALAKL KREA EKAKR ALSSQMSTR IEID
<i>C. famata</i> DEHA0A01749	DISGNSKAISKL KRETEKAKR TLSSQMSTR IEID
<i>Y. lipolytica</i> YALI0E13706	DITKNAKTIGKL KREVEKAKR TLSSQMSTR IEIE
<i>A. nidulans</i> AN2062	DVTKDLKAMGKL KREVEKAKR TLSSQMSTR IEIE
<i>A. gossypii</i> ACR038W	DVSSNAKAMAKL KREA EKAKR ALSSQMSTR VEID
<i>A. fumigatus</i> Afu2g04620	DVSKDLKAMGKL KREVEKAKR TLSSQMSTR IEIE
<i>M. grisea</i> MG02503	DVTKDAKAMGKL KREA EKAKR TLSSQKSTR IEIE
<i>F. graminearum</i> FG09471	DITKDLKAMGKL KREA EKAKR TLSSQKSTR IEIE
<i>K. lactis</i> KLLA0D09559	DVTKNDKAMAKL KREA EKAKR SLSSQTSTR IEID
<i>T. reesei</i> 42955	DISKDLKAMGKL KREA EKAKR TLSSQMSTR IEIE
<i>N. crassa</i> NCU03982	DVTKDAKAMGKL KREA EKAKR TLSSQMSTR IEIE
<i>Cr. neoformans</i> cn14166	DVSKNNRAMGKL KREVEKAKR TLSSQMSTK IEIE

Tabelle 39: Kex2-Schnittstellen in Kar2-homologen Proteinen. Blau: Potentielle Kex2-Schnittstellen.

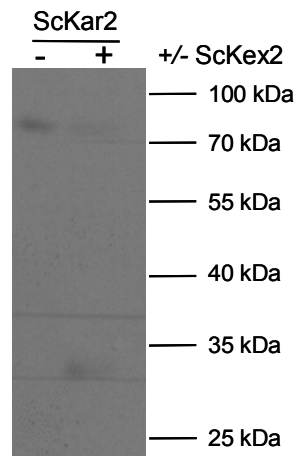


Abbildung 50: Prozessierung von ScKar2. Westernblot Analyse eines Kex2-Verdau von ScKar2. Nachgewiesen wurde das 6xHis-tag am 33,7 kDa großen N-Terminus des rekombinanten Proteins.



#### 4.4.6.10 Einige Proteine mit potentiellen Kex2 Schnittstellen werden nicht prozessiert

Außer den bereits oben beschriebenen, unprozessiert gebliebenen Proteinen Rbt4 und Tos1 wurde bei sieben weiteren erfolgreich aufgereinigten Proteinen trotz potentieller Kex2 Schnittstellen keine Prozessierung beobachtet: die drei Proteine unbekannter Funktion und ohne Homologien zu anderen Proteinen CAGL0A02277, CAGL0H08910, CA1394 und die vier Proteine CA2942 (Zellwand-Strukturprotein CCW14), CA0375 (putative Transglukosidase Crh11), CAGL0J11770 (Phospholipase B1) und CA5147 (saure Phosphatase Pho114).

#### CA0375/Crh11

Die Familie der Crh1-homologen Proteine (Abbildung 51) stellt eine Gruppe von Zellwandlokalisierten Transglykosidasen bei Pilzen dar, welche für die Vernetzung des Chitin- und Glukangerüsts der Zellwand benötigt werden, und diese enzymatische Reaktion möglicherweise selbst katalysieren [138].

Spezies	Protein	Schnittstelle
<i>S. cerevisiae</i>	Crh1	IKYGS DGLSMTLAKR/YDNPSLKS NFYIMYGKLE
<i>C. albicans</i>	Crh11	IDSGSNGLSLTMKKR/FDNPSFKSNFYIMFGRVE
<i>B. subtilis</i>	PDB:1CPM	LKLG---LTSSAYNK.FDCAEYRSTNIYGYGLYE
<i>C. glabrata</i>	CAGL0G09449g	IEYTS DGLAMSLTKR/FDNPSLKS NFYIMYGKTE
<i>K. lactis</i>	DEHA0G19877g	VSGDDGLELKLAKR/FDNPTLQSNFYIMFGKIE

**Abbildung 51: Putative Kex2-Schnittstellen und Struktur von Crh1-homologen Proteinen.** Potentielle Schnittstellen in Proteinen der *Saccharomyces*-Gruppe. Schnittstellen sind blau, N-Terminus des Proteins in der Kristallstruktur ist gelb unterlegt (siehe Abbildung 65).

#### CAGL0J11770g/Plb1

Bei dieser Phospholipase B1 aus *C. glabrata* handelt es sich ebenfalls um ein Enzym der Zellwand, dessen biologische Funktion allerdings unklar bleibt. In den gezeigten Plb1-homologen Proteinen befinden sich zwei konservierte potentielle Schnittstellen (Abbildung 52), die in CAGL0J11770g jedoch unprozessiert bleiben.

Spezies	Protein	Schnittstelle
<i>H. sapiens</i>	Pla2	SLEVCSCPDLRFSMALCDQEKTRFRQQRKEHIRESMK
<i>C. glabrata</i>	CAGL0J11770g / Plb1	SQSEIDWLKRRDVTREALESFLLKRVTSNFTSNSS
<i>C. glabrata</i>	CAGL0J11748g	SPAEKEWLKRRDPITRDALQTFLLRRRAFANVSTEIT
<i>C. glabrata</i>	CAGL0E02321g	SQNETQWLQRRDVRTKEALHNFLQRATSSSQNFQ
<i>C. albicans</i>	CA1975	NPNEQSYINARYPPIAKQALSFKFLHNANLQNFQDV
<i>C. albicans</i>	CA3834	SPQEQEWLKRSPIANENLISFLNSANMTDFDP
<i>C. albicans</i>	CA2223	SQQEKDYIHKRQETTNNKLNIDFLSKRANLSDFDA
<i>S. cerevisiae</i>	YOL011W	SQNESAWLEKRNKVTVALKDFLTRATANFSDSSE
<i>S. cerevisiae</i>	YMR008C	SDNETEWLKKRDAYTKEALHSFLNRATSNFSDTS
<i>S. cerevisiae</i>	YMR006C	STAETDWLKRRDAYTKEALHSFLSRATSNFSDTS
<i>C. famata</i>	DEHA0G09185g	SSSEKKWLKERQKKTNQALITYLDNANLTFDA
<i>C. famata</i>	DEHA0F04510g	SSSERDYMSNRHETTNNKLNIGFLSKRANLSDFDA
<i>C. famata</i>	DEHA0E05522g	SDEEKIEWIEGRNKITNENLIKFLNCDMKDFDA
<i>K. lactis</i>	KLLA0C05940g	SDDEADWLKVRHESTVPALKDFLQRFKGFKGFTNDTS
<i>A. fumigatus</i>	Afu4g08720	SPNETSWLETRRGKTTSAMKDFFNHVKIQDFDA
<i>Y. lipolytica</i>	YALI0E16060g	SSQEREWVQNRHEQTRPELLNYLKRVGFKSVDP
<i>A. fumigatus</i>	Afu5g01340	SSAETSWLKAARRNNTKDALKAFLSRVDLGSFNG
<i>A. fumigatus</i>	Afu3g14680	SPNETSWLEIRRKNTVQPMTDLLGRNLGFD
<i>U. maydis</i>	UM01035.1	SNGEADYISRRQNTSRGWQKWLSQLAKLDSVLPGG
<i>A. nidulans</i>	AN7792.2	SPNETKWLEPRRKEIISPMKNLLTRLNISDFDA
<i>A. gossypii</i>	AAL027W	SWTLMRLLLLTNFNIEELKEWNINSSLL-EGVPNFEI
<i>M. grisea</i>	MG07287.4	SPDEASWLPNRRKNTVKPMSDFLTRANIQGFNA

Abbildung 52: Putative Kex2-Schnittstellen und Struktur von Plb1-homologen Proteinen. In den zu CgPlb1 homologen Proteinen finden sich zwei konservierte Motive, die Kex2-Schnittstellen ähneln (blau unterlegt).

### CA5147/Pho114

Aus der Familie PHO der sauren Phosphatasen trägt lediglich das Protein aus *C. albicans* (Pho114/CA5147) Motive, die Kex2-Schnittstellen ähneln (Abbildung 53A und B) das Protein ist bereits vor der Zugabe von Kex2 aktiv (siehe Abschnitt 4.4.5), und wird durch Zugabe von Kex2-Enzymen nicht weiter prozessiert. Dies zeigt, dass es sich hier nicht um ein Substrat von Kex2 handelt.

Spezies	Protein	Schnittstelle 1	Schnittstelle 2
<i>C. albicans</i>	CA5147	RHGERFPSKGDGKYFNVSMEVFKRY-GEFHGDLISFLN	GTTNLLRHGAYFRKYQSLFDQKEK
<i>C. albicans</i>	CA0616	RHGERYPSKSNKGKSLKAIYAKFENYKGTFFKGDLAFLN	GTTNLRHGAAFRKRYGSLYKENST
<i>D. castellanii</i>	PDB:2GFI	RHGERYPSTGSGKKYKAVYEKLMSYNGTFFKGLAFLN	GTTDAMKHGIAFRKYGELFDTNDT
<i>S. cerevisiae</i>	YBR093C	RHGERYPTVSLAKTIKSTWYKLSNYTRQFNGSLSFLN	GEMNAKRHARDFLAQYGYMVENQTS
<i>S. cerevisiae</i>	YBR092C	RHGERYPTYSKGATIMKTWYKLSNYTRQFNGSLSFLN	GEMDAKRHAREFLAQYGYMFENQTS
<i>S. cerevisiae</i>	YHR215W	RHGERYPTVSKAKSIMTTWYKLSNYTGQFSGALSFLN	GEMNAKRHARDFLAQYGYMVENQTS
<i>S. cerevisiae</i>	YDL024C	RHGERYPTRSEAKDIFEVWYKISNYTGKYEGLSFLN	GEMNAKRHAREFLAKYGLMENCNTN
<i>S. cerevisiae</i>	YAR071W	RHGERYPTVSKAKSIMTTWYKLSNYTGQFSGALSFLN	GEMNAKRHARDFLAQYGYMVENQTS
<i>C. famata</i>	DEHA0C09460g	RHGERFPSSNAGKQYEHIMQKFEDYNGTFFKQLSFLN	GTTNALKHGAAFRKRYNSLYNDSNT
<i>C. famata</i>	DEHA0F08613g	RHGERYPSKKNVGKALEKINEKFKGYNGTFFKGEISFLN	GADNLRHGTTFRSKYNEIYNPDKP
<i>C. famata</i>	DEHA0F08525g	RHGERYPGISDGEKQKALVDKLSYNSLSGFLSFLN	GYETCNKAGLAFRAKYNDLYNENET
<i>C. famata</i>	DEHA0F08547g	RHGERYPSKKNVGKALEKINEKFKGYNGTLLKGEISFLN	GADNLRHGTTFRSKYNEIYNPDKP
<i>K. lactis</i>	KLLA0E24981g	RHAERFPKAAKGEKLEIMWNKFKEMKGEFKGPLAIFN	GSKTAQQLGNVIAVHYGELIGD-S
<i>K. lactis</i>	KLLA0B00286g	RHGERYPTASKGKLMIALWDKLFKQYNGPMEVFN	GAKTAQHLGKYIAYNYGDLFSD-S
<i>K. lactis</i>	KLLA0C19338g	RHGERYPSKSKGAKLSNVYQKLQNYTGALNGSLSFLN	GEQSARTHARQFISLYGDLLENST
<i>K. lactis</i>	KLLA0A00176g	RHGERYPTASKGKLMIALWDKLVFQGFQDFPLEIFN	GSKTAELGKYIAFNHYGELFGN--Y

A

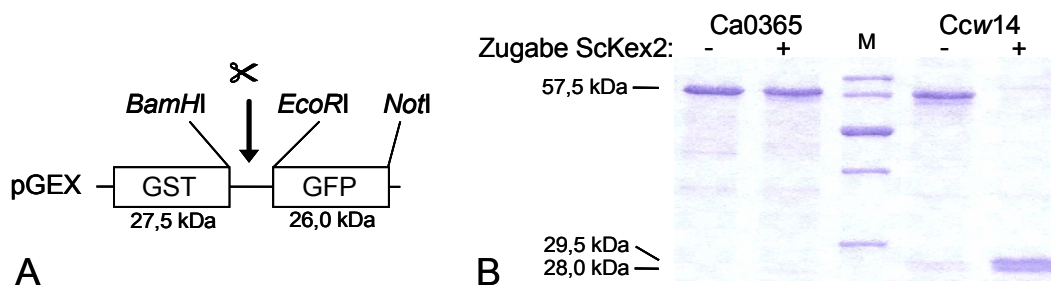
B

Abbildung 53: Putative Kex2-Schnittstellen in CaPho114 und Struktur der homologen Phosphatase aus *Debaryomyces castellanii*. A und B: Zu den CaPho14 Kex2-Schnittstellen-Motiven analoge Regionen der sauren Phosphatasen der *Saccharomyces* Gruppe.

#### 4.4.7 Untersuchung von Proteinen ohne Homologe bekannter Struktur

Für vier Proteine aus dem vorhergehenden Abschnitt existieren keine strukturellen Daten in der Literatur. Um diese Schnittstellen einer Überprüfung zugänglich zu machen, wurde ein System entwickelt, bei dem die Schnittstellen den Proteasen zugänglich gemacht werden. Dazu wurde ein Plasmidvektor (pGG4) konstruiert, bei dem kurze Fragmente mit der potentiellen Schnittstelle zwischen zwei kurzen, kompakt gefalteten und globulären Domänen (N-terminal Glutathion-S-Transferase, GST, und C-terminal das grün fluoreszierende Protein, GFP) exprimiert werden können. Dieses hat zur Folge, dass die Schnittstellen unabhängig von ihrer Struktur im nativen Protein der Protease zugänglich gemacht werden und dass überprüft werden kann, ob das Motiv prinzipiell ein Substrat für die Protease darstellt.

Versuche mit Schnittstellen aus CA0365 und CaCCW14 zeigten, dass dieses Prinzip funktioniert. Die im nativen Protein nur von ScKex2 nicht prozessierte Schnittstelle 3 aus CA0365 bleibt auch in diesem Konstrukt unprozessiert. Dies offenbart, dass diese Sequenz tatsächlich kein Substrat für ScKex2 darstellt. Die im nativen Protein unprozessierte Schnittstelle 2 aus CaCcw14 dagegen wird im Fusionsprotein geschnitten, was bedeutet, dass im nativen Protein strukturelle Gründe die Prozessierung verhindern.

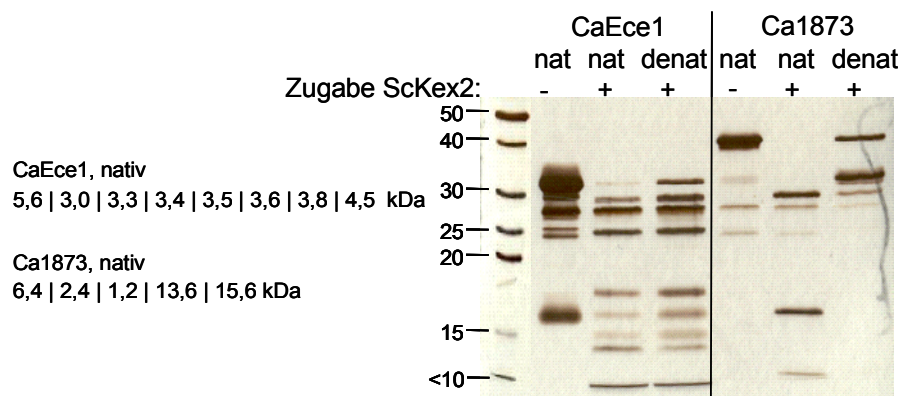


**Abbildung 54: Strukturunabhängige Proteolyse mittels GST-GFP-Fusionsproteinen.** A: Mit dem Vektor pGG4 lässt sich eine potentielle Proteaseschnittstelle, unabhängig von ihrer nativen Konformation, durch Fusion zwischen GST und GFP einer Protease zugänglich machen. B: Ein GG4-Fusionsprotein der Schnittstelle 3 aus CA0365 wird von ScKex2 nicht geschnitten, ein GG4-Fusionsprotein der Schnittstelle 2 von CaCCW14 hingegen wird prozessiert.

Zusätzliche Evidenz, dass die Struktur des Proteins einen Einfluss auf die Prozessierbarkeit eines Proteins hat, liefern Experimente mit denaturierten Proteinen. Im Laufe dieser Arbeit konnte beobachtet werden, dass renaturierte Proteine weniger effizient geschnitten wurden als solche, die in nativer Konformation aufgereinigt wurden.

Ein vergleichendes Experiment mit nativen und hitzedenaturierten Proteinen bestätigt dies: denaturiertes Ece1 oder Ca1873-Protein wird weniger effizient geschnitten als natives. Hierfür gibt es zwei mögliche Erklärungen: einerseits könnte die Ausbildung einer spezifischen Struktur für die Proteolyse nötig sein, andererseits könnten im denaturierten Protein die Schnittstellen größtenteils in unspezifischen Konglomeraten vor der Protease geschützt sein.

Die Beziehung zwischen Substraterkennung und Struktur der Schnittstelle wird unter Abschnitt 4.6 weiter untersucht.



**Abbildung 55: Denaturiertes Protein wird weniger effizient geschnitten als natives.** Die hitzedenaturierten Proteine („denat“) werden nach Zugabe der Protease weniger schnell geschnitten als die Nativen („nat“). Dieser Effekt ist bei Ece1 geringer als bei CA1873, da Ece1 über mehr Schnittstellen verfügt.

#### 4.4.8 In prozessierten Schnittstellen sind negative Ladungen an den Positionen P1'-P4' überrepräsentiert

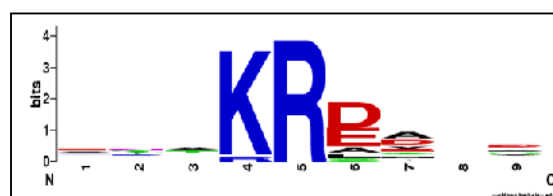
Alle in den vorhergehenden Kapiteln betrachteten Sequenzmotive für potentielle Kex2-Schnittstellen wurden aufgrund der biochemischen Aktivität von ScKex2 ausgewählt. Eine weitergehende Aufteilung in *in vitro* prozessierte und nicht prozessierte Schnittstellen (Abbildung 56A und B) zeigt, dass in prozessierten Motiven im P1'-P4'-Bereich negativ geladene Aminosäuren stärker repräsentiert sind als in nicht prozessierten (Abbildung 56C und D). Dies legt den Schluss nah, dass auch der P'-Bereich für die Erkennung der Substrate mitverantwortlich ist. Diese Beziehung zwischen P'-Bereich und Substraterkennung wird unter Abschnitt 4.5 weiter untersucht.

Prozessierte Motive			
Name	Pos.	Sequenz	PSSM
CaSun41	45	HKHKRAVAY	627
ScKar2	309	AKLKREAEK	564
Cgα-Pheromon	85	DFTKRDANP	402
Scα-Pheromon	103	PMYKREAEA	392
Scα-Pheromon	145	PMYKREADA	392
Scα-Pheromon	131	PMYKREADA	392
CgScw4	30	HKNKRDVVT	389
CgPir1	62	VVAKRAVSQ	312
CaEce1	30	FNMKRDVAP	303
CA1873	195	TMEKRLDLD	286
CaEce1	159	ENAKRDGVP	273
CA2974	54	YNAKRDLMs	273
Cαα-Pheromon	80	NNEKREAKS	264
CgPry1	80	AQKKRSNLS	256
CA1873	58	YKTRDAID	246
CaSun41	29	KNIKREDCD	243
CaEce1	193	QQAKRDGLE	228
CA0365	74	EQSKRLDAD	228
CaEce1	60	AITKRSIIG	187
CA1873	68	EIMKRDYAI	173
CaEce1	227	QPVKRDAGS	158
CA2974	65	EIIKRENQI	134
CA0365	31	EVDKRLDAD	131
Cαα-Pheromon	101	EPGKRDANA	122
Cαα-Pheromon	122	EPGKRDANA	122
CgSun4	30	HAIKRDVDV	119
CaEce1	125	ASTKRDGAN	114
CA0365	105	ASTKREDES	114
CA2974	180	ASAKREIMV	94
CA0365	62	VTAKREDES	94
CA0365	149	VSAKREDTV	94
CA1873	39	ALTKREEQD	60
Cgα-Pheromon	105	GLFRRSADA	39
Cgα-Pheromon	130	GLFRRSADA	39
CgPry1	75	DLDKRAQKR	21
Scα-Pheromon	84	SLDKREAEA	21
CaEce1	92	KGNKREDID	18
CaMns1	152	VSRARDWIK	2

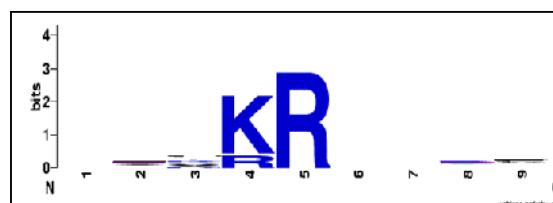
A

Unprozessierte Motive			
Name	Pos.	Sequenz	PSSM
CaCcw14	222	QHVKRIKKF	578
CaPho11	94	EVFKRYGEF	483
CAGL0H08910	157	KYVKRNYSF	372
CgEgt1	190	SQFKRPRI	353
CA1394	397	KQFKRSLMD	353
CaPho11	147	YFRKRYQSL	313
CAGL0A02277	98	TMAKRDKV	295
CaCrh1	76	TMKKRFDNP	277
CaRbt4	222	HNKKRARHG	256
CgPry2	132	HNKKRALHQ	234
CgPry1	98	HNKKRELHK	214
CaCcw14	174	IVIRRPVNV	136
CaCcw14	193	GNIRRVIIIT	119

B



C



D

**Abbildung 56: In prozessierten Substraten sind negativ geladene Aminosäuren an den Positionen P1'-P4' überrepräsentiert.** A: Aufstellung der aus der Literatur bekannten und in dieser Arbeit gefundenen Schnittstellen von Kex2-Enzymen und deren Bewertung mittels der PSSM. B: Aufstellung der in dieser Arbeit gefundenen, nicht prozessierten Motive. C: Sequenzlogo der prozessierten und D: unprozessierten Motive, blau: positiv geladene Aminosäuren (AS), rot: negativ geladene AS, grün: polare AS und schwarz: unpolare AS.

#	Theoretische Fragmentgrößen	Org.	Accession	Name	Sc	Ca	Cg	Pp	Pos.
1	5,6   3,2   1,4   3,1   1,5   3,1   2,8	Ca	CA0365	put. Prohormon	-	++	+	+	31 62 74 105 149
2	2,1   1,4   3,0   3,5   3,5   3,6   3,6   3,9   3,1   1,5	Ca	CA1402	Ece1	+	+	+	+	30 60 92 125 159 193 227
4	6,4   2,4   1,2   13,6   15,6	Ca	CA1873	OP4-like 1	+	+	+	+	39 58 68 195
5	13,0   2,4   2,3   2,2	Ca	CA2412	$\alpha$ -Pheromon	+	+	+	+	80 101 122
6	8   1,3   12,2   15,6	Ca	CA2974	OP4-like 2	+	+	+	+	54 65 180
10	13,3   2,2   35	Ca	CA2303	Tos1	-	-	-	-	105 123
11	4,9   1,9   38,5	Ca	CA0883	Sun41	+	+	+	+	29 45
12	17,4   0,8   1,5   9,0	Ca	CA2942	Ccw14	-	-	-	-	174 193 222
13	12,4   6,5   35,5	Ca	CA5147	Pho114	-	-	-	-	94 147
20	35   1,9   11,5   1,4   5,2	Ca	CA4679	Pga17	+	+	+	+	338 354 457 468
21	24   14	Ca	CA0104	Rbt4	-	-	-	-	222
23	9,3   36	Ca	CA0375	Crh1	-	-	-	-	76
27	45,1   2,1	Ca	CA1394	unbekannte Funktion	-	-	-	-	397
30	11,3   2,9   2,9   2,6	Cg	CAGL0H03135g	$\alpha$ -Pheromon	+	+	+	+	85 105 130
31	8,5   65	Cg	CAGL0J11770g	Plb-homolog	-	-	-	-	59 73
32	5,5   35	Cg	CAGL0M13805g	Scw4	+	+	+	+	30
33	7,6   31,5	Cg	CAGL0L05434g	SUN4-homolog	+	+	+	+	30
34	20,4   21,5	Cg	CAGL0H08910g	unbekannte Funktion	-	-	-	-	157
38	8,5   27,6	Cg	CAGL0M08492g	Pir1	+	++	++	++	62
39	12,3   19,4	Cg	CAGL0A02277g	unbekannte Funktion	-	-	-	-	98
40	10   0,6   2,2   15	Cg	CAGL0F05137g	Pry1	++	++	+	+	76 80 98
41	15   13	Cg	CAGL0G07667g	Pry2	+	+	+	+	132
43	10,9   2,4   2,5   2,5   2,3 kDa	Sc	YPL187W	$\alpha$ -Pheromon	+	+	+	+	84 103 145 131
45	33,7   40,6	Sc	YJL034W	Kar2	+	+	+	+	310

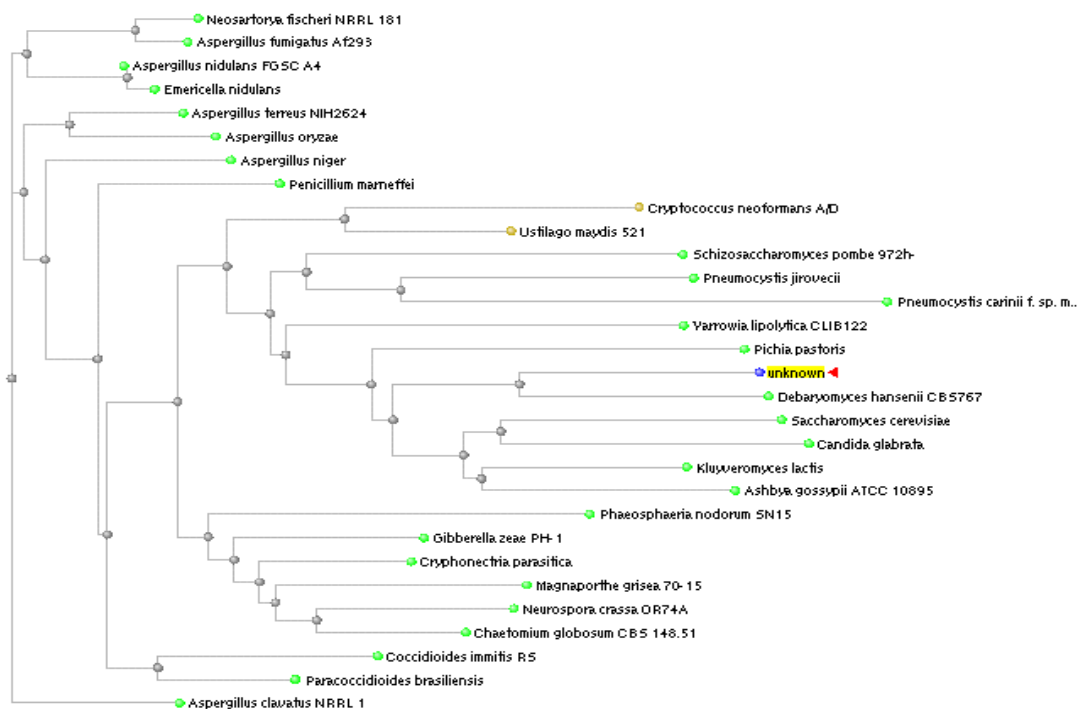
**Tabelle 40: Übersicht über die Prozessierung von vorhergesagten Kex2-Substraten.** Die Tabelle fasst die oben beschriebenen theoretischen Fragmentgrößen (in kDa) der Produkte bei Schnitt durch eine Kex2-Protease an den vorhergesagten Schnittstellen und die beobachteten Schnitte durch die vier Kex2-Enzyme bei den erfolgreich exprimierten und aufgereinigten Proteinen zusammen.

## 4.5 Vergleich der Struktur von Kexinen aus Pilzen und Säugern

Für eine Analyse des Substratspektrums von Kex2-ähnlichen Proteasen bei Pilzen ist es notwendig zu untersuchen, ob dieses Prozessierungssystem zwischen den verschiedenen Organismen konserviert ist. Dazu wurden aus verschiedenen Quellen die Proteinsequenzen möglichst vieler Kex2-ähnlicher Proteasen zusammengetragen und auf die Konservierung der relevanten Aminosäuren hin untersucht.

### 4.5.1 Identifizierung von *KEX2*-homologen Genen in Genomdaten verschiedener Pilze

Da in der Literatur nur relativ wenig Kex2-Proteasen beschrieben sind, wurden verschiedene Genomprojekte nach Kex2-ähnlichen durchsucht. Das geschah im Allgemeinen mittels BLAST-Analyse mit der Proteinsequenz der Protease Kex2 aus *S. cerevisiae*. In Fällen, bei denen ein Protein-BLAST auf der Website des Genomprojektes nicht zur Verfügung stand, wurde die Nukleotidsequenz des *KEX2*-homologen Gens des am nächsten verwandten Pilzes verwendet. Da die Sequenz des *KEX2*-Gens aus *Cr. neoformans* in der Genomdatenbank unvollständig war, wurde sowohl die genomische als auch die cDNA-Sequenz experimentell neu bestimmt (siehe Abschnitt 4.4.1). Es konnte in den Genomdaten jeder Spezies genau ein Kex2-orthologes Gen gefunden werden.



**Abbildung 57: Dendrogramm Kex2-homologer Proteine verschiedener Pilze.** Es findet sich in jedem Genom genau ein zu ScKex2 orthologes Protein. Das Alignment der Kex2-Proteine spiegelt die phylogenetische Nähe der Organismen zueinander wieder. Unknown= *C. albicans* Kex2.

#### 4.5.2 *In silico* Analyse der P'-Substratbindungsstellen von Prohormonkonvertasen

Im Laufe der letzten Jahre wurden die räumlichen Strukturen sowohl von *S. cerevisiae* Kex2 [28, 54] als auch von *M. musculus* Furin [53] aufgeklärt. Vergleiche dieser Strukturen zeigten, dass die Tertiärstruktur dieser Proteine extrem hoch konserviert ist [54] und dass die Unterschiede in der Substraterkennung an P6 bis P1 durch einzelne Aminosäureaustausche hervorgerufen werden (siehe Zusammenfassung in Tabelle 41). Zur Substratspezifität an P1' und folgenden Positionen existieren in der Literatur dagegen nur wenig experimentelle Daten, unter anderem da die zur Untersuchung der Substratspezifität verwendeten fluorogenen Peptide und die in den Strukturanalysen verwendeten Inhibitoren aus mechanistischen Gründen nicht über diese Positionen verfügen. Es gibt jedoch Arbeiten, die durch Modellierung der Bindung zwischen verschiedenen Varianten von Inhibitormolekülen (basierend auf dem Inhibitor Eglin C) und der Substratbindungstasche wahrscheinliche Bindungsregionen vorhersagen [58].

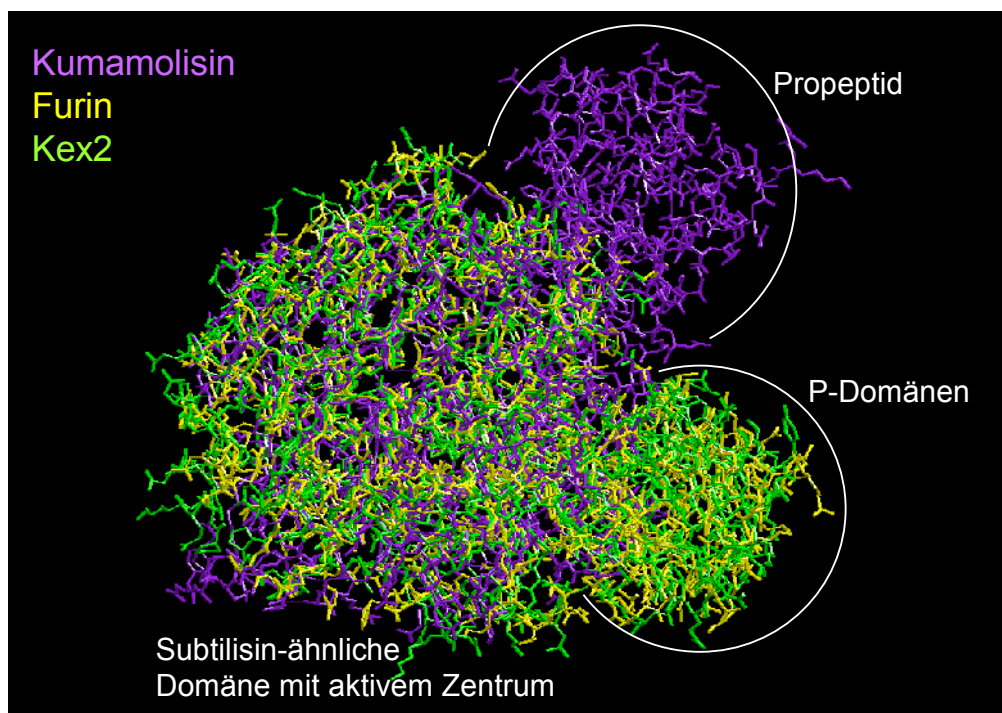
Hier soll mit einem anderen Ansatz versucht werden, Aussagen über die an einer möglichen P1' bis Pn' Erkennung beteiligten Aminosäuren zu treffen: das erste Substrat, welches von einer Kex2-ähnlichen Protease geschnitten wird, ist ihr eigenes Propeptid. Diese autokatalytische Reaktion wird durch denselben Mechanismus vermittelt, der auch für die Erkennung exogener Substrate verwendet wird. Daher sollte die Schnittstelle zwischen Propeptid und dem restlichen Protein einer optimalen Substratsequenz entsprechen. Durch Überlagerung der Strukturen von ScKex2 (PDB-Eintrag 1R64) und MmFurin (PDB-Eintrag 1P8J) mit der Subtilase Kumamolisin (PDB-Eintrag 1T1E) [139] aus dem thermo- und acidophilen Bakterium *Bacillus* novospec MN-32 kann die Ausrichtung der Substratseitenketten analysiert werden, da das zur Kristallisation verwendete Protein noch über sein eigenes Propeptid verfügt.

Die drei Strukturen 1R64, 1P8J und 1T1E wurden mit dem Programm MASS [95] überlagert und die resultierende Struktur (Abbildung 58) mit RasMol manuell inspiziert.

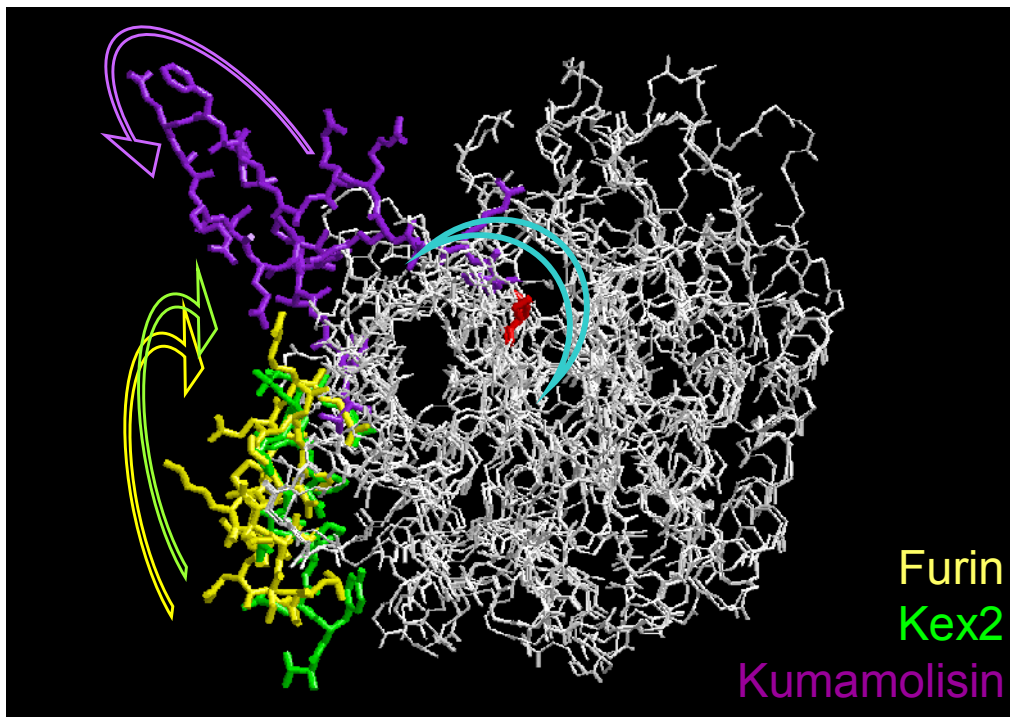


#### 4.5.2.1 *In silico* Analyse der S1'-S4' Substratbindungstaschen von Kex2 und Furin im Vergleich zu Kumamolisin

Der Vergleich der drei Raumstrukturen von ScKex2, MmFurin und Kumamolisin (Abbildung 58) zeigt zunächst, dass lediglich die aktive Subtilisin-ähnliche Domäne allen drei Strukturen gemein ist und den Kern des Alignments ausmacht. Die Prodomäne von Kumamolisin steht alleine, die beiden P-Domänen von Kex2 und Furin stehen gemeinsam aus dem Kern des Alignments heraus. Dies entspricht der Erwartung entsprechend dem Aufbau der Proteine.

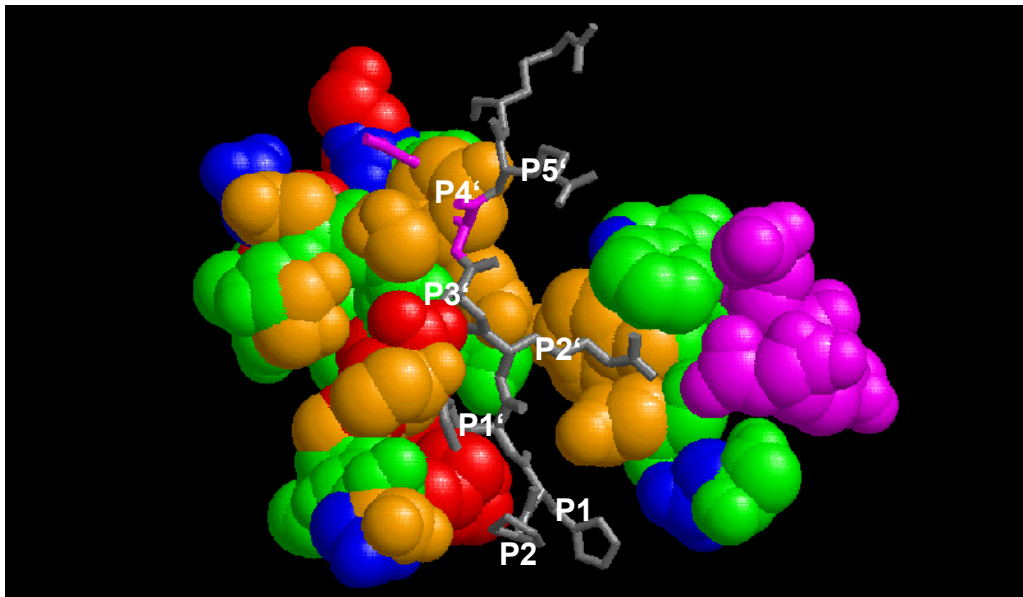


**Abbildung 58:** 3D-Strukturalignment der drei Subtilisin-ähnlichen Proteasen Furin (*M. musculus*), Kex2 (*S. cerevisiae*) und Kumamolisin (*Bacillus novospec* MN-32). Grün: Kex2 aus *S. cerevisiae* (1R64), gelb: Furin aus *M. musculus* (1P8J) und Kumamolisin aus *Bacillus novospec* MN-32 (1T1E). Das Alignment der drei Strukturen zeigt die starke Ähnlichkeit der aktiven Domänen und deutlich das Fehlen sowohl der Propeptide in Kex2 und Furin als auch das Fehlen der P-Domäne in Kumamolisin.



**Abbildung 59: Verlauf der Linkerdomänen entlang der aktiven Domäne vor und nach autokatalytischer Abspaltung des Propeptids.** Ein Vergleich des Verlaufs der Linkerdomänen, die die aktive Domäne mit dem Propeptid verbindet (Kumamolisin: lila), bzw. verbunden haben (Kex2: gelb, Furin: grün) zeigt, dass sich diese nach der Abtrennung der Prodomäne aus der Substratbindungsregion (türkis) bzw. dem aktiven Zentrum (rot: aktive Serine aller drei Proteasen) wieder entfernen und somit später nicht an einer Substraterkennung teilnehmen.

Aus dem Verlauf der Prodomäne mit der Erkennungssequenz durch das aktive Zentrum des Kumamolinsins ist ersichtlich, dass die autokatalytische Abspaltung der Prodomäne, wie erwartet, durch das aktive Zentrum und nicht durch einen anderen, unbekanntem Mechanismus erfolgt. Vergleicht man den Verlauf der sich an die Prodomäne anschließenden Domäne (im Folgenden Linkerdomäne genannt) des Kumamolinsins mit den entsprechenden Sequenzen von Kex2 und Furin (Abbildung 59), so wird deutlich, dass sich dieser Teil der Domäne nach der Abtrennung des Propeptids aus der Substratbindungsregion herausfaltet und nicht in dieser verbleibt. Diese Feststellung ist insofern wichtig, als dass man die Art der Substraterkennung des Kumamolinsins zu seiner eigenen Prodomäne auch auf andere Substrate übertragen kann. Gleichzeitig kann aus einer Struktur ohne die Prodomäne (so im Fall von Kex2 und Furin) nicht mehr auf die Substraterkennung an diesen Positionen geschlossen werden.



**Abbildung 60:** Verlauf der P2-P5' Substratdomäne des Kumamolisin Propeptids entlang der Kex2 und Furin Substratbindungsstasche. Das Strukturalignment zeigt, dass der theoretische Verlauf eines Substrats (grau) durch die Erkennungsregionen der P1'-P3' Positionen in Kex2 und Furin dem in der Literatur vorhergesagten entspricht. Eine andere Orientierung nimmt das Substrat an der P4' Position ein. (lila: vorhergesagter Bindungsbereich für Eglin C [58] und P4' Aminosäure der Linkerdomäne). Ab der Position P5' gibt es keine Interaktionen mehr.

Der Vergleich des Verlaufs der Schnittstelle zwischen Prodomäne und Linkerdomäne entlang der Substraterkennungsregionen S4-S1 von Kex2, Furin und Kumamolisin (Abbildung 60) zeigt, dass diese räumlich ähnlich aufgebaut sind und die Substratspezifität durch, wenn auch unterschiedliche, Aminosäuren generiert werden, die an analoger Position sowohl in der Struktur als auch im Sequenzalignment zu finden sind (Abbildung 60, [139]).

Bei Betrachtung der putativen Interaktionspartner der einzelnen Aminosäuren in den Linkerdomänen von Kex2 und Furin (Abbildung 61) lassen sich für die Bindungstaschen in Kex2 und Furin jeweils charakteristische Eigenschaften vorhersagen:

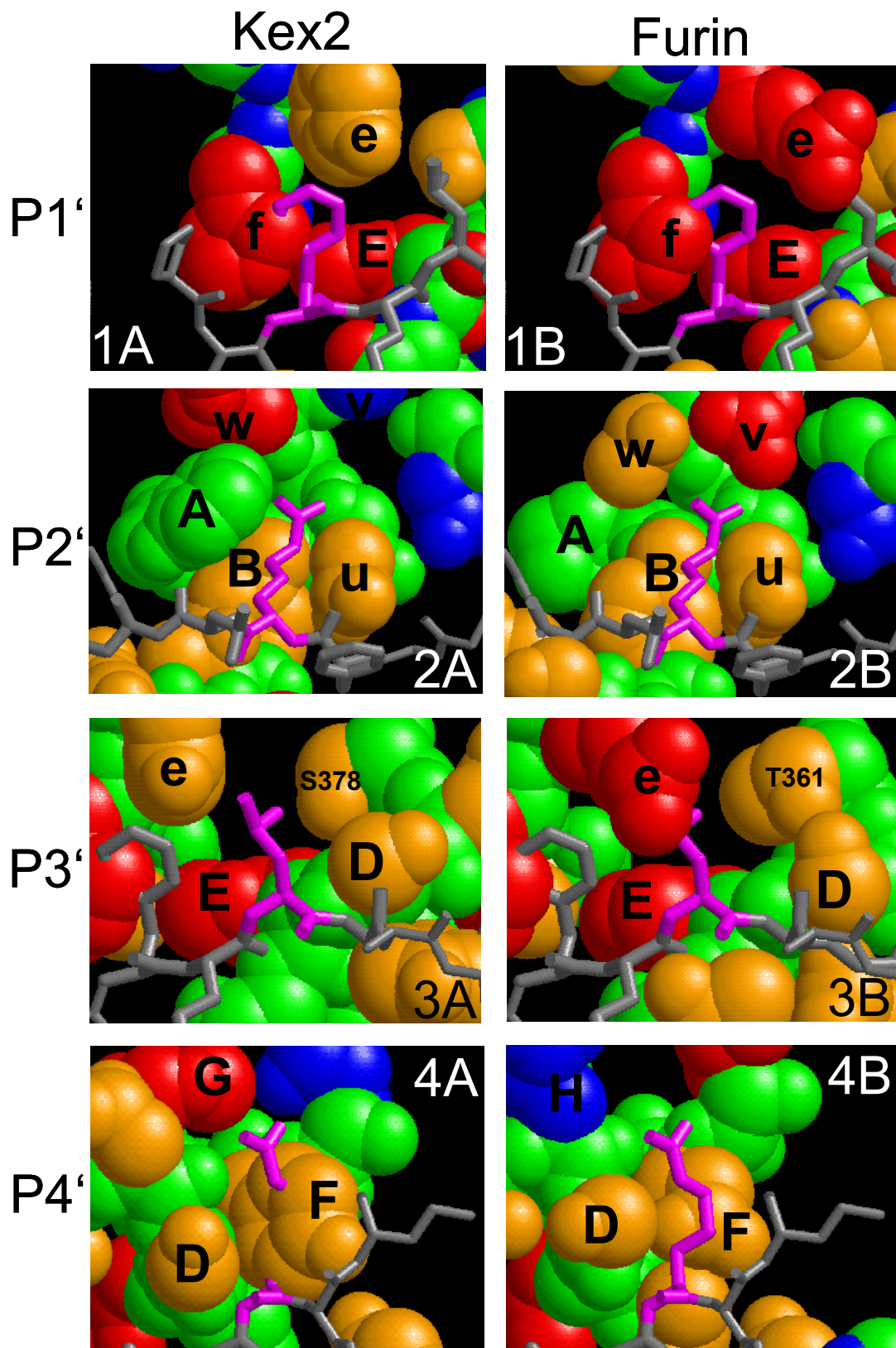
Die S1'-Tasche (Abbildung 61.1A und 1B) ist bei beiden Proteinen positiv geladen, zunächst durch jeweils zwei Histidine (Position f und E) und bei Furin zusätzlich durch ein weiteres Arginin, welches in Kex2 durch ein Tyrosin ersetzt ist (Position e). Außerdem ist diese Tasche in beiden Proteinen direkt neben der S2-Tasche gelegen. Hier würde eine positiv geladene Seitenkette des Substrats diese Ladung noch weiter verstärken.

Die putative S2'-Tasche (Abbildung 61 2A und 2B) verläuft entlang von polaren Aminosäuren besetzten Positionen (Positionen u, A und B) hin zu einem in beiden Proteinen unterschiedlich gebauten Bereich. Zwar ist in beiden Proteinen am Ende der Tasche eine positive Ladung durch ein Arginin zu finden, aber dieses befindet sich in zwei verschiedenen Positionen (Position w in Kex2, Position v in Furin). Bei Furin wird diese Ladung teilweise durch eine Glutaminsäure

(Position w) kompensiert, an der Position v in Kex2 findet sich die polare, aber ungeladene Aminosäure Threonin. Außerdem ist die Tasche bei Kex2 enger gebaut als bei Furin. Hier versperrt ein Prolin (Position A) den Zugang zu den terminal gelegenen Ladungen. Bei Furin befindet sich an analoger Position ein Tryptophan, welches aus der Tasche heraus steht.

Eine Bindung der P3'-Aminosäure in der S3'-Tasche, die direkt neben der S1'-Tasche gelegen ist, wird in beiden Proteinen auf ähnliche Weise vermittelt, wie die Bindung an P2'. Der Eingang zur Tasche wird in beiden Proteinen von einem Histidin (Position E) und einem Serin (Position D) flankiert, das Ende der Tasche wiederum ist unterschiedlich gebaut. Hier ist in Furin wieder ein Arginin dominant, wobei sich in Kex2 an analoger Stelle ein ungeladenes Tyrosin befindet.

Ein ähnlicher Unterschied findet sich auch in der putativen S4'-Tasche wieder. Wieder wird der Eingang der Tasche von zwei polaren, ungeladenen Aminosäuren (Serin und Threonin bzw. Glutamin, Positionen D und F) gebildet, wobei sich terminal eine geladene Aminosäure befindet. Bei Kex2 ist dies ein positiv geladenes Histidin (Position G), bei Furin eine negativ geladene Glutaminsäure. Interessant ist auch, dass beide Aminosäuren aus unterschiedlichen Kontexten in der Primärstruktur des Proteins kommen, obwohl sie in der Struktur eine fast identische Konformation annehmen.



**Abbildung 61: Vergleich der Struktur der an den S1'-S4'-Substratbindungsstellen beteiligten Regionen.** Grau: Kumamolisin Linkerdomäne, lila: betrachtete Kumamolisin Aminosäure. Seitenketten der Proteasen: rot: positiv geladene, blau: negativ geladene, orange: polare, grün: hydrophobe. Die Benennung der Seitenketten erfolgt gemäß der Auflistung in Tabelle 41.

#	Kex2	Furin	Tasche	Referenz
a	D175	D153	S2	[54]
b	D176	D154	S2	[54]
c	D210	(D191)	S2	[54]
d	D211	N192	S2	[54]
e	Y212	R193	S3'	[58, 140]
			S1'	Diese Arbeit
f	H213	H194	S1'	[140]
g	R216	R197	S3'	[140]
h	I245	M226	S4	[54]
i	L246	L227	S4	[54]
J	I250	V231	S4	[54]
k	T252	D233	S4, S6	[61]
l	D254	V235	S4	[54]
m	E255	E236	S4	[54, 61]
n	W273	W254	S4	[54]
o	A276	E257	S3, S5	[54, 140]
p	D277	D258	S1	[54]
q	D278	D259	S5	[140]
r	H281	T262	S5	[140]
s	Q283	D264	S4, S6	[61]
t	A311	A292	S1	[54]
u	N314	N295	S1	[60]
v	T317	R298	S2'	[58]
w	R318	E299	S2'	[58]
x	D320	D301	S1	[54]
y	D325	D306	S1	[54]
z	Y327	Y308	S4	[58]
A	P347	W328	S2'	[58, 140]
B	Y348	Y329	S2'	[58]
C	E350	E331	S1	[54]
F	Y367	Q350	S4'	Diese Arbeit
G	H369	-	S4'	Diese Arbeit
H	(N379)	E362	S4'	Diese Arbeit
D	S380	S363	S3'	[58]
			S4'	Diese Arbeit
E	H381	H364	S1'	[59, 140]

Tabelle 41: Aminosäuren, welche in *S. cerevisiae* Kex2 und *M. musculus* Furin zur Substratspezifität beitragen.



#### 4.5.2.2 Besetzung der an der Substraterkennung beteiligten Positionen bei Kex2 im Vergleich zu anderen Prohormonkonvertasen

Für die *in vitro* Proteolyse wurden Enzyme aus vier verschiedenen Hefen (*S. cerevisiae*, *C. albicans*, *C. glabrata* und *P. pastoris*) verwendet. Da sich die biochemischen Daten der Literatur jedoch lediglich auf Kex2 aus *S. cerevisiae* beziehen, es war daher nötig zu untersuchen, ob sich die Enzyme verschiedener Pilze in Regionen der Substraterkennung ähnlich genug sind, um diese Daten auf andere Spezies zu übertragen.

Der Vergleich der an der Substraterkennung beteiligten Bereiche zwischen den einzelnen Kex2-orthologen Proteinen (Abbildung 62) zeigt zunächst einen hohen Grad an Konservierung. Allerdings lassen sich bestimmte Aminosäuren aufzeigen, die zwischen Hemiascomyceten und Ascomyceten unterschiedlich sind: die Kex2-orthologen Proteasen der Ascomyceten tragen überall dort geladene Aminosäuren, wo sie entweder in Kex2 oder Furin vorkommen. Das spricht für eine wesentlich stärkere elektrostatische Selektion als bei jeweils Kex2 oder Furin.

Die Ladung der potentiellen S6-Tasche (Positionen k und s) der Kex2-Proteine von Ascomyceten entspricht mit zwei negativen Ladungen der von Furin. Die Hemiascomyceten besitzen an dieser Stelle zwei polare oder eine polare und eine hydrophobe Aminosäure.

Eine Interaktion von Kex2 mit der P5-Aminosäure würde potentiell durch eine Ladung an Position o, wie bei den Furinen (dort Glutaminsäure oder Asparaginsäure), unterstützt. Diese Position ist aber in allen Fällen durch hydrophobe oder schwach polare Aminosäuren besetzt.

Die S4-Tasche (Positionen h, i, j, k, l, m, n, s und z) der Ascomyceten ist wiederum wie die von Furin gestaltet. Die Positionen h, i, j, n und z tragen bei allen Proteasen hydrophobe oder schwach polare Reste; die Position m ist von einer konservierten Glutaminsäure besetzt. Unterschiede gibt es dagegen an den Positionen k, l und s. Bei den Hemiascomyceten ist an dieser Stelle nur eine negative Ladung (Position l) vertreten, bei den Furinen dagegen zwei an genau den anderen Positionen (Positionen k und s). Furine selektieren an P4 wesentlich stärker auf positive Ladungen, was genau auf diesen Unterschied zurückzuführen ist. Bei den Kex2-orthologen Proteinen aus Ascomyceten hingegen finden sich an allen drei Positionen negative Ladungen, was für eine noch weiter verstärkte Selektion auf positive Ladungen spricht.

Potentielle Interaktionen der Protease mit der P3-Aminosäure ergäben sich über die Positionen o, q und r. Hier wird vor allem bei Furin eine Stabilisierung von positiven Ladungen des Substrates an P3 durch die negative Ladung an Position o vermutet. Diese fehlt jedoch bei allen Kexinen, da sich hier sowohl polare als auch hydrophobe Aminosäuren befinden. Dagegen ähnelt sich der Aufbau an den Positionen q und r wieder stärker zwischen den Kexinen der Ascomyceten und

den Furinen der Säuger. Hier dominiert an Position q ein Aspartat, welches lediglich bei einigen Hemiascomyceten durch ein Asparagin ersetzt ist. An Position r finden sich bei den Furinen und den Kexinen der Ascomyceten Serine und Threonine, bei den Hemiascomyceten dagegen sehr verschiedene Aminosäuren, bis hin sowohl zu negativ als auch positiv Geladenen.

Die S2-Tasche ist hoch konserviert gebaut (Positionen a, b, c und d). Die stärkere Selektivität an dieser Position auf positive Ladungen bei ScKex2 als bei MmFurin liegt in dem Aspartat/Asparagin-Austausch an Position d begründet. Ein Aspartat findet sich auch bei den Ascomyceten.

Noch konservierter als die S2-Tasche ist die S1-Tasche (Positionen p, t, u, x, y und C). Dies gilt für alle Prohormonkonvertasen.

Im Gegensatz zu allen anderen Taschen weist die putative S1'-Tasche (Positionen e, f und E) positive Ladungen auf und selektiert auf diese Weise auf negative Ladungen im Substrat. Die beiden Histidine an den Positionen f und E sind in allen Prohormonkonvertasen konserviert. Die Position e wird in den Hemiascomyceten von dagegen lediglich von polaren Aminosäuren besetzt.

Die putative S2'-Tasche (Positionen v, w, A und B) ist bei den Furinen ausgeprägter als bei den Kexinen der Pilze. Allerdings heben sich bei den Furinen möglicherweise die konservierten Ladungen an den Positionen v und w gegenseitig auf, so dass die S2'-Tasche der Pilze mit ihren einzelnen, auf die Positionen v und w verteilten positiven Ladungen eine höhere Selektivität gegenüber negativen Ladungen ausprägen könnte. Die Positionen A und B sind konserviert von polaren, ungeladenen und teilweise von hydrophoben Aminosäuren besetzt.

Wie bereits für die S1'-Tasche gezeigt, ist die Position e auch für die S3'-Tasche (Positionen e und D) ausschlaggebend. Sie ist in den Furinen, wie den Kexinen der Ascomyceten, durch das positiv geladene Lysin besetzt, während in *Saccharomycetales* hier verschiedene polare Aminosäuren zu finden sind. Die die Tasche begrenzende Position D ist nicht konserviert.

Eine putative S4'-Tasche wird von den Positionen D und F flankiert. Beide Positionen sind nicht konserviert, aber es fällt auf, dass hier bei den Hemiascomyceten häufig wenigstens an einer Stelle eine positive Ladung zu finden ist. Der Boden der S4'-Tasche wird in Kex2 von der Position G besetzt, in den Furinen von Position H. Das Sequenzalignment legt nahe, dass hier die Kexine der Pilze gleich gebaut sind und durch die konservierte positive Ladung auf langkettige Aminosäuren mit negative Ladungen selektiert (Glutaminsäure und Asparaginsäure). Bei den Furinen ist an Position H dagegen eine negative Ladung konserviert, was für eine Bevorzugung von positiven Ladungen (Lysin und Arginin oder Histidinen) spricht.



Kex2	X	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	A	B	C	F	G	H	D	E
<i>Pneumocystis carinii</i>	QKLVKR-ENRG	D	N	Y	N	N	H	R	Y	I	H	S	N	E	W	A	D	T	L	Q	G	N	Y	F	D	D	Y	M	F	E	Y	Y	F	L	H
<i>Pneumocystis jereveci</i>	KQHLKR-RYRK	D	D	E	D	T	H	R	I	L	I	D	D	E	W	D	D	D	T	D	A	N	I	Y	D	D	Y	K	Y	E	Y	S	R	H	
<i>Pneumocystis muris</i>	LRREKR-NVIR	D	D	D	D	T	H	R	F	L	L	S	L	E	W	R	D	D	T	E	G	N	Y	Y	D	D	Y	Y	F	E	R	Y	T	V	H
<i>Schizosaccharomyces pombe</i>	RWRVKR-DASE	D	D	D	D	Q	H	R	I	L	I	D	V	E	W	A	D	D	A	D	A	N	H	Y	D	D	Y	F	Y	E	S	L	R	S	H
<i>Cryptococcus neoformans</i>	KQRAKR-HKSY	D	D	D	D	Q	H	R	I	L	I	D	D	E	W	P	D	D	S	E	A	N	G	S	D	D	Y	Y	Y	E	H	H	S	H	
<i>Saccharomyces cerevisiae</i>	NDLFKR-LPVP	D	D	D	D	Y	H	R	I	L	I	T	D	E	W	A	D	D	H	Q	A	N	T	R	D	D	Y	P	Y	E	Y	H	N	S	H
<i>Candida albicans</i>	KKLSKR-IPVP	D	D	D	D	Y	H	R	I	L	I	S	D	E	W	T	D	N	V	S	A	N	R	F	D	D	Y	Q	Y	E	H	H	A	T	H
<i>Candida glabrata</i>	NKLQKR-MPVP	D	D	D	D	N	H	R	I	L	L	A	D	E	W	T	D	D	H	Q	A	N	M	H	D	D	Y	P	Y	E	Y	H	D	R	H
<i>Pichia pastoris</i>	HYLHKR-APVV	D	D	D	D	H	H	R	I	L	I	P	D	E	W	A	D	N	T	Q	A	N	S	S	D	D	Y	P	Y	E	H	H	D	T	H
<i>Kluyveromyces lactis</i>	VQLHKR-LPIG	D	D	D	D	Y	H	R	I	L	I	A	D	E	W	S	D	D	T	Q	A	N	M	F	D	D	Y	P	Y	E	Y	K	N	T	H
<i>Ashbya gossipii</i>	HRLAKR-MPIR	D	D	D	D	T	H	R	I	L	I	P	D	E	W	T	D	N	E	Q	A	N	A	L	D	D	Y	T	Y	E	F	R	D	H	H
<i>Candida dubliensis</i>	KKLSKR-IPVP	D	D	D	D	Y	H	R	I	L	I	S	D	E	W	T	D	N	V	S	A	N	R	F	D	D	Y	Q	Y	E	H	H	A	T	H
<i>Debaryomyces hanseni</i>	KRLRKR-MPVP	D	D	D	D	Y	H	R	I	L	I	A	E	E	W	T	D	D	T	S	A	N	R	Y	D	D	Y	M	Y	E	H	H	A	L	H
<i>Saccharomyces paradoxus</i>	NDLFKR-LPVP	D	D	D	D	Y	H	R	I	L	I	T	D	E	W	A	D	D	H	Q	A	N	S	R	D	D	Y	P	Y	E	Y	H	N	S	H
<i>Yarrowia lipolytica</i>	RRIHKR-APVN	D	D	D	D	Y	H	R	I	L	I	E	I	E	W	P	D	N	T	A	A	N	S	R	D	D	Y	Y	Y	E	Y	V	N	Q	H
<i>Fusarium graminearum</i>	QHLFKR-DVPL	D	D	D	D	R	H	R	I	L	I	D	D	E	W	S	D	D	T	E	A	N	G	Q	D	D	Y	Y	Y	E	A	H	K	A	H
<i>Aspergillus fumigatus</i>	RKLHKR-VPPT	D	D	D	D	R	H	R	I	L	I	D	D	E	W	Y	D	D	T	E	A	N	I	H	D	D	Y	P	Y	E	A	H	T	T	H
<i>Aspergillus nidulans</i>	QRLVKR-TPPP	D	D	D	D	R	H	R	I	L	Y	T	P	T	W	P	D	D	T	D	A	N	A	S	D	D	Y	Q	Y	E	A	S	S	V	H
<i>Aspergillus oryzae</i>	QRLVKR-APPT	D	D	D	D	R	H	R	I	L	I	D	D	E	W	P	D	D	T	E	A	N	G	Y	D	D	Y	S	Y	E	A	H	S	L	H
<i>Aspergillus niger</i>	RKLHKR-VPPT	D	D	D	D	R	H	R	I	L	I	D	D	E	W	Y	D	D	T	E	A	N	I	H	D	D	Y	P	Y	E	A	H	T	T	H
<i>Magnaporthe grisea</i>	SPWPKR-VPPT	D	D	D	D	R	H	R	I	L	I	D	D	E	W	P	D	D	S	D	A	N	A	N	D	D	Y	Y	Y	E	A	H	N	G	H
<i>Neurospora crassa</i>	KPWEKR-MVPR	D	D	D	D	K	H	R	I	L	I	D	D	E	W	P	D	D	S	D	A	N	G	N	D	D	Y	Y	Y	E	S	H	D	T	H
<i>Paracoccidioides brasiliensis</i>	PFMSKR-APPL	D	D	D	D	K	H	R	I	L	V	D	D	E	W	V	D	D	T	D	A	N	L	H	D	D	Y	Y	Y	E	A	H	T	K	H
<i>Histoplasma capsulatum</i>	HYLHKR-APVV	D	D	D	D	K	H	R	I	L	V	D	D	E	W	V	D	D	T	E	A	N	L	H	D	D	Y	Y	Y	E	A	H	T	R	H
<i>Trichoderma reesei</i>	NHLQKR-IIPP	D	D	D	D	R	H	R	I	L	I	D	D	E	W	P	D	D	S	E	A	N	A	S	D	D	Y	Y	Y	E	G	H	G	M	H
<i>Stagonosporum nodorum</i>	PRHWKR-SVIP	D	D	D	D	R	H	R	I	L	I	D	D	E	W	P	D	D	T	Q	A	N	A	S	D	D	Y	Y	Y	E	S	H	S	Q	H
<b>Furin</b>																																			
<i>Mus musculus</i>	KRRAKR-DVYQ	D	D	D	N	R	H	R	M	L	V	D	V	E	W	E	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	Q	V	E	S	H
<i>Rattus norvegicus</i>	KQRAKR-DVYQ	D	D	D	N	R	H	R	M	L	V	D	V	E	W	E	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	Q	V	E	S	H
<i>Homo sapiens</i>	KRRTKR-DVYQ	D	D	D	N	R	H	R	M	L	V	D	V	E	W	E	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	Q	V	E	S	H
<i>Aplysia californica</i>	KRSRKR-DFHP	D	D	E	N	R	H	R	M	L	V	D	V	E	W	D	D	D	V	D	A	N	S	A	D	D	Y	W	Y	E	Q	A	T	T	H
<i>Bos taurus</i>	KRRAKR-DIYQ	D	D	D	N	R	H	R	M	L	V	D	V	E	W	E	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	Q	V	E	S	H
<i>Cricetulus griseus</i>	KRRTKR-DVYQ	D	D	D	N	R	H	R	M	L	V	D	V	E	W	E	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	Q	V	E	S	H
<i>Gallus gallus</i>	KRRTKR-DIFM	D	D	D	N	R	H	R	M	L	V	D	V	E	W	E	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	Q	V	E	S	H
<i>Lymnea stagnalis</i>	KSRQKR-SVHF	D	D	E	N	R	H	R	M	L	V	D	V	E	W	D	D	D	T	D	A	N	R	D	D	Y	W	Y	E	Q	V	E	T	H	
<i>Xenopus laevis</i>	KKRKKR-DIYT	D	D	D	N	R	H	R	M	L	V	D	V	E	W	E	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	Q	V	D	S	H
<i>Bombix mori</i>	LSRKKR-DFQI	D	D	S	N	R	H	R	M	L	V	D	V	E	W	D	D	D	T	D	A	N	K	E	D	D	Y	W	Y	E	Q	V	A	G	H
<b>PACE4</b>																																			
<i>Homo sapiens</i>	KRRVKR-QVRS	D	D	E	N	K	H	R	M	L	V	D	V	E	W	D	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	K	V	D	G	H
<i>Mus musculus</i>	KRRVKR-QARS	D	D	E	N	K	H	R	M	L	V	D	V	E	W	D	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	K	V	D	G	H
<i>Rattus norvegicus</i>	KRRVKR-QARS	D	D	E	N	K	H	R	M	L	V	D	V	E	W	D	D	D	T	D	A	N	R	E	D	D	Y	W	Y	E	K	V	D	G	H
<b>PC1-3</b>																																			
<i>Homo sapiens</i>	KERSKR-SALR	D	D	E	N	K	H	R	M	L	V	D	I	E	W	N	D	D	T	E	A	N	R	Q	D	D	Y	W	Y	E	R	T	E	T	H
<i>Mus musculus</i>	KERSKR-SVQK	D	D	E	N	K	H	R	M	L	V	D	I	E	W	N	D	D	T	E	A	N	R	Q	D	D	Y	W	Y	E	R	T	E	T	H
<i>Rattus norvegicus</i>	KERRKR-SVPR	D	D	E	N	K	H	R	M	L	V	D	I	E	W	N	D	D	T	E	A	N	R	Q	D	D	Y	W	Y	E	R	T	E	T	H
<b>PC2</b>																																			
<i>Homo sapiens</i>	FDRKKR-GYRD	D	D	F	N	S	H	R	M	L	M	D	I	E	W	T	D	N	T	D	A	D	.	S	D	D	Y	L	Y	E	G	A	L	R	H
<i>Mus musculus</i>	FDRKKR-GYRD	D	D	F	N	S	H	R	M	L	M	D	I	E	W	T	D	N	T	D	A	D	.	S	D	D	Y	L	Y	E	G	A	L	R	H
<i>Rattus norvegicus</i>	FDRKKR-GYRD	D	D	F	N	S	H	R	M	L	M	D	I	E	W	T	D	N	T	D	A	D	.	S	D	D	Y	L	Y	E	G	A	L	R	H

Abbildung 62: Vergleich der zur Substratspezifität beitragenden Aminosäuren bei Kex2-, Furin und homologen Proproteinkonvertasen. Rot: positiv geladene, blau: negativ geladene, orange: polare, grün: apolare Aminosäuren.

## 4.6 Strukturelle Determinanten der Proteasesuszeptibilität im Substrat

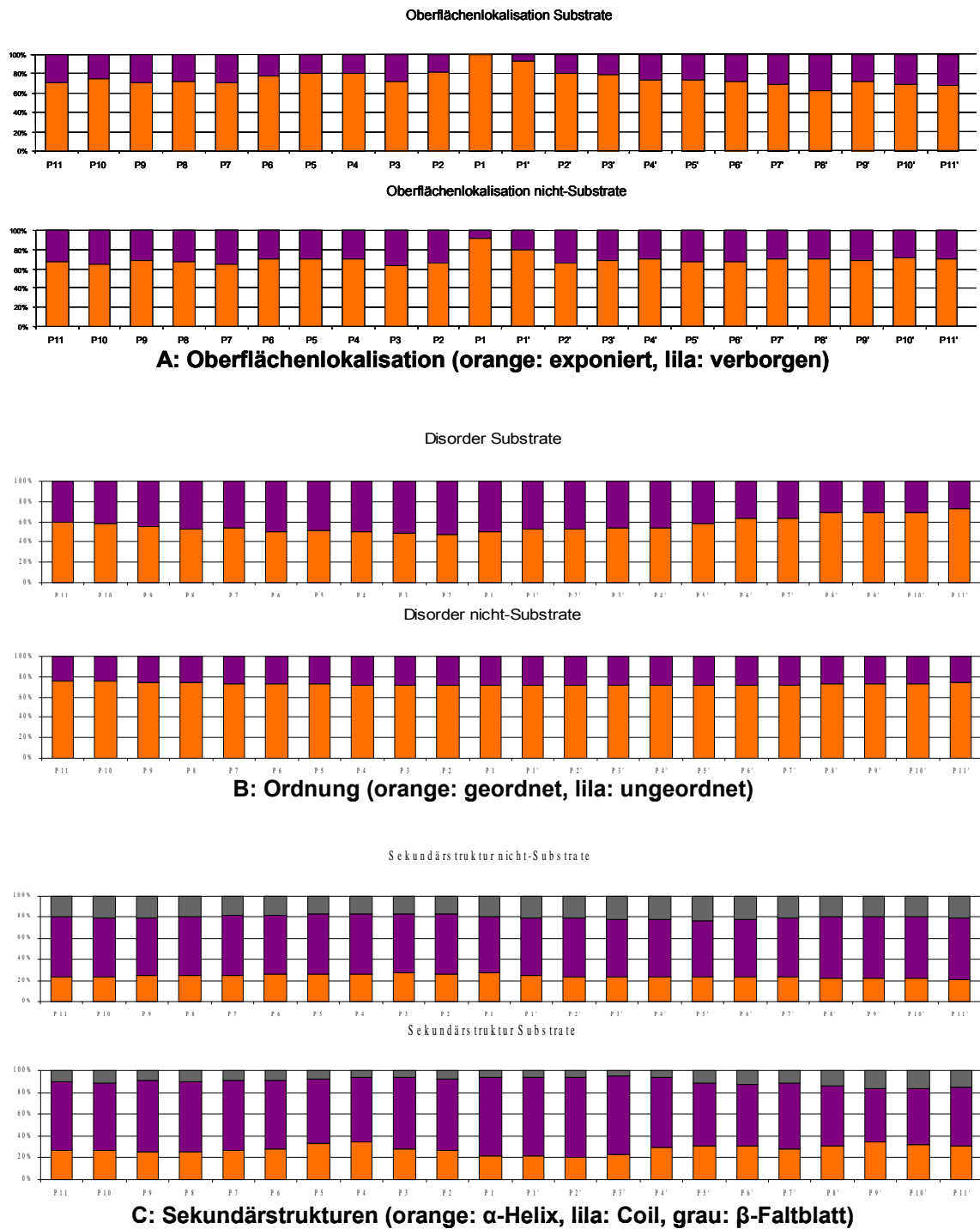
### 4.6.1 Die ausgebildete Struktur der Schnittstelle beeinflusst die Prozessierung

Wie unter Abschnitt 4.4.6 gezeigt, werden nicht alle durch ein vorhandenes Aminosäuremotiv gekennzeichneten potentiellen Schnittstellen eines Proteins auch tatsächlich von einer entsprechenden Protease prozessiert. Es kann daher postuliert werden, dass, bevor die Aminosäuren der Schnittstelle mit dem aktiven Zentrum der Protease interagieren können, diese erst für die Protease zugänglich werden müssen.

Im Folgenden soll daher überprüft werden, ob sich verschiedene Algorithmen der Strukturvorhersage generell eignen, um Proteaseschnittstellen vorherzusagen. Es gibt zur Zeit für drei prinzipiell verschiedene strukturelle Eigenschaften von Proteinen Vorhersage Algorithmen:

1. Sekundärstruktur von Proteinabschnitten ( $\alpha$ -Helix/ $\beta$ -Faltblatt/Coil)
2. Lokalisation von Aminosäuren an der Proteinoberfläche (verdeckt/exponiert)
3. Grad der Ordnung in Proteinabschnitten (ungeordnet/geordnet)

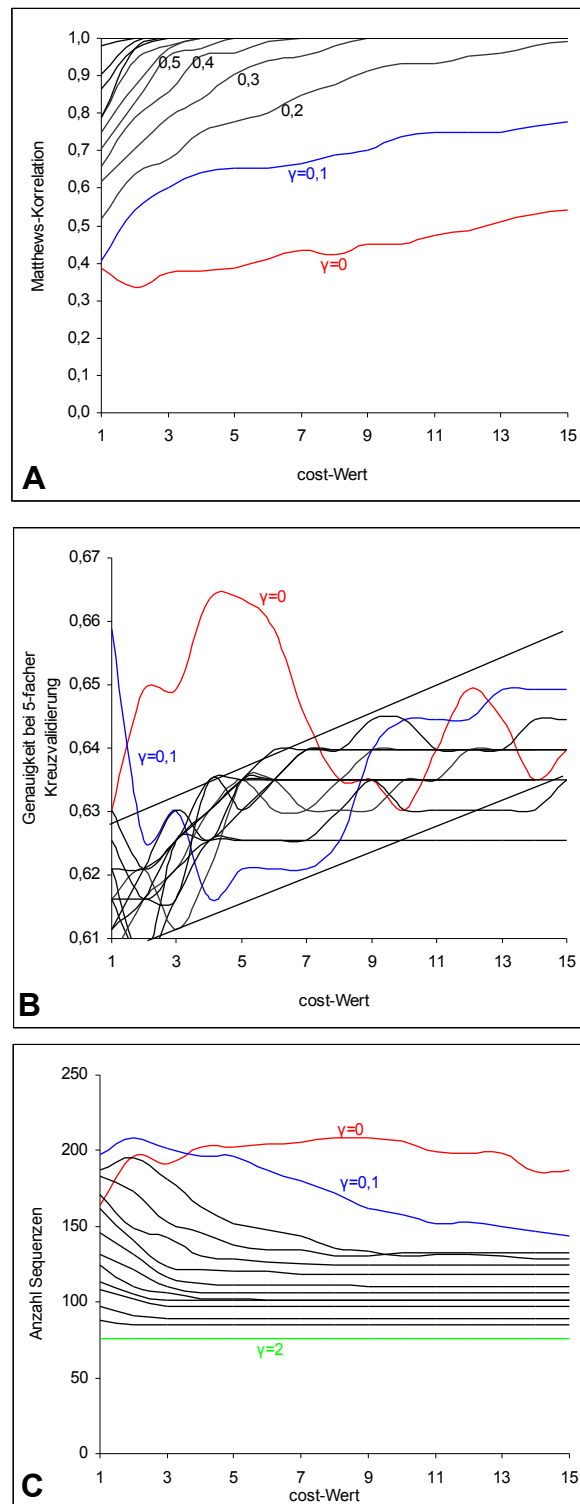
Für die einzelnen Proteine der Testsets wurden mit den Programmen Prof/PhD (Sekundärstruktur), rvp-net (Oberflächenlokalisierung) und Disopred2 (Unordnung) Strukturvorhersagen gemacht und getestet, ob im Bereich von P11-P12' um die geschnittene Bindung herum gehäuft die betrachteten Strukturelemente zu finden sind. Dazu wurden die Arginine nach geschnitten/nicht geschnitten aufgeteilt und die Frequenz der einzelnen Strukturelemente an den Positionen P11 bis P12' berechnet. Dabei ergaben sich für alle drei untersuchten Elemente signifikante Unterschiede zwischen Substraten und Nicht-Substraten (Abbildung 63). Allgemein sind die Aminosäuren in Substraten stärker exponiert (Abbildung 63A, orange), stärker ungeordnet (Abbildung 63B, lila) und prägen weniger  $\beta$ -Faltblatt und dafür mehr Coil-Strukturen aus (Abbildung 63C, grau und orange). Besonders auffällig ist, dass in Substraten das Arginin an P1 immer exponiert ist. Dies gilt verstärkt auch für die benachbarte P1' Position.



**Abbildung 63:** Assoziationen von Strukturelementen mit Schnittstellen in verschiedenen Substratsets. A: Verteilung von exponiert/verborgen, B: Verteilung von Ordnung/Unordnung, C: Verteilung von  $\alpha$ -Helix/Coil/ $\beta$ -Faltblatt.

#### 4.6.2 Berechnung der Verbesserung von Vorhersagealgorithmen durch Berücksichtigung struktureller Determinanten

Als abschließendes Experiment dieser Arbeit wurde getestet, ob sich die Vorhersage von Kex2-Schnittstellen mit den gegebenen Methoden (hier Support-Vektoren) unter Berücksichtigung der oben beschriebenen strukturellen Merkmal verbessern lässt. Dazu wurden die in dieser Arbeit experimentell gefundenen Schnittstellen mit den aus der Literatur bekannten zu einem Datensatz vereinigt und in die Klassen „ProHormon“ (PH) und „ProProtein“ (PP) aufgeteilt. Alle nicht geschnittenen Arginine dieser Proteine mit einem PSSM-Wert von über 10 wurden als Negativbeispiele verwendet. Tatsächlich lässt sich unter Verwendung aller durch die drei Strukturalgorithmen gegebenen Werte in einem Fenster von P8-P6' der Substrate problemlos ein SVM-Modell mit 100% Spezifität und 100% Sensitivität durch Erhöhen der Parameter „gamma“ und „cost“ gegenüber den Standardbedingungen (radialer Kernel, cost=1, gamma=1) generieren (Abbildung 64A). Dieses bewertet unter den in Abschnitt 4.3.1 dargestellten Substraten 76 als Motive mit einer zu den bekannten Substraten ähnlichen Struktur (Abbildung 64C). Optimale Bedingungen unter Kreuzvalidierungsbedingungen waren bei den Parametern gamma=1, cost =0 (65,87 % Genauigkeit) und bei gamma=0,1, cost=5 (66,35% Genauigkeit) erreicht (Abbildung 64B). Unter diesen Bedingungen wurde allerdings auch das Maximum an als „ähnlich“ klassifizierten Motiven erreicht (Abbildung 64C), so dass für die Vorhersage auch die Werte mit minimaler Menge (gamma=1,25, cost=2 und gamma=2, cost=15) herangezogen wurden. Die Bewertung der Struktur um die einzelnen Motive sind in Tabelle 42 dargestellt.



**Abbildung 64: Training eines Support Vektor-Modells mit strukturellen Daten der bekannten Kex2-Schnittstellen.** A: Durch Erhöhung der Parameter „cost“ (X-Achse) und  $\gamma$  (verschiedene Wertereihen) lassen sich Modelle generieren, die eine Matthews-Korrelation von 1 für die Trainingsdaten aufweisen. B: Bei einer 5-fachen Kreuzvalidierung zeigen lediglich die Parameterpaare  $\gamma=0$  | cost= 5 und  $\gamma=0,1$  | cost =0 maximale Genauigkeit des Modells. C: Unter den optimalen Bedingungen sind die Anzahl positiv klassifizierter Motive (Klassen „ProHormon“ und „ProProtein“) maximal.

	Proteinae/ Sc-Homolog	ORF Nummer	Position	Score	Klassifikation	ProProtein		ProHormon		Kein Substrat		Kompartiment	Art	Beschreibung
						P	+/-	P	+/-	P	+/-			
Richtig-Positiv	CA6162	CA6162	200	653	PP	0,45	0,11	0,10	0,01	0,45	0,12	Unbekannt		Opaque Phase Protein OP4 ähnlich
	Cis3	CA1334	135	746	PP	0,43	0,09	0,08	0,03	0,49	0,07	ZW	PIR	Pir-Protein
		CA3614	121	410	PP	0,48	0,10	0,08	0,01	0,44	0,09			
	Kar2	CA0915	308	954	PP	0,50	0,30	0,07	0,01	0,43	0,30	Unbekannt		DnaK-Typ Chaperon
		CAGL0D02948g	295	954	PP	0,52	0,32	0,06	0,01	0,43	0,32			
	Kex2	YNL238W	79	59	PP	0,36	0,06	0,20	0,09	0,44	0,14	Golgi	IMC	Subtilisin-ähnliche Serinprotease
	MFa2	YGL089C	79	68	PH	0,14	0,11	0,45	0,30	0,41	0,21	Sekretorisch	Prohormon	α-Kreuzungsspheromon
	Pir3	CA0856	198	63	PP	0,52	0,13	0,17	0,02	0,31	0,12	Zellwand	PIR	Pir-Protein
		CA0856	211	54	PP	0,36	0,11	0,28	0,17	0,35	0,23			
		CA2912	119	410	PP	0,37	0,18	0,14	0,07	0,49	0,13			
CA2912		129	119	PP	0,62	0,13	0,12	0,08	0,27	0,06				
Sap98	CA0549	40	130	PP	0,45	0,10	0,17	0,05	0,38	0,13	Sekretorisch	IMC	putative sekretorische Aspartylprotease	
Sun41	CA0883	45	678	PP	0,55	0,32	0,07	0,03	0,38	0,30	ZW/PM	His-Box	putative Zellwand β-Glukosidase	
Tos1	CA2303	123	848	PP	0,49	0,14	0,07	0,02	0,44	0,14	ZW/PM	His-Box	putative Ankeruntereinheit von α-agglutinin	
Falsch-Positiv	Ifo2	CA1394	31	25	PP	0,54	0,33	0,06	0,02	0,39	0,32	Sekretorisch		Unbekannte Funktion
	Pga17	CA4679	338	293	PP	0,58	0,29	0,14	0,11	0,27	0,22	ZW/PM		Unbekannte Funktion
	Pry1	CAGL0F05137g	98	346	PP	0,63	0,26	0,07	0,03	0,30	0,24	Unbekannt		Pr-1 verwandtes Protein
	Sap9	CA4700	105	185	PP	0,66	0,31	0,08	0,07	0,25	0,24	ZW/PM	IMC	GPI-verankerte Aspartatprotease 9
Proteasen	Kex2	CA2633	526	151	PP	0,32	0,10	0,23	0,09	0,45	0,10	Golgi	shedding	Subtilisin-ähnliche Serinprotease
	Prb1	CAGL0L13266g	45	247	PP	0,35	0,07	0,17	0,12	0,49	0,12	Vakuole	N-terminal	Vakuoläre Protease B
		YEL060C	114	70	PP	0,29	0,08	0,25	0,08	0,46	0,10		IMC	
	Pep4	CAGL0M02211g	188	522	PP	0,44	0,23	0,09	0,04	0,47	0,19		IMC	
			68	59	PP	0,58	0,09	0,12	0,01	0,31	0,09	Vakuole	IMC	Vakuoläre Aspartatprotease A
Ste13	YOR219C	162	59	PP	0,61	0,06	0,05	0,03	0,34	0,05	Golgi	shedding	Dipeptidyl-Aminopeptidase	

**Tabelle 42: Bewertung der Struktur um die einzelnen potentiellen Schnittstellen (Seiten 111-115).** Aufgelistet sind alle Motive, die durch den Algorithmus als ähnlich zu den Trainingsdaten bewertet werden. Richtig-Positiv: bekannte Schnittstellen, die nicht im Trainingsset enthalten waren, Falsch-Positiv: Motive für die keine Prozessierung beobachtet werden konnte. Alle anderen Motive sind nach funktionellen Gruppen der Proteine geordnet.

	Proteinae/ Sc-Homolog	ORF Nummer	Position	Score	Klassifikation	ProProtein		ProHormon		Kein Substrat		Kompartiment	Art	Beschreibung
						P	+/-	P	+/-	P	+/-			
ER-Residente	Gpi16	YHR188C	276	59	PP	0,32	0,08	0,24	0,11	0,44	0,15	ER		Untereinheit des GPI-Transamidase Komplex
	Gpi8	CAGL0M01298g	128	434	PP	0,42	0,09	0,09	0,03	0,49	0,09	ER		Untereinheit des GPI-Transamidase Komplex
	Hrd3	CAGL0J08756g	267	68	PP	0,41	0,14	0,18	0,08	0,41	0,19	ER		Teil der ER-lokalisierten Qualitätskontrolle
		CAGL0J08756g	718	95	PP	0,39	0,09	0,26	0,11	0,35	0,17			
		YLR207W	266	158	PP	0,44	0,13	0,07	0,02	0,49	0,13			
		YLR207W	269	118	PP	0,47	0,03	0,18	0,05	0,35	0,04			
	Nvj1	CAGL0M07205g	185	265	PP	0,41	0,16	0,26	0,09	0,33	0,19	ER		Beteiligt and der Bildung von Kontakten zwischen Nukleus und Vakuole
	Orm1	CA0894	105	341	PP	0,53	0,16	0,07	0,02	0,40	0,16	ER		Unbekannte Funktion
	Pdi1	CAGL0J02068g	128	301	PH	0,25	0,03	0,29	0,18	0,46	0,15	ER		Disulfidisomerase
	Rot1	CAGL0L10670g	207	102	PP	0,51	0,11	0,08	0,04	0,41	0,10	ER	Shedding	Co-Chaperon
		YMR200W	226	130	PP	0,49	0,08	0,07	0,02	0,43	0,06			
	Scs3	YGL126W	220	140	PP	0,38	0,09	0,12	0,03	0,49	0,10	ER		Protein der Inositol-Phospholipidsynthese
	Ted1	CAGL0L00979g	337	59	PP	0,50	0,05	0,22	0,05	0,28	0,06	ER		Unbekannte Funktion
Tyw1	YPL207W	155	96	PP	0,49	0,14	0,17	0,05	0,34	0,17	ER		Protein der Wybutosinsynthese	
YKL027w	CA2730	43	151	PP	0,40	0,16	0,12	0,03	0,48	0,19	ER		Molybdopterin-modifizierendes Enzym	
YOR238W	YOR238W	399	426	PP	0,33	0,18	0,24	0,11	0,43	0,19				
	YOR238W	253	48	PP	0,33	0,06	0,23	0,06	0,44	0,09	ER		Unbekannte Funktion	
Hlj1	CAGL0H04499g	193	140	PP	0,41	0,12	0,12	0,03	0,47	0,11	ER	Shedding	Teil der ER-lokalisierten Qualitätskontrolle Ähnlich zu <i>E. coli</i> DnaJ	
Protein Mannosylierung	Mnn10	YDR245W	109	142	PP	0,40	0,10	0,18	0,05	0,42	0,14	Golgi		Galaktosyltransferase
	Mnn6	CA0951	187	96	PP	0,43	0,13	0,08	0,03	0,49	0,13	Golgi	Shedding	$\alpha$ -1,2-Mannosyltransferase
	Mnt2	CA1820	307	39	PP	0,61	0,13	0,09	0,01	0,30	0,13	Golgi		Mannosyltransferase
			607	93	PP	0,33	0,03	0,23	0,10	0,44	0,13			
	Mnt3	CAGL0H09240g	505	93	PP	0,42	0,20	0,10	0,04	0,48	0,23	Golgi		$\alpha$ -1,3-Mannosyltransferase
Mnt4	CA2418	154	392	PP	0,46	0,13	0,13	0,03	0,41	0,14	Golgi	Shedding	Mannosyltransferase	

	Proteinae/ Sc-Homolog	ORF Nummer	Position	Score	Klassifikation	ProProtein		ProHormon		Kein Substrat		Kompartiment	Art	Beschreibung
						P	+/-	P	+/-	P	+/-			
Polypeptid Vorläufer	Ddr48	CA4336	16	99	PP	0,49	0,21	0,16	0,06	0,35	0,16	Sekretorisch (nicht-klassisch)	Prohormon	Stressantwortprotein
			27	70	PP	0,49	0,18	0,17	0,03	0,34	0,16			
			54	138	PP	0,53	0,13	0,07	0,03	0,40	0,15			
			73	88	PP	0,53	0,14	0,13	0,03	0,34	0,12			
			154	138	PP	0,57	0,12	0,09	0,03	0,34	0,13			
			181	70	PP	0,58	0,08	0,15	0,06	0,28	0,05			
	200	56	PP	0,54	0,13	0,09	0,03	0,37	0,15					
	YDR262w	CA3122	41	156	PP	0,54	0,05	0,13	0,08	0,33	0,07	ZW/PM	Prohormon	Erhöhte Expression bei Zymolyasebehandlung
			68	420	PP	0,36	0,19	0,20	0,10	0,43	0,20			
			92	114	PP	0,37	0,17	0,16	0,06	0,46	0,18			
			116	271	PP	0,42	0,07	0,16	0,10	0,42	0,14			
161			94	PP	0,55	0,12	0,09	0,03	0,36	0,15				
186	211	PP	0,44	0,17	0,16	0,07	0,40	0,16						
	CA3122	225	99	-	0,37	0,11	0,09	0,02	0,54	0,11				
	CAGL0M08734g	34	43	PP	0,49	0,13	0,06	0,02	0,45	0,12				
	YDR262W	44	85	-	0,23	0,04	0,22	0,11	0,54	0,07				
Zellwand- aufbau	CA3295	CA3295	551	123	PP	0,41	0,09	0,12	0,02	0,47	0,11	ZW/PM	Prohormon	Opaque Phasen spezifisch
	Chs2	CA5537	22	235	PP	0,46	0,10	0,18	0,05	0,37	0,11	Plasmamembran		Chitinsynthase
	Chs3	YBR023C	85	92	PP	0,62	0,13	0,08	0,01	0,30	0,13	Verschiedene		Chitinsynthase 3
	Crh12	CA1835	393	79	PP	0,46	0,12	0,06	0,04	0,48	0,08	ZW/PM	Shedding	Chitintransglykosidase
	Dan4	YJR151C	1077	64	PP	0,45	0,12	0,05	0,03	0,49	0,15	ZW/PM	Shedding	Nur unter anaeroben Bedingungen exprimiert
	Sps22	CAGL0B00616g	399	277	PP	0,38	0,10	0,13	0,05	0,49	0,11	ZW/PM	Shedding	β-Glukan Organisation
	YLR194C	CAGL0A04081g	222	351	PP	0,42	0,07	0,20	0,06	0,38	0,10	ZW/PM	Shedding	Zellwand Strukturprotein
Adhäsion	Rbt1	CA2830	59	731	PP	0,43	0,13	0,07	0,03	0,50	0,10	ZW/PM		Adhäsion ähnlich zu Hwp1
			73	164	PP	0,50	0,10	0,07	0,03	0,42	0,08			
	Flo10	CAGL0E00275	398	40	PP	0,51	0,16	0,17	0,07	0,32	0,13	ZW/PM		Adhäsion vom Flocculin-Typ
		449	283	PP	0,60	0,10	0,10	0,02	0,30	0,12				



	Proteinae/ Sc-Homolog	ORF Nummer	Position	Score	Klassifikation	Protein		ProHormon		Kein Substrat		Kompartiment	Art	Beschreibung
						P	+/-	P	+/-	P	+/-			
A n- d- e r- e n- z- y- m- e	Atg15	YCR068W	139	42	PP	0,31	0,08	0,24	0,10	0,45	0,16	Golgi		Lipase
	Csh1	YBR161W	348	165	PP	0,46	0,19	0,06	0,02	0,48	0,18	Vakuole		Mannosylinositol-phosphorylceramide Synthase
	Pfa4	YOL003C	248	48	PP	0,31	0,07	0,23	0,13	0,46	0,19	Vakuole	C-terminal	Palmitoyltransferase
	Pho8	CAGL0H07359g	224	79	PP	0,45	0,08	0,19	0,08	0,36	0,12	Vakuole		Alkalische Phosphatase
	Plb1	CAGL0J11770g	59	64	PP	0,38	0,08	0,13	0,03	0,49	0,08	ZW/PM		Lysophospholipase
	Plb2	YMR006C	61	64	PP	0,40	0,12	0,11	0,01	0,48	0,13	ZW/PM		Lysophospholipase
	Ppn1	CAGL0K06237g	80	0	PP	0,41	0,16	0,13	0,03	0,46	0,17	Vakuole	Shedding	Endopolyphosphatase
	Rny1	CAGL0J06820g	104	502	PP	0,37	0,09	0,17	0,07	0,46	0,15	sekretorisch		Ribonuklease
	Vtc4	YJL012C	399	259	PP	0,40	0,09	0,13	0,03	0,46	0,12	Vakuole		Polyphosphate
	YMR317w	CAGL0E00187g	179	277	PP	0,55	0,08	0,10	0,01	0,35	0,08	Sekretorisch		Putative Sialidase
Ynd1	YER005W	265	47	PP	0,31	0,04	0,20	0,10	0,49	0,13	Golgi		Apyrase	
A n- d- e r- e n- z- y- m- e	Bud8	CAGL0B03091g	83	54	PP	0,46	0,16	0,06	0,03	0,47	0,13	Verschiedene		Zellknospung
	Lem3	CAGL0D02442g	48	149	PP	0,46	0,10	0,17	0,05	0,38	0,15	Plasmamembran		Translokation von Phospholipiden über die Plasmamembran
	Mrl1	YPR079W	44	482	PP	0,50	0,12	0,05	0,03	0,45	0,10	Vakuole	Loop	Putativer Mannose-6-phosphate Rezeptor
	Nce102	CAGL0L08448g	159	288	PP	0,40	0,05	0,10	0,01	0,50	0,06	Plasmamembran	Loop	Beteiligt an der nicht-klassischen Proteinsekretion
		YPR149W	160	208	PP	0,65	0,06	0,07	0,02	0,28	0,06			
		CAGL0M01386g	7	407	PP	0,53	0,17	0,12	0,02	0,35	0,18			
S- p- u- r- e- n- - e- l- e- m- e- n- t- A- q- u- i- s- i- t- i- o- n	YDR089w	CAGL0C03696g	304	212	PP	0,67	0,03	0,06	0,02	0,26	0,04	Plasmamembran		Nickel-Suszeptibilität
	Ctr2	CAGL0I02508g	115	65	PP	0,55	0,14	0,06	0,03	0,39	0,11	Vakuole	Loop	Kupfertransport-Protein
	YJL132W	YJL132W	307	114	PP	0,38	0,08	0,13	0,03	0,50	0,10	Membranfraktion		Induziert durch Zinkmangel
	Fre5	YOR384W	26	96	PP	0,29	0,09	0,23	0,10	0,49	0,07	Plasmamembran		Eisen-Reduktase

	Proteinae/ Sc-Homolog	ORF Nummer	Po siti on	S c o r e	K l a s s i f i k a t i o n	ProProt ein		ProHor mon		Kein Substra t		K o m p a r t i m e n t	A r t	B e s c h r e i b u n g
						P	+/-	P	+/-	P	+/-			
Pr o- te- i- e u n b e k a n n t e r F u n k t i- o n	CA0338	CA0338	134	444	PP	0,39	0,14	0,23	0,08	0,38	0,18	Sekretorisch		Unbekannte Funktion
	CA0339	CA0339	165	295	PP	0,52	0,19	0,11	0,02	0,37	0,19	Sekretorisch		Unbekannte Funktion
	CA0340	CA0340	146	772	PP	0,37	0,13	0,15	0,06	0,48	0,17	Sekretorisch		Unbekannte Funktion
	CA2994	CA2994	322	498	PP	0,51	0,12	0,14	0,03	0,35	0,14	Plasmamembran	Shedding	Unbekannte Funktion
	Pga18	CA2405	688	490	PP	0,34	0,09	0,17	0,05	0,50	0,11	ZW/PM	Shedding	Unbekannte Funktion
	Pga49	CA1360	470	48	PP	0,55	0,12	0,05	0,03	0,40	0,09	ZW/PM		Unbekannte Funktion
	Pin2	CAGL0B04477g	228	61	PP	0,45	0,20	0,11	0,03	0,44	0,19	Plasmamembran		Unbekannte Funktion
	Prm5	CA2394	20	369	PP	0,35	0,05	0,18	0,06	0,47	0,10	Plasmamembran		Unbekannte Funktion
	YBR255W	CA4624	17	468	PP	0,47	0,10	0,06	0,03	0,47	0,07	Microsomes		Unbekannte Funktion
	YFL042C	YFL042C	137	52	PP	0,48	0,19	0,06	0,03	0,46	0,17	Golgi		Unbekannte Funktion
	YFL042C	YFL042C	162	41	PP	0,57	0,13	0,05	0,03	0,38	0,12	Golgi		Unbekannte Funktion
	YLR057W	YLR057W	188	265	PP	0,51	0,22	0,08	0,06	0,40	0,16	Golgi		Unbekannte Funktion
	YLR072w	CAGL0L13134g	125	68	PH	0,26	0,04	0,26	0,14	0,48	0,14	Microsom		Unbekannte Funktion
	YMR244W	YMR244W	46	440	PP	0,60	0,12	0,13	0,02	0,28	0,10	ZW/PM	Shedding	Unbekannte Funktion
			328	593	PP	0,54	0,16	0,16	0,05	0,30	0,12			

## 5 Diskussion

Die Protease Kex2 ist eine im Trans-Golgi-Netzwerk von Pilzen lokalisierte Endoprotease, welche im Transit befindliche Proteine an spezifischen Schnittstellen schneidet und auf diese Weise aktiviert [75]. Trotz der biochemisch gut charakterisierten Aktivität von Kex2 aus *Saccharomyces cerevisiae* [24, 46, 48, 49] sind über die zwei bekannten Substrate Killertoxin [141] und  $\alpha$ -Kreuzungspheromon [27] hinaus nur wenige Proteine bekannt, für die eine Prozessierung durch Kex2 erfolgt. Trotzdem ist es in der Literatur üblich, Proteine mit einem in relativer Nähe zum N-Terminus gelegenen dibasischen „KR“-Motiv als putative Kex2-Substrate zu annotieren bspw. das Adhesin Hwp1 [142]. Es gibt jedoch Beispiele dafür, dass Kex2 auch C-terminale [143] oder endogene Prozessierungsstellen in Proteinen mit mehreren Transmembrandomänen [144] erkennt und schneidet, so dass generell davon ausgegangen werden muss, dass Kex2-Prozessierungen nicht nur auf N-terminale Bereiche beschränkt sind. Gleichzeitig finden sich viele verschiedene Proteine mit „KR“-Motiven, für die in der Literatur keine Prozessierung beschrieben wurde. Daraus folgt, dass die in der Literatur übliche Beschreibung der Schnittstelle als ein in Nähe des N-Terminus gelegenes KR-Motiv weder ausreichend noch spezifisch ist.

Die wenigen über das Kreuzungspheromon und Killertoxin hinaus bekannten Substrate von Kex2 sind Vertreter wichtiger Enzym- und Strukturproteinklassen [32, 145, 146]. Dieses spiegelt sich auch in den pleiotropen Phänotypen von Kex2-Deletionsmutanten wieder [32, 35], die eine allgemeine Schwächung der Zellwand- und Plasmamembranintegrität aufweisen. Diese Schwächung der Zellwand, und potentiell eine ausbleibende Aktivierung verschiedener, an Infektionsprozessen beteiligter Proteine, führt bei den beiden humanpathogenen Pilzen *C. albicans* und *C. glabrata* zu einer reduzierten Virulenz [32, 36].

In dieser Arbeit wurden daher Substratproteine von Kex2-Proteasen aus Pilzen im Allgemeinen, sowie in *C. albicans* und *C. glabrata* im Speziellen, *in silico* identifiziert und experimentell überprüft. Aufgrund der erweiterten experimentellen Basis wurde anschließend ein erweitertes Modell zur Vorhersage von Kex2-Substraten in Genomdaten entwickelt.

## 5.1 Entwicklung einer neuen Methode zur Vorhersage subzellulärer Proteinlokalisierung

Das Protein Kex2 ist in der Zelle hauptsächlich in einem späten Kompartiment des Trans-Golgi-Netzwerks lokalisiert [75]. Daher kommen als Substrate nur solche Proteine in Frage, die dieses Kompartiment selbst passieren. Zur Vorhersage von Kex2-Substratproteinen musste daher eine Methode („SLoPF-DB“ für „Datenbank Subzellulärer Lokalisation von Proteinen aus Fungi“) entwickelt werden, die in der Lage ist, solche Proteine aus dem Genom vorherzusagen, die im sekretorischen Weg zumindest über das ER hinaus transportiert werden.

Innerhalb der Zelle werden Proteine aufgrund verschiedener, in der Proteinsequenz kodierten Lokalisationssequenzen sortiert und über die unterschiedlichen Membranen der Kompartimente transloziert. Das Sortierschema, welchem SLoPF-DB folgt, ähnelt im Wesentlichen einem angenäherten Zwei-Schritt Sortierungsprozess, wie er auch in der Zelle funktioniert. Es wird zunächst eine Klassifizierung in „nicht-sekretorisch“/„sekretorischer Weg“/„mitochondrial“ vorgenommen („primäre Ebene“), wobei in die Klasse „sekretorischer Weg“ auch alle nicht-mitochondrialen Transmembranproteine und Proteine der Mikrokörperkompartimente fallen. Die Klassen „nicht-sekretorisch“ und „sekretorischer Weg“ werden dann weiter aufgeteilt auf „zytosolisch“/„nukleär“ bzw. „ER“/„Golgi“/„Vakuole“/„Peroxisom“/„Extrazellularraum“ („sekundäre Ebene“).

Die verschiedenen, in der Literatur beschriebenen Lokalisationssignale weisen in der Regel keine einfache Konsensussequenz auf. Somit ist *in silico* eine genaue Klassifizierung anhand dieser nicht immer möglich. Daher gibt es für die verschiedenen Lokalisationssignale auch verschiedenste Ansätze und Algorithmen, diese vorherzusagen [147]. An dieser Stelle wurde eine Kombination aus verschiedenen Methoden entwickelt, deren Zusammenspiel zu einer erhöhten Qualität der Aussage führt.

Es wurden zu Vorhersagen nicht nur die verschiedenen Lokalisationssignale, sondern auch die Aminosäurezusammensetzung [123-126] und die Ergebnisse einer Textanalyse von Proteinannotationen [129] herangezogen.

Die Vorhersage mittels Analyse der Aminosäurezusammensetzung basiert darauf, dass aufgrund der verschiedenen physikochemischen Bedingungen (pH-Wert, Ionenstärke) in den verschiedenen Kompartimenten der Zelle sich die Proteine, die dort lokalisiert sind, in der Aminosäurezusammensetzung der Oberfläche unterscheiden [123]. Der in der Literatur beschriebene Algorithmus (Supportvektormodell) [126] war allerdings nicht geeignet, da er nur Ja/Nein-Aussagen zuließ, hier jedoch später Wahrscheinlichkeitswerte benötigt wurden. Daher

wurde aus den Trainingsdaten [126] ein neues Supportvektormodell [90] generiert, welches dies ermöglichte.

Der dritte hier verwendete Ansatz, die Textanalyse, beruht auf der statistischen Analyse von Annotationen eines Proteins und/oder seiner Orthologen [129]. Es werden Wahrscheinlichkeiten für die Lokalisationskategorien errechnet, indem für in den Annotationen der Proteine gefundene, vorher festgelegte, Vokabeln Punkte vergeben werden. Problematisch bei dieser Methode ist, dass sie nur für Proteine eine Aussage treffen kann für die auch Annotationen vorliegen. Somit sind aus dem Validierungsset (siehe unten) nur 3001 Proteine (60,5%) mit dieser Methode klassifizierbar, dafür allerdings mit einem sehr hohen positiv prädiktiven Wert.

Ein Teil der Problematik bei statistischen Verfahren, wie sie hier angewandt werden, sind Ausreißersequenzen, welche leicht vom Idealmodell abweichen und auf diese Weise „gerade eben“ falsch klassifiziert werden. Fehler durch solche Ausreißer können normalerweise durch Mittelwertbildung vermindert werden, wie es beispielsweise bei physikalischen Messverfahren durch mehrere Messwiederholungen üblich ist. Da eine schlichte „Messwiederholung“ der Lokalisationsvorhersage hier wieder zu einem identischen Ergebnis führt, ist dies nicht praktikabel. Allerdings besteht die Möglichkeit, dass orthologe Sequenzen eines Ausreißers richtig klassifiziert werden und daher als „quasi Messwiederholung“ herangezogen werden könnten.

Daher wurden Ergebnisse dieser verschiedenen Methoden mit der Gruppierung von orthologen Sequenzen [130] der 17 zum Zeitpunkt dieser Arbeit vollständig sequenzierten und veröffentlichten Pilzgenome gekoppelt und innerhalb solcher Gruppierungen gemittelt.

Abschließend wurden die hier durchgeführten Vorhersagen von subzellulären Lokalisationen mittels 4940 Proteinen bekannter Lokalisation von *S. cerevisiae* validiert: bei Anwendung dieser Methode auf das Testset zeigte sich (Abbildung 15), dass die Lokalisationsvorhersage aufgrund von Lokalisationssequenzen eine sehr hohe Effizienz bezüglich der Erkennung von Proteinen des sekretorischen Weges aufweist. Für mitochondriale Proteine ist diese Methode weniger spezifisch, weil viele der Proteine aufgrund eines von TargetP nicht erkannten mitochondrialen Signalpeptids als zytoplasmatisch eingestuft werden (nicht gezeigt).

Die Performance der kombinierten Lokalisationsvorhersage aus Lokalisationssequenzen, Aminosäurezusammensetzung und Textanalysen wurde unter Verwendung der verschiedenen Genomgruppierungen gemessen. Die Messungen der Performance (Matthews-Korrelation) für die drei verschiedenen Lokalisationen „sekretorisch“/„nicht-sekretorisch“/„mitochondrial“ (Abbildung 18 B, C und D) zeigten, dass die Klassifikationsgenauigkeit proportional zur Anzahl der verwendeten Genome steigt. Der Effekt ist bei der Klassifikation in die Klassen

„sekretorisch“ und „nicht-sekretorisch“ am stärksten, während er für die Gruppe „mitochondrial“ nur moderat ansteigt.

Den größten Anteil an Fehlklassifikationen mittels Aminosäurezusammensetzung auf der primären Ebene stellt die Verwechslung von mitochondrialen und nukleären Proteinen dar, was auf die Redundanz des transkriptionellen Apparates zurückzuführen ist (nicht gezeigt). Dieses war im Rahmen dieser Arbeit aber nicht von Bedeutung, da keine der beiden Klassen als Kex2-Substrat in Frage kam.

Tatsächlich konnte beobachtet werden, dass bei Verwendung der Proteome Analyst Daten die Gesamtperformance signifikant ansteigt (Abbildung 19A), sich aber spezifisch die Vorhersage von sekretorischen Proteinen verschlechtert (Abbildung 19B). Dies war vor allem auf eine wesentlich höhere Falsch-Positiv-Rate durch peripher assoziierte Proteine zurückzuführen. Eine Analyse des Validierungssets ohne Verwendung des Proteome Analyst mit der Orthologengruppe 3 zeigte eine nur moderate Vorhersagequalität für die Unterkompartimente (Abbildung 20). Daher wurden die Vorhersagen des Proteome Analyst lediglich für die Vorhersage innerhalb des sekretorischen Subproteoms verwendet, nicht jedoch für den Eintritt in den sekretorischen Weg. Ein weiteres Beispiel für die Qualität der hier vorgestellten Verfahrensweise zeigt, dass alle sekretorischen Proteine, die in einer genetischen Studie in *C. albicans* gefunden wurden [148], unabhängig davon auch hier gefunden wurden.

## 5.2 Schnittstellen im Gesamt- und sekretorischen Proteom

Die Suche nach Kex2-Schnittstellenmotiven im so vorhergesagten Subproteom des sekretorischen Weges inklusive der entsprechenden Membranen zeigt, dass auf diese Weise 2/3 der Motive ausgeschlossen werden können (Abschnitte 4.1 und 4.3). Aufgrund der geringen Größe des sekretorischen Subproteoms lässt sich aus dieser Menge folgern, dass diese Motive in Proteinen des sekretorischen Weges mit etwa der doppelten Frequenz auftreten, wie im restlichen Proteom. Die Bewertung bekannter Schnittstellen (Abbildung 22, 1-4) ist im sekretorischen Subproteom relativ zur Gesamtzahl der gefundenen Proteine höher als im Gesamtproteom. Dies zeigt, dass die hier entwickelte PSSM im Zusammenspiel mit der Lokalisationsvorhersage erfolgreich Schnittstellen von Kex2 aus Genomdaten anreichern kann, die so gefilterten Daten aber manuell weiter analysiert werden müssen.

Nach manueller Überprüfung der oberen 10% der gefundenen Proteine mit Motiven, die Ähnlichkeit zu Kex2-Schnittstellen aufwiesen, wurden bei *C. albicans* 214 Motive aus 132 Proteinen (163/116 bei *C. glabrata* und 177/139 bei *S. cerevisiae*) weiterverwendet. Die

Gruppen sind in den Tabelle 21 bis Tabelle 35 aufgelistet, die restlichen Motive finden sich in Anhang 11.2.

### 5.3 Substraterkennung an der P1'-P4'-Region

In der Literatur sind biochemische Daten für die Substratspezifität von ScKex2 nur für den P4-P1-Bereich, nicht aber für den P1'-P4'-Bereich vorhanden. Die Schnittstelle in Substraten von Kex2 wird in der Literatur im Allgemeinen als [K/R]R bzw. XX[K/R]R angegeben. Allerdings gibt es statistische Daten zu Schnittstellen von Furinen in Proteinen verschiedener höherer Organismen [57] die besagen, dass hier häufig negativ geladene Aminosäuren auftreten. Dieses findet sich auch in den hier untersuchten Substraten von Kex2 bei Pilzen (Abschnitt 4.4.8): insbesondere in den P1' und P2'-Positionen findet sich eine Überrepräsentation negativ geladener oder kleiner, aliphatischer Aminosäuren. Positiv geladene Aminosäuren sind stark unterrepräsentiert, bzw. kommen an P1' gar nicht vor. Tatsächlich führt die Deletion von 18 Aminosäuren C-terminal der prozessierten Kex2-Schnittstelle, darunter auch der an P1' gelegenen, negativ geladenen Aspartatseitenkette in ScPir4 (KAKR|DVIS → KAKR|AQAT) zum Verlust der Prozessierung des Proteins [149].

Um zu untersuchen, ob sich in der Struktur der Kex2 und Furin-Proteine Bindungstaschen identifizieren lassen, die diese potentielle Substrataffinität erklären würden, wurde ein Alignment der entsprechenden 3D-Strukturen [28, 53] angefertigt. Zusätzlich wurde die Raumstruktur einer bakteriellen Subtilase, des Kumamolisin [139], in das Alignment aufgenommen. Diese Struktur enthält aufgrund einer Mutation im aktiven Zentrum noch das Propeptid. Das Propeptid wird intramolekular in analoger Weise zur intermolekularen Hydrolyse abgespalten [131, 150], und so lässt sich die Enzym-Propeptid-Bindung auf die intermolekulare Substraterkennung übertragen. Die Propeptid-Schnittstelle verläuft durch das aktive Zentrum, daher ist eine intermolekulare Hydrolyse nicht möglich (Abbildung 59).

Folgt man dem Verlauf der Prodomäne entlang der Substraterkennungsregionen S4-S1 von Kex2, Furin und Kumamolisin (Abbildung 60), so zeigt sich, dass diese Regionen räumlich ähnlich aufgebaut und die Substratspezifität durch, wenn auch unterschiedliche, Aminosäuren generiert werden, die an analoger Position sowohl in der Struktur als auch im Sequenzalignment zu finden sind (Abbildung 60 und [139]). Die Taschen S1' bis S3' finden sich bis auf kleinere Abweichungen dort, wo sie auch in der Literatur vorhergesagt wurden. Ein wesentlicher Unterschied zeigt sich bei der Bindung der P4' Position: hier nimmt die Linkerdomäne einen vollständig anderen Verlauf, als für den Inhibitor Eglin C vorhergesagt (lila eingefärbter Bereich in (Abbildung 60 und [58])). Die Aminosäure an P5' ist vom Protein weggedreht, so dass sie

keine Interaktionen ausbilden kann. Die Aminosäure an der P6'-Position ist zwar zum Protein hin ausgerichtet aber bereits nicht mehr in Reichweite.

Bei Betrachtung der putativen Interaktionspartner der einzelnen Aminosäuren in den Substratbindedomänen von Kex2 und Furin (Abbildung 61) lassen sich für die Bindungstaschen in Kex2 und Furin jeweils charakteristische Eigenschaften vorhersagen (Abschnitt 4.5.2.1). Von besonderer Bedeutung ist hier die auffällig starke positive Ladung der S1' und etwas weniger die der S2'-Tasche. Diese Ladung erklärt die Präferenz der Proteasen gegenüber negativ geladenen Aminosäuren insbesondere an P1', und ebenfalls die etwas abgeschwächtere an P2'. Die positive Ladung wird durch zwei Histidine (Positionen f und E) und bei Furin zusätzlich durch ein weiteres Arginin (Position e, Tyr in Kex2) vermittelt. Tatsächlich findet sich diese Ladungsverteilung auch in der Schnittstelle der Propeptide wieder: diese autokatalytisch prozessierte Schnittstelle trägt in den Furinen ein konserviertes Aspartat, bei Kex2 nicht.

Des Weiteren ist diese Tasche in beiden Proteinen direkt neben der S2-Tasche gelegen. Hier würde eine positiv geladene Seitenkette des Substrats diese Ladung noch weiter verstärken. Diese ist insbesondere für Kex2 von Bedeutung, da dieses Enzym im Gegensatz zu Furin an P2 stark auf positive Ladungen (hauptsächlich Lysin) selektiert.

Ein ähnlicher Unterschied ist auch in der putativen S4'-Tasche zu beobachten. Wieder wird der Eingang der Tasche von zwei polaren, ungeladenen Aminosäuren (Serin und Threonin bzw. Glutamin, Positionen D und F) gebildet, wobei sich terminal eine geladene Aminosäure befindet. Bei Kex2 ist dies ein positiv geladenes Histidin (Position G), bei Furin eine negativ geladene Glutaminsäure. Interessant ist auch, dass beide Aminosäuren aus unterschiedlichen Kontexten in der Primärstruktur des Proteins kommen, obwohl sie in der Struktur eine fast identische Konformation annehmen.

#### **5.4 Die biochemischen Daten von *S. cerevisiae* Kex2 sind nur bedingt auf die Kex2-Proteine anderer Pilze übertragbar**

Aus der Literatur ist bekannt, dass das CaKEX2-Gen die Deletion von ScKEX2 komplementieren kann [32] und ScKEX2 wiederum die Deletion von SpKRPI aus *S. pombe* [41]. Das spricht zunächst dafür, dass die entsprechenden Proteine eine identische Funktion und auch Substratspezifität besitzen.

Ein Vergleich der *in vitro* Aktivität der in dieser Arbeit hergestellten rekombinanten Proteasen aus *S. cerevisiae*, *C. albicans*, *C. glabrata* und *P. pastoris* zeigt, dass die Aktivität gegenüber dem Substrat KVKR-pNA bei allen vier Proteasen bei pH 7,3-7,5 (Abbildung 33A) am stärksten und damit ähnlich wie für ScKex2 bereits beschrieben [135] ist. Auch das Temperaturoptimum



war bei allen vier Protease vergleichbar. Allerdings lag es nicht, wie angenommen, in einem Bereich um 30°C-37°C, welches einer idealen Umgebungstemperatur unter Kulturbedingungen oder in der Infektionssituation entspräche, sondern bei unphysiologisch hohen Temperaturen um 50°C (Abbildung 33B).

Die Tatsache, dass die Enzyme bei solch hohen Temperaturen noch stabil sind, ist möglicherweise auf die strukturell stabilisierende Wirkung der P-Domäne zurückzuführen [54]. Eine biologische Signifikanz lässt sich diesem Phänomen wahrscheinlich nicht zuordnen, da die Organismen bei diesen Temperaturen nicht wachsen können. Das gilt insbesondere für *P. pastoris*. Dieser Pilz kann nur bis zu einer maximalen Temperatur von 30°C wachsen. Auch sind solche Bedingungen *in vivo* nicht zu finden, da beispielsweise für *C. albicans* und *C. glabrata* die Körpertemperatur des Wirtes, auch bei extremen Fieber, nicht über 44°C steigt.

Einige der hier untersuchten Substrate werden jedoch nicht von allen gleich effizient geschnitten. Bei der Prozessierung der  $\alpha$ -Pheromon Vorläuferproteine und dem Prohormon CaEce1 konnte zwar beobachtet werden, dass alle Proteine von allen Proteasen geschnitten werden und dass die der Vorhersage entsprechenden Zwischenprodukte in Erscheinung traten, dennoch war die jeweils aus dem entsprechenden Organismus stammende Protease gegenüber dem Substrat am aktivsten. Das lässt darauf schließen, dass die Kex2-Proteasen und die Schnittstellen der Pheromonvorläufer und andere Prohormone zueinander optimal angepasst sind.

Besonders auffällig war der Unterschied bei der Prozessierung des Prohormons CA0365: wiederum wurde das Polypeptid am effizientesten von der Protease des Herkunftsorganismus (hier CaKex2) prozessiert. Im Gegensatz zu den vorigen Proteinen wurde CA0365 jedoch gar nicht von ScKex2 prozessiert (Abbildung 39). Dies weist darauf hin, dass es im Einzelfall zwischen den Proteasen gravierende Unterschiede in der Substraterkennung geben muss.

Um zu untersuchen, ob dies mit den in der Literatur beschriebenen Substraterkennungsmechanismen zu erklären ist, wurde ein Vergleich der an der Substraterkennung beteiligten Bereiche zwischen den einzelnen Kex2-orthologen Proteinen angefertigt (Abbildung 62). Dieser zeigte zunächst einen hohen Grad an Konservierung zwischen den einzelnen Enzymen, insbesondere zwischen den evolutionär näher stehenden. Daher konnten hier keine Unterschiede aufgezeigt werden, die die oben beschriebenen Unterschiede in der Substraterkennung erklären würden.

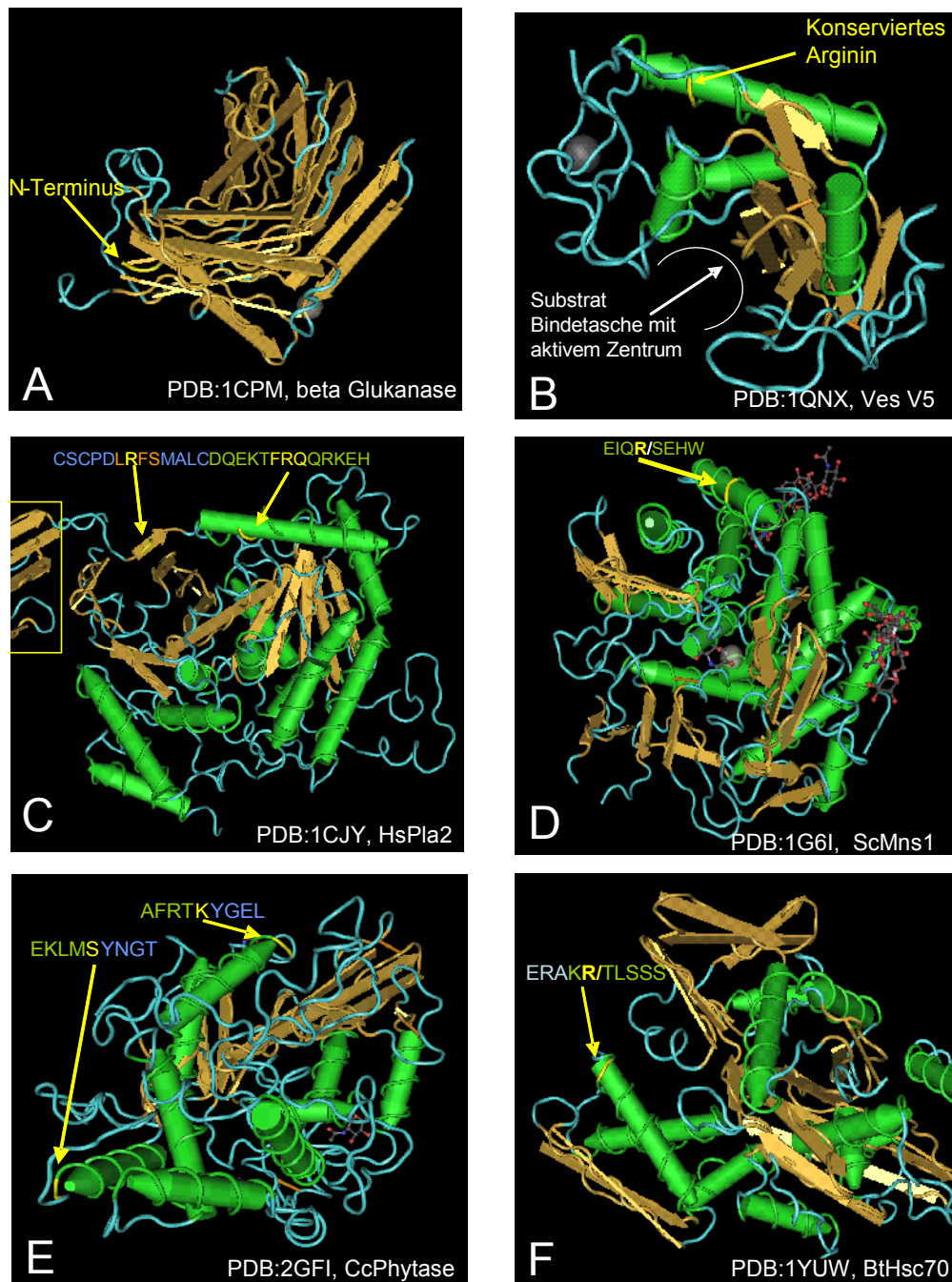
Allerdings lassen sich bestimmte Aminosäuren aufzeigen, die zwischen Hemiascomyceten und Ascomyceten unterschiedlich sind: die Kex2-orthologen Proteasen der Ascomyceten tragen überall dort geladene Aminosäuren, wo sie entweder in ScKex2 oder MmFurin vorkommen

(Abbildung 62). Das spricht für eine wesentlich stärkere elektrostatische Selektion bei diesen als bei jeweils ScKex2 oder MmFurin.

### **5.5 Die dreidimensionale Struktur des Substrats beeinflusst seine Prozessierbarkeit**

Die in den Experimenten beobachtete Unterteilung in prozessierte und nicht prozessierte Schnittstellen lässt sich mit der Ladungsverteilung um die Schnittstelle herum nur annähernd erklären. Einige Proteine besitzen Motive, die dem Modell sehr gut entsprechen, aber nicht geschnitten werden. Eine mögliche Erklärung für dieses Phänomen ist, dass die entsprechenden Aminosäuren nicht für die Protease zugänglich sind.

Ein Experiment mit denaturiertem CaEcel oder Ca1873-Protein zeigt, dass dieses weniger effizient geschnitten wird als natives Protein (Abbildung 55). Hierfür gibt es zwei mögliche Erklärungen: einerseits könnte die Ausbildung einer spezifischen Struktur für die Proteolyse nötig sein, andererseits könnten im denaturierten Protein die Schnittstellen größtenteils in unspezifischen Konglomeraten vor der Protease geschützt sein.



**Abbildung 65: Strukturen von Substratproteinen.** A: Räumliche Darstellung der bakteriellen Glukanase 1CPM. N-Terminus gelb unterlegt. B: Die Raumstruktur des zu den Pry-Proteinen homologen Proteins VesV5 (PDB:1QNX) zeigt, dass das konservierte Arginin nicht für die Protease zugänglich ist, da es zum Inneren des Proteins hin orientiert ist. C: Raumstruktur der Phospholipase A2 aus *Homo sapiens*. Die zu den Kex2-Schnittstellen-Motiven analogen Bereiche (gelb, Pfeile) liegen in unflexiblen Bereichen des Proteins am Übergang zu einer N-terminalen Domäne, welche in Proteinen aus Pilzen fehlt (gelber Kasten). D: Die Raumstruktur des Mns1-Proteins aus *S. cerevisiae* (PDB: 16GI) zeigt, dass sich die Schnittstelle an einer exponierten Stelle der Proteinoberfläche befindet und dass das P1-Arginin für die Protease erreichbar ist. E: Raumstruktur der entsprechenden Phosphatase aus *Debaryomyces castellanii*. Die zu den nicht geschnittenen Motiven analogen Bereiche (gelb) befinden sich an exponierten Stellen der Struktur. F: Raumstruktur des zu ScKar2 homologen Proteins Hsc70 aus *B. taurus*. Das Motiv mit Kex2-Schnittstellen geht aus einer exponierten Schleife an der Oberfläche des Proteins in eine  $\alpha$ -Helix über.

Die Betrachtung der dreidimensionalen Raumstruktur der Pry-homologen Proteine lässt hier eine differenzierte Interpretation der Ergebnisse zu: da CaRbt4 nicht geschnitten wurde und nach Proteolyse von CgPry1 nur die Produktbanden bei 17 und 18 kDa (Schnitte vor der konservierten Region), nicht aber die bei 15 kDa (Schnitt in der konservierten Region) nachweisbar waren (Abbildung 48), ist davon auszugehen, dass Pry-Proteine nicht an dem konservierten Arginin prozessiert werden, sondern nur in den vorderen Regionen. Dagegen spricht, dass CgPry2 auch prozessiert wurde, und zwar zwangsläufig an der einzigen Schnittstelle in Region 3. Allerdings fand die Prozessierung mit niedrigerer Effizienz statt. Beides könnte darauf hinweisen, dass die Renaturierung von CgPry2 nicht korrekt stattgefunden hat, und die Prozessierung durch Kex2 an dieser Stelle ein Artefakt ist. Diese Hypothese wird auch durch Strukturanalysen von homologen Proteinen aus anderen Organismen unterstützt: in dem Modell für das Protein Ves V5, das Hauptallergen aus Wespengift der Art *Vespula vulgaris* (PDB:1qnx, [151]), ist das konservierte Arginin innerhalb einer  $\alpha$ -Helix Struktur lokalisiert und die Seitenkette zum Inneren des Proteins hin gewandt (Abbildung 65B, gelb, mittig oben). Eine identische strukturelle Orientierung lässt sich auch in weiteren Proteinen beobachten (bspw. PDB:1cfe und 1smb, Daten nicht gezeigt).

### 5.5.1 Schnittstellen prozessierter Proteine sind exponiert

Die in dem Chaperon ScKar2 [152] und seinen orthologen Proteinen gefundenen Motive weisen von allen untersuchten Proteinen die höchsten mit der PSSM ermittelten Werte auf. Tatsächlich wird das rekombinante Protein ScKar2 auch von Kex2 geschnitten. Aus dem hohen Grad der Konservierung der Sequenz um die Schnittstelle zwischen den Orthologen kann man eigentlich annehmen, dass sich das Motiv in einem strukturell unzugänglichen Teil des Proteins befindet [153]. Die Raumstruktur des strukturell homologen Proteins Hsc70 aus *B. taurus* (Abbildung 65F) [154] zeigt allerdings, dass sich das entsprechende Motiv (ENKRAV**RR**LRITAC) exponiert in einer Schleife an der Oberfläche des Proteins befindet. Der hintere Teil des Motivs (ERAK**R**TLSS) geht dort in eine  $\alpha$ -Helix über, ist aber immer noch stark exponiert.

Ebenso ist die prozessierte Schnittstelle bei der  $\alpha$ -Mannosyltransferase Mns1 zwischen den Orthologen konserviert (Tabelle 38). Die Raumstruktur von ScMns1 (Abbildung 65D) zeigt, dass das Arginin an P1 exponiert an der Oberfläche des Proteins liegt. Obwohl es sich hier innerhalb einer Helixstruktur befindet, ist die Seitenkette vom Protein wegorientiert und somit für die Protease erreichbar. In den Spezies der *Saccharomyces* Gruppe befindet sich eine unkonservierte Insertion von ca. 9-12 Aminosäuren um die Schnittstelle herum, was nahe legt, dass dieser Abschnitt eine Schleife an der Oberfläche des Proteins bildet [153, 155] und damit

suszeptibel gegenüber Proteolyse ist. Dies bestätigt sich auch in der Raumstruktur von ScMns1 [156].

Die Exoglukanase Exg1 aus *C. albicans* ist eines der wenigen bekannten Kex2-Substrate [145]. Es wird in der Nähe des N-Terminus (\*VISNPFKPNGNLKFKR|GGGHNVVA...) prozessiert. In der Raumstruktur des reifen Proteins befindet sich dieser N-Terminus an der Oberfläche des Proteins [99]. Ob diese Prozessierung für die Funktion der Glukanase eine Rolle spielt, ist nicht klar.

### 5.5.2 Unprozessierte Motive sind für die Proteasen nicht erreichbar

Es wurde bei sieben weiteren erfolgreich aufgereinigten Proteinen trotz potentieller Kex2 Schnittstellen keine Prozessierung beobachtet: die drei Proteine unbekannter Funktion und ohne Homologien zu anderen Proteinen CAGL0A02277, CAGL0H08910, CA1394, die vier Proteine CA2942 (Zellwand-Strukturprotein CCW14), CA0375 (putative Transglukosidase Crh11), CAGL0J11770 (PhospholipaseB1) und CA5147 (saure Phosphatase Pho114). Zu den drei Proteinen unbekannter Funktion CAGL0A02277, CAGL0H08910, CA1394 und zu CaCcw14 oder entsprechenden sequenzhomologen Proteinen gibt es in der Literatur keine experimentellen Daten bezüglich ihrer Raumstruktur.

Die Familie der Crh1-homologen Proteine stellt eine Gruppe von zellwandlokalisierten Transglykosidasen bei Pilzen dar, welche für die Vernetzung des Chitin- und Glukangerüsts der Zellwand benötigt werden, und diese enzymatische Reaktion möglicherweise selbst katalysieren [138]. Ein zu den Crh1-homologen Proteine strukturell homologes Protein ist die  $\beta$ -D-Glukan 4-Glukanohydrolase 1CPM aus *Bacillus spec.*[157]. Der zur potentiellen Kex2-Schnittstelle der Crh1-Homologen analoge Bereich ist bei dem für die Kristallisation verwendeten Konstrukt nicht mit verwendet worden, tatsächlich fängt das reife Protein gerade hinter diesen Aminosäuren an (Abbildung 65A). Es handelt sich hierbei allerdings nicht um eine proteolytische Prozessierung des Proteins, sondern lediglich um einen „Artefakt“ der Klonierungsstrategie. Man kann aus der Raumstruktur erkennen, dass der zur Schnittstelle analoge Bereich gerade im Proteininneren verborgen liegt (Abbildung 65A), und somit im reifen Protein nicht von der Protease erreicht werden kann. Diese Daten werden indirekt durch unveröffentlichte Experimente aus *S. cerevisiae* bestätigt (A. Engel, persönliche Mitteilung) [158]. In dieser Arbeit wurde gezeigt, dass ScCrh1, ebenso wie ScTos1 (siehe auch Abschnitt 4.4.6.7), *in vivo* nicht von Kex2 prozessiert wird.

Bei dieser Phospholipase B aus *C. glabrata* handelt es sich ebenfalls um ein Enzym der Zellwand, dessen biologische Funktion allerdings unklar bleibt. In den gezeigten Plb1-homologen Proteinen befinden sich zwei konservierte potentielle Schnittstellen (Abbildung 52), die in CAGL0J11770g jedoch unprozessiert bleiben.

Die zu diesen Proteinen am nächsten gelegene Raumstruktur stammt aus einer Phospholipase A2 aus *Homo sapiens*. In dieser Struktur befinden sich die Motive zwar an oberflächlich gelegenen, aber rigide verankerten Regionen der Struktur (Abbildung 65C). Die erste in einer  $\beta$ -Faltblatt-Struktur an der Grenze zu einer N-terminalen Domäne, welche in Pilzen nicht vorkommt. Die zweite innerhalb einer  $\alpha$ -helikalen Struktur, bei der die Seitenkette zum Inneren des Proteins gewandt ist. Beides legt auch hier strukturelle Gründe für das Ausbleiben der Prozessierung nahe.

Aus der Familie PHO der sauren Phosphatasen trägt lediglich das Protein aus *C. albicans* (Pho114/CA5147) Motive, die Kex2-Schnittstellen ähneln (Abbildung 53A und B). Das Protein ist bereits vor der Zugabe von Kex2 aktiv (siehe Abschnitt 4.4.5), und wird durch Zugabe von Kex2-Enzymen nicht weiter prozessiert, was zeigt, dass es sich hier nicht um ein Substrat von Kex2 handelt. In der Raumstruktur des strukturell homologen Proteins Phytase aus dem nahverwandten Pilz *Debaryomyces castellanii* befinden sich die entsprechenden Sequenzmotive allerdings an exponierten Stellen an der Oberfläche des Proteins (Abbildung 65E). Es muss gefolgert werden, dass sie entweder trotz ihrer Lage strukturell fixiert sind oder dass die Primärsequenz bei CA5147 kein Substrat für Kex2-Enzyme darstellt.

### 5.5.3 Computermodell zur verbesserten Vorhersage

Wie oben gezeigt, unterscheiden sich die hier prozessierte Motive von unprozessierten vornehmlich durch ihre Lage in der Struktur des Substratproteins. Um zu einer verbesserten Vorhersage von Proteasesubstraten zu kommen wurde daher untersucht, ob sich mit konventionellen Strukturvorhersage-Algorithmen solche suszeptiblen Regionen finden lassen. Die hier verwendeten Algorithmen sagen unterschiedliche Sekundärmerkmale vorher. Der rvp-net-Algorithmus berechnet den Grad der Exposition von Seitenketten und damit die Wahrscheinlichkeit der Oberflächenlokalisierung für einen Abschnitt eines Proteins. Der Prof/PhD-Algorithmus berechnet die Wahrscheinlichkeiten für die Sekundärstrukturelemente  $\alpha$ -Helix/ $\beta$ -Faltblatt/Coil und der di-sopred2-Algorithmus berechnet die Flexibilität eines Proteinsabschnitts und damit ein Maß für die Wahrscheinlichkeit ob eine Aminosäure in einer exponierten Schleife liegt.

Tatsächlich konnte gezeigt werden, dass bestimmte Strukturmerkmale in prozessierbaren Regionen gehäuft auftreten: allgemein sind die Aminosäuren in Substraten stärker exponiert (Abbildung 63A, orange), stärker ungeordnet (Abbildung 63B, lila) und prägen weniger  $\beta$ -Faltblatt und dafür mehr Coil-Strukturen aus (Abbildung 63C grau und orange). Besonders auffällig ist, dass in Substraten das Arginin an P1 immer exponiert ist. Dies gilt verstärkt auch für die benachbarte P1' Position.

Ein Beispiel, dass dieser Ansatz erfolgreich ist, um Prozessierungen zu verstehen, die mit Sequenzmotiven allein nicht erklärbar sind, liefert die Betrachtung des Proteinvorläufers des K28 Killertoxins aus *S. cerevisiae*: Dieses Proprotein wird von Kex2 an drei Stellen prozessiert, ER<sub>49</sub>, SR<sub>149</sub> und KR<sub>245</sub>. Gleichzeitig wird die potentiell bevorzugte Sequenz KR<sub>192</sub> nicht geschnitten. Vier weitere Arginine bleiben unprozessiert: VR<sub>118</sub>, QR<sub>316</sub> YR<sub>329</sub> und AR<sub>340</sub> [107]. Betrachtet man nun die Vorhersage des Diospred2-Algorithmus für potentiell ungeordnete und flexible Regionen im Protein (Abbildung 66), so wird ersichtlich, dass sich die prozessierten Arginine jeweils nach kurzen flexiblen Regionen befinden, während die unprozessierten Bereiche als geordnet vorhergesagt werden.





## 5.6 Korrelation von Phänotypen mit Substraten

Deletionsmutanten von *ScKEX2* und *ScKEX2*-ähnlichen Genen bei verschiedenen Pilzen zeigen ein breites Spektrum an phänotypischen Veränderungen gegenüber Wildtypstämmen. Eine direkte Folge des Kex2-Verlustes ist die Sekretion und Anreicherung unprozessierter Proteinvorläufer im Kulturmedium [27, 32, 40, 42]. Die Folgen einer ausbleibenden Prozessierung lassen sich in zwei Fällen direkt beobachten: die fehlende Prozessierung des  $\alpha$ -Pheromons führt zum Ausbleiben der sexuellen Konjugation der Paarungstypen [27]. Am sexuellen Reproduktionszyklus bei *S. cerevisiae* müssen aber noch mehr Kex2-Substrate als die bisher Bekannten beteiligt sein [158].

Andere Phänotypen sind weit pleiotroperer Natur (Tabelle 1): alle bisher hergestellten *KEX2*-Mutanten zeigen morphologische Defekte der Zellwand [35, 37, 38], insbesondere bei der Ausbildung von Myzelstrukturen [33, 37-39] und bei der Separation von Tochter- und Mutterzelle nach der Zellteilung. Die Summe dieser Effekte und die potentiell fehlende Prozessierung von Virulenzproteinen, wie bspw. den sekretorischen Aspartatproteasen bei *C. albicans*, führen zu einer Reduktion der Virulenz [33].

Allerdings müssen einzelne ausbleibenden Prozessierungen nicht zwangsläufig zu einem beobachtbaren Phänotyp führen. Ein Beispiel für eine Kex2-Prozessierung, welche offenbar ohne Konsequenz ist, findet sich bei dem Transmembranprotein ScGef1. Dieses Kanalprotein, welches in den Eisenstoffwechsel involviert ist, besitzt eine Prozessierungsstelle in der ersten luminalen Schleife [159]. Dabei ist für die Funktion des Kanals egal, ob das Protein prozessiert wird oder nicht, es ist auch möglich die beiden Untereinheiten getrennt zu exprimieren um funktionelles Protein zu erhalten.

Bei den Proteinen, für die in dieser Arbeit gezeigt wurde, dass sie von Kex2 geschnitten werden können, handelt es sich um verschiedenartige sekretorische Proteine, bei denen ein Ausbleiben der Prozessierung jeweils Teile dieser phänotypischen Erscheinungen in *kex2*-Deletionsmutanten erklären könnten, da die phänotypischen Erscheinungen der entsprechenden Deletionsmutanten denen von *kex2*-Deletionmutanten ähneln.

### 5.6.1 Zellwand assoziierte Phänotypen der *kex2*-Deletionsmutanten

Zu den offensichtlichsten Phänotypen der Kex2-Deletionen in *Candida* gehören morphologische Defekte bei der Ausbildung von Myzel [32] und die Hypersensitivität gegenüber Verbindungen und Wirkstoffen, die die Zellwandintegrität herabsetzen [35], bspw. Calcofluor Weiß.

Verschiedene Kex2-Substrate interagieren direkt mit der Zellwand oder sind ein Teil davon: die strukturellen Pir-Proteine, Glukanasen wie CaExg1 oder die Proteine der Sun/Scw Familie.

Von Proteinen der Pir-Familie (Proteins with internal repeats, Abbildung 45) ist bekannt, dass sie in *kex2*-Deletionsmutanten anomal prozessiert werden [102, 103]. Zusätzlich zu den Pir-Proteinen gibt es bei *C. albicans* eine Gruppe zum C-Terminus der Proteine der Pir-Familie ähnlicher Proteine (Abbildung 44). Diese verfügen zwar nicht über die charakteristisch kurzen internen Wiederholungen, aber über eine im Vergleich zu den Pir-Proteinen weiter C-terminal gelegene potentielle Kex2-Schnittstelle. Da bisher nicht gezeigt wurde, ob Pir-Proteine von Kex2 oder einer anderen, aber selbst Kex2-abhängigen, Protease geschnitten werden, wurden zwei Pir-homologe Proteine in das Spektrum der zu untersuchenden Substrate aufgenommen (CA0811, Substrat #22, und CAGL0M08492, Substrat #38). Eines davon (CAGL0M08492) konnte teilweise aufgereinigt und für Versuche verwendet werden (Abbildung 46). Während sich das von CA0811 kodierte Protein nicht in löslicher Form herstellen ließ, konnte CgPir1 in ausreichend sauberer Form aufgereinigt werden. Der Verdau mit den verschiedenen Kex2-Enzymen zeigte, dass es von allen geschnitten wurde (Abbildung 46). ScKex2 zeigte dabei eine wesentlich geringere Effizienz als die anderen Enzyme. Es kann also angenommen werden, dass die Pir-Proteine tatsächlich von Kex2 geschnitten werden.

Eine besonders auffällige Struktur mit potentiellen Kex2-Schnittstellen ist bei Proteinen der Sun/Scw-Familie und bei ScTos1 und seinen orthologen Proteinen vorhanden. Sie besteht aus einem KR-Motiv, welches N-terminal von einer Reihe positiv geladener Aminosäuren, meist Histidin, flankiert wird. Bei Tos1 befindet sich das Motiv, nicht wie bei Sun/Scw, direkt am N-Terminus, sondern ca. 120 Aminosäuren in das Protein hinein. Allerdings befindet sich direkt vor dem Motiv eine weitere potentielle Schnittstelle, so dass bei einer Prozessierung an dieser vorderen Schnittstelle die histidinreiche Struktur wieder am N-Terminus des Proteins gelegen wäre (Abbildung 47A). In einer Westernblot Analyse konnten für CaSun41 die N-terminal abgespaltenen Produkte von 5 und 7 kDa nachgewiesen werden. Ebenso konnte in einer SDS-PAGE eine Prozessierung bei CgScw4 gezeigt werden.

Proteine der Scw4-Familie sind potentielle Glukanasen, deren Ausschalten zu einer Verringerung der Kreuzungseffizienz führt [160] (siehe unten). Ebenso wird auch von den Sun-Proteinen vermutet, dass es sich um Glukanasen handelt. Bei massenspektrometrischen Analysen von isolierten Zellwandproteinen wurden die N-terminalen, histidinreichen Regionen nicht gefunden (Yin, persönliche Mitteilung). Da für ScSun4 gezeigt wurde, dass es sich außer in der DTT-extrahierbaren Zellwandfraktion auch in Mitochondrien befindet [161], besteht die

Möglichkeit, dass *in vivo* eine Prozessierung hier eine Funktion bei der Lokalisation des Proteins übernimmt.

Die Konsequenz einer ausbleibenden Prozessierung durch Kex2 für diese Proteine *in vivo* ist nicht bekannt. Dennoch ähneln die Phänotypen der entsprechenden Deletionsmutanten denen der *kex2*-Deletionsmutante: Mutanten der zellwandlokalisierten Glukanase ScExg1 [162], der sekretorischen Glukanase CaBgl2 [163] oder Stämme, in denen Mitglieder der Pir- [164] oder Sun-Familien [165, 166] ausgeschaltet wurden, zeigen ähnlich erhöhte Suszeptibilität gegenüber zellwandperturbierenden Wirkstoffen wie die *kex2*-Mutanten von *C. glabrata* oder *C. albicans* [33, 35]. Hier ist es interessant, dass nicht alle Glukanasen eine solche Schnittstelle besitzen. Möglicherweise bedeutet dies, dass entweder die Prozessierung ohne Konsequenz für die enzymatische Funktion des Proteins ist oder es sich um zwei verschiedenen Gruppen von Glukanasen handelt.

Zusätzlich resultiert die Deletion von Pir-Proteinen in der Ausbildung von Zellaggregaten [164], ein Phänotyp der auch in *S. cerevisiae* *SUN4* und *C. albicans* *SUN41* Deletionen beobachtet wird [165-167]. Eine *S. cerevisiae* *scw4/scw10* Doppelmutante [165] und ein *C. albicans* *sun41* Stamm zeigten stark vergrößerte Zellen [166], ein Phänotyp, der sich auch in der *C. glabrata* *kex2* Mutante (Daten nicht gezeigt) beobachten lässt. Des Weiteren zeigen Calcofluor Weiß gefärbte Zellen ein verändertes Färbemuster [32]. Diese Veränderungen in der Chitinverteilung der Zellwand ähneln denen die auf eine veränderte Prozessierung der Septen in *C. albicans* *sun41* Stämmen [165] zurückzuführen ist. Den Schnittstelle in Sun4- und Scw10-ähnlichen Proteinen geht ein N-terminaler Bereich von positiv geladenen Aminosäuren, hauptsächlich Histidinen (Abbildung 47), voraus. Dieses Motiv, welches hier „His-Box“ benannt wurde, findet sich auch bei Tos1-homologen Proteine, allerdings liegt es hier nicht am N-Terminus sondern weiter in der Mitte des Proteins und wird in Richtung N-Terminus von einem weiteren KR-Motiv flankiert. Würde dieses Motiv als erstes gespalten, so befände sich das His-Box-Motiv wieder am (jetzt neuen) N-Terminus des Proteins. Es kann spekuliert werden, dass wenn dieses Motiv in die Zellwandverankerung dieser Proteine involviert wäre, eine Prozessierung hier zur Sekretion der Proteine führe. Solche duale Lokalisierungen wurden bei *C. albicans* *Sun41* und *Tos1* [166] beobachtet.

## 5.6.2 Manche Proteine der Pry-Proteinfamilie besitzen glykosylierte potentielle Propeptide

Die Familie der Pry-Proteine (Familie der PR-1-homologen Proteine) besteht aus 2-5 Mitgliedern pro untersuchter Spezies. Sie kommt in allen hier untersuchten Spezies vor. Einige weisen N-terminal vor der konservierten Domäne potentielle Kex2-Schnittstellen auf (Tabelle 37). Die Funktion von Pry-homologen Proteinen bei Pilzen ist nicht bekannt. Es gibt aber strukturell homologe Proteine anderer Organismen mit bekannter Funktion, beispielsweise das in Wespengift vorkommende Protein V5/Tpx1 (siehe potentielle Substratbindungstasche in Abbildung 48B), für das gezeigt worden ist, dass es sich um eine Protease handelt. Tpx1 schneidet, wie auch Kex2, hinter Arginin.

PR1-homologe Proteine zeichnen sich durch einen hoch konservierten Bereich aus, der homolog zu der aktiven Domäne von Tpx1 ist und entweder N- oder C-terminal von einer S/T-reichen Region flankiert wird. Bei wenigen findet sich auch ein GPI-Verankerungsmotiv. Da es sich auch bei den Pry-Proteinen um Proteasen handeln kann, lag die Vermutung nahe, dass sie über abspaltbare Propeptide verfügen. Tatsächlich gibt es am Anfang der konservierten Domäne ein konserviertes Arginin, welches in den Pry-Proteinen der Pilze einer Kex2-Schnittstelle ähnelt (Abbildung 48A, Region 3). Andere Proteine der Familie tragen potentielle Schnittstellen am N-Terminus und CgPry1 kann kurz vor der konservierten Domäne prozessiert werden. Dies würde zum Verlust von ~60 Aminosäuren N-terminal zur konservierten Domäne führen, welche möglicherweise für eine katalytische Funktion des Proteins nicht benötigt werden, da sie bei anderen Proteinen der Familie vollständig fehlen (z. B. ScPry3).

Aus dieser Familie wurden in dieser Arbeit die drei Proteine CaRbt4, CgPry1 und CgPry3 erfolgreich exprimiert und untersucht. CaRbt4 wurde dabei nicht von Kex2-Enzymen geschnitten (nicht gezeigt). Bei CgPry1 wurden zwei Fragmente von jeweils ~17 und ~18 kDa beobachtet. Bei CgPry2 wurde nur eine Produktbande von ca. ~13-15 kDa beobachtet, welche sich jedoch nicht eindeutig einem spezifischen Produkt zuordnen ließ. Die Fehlende Prozessierung von CaRbt4 lässt den Schluss zu, dass es sich bei dem konservierten Motiv in Region 3 nicht um eine Kex2-Schnittstelle handelt. Es ist daher zu vermuten, dass die Prozessierung bei CgPry2 lediglich ein Artefakt aufgrund schlechter Renaturierung ist und die Prozessierung bei CgPry1 an Region 1 erfolgt (siehe Abschnitt 5.5).

Eine Analyse der in der Literatur verfügbaren Expressionsanalysen zeigt, dass Pry-Proteine möglicherweise in verschiedene Prozesse involviert sein können: *ScPRY1* wird sowohl in den Rim101- und MVB-abhängigen pH-Regulationskaskaden aktiviert [168] als auch über das Aft1-abhängige Eisenmangel-Regulon [169]. Alle drei *PRY*-Gene aus *S. cerevisiae* wurden in

Experimenten zur Isooktan-Resistenz differenziell reguliert [170] und bei *ScPRY2* konnte eine Induktion unter Glukose-Mangel Bedingungen beobachtet werden [171]. Für die orthologen Gene aus *C. albicans* und *C. glabrata* lagen bisher keine publizierten Expressionsdaten vor.

### 5.6.3 CaEce1 und CA0365 sind Polypeptidvorläufer mit Ähnlichkeit zu UmRep1

Das bei der Induktion von Hyphenwachstum bei *C. albicans* am stärksten hoch regulierte Gen ist *CaECE1*. Es wurde bereits 1993 beschrieben und war das erste mit der Ura-Blaster-Methode deletierte Gen [136]. Bei der Betrachtung der von *CaECE1* kodierten Proteinsequenz fällt auf, dass sie acht degenerierte Peptidwiederholungen enthält [136]. *CaEce1* besitzt keine Sequenzhomologien zu anderen Proteinen außer zu einem orthologen Protein in *C. dubliniensis* (Abbildung 37). Andere *Candida* Spezies besitzen kein solches Ortholog (Daten nicht gezeigt). Das zusätzliche KR-Motiv an Position 253 bei *CdEce1* zeigt, dass möglicherweise auch das an der entsprechenden Position in *CaEce1* vorhandene ER-Motiv eine Kex2-Schnittstelle darstellt. Ein Zeitreihenexperiment zeigte, dass das Protein bevorzugt an den Enden prozessiert wurde, da anfänglich genau diese Intermediate entstanden, nicht aber solche, die auf eine Prozessierung in der Mitte schließen ließen (Abbildung 38B). Die Zuordnung der N-terminalen Fragmente konnte im Westernblot bestätigt werden (Abbildung 38C). Trotz der starken Expression konnte für die *Caece1*-Mutante jedoch kein Phänotyp beobachtet werden [136].

Auch das von ORF CA0365 kodierte Protein zeigt, wie *CaEce1*, die Struktur eines Prohormonvorläufers. Vier konservierte Wiederholungen werden von potentiellen Kex2-Schnittstellen flankiert, die ersten drei Wiederholungen beinhalten dabei eine weitere Schnittstelle. Die vollständige Prozessierung würde so zu drei kurzen (~1,5 kDa) und vier längeren (~3 kDa) Peptiden führen (Abbildung 39A). Der Verdau mit den vier Kex2 Proteasen führte tatsächlich, ohne nennenswerte Bildung von Intermediaten, zu peptidgroßen Abbauprodukten (Abbildung 39C).

Wie *CaEce1* besitzt auch CA0365 keine Sequenzhomologien zu anderen Proteinen außer zu genau einem Ortholog bei *C. dubliniensis*. Dieses kodiert allerdings für insgesamt neun Wiederholungen, wovon fünf denen von CA0365 entsprechen. Die anderen vier sind leicht unterschiedlich, aber in sich wieder konserviert (Abbildung 39B).

CA0365, *CaEce1* und die beiden entsprechenden orthologen Proteine aus *C. dubliniensis* weisen keine längeren Sequenzhomologien zu anderen bekannten Proteinen auf. Das einzige bekannte Protein mit einem von Kex2 prozessierten ähnlichen Aufbau ist das Hydrophobin *UmRep1* aus *Ustilago maydis*. *UmRep1* kodiert für zehn konservierte Wiederholungen (Abbildung 40) und

eine längere C-terminale Domäne [112]. Es handelt sich hierbei um sekretorische Peptide, die in Lufthyphen in der Zellwand verankert werden und der Oberfläche hydrophobe Eigenschaften vermitteln. Eine weitere Möglichkeit wäre, dass es sich um Signalmoleküle ähnlich den Kreuzungspheromonen handelt. Hydropathieplots zeigen, dass die einzelnen Untereinheiten von CA0365 und CaEce1 mehr Ähnlichkeiten zu Rep1 als zu denen aus Pheromonvorläuferproteinen zeigen (Abbildung 41).

#### 5.6.4 Matingassoziierte Phänotypen der *kex2*-Deletionsmutanten

Das *KEX2*-Gen wurde ursprünglich durch seine Beteiligung an der sexuellen Reproduktion („Mating“) und in der Prozessierung des Killertoxins isoliert [27]. In der *S. cerevisiae* *kex2*-Deletionsmutante wird das  $\alpha$ -Pheromon Vorläuferprotein nicht mehr zu den funktionellen Peptiden hydrolysiert und somit findet keine Konjugation der Kreuzungspartner mehr statt. Es muss allerdings über das Pheromon hinaus weitere Kex2-Substrate geben, die an der Zellfusion während des Matings beteiligt sind [158]: in jener Arbeit konnte gezeigt werden, dass sowohl eine *prm1/kex2* als auch eine *prm1/kex1*-Mutante keine Zellfusionen mehr durchführen konnte. Um das entsprechende Kex2-Substrat zu finden, welches diese Disfunktion induziert, wurden 11 Deletionen von bekannten Kex2-Substraten im *S. cerevisiae* *prm1*-Hintergrund erzeugt und auf die Fähigkeit der Zellfusion hin getestet. Darunter waren auch verschiedene Glukanasen und Pir-Proteine. Das Ausschalten von Glukanasen der Scw4-Familie führt zu einer Verringerung der Matingeffizienz [160]. Trotzdem konnte weder hiermit noch mit keinem der weiteren getesteten Kex2-Substrate aus *S. cerevisiae*, der Zellfusionsphänotyp erklärt werden, so dass die Autoren folgern, dass es weitere, unbekannte an diesem Prozess beteiligten Kex2-Substratproteine geben muss.

Für die in der Vergangenheit als „asexuell“ bezeichnete Hefe *C. albicans* wurde mittlerweile *in vitro* ein Kreuzungsprozess gezeigt [8, 9]. Auch hier ist die Kex2 bei der Prozessierung des  $\alpha$ -Kreuzungspheromons aktiv und essentiell für das Mating [172]. Bei *C. glabrata* ist bisher noch keine sexuelle Reproduktion gezeigt worden, aber ein  $\alpha$ -Pheromonvorläuferprotein findet sich auch hier im Genom [19]. Ein weiterer Hinweis für ähnliche Prozesse in der *C. glabrata* *kex2*-Deletionsmutante wie bei *S. cerevisiae*, sind teilweise ähnliche Abberationen in der Struktur der Plasmamembran [35].

Eine Gruppe von Proteinen, die potentielle Schnittstellen trägt und aufgrund ihrer Transkriptionsprofile (s. u.) möglicherweise in die Kreuzungsprozesse involviert ist, besteht aus CaOps4, den zwei zu Ops4 homologen Proteinen CA1873 und CA2974, und dem GPI-

verankerten Protein Pga17. Diese Proteine weisen N-terminal eine Gruppe aus zwei oder drei potentiellen Schnittstellen und eine weitere Schnittstelle in der Mitte auf. Bei Pga17 ist die gesamte Anordnung zum C-Terminus hin verschoben.

Für die Proteine CA1873, CA2974 und Pga17 konnte in dieser Arbeit eine Prozessierung an diesen Schnittstellen gezeigt werden. Ops4 konnte in *E. coli* nicht exprimiert werden. Aufgrund der hohen Sequenzhomologien lässt sich aber vermuten, dass auch Ops4 analog zu den anderen Proteinen geschnitten wird. Bei Pga17 konnte lediglich die Prozessierung in der Gruppe der ersten drei Schnittstellen festgestellt werden. Da aber keine weiteren größeren Produkte beobachtet werden konnten, ist davon auszugehen, dass Pga17, wie die anderen Proteine der Gruppe, vollständig prozessiert wird.

Ein Vergleich der Proteinsequenzen zeigt nicht nur eine ähnliche Anordnung der Kex2-Schnittstellen, sondern auch weitere starke Sequenzähnlichkeiten. Der zwischen beiden Gruppen von Kex2-Schnittstellen gelegene mittlere Teil der Proteine trägt potentielle N-Glykosylierungsstellen, während der C-terminale Teil mögliche O-Glykosylierungsstellen aufweist. Pga17 besitzt ein C-terminales GPI-Verankerungssignal [120]. Bei den drei anderen Proteinen ist der C-Terminus auffällig positiv geladen und endet in einem konservierten Tyrosin. Dieses Motiv könnte ein bisher unbekanntes subzelluläres Lokalisationssignal darstellen.

Die biologische Funktion ist für keines dieser Proteine bekannt. Aus verschiedenen Transkriptionsprofilen aus der Literatur lässt sich aber eine Beteiligung an Kreuzungs- und „switching“-Prozessen ableiten: *OPS4* wird spezifisch in „opaque“-Zellen transkribiert und nach Kontakt der Zelle mit  $\alpha$ -Pheromon in opaque MTLa/a Zellen herunterreguliert [173]. Ebenso werden auch *CA1873* und *PGA17* nach Kontakt mit  $\alpha$ -Pheromon reprimiert [174]. Zusätzlich wird *PGA17* durch Kontakt mit Makrophagen verstärkt exprimiert [175]. *CA2479* wurde bisher nicht mit diesen Prozessen in Verbindung gebracht, jedoch wurde die Expression von *CA2974* in Gegenwart von Fluconazol um den Faktor 9,4 reprimiert [176].

### 5.6.5 Auch Proteine aus dem ER tragen prozessierbare Schnittstellen

Das Protein Kex2 ist in einem späten Kompartiment des Golgi lokalisiert, und wird auch dort erst aktiviert. Trotzdem konnte in Zusammenarbeit mit H. Mora, Universität Aberdeen, UK, [137] gezeigt werden, dass auch ER-lokalisierte Proteine durch Kex2 *in vivo* prozessiert werden können. In diesem Fall handelte es sich um das Membranprotein CaMns1, eine  $\alpha$ -1,2-Mannosidase welche an der Glykosylierung von Proteinen im ER beteiligt ist. Mittels Kombination von Proteaseinhibitoren wurde zunächst eine Prozessierung durch eine

Serinprotease nachgewiesen, später die Prozessierung durch rekombinantes Kex2 (Abbildung 49A). Durch N-terminale Sequenzierung (Tabelle 38, identifizierte Sequenz fett gedruckt) konnte hier gezeigt werden, dass der Schnitt hinter der Membrandomäne an der unkonventionellen Schnittstelle SRAR<sub>152</sub>/DWIK stattfindet. Die Schnittstelle befindet sich zwischen zwei konservierten Teilen des Proteins (Tabelle 38) und durch diese Prozessierung wird das Membranprotein in ein lösliches konvertiert und der Re-Import in das Zytosol eingeleitet. Für *C. glabrata* Mns1 konnten Mora *et al.* eine entsprechende Prozessierung zeigen [137].

Das Mns1-Protein ist nicht das einzige ER-lokalisierte Protein, welches potentielle Kex2-Schnittstellen besitzt: die Schnittstelle in den untersuchten Organismen mit der höchsten Bewertung durch die PSSM ist Kar2. Kar2 (auch „BiP“) ist ein Hitzeschockprotein, welches über das Ire1-Hac1-System die „unfolded protein response“-Signalkaskade auslöst. Das Motiv ist hoch konserviert (Tabelle 39), was zunächst eine Lokalisation in einer strukturell unzugänglichen Region des Proteins nahe legt.

Aus dieser Gruppe konnte ScKar2 erfolgreich exprimiert werden, die orthologen Proteine aus *C. albicans* und *C. glabrata* hingegen nicht. Der *in vitro* Verdau zeigt, dass ScKar2 tatsächlich von Kex2 geschnitten werden kann (Abbildung 50A). Ob die Proteine des ER teilweise das Kex2-enhaltende TGN-Kompartiment erreichen, oder ob Kex2 auch bereits im ER aktiv sein kann, ist hingegen nicht klar. Eine theoretische Möglichkeit wäre, dass Kex2 mittels Abbau bestimmter eigentlich ER-residenter Proteine eine Kontrollfunktion im Golgi ausübt.

## 5.7 Relevanz der *in vitro* Daten für die *in vivo* Situation

In der Literatur wurden verschiedene Methoden der Analyse der Substratspezifität von Kex2 beschrieben (siehe Abschnitt 1.6). Diese fallen grob in die zwei Kategorien *in vitro* und *in vivo*. Bei *in vitro* Versuchen wurden in der Regel synthetische Peptide mit der rekombinanten Kex2-Protease aus *S. cerevisiae* verdaut [24, 46]. Diese Versuche ließen einen Rückschluss auf die biochemische Affinität der S4-S1-Region zu. Da geeignete Substrate für die S1'-S4'-Region fehlen, war bislang keine Aussage über den Beitrag dieser potentiellen Bindungsregion gemacht werden. Außerdem lassen Versuche mit Peptiden keinen Schluss über den Einfluss der Faltung von Proteinen auf die Suszeptibilität gegenüber Proteolyse zu, die aber *in vivo* eine große Rolle spielt [70]. Bei *in vivo* Versuchen besteht dieses Problem nicht: nur korrekt gefaltete Proteine werden auch von Kex2 hydrolysiert. Eine elegante Möglichkeit besteht hier darin, die Zielproteine mit einem Marker (bspw. HA-tag, His-tag, etc.) zu versehen und die Prozessierung in Wildtyp und *kex2*-Mutante im Westernblot zu analysieren [158]. Allerdings kann durch die



Markierung eine Fehlfaltung nicht ausgeschlossen werden. Außerdem besteht jedoch die Möglichkeit, dass alternative Prozessierungen durch andere Proteasen (beispielsweise die Yapsine) stattfinden, und somit zu der falschen Annahme führen es handle sich primär um das Substrat einer anderen Protease. Auch kann die im Wildtyp beobachtete Prozessierung auf eine andere Protease zurückzuführen sein, die in der *kex2*-Mutante nur ausbleibt, das die Protease selbst ein Kex2-Substrat ist.

Ein weiteres Problem ergibt sich hier, wenn die detektierten Fragmente post-translational etwa durch Glykosylierungen modifiziert sind: durch die so veränderten Molekulargewichte lassen sich die Schnittstellen nicht mehr durch Messung der Fragmentgrößen bestimmen (siehe bspw. [158]).

Auch ein proteomischer Ansatz ist denkbar. Eine Auftrennung von sekretorischen Proteinen könnte hier unprozessierte Vorläuferproteine aufzeigen [177]. Eine Analyse des Zellwandproteoms ist allerdings schwierig, da es sich hier um stark glykosylierte und kovalent verankerte Protein handelt, die mit einfachen Methoden nicht zugänglich sind [178, 179].

In dieser Arbeit wurden daher sowohl Proteasen als auch ganze Substratproteine rekombinant hergestellt und der Verdau der Substrate unter *in vitro* Bedingungen analysiert. Bei diesem Ansatz bleibt die Frage, ob diese Proteine auch *in vivo* prozessiert werden, zwar ungeklärt, dafür verbindet dieses Vorgehen gegenüber den anderen möglichen Ansätzen die Vorteile der Kontrolle über das prozessierende Enzym mit einer Erweiterung der Aussagekraft über die Selektivität der S'-Region und die räumliche Struktur des Substrates.

## 6 Zusammenfassung

Die Protease Kex2 ist eine im Trans-Golgi-Netzwerk von Pilzen lokalisierte Subtilisin-ähnliche Endoprotease. Sie prozessiert im Transit befindliche Proteine an spezifischen Lysin-Arginin-Motive („KR“) schneidet und aktiviert jene auf diese Weise. Die biochemische Aktivität des Modellenzym Kex2 aus *Saccharomyces cerevisiae* ist in der Literatur gut charakterisiert, dennoch sind über die beiden klassischen Substrate Killertoxin und  $\alpha$ -Kreuzungspheromon hinaus nur wenige weitere bekannt. *KEX2*-Deletionsmutanten besitzen ein sehr pleiotropes Spektrum an Phänotypen, die auf eine allgemeine Schwächung der Zellwand- und Plasmamembranintegrität hinweisen. Bei den beiden humanpathogenen Pilzen *Candida albicans* und *C. glabrata* führt dies zu einer reduzierten Virulenz. Die Phänotypen der Mutanten lassen sich jedoch nicht durch einen fehlende Aktivierung der wenigen bekannten Substrate von Kex2 erklären.

Ziel dieser Arbeit war es daher, die Prozessierung vermuteter Substraten zu verifizieren und neue Kex2-Substrate bei den humanpathogenen Hefen *C. albicans* und *C. glabrata* zu identifizieren, um so das breite Spektrum an Phänotypen von Kex2-defizienten Mutanten besser erklären zu können.

Die in einem einzelnen Pilzgenom kodierten Proteine enthalten ca. 20.000 bis 25.000 Motive, die Kex2-Schnittstellen ähneln. Das Protein Kex2 ist in der Zelle allerdings in einem späten Kompartiment des Trans-Golgi-Netzwerks lokalisiert. Mit Hilfe eines in dieser Arbeit entwickelten Algorithmus zur subzellulären Lokalisationsvorhersage konnte die Anzahl an potentiellen Schnittstellen auf 214 Motive bei *C. albicans* und 163 in bei *C. glabrata* reduziert werden. Zur funktionellen Überprüfung der Schnittstellen in diesen Proteinen wurden zunächst sowohl die Kex2-Proteasen aus *C. albicans* und *C. glabrata* als auch zur Kontrolle die Kex2-Proteasen aus *S. cerevisiae* und der biotechnologisch relevanten Hefe *Pichia pastoris* rekombinant exprimiert und aufgereinigt. Im zweiten Schritt wurden dann 23 ebenfalls rekombinante potentielle Substratproteine durch *in vitro* Verdau analysiert, wobei zwölf davon die erwartete Prozessierung zeigten. Eine statistische Auswertung der die Schnittstelle umgebenden Aminosäuren zeigte, dass im C-terminal zum KR-Motiv gelegenen Bereich der tatsächlich prozessierten Substrate negativ geladene Aminosäuren überrepräsentiert sind.

Dieses spiegelt sich auch in der räumlichen Struktur des Substratbindungszentrums von ScKex2 wieder. Ein Vergleich der Struktur mit bekannten Substratbindungsstellen anderer Subtilisine zeigte, dass es auch in diesem Bereich potentielle Substratbindungsstellen gibt, die eine Präferenz für negativ geladene Aminosäuren erklären würden.

Obwohl sich die KEX2-Gene untereinander komplementieren können, sind sie doch evolutionär entfernt und könnten unterschiedliche Spezifitäten besitzen. Dies wird durch die Tatsache unterstrichen, dass bei dem Substrat CA0365 durch ScKex2 keine Prozessierung zu beobachten war, während es von CaKex2 mit sehr hoher Effizienz geschnitten wurde. Hierfür konnte allerdings unter den bekannten, an der Substraterkennung beteiligten Aminosäuren keine Erklärung gefunden werden. Dies weist darauf hin, dass es weitere, bisher unbekannte Mechanismen zur Substraterkennung geben muss.

Als ein entscheidender Faktor für die Prozessierung über die Aminosäuresequenz der Schnittstelle hinaus, und damit ein weiteres Kriterium für die *in silico* Suche nach Schnittstellen, hat sich die räumliche Struktur des Substrates herausgestellt. Bei allen nicht prozessierten Motiven konnte in den Strukturen homologer Proteine gezeigt werden, dass sie sich in unflexiblen Regionen befinden, während sich prozessierbare Schnittstellen in zugänglichen Bereichen befanden. Ebenso konnte gezeigt werden, dass hitzedenaturierte Substratproteine weniger effizient geschnitten wurde als native. Dies deutet darauf hin, dass ein Protein an einem prozessierbaren Motiv auch eine prozessierbare Struktur ausbilden muss um zum Substrat zu werden. Tatsächlich konnte eine Überrepräsentation von flexiblen, an der Proteinoberfläche gelegenen „Coil“-Strukturen und eine Unterrepräsentation der eher rigiden  $\beta$ -Faltblattstruktur in prozessierbaren Motiven gefunden werden.

In dieser Arbeit konnte gezeigt werden, dass Proteine aus verschiedenen funktionellen Gruppen von Kex2 prozessiert werden können. Von besonderem Interesse ist darunter das putative Prohormon CaEce1. CaEce1 ist das am stärksten exprimierte hyphenassoziierte Protein aus *C. albicans*, dessen Funktion bislang ungeklärt ist. Aufgrund der Kex2-prozessierbaren Prohormonstruktur kann postuliert werden, dass es sich möglicherweise um ein Hydrophobin, ähnlich Rep1 aus *Ustilago maydis*, handelt. Weiterhin konnte eine Prozessierung für die Glykosidase-ähnlichen Proteine der Sun-Familie gezeigt werden. Deren Deletionsmutanten besitzen Phänotypen die denen der Deletionsmutanten von *KEX2* sehr ähnlich sind. In *kex2*-Deletionsmutanten ist auch die sexuelle Reproduktion beeinträchtigt. Hier konnte gezeigt werden, dass eine Gruppe von Ops4-homologen Proteinen durch Kex2 prozessiert werden können, die während dieses Prozesses differentiell reguliert werden. Daher könnte die Kreuzungsdefizienz der *kex2*-Mutanten wesentlich mehr Ursachen haben als nur das Ausbleiben der Prozessierung des  $\alpha$ -Pheromons.

## 7 Summary

The fungal Kex2 protease is a subtilisin-like endoprotease localized in the trans-Golgi network, which proteolytically activates other proteins in transit at specific lysine-arginine-motifs ("KR"). Despite the well characterized biochemical activity of Kex2 from *Saccharomyces cerevisiae*, only few substrates are known besides killer toxins and the  $\alpha$ -mating pheromones. Among those, however, are representatives of important secretory enzyme classes and structural protein families. This is also reflected in the pleiotropic phenotypes of *kex2*-deletion. Mutants show a general weakening of cell wall and plasma membrane integrity, which in the case of the two human pathogenic fungi *Candida albicans* and *C. glabrata* leads to severely reduced virulence.

The aim of this work was to confirm suspected substrates of the human pathogenic yeasts *C. albicans* and *C. glabrata* and to identify new, previously unknown Kex2 substrates, thus explaining the wide range of phenotypes of Kex2-deficient mutants.

The proteins encoded in a single fungal genome contain together about 20,000 to 25,000 motifs similar to Kex2 cleavage sites. However, as the Kex2 protein is localized in a late compartment of the trans-Golgi network, it can only cleave proteins passing through this compartment. Using a prediction algorithm for protein subcellular localization developed in this work, the number of potential cleavage sites could thus be reduced to 214 in *C. albicans* and 163 in *C. glabrata*.

To verify the predicted cleavage sites, the Kex2 proteases from *C. albicans* and *C. glabrata* and as controls the Kex2 proteases from *S. cerevisiae* and the biotechnologically important yeast *Pichia pastoris* were recombinantly expressed and purified. In a second step, 23 recombinant potential protein substrates were analysed by *in vitro* digestion. Twelve of them showed the expected processing. A statistical analysis of the surrounding amino acids showed that negatively charged amino acids are overrepresented C-terminal to the KR-motif of actually processed substrates. This is also reflected in the spatial structure of the substrate binding cleft of ScKex2. A comparison of this part of the Kex2 structure with known substrate binding sites of other subtilisins showed, that potential substrate binding sites with preference for negatively charged amino acids are present.

Although the *KEX2* genes can complement each other, they are evolutionary distant and may have different specificities. This is stressed by the fact that one of the substrates tested here (CA0365) was processed differently by the proteinases: no processing at all was observed with

---

ScKex2 while CaKex2 cleaved it efficiently. An inspection of the amino acids involved in substrate recognition showed no differences. This might indicate, that other, previously unknown mechanisms of substrate recognition exist.

Besides the amino acid sequence of the cleavage site, the spatial structure of the substrate was found to be a crucial factor in the processing of the substrates. For all motifs not processed, it was demonstrated in the solved spatial structures of homologous proteins, that they lay in inflexible regions. Cleavable motifs were, in comparison, localized in accessible and flexible areas. Similarly, it was shown that heat denatured substrate proteins were cut less effective than native ones.

This suggests that a protein must form an accessible structure in order to become a substrate to Kex2. In fact, an over-representation of flexible, protein surface localized "coil"-structures and an under-representation of the more rigid  $\beta$ -sheet structure were found in cleavable motifs. Structural factors are thus a further criterion to be included to *in silico* searches for substrates.

It was shown in this work that proteins from various functional groups can be processed by Kex2. Here, the putative pro-hormone CaEce1 is of particular interest: it is the most highly expressed hyphae associated protein of *C. albicans*, whose function is currently unknown. Due to the Kex2-cleavable pro-hormone like structure, it can be postulated that it may represent a hydrophobin similar to Rep1 from *Ustilago maydis*. Furthermore, a processing for the glycosidase-like proteins of the Sun-family was shown. Their deletion mutants show phenotypes of reduced cell wall integrity similar to those of *KEX2* deletion mutants. The deletion of Kex2 also affects the sexual reproduction, which has been attributed to the lack of properly processed  $\alpha$ -pheromone. Here, it was shown that a group of Ops4-homologous proteins are processed by Kex2, which are differentially regulated during this process. Thus, possibly the mating deficiency of the *KEX2*-deletion mutant is more severe than thought.

## 8 Danksagung

Die praktischen Arbeiten für diese Dissertation wurden am Robert Koch-Institut in Berlin in der Gruppe „NG4/FG16“ unter der Leitung von Prof. Dr. Bernhard Hube durchgeführt.

Mein besonderer Dank gilt meinem Doktorvater Bernhard Hube für die Möglichkeit, dieses Thema aus meiner Diplomarbeit weiterzuführen, für die Bereitstellung des Arbeitsplatzes sowie die Finanzierung und die Betreuung dieser Arbeit. Ferner möchte ich mich für die Ermöglichung der Teilnahme an verschiedenen Tagungen im In- und Ausland bedanken.

Herrn Prof. Dr. Rupert Mutzel möchte ich für die Übernahme des Gutachtens dieser Arbeit danken.

Verschiedene Leute haben mir an unterschiedlichen Stellen dieser Arbeit helfend zur Seite gestanden:

Anti übernahm die Kultivierung von *Cryptococcus* Zellen. Yannick Krauke stellte im Rahmen seiner Diplomarbeit das Kex2 Enzym aus *Pichia pastoris* her. Guy Boileau stellte den Klon zur Herstellung des *S. cerevisiae* Kex2 Enzyms und Jeff Brodsky das Plasmid zur Expression von ScKar2 zur Verfügung. Stefan Bentink, Sascha Brunke und Stefan Biere halfen, diverse Computerprobleme zu überwinden. Tobi, Marcel und Axel halfen, die Westernblots und die bakterielle Expression zum Laufen zu bringen und häufig mit Antikörper und anderem Material aus.

Mein Dank gilt allen Kollegen der Arbeitsgruppe (Abi, Antje, Chantal, Donika, Fred, Rebekka, Ricardo, Sabine, Sascha B., Sascha T., Stefan, Wiebke und Yannick) für die gute Zusammenarbeit, und auch allen, die ich vielleicht vergessen habe zu erwähnen!

Mein besonderer Dank gilt Stefan und Katherina, Higi und Johannes, Marcel und Stefan Bentink für die persönliche Unterstützung in den etwas dunkleren Stunden!

Mein größter Dank jedoch gilt meiner Mutter und meiner Frau Maram für uneingeschränkte Unterstützung während der ganzen Zeit.

## 9 Literatur

1. Roberts, S., *A Clinician's Guide to Fungal Disease*. Infectious Diseases and Antimicrobial agents, 1984. **5**.
2. Borg-von Zepelin, M., et al., *Epidemiology and antifungal susceptibilities of Candida spp. to six antifungal agents: results from a surveillance study on fungaemia in Germany from July 2004 to August 2005*. J Antimicrob Chemother, 2007. **60**(2): p. 424-8.
3. Fridkin, S.K., et al., *The role of understaffing in central venous catheter-associated bloodstream infections*. Infect Control Hosp Epidemiol, 1996. **17**(3): p. 150-8.
4. Staib, P., et al., *Transcriptional regulators Cph1p and Efg1p mediate activation of the Candida albicans virulence gene SAP5 during infection*. Infect Immun, 2002. **70**(2): p. 921-7.
5. Abaitua, F., et al., *In vitro survival and germination of Candida albicans in the presence of nitrogen compounds*. Microbiology, 1999. **145 ( Pt 7)**: p. 1641-7.
6. Nantel, A., et al., *Transcription profiling of Candida albicans cells undergoing the yeast-to-hyphal transition*. Mol Biol Cell, 2002. **13**(10): p. 3452-65.
7. Hull, C.M. and A.D. Johnson, *Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans*. Science, 1999. **285**(5431): p. 1271-5.
8. Hull, C.M., R.M. Raisner, and A.D. Johnson, *Evidence for mating of the "asexual" yeast Candida albicans in a mammalian host*. Science, 2000. **289**(5477): p. 307-10.
9. Magee, B.B. and P.T. Magee, *Induction of mating in Candida albicans by construction of MTL $\alpha$  and MTL $\alpha$  strains*. Science, 2000. **289**(5477): p. 310-3.
10. Bennett, R.J., et al., *Identification and characterization of a Candida albicans mating pheromone*. Mol Cell Biol, 2003. **23**(22): p. 8189-201.
11. Slutsky, B., et al., *"White-opaque transition": a second high-frequency switching system in Candida albicans*. J Bacteriol, 1987. **169**(1): p. 189-97.
12. Miller, M.G. and A.D. Johnson, *White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating*. Cell, 2002. **110**(3): p. 293-302.
13. Odds, F.C., et al., *Candida concentrations in the vagina and their association with signs and symptoms of vaginal candidosis*. J Med Vet Mycol, 1988. **26**(5): p. 277-83.
14. Haley, L.D., *Yeasts of medical importance*. Am J Clin Pathol, 1961. **36**: p. 227-34.
15. Stenderup, A. and G.T. Pedersen, *Yeasts of human origin*. Acta Pathol Microbiol Scand, 1962. **54**: p. 462-72.
16. Pfaller, M.A., et al., *Hospital specificity, region specificity, and fluconazole resistance of Candida albicans bloodstream isolates*. J Clin Microbiol, 1998. **36**(6): p. 1518-29.
17. Pfaller, M.A., et al., *Trends in species distribution and susceptibility to fluconazole among blood stream isolates of Candida species in the United States*. Diagn Microbiol Infect Dis, 1999. **33**(4): p. 217-22.
18. Bodey, G.P., et al., *The epidemiology of Candida glabrata and Candida albicans fungemia in immunocompromised patients with cancer*. Am J Med, 2002. **112**(5): p. 380-5.
19. Wong, S., et al., *Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata*. Genome Biol, 2003. **4**(2): p. R10.
20. Pfaller, M.A., et al., *In vitro susceptibilities of rare Candida bloodstream isolates to ravuconazole and three comparative antifungal agents*. Diagn Microbiol Infect Dis, 2004. **48**(2): p. 101-5.
21. Rawlings, N.D., et al., *MEROPS: the peptidase database*. Nucleic Acids Res, 2007.

22. Turba, F. and G. Gundlach, [*Amino acid sequence in the area of the reactive serine group of the chymotrypsin molecule.*]. *Biochem Z*, 1955. **327**(3): p. 186-8.
23. Blow, D.M. and J.M. Smith, *Enzyme substrate and inhibitor interactions*. *Philos Trans R Soc Lond B Biol Sci*, 1975. **272**(915): p. 87-97.
24. Bevan, A., C. Brenner, and R.S. Fuller, *Quantitative assessment of enzyme specificity in vivo: P2 recognition by Kex2 protease defined in a genetic system*. *Proc Natl Acad Sci U S A*, 1998. **95**(18): p. 10384-9.
25. Wilcox, C.A. and R.S. Fuller, *Posttranslational processing of the prohormone-cleaving Kex2 protease in the Saccharomyces cerevisiae secretory pathway*. *J Cell Biol*, 1991. **115**(2): p. 297-307.
26. Brenner, C. and R.S. Fuller, *Structural and enzymatic characterization of a purified prohormone-processing enzyme: secreted, soluble Kex2 protease*. *Proc Natl Acad Sci U S A*, 1992. **89**(3): p. 922-6.
27. Julius, D., et al., *Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor*. *Cell*, 1984. **37**(3): p. 1075-89.
28. Holyoak, T., et al., *2.4 A resolution crystal structure of the prototypical hormone-processing protease Kex2 in complex with an Ala-Lys-Arg boronic acid inhibitor*. *Biochemistry*, 2003. **42**(22): p. 6709-18.
29. Seidah, N.G., A.M. Khatib, and A. Prat, *The proprotein convertases and their implication in sterol and/or lipid metabolism*. *Biol Chem*, 2006. **387**(7): p. 871-7.
30. Bussey, H., *Proteases and the processing of precursors to secreted proteins in yeast*. *Yeast*, 1988. **4**(1): p. 17-26.
31. Shinde, U. and M. Inouye, *Intramolecular chaperones and protein folding*. *Trends Biochem Sci*, 1993. **18**(11): p. 442-6.
32. Newport, G. and N. Agabian, *KEX2 influences Candida albicans proteinase secretion and hyphal formation*. *J Biol Chem*, 1997. **272**(46): p. 28954-61.
33. Newport, G., et al., *Inactivation of Kex2p diminishes the virulence of Candida albicans*. *J Biol Chem*, 2003. **278**(3): p. 1713-20.
34. Bader, O., *unveröffentlichte Daten*. 2000.
35. Bader, O., et al., *The KEX2 gene of Candida glabrata is required for cell surface integrity*. *Mol Microbiol*, 2001. **41**(6): p. 1431-44.
36. Costi, *persönliche Mitteilung*. 2002.
37. Sohn, Y.S., et al., *Disruption of PMRI, encoding a Ca<sup>2+</sup>-ATPase homolog in Yarrowia lipolytica, affects secretion and processing of homologous and heterologous proteins*. *J Bacteriol*, 1998. **180**(24): p. 6736-42.
38. Mizutani, O., et al., *Disordered cell integrity signaling caused by disruption of the kexB gene in Aspergillus oryzae*. *Eukaryot Cell*, 2004. **3**(4): p. 1036-48.
39. Punt, P.J., et al., *The role of the Aspergillus niger furin-type protease gene in processing of fungal proproteins and fusion proteins. Evidence for alternative processing of recombinant (fusion-) proteins*. *J Biotechnol*, 2003. **106**(1): p. 23-32.
40. Werten, M.W. and F.A. de Wolf, *Reduced proteolysis of secreted gelatin and Yps1-mediated alpha-factor leader processing in a Pichia pastoris kex2 disruptant*. *Appl Environ Microbiol*, 2005. **71**(5): p. 2310-7.
41. Ladds, G. and J. Davey, *Identification of proteases with shared functions to the proprotein processing protease Krp1 in the fission yeast Schizosaccharomyces pombe*. *Mol Microbiol*, 2000. **38**(4): p. 839-53.
42. Goller, S.P., et al., *Role of endoproteolytic dibasic proprotein processing in maturation of secretory proteins in Trichoderma reesei*. *Appl Environ Microbiol*, 1998. **64**(9): p. 3202-8.



43. Schechter, I. and A. Berger, *On the size of the active site in proteases. I. Papain*. Biochem Biophys Res Commun, 1967. **27**(2): p. 157-62.
44. Sun, H., R.C. Panicker, and S.Q. Yao, *Activity based fingerprinting of proteases using FRET peptides*. Biopolymers, 2007. **88**(2): p. 141-9.
45. Patel, D., et al., *In vitro assay for site-specific proteases using bead-attached GFP substrate*. Biotechniques, 2001. **31**(5): p. 1194, 1196, 1198 passim.
46. Rockwell, N.C., et al., *Internally consistent libraries of fluorogenic substrates demonstrate that Kex2 protease specificity is generated by multiple mechanisms*. Biochemistry, 1997. **36**(7): p. 1912-7.
47. Yun Kim, S., et al., *Detection of site-specific proteolysis in secretory pathways*. Biochem Biophys Res Commun, 2002. **296**(2): p. 419-24.
48. Suzuki, Y., et al., *Effect of amino acid substitution at the P(3) and P(4) subsites of fusion proteins on kex2 protease activity*. Biotechnol Appl Biochem, 2000. **32** ( Pt 1): p. 53-60.
49. Krysan, D.J., N.C. Rockwell, and R.S. Fuller, *Quantitative characterization of furin specificity. Energetics of substrate discrimination using an internally consistent set of hexapeptidyl methylcoumarinamides*. J Biol Chem, 1999. **274**(33): p. 23229-34.
50. Siezen, R.J., et al., *Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases*. Protein Eng, 1991. **4**(7): p. 719-37.
51. Siezen, R.J., J.W. Creemers, and W.J. Van de Ven, *Homology modelling of the catalytic domain of human furin. A model for the eukaryotic subtilisin-like proprotein convertases*. Eur J Biochem, 1994. **222**(2): p. 255-66.
52. Lipkind, G., Q. Gong, and D.F. Steiner, *Molecular modeling of the substrate specificity of prohormone convertases SPC2 and SPC3*. J Biol Chem, 1995. **270**(22): p. 13277-84.
53. Henrich, S., et al., *The crystal structure of the proprotein processing proteinase furin explains its stringent specificity*. Nat Struct Biol, 2003. **10**(7): p. 520-6.
54. Holyoak, T., et al., *Structural basis for differences in substrate selectivity in Kex2 and furin protein convertases*. Biochemistry, 2004. **43**(9): p. 2412-21.
55. Than, M.E., et al., *The endoprotease furin contains two essential Ca<sup>2+</sup> ions stabilizing its N-terminus and the unique S1 specificity pocket*. Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 5): p. 505-12.
56. Fuller, R.S., A. Brake, and J. Thorner, *Yeast prohormone processing enzyme (KEX2 gene product) is a Ca<sup>2+</sup>-dependent serine protease*. Proc Natl Acad Sci U S A, 1989. **86**(5): p. 1434-8.
57. Rholam, M., et al., *Role of amino acid sequences flanking dibasic cleavage sites in precursor proteolytic processing. The importance of the first residue C-terminal of the cleavage site*. Eur J Biochem, 1995. **227**(3): p. 707-14.
58. Cai, X.H., Q. Zhang, and D.F. Ding, *Rational redesign of inhibitors of furin/kexin processing proteases by electrostatic mutations*. Acta Pharmacol Sin, 2004. **25**(12): p. 1712-818.
59. Komiyama, T., *Interaction of EGLIN C variants with the extended subsites of the precursor processing proteases*. Protein Pept Lett, 2005. **12**(5): p. 415-20.
60. Brenner, C., A. Bevan, and R.S. Fuller, *One-step site-directed mutagenesis of the Kex2 protease oxyanion hole*. Curr Biol, 1993. **3**(8): p. 498-506.
61. Rozan, L., et al., *Plasticity of extended subsites facilitates divergent substrate recognition by Kex2 and furin*. J Biol Chem, 2004. **279**(34): p. 35656-63.
62. Han, H.E., et al., *Engineering of Kex2 variants exhibiting altered substrate specificity*. Biochem Biophys Res Commun, 2005. **337**(4): p. 1102-6.
63. Kacprzak, M.M., et al., *Mutations of the PC2 substrate binding pocket alter enzyme specificity*. J Biol Chem, 2005. **280**(36): p. 31850-8.

64. Duckert, P., S. Brunak, and N. Blom, *Prediction of proprotein convertase cleavage sites*. Protein Eng Des Sel, 2004. **17**(1): p. 107-12.
65. Yang, Z.R., et al., *Searching for discrimination rules in protease proteolytic cleavage activity using genetic programming with a min-max scoring function*. Biosystems, 2003. **72**(1-2): p. 159-76.
66. Backes, C., et al., *GraBCas: a bioinformatics tool for score-based prediction of Caspase- and Granzyme B-cleavage sites in protein sequences*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W208-13.
67. Yang, Z.R., et al., *A bio-basis function neural network for protein peptide cleavage activity characterisation*. Neural Netw, 2006. **19**(4): p. 401-7.
68. Yang, Z.R. and K.C. Chou, *Bio-support vector machines for computational proteomics*. Bioinformatics, 2004. **20**(5): p. 735-41.
69. Blom, N., et al., *Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks*. Protein Sci, 1996. **5**(11): p. 2203-16.
70. Hubbard, S.J., *The structural aspects of limited proteolysis of native proteins*. Biochim Biophys Acta, 1998. **1382**(2): p. 191-206.
71. Stultz, C.M., *Localized unfolding of collagen explains collagenase cleavage near imino-poor sites*. J Mol Biol, 2002. **319**(5): p. 997-1003.
72. Dosztanyi, Z., et al., *Prediction of protein disorder at the domain level*. Curr Protein Pept Sci, 2007. **8**(2): p. 161-71.
73. Kleywegt, G.J. and A.T. Brunger, *Checking your imagination: applications of the free R value*. Structure, 1996. **4**(8): p. 897-904.
74. Novotny, J. and R.E. Bruccoleri, *Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility*. FEBS Lett, 1987. **211**(2): p. 185-9.
75. Cunningham, K.W. and W.T. Wickner, *Yeast KEX2 protease and mannosyltransferase I are localized to distinct compartments of the secretory pathway*. Yeast, 1989. **5**(1): p. 25-33.
76. Henkel, M.K., et al., *Endocytic delivery of intramolecularly quenched substrates and inhibitors to the intracellular yeast Kex2 protease1*. Biochem J, 1999. **341** ( Pt 2): p. 445-52.
77. Wilcox, C.A., et al., *Mutation of a tyrosine localization signal in the cytosolic tail of yeast Kex2 protease disrupts Golgi retention and results in default transport to the vacuole*. Mol Biol Cell, 1992. **3**(12): p. 1353-71.
78. Bryant, N.J. and A. Boyd, *Immunoisolation of Kex2p-containing organelles from yeast demonstrates colocalisation of three processing proteinases to a single Golgi compartment*. J Cell Sci, 1993. **106** ( Pt 3): p. 815-22.
79. Dmochowska, A., et al., *Yeast KEX1 gene encodes a putative protease with a carboxypeptidase B-like function involved in killer toxin and alpha-factor precursor processing*. Cell, 1987. **50**(4): p. 573-84.
80. Achstetter, T., *Regulation of alpha-factor production in Saccharomyces cerevisiae: a-factor pheromone-induced expression of the MF alpha 1 and STE13 genes*. Mol Cell Biol, 1989. **9**(10): p. 4507-14.
81. Sipos, G., et al., *Soi3p/Rav1p functions at the early endosome to regulate endocytic trafficking to the vacuole and localization of trans-Golgi network transmembrane proteins*. Mol Biol Cell, 2004. **15**(7): p. 3196-209.
82. Wybranietz, W.A. and U. Lauer, *Distinct combination of purification methods dramatically improves cohesive-end subcloning of PCR products*. Biotechniques, 1998. **24**(4): p. 578-80.
83. Hanahan, D. and M. Meselson, *Plasmid screening at high colony density*. Methods Enzymol, 1983. **100**: p. 333-42.

84. Kimata, K., et al., *cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose transporter gene in Escherichia coli*. Proc Natl Acad Sci U S A, 1997. **94**(24): p. 12914-9.
85. Studier, F.W., *Protein production by auto-induction in high density shaking cultures*. Protein Expr Purif, 2005. **41**(1): p. 207-34.
86. Franken, K.L., et al., *Purification of his-tagged proteins by immobilized chelate affinity chromatography: the benefits from the use of organic solvent*. Protein Expr Purif, 2000. **18**(1): p. 95-9.
87. Oganessian, N., S.H. Kim, and R. Kim, *On-column protein refolding for crystallization*. J Struct Funct Genomics, 2005. **6**(2-3): p. 177-82.
88. Lin-Cereghino, J., et al., *Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris*. Biotechniques, 2005. **38**(1): p. 44, 46, 48.
89. Simpson, *trapezoid approximation by trapezoid rule*, in Proc. R. Soc. 1956. p. 402.
90. Chang, C. and C. Lin, *LIBSVM: a library for support vector machines*. 2001.
91. Sonnhammer, E.L., G. von Heijne, and A. Krogh, *A hidden Markov model for predicting transmembrane helices in protein sequences*. Proc Int Conf Intell Syst Mol Biol, 1998. **6**: p. 175-82.
92. Emanuelsson, O., et al., *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence*. J Mol Biol, 2000. **300**(4): p. 1005-16.
93. Nielsen, H., et al., *A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites*. Int J Neural Syst, 1997. **8**(5-6): p. 581-99.
94. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder*. Bioinformatics, 2004. **20**(13): p. 2138-9.
95. Dror, O., et al., *MASS: multiple structural alignment by secondary structures*. Bioinformatics, 2003. **19** Suppl 1: p. i95-104.
96. Ahmad, S., M.M. Gromiha, and A. Sarai, *RVP-net: online prediction of real valued accessible surface area of proteins from single sequences*. Bioinformatics, 2003. **19**(14): p. 1849-51.
97. Sayle, R.A. and E.J. Milner-White, *RASMOL: biomolecular graphics for all*. Trends Biochem Sci, 1995. **20**(9): p. 374.
98. Sundaramoorthy, M., J. Ternner, and T.L. Poulos, *The crystal structure of chloroperoxidase: a heme peroxidase--cytochrome P450 functional hybrid*. Structure, 1995. **3**(12): p. 1367-77.
99. Cutfield, S.M., et al., *The structure of the exo-beta-(1,3)-glucanase from Candida albicans in native and bound forms: relationship between a pocket and groove in family 5 glycosyl hydrolases*. J Mol Biol, 1999. **294**(3): p. 771-83.
100. Hung, C.Y., et al., *A parasitic phase-specific adhesin of Coccidioides immitis contributes to the virulence of this respiratory Fungal pathogen*. Infect Immun, 2002. **70**(7): p. 3443-56.
101. Stark, M.J. and A. Boyd, *The killer toxin of Kluyveromyces lactis: characterization of the toxin subunits and identification of the genes which encode them*. EMBO J, 1986. **5**(8): p. 1995-2002.
102. Castillo, L., et al., *Functional analysis of the cysteine residues and the repetitive sequence of Saccharomyces cerevisiae Pir4/Cis3: the repetitive sequence is needed for binding to the cell wall beta-1,3-glucan*. Yeast, 2003. **20**(11): p. 973-83.
103. Weig, M., *persönliche Mitteilung*. 2007.
104. Bostian, K.A., et al., *Sequence of the preprotoxin dsRNA gene of type I killer yeast: multiple processing events produce a two-component toxin*. Cell, 1984. **36**(3): p. 741-51.

105. Zhu, Y.S., et al., *Kex2-dependent processing of yeast K1 killer preprotoxin includes cleavage at ProArg-44*. Mol Microbiol, 1992. **6**(4): p. 511-20.
106. Dignard, D., et al., *Expression in yeast of a cDNA copy of the K2 killer toxin gene*. Mol Gen Genet, 1991. **227**(1): p. 127-36.
107. Riffer, F., et al., *Mutational analysis of K28 preprotoxin processing in the yeast Saccharomyces cerevisiae*. Microbiology, 2002. **148**(Pt 5): p. 1317-28.
108. Kohno, M., et al., *Thermal stability of Rhizopus niveus lipase expressed in a kex2 mutant yeast*. J Biotechnol, 2000. **81**(2-3): p. 141-50.
109. Takahashi, S., M. Ueda, and A. Tanaka, *Independent production of two molecular forms of a recombinant Rhizopus oryzae lipase by KEX2-engineered strains of Saccharomyces cerevisiae*. Appl Microbiol Biotechnol, 1999. **52**(4): p. 534-40.
110. Kurjan, J. and I. Herskowitz, *Structure of a yeast pheromone gene (MF alpha): a putative alpha-factor precursor contains four tandem copies of mature alpha-factor*. Cell, 1982. **30**(3): p. 933-43.
111. Tao, J., et al., *Ustilago maydis KP6 killer toxin: structure, expression in Saccharomyces cerevisiae, and relationship to other cellular toxins*. Mol Cell Biol, 1990. **10**(4): p. 1373-81.
112. Wosten, H.A., et al., *A novel class of small amphipathic peptides affect aerial hyphal growth and surface hydrophobicity in Ustilago maydis*. EMBO J, 1996. **15**(16): p. 4274-81.
113. Bendtsen, J.D., et al., *Feature-based prediction of non-classical and leaderless protein secretion*. Protein Eng Des Sel, 2004. **17**(4): p. 349-56.
114. Hong, E.L., et al., *Gene Ontology annotations at SGD: new data sources and annotation methods*. Nucleic Acids Res, 2008. **36**(Database issue): p. D577-81.
115. Sickmann, A., et al., *The proteome of Saccharomyces cerevisiae mitochondria*. Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13207-12.
116. Nielsen, H. and A. Krogh, *Prediction of signal peptides and signal anchors by a hidden Markov model*. Proc Int Conf Intell Syst Mol Biol, 1998. **6**: p. 122-30.
117. Cokol, M., R. Nair, and B. Rost, *Finding nuclear localization signals*. EMBO Rep, 2000. **1**(5): p. 411-5.
118. Townsley, F.M., G. Frigerio, and H.R. Pelham, *Retrieval of HDEL proteins is required for growth of yeast cells*. J Cell Biol, 1994. **127**(1): p. 21-8.
119. Kall, L., A. Krogh, and E.L. Sonnhammer, *A combined transmembrane topology and signal peptide prediction method*. J Mol Biol, 2004. **338**(5): p. 1027-36.
120. De Groot, P.W., K.J. Hellingwerf, and F.M. Klis, *Genome-wide identification of fungal GPI proteins*. Yeast, 2003. **20**(9): p. 781-96.
121. Kronegg, J. and D. Buloz, *Detection/prediction of GPI cleavage site (GPI-anchor) in a protein (DGPI)*. 1999.
122. Borgese, N., et al., *Biogenesis of tail-anchored proteins*. Biochem Soc Trans, 2003. **31**(Pt 6): p. 1238-42.
123. Andrade, M.A., S.I. O'Donoghue, and B. Rost, *Adaptation of protein surfaces to subcellular location*. J Mol Biol, 1998. **276**(2): p. 517-25.
124. Hua, S. and Z. Sun, *Support vector machine approach for protein subcellular localization prediction*. Bioinformatics, 2001. **17**(8): p. 721-8.
125. Cui, Q., et al., *Esub8: a novel tool to predict protein subcellular localizations in eukaryotic organisms*. BMC Bioinformatics, 2004. **5**: p. 66.
126. Chen, S., et al., *Roles for the Drs2p-Cdc50p complex in protein transport and phosphatidylserine asymmetry of the yeast plasma membrane*. Traffic, 2006. **7**(11): p. 1503-17.

127. Brown, L.A. and A. Baker, *Peroxisome biogenesis and the role of protein import*. J Cell Mol Med, 2003. **7**(4): p. 388-400.
128. Heiland, I. and R. Erdmann, *Biogenesis of peroxisomes. Topogenesis of the peroxisomal membrane and matrix proteins*. Febs J, 2005. **272**(10): p. 2362-72.
129. Szafron, D., et al., *Proteome Analyst: custom predictions with explanations in a web-based tool for high-throughput proteome annotations*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W365-71.
130. Remm, M., C.E. Storm, and E.L. Sonnhammer, *Automatic clustering of orthologs and in-paralogs from pairwise species comparisons*. J Mol Biol, 2001. **314**(5): p. 1041-52.
131. Lesage, G., et al., *Mechanism of Kex2p inhibition by its proregion*. FEBS Lett, 2001. **508**(3): p. 332-6.
132. Murad, A.M., et al., *Clp10, an efficient and convenient integrating vector for Candida albicans*. Yeast, 2000. **16**(4): p. 325-7.
133. Stoldt, V.R., et al., *Efg1p, an essential regulator of morphogenesis of the human pathogen Candida albicans, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi*. Embo J, 1997. **16**(8): p. 1982-91.
134. Krauke, Y., *Heterologe Expression Subtilisin-ähnlicher Serinproteasen und Untersuchung der Substratspezifität*, in Diplomarbeit, FU Berlin. 2006.
135. Achstetter, T. and D.H. Wolf, *Hormone processing and membrane-bound proteinases in yeast*. EMBO J, 1985. **4**(1): p. 173-7.
136. Birse, C.E., et al., *Cloning and characterization of ECE1, a gene expressed in association with cell elongation of the dimorphic pathogen Candida albicans*. Infect Immun, 1993. **61**(9): p. 3648-55.
137. Mora, H., eingereicht bei Glycobiology, 2008.
138. Cabib, E., et al., *Crh1p and Crh2p are required for the cross-linking of chitin to beta(1-6)glucan in the Saccharomyces cerevisiae cell wall*. Mol Microbiol, 2007. **63**(3): p. 921-35.
139. Comellas-Bigler, M., et al., *1.2 A crystal structure of the serine carboxyl proteinase pro-kumamolisin; structure of an intact pro-subtilase*. Structure, 2004. **12**(7): p. 1313-23.
140. Henrich, S., et al., *Proprotein convertase models based on the crystal structures of furin and kexin: explanation of their specificity*. J Mol Biol, 2005. **345**(2): p. 211-27.
141. Bussey, H., et al., *Secretion of Saccharomyces cerevisiae killer toxin: processing of the glycosylated precursor*. Mol Cell Biol, 1983. **3**(8): p. 1362-70.
142. Staab, J.F., et al., *Adhesive and mammalian transglutaminase substrate properties of Candida albicans Hwp1*. Science, 1999. **283**(5407): p. 1535-8.
143. Conesa, A., et al., *C-terminal propeptide of the Caldariomyces fumago chloroperoxidase: an intramolecular chaperone?* FEBS Lett, 2001. **503**(2-3): p. 117-20.
144. Schwappach, B., et al., *Golgi localization and functionally important domains in the NH2 and COOH terminus of the yeast CLC putative chloride channel Gef1p*. J Biol Chem, 1998. **273**(24): p. 15110-8.
145. Larriba, G., et al., *Yeast exoglucanases. Where redundancy implies necessity*. Arch Med Res, 1993. **24**(3): p. 293-9.
146. Mrsa, V., et al., *Specific labelling of cell wall proteins by biotinylation. Identification of four covalently linked O-mannosylated proteins of Saccharomyces cerevisiae*. Yeast, 1997. **13**(12): p. 1145-54.
147. Emanuelsson, O., et al., *Locating proteins in the cell using TargetP, SignalP and related tools*. Nat Protoc, 2007. **2**(4): p. 953-71.
148. Monteoliva, L., et al., *Large-scale identification of putative exported proteins in Candida albicans by genetic selection*. Eukaryot Cell, 2002. **1**(4): p. 514-25.

149. Ecker, M., et al., *Pir proteins of Saccharomyces cerevisiae are attached to beta-1,3-glucan by a new protein-carbohydrate linkage*. J Biol Chem, 2006. **281**(17): p. 11523-9.
150. Gluschankof, P. and R.S. Fuller, *A C-terminal domain conserved in precursor processing proteases is required for intramolecular N-terminal maturation of pro-Kex2 protease*. EMBO J, 1994. **13**(10): p. 2280-8.
151. Henriksen, A., et al., *Major venom allergen of yellow jackets, Ves v 5: structural characterization of a pathogenesis-related protein superfamily*. Proteins, 2001. **45**(4): p. 438-48.
152. Nishikawa, S., J.L. Brodsky, and K. Nakatsukasa, *Roles of molecular chaperones in endoplasmic reticulum (ER) quality control and ER-associated degradation (ERAD)*. J Biochem, 2005. **137**(5): p. 551-5.
153. Pascarella, S. and P. Argos, *Analysis of insertions/deletions in protein structures*. J Mol Biol, 1992. **224**(2): p. 461-71.
154. Jiang, J., et al., *Structural basis of interdomain communication in the Hsc70 chaperone*. Mol Cell, 2005. **20**(4): p. 513-24.
155. Panchenko, A.R. and T. Madej, *Structural similarity of loops in protein families: toward the understanding of protein evolution*. BMC Evol Biol, 2005. **5**(1): p. 10.
156. Vallee, F., et al., *Crystal structure of a class I alpha1,2-mannosidase involved in N-glycan processing and endoplasmic reticulum quality control*. Embo J, 2000. **19**(4): p. 581-8.
157. Hahn, M., et al., *Native-like in vivo folding of a circularly permuted jellyroll protein shown by crystal structure analysis*. Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10417-21.
158. Heiman, M.G., A. Engel, and P. Walter, *The Golgi-resident protease Kex2 acts in conjunction with Prm1 to facilitate cell fusion during yeast mating*. J Cell Biol, 2007. **176**(2): p. 209-22.
159. Wachter, A. and B. Schwappach, *The yeast CLC chloride channel is proteolytically processed by the furin-like protease Kex2p in the first extracellular loop*. FEBS Lett, 2005. **579**(5): p. 1149-53.
160. Cappellaro, C., V. Mrsa, and W. Tanner, *New potential cell wall glucanases of Saccharomyces cerevisiae and their involvement in mating*. J Bacteriol, 1998. **180**(19): p. 5030-7.
161. Velours, G., et al., *Dual cell wall/mitochondria localization of the 'SUN' family proteins*. FEMS Microbiol Lett, 2002. **207**(2): p. 165-72.
162. Larriba, G., et al., *Molecular biology of yeast exoglucanases*. FEMS Microbiol Lett, 1995. **125**(2-3): p. 121-6.
163. Kalebina, T.S., et al., *Deletion of BGL2 results in an increased chitin level in the cell wall of Saccharomyces cerevisiae*. Antonie Van Leeuwenhoek, 2003. **84**(3): p. 179-84.
164. Martinez, A.I., et al., *Role of Pir1 in the construction of the Candida albicans cell wall*. Microbiology, 2004. **150**(Pt 10): p. 3151-61.
165. Firon, A., et al., *The SUN41 and SUN42 genes are essential for cell separation in Candida albicans*. Mol Microbiol, 2007. **66**(5): p. 1256-75.
166. Hiller, E., et al., *Candida albicans Sun41p, a putative glycosidase, is involved in morphogenesis, cell wall biogenesis, and biofilm formation*. Eukaryot Cell, 2007. **6**(11): p. 2056-65.
167. Mouassite, M., et al., *The 'SUN' family: yeast SUN4/SCW3 is involved in cell septation*. Yeast, 2000. **16**(10): p. 905-19.
168. Barwell, K.J., et al., *Relationship of DFG16 to the Rim101p pH response pathway in Saccharomyces cerevisiae and Candida albicans*. Eukaryot Cell, 2005. **4**(5): p. 890-9.

169. Stadler, J.A. and R.J. Schweyen, *The yeast iron regulon is induced upon cobalt stress and crucial for cobalt tolerance*. J Biol Chem, 2002. **277**(42): p. 39649-54.
170. Miura, S., et al., *Screening of genes involved in isooctane tolerance in Saccharomyces cerevisiae by using mRNA differential display*. Appl Environ Microbiol, 2000. **66**(11): p. 4883-9.
171. Slattery, M.G., D. Liko, and W. Heideman, *The function and properties of the Azf1 transcriptional regulator change with growth conditions in Saccharomyces cerevisiae*. Eukaryot Cell, 2006. **5**(2): p. 313-20.
172. Panwar, S.L., et al., *MFalpha1, the gene encoding the alpha mating pheromone of Candida albicans*. Eukaryot Cell, 2003. **2**(6): p. 1350-60.
173. Lockhart, S.R., et al., *Alpha-pheromone-induced "shmooing" and gene regulation require white-opaque switching during Candida albicans mating*. Eukaryot Cell, 2003. **2**(5): p. 847-55.
174. Moreno, I., et al., *Global transcriptional profiling of Candida albicans cwt1 null mutant*. Yeast, 2007. **24**(4): p. 357-70.
175. Fernandez-Arenas, E., et al., *Integrated proteomics and genomics strategies bring new insight into Candida albicans response upon macrophage interaction*. Mol Cell Proteomics, 2007. **6**(3): p. 460-78.
176. Copping, V.M., et al., *Exposure of Candida albicans to antifungal agents affects expression of SAP2 and SAP9 secreted proteinase genes*. J Antimicrob Chemother, 2005. **55**(5): p. 645-54.
177. Schilling, O. and C.M. Overall, *Proteomic discovery of protease substrates*. Curr Opin Chem Biol, 2007. **11**(1): p. 36-45.
178. Rupp, S., *Proteomics on its way to study host-pathogen interaction in Candida albicans*. Curr Opin Microbiol, 2004. **7**(4): p. 330-5.
179. Yin, Q.Y., et al., *Mass spectrometry-based proteomics of fungal wall glycoproteins*. Trends Microbiol, 2008. **16**(1): p. 20-6.

## 10 Publikationsliste

### Wissenschaftliche Publikationen

- C Fradin, S Thewes, K Zakikhany, A Albrecht, **O Bader**, D Kunze und B Hube (2004). Transcriptional profiling of *Candida albicans* during infections. *Mikol Lek* **11**:157-163.
- C d'Enfert, S Goyard, S Rodriguez-Arnaveilhe, L Frangeul, L Jones, F Tekaia, **O Bader et al.** (2005) CandidaDB: a genome database for *Candida albicans* pathogenomics. *Nucleic Acids Res* **33**:D353-7.
- H Mora-Montes, **O Bader**, E Lopez-Romero, S Zinker, P Ponce-Noyola, B Hube, N Gow und A Flores-Carreón. Cytoplasmic alpha-1,2-mannosidase E-I from *Candida albicans*. Eingereicht bei *Glycobiology*.
- O Bader**, Y Krauke und B Hube. Processing of predicted substrates of fungal Kex2 proteinases from *Candida albicans*, *C. glabrata*, *Saccharomyces cerevisiae* and *Pichia pastoris*. Eingereicht bei *BMC Microbiology*.
- M Girgi, F Navid, U Klampferer, D Knoll, **O Bader**, W Schäfer und H Lörz. Isolation and expression regulation of a chitinase and a chitosanase gene from *Trichoderma harzianum*. In Vorbereitung.

### Übersichtsartikel

- J Naglik, A Albrecht, **O Bader** und B Hube (2004). *Candida albicans* proteinases and host/pathogen interactions. *Cell Microbiol* **6**:915-26
- CA Munro, C Fradin, **O Bader**, und B Hube (2005). Postgenomic approaches to analyse *Candida albicans* pathogenicity. In *The Mycota*, K Esser and AJP Brown Eds; Springer Heidelberg, Vol. XIII.

### Tagungsbeiträge

- Vortrag **O Bader** und B Hube (2004). DGHM Fachgruppe Eukaryontische Krankheitserreger. "In silico prediction of Kex2 processing sites in protein sequences: Extraction of sequences for ER-luminal and secretory proteins from genome databases".
- Poster MA Müller, **O Bader**, A Schulz, C Linke, L van der Hoek und M Niedrig (2005). Jahrestagung der Deutschen Virologischen Gesellschaft. "Heterologous expression of recombinant open reading frame 3 of human coronavirus NL63".
- Vortrag **O Bader** und B Hube (2005). British Society for Medical Mycology annual meeting. "Zymogen activation in pathogenic fungi: assembling the substratome of the Kex2 protease system".
- Poster **O Bader**, Y Krauke und B Hube (2007). Advanced FEBS Lecture Course on human fungal pathogens FEBS (HFP2007). "Predicting substrates of fungal Kex2 proteinases on genomic scale".



## 11 Anhang

### 11.1 Genebank Eintrag für *Cryptococcus neoformans KEX2*

LOCUS AY184224 3550 bp DNA linear PLN 29-JAN-2003  
 DEFINITION *Cryptococcus neoformans* A/D Kex2 gene, complete cds.  
 ACCESSION AY184224  
 VERSION AY184224.1 GI:28142391  
 KEYWORDS .  
 SOURCE *Cryptococcus neoformans* A/D  
 ORGANISM [Cryptococcus neoformans A/D](#)  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes;  
 Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae;  
 Filobasidiella.  
 REFERENCE 1 (bases 1 to 3550)  
 AUTHORS Bader,O., Antweiler,E., Lemmer,K., Tintelnot,K. and Hube,B.  
 TITLE Cloning and cDNA sequencing of the KEX2 gene from *Cryptococcus neoformans*  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3550)  
 AUTHORS Bader,O., Antweiler,E., Lemmer,K., Tintelnot,K. and Hube,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-NOV-2002) NG4, Robert Koch Institut, Nordufer 20, Berlin 13353, Germany  
 FEATURES Location/Qualifiers  
 source 1..3550  
 /organism="Cryptococcus neoformans A/D"  
 /mol\_type="genomic DNA"  
 /strain="ATCC 34544"  
 /serotype="AD"  
 /db\_xref="taxon:[192011](#)"  
 /note="teleomorph: *Filobasidiella neoformans* Kwon-Chung"  
[mRNA](#) join(<243..857,910..1443,1491..1710,1758..2214,2267..>3194)  
 /product="Kex2"  
[CDS](#) join(243..857,910..1443,1491..1710,1758..2214,2267..3194)  
 /note="prohormone convertase"  
 /codon\_start=1  
 /product="Kex2"  
 /protein\_id="[AAO25531.1](#)"  
 /db\_xref="GI:28142392"  
 /translation="MRTLSSLWGILLLALIVPPSLALQRPQPRSYDTHAYALELDP  
 SPAAALQLSKSLGVELVERIGELDGHVLRTEGWTPHEASITKRSVSHDPILKRWEAL  
 PSSLGKKSILTPLSLKQRAKRHSYSPRSRHSRDDRTELLYAQNELHLADPMLDQQWHL  
 INTQMKDIELNVTGLWGRGITGEGVHVVIIDDLGDLVESKDLKDNFFAEGSYDFNDHTE  
 LPILRLKDDQHGTTRCAGEIAAVPNDVCGVGVAYDSKIAGVRIILSAPISDADEAAALNY  
 AYQLNDIYSCSWGPPDDGRSMEAPDGLILKAMVNGVQKGRDGKGSVFVFAAGNGGGSD  
 DQC�FDGYTNSIFSVTVGAVDRKGLHPYYSEMCAAMMVVAPSSGSGDHIHTTDVGGKDK  
 CSHSHGGTSAAPLAVGVFALALSVRPDLTWRDIQHHLAVRHAVFFNPDDPAWELTAAG  
 RHFSYKYGYGKLDAGLFVEAAEKWQLVKEQTWYDPSVYLPTTSPADVTRRQDEAADG  
 PTSSDEETSNNPPVVEPSGSFITEDGVIISTYEVTSMLFDANFERLEHVTVRVWIDHQ  
 RRGDVEVELTSPNGVSVLCRQRREFDNADSGFPGWKFMSLKHWDENPVGTWTIKVKDQ  
 VNPDKTGRFVAVSLQLWGESVDPALAKLWAPAEEGQPDEEQTGSNPSTTVSQPKPFTA  
 LLPGDHGEASGEATQPGLGSATAHPQPTSTTGAGNVAEPTSPDADADEGFFSGISN  
 LASSSTWLAGAGAI I ILSGAAIGAFFFIRARRQKRNLFGLSNNGQARGAYEPVDDVQ  
 MSLLERGRKFKGSKSESQGTKDLYDAFGDGPSEDEEEDLDERALRYHDGFLEDDP  
 NEVGPKTEYKDEPESEPETFKDGEETVGTCKDKGKGPSEGESGSGSSSSWQDAADEE  
 ARV"

## ORIGIN

```

1 aagaaataag ctgcaaaaa gtcttaagct tgggatcaca acgaaacgaac gacgtgcgcy
61 ttgtcgaata cctatctaaa gaggaagtcg ttacgcaatg acctgtatat tcccacggta
121 ggtcaccacc cgcccgcccc cgcccgagcc aagttccagg cagagcagca ggtgaaatc
181 aagtatttat ttacatttgt atttacattt ccaattcaga tcgcatctgt aactcttgca
241 aaatgcgcac cttattatcc ctttggggga ttctctctggc gctcatagtg cctccatcgc
301 tcgcccgtca aagacctcaa ccaaggtcct acgatacgcg cgcttattac gccttgagc
361 tcgaccatc aatatcacca gcagccgccc tgcaactctc aaaatcttta ggcgtcgagc
421 tggtggaacg tataggagag ttggacggac attggcttgt caggactgaa ggggtggac
481 cagagcatgc gtcaataaca aaaagaagtg tttctcatga tccaatattg aagcgatggg
541 aagcattgcc ttcaagtctt ggcaagaaat ccctcacgcc cttgtcactc aagcaacgtg
601 ccaagcgaca taaatcatat tctcctcggt cccgtcattc aagagacgat agaacagagc
661 ttttatatgc ccaaaatgag ctgcatttgg cagaccctat gctcgatcag caatggcatc
721 tcatacaatc ccagatgaag gacatcagag tcaatgtcac tggcctttgg gggaggggta
781 ttactgggtg ggggtgtcac gtgggtgatc tagacgatgg actggatgta gagagcaaag
841 atttgaagga taattctcgt cgctcatcgt cgtcagctc atttagttaa actcactggc
901 tctatgcagt tcgctgaagg atcttacgac ttcaacgacc acactgagct tccgattcct
961 cgctcaaaag acgaccaaca tgggtactaga tgtgctggcg agattgctgc tgttcccaac
1021 gacgtgtgtg gagtcggcgt agcatatgat agcaaaatcg ccggtgtccg tatcctttcg
1081 ccaagcgaca cagatgcccga cgaagcagct gctctcaatt atgcctatca actcaacgac
1141 atttattctt gctcatgggg tctcctccgac gatgggaggt caatggaagc ccctgatggg
1201 ttgatcctca agcgatgggt gaacggtggt caaaagggac gagacggtaa aggatcgggt
1261 ttctgtgttcg ctgctggcaa cgggtgggga tcagacgac agtgtaat tgcaggat
1321 acgaaactga ttttctctgt cactgttggg gcggtagata gaaaaggatt acatccttac
1381 tattcagaga tgtgtgcagc catgatgggt gttgctcctt cttcaggcag tggagatcac
1441 attgtgggtc attgtgtctg tcccttgccg tcaattgact ttctctgtag catacaacag
1501 acgttggaaa ggataagtgc tcacacagcc atggcggaac ttctgctggc gcacctctcg
1561 ctggttgagt cttcgctctc gccctttccg tgcgccccga ccttacttgg cgagacattc
1621 aacatcttgc cgtgcggcat gctgttttct tcaaccctga tgatccagct tgggagctaa
1681 ctgctgctgg aagacatctt agctataaat gcaagttcct ttctacataa cgcatgtg
1741 ccgatatcat tctgcagatg gttatggaaa gcttgacgca ggtttgttcc ttgaagctgc
1801 taggtaacat caactcaaaa agccccaaac gtggatgac tctccatcgg tttatcttcc
1861 taccactctc cctgccgatg tcaccagacg tcaagacgaa gctgccgacg gccccacaag
1921 ctctgacgag gagacctcca acccgccgcc tgtggtcgag cccagtggtt ctttcattac
1981 agaagatggt gttatctcca cgtatgaagt cactcagctc atgctttttg atgccaactt
2041 tgagagactg gagcatgtca ccggttaggg ttggatagac catcagagga ggggtgatgt
2101 tgaggtggag cttaccagtc ccaatggggg ggtagtgct ttgtgcaggc agaggaggtt
2161 tgacaatgca gatagtggtt tccctggctg gaaatttatg tctttgaaagc attggtatgt
2221 ttaaatgctg ttttgatgta gaccaaattg atctgatgaa atgtagggat gagaaccggg
2281 taggtacatg caactcaaaa gtcaaaagacc aagtcaacc cgacaaaacc ggccgtttcg
2341 tcgcatggtc acttcagttg tggggagaa ctggtgatcc tgcccttggc aaactctggg
2401 caccgcgaga agaaggtcaa ccggatgaag agcaaacagg ttccaacccc agtactactg
2461 tcagccaaaa gcccaaacct acggcactcc ttcttgggga tcatggtgag gcttctgggt
2521 aagcgacca gccaggactc ggatctgcta cagcccatcc tcaaccacaa agcacgactg
2581 gtaacgctgg aaatgtcggg gagccaacca gccccacaga tgccgatgcc gacgaagggg
2641 tcttcagcgg tatttccaac ctgcgttcat cctctacatg gcttgaggcc gcaggtgcca
2701 ttatcatcct ctctgggtgt gctattgggt cctttttctt catccgtgcc cgacgacaga
2761 aagcgaacct ctttggctc tccaacaacg gtcaaggagc tcgcggtgct tacgagcctg
2821 ttgatgacgt gcagatgagt ctccctgaga gaggcaggag gaagtttggc aagagcaaga
2881 gtgagagtca aggaacgaaa gatttgtatg atgcctttgg agatggtccg agtgatgaag
2941 aggaggagga tttggatgag aggactgcat tgaggatca tgatggtttc ttagaggatg
3001 acgagccgaa tgaggtaggg cccaagacag agtacaagga tgagcctgag tctgagcctg
3061 agacctttaa agatggagag gaaactgtgg gaacaaaaga taagggtaaag gggaaaggtc
3121 caagtgaggg agagagtggt agcgggagtt ccagcagttg gcaagatgcc gccgacgaag
3181 aagcgcgtgt gtaagatggg ggtatcaacg agtcatggat gatgtactgt catatgtata
3241 tatcgcaatg tcaatcttt gcatcacatt cggaggtcat tcgagcttta tatgaatgcc
3301 ttgttattgt cctgctttcc acgtattcct tgaatatatc cagtccgcca gtccttctg
3361 tttcctcctc gcgcatttcc gttcagatgg ggtggagaaa gttatgacgt cacctttagg
3421 tgagttttgc cacttattgc aacgctgggt gttgtcgtcg ttcgttaata aaaatcaata
3481 gacagatgtc tgttcgagag aagtgaatag ccattgctcg aaagcggaaag agtgtagagc
3541 gatccagcaa

```

//

## 11.2 Bewertung von KR-Motiven

Tabelle 43: Bewertung putativer Kex2-Schnittstellen, sortiert nach Bewertung der Struktur.

	Pos	Schnittstelle	PSSM-Score	Experimentell	ProProtein		ProHormon		kein Substrat		Topologie
					p	+/-	p	+/-	p	+/-	
CAGLOE01815g	33	*KLDFEKSPKASDDLVKR/DDEYVNVPLK	151	1	78%	9%	9%	3%	13%	6%	16/(no TMs)(503-GPI)o519
CA1266	74	AHGQHHQSQQQQQPAQKR/GTVQTSLINE	167	1	78%	10%	6%	1%	16%	9%	26/(no TMs)405
CAGLOH03135g	85	LFVNTTLYNQATKGEKLSDFTKR/DANPDAAEAEA	544	2	36%	29%	48%	30%	16%	7%	21/(no TMs)159
CAGLOE01859g	29	*ALDFEKSHGDDLVKR/DVQDVKLINS	96	1	78%	9%	5%	1%	17%	9%	14/(no TMs)(484-GPI)o505
CAGLOE01837g	32	*KLDFTKTPGSDLAKR/DVVDPEAAQL	68	1	74%	13%	9%	4%	17%	10%	17/(no TMs)(498-GPI)o521
CA0968	75	FLDLLPTKIPVDDTGRNDGVDKR/GPVAVKLDNE	142	1	76%	13%	6%	1%	17%	11%	24/(no TMs)418
CA4863	37	*KLDFNKVSTPSKYTKR/DALPMLIND	426	1	72%	14%	10%	4%	17%	10%	21/(no TMs)(428-GPI)o453
CA2499	75	FLEFTPSEFPVDETGRDGDVDKR/GPVAVTLHNE	142	1	75%	13%	8%	2%	17%	11%	24/(no TMs)418
CAGLOE01881g	30	TAAAAGYLRLPFTKI TDNSLAKR/ADDDYLVNQL	85	1	76%	10%	6%	1%	18%	9%	(no TMs)(469-GPI)o508
CA2055	74	LFLDIEPTEIPVDDTGRNDVGKR/GPVAVKLDNE	369	1	76%	12%	6%	1%	18%	11%	19/(no TMs)417
CA1266	32	*EDIDKR/AEKVVSLDFT	45	1	74%	12%	6%	2%	20%	11%	26/(no TMs)405
CA2202	22	*IKR/GLLDDDFGTS	104	1	73%	14%	7%	2%	20%	12%	19/(no TMs)578
CA3138	55	FSVVKTPKAFPVTNGQEGKTSKR/QAVPVLHNE	103	1	76%	12%	4%	1%	20%	12%	19/(no TMs)398
YLR121C	46	DGKYVKIPFTKKKNGDNGELSKR/SNGHEKFLVA	54	1	74%	13%	5%	1%	22%	13%	21/(no TMs)(483-GPI)o508
CAGLOE01771g	33	*DFEKTPKISDELTKR/ADNDSFIAPL	82	1	70%	15%	7%	2%	22%	13%	18/(no TMs)(501-GPI)o519
CAGLOM13805g	30	*ANVHHEHKNKR/DVVTKTVHNQ	585	1	73%	13%	4%	1%	22%	13%	19/(no TMs)371
CA1402	227	QRLPQLITRSAESALKDSQPVKR/DAGSVALSNL	214	2	25%	22%	52%	25%	23%	10%	19/(no TMs)271
CAGLOE01749g	33	*KLDFEKVPEQNLAENKR/DDKGLQLELE	46	1	65%	19%	12%	7%	23%	13%	16/(no TMs)482
CAGLOL05434g	30	*SPKHDHEDRHAIKR/DVDVTVVVEY	202	1	70%	15%	6%	2%	24%	14%	16/(no TMs)(326-GPI)o346
CA0365	105	LAGNILNLIQLGAGAEISASTKR/DESAVDVEE	193	2	25%	19%	51%	27%	24%	10%	19/(no TMs)(145-GPI)o178
CA4700	105	TGSSDLWVMSHDLKCVSAPNSKR/NERSFGHGTTG	185	0	66%	31%	8%	7%	25%	24%	18/(no TMs)(520-GPI)o544
CA2660	49	VTLDLDFVIKTPVNATGQEGVKR/QALPVTLNNE	617	1	71%	15%	4%	2%	25%	16%	24/(no TMs)391
CAGLOC03696g	304	NSEVRSLSKSKSFYESDRPALKR/LHTSSTEYLN	212		67%	3%	6%	2%	26%	4%	o735-761i773-799o805-827i831

Anhang

CAGLOJ02288g	37	VVNCRLRTIEKRIVSSHASLSKR/SAVDLQFRRF	54	1	69%	16%	4%	1%	27%	17%	(no TMs)(508-GPI)o541
CA2912	129	DEHRKKGSKQYKRGEVENPRETR/DCDFCTIEKS	119		62%	13%	12%	8%	27%	6%	19/(no TMs)247
CA1402	193	ANAPRLISNVFDGVSSETVQQAQR/DGLEDFLDEL	309	2	19%	12%	54%	23%	27%	12%	19/(no TMs)271
CA4679	338	SGALRKREYNDAVEAALREIQKR/EEGIDDVEIA	293	0	58%	29%	14%	11%	27%	22%	20/(no TMs)(535-GPI)o557
CA4336	181	YGSSNDDSYGSSNDDSYGSSNRR/GNDSYGSSND	70		58%	8%	15%	6%	28%	5%	(no TMs)212
YMR244W	46	*SPSIQMTENTNQDHHEHAKR/GGTCAFPNYD	440		60%	12%	13%	2%	28%	10%	26/(no TMs)355
CAGLOE01419g	49	YCKNSDNKSRNLDALGSNRFKR/ASSVASASPS	362	1	65%	20%	6%	2%	28%	18%	o500-518i519
CAGLOL00979g	337	FYKEEGLCVDGPEFRYYPEVFER/EPYKANLLRS	59		50%	5%	22%	5%	28%	6%	26/(no TMs)475
YPR149W	160	LMSVFNMISNGAFSGSFSKRRR/TGQVGVPTIS	208		65%	6%	7%	2%	28%	6%	o12-28i40-59o65-93i125-146o173
CA1402	60	AAPADQAPTVPAPQEFNTAITKR/SIIGIIMGIL	203	2	25%	22%	47%	31%	28%	23%	19/(no TMs)271
CA0365	31	*IPQESTTEVDKR/LDADVAQQLA	178	2	21%	20%	50%	28%	29%	13%	19/(no TMs)(145-GPI)o178
CAGLOE01793g	32	*KLDFTTRVPSASVLEKR/ADDLSPVPLR	82	1	65%	20%	6%	1%	29%	19%	16/(no TMs)(480-GPI)o516
CA1873	68	ELVQHINNYKTRRDAIDEEIMKR/DYAIVTDVLA	234	1	59%	27%	12%	9%	29%	20%	21/(no TMs)351
CA1820	307	SYYYPPNYEEIFKDKNEPLKSER/STVRDSTSSP	39		61%	13%	9%	1%	30%	13%	i12-31o765
CAGLOF05137g	98	AQKKRSNLSEWQKMLDQHNKKR/ELHKDTDSL	346	0	63%	26%	7%	3%	30%	24%	18/(no TMs)227
CAGLOE00275g	449	RSHLAQFSVNLPSNDTQRPYSKR/NSFFPSNLS	283		60%	10%	10%	2%	30%	12%	20/(no TMs)880
YBR023C	85	QEQMNHLDVLPSSSTGVNPNATR/SGSLRSKGSV	92		62%	13%	8%	1%	30%	13%	o169-191i203-223o453-476i1029-1051o1057-1076i1083-1107o1165
CAGLOJ09922g	42	*MPDLKDCTTTAQGNHQHQR/AVAVEVYVET	386	1	66%	19%	4%	2%	30%	20%	23/(no TMs)427
YMR244W	328	RNGAKANIEVFDVSGSSSKGKR/ELNPLDVIT	593		54%	16%	16%	5%	30%	12%	26/(no TMs)355
CA6065	57	VIKTHKNVTGPQGEINTNVNKKR/QTVPVKLINE	420	1	66%	20%	4%	2%	30%	21%	19/(no TMs)398
CA0856	198	ECKKRKKVHRNYKRGYGSERR/QSDCESDCER	63		52%	13%	17%	2%	31%	12%	o28-45i57-78o356
CAGLOM02211g	68	VQQLGHKYVSQYKANPGTVLPR/DHLFPDQVH	59		58%	9%	12%	1%	31%	9%	23/(no TMs)415
CA1402	159	VRLPEIVARVATGVQSIENAKR/DGVPDVGVLN	369	2	19%	13%	49%	29%	31%	22%	19/(no TMs)271
CAGLOM04191g	44	ADLSYVKLDFDKYYGETFETAKR/GRSQADIRVN	51	1	62%	24%	7%	2%	31%	22%	19/(no TMs)(580-GPI)o601
YIR039C	51	FPVQKLADIINICTQDVSTVFKR/NEVLNTVIN	653	1	65%	21%	4%	1%	32%	22%	22/(no TMs)537
CA1402	125	IADMPFVVRAVDTAMTSVASTKR/DGANDDVANA	154	2	16%	9%	53%	25%	32%	19%	19/(no TMs)271
YLR120C	29	*IPAANKR/DDDSNSKFVK	317	1	53%	35%	16%	12%	32%	23%	22/(no TMs)(542-GPI)o569
YPL187W	103	LDKREAEAWHLQLKPGQPMYKR/EAEAEAWHL	663	2	12%	5%	56%	21%	32%	16%	20/(no TMs)165
CA1873	195	QLITRDETDIDALAPYVVTMEKR/LDLGVDVNL	387	1	60%	26%	9%	6%	32%	22%	21/(no TMs)351
CA1873	58	LTKREEQDIQELVQHINNYKTRR/DAIDEEIMKR	333	1	56%	30%	12%	7%	32%	24%	21/(no TMs)351
CAGLOG00308g	28	*HAAHTHKEKR/DVVTKTMHAV	526	1	59%	26%	9%	4%	32%	23%	18/(no TMs)374
CAGLOE00275g	398	IPMPSVTTSKIMSSRFEAPFGPR/IGTWDLEDKP	40		51%	16%	17%	7%	32%	13%	20/(no TMs)880
CA1873	39	*EQQQQESQFASKALTKR/EEQDIQELVQ	102	1	58%	27%	9%	5%	32%	22%	21/(no TMs)351
CA3122	41	*IFDLESKQGLQDEETVNNDKR/EPVNLLYLDR	156		54%	5%	13%	8%	33%	7%	19/(no TMs)(?GPI)o410
CAGLOM07205g	185	PVATAEVELMHETFGAVPALKR/EPNPSHTASS	265		41%	16%	26%	9%	33%	19%	o6-23i88-105o241
YGL028C	31	*ILPKREVVTR/VHTASTNVV	62	1	62%	23%	4%	2%	34%	24%	21/(no TMs)542
CA4336	154	YGSSNTDSYGSSNTDSYGSSNKR/GNDSYGSSND	138		57%	12%	9%	3%	34%	13%	(no TMs)212
CA4336	73	SNKRSNDSYGSSNTDSYGSSNRR/DNDSYGSSNN	88		53%	14%	13%	3%	34%	12%	(no TMs)212
YOR219C	162	SKLLPTKITRPKTSAGDSSLGKR/SFSIENVLNG	59		61%	6%	5%	3%	34%	5%	i120-139o931
CA2412	101	KREAKSKGGFRLTNFGYFEPGKR/DANADAGFRL	165	2	17%	14%	49%	28%	34%	21%	19/(no TMs)143

CA4336	27	GKDDDKDKNDRRDNDSYGSSNRR/NNDDSYGSSS	70		49%	18%	17%	3%	34%	16%	(no TMs)212
CAGLOJ02288g	46	EKRIVSSHASLSKR/SAVDLQFRR/FNNLYYESVL	195	1	60%	26%	6%	1%	34%	25%	(no TMs)(508-GPI)o541
YPL207W	155	EKAVNYPVTVDFTEVFRKPTKKR/SSTPKVFSKN	96		49%	14%	17%	5%	34%	17%	i7-39o810
YJL158C	63	GAAEYTTTTFGIAVQAITSSAKR/DVISQIGDGO	543	1	60%	25%	5%	1%	34%	25%	22/(no TMs)(209-GPI)o227
CA0365	62	LALNILNLIQLGIGANVNVTAKR/DESAVTIEQS	160	2	16%	10%	49%	28%	34%	24%	19/(no TMs)(145-GPI)o178
CA2994	322	ILGAATIFNTLLVINSTDSNNKR/DDSYPHINLYD	498		51%	12%	14%	3%	35%	14%	o367-390i671
CAGLOM01386g	7	*VDKKKKR/SFSETGIFKL	407		53%	17%	12%	2%	35%	18%	o1284-1303i1427
CA0365	74	IGANVNVTAKRDESAVTIEQSKR/LDADVAQLA	309	2	12%	6%	53%	24%	35%	20%	19/(no TMs)(145-GPI)o178
CAGLOJ08756g	718	STVVRFFFDRLVQLLKNLSRETR/GDNKKKNQHR	95		39%	9%	26%	11%	35%	17%	20/o752-772i833
CA4336	16	*VFGFGKDDDKDKNDRR/DNDSYGSSNR	99		49%	21%	16%	6%	35%	16%	(no TMs)212
YLR207W	269	ESGLLGKGLNSVPSSTVRKRTR/PDIGSPFIAQ	118		47%	3%	18%	5%	35%	4%	21/o768-789i833
CAGLOE00187g	179	SNANLDATGSLVVSRETTTKSRR/QDASEAITGK	277		55%	8%	10%	1%	35%	8%	o216-237i238
CA0856	211	KRGGYGSERRQSDCESDCERSER/CNYPFCELYE	54		36%	11%	28%	17%	35%	23%	o28-45i57-78o356
CA0822	37	*ISNPFKPNGNLKFKR/GGGHNVADYD	671	1	58%	29%	7%	3%	35%	26%	22/(no TMs)438
CA2412	80	DNKPVILIVNGTTLTSGANNEKR/EAKSKGGFRL	179	2	16%	16%	48%	29%	35%	21%	19/(no TMs)143
CA3122	161	KDAKNLIDLVALKQLEKEAAKR/DAKNI PDLEA	94		55%	12%	9%	3%	36%	15%	19/(no TMs)(?-GPI)o410
CA2974	180	NLIGNALGLNANTVSSKMASAKR/EIMVESAPEP	128	0	55%	31%	9%	8%	36%	29%	21/(no TMs)328
YLR337C	282	VSPPSSEVPAGGLPFLAEINARR/SERGAVEGVS	94		55%	12%	9%	2%	36%	12%	(noTMs)817
CA2974	54	LTVRESNLVNSALANLQHYNKR/DLMSQEEI IK	369	1	54%	33%	10%	9%	36%	27%	21/(no TMs)328
CAGLOH07359g	224	LMMGGGRTHFYPSGTNDPRYGRR/GLRQDDRNLI	79		45%	8%	19%	8%	36%	12%	i21-44o545
CA4336	200	SNRRGNDSYGSSNDDSYGSSNRR/GNSDSYGSSD	56		54%	13%	9%	3%	37%	15%	(no TMs)212
CA5537	22	*SYNNPNNSNSHLRPHAYNSRR/DDS DGESSI	235		46%	10%	18%	5%	37%	11%	o647-666i678-702o722-741i753-777o804-823i930-950o970-994i1009
CA3321	68	DVESSPFQIDYNNSSSPVEASRR/SHYNYFPNSL	77		38%	11%	24%	7%	37%	11%	18/(no TMs)535
CA0339	165	NQVKVKLPQAI SEFEQNSQNSKR/IGLDSIAQ*	295		52%	19%	11%	2%	37%	19%	19/(no TMs)174
YNL323W	51	EEDVDASEFEDEVKPVRTKNRR/PKEDAF TQQR	149		44%	12%	18%	7%	38%	16%	o76-95i373-398o414
CA0549	40	*SNPKVISFDFELRFS DPIP KR/DTNFGTG TAA	130		45%	10%	17%	5%	38%	13%	20/(no TMs)364
CAGLOD02442g	48	EEDVDASEFEDDDYVPVKTKNRR/PKEDNFTQQR	149		46%	10%	17%	5%	38%	15%	o73-92i374-395o417
CA0338	134	NEVKKVLPQAISSVQQNAGNTR/DTTTELIQAI	444		39%	14%	23%	8%	38%	18%	19/(no TMs)(?-GPI)o183
YFL042C	162	NATIAEIGSPLQVQEKPDVKT R/LTPSQMKEDN	41		57%	13%	5%	3%	38%	12%	o636-654i674
CA0883	45	*NKNIKREDCDKTSFHGHHKHKR/AVAYDYAVVT	678		55%	32%	7%	3%	38%	30%	23/(no TMs)418
YPL187W	124	KREAEAEAWHLQLKPGQPMYKR/EADAEAWHWL	663	2	12%	6%	50%	28%	38%	22%	20/(no TMs)165
CAGLOA04081g	222	SASKGKGGKDS DASSVSSHRR/PDEPITVLT Y	351		42%	7%	20%	6%	38%	10%	18/(no TMs)(352-GPI)o374
YMR305C	28	*AVRHKHEKR/DVVTATVHAQ	533	1	54%	33%	7%	3%	38%	30%	19/(no TMs)389
CA1402	30	*IHHAPEFNMKR/DVAPAAPAAP	410	2	14%	7%	47%	30%	39%	24%	19/(no TMs)271
CA0968	35	*GFVTLDNFV KR/SLVDPDDPTV	420	1	56%	31%	5%	1%	39%	31%	24/(no TMs)418
CA2499	35	*GFVTLDNFV KR/SLVDPDDPTV	420	1	56%	31%	5%	1%	39%	31%	24/(no TMs)418
CA1334	116	YRNENSEECSDEDDHKKKKKR/PHRHGKSD	203	1	55%	32%	6%	2%	39%	31%	19/(no TMs)201
CA2055	35	*VKRSTGFVTLDNFV KR/SLVDPK DPTV	420	1	56%	31%	5%	1%	39%	31%	19/(no TMs)417
CAGLOI02508g	115	EAVVTSTSATNGSLNSP SPTLKR/YAVSRSLWYG	65		55%	14%	6%	3%	39%	11%	i64-82o131-156i175
YPL187W	145	KREAEAEAWHLQLKPGQPMYKR/EADAEAWHWL	663	2	12%	7%	49%	29%	39%	23%	20/(no TMs)165

Anhang

CA2247	130	RNQHRHDKLMINREKINDQQSKR/DINDQYEKEY	309		51%	17%	9%	2%	39%	18%	(noTMs)200
CA1394	31	*IQKDSLGLKR/TLSSTFDPH	25	0	54%	33%	6%	2%	39%	32%	21/(no TMs)418
CA1872	159	TSTGSTHPPIPAASVLSSSSQRR/ATIEDFADTE	58		55%	5%	5%	2%	40%	6%	o895-919i943-968o972
CA1360	470	RYSNHHSEFTVERPPRPSRTKRR/PRIKAKKTMK	48		55%	12%	5%	3%	40%	9%	17/o712-733i734
CA3122	186	KNIPDLEALKTGIEEEEGQVAKR/DAKNVINLSN	211		44%	17%	16%	7%	40%	16%	19/(no TMs)(?-GPI)o410
YJL068C	47	TSMNVNIYLPKHYYAQDFPRNKR/IPTVFYLSGL	650		47%	8%	13%	3%	40%	11%	(no TMs)299
CA4336	54	SYGSSSFGSSNDDNSYSSSNKR/SNDSYSSSNT	138		53%	13%	7%	3%	40%	15%	(no TMs)212
CAGLOM08492g	62	TDYSSTFGIAVQPIATSSSVAKR/AVSQIGDGQV	338	1	56%	31%	4%	2%	40%	33%	20/(no TMs)335
CA0894	105	SANALTPTPSHHQPAHSITRQRR/SSSLIQHLEP	341		53%	16%	7%	2%	40%	16%	o142-162i219-235o241-259i268
YLR057W	188	QIKDIFMKSWNQEQLLKSNLRR/ESTWPIDLID	265		51%	22%	8%	6%	40%	16%	(no TMs)849
CA2418	154	IGQPKDAKLVPPPPPEELPNYKR/ANATILSLVR	392		46%	13%	13%	3%	41%	14%	i12-32o480
YDR326C	9	*RDEATRKKR/SFSDGHFFKK	565		47%	8%	12%	1%	41%	9%	o1281-1302i1438
CAGLOL10670g	207	STGSHPTGSSESESESQESDDTSKR/SLLKRSRLE	102		51%	11%	8%	4%	41%	10%	21/(no TMs)(241-GPI)o263
YJL034W	309	KAFKKKHGIDVSDNNKALAKLKR/EAEKARALS	954	1	54%	34%	6%	1%	41%	33%	i12-29o682
CAGLOJ08756g	267	DEGLLGKGLSTVQPQSVRRKKTTR/PFFAGSSNLK	68		41%	14%	18%	8%	41%	19%	20/o752-772i833
CA2412	122	KRDANADAGFRLTNFGYFEPGKR/DANAEGFRL	165	2	9%	4%	50%	27%	42%	24%	19/(no TMs)143
CAGLOI06182g	61	TDYASTFGIAVQPIATPASVAKR/AVSQIGDGQV	338	1	54%	33%	4%	2%	42%	35%	19/(no TMs)340
CA3122	116	PKNLFNLQALHEGLKDEETKSKR/EAKNLPNLEA	271		42%	7%	16%	10%	42%	14%	19/(no TMs)(?-GPI)o410
YGR279C	29	*ANHEHKDKR/AVVTTTVQKQ	183	1	53%	34%	4%	1%	42%	35%	20/(no TMs)386
YDR245W	109	DNKLSNTGSSINSKRYVPPYSKR/SRWSFWNQDP	142		40%	10%	18%	5%	42%	14%	i50-67o393
CA2830	73	SLSGSKNKREAEIANKDGTIEKR/TFGSAQVNAG	164		50%	10%	7%	3%	42%	8%	18/(no TMs)(684-GPI)o714
CA0365	149	IAANILNLIQLGIGADVNVSAKR/EDTVAAQIGA	160	2	10%	5%	47%	26%	43%	22%	19/(no TMs)(145-GPI)o178
CAGLOD02948g	295	KTFKKKHGIDVSDNEKALAKLKR/EAEKARALS	954		52%	32%	6%	1%	43%	32%	29/(no TMs)667
CA2730	399	SKEISFQNVVMTKDEQRNHEKR/VLNGGEKIED	426		33%	18%	24%	11%	43%	19%	24/(no TMs)434
CA0915	308	KQFKKKHNDITANAKAISLKR/EAEKARRTLS	954		50%	30%	7%	1%	43%	30%	35/(no TMs)687
YMR200W	226	KRSLRSLVRRSLENRHKTNAIKR/QNTSFLTSNA	130		49%	8%	7%	2%	43%	6%	25/(no TMs)(235-GPI)o256
CA3122	68	LLYLDRFKMGVSDAKGNAKFKR/DPKNVIDPAS	420		36%	19%	20%	10%	43%	20%	19/(no TMs)(?-GPI)o410
YNL238W	79	EHDVRLPNHYVFSKELLKLGKR/SSLEELQGDN	59		36%	6%	20%	9%	44%	14%	24/o676-699i814
CA1387	266	RRDAAQKKISATYMKIEIKSRER/GDIDPNRDLI	64		31%	8%	25%	10%	44%	15%	o13-36i48-65o528
CA2303	123	GNSKRQIDDQDCNVKHVHHKHKR/ATEVVQVTQT	848		49%	14%	7%	2%	44%	14%	22/(no TMs)468
YIL136W	90	SLKDALLGVSQKAREEAPKVTKR/VISPEEDAQT	406		50%	16%	6%	4%	44%	12%	19/(no TMs)393
YOR238W	253	KNALSLSFDWYATKDRLLTKKR/SRNPFNRTAP	48		33%	6%	23%	6%	44%	9%	(no TMs)303
CA3614	121	DNDECEDEDCDEDEKKGHKYKR/GEVEEPCETS	410		48%	10%	8%	1%	44%	9%	19/(no TMs)251
YHR188C	276	HYEMDARIEMVTHVPSALARGER/PIPKPLDGNT	59		32%	8%	24%	11%	44%	15%	20/o548-568i610
CA1820	607	KFNNHFAAAIGQLTSNQYNKDRR/TPNLSKEICS	93		33%	3%	23%	10%	44%	13%	i12-31o765
CA1872	116	ERTRTKKSSSIGSAELSPRSRPR/MMNFSLRPKF	60		48%	19%	8%	2%	44%	17%	o895-919i943-968o972
CAGLOB04477g	228	SDVSDISYEMEPFPPAQSPSRR/QHLSTNYNSS	61		45%	20%	11%	3%	44%	19%	i41-62o283
CA2633	526	YSDVIQVNQTITTTPEQKAPSKR/DSPQKI IHSS	151		32%	10%	23%	9%	45%	10%	21/o774-794i936
CAGLOM08734g	34	LVAANKIVEMITFRDENGKLHRR/LAPEEYRGGV	43		49%	13%	6%	2%	45%	12%	(no TMs)253
YPR079W	44	HVYNGPGLSHEANEHRASHKQKR/TLANPKPKS	482		50%	12%	5%	3%	45%	10%	i7-25o237-257i381
YCR068W	139	TNPFDFEFELRRMPLLMKRMKER/DPEFIESYIY	42		31%	8%	24%	10%	45%	16%	i12-34o520

Anhang

CA6162	200	VRETKDLADQSPEDSTSIDVFKR/ENSLLVLLFT	653	45%	11%	10%	1%	45%	12%	21/(no TMs)296
CAGLOM13453g	319	GICKFEMEAIIVDMVMTQPQLRR/GTNPIPRPLS	74	34%	6%	21%	9%	45%	14%	22/o582-602i646
YOL003C	248	RIYPESRSFQNKKAEEHLTKKR/PRFDELVNFP	48	31%	7%	23%	13%	46%	19%	o6-29i41-58o126-143i164-186o378
CA1872	337	TQIAQQYYGDARKHIEVVDPTKR/EKNSVIATT	91	31%	6%	24%	10%	46%	11%	o895-919i943-968o972
CAGLOL13266g	114	VAKLQLRGSISLAPNDPFFHHTR/DKEFSTEELG	70	29%	8%	25%	8%	46%	10%	20/(no TMs)579
CAGLOJ06820g	104	DSFTIHGLWPDDCMGGYPQFCKR/DLFIDDVDYL	502	37%	9%	17%	7%	46%	15%	(no TMs)433
CAGLOL11528g	53	STRLSEGILEYYDMFDGAVVPR/EDFNTVCKTL	75	28%	1%	26%	15%	46%	16%	19/o275-29i319
CAGLOJ02068g	128	IRAYPTIRLYGPDGFLEEYHGKR/TKEEFKLFAR	301	25%	3%	29%	18%	46%	15%	19/o662-678i708
CA3122	92	PKNVIDPASLKEGSAEEEQDKR/EPKNLNLQA	114	37%	17%	16%	6%	46%	18%	19/(no TMs)(?-GPI)o410
YFL042C	137	SQPSIISHRREPSMGSKRRSSRR/ISNATIAEIG	52	48%	19%	6%	3%	46%	17%	o636-654i674
CAGLOK06237g	80	QISLGLTPRKPVVITDLRTGNKR/KLHGRFLHIT	0	41%	16%	13%	3%	46%	17%	i12-29o663
CAGLOE00385g	591	RVIVIEDGRVVEGSGFDQLRSTR/DSELNKLKE	122	31%	12%	23%	12%	46%	12%	i47-73o93-115i191-210o269-293i305-326o608
YJL012C	399	TELTMVREDNFDGVDRTHKNWRR/TDIGVDWPFK	259	40%	9%	13%	3%	46%	12%	(no TMs)648
CA3295	551	APSSEAPSSSSSEALSSSSTTKR/PTAAAKGFFG	123	41%	9%	12%	2%	47%	11%	23/(no TMs)(?-GPI)o1249
CAGLOH04499g	193	PFQRARFRRADPREAYRQQQRR/QGNQENELVN	140	41%	12%	12%	3%	47%	11%	o207-226i232
CA4624	17	*SRNVSLPSTTESTKNKR/VVSSQLRLAQ	468	47%	10%	6%	3%	47%	7%	o723-74i757
CA2394	20	*NDNVVILTLRAANHITNAKR/ADSTTSTEAM	369	35%	5%	18%	6%	47%	10%	o66-88i288
YEL060C	188	TAQFNPDAISKIIPNRYIIIVFKR/GAPQEEIDFH	522	44%	23%	9%	4%	47%	19%	19/(no TMs)635
CAGLOB03091g	83	ANTSSSKISSYEKLESENNGRSER/LYPVYVDDSK	54	46%	16%	6%	3%	47%	13%	o605-626i670-694o696
YML012W	77	VGDGQKLNLFVRDSVGNERYRKR/DFAGDVRVAF	707	43%	5%	10%	1%	47%	5%	21/o182-20i211
YBR161W	348	SSYKNKEKRRNPTRHEYNRSRGR/LRKDSNIPYD	165	46%	19%	6%	2%	48%	18%	i7-24o200-220i274-299o376
CAGLOL13134g	125	FANGYHSSDLFADSKYHYASDKR/DEAFHAIFRS	68	26%	4%	26%	14%	48%	14%	o622-644i725
YOL003C	250	YPESRSFQNKKAEEHLTKKR/DFDELVNFPYD	64	31%	6%	22%	11%	48%	16%	o6-29i41-58o126-143i164-186o378
YLR337C	285	PSSEVPAGGLPFLAEINARRSER/GAVEGVSTK	54	39%	18%	13%	6%	48%	16%	(noTMs)817
CA1835	393	DNEEEESITAIPIKSRKGRSTR/LDISTQLPPL	79	46%	12%	6%	4%	48%	8%	19/(no TMs)(481-GPI)o504
CA2730	43	*CYERKKIHPQPKETPAKR/DHSELIREQ	151	40%	16%	12%	3%	48%	19%	24/(no TMs)434
YHR080C	50	RNFLNNTTGENVSTTASAERFR/VGGNPDIPSL	476	35%	11%	16%	7%	48%	15%	o1204-1223i1345
CA0340	146	NLPQAITTVQQQQGQSQVVKR/DTIGDIVNEI	772	37%	13%	15%	6%	48%	17%	22/(no TMs)195
YMR006C	61	DDTSLVRNASGLSTAETDWLKRR/DAYTKEALHS	64	40%	12%	11%	1%	48%	13%	20/(no TMs)(680-GPI)o706
CAGLOH09240g	505	DYVIEPEDAATIGKLEYSEKDRR/YSICAIQMAH	93	42%	20%	10%	4%	48%	23%	33/(no TMs)642
YOR384W	26	*ASTKKR/TQWDQIAIDA	96	29%	9%	23%	10%	49%	7%	20/o162-184i205-225o237-254i266-288o350-368i375-392o398-417i694
YER005W	265	LAQLINTLPENTNDYENDDFSTR/NLNDPCMPRG	47	31%	4%	20%	10%	49%	13%	o501-517i630
CAGLOJ11770g	59	NDINLLRNATGLSQSEIDWLKRR/DVNTREALE	64	38%	8%	13%	3%	49%	8%	23/(no TMs)(633-GPI)o659
CAGLOL13266g	45	DINTLEQTSESQEYDNFFKQSKR/PGNVKAPLVA	247	35%	7%	17%	12%	49%	12%	20/(no TMs)579
CAGLOM01298g	128	VTVENFIRLLTDRWTEQPKSKR/LQTDENSNIF	434	42%	9%	9%	3%	49%	9%	22/o359-377i390
YLR207W	266	SDFESGLLGKGLNSVPSSTVRKR/TRRPDIGSPF	158	44%	13%	7%	2%	49%	13%	21/o768-789i833
CA2912	119	GNEECDEDCDEHRKKGSKYKR/GEVENPRETR	410	37%	18%	14%	7%	49%	13%	19/(no TMs)247
CAGLOB00616g	399	EYNQAKPLMNSLIESENSNKKR/LSLLTNFLLM	277	38%	10%	13%	5%	49%	11%	17/(no TMs)(394-GPI)o416
CA1334	135	KKKRPHRHGGKSDDDDDKKWKR/GGDYSDDNDN	746	43%	9%	8%	3%	49%	7%	19/(no TMs)201
YJR151C	1077	VSCSAGACTTLTTVDAGISMFTR/TGLSITQTTV	64	45%	12%	5%	3%	49%	15%	20/(no TMs)(1127-GPI)o1161

Anhang

CA0951	187	IFKIFYQAKPSVSQLNTYPSKRR/IYHARFDSL	96	43%	13%	8%	3%	49%	13%	i18-38o660
YGL126W	220	NTFIHDQMHRWQGPLTTSAQCRR/FGGHWAGGHD	140	38%	9%	12%	3%	49%	10%	i12-31o51-66i87-105o236-254i326-349o355-371i380
CA2405	688	TTYCPLTTTKAVYRVAKLANYKR/DQTEFIATET	490	34%	9%	17%	5%	50%	11%	19/(no TMs)(721-GPI)o753
CAGLOL08448g	159	MIMSLLNIFSNAGFGTKFIRRRR/NNAEVGVPSV	288	40%	5%	10%	1%	50%	6%	o12-28i40-58o64-93i123-146o173
YJL132W	307	NTNEAQAEELRLCANLPPVSQKR/INARPLVSSL	114	38%	8%	13%	3%	50%	10%	26/(no TMs)750
CA2830	59	PVCTDLGLYHDSISLSGSKNKR/EAEIANKDGT	731	43%	13%	7%	3%	50%	10%	18/(no TMs)(684-GPI)o714
CA3629	272	KPTTEPHVYFGTEAFDHDGIEKR/GQIVKIEDHT	164	38%	11%	11%	2%	51%	12%	(no TMs)415
CA1039	278	TGGSPTTYFISSTKQVLVQKHKR/SLDYVAGVAK	848	28%	3%	21%	12%	51%	10%	24/o170-190i202-220o327-351i372-394o406-426i466-485o491-510i522-541o553-579i808
CA2714	872	IVSQSGSSFTTITTFEPPVVCRR/DDVNCGSSIG	585	41%	6%	8%	2%	51%	5%	21/(no TMs)(1225-GPI)o1249
CAGLOG02959g	86	FFLSKKPSPIQVVSNSAKNKR/YVTPPTLYKL	277	41%	9%	8%	2%	51%	8%	i27-51o114-134i146-173o179-199i239-258o918
YMR306W	482	FFAVRPLGGLFRPYLNKDKKRR/YISSQFTTAS	346	38%	6%	11%	5%	51%	9%	o338-357i377-395o415-437i444-464o503-526i547-567o579-604i1215-1234o1264-1288i1358-1378o1384-1402i1473-1495o1515-1539i1551-1572o1584-1606i1649-1670o1676-1692i1713-1733o1785
YHR080C	147	NSNSVPSLSVQALEKEKLQSGKR/EGSSNQAEK	140	35%	12%	14%	4%	51%	16%	o1204-1223i1345
CAGLOB04235g	242	TEQALVNVNQLGYELCPECEKQRR/EEMETEHLWQ	384	37%	14%	12%	3%	51%	16%	22/o268-289i313
CAGLOJ06820g	297	PLRMSRCKDQGIKYYPKGYQRR/DNRGPNKKPI	100	35%	6%	13%	7%	51%	12%	(no TMs)433
CA5996	223	NWPIFEFEKGFTEDDELWPNKR/ETGQHRKYA	163	21%	9%	28%	4%	51%	11%	27/(no TMs)327
YGL156W	547	GNVIPKLQVGITVDEFYDDILKR/TNQGHDLPW	237	32%	6%	17%	8%	51%	14%	(no TMs)1083
CAGLOK11880g	8	*SYVDLYKR/GGNEAVKINP	72	35%	6%	14%	5%	51%	5%	o34-55i62-81o116-134i141-161o167-192i204-225o237-262i316
CA1379	38	*VENNQDLEATFTRIKR/SLVHQDIFEI	518	34%	17%	14%	6%	52%	20%	22/o334-355i784-804o816-839i860-884o896-921i928-946o1012-1035i1038
CA1345	39	*QENEITPADNIDKR/AGAIGNFFRD	71	28%	4%	20%	7%	52%	6%	25/o1297-1319i1409
CAGLOG05896g	28	*QIQKR/DDDNVRYITS	366	29%	2%	19%	9%	52%	10%	23/(no TMs)294
CA0371	554	GKGKNIHEIFGQDLVIPTKSRR/GPGGYKVKTS	221	30%	16%	19%	8%	52%	8%	21/(no TMs)806
CAGLOG02717g	34	*KFLRSDAVSWHDRVPR/GNFTEWVHAQ	40	27%	7%	21%	10%	52%	4%	17/(no TMs)515
CA1872	358	KREKSNSVIATTVSGDSPTALKR/LSGGSNSSGS	212	42%	9%	6%	3%	52%	11%	o895-919i943-968o972
CAGLOI06204g	61	TDYASTFGIAVQPIATPASVAKR/AVSQIGDGQV	338	42%	23%	6%	3%	52%	26%	19/(no TMs)349
YJR126C	363	TTPGYPSKSDTEHMSPVGRVPR/IPSPVMSARD	86	17%	6%	31%	9%	52%	9%	i90-109o811
CA3105	279	SQTFDIIVVTFPQIQELSHLNKR/NVELSVKSFQ	73	41%	10%	7%	3%	52%	8%	23/o351-371i393
CA5699	1405	ESSIGTILSESSDSIPTTFSTR/YWSPSGMSSR	47	42%	4%	6%	3%	52%	7%	19/(no TMs)(1980-GPI)o2000
CA5115	242	RNGNNSNGNASNSTLRGGNIKR/SSIKQLGLKF	254	31%	7%	17%	6%	52%	10%	18/o310-331i343-362o374-393i414-440o465
YMR306W	124	IGFQNMRLHGGKARQNRVVKMAKKR/GVSIKEQIKQ	141	39%	10%	9%	4%	52%	12%	o338-357i377-395o415-437i444-464o503-526i547-567o579-604i1215-1234o1264-1288i1358-1378o1384-1402i1473-1495o1515-1539i1551-1572o1584-1606i1649-1670o1676-1692i1713-1733o1785
CA0397	27	*PRKYKR/YGQLKSFYAC	577	32%	2%	16%	5%	52%	6%	21/o158-181i318-340o346-368i375-393o399-417i710
CAGLOK12914g	31	*INKSGDETLR/DISLLDLVKA	92	42%	17%	5%	3%	52%	15%	19/o397-416i428
CA5614	56	VTVSTYFVNSYTCNQPHNHSTR/PSHYLPINGT	57	32%	3%	15%	6%	53%	9%	i20-42o270-293i337



Anhang

CAGL0L03333g	344	NEYHWSNEREIQDSTTNDNLPR/SYEYAQALKN	52		30%	8%	17%	7%	53%	13%	43/o410-429i450
YIL123W	34	*HVDVHQEDAHQHKKR/AVAYKYVYET	386		38%	7%	9%	2%	53%	7%	20/(no TMs)475
CA0932	104	FFYAFLFMYQTFDGDGCHARR/TGQSGPLGEL	224		31%	4%	16%	6%	53%	8%	i50-69o138-158i179-200o220-240i274-292o298-322i334-353o359-379i402
CA1840	35	*VPNVEQLKAQSLQKR/DEDLMEPVFK	118		36%	10%	11%	2%	53%	13%	20/(no TMs)712
YNR019W	26	KDLLENEQFLRIQKLNADAGKR/QSITVDEGE	165		41%	6%	6%	3%	53%	4%	o218-239i259-281o293-312i324-343o404-422i434-459o479-499i567-586o592-610i622-640o642
CA3529	213	KQFQSSISQEHETLQRDLSSHKKR/EFEEESDNKS	107 5		35%	15%	13%	6%	53%	20%	25/(no TMs)837
CAGL0M01386g	562	KRNSRITTLNDEIVGDNDRKPR/SLSRKFLNRR	57		34%	15%	13%	2%	53%	16%	o1284-1303i1427
CA5699	1915	KSVQAESVQTDSTTYSEMMSKKR/NKNSGFGTSS	51		39%	6%	8%	3%	53%	4%	19/(no TMs)(1980-GPI)o2000
CAGL0M05599g	120	SKRAVAEEDCDSVVAKPVHKKR/DVAVEVIEVT	848		40%	5%	7%	2%	53%	3%	18/(no TMs)446
YGR176W	95	HSRLWSKRLGRTQISRRLQTARR/LNPERSTAWS	52		31%	3%	15%	4%	53%	5%	26/(no TMs)115
CAGL0I10296g	33	VYIKADRKEFMDPIGKSDLWKR/DISRNYDKL	93		27%	1%	20%	11%	53%	10%	(no TMs)446
YPL170W	46	GASNTNDSNKGSEPVVAGNFFPR/TLKSFNGHDD	56		25%	4%	22%	10%	53%	8%	(no TMs)152
CAGL0H09130g	243	DRKKYSQASWYKMERKKADFHRR/NKISDYCEPF	144		30%	6%	16%	9%	54%	7%	i7-25o1015
CA5819	1208	SPHTSESDITSQYNHSTSVATR/SDSISLTDTF	55		41%	4%	6%	3%	54%	6%	21/(no TMs)(1502-GPI)o1526
CA2007	143	VTVENFMRLLTDKWSDQPRSKR/LLSDENSNIF	603		38%	7%	9%	2%	54%	6%	21/(no TMs)(?-GPI)o383
CAGL0J07546g	89	YVFSVAKSENEDDIDSGNSLRKR/AIENEDHILS	64		33%	8%	13%	8%	54%	8%	24/o671-693i789
CA3867	73	QQNNLDSNESFVDPLANPEHCKR/DIPYLEAVDT	610		28%	5%	18%	9%	54%	7%	23/(no TMs)(510-GPI)o544
YIL077C	190	LDSGSPSRWSMYFYITYQNPERR/LPDPKVKLQQ	74		24%	8%	22%	15%	54%	18%	o67-86i98-119o320
CAGL0M05599g	100	HFRGPLKLSQFAVYYPASSNSKR/AVAEEDCDV	295		39%	11%	7%	3%	54%	10%	18/(no TMs)446
YDR262W	44	*VVNLITFKDSNGKLHKKR/LAPEEIPPRL	85		23%	4%	22%	11%	54%	7%	27/(no TMs)(256-GPI)o272
CAGL0C03333g	24	*FQKR/DVWEVNAVMS	502		38%	6%	8%	2%	54%	5%	20/o186-206i257-276o296-316i337-358o370-393i400-420o426-445i727
YDR245W	101	SFQQQNSNDNKLNTGSSINSKR/YVPPYSKRSR	295		37%	8%	9%	1%	54%	8%	i50-67o393
CA3122	225	KREGKNLFDLTKFQSGQPIKKR/DQKILKQEK	99		37%	11%	9%	2%	54%	11%	19/(no TMs)(?-GPI)o410
CAGL0H07403g	139	VKDMTKKDIEEYLAPSFANKKR/PKAAFVTLIR	237		20%	4%	25%	5%	55%	4%	29/(no TMs)463
YFL042C	124	SPVSRSTTPTSPVSPSIIISHRR/EPSMGSKRRS	102		37%	6%	8%	4%	55%	2%	o636-654i674
YHR009C	303	LKCEMDCVHSLKYLPSVVKNRR/NSRGHAENPD	119		29%	8%	17%	7%	55%	11%	i24-41o523
CA2912	84	DDGQLERNDKKLDNCCKSERVSR/PAPSPSAIAV	34		27%	8%	18%	7%	55%	11%	19/(no TMs)247
CAGL0L04466g	323	TEDIANEAYAKFTEEYETEKR/INDKTELVS	124		40%	5%	5%	3%	55%	3%	o678-694i701-718o783
CA5572	59	TSSINHKNQPLSLRNNPQLIQKR/SRHAESPLFI	94		37%	4%	8%	1%	55%	4%	21/(no TMs)924
CAGL0H04499g	77	PKAHEAFKLINRAFEVLSDSQKR/QIYDQIGRDP	114		33%	10%	12%	6%	55%	4%	o207-226i232
CAGL0K02937g	48	WKKYVEQKDAITQQVSVKMLKR/EYDDNLFREQ	702		37%	7%	8%	2%	55%	7%	24/o79-97i285-305o438
YOR041C	110	RSTSVVFEISSLIDSGTLLFSRR/EESLIPFS	288		15%	3%	30%	9%	55%	9%	(no TMs)143
YER076C	74	MVEFYQTIGLTANGTVPEFNKR/DATEYPNIIS	488		26%	8%	19%	5%	55%	7%	24/(no TMs)302
YER001W	596	ALQGVYSGAIGDYVKKTDLNGKR/QEAAVEICSG	404		25%	5%	20%	8%	55%	6%	i12-36o762
CA3295	354	SVNKVIARRQQAIAEKKQANAKR/ESISFVSFEA	369		39%	14%	6%	3%	55%	11%	23/(no TMs)(?-GPI)o1249
CA2633	128	DYDELISHVESIHLPMKLSKR/IPVPIEMEDV	54		36%	7%	9%	6%	55%	11%	21/o774-794i936
CA2303	105	FRGPLTLLQFGVYYPSSSGNSKR/QIDDDQCNVK	369		36%	17%	9%	3%	55%	14%	22/(no TMs)468

CA2085	407	DSIPERWIYNHQHHHQGNKKVKKR/NKKKFKAEADS	154		32%	3%	13%	3%	55%	6%	25/(no TMs)842
CAGLOB03091g	257	QSHDKIMDNNNDTNLMEKIKSRR/EENSANFERE	346		33%	4%	12%	6%	55%	6%	o605-626i670-694o696
CA3321	135	YSGGTGVEVSAVIEAPNVEVAKR/NWLKLTPTLS	338		27%	2%	17%	8%	56%	7%	18/(no TMs)535
CAGLOL03289g	29	*AVQHQDKHAHEKR/DLVVVTEYVD	533		38%	10%	7%	3%	56%	7%	16/(no TMs)366
CAGLOJ11748g	201	YDKDFSKTTLARWTAIGSQVQGKR/DAGFNVTITD	338		25%	8%	19%	8%	56%	14%	20/(no TMs)(667-GPI)o695
YMR101C	222	SRNDMLHTIRDSDVEDHLENKSPR/INIRKFTNKM	43		35%	7%	9%	2%	56%	6%	i32-51o274-292i343
YKL163W	66	SFGIAIEAVATSASSVASSAKR/AASQIGDGQV	434		39%	12%	5%	3%	56%	16%	19/(no TMs)325
CAGLOF09053g	781	STLFKEVVSFEELYEDPEILHKKR/DILRAFVGNK	106		24%	4%	20%	9%	56%	8%	17/(no TMs)809
CAGLOH00330g	2010	THQFMELASHKTNRMFNDIINKR/QQKDNVNEEE	91		34%	13%	10%	3%	56%	15%	i35-59o2635
CAGLOK07359g	161	TAKNGAKAKIVVFEVGGGNVKKR/NIAIEEKKV	420		37%	9%	6%	3%	56%	6%	(no TMs)190
CA0894	61	TPTPIAEKLTSTPIPTPSPSHRR/GSSIGSSTSP	82		38%	22%	5%	3%	56%	21%	o142-162i219-235o241-259i268
CAGLOH01705g	68	MVSVHYTGSLAENGKVFDSLLRR/NEPIQFKLGA	92		26%	4%	18%	7%	56%	4%	23/(no TMs)136
CA4418	510	SEEIKVFATIFPKDDLILEWKRR/DPLKSL* . . .	226		31%	12%	12%	10%	57%	7%	(no TMs)517
YLR084C	1052	NASSISSVSAIYTDIESNTLNKR/GINNVNNGSI	59		32%	3%	11%	2%	57%	2%	21/o1161-1188i1220
CA5423	181	YAGKEQLRINDKQFLNIEHRTR/DENDNMLPQ	115		23%	8%	21%	10%	57%	8%	17/(no TMs)871
YFL051C	99	LSIYSPPESTPTCVTYAVLKR/DEGDYPCGP	927		28%	12%	15%	4%	57%	11%	24/(no TMs)160
YCL048W	432	QDGTIIEKGYKDLTQEGEDSKKR/VISKYANSAN	96		35%	6%	8%	5%	57%	7%	26/(no TMs)(440-GPI)o463
CAGLOH10076g	9	*STFVDLYKR/GGNEAVKINP	72		28%	4%	15%	6%	57%	4%	o35-55i62-82o117-135i142-162o168-193i205-227o239-258i325
CAGLOG07887g	153	ETKSQTLPLLALELTSNKKVRR/DRIIDEIEEI	197		31%	6%	12%	4%	57%	10%	20/o199-221i242
CA1360	254	LKSKSKSKSKLKHRSRSHKRR/PKLLKSKDVE	105		37%	8%	6%	3%	57%	5%	17/o712-733i734
YGL075C	268	FRRRLQRQTDITISFLKDQIRRR/GLNCSNDKVS	51		28%	3%	15%	3%	57%	3%	o305-327i387
CA3160	15	*NSDSISQVFEYLNKR/DECPTDNDYN	91		28%	1%	15%	5%	57%	6%	o33-52i64-85o105-127i212-235o275-295i307-326o346-366i370
CAGLOJ07546g	112	AIENEDHILSSEFLTSPSNKLQKR/MPVPAPPLDS	60		33%	6%	9%	5%	58%	3%	24/o671-693i789
YDR420W	210	EEGTLAQTSSISETRIAQMVTR/VSQISSITAA	59		36%	15%	6%	4%	58%	19%	22/o1486-1508i1802
CA5814	104	IAIGDFKGFSGRFSKDIIDRLKR/CPLVQEITED	848		25%	4%	18%	6%	58%	5%	i21-40o421
CAGLOB03091g	414	SGSSSRRTTSNNSQQSLIRQRR/NGDYVAIPVL	426		33%	5%	10%	5%	58%	9%	o605-626i670-694o696
CAGLOK06237g	454	SDLIEKKKKKGGKKKGGKKEKR/TLEEIYDQHS	526		28%	2%	14%	5%	58%	4%	i12-29o663
CAGLOH06919g	428	NTFDKLSVESVEYINNFEKLYKRR/EGGDSYTNFF	169		27%	2%	15%	4%	58%	3%	19/(no TMs)627
CA1583	230	KIKRSRMRIKRSLENQYRTNAKR/EIYTEKFDKI	369		34%	6%	8%	2%	58%	6%	25/(no TMs)(244-GPI)o260
CA5418	69	TSRSKPKETGTHRYGKFNKTPR/PVTTTLVKE	52		33%	10%	9%	2%	58%	8%	19/(no TMs)(105-GPI)o131
YIL037C	400	NLSTLYPIPYQLTKLINQKIFKR/EPEKIDDKV	408		27%	1%	15%	6%	58%	5%	o20-43i318-336o421-440i631-654o656
YBR162C	134	KQSIESCKEGEAVVSRHKHQHQR/DVAVEYVQVT	483		35%	15%	7%	1%	58%	15%	24/(no TMs)455
YDL221W	106	PASSLSSSFFVTVPFCAARSRR/GLTVALIAQS	307		31%	13%	11%	2%	58%	11%	(no TMs)183
YHR009C	142	VNIENYERLSKKAYNLNVPPPKR/PGYISNKFNI	113		26%	8%	16%	8%	58%	15%	i24-41o523
CAGLOH07997g	244	DYNYAPPAPFSDNGGWYDPKRR/QSFTTRKLM	113		30%	12%	12%	2%	58%	10%	15/(no TMs)265
CA3713	191	SSSSSAAPAAAAAPATNGASKR/EIMEAAEYLS	188		26%	6%	16%	9%	58%	12%	21/(no TMs)403
CA3713	33	*ANIVSEQTMVKR/EDVNAIVELI	710		27%	6%	14%	6%	58%	5%	21/(no TMs)403
YLR396C	624	ESIWVPGTITTKINASIKSNRR/SIDGSNGTFH	203		29%	4%	13%	7%	58%	11%	o640-657i691
CAGLOH09262g	41	*QRRSTR/YEEDLLFSSG	122		33%	4%	8%	2%	58%	4%	35/(no TMs)625

CAGL0M05841g	44	DIDRNYFERSVAVKTKYDGVRRR/EYFVPLKLAT	202		35%	15%	7%	2%	58%	16%	(no TMs)411
CA1371	141	GLIGLSKITHLIINGINGVQKR/GINRDSFINK	375		18%	4%	24%	10%	58%	8%	i103-125o375
YER001W	519	LNECDNTLTTEEYIFKAFHHRR/QHLEAGLFA	351		28%	6%	14%	3%	58%	7%	i12-36o762
CA1676	31	HPTLKAIVQLFGFRQLPHGKR/DVNDKELVRC	165		26%	5%	15%	7%	59%	3%	o430-447i448
CAGL0M08734g	211	LLQSKQYSMQHSNGKGGDVLKR/EGEKYCVASV	119		33%	9%	8%	3%	59%	7%	(no TMs)253
YKR102W	131	VPVKGKVLCSDNNTLSSKTEKR/ENDDCDQGA	155		23%	5%	18%	7%	59%	4%	25/(no TMs)1169
CA1313	155	NDKSDATFKSLSSSAPTUVHSRR/SWIKDFLIFK	224		22%	6%	20%	7%	59%	4%	22/(no TMs)332
YHR204W	69	DNYLEHGYPYDEVKPI SCVPKRR/NFEDPTDQGT	141		21%	9%	21%	12%	59%	12%	21/(no TMs)796
CAGL0G04279g	578	DIESEVGIRKRSLPSTDVNVVRR/DEQQFASKEP	383		26%	6%	15%	6%	59%	3%	i27-44o660
CA5699	1537	ESSSGTILSEENSISPTTFSTR/YVSVSLTVGE	47		36%	2%	5%	3%	59%	4%	19/(no TMs)(1980-GPI)o2000
CAGL0K06963g	815	PLDTIPIFLEGGHILFQKEKYRR/SSELSNDPY	368		26%	6%	16%	6%	59%	9%	21/(no TMs)941
YMR101C	84	VGPVPEHVSFIMDGNRRYAKSRR/LPVKKGHEAG	221		26%	3%	15%	4%	59%	3%	i32-51o274-292i343
CA1939	94	EYGHLLSVHQTYFHNLSINSKR/AGVKFEQDE	148		29%	2%	12%	2%	59%	4%	16/(no TMs)(?-GPI)o358
CA0825	299	LRSFVDTKYIGTRLDDGAPVSKR/CVNGFDNAGF	338		23%	5%	18%	8%	59%	7%	18/(no TMs)609
YLR337C	485	SSSSAVTPGGPLPLAEIQKRR/DRFVVGDDT	181		32%	13%	9%	5%	59%	17%	(noTMs)817
CA4418	445	VYDEPDSFIPERWENASGDMYKR/NWLVFGTGPH	424		31%	7%	10%	5%	59%	4%	(no TMs)517
YNL066W	44	. /AVAVTYVYET	386		35%	11%	6%	3%	59%	8%	23/(no TMs)420
CAGL0H10032g	23	ARVAII IYSMYGHTAALAEAEKR/GVESAGGKAD	183		25%	2%	16%	6%	59%	7%	(no TMs)198
YMR200W	205	PTETLNPTSSATSTDDPSSNKRR/SLRSLVRRSL	138		34%	4%	7%	4%	59%	6%	25/(no TMs)(235-GPI)o256
CA1387	264	YWRRDAAQKKISATYMEIKSRR/ERGDIDPNRD	138		27%	2%	14%	3%	59%	1%	o13-36i48-65o528
YHR080C	444	RLSFNSSNGLTNNDEPEYEDREPR/EMSCKFLNRR	73		29%	6%	12%	4%	59%	3%	o1204-1223i1345
CAGL0I00506g	379	QLVETMKWMSARSRLGDYVGKR/FPDNIQEVLN	462		22%	6%	19%	5%	60%	7%	22/(no TMs)621
YJR126C	327	VTEENGFLHYPGPARNPSYIRR/DSVNYFSOTP	155		24%	4%	16%	7%	60%	6%	i90-109o811
CAGL0I04532g	320	FWRRTMPAYYEIANLPYNQNR/VRNAFDDITP	133		19%	2%	22%	4%	60%	5%	i12-30o612
YKL128C	107	SGDEHGEWLDKLTPLGKDQVRR/TGSNVLLPMA	179		26%	1%	15%	5%	60%	4%	(no TMs)295
YNL238W	503	TEETLESVITISEKSLQDANFKR/IEHVTVTVDI	571		27%	7%	14%	10%	60%	16%	24/o676-699i814
CA1849	211	AYFAWGHLKSRKKRYTEIMNKR/NVRQVATLES	172		30%	6%	10%	1%	60%	6%	o15-39i51-70o77-96i108-128o148-168i174-194o237
CAGL0H09240g	93	INTRIMKLKEEKDDPKKAGKTR/TESEEDKILL	66		33%	21%	7%	3%	60%	23%	33/(no TMs)642
YDR230W	34	PASTFCFPLVLFTRTCTISSQRR/PEINSYGFVQ	72		22%	10%	19%	10%	60%	1%	(no TMs)115
YGL125W	336	PWRKSLNPKRKNNEVRPIFWKRR/PYSYVARTSQ	92		28%	2%	12%	2%	60%	4%	(no TMs)600
CAGL0H00330g	1131	LHGRAGATSHTDISVEELIWKRR/LKINKKNTPD	90		35%	12%	5%	3%	60%	14%	i35-59o2635
CA1678	132	QLEANSTAAIRRRMLGETPIVKR/DQIDDFIASI	300		15%	7%	25%	11%	60%	5%	22/(no TMs)(499-GPI)o533
YMR008C	59	DDINLVREASGLSDNETEWLKKR/DAYTKEALHS	64		31%	14%	9%	1%	60%	15%	23/(no TMs)(637-GPI)o664
CAGL0M12320g	284	VVVSNDILSISVQRRANLLKR/GISVASSSDS	76		28%	4%	12%	3%	60%	2%	20/o167-19i203-220o341-367i396-416o428-448i486-506o512-531i543-564o576-602i769
CAGL0L02365g	92	DAGEKKYDHLIVFPATKSLNKR/LSMNHLLNFY	59		29%	7%	11%	2%	60%	6%	20/o391-413i428
CA5232	30	*LQHQHHEHHEHARR/AEVTKVYVYN	281		34%	3%	6%	3%	60%	1%	16/(no TMs)372
YDR452W	448	NKEAYMNTVRDYYRQVWVKLER/VDEKNVENEK	86		27%	4%	12%	3%	60%	2%	i21-40o674
YLR300W	39	*VPARDPSSIQFVHEENKRR/YDYDHGSLG	346		32%	4%	8%	2%	60%	4%	20/(no TMs)448
CAGL0M00132g	872	AIFRSATSRHDTYTTISAFFTPR/LSLQITTLVQ	44		34%	14%	5%	3%	60%	15%	24/(no TMs)(892-GPI)o922
CA2830	243	FPAKRELDEFEELSNDGVTHSKR/TLGLLGLL	440		26%	9%	13%	5%	60%	4%	18/(no TMs)(684-GPI)o714

Anhang

CAGLOA01738g	310	LKSLIDNSADEKDFNINRHRRR/FDKEYDHSQK	132		25%	5%	14%	4%	60%	6%	25/(no TMs)469
CAGLOI01848g	14	*ICADGRKPAMI FRR/DSDSDSDGDG	208		32%	9%	7%	2%	60%	10%	o53-76i701
YLR353W	119	QFHDTQDNNIPSNNTTVRPTSLRR/DNEDTVPLRN	115		23%	4%	17%	9%	60%	12%	o516-536i576-600o603
YHR204W	511	DVIFSTEAHMWPQEVRSNYKR/NAKFNNVSVYS	196		31%	7%	8%	1%	60%	8%	21/(no TMs)796
YIL014W	61	KLNNHASERTAVESSAFNWIEKR/QHQVRSENLM	164		31%	11%	8%	2%	61%	12%	31/(no TMs)630
CA3122	204	QVAKRDAKNVINLSNFIETPSKR/EGKNLFDLTK	75		29%	6%	10%	8%	61%	3%	19/(no TMs)(?-GPI)o410
CAGLOF06369g	209	SVKDEAVLVSDGLSVAIDYALKR/PDLEINVPQY	265		26%	6%	14%	6%	61%	3%	16/(no TMs)889
YBR229C	591	YGLSVHEATYDAIKSIYSPSDKR/PFLLTRAFFA	43		29%	4%	10%	3%	61%	3%	23/(no TMs)954
YHR042W	522	VHYDLNGPRKLFANYKLPVHVR/ SNFRLPSNPS	319		24%	8%	15%	5%	61%	10%	(no TMs)691
CA2633	99	DKSHPHNSFLGNHNSNEYNLMKR/QLGHEQDYDE	60		27%	7%	12%	8%	61%	5%	21/o774-794i936
CA3643	72	FSQVVLSTPNWDQTLTKYKRKR/DYNDDSLASM	707		26%	2%	13%	3%	61%	3%	23/(no TMs)(?-GPI)o229
YFL042C	467	EGSQITTIPNFTSIDGSSMTLKR/EYSYEKALHF	180		28%	10%	11%	5%	61%	5%	o636-654i674
YHR079C	34	*PLSSRTSRR/QIVEDEVAST	52		32%	5%	7%	3%	61%	7%	25/o534-555i897-914o1115
CA4913	629	AMNGIKRVKGGKLGHVSDWWRKR/GESDGY* . . .	226		27%	2%	12%	2%	61%	1%	i21-39o636
CAGLOK05137g	946	RIKVPNRDILHDRNITSALYSKR/DDDLSATDDY	553		25%	9%	14%	7%	61%	4%	i83-102o1212
CA3529	494	YQASIEFGGFSNALPKPKFNKR/NLSPTEDHLK	323		28%	13%	10%	6%	61%	19%	25/(no TMs)837
YER019C-A	16	*AASVPPGGQRILQKRR/QAQSIKEKQA	118		32%	9%	7%	3%	61%	11%	o62-81i88
CA2830	225	NAIGTFDASALFNAQVKAFAKRR/ELDEFEELSN	188		18%	5%	21%	8%	61%	4%	18/(no TMs)(684-GPI)o714
CA4295	188	KDTSIPDLENYLTKANDKYFKR/QEYWRYLKLD	548		20%	2%	18%	11%	62%	10%	i12-29o756
YDR331W	140	VTVENFIRLTLTDRWTEHDHPSKR/LTDENSNIF	434		32%	5%	6%	3%	62%	3%	23/o378-397i411
YLR299W	536	QFNKNSFELAPSIYNFPEPGKR/PLSSTAPTIV	132		18%	5%	20%	5%	62%	5%	i12-33o660
CA3295	318	GGILGSSGYTGSTSKVCTLKRR/SLEKKEIRKL	64		20%	7%	19%	8%	62%	4%	23/(no TMs)(?-GPI)o1249
CA1621	154	PVFSITAADWHNGLLVLAQAKR/TALANLGF	247		22%	8%	16%	6%	62%	6%	20/(no TMs)(1063-GPI)o1086
CAGLOE01353g	9	*DIVQLLFKR/DDDDGPPPAC	164		25%	6%	14%	4%	62%	3%	o34-53i65-86o106-126i277-297o338-358i370-389o409-429i433
YOL075C	352	LQLQAESYISNATEIQIQNMTTR/LPFWKQVTVL	50		28%	7%	11%	1%	62%	7%	o375-397i409-428o469-485i492-517o523-542i554-572o601-625i1038-1059o1121-1141i1148-1166o1178-1196i1208-1226o1266-1285i1294
YJR075W	292	KPGHPVLREIISRIETTLQRKR/DDQLNVNLRN	362		24%	4%	14%	4%	62%	4%	29/(no TMs)396
CA1360	222	DRAELASTKNNLVRSPKSNRR/LIGSSRKKTL	162		33%	9%	5%	3%	62%	6%	17/o712-733i734
CA5859	434	QTVYVTLTTTVGSQLQSLFDKR/DVIAEAKSTN	191		26%	4%	12%	4%	62%	8%	23/(no TMs)(?-GPI)o462
YLR057W	668	NLDLVGIKVSNYKGGNDQAKR/NVLSKDGITE	434		28%	23%	10%	5%	62%	19%	(no TMs)849
YPR026W	449	GLAKMIALNSKGNYEKLLSSHKR/AWYDLYNDAF	160		26%	5%	12%	3%	62%	7%	i47-69o1211
CAGLOD05852g	406	EITSEGSQASKLYDSAFCSNKR/LGGHCQKFKI	111		18%	8%	20%	10%	62%	5%	23/o753-775i864
YLR396C	582	YDRSILFHNYSQQPFILSREPR/VSQTEDLIEQ	58		28%	3%	10%	7%	62%	10%	o640-657i691
CAGLOJ11748g	53	NANFIRNAADGLSPAEEKWLKRR/DPITRDALQT	64		24%	12%	14%	7%	62%	19%	20/(no TMs)(667-GPI)o695
YFL027C	357	LWEEVLQITRKDATTASKKALKR/VSLNKYSALL	212		32%	9%	6%	1%	62%	10%	o429-445i497
CA1387	156	PNSRLMEQKFAKFAKALTDSFKR/YVPKIREEIL	159		29%	9%	9%	1%	62%	9%	o13-36i48-65o528
CA3758	344	IAEITATNGAQIAYEVLTSVMKR/GNIDKSLVKS	454		24%	4%	14%	5%	62%	3%	16/(no TMs)361
CAGLOE01727g	513	TSVPSDDKKI IDDDHSHGDHAKR/DAYSNTAGIL	550	0	30%	25%	8%	3%	62%	27%	15/(no TMs)(519-GPI)o539
CAGLOG02717g	303	NRTGFISKSANHIISCPPIYGKR/SGLDIATILA	310		23%	8%	15%	4%	62%	6%	17/(no TMs)515

Anhang

CA3925	18	*PIPRSFIFEPELEQLLKR/DENGCVSCPS	119	26%	3%	12%	2%	62%	5%	o69-94i527
CAGLOJ08910g	458	SPNSEYSNENDSEEINVEEWYRR/SMMFRPKRQK	131	31%	6%	7%	3%	62%	7%	23/(no TMs)(470-GPI)o494
CA5699	1268	ESSSGTILSESSSDSIPTTFSTR/YWSPSGMSSR	47	28%	4%	9%	5%	62%	4%	19/(no TMs)(1980-GPI)o2000
CAGLOF03575g	152	NNKARGENGKFSHDRILLDYARR/SVDLSKLEKD	181	27%	7%	11%	2%	63%	7%	i106-124o755
CA2825	38	*VPQVDGQGETEEALIQKR/SYDYYQEPD	234	27%	1%	11%	1%	63%	2%	20/(no TMs)(604-GPI)o634
CA5928	87	KTETYSVEVINATTLSENTRKRR/FDVSGPVRTP	707	32%	9%	5%	3%	63%	6%	26/(no TMs)232
CAGLOM08734g	56	RLAPEEYRGGVLDGEESRLQKR/DLEMHPVVDL	94	29%	10%	9%	2%	63%	9%	(no TMs)253
YKL201C	661	MATLNITELRDYIITSDENKNHKKR/VPTDIDLKDL	462	29%	11%	8%	3%	63%	10%	i28-45o1178
YBR015C	316	FWRRTTHPLYDYDIAGIAVDKRR/VRNSRDDITP	203	23%	6%	14%	5%	63%	2%	(no TMs)597
YNL238W	108	QGDNDHILSVHDLFPRNDLFKR/LPVPAPPMDS	84	29%	10%	8%	5%	63%	14%	24/o676-699i814
YJL159W	71	VQPISTTSSASSAATTASSAKR/AASQIGDGQV	434	31%	2%	6%	4%	63%	4%	24/(no TMs)(280-GPI)o310
CAGLOL01727g	385	GMYGLGEQMAALEVMVNTQALKR/APPYNATNGG	212	25%	3%	12%	4%	63%	3%	19/(no TMs)(420-GPI)o446
CAGLOM10571g	172	KFAGPTLEASIKSKEKELRKLRR/EQKRQLDSGE	195	31%	6%	6%	3%	63%	9%	o200-221i242-262o274-293i305-324o380-397i418-442o462-482i544-563o569-587i599-617o619
CAGLOI00484g	38	*VIAPKNKDTSLHFVNEKR/YYDYDSKAIG	358	27%	2%	10%	1%	63%	2%	20/(no TMs)443
CAGLOH02893g	618	PVAYMYIYQRLGHFNMMNRR/SEPCITLRCY	220	23%	7%	14%	7%	63%	3%	o678-700i867
CAGLOC00363g	260	AVNYATPAAMPSPDNGGWYNIPIKR/QSLSTRKLVN	104	17%	3%	20%	7%	63%	5%	17/(no TMs)276
CA3295	401	VTAVSSVLLTVAALTNANEIDKR/SFFGDLFSGL	71	25%	4%	12%	2%	63%	5%	23/(no TMs)(?-GPI)o1249
CAGLOM00132g	324	VGIPCSSSQSIYGSSSSSSSKR/LDCPSGLFFD	128	31%	16%	6%	3%	63%	15%	24/(no TMs)(892-GPI)o922
CA0267	36	*GNDAEYLLNTVLKR/DATSLSRNAY	593	27%	4%	10%	1%	63%	4%	22/(no TMs)(1091-GPI)o1114
CA0875	304	ASSFTLWDWLFQTECGTYARKRR/ERNSRASAES	180	29%	1%	7%	2%	63%	2%	(no TMs)321
CAGLOE05412g	1146	IHVEFVKYEWPPWLRPQRFRQKR/LAISKLLFLD	669	30%	12%	7%	2%	63%	14%	19/(no TMs)1326
CA1032	531	DYAVSVLAKDFGSSDDHDKYLKR/SANWQNIWNF	398	24%	4%	13%	6%	63%	4%	21/(no TMs)795
YIL005W	138	SGYIKSFTETPRTKESLIAFARR/ESMDPNLDT	231	27%	2%	10%	3%	63%	4%	28/o652-672i701
CA0403	391	GKWGLGEQMSALEVMQNLMVHKKR/PAPYTADTGG	528	23%	6%	14%	5%	63%	1%	19/(no TMs)(425-GPI)o452
YIL077C	25	KEEEQQYQGNGKGMENELPFMKR/PWFKKAYENA	401	18%	6%	19%	17%	63%	22%	o67-86i98-119o320
CAGLOH04499g	179	GGPGGFRMYTTGNGPFQRRFR/ADPREAYRQQ	595	26%	17%	10%	3%	64%	18%	o207-226i232
CA4176	193	SEIMKPIGLKNNCEWLQKEANKR/TRFPDFCKDL	81	24%	5%	12%	4%	64%	3%	i7-26o808
CA5418	86	FNKTPRPVTTVLVKESDLPKKR/DAVVARDSKN	141	28%	3%	9%	3%	64%	6%	19/(no TMs)(105-GPI)o131
CAGLOK05137g	482	KFASNIANTAKGNYKSLLSNHKR/AWYDLYNDAF	462	26%	5%	10%	1%	64%	6%	i83-102o1212
CAGLOH00330g	2514	MSFSQDTRIHTGTKIDEMFKR/STNFLT IENM	494	28%	16%	8%	2%	64%	16%	i35-59o2635
CA1975	193	GIEHPKIQVEEWNVGNQVASKR/NANFNVS LTD	151	23%	2%	14%	4%	64%	5%	18/(no TMs)605
CA5363	100	GKDFDYICCSPEQIDILES NLKR/VDPLISSCPA	519	24%	3%	12%	3%	64%	1%	29/o272-293i355-375o587-606i618-642o648-670i700-719o725-750i810-832o1068-1091i1098-1118o1124-1145i1196-1222o1234-1254i1256
CA5814	67	NDSYFVSLKASESIDHFMEYDKR/YPKHLQVRDF	60	21%	3%	15%	10%	64%	11%	i21-40o421
CAGLOK00297g	471	DAVSFSYSGTGALKTDFTRTGKR/TKAGAFNDFV	56	24%	7%	11%	2%	64%	8%	o526-548i560-577o627
YBR054W	9	*SDYVELLKR/GGNEAIKINP	76	24%	3%	11%	4%	64%	7%	o35-56i63-82o117-135i142-162o168-193i200-219o239-260i344
CA0825	37	*PTNLYTPGYVQCPEGKLR/SSLDGINSNE	79	17%	4%	19%	5%	64%	3%	18/(no TMs)609
YGL156W	74	HVKVWVYQVSFERGSSPVPDKR/PSWKSIIERD	51	22%	9%	13%	14%	64%	22%	(no TMs)1083

Anhang

CA0380	84	NVFNVDQADGQVDGQVLKVFIEKR/DANTEQTIEG	438		21%	7%	14%	5%	65%	3%	18/(no TMs)(431-GPI)o462
CA3805	63	RLIKLGPENYRIVSENEKLLKLR/NKINFIDVTN	305		25%	4%	10%	1%	65%	5%	19/(no TMs)417
CA1379	123	DGFGGNDCEPLCGSLGDGNSKR/PVRKKKQCAC	148		26%	19%	10%	6%	65%	25%	22/o334-355i784-804o816-839i860-884o896-921i928-946o1012-1035i1038
YJL079C	173	ASSSDSDLSDFASSVLAHEHNKR/ALHKDTPALS	277		27%	1%	8%	1%	65%	2%	20/(no TMs)299
YDL241W	102	DFYACGFLQKGTETNAGIHNR/PSHSLAKFTI	62		20%	5%	16%	7%	65%	5%	(no TMs)(96-GPI)o123
YDR003W	13	*ILREQIDFLIHKR/QDDNNNNGEA	412		24%	4%	11%	3%	65%	2%	o38-59i210
CA1507	13	*SAAVSTLSDI IKR/NDAVNVNPPN	181		24%	3%	11%	3%	65%	5%	o38-59i71-89o117-139i159-179o185-203i215-233o253-275i342
CAGLOJ11770g	64	LRNATGLSQSEIDWLKRDVNTR/EALESFLKRV	59		24%	4%	12%	3%	65%	3%	23/(no TMs)(633-GPI)o659
YAL058W	461	GPTERKFMNSRLGNLQTTFHNER/ESPMPFDRII	53		22%	4%	13%	6%	65%	8%	30/(no TMs)502
YHR042W	25	GIDNTDFTVLAGLVAVLLVYVNR/NSIKELMSD	403		22%	8%	13%	4%	65%	5%	(no TMs)691
CAGLOF06369g	734	IGLNVTEKYNSILSKMSFSSIRR/SSEENIKTLA	56		30%	4%	5%	3%	65%	4%	16/(no TMs)889
CAGLOF03575g	312	YGASGILFVSNNSWDDNDIVEKR/SVALPQYGLG	327		26%	6%	9%	4%	65%	10%	i106-124o755
CA2786	1390	DAKMIDLCNNPLTRENKLDAAKR/LIPEWIEYEQ	151		28%	3%	7%	2%	65%	3%	23/(no TMs)1447
CA5814	140	NALEFEIQEDAPRHLARISRRR/MKPNKPYPYM	144		24%	4%	11%	2%	65%	3%	i21-40o421
YBR290W	98	QAARHIGKIGRHFNLDRLFKRR/TQQSSDIQQG	339		23%	4%	11%	2%	65%	4%	o179-200i212-230o281-301i321
YHR149C	31	LISVTSTVSVSDASYAYARLTR/DDSDSSSSSA	52		24%	5%	10%	1%	65%	5%	o72-97i734
CA4176	76	YLKEEDSLPEEQELQLLHKVNR/NENNTYWLAN	772		22%	3%	13%	6%	65%	6%	i7-26o808
YCR089W	1039	SEMSTFSVSTQSLPLAFTSEKR/STTSVSQWSN	99		26%	3%	8%	2%	66%	2%	23/(no TMs)(1588-GPI)o1609
CAGLOA00253g	50	LQSLFFKNSPISVKFETLVKERR/SIQEENKSIS	214		27%	1%	7%	3%	66%	4%	20/(no TMs)223
CA2769	21	*FIKR/YENTTAPASQ	389		20%	3%	14%	4%	66%	2%	17/(no TMs)342
CA1371	248	DDDTDIQQIKNDKDFHSHKSKR/NLSSLSEEEE	434		29%	5%	6%	3%	66%	4%	i103-125o375
CAGLOI00506g	614	SVVNAIQRIAGEWHAVSDYWRKR/GISAVY* . . .	181		19%	6%	15%	6%	66%	5%	22/(no TMs)621
CA1360	690	HSRIRFPTADDNNNAGNHYHRR/FTGVVLPENR	226		29%	3%	6%	3%	66%	3%	17/o712-733i734
YOL105C	170	STSYMPSTTSSLSSAQISSTRR/TSTDMSSEM	63		27%	2%	7%	3%	66%	3%	39/o381-405i556
CAGLOH02893g	179	DDKRAFVYPKPLPKFWKYENDKR/LTLEEQKEEK	125		26%	6%	8%	2%	66%	6%	o678-700i867
CA3931	223	RPIVCEKLRKIGKNLSDKRSTR/SEIHEKFGDT	98		18%	4%	17%	5%	66%	4%	(no TMs)330
CAGLOL08096g	132	YCFHLSPTQTNNLKTAHNGYCRR/VETLFGSLDE	200		23%	4%	11%	2%	66%	3%	27/(no TMs)564
CA5423	766	QLSDIPMLLKGGSIIPMKTRYRR/SSKLMKSDPY	204		20%	7%	14%	6%	66%	8%	17/(no TMs)871
CAGLOK00297g	99	ESGRLDGHI IYKVQHSIVNAKR/SSRIDKDEAE	148		23%	7%	11%	1%	66%	7%	o526-548i560-577o627
CA0856	190	SGSDCEDDECKRRKKVHRNYKR/GYGSERRQS	392		26%	6%	8%	2%	66%	8%	o28-45i57-78o356
CAGLOH09240g	187	VNPRELFFKLDLQATDAYDFQKR/MFFFLKNVTQ	401		27%	9%	7%	2%	66%	11%	33/(no TMs)642
CAGLOA01265g	35	*VEDKLNKIKVFEKR/AVDFTPPFAHY	438		22%	5%	12%	5%	66%	5%	21/(no TMs)727
CA4846	22	*SLIDTNSVNGIITVREESMGKR/WESLSPIIWQ	437		20%	5%	14%	3%	66%	3%	(no TMs)232
CA0229	193	GTNDDSQLSSFSRQILEAHNIKR/ASHGVNPLTW	254		23%	4%	11%	2%	66%	3%	24/(no TMs)317
YLR250W	138	MRLLDVDDNTKITKEEYLQFAKR/GNKFPDLGVG	181		22%	2%	12%	3%	66%	4%	21/(no TMs)234
YOR115C	86	GDVISSDSNVTSSENGNINMIKR/LKIEEHTVD	137		28%	4%	5%	4%	66%	2%	(no TMs)268
CA1166	392	SINYQFNGDDYCDGNSDGGQIKR/SLMENASAAQ	212		25%	3%	9%	3%	66%	3%	22/(no TMs)412
CAGLOL13244g	352	LEKEATDFILQMVGLLDIIVKRR/SEVFNQSDEN	396		22%	5%	12%	2%	66%	4%	19/o412-435i447
CA4624	384	HNGVPSNEGSSSDSDFHKFRITR/SKSPHKRKIR	34		27%	5%	6%	3%	66%	8%	o723-741i757

Anhang

CA5160	291	STDNIPPIPEGQTLTKDYKGRR/YDEAGNLVVL	378		23%	10%	10%	1%	66%	10%	o7-24i44-64o85-104i146-166o186-205i225-244o476-495i56 6-590o600
CA4251	183	DWLCKRRNALLKKQVHDSVQKR/VIPAQYFRDP	375		23%	1%	11%	3%	66%	4%	(no TMs)316
YNL158W	41	*ESILYKVPHNFLPKKR/DSSTYARDVN	51		22%	6%	12%	3%	66%	3%	24/o168-187i198
CAGLOH00330g	1228	LSDFTYVWGRVNSKLPDDINKR/VLDFSTNPFLL	228		27%	4%	6%	4%	66%	7%	i35-59o2635
CA1646	33	*LPIVEGLLNCVFPR/PPPPTIPPSP	52		24%	4%	10%	1%	66%	4%	19/(no TMs)725
YOL046C	165	FGDMGLYENACGDSLGTATRSE/SSSLNVDENE	54		28%	10%	5%	2%	66%	8%	(no TMs)(204-GPI)o224
CA2730	232	LADISKTEEDPLMKKIRVVLKRR/GIINGIPVVF	51		22%	8%	12%	3%	67%	6%	24/(no TMs)434
YFL051C	116	YAVLKRDEDEGYPGPLYETKRR/DTEYCDPNTA	150		14%	5%	19%	2%	67%	6%	24/(no TMs)160
CAGLOG08338g	176	DNAVTKAGISLLKRYIQTSNSKR/DHNSGSSNPE	369		16%	4%	17%	4%	67%	4%	o32-57i120-139o191-212i449-471o499-523i544-567o587-6 14i635-656o662-682i703-721o727-745i794
CAGLOL04466g	69	GKSTLLNLFNTQFDTMDAQVQR/QQTTKGIWLA	351		24%	6%	9%	3%	67%	5%	o678-694i701-718o783
CA4759	294	TTFCPYDTCYDYLHHTENNRR/VDQGGLEGNT	106		27%	1%	6%	3%	67%	4%	20/(no TMs)575
YNL012W	182	SEEEIVTNAKKEYNDNDLKMKKR/QGGSLITSSS	300		22%	3%	11%	3%	67%	2%	18/o374-396i631
YJR038C	106	VKPTSASFGGIRLSQLTMIMERR/ATPTCQDPSL	158		21%	5%	12%	3%	67%	3%	o20-42i54-72o120
YOL105C	346	TSVVTKAIVSTSDQHQETIFVTR/TSVVERSSSEV	67		27%	8%	6%	3%	67%	11%	39/o381-405i556
YJL216C	415	FKNRKIEQIKCVEGTGYAAIKR/DYGEDSEKMK	162		21%	7%	12%	2%	67%	6%	(noTMs)581
CA3713	53	VKREDVNAIVELINEIKHINQKR/DLAEGEDLLE	410		22%	6%	11%	2%	67%	6%	21/(no TMs)403
CA1159	221	KRCSWYTASISPDDKVNSSYNKR/PENINCKYIK	382		17%	4%	16%	7%	67%	6%	o270-290i434
YDL024C	175	LAKYKLMENCTNFPIFTNSKR/IYDTAQYFAE	369		18%	8%	15%	5%	67%	4%	21/(no TMs)468
YHR045W	422	VKDAEVISTFPEKIPQLKKNRR/LTPSTQLNK	56		26%	9%	7%	2%	67%	10%	20/(no TMs)560
CA4700	49	VRRGESKDDLSPEDDSNPRFVKR/DGSLDMTLTN	643	1	23%	30%	10%	6%	67%	27%	18/(no TMs)(520-GPI)o544
YDR245W	173	NKKAYAKRHGYALTIKDLTTSKR/YSHEYREGWQ	102		16%	4%	17%	4%	67%	1%	i50-67o393
YIL117C	8	*TVITIAKR/GLPKLTSTTS	169		21%	4%	12%	3%	67%	3%	o77-98i318
YLR083C	249	YHDRGEGNYRVVGVIVEPVSIKR/SSPGTCETTG	88		20%	8%	13%	3%	67%	5%	23/o305-327i371-395o401-426i438-462o474-496i521-548o 560-583i595-615o627-657i667
YEL060C	86	HGKFHKKGRKQDKESPEFNGKR/ASGSHGSAHE	323		25%	26%	8%	3%	67%	24%	19/(no TMs)635
YOL105C	352	AIVSTSDQHQETIFVTRTSVVER/SSEVATATAA	54		28%	6%	5%	2%	67%	8%	39/o381-405i556
YLR299W	441	LPKHIEEVLDPWALKAVKSIKR/NSQDGNFKTL	88		21%	5%	11%	6%	68%	10%	i12-33o660
YGL089C	100	KREAVADAWHNLNLRPGQPMYKR/EANADAWHWL	530		21%	2%	12%	6%	68%	5%	20/(no TMs)120
CAGLOC02607g	11	*STNWNTWKLKR/ISASNICQAF	610		25%	4%	8%	2%	68%	4%	o634-65i677
CAGLOL12804g	207	DSKGFDKLQEKERHALAVQKERR/SAMALARNEL	214		26%	7%	6%	3%	68%	10%	(no TMs)393
CAGLOK03927g	156	TILERPLSFHIEFTPEEWELEKR/PEFGPLSLL	66		20%	5%	12%	5%	68%	4%	i37-57o235
YKR020W	20	*AEQISHKSLRVSSLNKDRR/LLREFYNLE	93		21%	5%	11%	2%	68%	4%	(no TMs)164
YNL160W	175	PIVVTEDSAYAI PVANNKNATKR/GVLSVTSDKL	181		26%	2%	6%	2%	68%	4%	20/(no TMs)(322-GPI)o354
CAGLOF09053g	608	KVSKYRQNDAGFSFSNMSSFKR/ISEPERIFNT	198		21%	1%	11%	3%	68%	3%	17/(no TMs)809
CAGLOA04081g	345	VHTSVSHSASHSTQS SVSEVAKR/AVHFENAGVV	338		26%	5%	6%	4%	68%	6%	18/(no TMs)(352-GPI)o374
YNL012W	217	YEQIEEIMNSIEEIPEDYMITKR/NLNPLARLKK	203		20%	5%	12%	2%	68%	3%	18/o374-396i631
YNR060W	33	*PPSKTSLINTHERR/SIYSCYVGLR	217		24%	8%	7%	1%	68%	9%	19/o157-180i229-249o269-290i311-332o344-364i371-388o 394-412i719
CA1166	31	*ISTTKFLKR/GSLPSSSSFD	509		23%	4%	9%	4%	68%	7%	22/(no TMs)412

Anhang

CAGLOG08668g	44	./AVTVEYVYAT	232		26%	1%	5%	4%	68%	4%	23/(no TMs)(449-GPI)o469
CA2234	34	*ISDLNNIASNQVTKR/LGGGSRGGSS	406		25%	3%	7%	3%	68%	6%	19/(no TMs)(185-GPI)o208
CAGLOF05137g	80	TKSQKTKPTPSTDLDKRAQKQR/SNLSEWQQKM	232	1	24%	23%	7%	2%	69%	24%	18/(no TMs)227
YIL014W	387	NEQTFQYCIEMLNEVEPSAQERR/FIGSRLVFDS	122		23%	4%	8%	4%	69%	8%	31/(no TMs)630
YAL034C-B	85	LIFNASYLPFLFTRTIKISKVER/QQGERSKMAM	56		20%	7%	12%	2%	69%	6%	23/(no TMs)117
CA0416	67	KQRELDGSSTDIELKPKFRISKR/YSNKSIATL	169		23%	2%	8%	1%	69%	2%	25/(no TMs)207
YHR204W	764	ILPRHGSVQIKKHSPVLTSNGTR/EEDEFKMDGI	82		14%	4%	18%	13%	69%	12%	21/(no TMs)796
CA5815	116	KAYDAVFQDSAPRHLARISRRKR/MKPIKKYSYI	283		20%	3%	11%	1%	69%	3%	22/(no TMs)379
CAGLOI07777g	164	GSRSVNPLQGFITMVVNKDVMKR/QLMDIIEGRP	375		24%	5%	8%	2%	69%	7%	16/(no TMs)240
CAGLOL13134g	276	VHPEKVGAAEVNGLSASNLTSRR/SSTSSSPLQK	52		27%	11%	4%	3%	69%	9%	o622-644i725
CA4295	642	HPKIDPLKNSKEADAIEFGVYSRR/NMGLDPKVEK	145		17%	6%	14%	4%	69%	3%	i12-29o756
CA0340	47	ETKKDVVAANADGTVGFNLVGKR/EVSEGEISKR	462		22%	5%	9%	1%	69%	5%	22/(no TMs)195
CA3354	74	STTTASILSALFNDIILPTIFKR/DYSTRDHKAN	326		17%	2%	14%	6%	69%	5%	o100-122i134-152o172-199i220-241o261-287i440-460o472-495i823
CA5423	544	NVFGLSYHETTFNSLLNRSPEKR/PFILTRSYFA	117		25%	5%	6%	3%	69%	6%	17/(no TMs)871
CA5344	134	SDATPTADVEFAEELKEHNVKR/ALHGVPAISW	420		23%	4%	8%	2%	69%	6%	20/(no TMs)271
YIL014W	557	EDDFGREPEYFKSRYGDISLKR/IYDASLNVEG	763		23%	4%	8%	3%	69%	7%	31/(no TMs)630
CAGLOE01881g	479	SASSSGNTSSNGSTNLTRRFERR/DIPLNKGVP	223	0	24%	17%	7%	1%	69%	18%	(no TMs)(469-GPI)o508
YOR365C	313	ATSGSPTVVYRNKEVLSISVQKR/SLNSKIISAS	375		23%	8%	8%	2%	69%	9%	25/o203-222i234-259o374-396i417-436o456-478i517-537o543-561i573-595o607-634i703
CAGLOL04466g	165	YQGNMGLLKTVEVNLISLFGKR/GNDHKVLLLF	495		20%	5%	10%	5%	69%	6%	o678-694i701-718o783
CAGLOJ00583g	516	ALLSYDGPDPVYQKIIELDTKR/DEHIHNSIG	68		16%	5%	15%	6%	69%	2%	i24-45o817
CA4679	354	LREIQKREEGIDDEIARLKMKR/DNIEDLLTTI	602	1	17%	9%	14%	14%	69%	21%	20/(no TMs)(535-GPI)o557
YBR015C	321	THPLYDIAGIAVDKRRVNSR/DDITPPAVYT	41		19%	6%	12%	2%	69%	4%	(no TMs)597
CA3642	56	EPNDIQFFNQVRSLSYIYQKR/DVSDSDNEQV	393		20%	9%	11%	1%	69%	8%	21/(no TMs)(210-GPI)o229
CA2223	138	SKKTFIRTASELSQEKDYIHKR/QETTNKNLID	330	0	24%	17%	7%	2%	69%	18%	20/(no TMs)(720-GPI)o754
YKL212W	218	TPIVLGLITRRSIFRAGTRYFRR/GVDKDGNGVN	279		22%	5%	8%	3%	69%	7%	o522-543i555-573o623
YEL028W	135	YMAKLGDATVNHTDLPHFRNKR/LTTQELNAFQ	319		20%	5%	10%	2%	69%	4%	22/(no TMs)153
YFL034W	62	GLSNQDDLTVMHTQAKEEVFKRR/EEEDGTRTEDA	338		23%	4%	8%	2%	69%	6%	o461-479i542-565o585-606i1073
YOR389W	377	IDYRYLVSGINRTYIAPDPNRR/LLDEGMTWEK	203		21%	9%	10%	3%	69%	7%	24/(no TMs)624
CAGLOB04565g	159	GLIPREHWSYPEWIDQEKARLTR/EEMKKQKIY	172		22%	6%	8%	2%	69%	7%	37/(no TMs)401
YNL283C	266	STATVTSTPSSTSIGTSTHYTTR/VVTQSVVSQA	44		24%	7%	7%	4%	70%	11%	24/o322-347i503
CA3825	223	TNLTDSMSEENKIVDLRIPMKR/WGNPDDFMGP	134		17%	9%	14%	3%	70%	6%	(no TMs)260
CA3347	141	VHRIFFDVAQGSANEYVDSKSTR/MVDDLTKGVN	71		25%	10%	6%	3%	70%	9%	23/o186-205i218
CAGLOA02255g	98	EKTGYVLLNITQNYAGNDTMAKR/DVKVYGYDGW	200		26%	19%	4%	2%	70%	18%	25/(no TMs)272
CAGLOA02277g	98	EKTGYVLLNITQNYAGNDTMAKR/DVKVYGYDGW	200	0	26%	19%	4%	2%	70%	18%	25/(no TMs)272
CA3518	605	NIYGLIQERAIYEALLEIHPNKR/PFIIGRSSFA	130		24%	7%	7%	3%	70%	9%	21/(no TMs)(?-GPI)o862
YDR415C	217	DNGSGVTNMEALRLYTENFLKR/GFRPNNTVEF	254		21%	2%	10%	1%	70%	3%	19/(no TMs)374
YKL212W	467	DAVSVAISGTGALKTDFTRTGKR/TRLGAFNDFL	56		21%	4%	9%	5%	70%	8%	o522-543i555-573o623
CA3518	775	QFFVGDALLVTPVLEPGVNHTRK/YFPQARMSCT	530		15%	3%	15%	6%	70%	5%	21/(no TMs)(?-GPI)o862
CAGLOC03696g	96	SIEHSITKLLDNEINDNRKVLRR/IKTIKNYLET	121		23%	5%	7%	3%	70%	7%	o735-761i773-799o805-827i831



Anhang

YAL053W	291	ATNGVSNVVVANKDVLISISVQKR/AISSMASSDY	375		24%	5%	6%	2%	70%	6%	25/o348-374i403-424o430-455i493-513o519-538i550-571o583-609i783
CA0854	800	CSGYMWCAYDKIGGKTKSGKNTR/LEGKIINFDE	76		24%	9%	6%	2%	70%	9%	i14-32o828
CA0855	605	NIYGLIQERAIYEALLEIHPNKR/PFIIGRSSFA	130		24%	7%	6%	3%	70%	9%	21/(no TMs)946
YGR141W	416	SLCQRFKWWNFWKGCPCARRYIKR/GEGLDAEKND	304		18%	0%	11%	3%	70%	3%	i15-37o467
YBR042C	96	VRITTENSSVPKGTFFLDLKKKR/ILSHLKSNSV	407		20%	8%	10%	2%	70%	9%	o13-37i58-77o120-138i377-396o397
CAGLO09449g	74	TAEDNPGKIEYTSGLAMSLTKR/FDNPSLKS NF	82		20%	4%	10%	1%	70%	3%	21/(no TMs)(421-GPI)o452
CAGLO04279g	642	CTFVDRIGDADSKLVRFDKKTKR/FVNHLAQIWR	522		21%	7%	8%	2%	70%	7%	i27-44o660
YBR229C	816	PLDKIPLFIEGGHII TMKDKYRR/SSMLMKNDPY	294		22%	5%	8%	3%	70%	8%	23/(no TMs)954
YLR120C	546	QTASFSGNLTSTASATSTSSKR/NVGDHIVPSL	102	0	24%	18%	6%	2%	70%	19%	22/(no TMs)(542-GPI)o569
CA0371	32	*TAGASSRRRRR/GSLPNEKTIY	288		25%	6%	4%	3%	70%	4%	21/(no TMs)806
CA1975	552	VGCAIIRREQERGIEQTEQCKR/CFENYCW DGT	343		23%	4%	7%	3%	70%	5%	18/(no TMs)605
CAGLO01991g	121	FGSMSMIKPDSSNAHQIGSEKR/YPYDNL LYFP	99		12%	2%	17%	3%	70%	2%	26/o49-70i82-107o176-195i216-239o330
YJR151C	652	LSATSTFTSAVVSSSKAADFFTR/STVSAKSDVS	73		24%	3%	5%	3%	71%	4%	20/(no TMs)(1127-GPI)o1161
CA1676	419	LPENKFTKIVASRESIDLSVQRR/IEQENFNCNV	239		23%	4%	7%	3%	71%	6%	o430-447i448
YLR213C	127	EGNGPNKANTKVGVEGELNSKR/IIHYAKFLVT	295		23%	7%	6%	3%	71%	10%	21/(no TMs)422
YLR120C	66	KLYGDSLENVGSDDKPEVRL LKR/ADGYEIIIT	95	0	24%	17%	6%	2%	71%	18%	22/(no TMs)(542-GPI)o569
YDR107C	74	EEGNDVSGDKEHFLYSYDYYNKR/FHFCRPEHVE	306		23%	6%	6%	2%	71%	8%	21/o310-332i373-399o411-431i443-467o479-501i533-553o565-591i603-621o633-662i672
CA0267	186	YTDINSESTITGDYTFIKNKNTR/TVIITDYVTS	61		22%	6%	7%	4%	71%	10%	22/(no TMs)(1091-GPI)o1114
YHR079C	98	SSYPTPNLLNTADNRANKGR/AANSISVPYL	242		24%	12%	4%	3%	71%	9%	25/o534-555i897-914o1115
CAGLOI00506g	497	QPNRSNSFDLAASIYNLIEPGKR/PLSSTAPSI I	132		18%	5%	11%	6%	71%	4%	22/(no TMs)621
CA3952	160	GSTPASLTFQKKKTISQYLNKR/QIELKSDTIN	73		22%	5%	6%	3%	71%	6%	i21-44o600
YKR013W	203	ASSTQSSSSDFSTSMVNEHNTKR/ALHKDTGSLT	355		23%	5%	6%	4%	71%	8%	19/(no TMs)329
CAGLOE01595g	388	EISGKTENDIPAKAIKTGIQV KR/GHILELLNSN	351		20%	3%	9%	1%	71%	4%	23/(no TMs)(439-GPI)o468
CAGLOF01793g	315	LYFNYNYGQFTTLWDRLGGSYRR/PEDSLFDPKL	108		12%	2%	16%	7%	71%	7%	o95-121i142-160o183-201i255-272o364
CAGLOJ11770g	73	SEIDWLKRDVNTREALESFLKR/VTSNFTSNSS	509		22%	2%	7%	3%	71%	5%	23/(no TMs)(633-GPI)o659
CA0349	292	TDILKSSYLSVSVQRFVKKFIRR/SVLAMPDQQR	159		21%	6%	8%	4%	72%	8%	23/o373-396i420-440o446-472i510-530o536-555i567-588o600-627i811
CA4882	30	*VVVVEKR/AINETALAED	327		16%	3%	12%	2%	72%	2%	23/(no TMs)406
CA1548	740	QDRKNAMDEPTSGWLWESGYCKR/YMWCAYSSVG	314		22%	6%	6%	3%	72%	8%	i12-31o800
CAGLOE05940g	287	ATTMSVLTQRAIDYVRNLEYVKR/AENILYGNQN	504		21%	5%	7%	3%	72%	8%	24/o169-192i204-229o317-340i373-393o399-421i462-483o489-506i518-538o550-579i840
YGR141W	267	MSAHGGGSAYKFEAIEKIKRLRR/VDGKLTNEVI	541		20%	1%	8%	1%	72%	2%	i15-37o467
YNL327W	292	SISNTVELTYPDEGTSVRL LGKR/DTSTTLASEL	74		19%	6%	9%	0%	72%	5%	21/(no TMs)(1019-GPI)o1041
YNR059W	279	PQELWFLDVKDMLNDEYLARFKR/FSNKWLAITF	933		22%	3%	6%	1%	72%	3%	31/(no TMs)580
CA1548	755	WESGYCKRYMWCAYSSVGGPQKR/PDEKDDTVDN	209		17%	4%	11%	4%	72%	2%	i12-31o800
CAGLOL04466g	657	HSNRFABILNETQKEKVQQFRR/QINVTVLDCK	195		20%	2%	8%	2%	72%	3%	o678-694i701-718o783
CA0337	65	EVANYLQQLVVAAGLDSQNLTKR/DYPDNSLLTA	82		12%	3%	15%	4%	72%	6%	19/(no TMs)(264-GPI)o287
YBR093C	174	LAQYGYMVENQTSFAVFTSN SKR/CHDTAQYFID	369		15%	6%	13%	3%	72%	4%	18/(no TMs)467
CAGLOH09262g	92	PFPSLENMLPSWMLSTHTAREKR/VLLNWDKSSG	584		22%	2%	5%	3%	72%	5%	35/(no TMs)625

YIR007W	373	PEWKPDTCIWKLHGVEIQNGKR/PVLLKPNYFS	323		21%	6%	7%	3%	72%	9%	(no TMs)764
CAGLOA01265g	47	KLNKIKVFEKRAVDFTFPAHYPR/PEGCSSPPNS	73		17%	3%	10%	4%	72%	1%	21/(no TMs)727
YKL164C	62	DYSSTFGIAVEPIATASSAKR/AAAISQIGDG	434		22%	4%	6%	4%	72%	8%	19/(no TMs)341
YOL031C	310	LISKILKADMYENDTNLILFKR/NAENWSSNLQ	105		15%	2%	13%	2%	72%	3%	19/(no TMs)421
CA3182	408	ALTWIINRFNGKQTVSGCQHVKR/TSNFEPNIP	626		17%	3%	11%	3%	73%	5%	15/(no TMs)459
CA0883	29	*NKNIKR/EDCDKTSFHG	397	1	19%	8%	8%	2%	73%	11%	23/(no TMs)418
CA3899	16	*TQNKYSVILPTYNEKR/NLPILYLLN	286		14%	3%	13%	2%	73%	3%	(no TMs)239
YBR265W	60	VSRSEARLLDTCNEIRIEAHLRR/ETTDEGQVQH	395		14%	2%	13%	1%	73%	1%	o285-307i320
CA4913	507	LPNVSNAFNLTPSIFNFIHPGKR/PLSSTAPTII	132		10%	1%	17%	4%	73%	4%	i21-39o636
CAGLOE02915g	93	VGEGGDINAEPTTTTITSIIQVQR/VSLANEQPTA	351		21%	9%	6%	4%	73%	12%	21/(no TMs)498
CA1394	397	HKQSSGSTHHNYCNHSHKQFKR/SLMDDGSAAQ	382	1	17%	6%	10%	2%	73%	4%	21/(no TMs)418
CAGLOL12804g	166	GFIIPKSASGDAALKQLEAAVQR/AQTDKKNRFR	214		21%	6%	6%	4%	73%	10%	(no TMs)393
YHR003C	397	NKEISLTVVLMTKEEQEIHEKR/ILLDGEKLTA	426		20%	6%	7%	4%	73%	10%	24/(no TMs)429
CA4910	16	*IEPITNSFQSFDFNKR/DDGLDNCVSVY	609		13%	2%	14%	5%	73%	3%	o32-50i70-91o111-130i150-169o175-197i209-232o244-263i309
CAGLOC03487g	291	MRTAALPSFYKLYMKNDNLDLPR/GTYEISIGMN	42		19%	4%	7%	2%	74%	5%	i43-66o327-353i388
CAGLOG04279g	566	FYECESDKLQEIIESEVGIKR/SLPSTDVNVV	158		12%	3%	14%	4%	74%	2%	i27-44o660
CAGLOA03806g	430	LEFLRPDSKVDPRDPIITYVNIKR/FDTDAI LCLK	317		14%	3%	12%	2%	74%	2%	o20-42i51-69o470
YNL283C	287	TRVVTQSVVSQANQASTIFTTR/TSVIATVSSST	57		20%	5%	6%	4%	74%	9%	24/o322-347i503
YHR076W	60	IDSETNFSHIIGSCSQIISFSKR/TFYSSAKSGY	362		20%	8%	6%	4%	74%	12%	i17-39o374
CA4679	468	VLWNDITSIFKRDMAFRDEIVKR/SNGTITSLPV	240	0	18%	9%	8%	3%	74%	11%	20/(no TMs)(535-GPI)o557
YJL160C	61	TEYRTSFGLAVIPFTVTESVKR/NVISQINDGQ	617		19%	6%	6%	3%	74%	8%	24/(no TMs)180
CA2288	381	MESHGYTRKQDSNIDTVRNGKR/FFDSSSEDQN	323		20%	4%	6%	4%	74%	8%	i16-36o872
CAGLOI07777g	122	KPPTLDANTTVTDTPTLFRVQR/AIFGENTWTI	857		19%	2%	6%	2%	75%	4%	16/(no TMs)240
YFL051C	33	*EACLPTNKR/EDGMNINFYE	173		12%	3%	13%	4%	75%	5%	24/(no TMs)160
CA1583	211	LNPTDEASETASTLGSVSHKIKR/SRMRIKRSLE	187		20%	3%	5%	3%	75%	2%	25/(no TMs)(244-GPI)o260
YKL201C	1039	DKDPIIVYEDYAYAKLLEERKR/EKKEEEEEK	90		19%	11%	6%	3%	75%	9%	i28-45o1178
CAGLOJ08481g	237	KIVLSPDGTRDPRLLDDSLINGKR/STFQVIDIPE	323		20%	5%	5%	3%	75%	7%	29/(no TMs)(?GPI)o583
CA3460	50	IVMSTCTTFIGKYGTCTSTGKR/STNWNVCYCKT	112		15%	6%	10%	2%	75%	5%	19/o212-233i288-306o326-347i368-387o402-424i431-449o455-473i760
YFL038C	128	DRYATSTVLKLLVGNKCDLKDKR/VVEYDVAKEF	229		19%	3%	5%	3%	75%	6%	(noTMs)206
CA5713	918	LSESTDTTSSIGTDSSTVTFCRR/DNGDGCIVTG	256		18%	3%	6%	1%	75%	2%	19/(no TMs)(1345-GPI)o1366
CA3720	128	DDDDNCDDDCKKKKKVYFAKR/GDDDDDKCD	565		17%	3%	8%	1%	75%	4%	25/(no TMs)274
CAGLOE04664g	30	*LSLFASTRDNGKR/VSLGEITVDE	323		19%	2%	5%	3%	76%	3%	17/(no TMs)200
YHR101C	103	GLRKLDFLEFETEFVSLQRYIRR/SSTAIFKEKV	124		16%	3%	8%	2%	76%	4%	20/o276-295i335
YKL220C	55	LVTNACTRIFQKVWTEYTSKSKR/SSPVCSEYEP	434		11%	2%	13%	4%	76%	3%	21/o165-184i236-256o276-297i318-339o351-370i377-395o401-419i711
CA2024	28	*TNGKR/GNVVVVKVKT	323		15%	2%	8%	2%	77%	2%	23/(no TMs)250
CA0046	395	SDALQYILTPTWSPGLATFFKR/MDDAGAYQRL	874		16%	6%	7%	1%	77%	7%	24/(no TMs)415
CAGLOE01815g	492	PINKTNGSSNRTLNSRRFDLQKR/VYSNQASLKN	60	0	19%	10%	4%	2%	77%	8%	16/(no TMs)(503-GPI)o519
CA2974	65	ALANLQHYNKRDLMSQEEI IKR/ENQIVTDVLT	181	1	18%	10%	5%	1%	77%	10%	21/(no TMs)328

Anhang

YLR299W	653	SVVNAIRNVRGEWHAVSDYWRKR/GISSVY*...	181		15%	3%	8%	3%	77%	3%	i12-33o660
YLR155C	308	ATGSIVNEQLYEEYGPIVHSRR/TADGTVPPDD	281		12%	2%	10%	1%	78%	1%	19/(no TMs)362
YLR157C	308	ATGSIVNEQLYEEYGPIVHSRR/TADGTVPPDD	281		12%	2%	10%	1%	78%	1%	19/(no TMs)362
YLR158C	308	ATGSIVNEQLYEEYGPIVHSRR/TADGTVPPDD	281		12%	2%	10%	1%	78%	1%	19/(no TMs)362
YLR160C	308	ATGSIVNEQLYEEYGPIVHSRR/TADGTVPPDD	281		12%	2%	10%	1%	78%	1%	19/(no TMs)362
CAGL0G07667g	132	SPAQDSNLSDFAKSMLNEHNIKR/ALHQDTNPLT	254	0	14%	5%	8%	2%	78%	7%	18/(no TMs)258
CAGL0M04191g	56	YYGETFETAKR/GRSQADIRVNKR/ANGYEEVQIT	364	0	16%	7%	6%	0%	78%	7%	19/(no TMs)(580-GPI)o601
CAGL0K08976g	670	LEISHKGQIRFDTIKGLLMWKR/LGKECTAWKH	220		14%	5%	7%	1%	79%	5%	i123-141o793
CAGL0E01837g	492	RNSSLNSTFNSSSNYSRVQMKKR/TYPNSSSSLQ	300	0	14%	6%	7%	1%	79%	7%	17/(no TMs)(498-GPI)o521
CA0375	76	ESLKKQGTIDSGSNGLSLTMKKR/FDNPSFKSNF	374	0	11%	4%	10%	2%	79%	6%	22/(no TMs)(430-GPI)o453
CA5147	94	HGERFPSKGDGKYFNSVMEVFKR/YGEFHGDLSF	653	0	15%	6%	6%	1%	79%	6%	(no TMs)456
CAGL0H08910g	157	AAINDINNYHCFLARRKKYVKR/NYSFFLPGSF	403	0	15%	6%	5%	0%	79%	6%	19/(no TMs)346
CA2055	22	*VKR/STGFVTLDFN	171	?	12%	3%	8%	3%	80%	4%	19/(no TMs)417
CAGL0E01749g	474	TGDVFTPEATMSMSNNSVALRKR/TTLKPLR*..	51	0	16%	7%	4%	1%	80%	6%	16/(no TMs)482
YLR121C	474	TRSTTTKKTNSTTTAKSTHKSQR/ALQRAATNSA	434	0	15%	6%	4%	2%	80%	6%	21/(no TMs)(483-GPI)o508
CAGL0D02530g	190	GGIYQTNNTAWFTLDIPISQFKR/PIRIQSTPSV	191	0	15%	6%	5%	1%	81%	5%	23/(no TMs)(656-GPI)o676

Tabelle 44: Bewertung putativer Kex2-Schnittstellen, Sortierung alphabetisch.

	Pos	Schnittstelle	P S S M - S c o r e	E x p e r i m e n t e l	ProProtein		ProHormon		kein Substrat		Topologie
					p	+/-	p	+/-	p	+/-	
CA0046	395	SDALQYILTDPTWSPGLATFFKR/MDDAGAYQRL	874		16%	6%	7%	1%	77%	7%	24/(no TMs)415
CA0229	193	GTNDDSQLSSFSRQILEAHNIKR/ASHGVNPLTW	254		23%	4%	11%	2%	66%	3%	24/(no TMs)317
CA0267	36	*GNDAEYLLNTVLKR/DATSLSRNAY	593		27%	4%	10%	1%	63%	4%	22/(no TMs)(1091-GPI)o1114
CA0267	186	YTDINSESTITGDYTFIKNKTR/TVIITDYVTS	61		22%	6%	7%	4%	71%	10%	22/(no TMs)(1091-GPI)o1114
CA0337	65	EVANYLQQLVVAAGLDSQNLTKR/DYPDNSLLTA	82		12%	3%	15%	4%	72%	6%	19/(no TMs)(264-GPI)o287
CA0338	134	NEVKVLPQAISSVQQNAGNTKR/DTTTELIQAI	444		39%	14%	23%	8%	38%	18%	19/(no TMs)(?-GPI)o183
CA0339	165	NQVKVLPQAISEFEQNSQNSKR/IGLDSIAQ*.	295		52%	19%	11%	2%	37%	19%	19/(no TMs)174
CA0340	47	ETKKDVVAANADGTVGFNLVGKR/EVSEGEISKR	462		22%	5%	9%	1%	69%	5%	22/(no TMs)195
CA0340	146	NLPQAITTVQQQQGQQSQVKR/DTIGDIVNEI	772		37%	13%	15%	6%	48%	17%	
CA0349	292	TDILKSSYLSVSVQKFKVFIRR/SVLAMPDQQR	159		21%	6%	8%	4%	72%	8%	23/o373-396i420-440o446-472i510-530o536-555i567-588o600-627i811
CA0365	31	*IPQESTTEVDKR/LDADVAAQLA	178	2	21%	20%	50%	28%	29%	13%	19/(no TMs)(145-GPI)o178
CA0365	62	LALNILNLIQLGIGANVNVTAQR/DESAVTIEQS	160	2	16%	10%	49%	28%	34%	24%	
CA0365	74	IGANVNVTAQRDESAVTIEQSKR/LDADVAAQLA	309	2	12%	6%	53%	24%	35%	20%	
CA0365	105	LAGNILNLIQLGAGAEISASTKR/DESAVDVTEE	193	2	25%	19%	51%	27%	24%	10%	
CA0365	149	IAANILNLIQLGIGADVNVSAKR/EDTVAAQIGA	160	2	10%	5%	47%	26%	43%	22%	
CA0371	32	*TAGASSRRRRR/GSLPNEKTIY	288		25%	6%	4%	3%	70%	4%	21/(no TMs)806
CA0371	554	GKGKNIHEIFGQDLVIPTKSRR/GPGGYKVKTS	221		30%	16%	19%	8%	52%	8%	
CA0375	76	ESLKKQGTIDSGSNGLSLTMKKR/FDNPSFKSNF	374	0	11%	4%	10%	2%	79%	6%	22/(no TMs)(430-GPI)o453
CA0380	84	NVFNVDQADGQVDGQVLKVFQKR/DANTEQTIEG	438		21%	7%	14%	5%	65%	3%	18/(no TMs)(431-GPI)o462
CA0397	27	*PRKYKR/YGQLKSFYAC	577		32%	2%	16%	5%	52%	6%	21/o158-181i318-340o346-368i375-393o399-417i710
CA0403	391	GKWGLGEQMSALEVMQNLVHVKR/PAPYTADTGG	528		23%	6%	14%	5%	63%	1%	19/(no TMs)(425-GPI)o452
CA0416	67	KQRELDGSSDIELKPKFRISKR/YSNKSIAITL	169		23%	2%	8%	1%	69%	2%	25/(no TMs)207
CA0549	40	*SNPKVISFDFELRFSDPIKR/DTNFGTGTAA	130		45%	10%	17%	5%	38%	13%	20/(no TMs)364
CA0822	37	*ISNPFKPNGNLKFKR/GGHNVAWDY	671	1	58%	29%	7%	3%	35%	26%	22/(no TMs)438
CA0825	37	*PTNLYTPGYVQCPEGKLTR/SSLDGINSNE	79		17%	4%	19%	5%	64%	3%	18/(no TMs)609
CA0825	299	LRSFVDTKYIGTRLDDGAPVSKR/CVNGFDNAGF	338		23%	5%	18%	8%	59%	7%	

Anhang

CA0854	800	CSGYMWCAYDKIGGKTKSGKNTR/LEGKINFDE	76		24%	9%	6%	2%	70%	9%	i14-32o828
CA0855	605	NIYGLIQERAIYEALLEIHPNKR/PFIIGRSSFA	130		24%	7%	6%	3%	70%	9%	21/(no TMs)946
CA0856	190	SGSDCEDDECKRKKKVHRNYKR/GGYGSERRQS	392		26%	6%	8%	2%	66%	8%	o28-45i57-78o356
CA0856	198	ECKKRKKKVHRNYKRGGYGSERR/QSDCESDCER	63		52%	13%	17%	2%	31%	12%	o28-45i57-78o356
CA0856	211	KRGGYGSERRQSDCESDCERSER/CNYPFCELYE	54		36%	11%	28%	17%	35%	23%	
CA0875	304	ASSFTLWDWLFQTECGTYARKRR/ERNSRASAES	180		29%	1%	7%	2%	63%	2%	(no TMs)321
CA0883	29	*NKNIKR/EDCDKTSFHG	397	1	19%	8%	8%	2%	73%	11%	23/(no TMs)418
CA0883	45	*NKNIKREDCDKTSFHGHKHKR/AVAYDYAVT	678		55%	32%	7%	3%	38%	30%	23/(no TMs)418
CA0894	61	TPTPIAEEKLSTPIPTPSPSHRR/GSSIGSSTSP	82		38%	22%	5%	3%	56%	21%	o142-162i219-235o241-259i268
CA0894	105	SANALTPTPSHHQPAHSITRQRR/SSSLIQHLEP	341		53%	16%	7%	2%	40%	16%	o142-162i219-235o241-259i268
CA0915	308	KQFKKHNIDITANAKAISKLR/EAERAKRTLS	954		50%	30%	7%	1%	43%	30%	35/(no TMs)687
CA0932	104	FFYAFGLFMYQTFDGCDCGHARR/TGQSGPLGEL	224		31%	4%	16%	6%	53%	8%	i50-69o138-158i179-200o220-240i274-292o298-322i334-353o359-379i402
CA0951	187	IFKIFYQAKPSVSQLNTYPSKRR/IYHARFDSL	96		43%	13%	8%	3%	49%	13%	i18-38o660
CA0968	35	*GFVTLDFNVKR/SLVDPDDPTV	420	1	56%	31%	5%	1%	39%	31%	24/(no TMs)418
CA0968	75	FLDLLPTKIPVDDTGRNDGVDKR/GPVAVKLDNE	142	1	76%	13%	6%	1%	17%	11%	24/(no TMs)418
CA1032	531	DYAVSVLAKDFGSSDDHDKYLKR/SANWQNIWNF	398		24%	4%	13%	6%	63%	4%	21/(no TMs)795
CA1039	278	TGGSPPTYFISSTKQVLVQKHKR/SLDYVAGVAK	848		28%	3%	21%	12%	51%	10%	24/o170-190i202-220o327-351i372-394o406-426i466-485o491-510i522-541o553-579i808
CA1159	221	KRCSWYTASISPDDKVNSSYNKR/PENINCKYIK	382		17%	4%	16%	7%	67%	6%	o270-290i434
CA1166	31	*ISTTKFLKR/GSLPSSSSFD	509		23%	4%	9%	4%	68%	7%	22/(no TMs)412
CA1166	392	SINYQFNGDDYCDGNSDSGQIKR/SLMENASAAQ	212		25%	3%	9%	3%	66%	3%	22/(no TMs)412
CA1266	32	*EDIDKR/AEKVVSLDFT	45	1	74%	12%	6%	2%	20%	11%	26/(no TMs)405
CA1266	74	AHGQHHQSQQQQQQQPAQKR/GTVQTSLINE	167	1	78%	10%	6%	1%	16%	9%	26/(no TMs)405
CA1313	155	NDKSDATFKSLSSSAPTUVHSRR/SWIKDFLIFK	224		22%	6%	20%	7%	59%	4%	22/(no TMs)332
CA1334	116	YRNENSEECSDEDDHHKHKKR/PHRHGKSD	203	1	55%	32%	6%	2%	39%	31%	19/(no TMs)201
CA1334	135	KKKRPHRHGKSDDDDDKKWKR/GGDYSDDNDN	746		43%	9%	8%	3%	49%	7%	19/(no TMs)201
CA1345	39	*QENEITPADNIDKR/AGAIGNFFRD	71		28%	4%	20%	7%	52%	6%	25/o1297-1319i1409
CA1360	222	DRAELASTKNNLLVRSKSNRR/LIGSSRKKTL	162		33%	9%	5%	3%	62%	6%	17/o712-733i734
CA1360	254	LKSKSKSKSLKHKSRKSHKRR/PKLLSKDVE	105		37%	8%	6%	3%	57%	5%	17/o712-733i734
CA1360	470	RYSNHHSFTVERPPRSRTKRR/PRIKAKTMK	48		55%	12%	5%	3%	40%	9%	17/o712-733i734
CA1360	690	HSRIRFPTADDNNNAGNNYHRR/FTGVVLPENR	226		29%	3%	6%	3%	66%	3%	17/o712-733i734
CA1371	141	GLIGLSKITHLINGINGVQKR/GINRDSFINK	375		18%	4%	24%	10%	58%	8%	i103-125o375
CA1371	248	DDTDIQQIKNDDDKFHSKSKR/NLSSLSEEEE	434		29%	5%	6%	3%	66%	4%	i103-125o375
CA1379	38	*VENNQDLEATFTRIKR/SLVHQDIFEI	518		34%	17%	14%	6%	52%	20%	22/o334-355i784-804o816-839i860-884o896-921i928-946o1012-1035i1038
CA1379	123	DGFGGNDCEPLCGSLGDGNSKR/PVRKKKQCAC	148		26%	19%	10%	6%	65%	25%	22/o334-355i784-804o816-839i860-884o896-921i928-946o1012-1035i1038
CA1387	156	PNSRLMEQKFKAFALTDSFKR/YVPKIREEIL	159		29%	9%	9%	1%	62%	9%	o13-36i48-65o528
CA1387	264	YWRRDAAQKKISATYMKIERSR/ERGDIDPNRD	138		27%	2%	14%	3%	59%	1%	o13-36i48-65o528
CA1387	266	RRDAAQKKISATYMKIERSR/ERGDIDPNRDLI	64		31%	8%	25%	10%	44%	15%	o13-36i48-65o528

Anhang

CA1394	31	*IQKDSLGLKR/TLSSTFDPH	25	0	54%	33%	6%	2%	39%	32%	21/(no TMs)418
CA1394	397	HKQSSGSTHHNYCENNHSKQFKR/SLMDDGSAAQ	382	1	17%	6%	10%	2%	73%	4%	21/(no TMs)418
CA1402	30	*IHHAPEFNMKR/DVAPAAPAAP	410	2	14%	7%	47%	30%	39%	24%	19/(no TMs)271
CA1402	60	AAPADQAPTVPAPQEFNTAITKR/SIIGIIMGIL	203	2	25%	22%	47%	31%	28%	23%	19/(no TMs)271
CA1402	125	IADMPFVVRAVDTAMTSVASTKR/DGANDDVANA	154	2	16%	9%	53%	25%	32%	19%	19/(no TMs)271
CA1402	159	VRLPEIVARVATGVQOSIENAKR/DGVDPDVLNL	369	2	19%	13%	49%	29%	31%	22%	19/(no TMs)271
CA1402	193	ANAPRLISNVFDGVSETVQQAQR/DGLEDFLDEL	309	2	19%	12%	54%	23%	27%	12%	19/(no TMs)271
CA1402	227	QRLPQLITRSAESALKDSQPVKR/DAGSVALSNL	214	2	25%	22%	52%	25%	23%	10%	19/(no TMs)271
CA1507	13	*SAAVSTLSDI IKR/NDAVNVNPPN	181		24%	3%	11%	3%	65%	5%	o38-59i71-89o117-139i159-179o185-203i215-233o253-275i342
CA1548	740	QDRKNAMDEPTSGWLWESGYCKR/YMWCAYSSVG	314		22%	6%	6%	3%	72%	8%	i12-31o800
CA1548	755	WESGYCKRYMWCAYSSVGGPQKR/PDEKDDTVDN	209		17%	4%	11%	4%	72%	2%	i12-31o800
CA1583	211	LNPTDEASETASTLGSVSHKIKR/SRMRIKRSLE	187		20%	3%	5%	3%	75%	2%	25/(no TMs)(244-GPI)o260
CA1583	230	KIKRSRMRIKRSLENQYRTNAKR/EIYTEKFDKI	369		34%	6%	8%	2%	58%	6%	25/(no TMs)(244-GPI)o260
CA1621	154	PVFSITAADWHNGLLVLAQAKR/TSALANLGF	247		22%	8%	16%	6%	62%	6%	20/(no TMs)(1063-GPI)o1086
CA1646	33	*LPIVEGLLNCVFPR/PPPPTIPSP	52		24%	4%	10%	1%	66%	4%	19/(no TMs)725
CA1676	31	HPTELKAVIQLFGRQPLHPGKR/DVNDKELVRC	165		26%	5%	15%	7%	59%	3%	o430-447i448
CA1676	419	LPENKFTKIVASRESIDLSVQRR/IEQENFNVCNV	239		23%	4%	7%	3%	71%	6%	o430-447i448
CA1678	132	QLEANSTAAIRRLMGETPIVKKR/DQIDDFIASI	300		15%	7%	25%	11%	60%	5%	22/(no TMs)(499-GPI)o533
CA1820	307	SYYYPPNYEEIFKDKNEPLKSER/STVRDSTSSP	39		61%	13%	9%	1%	30%	13%	i12-31o765
CA1820	607	KFNNHFAAAIGQLTSNQYNKDRR/TPLNSEICIS	93		33%	3%	23%	10%	44%	13%	i12-31o765
CA1835	393	DNEEEESITAIPISKRKRGSTRR/LDISTQLPPL	79		46%	12%	6%	4%	48%	8%	19/(no TMs)(481-GPI)o504
CA1840	35	*VPNVEQLKAQSLQKR/DEDLMEPVFK	118		36%	10%	11%	2%	53%	13%	20/(no TMs)712
CA1849	211	AYFAWGHLSRKKRYTEIMNKR/NVRQVATLES	172		30%	6%	10%	1%	60%	6%	o15-39i51-70o77-96i108-128o148-168i174-194o237
CA1872	116	ERTRTKKSSSIGSAELSPRSRPR/MMNFSLRPKF	60		48%	19%	8%	2%	44%	17%	o895-919i943-968o972
CA1872	159	TSTGSTHPPIPAASVLSSSQRR/ATIEDFADTE	58		55%	5%	5%	2%	40%	6%	o895-919i943-968o972
CA1872	337	TQIAQQYYGDARKHIEVVDPTKR/EKSNVVIATT	91		31%	6%	24%	10%	46%	11%	o895-919i943-968o972
CA1872	358	KREKSNVVIATTVSGDSPTALKR/LSGGNSSSGS	212		42%	9%	6%	3%	52%	11%	o895-919i943-968o972
CA1873	39	*EQQQQESQFASKALTKR/EEQDIQELVQ	102	1	58%	27%	9%	5%	32%	22%	21/(no TMs)351
CA1873	58	LTKREEQDIQELVQHINNYKTRR/DAIDEEIMKR	333	1	56%	30%	12%	7%	32%	24%	21/(no TMs)351
CA1873	68	ELVQHINNYKTRRDAIDEEIMKR/DYAIIVTDVLA	234	1	59%	27%	12%	9%	29%	20%	21/(no TMs)351
CA1873	195	QLITRDETDIDALAPYVVTMEKR/LDLGVDNL	387	1	60%	26%	9%	6%	32%	22%	21/(no TMs)351
CA1939	94	EYGHLLSVHQTYFHNLRINSKR/AGVKFEQDE	148		29%	2%	12%	2%	59%	4%	16/(no TMs)(?-GPI)o358
CA1975	193	GIEHPKQVEEWNVGNQVASKR/NANFNVSLTD	151		23%	2%	14%	4%	64%	5%	18/(no TMs)605
CA1975	552	VGCAIIRREQERGIEQTEQCKR/CFENYCWGDT	343		23%	4%	7%	3%	70%	5%	18/(no TMs)605
CA2007	143	VTVENFMRLLTDKWSDQPRSKR/LLSDENSNIF	603		38%	7%	9%	2%	54%	6%	21/(no TMs)(?-GPI)o383
CA2024	28	*TNGKR/GNVVVVKVKT	323		15%	2%	8%	2%	77%	2%	23/(no TMs)250
CA2055	22	*VKR/STGFVTLDFN	171	?	12%	3%	8%	3%	80%	4%	19/(no TMs)417
CA2055	35	*VKRSTGFVTLDFNVKR/SLVDPKDPTV	420	1	56%	31%	5%	1%	39%	31%	19/(no TMs)417
CA2055	74	LFLDIEPTEIPVDDTGRNDVVKR/GPVAVKLDNE	369	1	76%	12%	6%	1%	18%	11%	19/(no TMs)417
CA2085	407	DSIPERWIYNHQQHHQGNKKVKR/NKKKKAEDS	154		32%	3%	13%	3%	55%	6%	25/(no TMs)842

Anhang

CA2202	22	*IKR/GLLDDDLFGTS	104	1	73%	14%	7%	2%	20%	12%	19/(no TMs)578
CA2223	138	SKKTFIRTASELSQQEKDYIHKR/QETTNKNLID	330	0	24%	17%	7%	2%	69%	18%	20/(no TMs)(720-GPI)o754
CA2234	34	*ISDLNNTASNQVTKR/LGGGSRGGSS	406		25%	3%	7%	3%	68%	6%	19/(no TMs)(185-GPI)o208
CA2247	130	RNQHHRDKLMINREKINDQQSKR/DINDQYEKEY	309		51%	17%	9%	2%	39%	18%	(noTMs)200
CA2288	381	MESHGYTRKQDSNIDTVRNGKR/FFFDSSSEDQN	323		20%	4%	6%	4%	74%	8%	i16-36o872
CA2303	105	FRGPLTLLQFGVYYPSSSGNSKR/QIDDQDCNVK	369		36%	17%	9%	3%	55%	14%	22/(no TMs)468
CA2303	123	GNSKRQIDDQDCNVKHVHHKHKR/ATEVVQVTQT	848		49%	14%	7%	2%	44%	14%	22/(no TMs)468
CA2394	20	*NDNVVILTLRAANHITNAKR/ADSTTSTEAM	369		35%	5%	18%	6%	47%	10%	o66-88i288
CA2405	688	TTYCPLTTTKAVYRVAKLANYKR/DQTEFIATET	490		34%	9%	17%	5%	50%	11%	19/(no TMs)(721-GPI)o753
CA2412	80	DNKPVILIVNGTTLTSGANNEKR/EAKSKGGFRL	179	2	16%	16%	48%	29%	35%	21%	19/(no TMs)143
CA2412	101	KREAKSKGGFRLTNFGYFEPGKR/DANADAGFRL	165	2	17%	14%	49%	28%	34%	21%	19/(no TMs)143
CA2412	122	KRDANADAGFRLTNFGYFEPGKR/DANAEAGFRL	165	2	9%	4%	50%	27%	42%	24%	19/(no TMs)143
CA2418	154	IGQPKDAKLVPFPPEELPNYKR/ANATILSLVR	392		46%	13%	13%	3%	41%	14%	i12-32o480
CA2499	35	*GFVTLDFNVKR/SLVDPDPTV	420	1	56%	31%	5%	1%	39%	31%	24/(no TMs)418
CA2499	75	FLEFTPSEFPVDETGRDGDVDKR/GPVAVTLHNE	142	1	75%	13%	8%	2%	17%	11%	24/(no TMs)418
CA2633	99	DKSHPHNSFLGNHNSNEYNLMKR/QLGHEQDYDE	60		27%	7%	12%	8%	61%	5%	21/o774-794i936
CA2633	128	DYDELISHVESIHLLPMMKLSKR/IPVPIEMEDV	54		36%	7%	9%	6%	55%	11%	21/o774-794i936
CA2633	526	YSDVIQVNQTTITTTPEQKAPSKR/DSPQKI IHSS	151		32%	10%	23%	9%	45%	10%	21/o774-794i936
CA2660	49	VTLDFDVIKTPVNTAGQEGVKR/QALPVTLNNE	617	1	71%	15%	4%	2%	25%	16%	24/(no TMs)391
CA2714	872	IVSQSGSSFTTITTFEPPVCKR/DDVNCGSSIG	585		41%	6%	8%	2%	51%	5%	21/(no TMs)(1225-GPI)o1249
CA2730	43	*CYERKKIHPQPKETPAKR/DHskELIREQ	151		40%	16%	12%	3%	48%	19%	24/(no TMs)434
CA2730	232	LADISKTEEDPLMKKIRVVLKKR/GIINGIPVVF	51		22%	8%	12%	3%	67%	6%	24/(no TMs)434
CA2730	399	SKEISFQNVVMTKDEQRNHEKR/VLNGGEKIED	426		33%	18%	24%	11%	43%	19%	24/(no TMs)434
CA2769	21	*FIKR/YENTTAPASQ	389		20%	3%	14%	4%	66%	2%	17/(no TMs)342
CA2786	1390	DAKMIDLCNNPLTRENKLDAAKR/LIPEWIEYEQ	151		28%	3%	7%	2%	65%	3%	23/(no TMs)1447
CA2825	38	*VPQVDGQGETEELIQKR/SYDYYQEPD	234		27%	1%	11%	1%	63%	2%	20/(no TMs)(604-GPI)o634
CA2830	59	PVCTDLGLYHDSSISLSGSKNKR/EAEIANKDGT	731		43%	13%	7%	3%	50%	10%	18/(no TMs)(684-GPI)o714
CA2830	73	SLSGSKNKREAEIANKDGTIEKR/TFGSAGVNAG	164		50%	10%	7%	3%	42%	8%	18/(no TMs)(684-GPI)o714
CA2830	225	NAIGTFDASALFNAQVKAFAKRAK/ELDEFEELSN	188		18%	5%	21%	8%	61%	4%	18/(no TMs)(684-GPI)o714
CA2830	243	FPAKRELDEFEELSNDGVTHSKR/TLGLLGLLLK	440		26%	9%	13%	5%	60%	4%	18/(no TMs)(684-GPI)o714
CA2912	84	DDGQLERNDKKLDNCCKSERVSR/PAPSPSAIAV	34		27%	8%	18%	7%	55%	11%	19/(no TMs)247
CA2912	119	GNEECDEDCDEHRKKGSKQYKR/GEVENPRETR	410		37%	18%	14%	7%	49%	13%	19/(no TMs)247
CA2912	129	DEHRKKGSKQYKRGEVENPRETR/DCDFCTIEKS	119		62%	13%	12%	8%	27%	6%	19/(no TMs)247
CA2974	54	LTVRESNLVNSALANLQHYNAKR/DLMSQEEI IK	369	1	54%	33%	10%	9%	36%	27%	21/(no TMs)328
CA2974	65	ALANLQHYNAKRDLMSQEEI IKR/ENQIVTDVLT	181	1	18%	10%	5%	1%	77%	10%	21/(no TMs)328
CA2974	180	NLIGNALGLNANTVSSKMASAKR/EIMVESAPEP	128	0	55%	31%	9%	8%	36%	29%	21/(no TMs)328
CA2994	322	ILGAATIFNTLLVINSTDSNNKR/DSYHINLYD	498		51%	12%	14%	3%	35%	14%	o367-390i671
CA3105	279	SQTFDIIVTFPQIQELSHLNKR/NVELSVKSFQ	73		41%	10%	7%	3%	52%	8%	23/o351-371i393
CA3122	41	*IFDLESKQGLQDEETVNNDKR/EPVNLLYLDR	156		54%	5%	13%	8%	33%	7%	19/(no TMs)(?-GPI)o410
CA3122	68	LLYLDRFKMGVSDAEAGNAKFKR/DPKNVIDPAS	420		36%	19%	20%	10%	43%	20%	19/(no TMs)(?-GPI)o410
CA3122	92	PKNVIDPASLKEGSAEEQDKR/EPKNLFLNQA	114		37%	17%	16%	6%	46%	18%	19/(no TMs)(?-GPI)o410

Anhang

CA3122	116	PKNLFNLQALHEGLKDEETKSKR/EAKNLPNLEA	271		42%	7%	16%	10%	42%	14%	19/(no TMs)(?-GPI)o410
CA3122	161	KDAKNLIDLVALKQSLKEAAKR/DAKNIPDLEA	94		55%	12%	9%	3%	36%	15%	19/(no TMs)(?-GPI)o410
CA3122	186	KNIPDLEALKTGIEEEEGQVAKR/DAKNVINLSN	211		44%	17%	16%	7%	40%	16%	19/(no TMs)(?-GPI)o410
CA3122	204	QVAKRDAKNVINLSNFIETPSKR/EGKNLFDLTK	75		29%	6%	10%	8%	61%	3%	19/(no TMs)(?-GPI)o410
CA3122	225	KREGKNLFDLTKFQQSQGQPIKRR/DQKILKQEK	99		37%	11%	9%	2%	54%	11%	19/(no TMs)(?-GPI)o410
CA3138	55	FSVVKTPKAFPVTNGQEGKTSKR/QAVPVLHNE	103	1	76%	12%	4%	1%	20%	12%	19/(no TMs)398
CA3160	15	*NSDSISQVFEYLNKR/DECPDNDYN	91		28%	1%	15%	5%	57%	6%	o33-52i64-85o105-127i212-235o275-295i307-326o346-366i370
CA3182	408	ALTWIINRFNGKQTVSGCQHVKR/TSNFEPNIP	626		17%	3%	11%	3%	73%	5%	15/(no TMs)459
CA3295	318	GGILGSSGYTGSTSKVCTLKRR/SLEKKEIRKL	64		20%	7%	19%	8%	62%	4%	23/(no TMs)(?-GPI)o1249
CA3295	354	SVNKVIARRQQAIAEKKQANAKR/ESISFVSFEA	369		39%	14%	6%	3%	55%	11%	23/(no TMs)(?-GPI)o1249
CA3295	401	VTAVSSVLLTVAALTNANEIDKR/SFFGDLFSG	71		25%	4%	12%	2%	63%	5%	23/(no TMs)(?-GPI)o1249
CA3295	551	APSSEAPSSSSSEALSSSSTTKR/PTAAAKGFFG	123		41%	9%	12%	2%	47%	11%	23/(no TMs)(?-GPI)o1249
CA3321	68	DVESPFFQIDYNNSSPVEASRR/SHYNYFPNSL	77		38%	11%	24%	7%	37%	11%	18/(no TMs)535
CA3321	135	YSGGTGVEVSAVIEAPNVEVAKR/NWLKLTKTL	338		27%	2%	17%	8%	56%	7%	18/(no TMs)535
CA3347	141	VHRIFFDVAQGSANEYVDSKSTR/MVDDLTKGVN	71		25%	10%	6%	3%	70%	9%	23/o186-205i218
CA3354	74	STTTASILSALFNDIILPTIFKRR/DYSTRDHKAN	326		17%	2%	14%	6%	69%	5%	o100-122i134-152o172-199i220-241o261-287i440-460o472-495i823
CA3460	50	IVMSTCTTFIGKYGTCTSTGKR/STNWNVCYCKT	112		15%	6%	10%	2%	75%	5%	19/o212-233i288-306o326-347i368-387o402-424i431-449o455-473i760
CA3518	605	NIYGLIQERAIYEALLEIHPNKR/PFIIGRSSFA	130		24%	7%	7%	3%	70%	9%	21/(no TMs)(?-GPI)o862
CA3518	775	QFFVGDALLVTPVLEPGVNHTRK/YFPQARMSCT	530		15%	3%	15%	6%	70%	5%	21/(no TMs)(?-GPI)o862
CA3529	213	KQFQSSISQEHETLQRDLSSHHR/EFEEESDNKS	107		35%	15%	13%	6%	53%	20%	25/(no TMs)837
CA3529	494	YQASIEFGGFSNALPKPKFNGKR/NLSPTEDHLK	323		28%	13%	10%	6%	61%	19%	25/(no TMs)837
CA3614	121	DNDEDEDCDEDEKKGKQYKR/GEVEEPCETS	410		48%	10%	8%	1%	44%	9%	19/(no TMs)251
CA3629	272	KPTTEPHVYFGTEAFDHDGIEKR/GQIVKIEDHT	164		38%	11%	11%	2%	51%	12%	(no TMs)415
CA3642	56	EPNDIQFFNQGVRSLSYIYQKR/DVSDSDNEQV	393		20%	9%	11%	1%	69%	8%	21/(no TMs)(210-GPI)o229
CA3643	72	FSQVVLSTPNWDQTLTKYKRKKR/DYNDDSLASM	707		26%	2%	13%	3%	61%	3%	23/(no TMs)(?-GPI)o229
CA3713	33	*ANIVSEQTMVKKR/EDVNAIVELI	710		27%	6%	14%	6%	58%	5%	21/(no TMs)403
CA3713	53	VKREDVNAIVELINEIKHINQKR/DLAEGEDLLE	410		22%	6%	11%	2%	67%	6%	21/(no TMs)403
CA3713	191	SSSSSSAAPAAAAAPATNGASKR/EIMEAAEYLS	188		26%	6%	16%	9%	58%	12%	21/(no TMs)403
CA3720	128	DDDDDDCDDCKKKKKVYFAKR/GDDDDDKCD	565		17%	3%	8%	1%	75%	4%	25/(no TMs)274
CA3758	344	IAEITATNGAQIAYEVLTSMVKR/GNIDKSLVKS	454		24%	4%	14%	5%	62%	3%	16/(no TMs)361
CA3805	63	RLIKLGPENYRIVSENEKLLKRR/NKINFIDVTN	305		25%	4%	10%	1%	65%	5%	19/(no TMs)417
CA3825	223	TNLTDSMSEENKIVDLRIPMKR/WGNPDDFMGP	134		17%	9%	14%	3%	70%	6%	(no TMs)260
CA3867	73	QQNNLDSNESFVDPLANPEHCKR/DIPYLEAVDT	610		28%	5%	18%	9%	54%	7%	23/(no TMs)(510-GPI)o544
CA3899	16	*TQNKYSVILPTYNEKR/NLPILYLLN	286		14%	3%	13%	2%	73%	3%	(no TMs)239
CA3925	18	*PIPRSFIFEPELEQLLKR/DENGCVSCPS	119		26%	3%	12%	2%	62%	5%	o69-94i527
CA3931	223	RPIVCEKLRKIGKLNLSDKRSTR/SEIHEKFGDT	98		18%	4%	17%	5%	66%	4%	(no TMs)330
CA3952	160	GSTPASLTFQKKKKTISQYLNKR/QIELKSDTIN	73		22%	5%	6%	3%	71%	6%	i21-44o600



CA4176	76	YLKEEDSLPEEQELLQLLHKVKR/NENNTYWLAN	772		22%	3%	13%	6%	65%	6%	i7-26o808
CA4176	193	SEIMKPIGLKNNCEWLQKEANKR/TRFPDFCKDL	81		24%	5%	12%	4%	64%	3%	i7-26o808
CA4251	183	DWLCKRRNALLKKQVWHDSVQKR/VIPAQYFRDP	375		23%	1%	11%	3%	66%	4%	(no TMs)316
CA4295	188	KDTDSIPDLENYLTKANDKYFKR/QEYWRYLLKD	548		20%	2%	18%	11%	62%	10%	i12-29o756
CA4295	642	HPKIDPLKNSKEADAIEFGVYSRR/NMGLPKVEK	145		17%	6%	14%	4%	69%	3%	i12-29o756
CA4336	16	*VFGFGKDDDKDKNDRR/DNDSYGSSNR	99		49%	21%	16%	6%	35%	16%	(no TMs)212
CA4336	27	GKDDDKDKNDRRDNDSDYSSNRR/NNDDSYGSSS	70		49%	18%	17%	3%	34%	16%	(no TMs)212
CA4336	54	SYGSSSFGSSNDDSDYSSNKR/SNDSYGSSNT	138		53%	13%	7%	3%	40%	15%	(no TMs)212
CA4336	73	SNKRSNDSYGSSNTDSYGSSNRR/DNDSYGSSNN	88		53%	14%	13%	3%	34%	12%	(no TMs)212
CA4336	154	YGSSNTDSYGSSNTDSYGSSNKR/GNDSYGSSND	138		57%	12%	9%	3%	34%	13%	(no TMs)212
CA4336	181	YGSSNDDSYGSSNDDSDYSSNRR/GNDSYGSSND	70		58%	8%	15%	6%	28%	5%	(no TMs)212
CA4336	200	SNRRGNDSYGSSNDDSDYSSNRR/GNSDSYGSSD	56		54%	13%	9%	3%	37%	15%	(no TMs)212
CA4418	445	VYDEPDSFIPERWENASGDMYKR/NWLVTGTGPH	424		31%	7%	10%	5%	59%	4%	(no TMs)517
CA4418	510	SEEIKVFATIFPKDDLILEWKKR/DPLKSL* . . .	226		31%	12%	12%	10%	57%	7%	(no TMs)517
CA4624	17	*SRNVSLPSTTESTKNKR/VVSSQLRLAQ	468		47%	10%	6%	3%	47%	7%	o723-74i757
CA4624	384	HNGVPSNEGSSSDSFHKFRITR/SKSPHKRKIR	34		27%	5%	6%	3%	66%	8%	o723-74i757
CA4679	338	SGALRKREYNDAVEAALREIQKR/EEGIDDVEIA	293	0	58%	29%	14%	11%	27%	22%	20/(no TMs)(535-GPI)o557
CA4679	354	LREIQKREEGIDDVEIALRKMKR/DNIEDLLTTI	602	1	17%	9%	14%	14%	69%	21%	20/(no TMs)(535-GPI)o557
CA4679	468	VLWNDITSIFKRDMAFRDEIVKR/SNGTITSLPV	240	0	18%	9%	8%	3%	74%	11%	20/(no TMs)(535-GPI)o557
CA4700	49	VRRGESKDDLSPEDDSNPRFVKR/DGSLDMLTN	643	1	23%	30%	10%	6%	67%	27%	18/(no TMs)(520-GPI)o544
CA4700	105	TGSSDLVWMSHDLKCVSAPNSKR/NERSFGHGTG	185	0	66%	31%	8%	7%	25%	24%	18/(no TMs)(520-GPI)o544
CA4759	294	TTFCPYDTCYDYLHHTENNRNTR/VDQGGLEGNT	106		27%	1%	6%	3%	67%	4%	20/(no TMs)575
CA4846	22	*SLIDTNSVNGIITVREESMGKR/WESLSPIIWQ	437		20%	5%	14%	3%	66%	3%	(no TMs)232
CA4863	37	*KLDFNKVSTPSKYTKR/DALPMLIND	426	1	72%	14%	10%	4%	17%	10%	21/(no TMs)(428-GPI)o453
CA4882	30	*VVVVEKR/AINETALAED	327		16%	3%	12%	2%	72%	2%	23/(no TMs)406
CA4910	16	*IEPITNSFQSFDFNKR/DDGLDNCVY	609		13%	2%	14%	5%	73%	3%	o32-50i70-91o111-130i150-169o175-197i209-232o244-263i309
CA4913	507	LPNVSNANLTPSIFNFIHPGKR/PLSSTAPTII	132		10%	1%	17%	4%	73%	4%	i21-39o636
CA4913	629	AMNGIKRVKGGKHLHGVSDDWRKR/GESDGY* . . .	226		27%	2%	12%	2%	61%	1%	i21-39o636
CA5115	242	RNGNNSNGNASNSTLRGGNIKR/SSIKQLGLKF	254		31%	7%	17%	6%	52%	10%	18/o310-33i343-362o374-393i414-440o465
CA5147	94	HGERFPSKGDGKYFNSVMEVFKR/YGEFHGDLSE	653	0	15%	6%	6%	1%	79%	6%	(no TMs)456
CA5160	291	STDNIPPIPEGQTLTKDYKGRR/YDEAGNLVVL	378		23%	10%	10%	1%	66%	10%	o7-24i44-64o85-104i146-166o186-205i225-244o476-495i566-590o600
CA5232	30	*LQHQHHEHHEHARR/AEVTKVYVYV	281		34%	3%	6%	3%	60%	1%	16/(no TMs)372
CA5344	134	SDATPTADVEFAEEILKEHNVKR/ALHGVPAISW	420		23%	4%	8%	2%	69%	6%	20/(no TMs)271
CA5363	100	GKDFDYICCSPEQIDILESNLKR/VDPLISSCPA	519		24%	3%	12%	3%	64%	1%	29/o272-293i355-375o587-606i618-642o648-670i700-719o725-750i810-832o1068-1091i1098-1118o1124-1145i1196-1222o1234-1254i1256
CA5418	69	TSRSKPKETGTHRYGKFNKTPR/PVTTTTVLVKE	52		33%	10%	9%	2%	58%	8%	19/(no TMs)(105-GPI)o131
CA5418	86	FNKTPRPVTTTTVLVKESDLPKKR/DAVVARDSKN	141		28%	3%	9%	3%	64%	6%	19/(no TMs)(105-GPI)o131
CA5423	181	YAGKEQLRINDKQFLNIEHRTR/DENDNMLPQ	115		23%	8%	21%	10%	57%	8%	17/(no TMs)871

Anhang

CA5423	544	NVFGLSYHETTFNSLLNRSPEKR/PFILTRSYFA	117		25%	5%	6%	3%	69%	6%	17/(no TMs)871
CA5423	766	QLSDIPMLLKGGSIIPMKTRYRR/SSKLMKSDPY	204		20%	7%	14%	6%	66%	8%	17/(no TMs)871
CA5537	22	*SYNNPNNSNSHLRPHAYNNSRR/DSDGDDESSI	235		46%	10%	18%	5%	37%	11%	o647-666i678-702o722-741i753-777o804-823i930-950o970-994i1009
CA5572	59	TSSINHKNQPLSLRNNPQLIQKR/SRHAESPLFI	94		37%	4%	8%	1%	55%	4%	21/(no TMs)924
CA5614	56	VTVSTYFVNSYTCNQPHHNHSTR/PSHYLPINGT	57		32%	3%	15%	6%	53%	9%	i20-42o270-293i337
CA5699	1268	ESSSGTILSEESSDSIPTTFSTR/YWSPSGMSSR	47		28%	4%	9%	5%	62%	4%	19/(no TMs)(1980-GPI)o2000
CA5699	1405	ESSIGTILSEESSDSIPTTFSTR/YWSPSGMSSR	47		42%	4%	6%	3%	52%	7%	19/(no TMs)(1980-GPI)o2000
CA5699	1537	ESSSGTILSEENSISPTTFSTR/YVSVSLTVGE	47		36%	2%	5%	3%	59%	4%	19/(no TMs)(1980-GPI)o2000
CA5699	1915	KSVQAESVQTDSTTYSEMSSKR/NKNSGFGTSS	51		39%	6%	8%	3%	53%	4%	19/(no TMs)(1980-GPI)o2000
CA5713	918	LSESTDTTSSIGTDSSTVTFCCR/DNGDGCIVTG	256		18%	3%	6%	1%	75%	2%	19/(no TMs)(1345-GPI)o1366
CA5814	67	NDSYFVSLKASESIDHFMEYDKR/YPKHLQVRDF	60		21%	3%	15%	10%	64%	11%	i21-40o421
CA5814	104	IAIGDFKGFSGRFSKDIIDRLKR/CPLVQEITED	848		25%	4%	18%	6%	58%	5%	i21-40o421
CA5814	140	NALEFEIQEDAPRHLARISRRR/MKPNKPYPYM	144		24%	4%	11%	2%	65%	3%	i21-40o421
CA5815	116	KAYDAVFQDSAPRHLARISRRR/MKPIKKYSYI	283		20%	3%	11%	1%	69%	3%	22/(no TMs)379
CA5819	1208	SPHTSES DITSQYNHSTSVATR/SDSISLTDTF	55		41%	4%	6%	3%	54%	6%	21/(no TMs)(1502-GPI)o1526
CA5859	434	QTVYVTLTTTVGSQMLQSLFDKR/DVIAEAKSTN	191		26%	4%	12%	4%	62%	8%	23/(no TMs)(?-GPI)o462
CA5928	87	KTETYSVEVINATTLSENTRKKR/FDVSGPVRTP	707		32%	9%	5%	3%	63%	6%	26/(no TMs)232
CA5996	223	NWPIFEFEKGFTEDDELWKNKR/ETGQHRKYRA	163		21%	9%	28%	4%	51%	11%	27/(no TMs)327
CA6065	57	VIKTHKNVTGPQGEINTNVNKR/QTVPVKLINE	420	1	66%	20%	4%	2%	30%	21%	19/(no TMs)398
CA6162	200	VRETKDLADQSPEDSTSIDVFKR/ENSLLVLLFT	653		45%	11%	10%	1%	45%	12%	21/(no TMs)296
CAGLOA00253g	50	LQSLFFKNSPISVKFETLVKERR/SIQEENKISIS	214		27%	1%	7%	3%	66%	4%	20/(no TMs)223
CAGLOA01265g	35	*VEDKLNKVKFEKR/AVDFTPPFAHY	438		22%	5%	12%	5%	66%	5%	21/(no TMs)727
CAGLOA01265g	47	KLNIKVKFEKRAVDFTPPFAHYPR/PEGCSSPPNS	73		17%	3%	10%	4%	72%	1%	21/(no TMs)727
CAGLOA01738g	310	LKSLIDNSADEKDFNINRHRRR/FDKEYDHSQK	132		25%	5%	14%	4%	60%	6%	25/(no TMs)469
CAGLOA02255g	98	EKTGYVLLNITQNYAGNDTMAKR/DVKVYGYDGW	200		26%	19%	4%	2%	70%	18%	25/(no TMs)272
CAGLOA02277g	98	EKTGYVLLNITQNYAGNDTMAKR/DVKVYGYDGW	200	0	26%	19%	4%	2%	70%	18%	25/(no TMs)272
CAGLOA03806g	430	LEFLRPDSKVD RDPIITYVNIKR/FD TDAILCLK	317		14%	3%	12%	2%	74%	2%	o20-42i51-69o470
CAGLOA04081g	222	SASKGKKGKDSASSVSSHRR/PDEPITVLTY	351		42%	7%	20%	6%	38%	10%	18/(no TMs)(352-GPI)o374
CAGLOA04081g	345	VHTSVSHSASHSTSQSVSEVAKR/AVHFENAGVV	338		26%	5%	6%	4%	68%	6%	18/(no TMs)(352-GPI)o374
CAGLOB00616g	399	EYNQAKPLMNSLIESENSNKKR/LSLLTNFLLM	277		38%	10%	13%	5%	49%	11%	17/(no TMs)(394-GPI)o416
CAGLOB01991g	121	FGSMSMIKPDSSNAHQIGSEKR/YPYDNLLYFP	99		12%	2%	17%	3%	70%	2%	26/o49-70i82-107o176-195i216-239o330
CAGLOB03091g	83	ANTSSKISSYEKLESENNGR SER/LYPVYVDDSK	54		46%	16%	6%	3%	47%	13%	o605-626i670-694o696
CAGLOB03091g	257	QSHDKIMDNNNDTNLMEKIKSRR/EENSANFERE	346		33%	4%	12%	6%	55%	6%	o605-626i670-694o696
CAGLOB03091g	414	SGSSSR YRTSNNSQQSLIRQRR/NGDYVAIPVL	426		33%	5%	10%	5%	58%	9%	o605-626i670-694o696
CAGLOB04235g	242	TEQALVVNQLGYELCPECEKQRR/EEMETEHLWQ	384		37%	14%	12%	3%	51%	16%	22/o268-289i313
CAGLOB04477g	228	SDVSDISYEMEPFPPAQSPSRR/QHLSTNYNSS	61		45%	20%	11%	3%	44%	19%	i41-62o283
CAGLOB04565g	159	GLIPREHWSYFEWIDQEKARLTR/EEMKKQKIY	172		22%	6%	8%	2%	69%	7%	37/(no TMs)401
CAGLOC00363g	260	AVNYATPAAMPSDNGGWYNPIKR/QSLSTRKLVN	104		17%	3%	20%	7%	63%	5%	17/(no TMs)276
CAGLOC02607g	11	*STNWNTWKLKR/ISASNICQAF	610		25%	4%	8%	2%	68%	4%	o634-65i677

Anhang

CAGL0C03333g	24	*FQKR/DVWEVNAVMS	502		38%	6%	8%	2%	54%	5%	20/o186-206i257-276o296-316i337-358o370-393i400-420o426-445i727
CAGL0C03487g	291	MRTAALPSFYKLYMKNDLPR/GTYEISIGMN	42		19%	4%	7%	2%	74%	5%	i43-66o327-353i388
CAGL0C03696g	96	SIEHSITKLLDNEINDNRKVLRR/IKTIKNYLET	121		23%	5%	7%	3%	70%	7%	o735-761i773-799o805-827i831
CAGL0C03696g	304	NSEVRSRLRSKKSFYESDRPALKR/LHTSSSTEYLN	212		67%	3%	6%	2%	26%	4%	o735-761i773-799o805-827i831
CAGL0D02442g	48	EEDVDASEFEDDDYVPVKTNR/ PKEDNFTQQR	149		46%	10%	17%	5%	38%	15%	o73-92i374-395o417
CAGL0D02530g	190	GGIYQTNNTAWFTLDIPISQFKR/PIRIQSTPSV	191	0	15%	6%	5%	1%	81%	5%	23/(no TMs)(656-GPI)o676
CAGL0D02948g	295	KTFKKKHGIDVSDNEKALAKLKR/EAEKAKRALS	954		52%	32%	6%	1%	43%	32%	29/(no TMs)667
CAGL0D05852g	406	EITSEGSQASKLYDSAFCSNKR/LGGHCQKFKI	111		18%	8%	20%	10%	62%	5%	23/o753-775i864
CAGL0E00187g	179	SNANLDATGSLVVSRETTTKSR/ QDASEAITGK	277		55%	8%	10%	1%	35%	8%	o216-237i238
CAGL0E00275g	398	IPMPVTTSKIMSSRFEAPFGR/IGTWLEDKPK	40		51%	16%	17%	7%	32%	13%	20/(no TMs)880
CAGL0E00275g	449	RSHLAQFVSNLPSNDTQRPYSKR/NSSFFPSNLS	283		60%	10%	10%	2%	30%	12%	20/(no TMs)880
CAGL0E00385g	591	RVIVIEDGRVVEGSFDQLRSTR/DSELNKLKE	122		31%	12%	23%	12%	46%	12%	i47-73o93-115i191-210o269-293i305-326o608
CAGL0E01353g	9	*DIVQLLFKR/DDDDGPPAC	164		25%	6%	14%	4%	62%	3%	o34-53i65-86o106-126i277-297o338-358i370-389o409-429i433
CAGL0E01419g	49	YCKNSDNKSRNLDALGSNRFKR/ASSVASASPS	362	1	65%	20%	6%	2%	28%	18%	o500-518i519
CAGL0E01595g	388	EISGKTENDIPAKAIKTGIQVQR/GHILELLNSN	351		20%	3%	9%	1%	71%	4%	23/(no TMs)(439-GPI)o468
CAGL0E01727g	513	TSVPSDDKKIIDDHSGHDHAKR/DAYSNTAGIL	550	0	30%	25%	8%	3%	62%	27%	15/(no TMs)(519-GPI)o539
CAGL0E01749g	33	*KLDFEKVPENLALNKR/DDKGLQLE	46	1	65%	19%	12%	7%	23%	13%	16/(no TMs)482
CAGL0E01749g	474	TGDVFTPEATMSMSNSVALNKR/TTLKPLR*..	51	0	16%	7%	4%	1%	80%	6%	16/(no TMs)482
CAGL0E01771g	33	*DFEKTPISELTQR/ADNDSFIAPL	82	1	70%	15%	7%	2%	22%	13%	18/(no TMs)(501-GPI)o519
CAGL0E01793g	32	*KLDFTRVSPASVLEKR/ADDLSPVPLR	82	1	65%	20%	6%	1%	29%	19%	16/(no TMs)(480-GPI)o516
CAGL0E01815g	33	*KLDFEKSPKASSDLVQR/DDEYVNVPLK	151	1	78%	9%	9%	3%	13%	6%	16/(no TMs)(503-GPI)o519
CAGL0E01815g	492	PINKTNGSSNRTLNSRFDLQKR/VYSNQASLKN	60	0	19%	10%	4%	2%	77%	8%	16/(no TMs)(503-GPI)o519
CAGL0E01837g	32	*KLDFTKTPGSDLAKR/DVVDPEAAQL	68	1	74%	13%	9%	4%	17%	10%	17/(no TMs)(498-GPI)o521
CAGL0E01837g	492	RNSSLNSTFNSSSNYSRVQMKKR/TYPNSSSLQ	300	0	14%	6%	7%	1%	79%	7%	17/(no TMs)(498-GPI)o521
CAGL0E01859g	29	*ALDFEKSHGDDLVR/DVQDVKLINS	96	1	78%	9%	5%	1%	17%	9%	14/(no TMs)(484-GPI)o505
CAGL0E01881g	30	TAAAAGYLRFPFTKITDNLAKR/ADDDYLVNQL	85	1	76%	10%	6%	1%	18%	9%	(no TMs)(469-GPI)o508
CAGL0E01881g	479	SASSSGNTSSNGSTNLTRRFERR/DIPLNKGVP	223	0	24%	17%	7%	1%	69%	18%	(no TMs)(469-GPI)o508
CAGL0E02915g	93	VGEGGDINAEPTTTTITSIQVQR/VSLANEQPTA	351		21%	9%	6%	4%	73%	12%	21/(no TMs)498
CAGL0E04664g	30	*LSLFAASNDRNGKR/VSLGEITVDE	323		19%	2%	5%	3%	76%	3%	17/(no TMs)200
CAGL0E05412g	1146	IHVEFVKYEWPPWLRPQRFQRK/LAISKLFLD	669		30%	12%	7%	2%	63%	14%	19/(no TMs)1326
CAGL0E05940g	287	ATTMSVLTQRAIDYVRNLEYVQR/AENILYGNQN	504		21%	5%	7%	3%	72%	8%	24/o169-192i204-229o317-340i373-393o399-421i462-483o489-506i518-538o550-579i840
CAGL0F01793g	315	LYFNYYGQFTTLWDRLGGSYRR/PEDSLFDPKL	108		12%	2%	16%	7%	71%	7%	o95-121i142-160o183-201i255-272o364
CAGL0F03575g	152	NNKARGENGKFSHDRILLDYARR/SVDLSKLEKD	181		27%	7%	11%	2%	63%	7%	i106-124o755
CAGL0F03575g	312	YGASGILFVSNWNNNDIVEKR/SVALPQYGLG	327		26%	6%	9%	4%	65%	10%	i106-124o755
CAGL0F05137g	80	TKSQKTKPTPSTDLDKRAQKQR/SNLSEWQKQM	232	1	24%	23%	7%	2%	69%	24%	18/(no TMs)227
CAGL0F05137g	98	AQKRSNLSEWQKMLDQHNKQR/ELHKDITSLV	346	0	63%	26%	7%	3%	30%	24%	18/(no TMs)227
CAGL0F06369g	209	SVKDEAVLVSDGLSVAIDYALKR/PDLEINVPQY	265		26%	6%	14%	6%	61%	3%	16/(no TMs)889
CAGL0F06369g	734	IGLNVTEKYNSILSKMFSIRS/SSEENIKTLA	56		30%	4%	5%	3%	65%	4%	16/(no TMs)889

Anhang

CAGLOF09053g	608	KVSKYRQNDAGFSFSNIMSSFKR/ISEPERIFNT	198		21%	1%	11%	3%	68%	3%	17/(no TMs)809
CAGLOF09053g	781	STLFKEVVSFEELYEDPEILHKR/DILRAFVNK	106		24%	4%	20%	9%	56%	8%	17/(no TMs)809
CAGLOG00308g	28	*HAAHTHEKRR/DVVTKTMHAV	526	1	59%	26%	9%	4%	32%	23%	18/(no TMs)374
CAGLOG02717g	34	*KFLRSDAVSWHDRVPR/GNFTWVHAQ	40		27%	7%	21%	10%	52%	4%	17/(no TMs)515
CAGLOG02717g	303	NRTGFISKSANHIISCPETIYGR/SGLDIATILA	310		23%	8%	15%	4%	62%	6%	17/(no TMs)515
CAGLOG02959g	86	FFLSKKPSPIQVVSNSAKNKKR/YVTPPTLYKL	277		41%	9%	8%	2%	51%	8%	i27-51o114-134i146-173o179-199i239-258o918
CAGLOG04279g	566	FYECESDKLQEIIESEVGIKRR/SLPSTDVNVV	158		12%	3%	14%	4%	74%	2%	i27-44o660
CAGLOG04279g	578	DIESEVGIKRRSLPSTDVNVVRR/DEQQFASKEP	383		26%	6%	15%	6%	59%	3%	i27-44o660
CAGLOG04279g	642	CTFVDRIGDADSKLVRFDKKTKR/FVNHLAQIWR	522		21%	7%	8%	2%	70%	7%	i27-44o660
CAGLOG05896g	28	*QIQKR/DDDNVRYITS	366		29%	2%	19%	9%	52%	10%	23/(no TMs)294
CAGLOG07667g	132	SPAQDSNLSDFAKSMLNEHNIKR/ALHQDTNPLT	254	0	14%	5%	8%	2%	78%	7%	18/(no TMs)258
CAGLOG07887g	153	ETKSQTLPLLALELTSNDKKVRR/DRIIDEIEEI	197		31%	6%	12%	4%	57%	10%	20/o199-22i242
CAGLOG08338g	176	DNAVTKAGISLLKRYIQTSNSKR/DHNSGSSNPE	369		16%	4%	17%	4%	67%	4%	o32-57i120-139o191-212i449-471o499-523i544-567o587-614i635-656o662-682i703-721o727-745i794
CAGLOG08668g	44	. /AVTVEYVYAT	232		26%	1%	5%	4%	68%	4%	23/(no TMs)(449-GPI)o469
CAGLOG09449g	74	TAEDNPGKIEYTSGLAMSLTKR/FDNPSLKSNF	82		20%	4%	10%	1%	70%	3%	21/(no TMs)(421-GPI)o452
CAGLOH00330g	1131	LHGRAGATSHTDISVEELIWKRR/LKINKKNTPD	90		35%	12%	5%	3%	60%	14%	i35-59o2635
CAGLOH00330g	1228	LSDFTYVWGRVNSKLPDDINKR/VLDFSTNPFLL	228		27%	4%	6%	4%	66%	7%	i35-59o2635
CAGLOH00330g	2010	THQFMELASHKTNRMFNDI INKR/QQKDNVNEEE	91		34%	13%	10%	3%	56%	15%	i35-59o2635
CAGLOH00330g	2514	MSFSQDTRIHTGTKIDIDEMFKR/STNFLTIENM	494		28%	16%	8%	2%	64%	16%	i35-59o2635
CAGLOH01705g	68	MVSVHYTGSLAENGKVFDSLSLR/NEPIQFKLGA	92		26%	4%	18%	7%	56%	4%	23/(no TMs)136
CAGLOH02893g	179	DDKRAFVYPKPLPKFKWYENDKR/LTLEEQKEEK	125		26%	6%	8%	2%	66%	6%	o678-700i867
CAGLOH02893g	618	PVAYMYIYQRLGHFNMMNRR/SEPCITLRCY	220		23%	7%	14%	7%	63%	3%	o678-700i867
CAGLOH03135g	85	LFVNTTLYNQATKGEKLSDFTKR/DANPDAEAEA	544	2	36%	29%	48%	30%	16%	7%	21/(no TMs)159
CAGLOH04499g	77	PKAHEAFKLINRAFEVLSDSQKR/QIYDQIGRDP	114		33%	10%	12%	6%	55%	4%	o207-226i232
CAGLOH04499g	179	GGPGGFRMYTTGNGPFQRRFR/ADPREAYRQQ	595		26%	17%	10%	3%	64%	18%	o207-226i232
CAGLOH04499g	193	PFQRRFRADPREAYRQQQRR/QGNQENELVN	140		41%	12%	12%	3%	47%	11%	o207-226i232
CAGLOH06919g	428	NTFDKLSSEVEYINNFEKLYKRR/EGGDSYTNFF	169		27%	2%	15%	4%	58%	3%	19/(no TMs)627
CAGLOH07359g	224	LMMGGGRTHFYPSGTNDPRYGR/GLRQDDRNLI	79		45%	8%	19%	8%	36%	12%	i21-44o545
CAGLOH07403g	139	VKDMTKKDIEEYLAPSFANKGKR/PKAAFVTLIR	237		20%	4%	25%	5%	55%	4%	29/(no TMs)463
CAGLOH07997g	244	DYNYAPPAPFSDNGGWYDPKRR/QSFTTRKLM	113		30%	12%	12%	2%	58%	10%	15/(no TMs)265
CAGLOH08910g	157	AAINDINNYHCFKLARRKKYVRR/NYSFFLPGSF	403	0	15%	6%	5%	0%	79%	6%	19/(no TMs)346
CAGLOH09130g	243	DRKKYSQASWYKMERKKADFHRR/NKISDYCEPF	144		30%	6%	16%	9%	54%	7%	i7-25o1015
CAGLOH09240g	93	INTRIMKLKEEKDDPKKAGKTR/TESEEDKILL	66		33%	21%	7%	3%	60%	23%	33/(no TMs)642
CAGLOH09240g	187	VNPRELFKLDLQATDAYDFQKR/MFFFLKNVTQ	401		27%	9%	7%	2%	66%	11%	33/(no TMs)642
CAGLOH09240g	505	DYVIEPEDAATIGKLEYSEKDRR/YSCAIQMAH	93		42%	20%	10%	4%	48%	23%	33/(no TMs)642
CAGLOH09262g	41	*QRRSTR/YEEDLLFSSG	122		33%	4%	8%	2%	58%	4%	35/(no TMs)625
CAGLOH09262g	92	PFFSLENMLPSWMLSTHTAREKR/VLLNWDKSSG	584		22%	2%	5%	3%	72%	5%	35/(no TMs)625
CAGLOH10032g	23	ARVAII IYSMYGHTAALAEAEKR/GVESAGGKAD	183		25%	2%	16%	6%	59%	7%	(no TMs)198
CAGLOH10076g	9	*STFVDLYKR/GGNEAVKINP	72		28%	4%	15%	6%	57%	4%	o35-55i62-82o117-135i142-162o168-193i205-227o239-258i325

Anhang

CAGLOI00484g	38	*VIAPKNKDTSLHFVNEKR/YYDYDSKAIG	358		27%	2%	10%	1%	63%	2%	20/(no TMs)443
CAGLOI00506g	379	QLVETMKWMSARSRLGDYVGRK/FPDNIQEVLN	462		22%	6%	19%	5%	60%	7%	22/(no TMs)621
CAGLOI00506g	497	QPNRSNSFDLAASIYNLIEPGKR/PLSSTAPSII	132		18%	5%	11%	6%	71%	4%	22/(no TMs)621
CAGLOI00506g	614	SVVNAIQRIAGEWHAVSDYWRKR/GISAVY*...	181		19%	6%	15%	6%	66%	5%	22/(no TMs)621
CAGLOI01848g	14	*ICADGRKPAMIFRR/DSDSDSDGDG	208		32%	9%	7%	2%	60%	10%	o53-76i701
CAGLOI02508g	115	EAVVTSTSATNGSLNSPPLTKR/YAVSRSLWYG	65		55%	14%	6%	3%	39%	11%	i64-82o131-156i175
CAGLOI04532g	320	FWRRTTMPAYYEIANLPYNQNK/VRNAFDDITP	133		19%	2%	22%	4%	60%	5%	i12-30o612
CAGLOI06182g	61	TDYASTFGIAVQPIATPASVAKR/AVSQIGDGQV	338	1	54%	33%	4%	2%	42%	35%	19/(no TMs)340
CAGLOI06204g	61	TDYASTFGIAVQPIATPASVAKR/AVSQIGDGQV	338		42%	23%	6%	3%	52%	26%	19/(no TMs)349
CAGLOI07777g	122	KPPTLDANTTVDTPTLFERVVKR/AIFGENTWTI	857		19%	2%	6%	2%	75%	4%	16/(no TMs)240
CAGLOI07777g	164	GSRSVNPLQGFITMVVNKDMVKR/QLMDIEGRP	375		24%	5%	8%	2%	69%	7%	16/(no TMs)240
CAGLOI10296g	33	VYIKADRKEFMDMPIGKSDLWKR/DISRNMYDKL	93		27%	1%	20%	11%	53%	10%	(no TMs)446
CAGLOJ00583g	516	ALLSYDGPDPVYQKIELDTRDKR/DEHIHGNSIG	68		16%	5%	15%	6%	69%	2%	i24-45o817
CAGLOJ02068g	128	IRAYPTIRLYGPDGFLEEYHGKR/TKEEFLKFAR	301		25%	3%	29%	18%	46%	15%	19/o662-678i708
CAGLOJ02288g	37	VVNCRLRTIEKRIVSSHASLSKR/SAVDLQFRRF	54	1	69%	16%	4%	1%	27%	17%	(no TMs)(508-GPI)o541
CAGLOJ02288g	46	EKRIVSSHASLSKR/SAVDLQFRRF/FNNLYYESVL	195	1	60%	26%	6%	1%	34%	25%	(no TMs)(508-GPI)o541
CAGLOJ06820g	104	DSFTIHGLWPDCCMGYPQFCKR/DFLIDDVDYL	502		37%	9%	17%	7%	46%	15%	(no TMs)433
CAGLOJ06820g	297	PLRMSRCKDQGIKYYPKGYQRR/DNRGPNKKPI	100		35%	6%	13%	7%	51%	12%	(no TMs)433
CAGLOJ07546g	89	YVFSVAKSENEDDIDSGNSLRKR/AIENEDHILS	64		33%	8%	13%	8%	54%	8%	24/o671-693i789
CAGLOJ07546g	112	AIENEDHILSSEFLTPSNLQKR/MPVPAPPLDS	60		33%	6%	9%	5%	58%	3%	24/o671-693i789
CAGLOJ08481g	237	KIVLSEPDGTRDRLDSDLINGKR/STFQVLDIPE	323		20%	5%	5%	3%	75%	7%	29/(no TMs)(?-GPI)o583
CAGLOJ08756g	267	DEGLLGKGLSTVPQSVRRKKTTR/PPFAGSSNLK	68		41%	14%	18%	8%	41%	19%	20/o752-772i833
CAGLOJ08756g	718	STVVRRPFFDRLVQLLKNLSRETR/GDNKKKNQHR	95		39%	9%	26%	11%	35%	17%	20/o752-772i833
CAGLOJ08910g	458	SPNSEYSNENDSEEINVEEWYRR/SMMFRPKRQK	131		31%	6%	7%	3%	62%	7%	23/(no TMs)(470-GPI)o494
CAGLOJ09922g	42	*MPDLKDCITTAQGNHQHQR/AVAVEVYET	386	1	66%	19%	4%	2%	30%	20%	23/(no TMs)427
CAGLOJ11748g	53	NANFIRNAADGLSPAEEKWLKRR/DPITRDALQT	64		24%	12%	14%	7%	62%	19%	20/(no TMs)(667-GPI)o695
CAGLOJ11748g	201	YDKDFSKTLARWTAIGSQVQGKR/DAGFNVTITD	338		25%	8%	19%	8%	56%	14%	20/(no TMs)(667-GPI)o695
CAGLOJ11770g	59	NDINLLRNATGLSQSEIDWLKRR/DVNTREALES	64		38%	8%	13%	3%	49%	8%	23/(no TMs)(633-GPI)o659
CAGLOJ11770g	64	LRNATGLSQSEIDWLKRRDVNTR/EALESFLKRV	59		24%	4%	12%	3%	65%	3%	23/(no TMs)(633-GPI)o659
CAGLOJ11770g	73	SEIDWLKRRDVNTREALESFLKR/VTSNFTSNSS	509		22%	2%	7%	3%	71%	5%	23/(no TMs)(633-GPI)o659
CAGLOK00297g	99	ESGRLDGHI IYKVVQHSIVNAKR/SSRIDKDEAE	148		23%	7%	11%	1%	66%	7%	o526-548i560-577o627
CAGLOK00297g	471	DAVSFSYSGTGALKTDFTRTGKR/TKAGAFNDFV	56		24%	7%	11%	2%	64%	8%	o526-548i560-577o627
CAGLOK02937g	48	WKKYVEQKDAITQQVSVKMLKR/EYDDNLFREQ	702		37%	7%	8%	2%	55%	7%	24/o79-97i285-305o438
CAGLOK03927g	156	TILERPLSFHIEFTPEEWELEKR/PEFGCPLSLL	66		20%	5%	12%	5%	68%	4%	i37-57o235
CAGLOK05137g	482	KFASNIANTAKGNYKSLNSHKR/AWYDLYNDAF	462		26%	5%	10%	1%	64%	6%	i83-102o1212
CAGLOK05137g	946	RIKVPNRDILHDRNITSALYSKR/DDDLSATDDY	553		25%	9%	14%	7%	61%	4%	i83-102o1212
CAGLOK06237g	80	QISLGLTPRKPVVITDLRGTGNKR/KLHGRFLHIT	0		41%	16%	13%	3%	46%	17%	i12-29o663
CAGLOK06237g	454	SDLDEIKKKKKKKKKKKGKKEKR/TLEEIYDQHS	526		28%	2%	14%	5%	58%	4%	i12-29o663
CAGLOK06963g	815	PLDTIPIFLEGGHILFQKEKYRR/SSELSNDPY	368		26%	6%	16%	6%	59%	9%	21/(no TMs)941
CAGLOK07359g	161	TAKNGAKAKIVVFEVGGGVVVKR/NIAIEEEKKV	420		37%	9%	6%	3%	56%	6%	(no TMs)190
CAGLOK08976g	670	LEISHKGQIRFDTIIKGLLMWKR/LGKECTAWKH	220		14%	5%	7%	1%	79%	5%	i123-141o793

Anhang

CAGL0K11880g	8	*SYVDLYKR/GGNEAVKINP	72		35%	6%	14%	5%	51%	5%	o34-55i62-81o116-134i141-161o167-192i204-225o237-262i316
CAGL0K12914g	31	*INKSGDETTLRR/DISLLDLVKA	92		42%	17%	5%	3%	52%	15%	19/o397-416i428
CAGL0L00979g	337	FYKEEGLCVDGPEFRYYPEVFER/EPYKANLLRS	59		50%	5%	22%	5%	28%	6%	26/(no TMs)475
CAGL0L01727g	385	GMYGLGEQMAALEVMVNTQALKR/APPYNATNGG	212		25%	3%	12%	4%	63%	3%	19/(no TMs)(420-GPI)o446
CAGL0L02365g	92	DAGEKKYDHLIVFPATTKSLNKR/LSMNHLLNFY	59		29%	7%	11%	2%	60%	6%	20/o391-413i428
CAGL0L03289g	29	*AVQHDKHAHEKR/DLVVVEYVD	533		38%	10%	7%	3%	56%	7%	16/(no TMs)366
CAGL0L03333g	344	NEYHWSNEREQDSTTNDNLPR/SYEYAQALKN	52		30%	8%	17%	7%	53%	13%	43/o410-429i450
CAGL0L04466g	69	GKSTLLNILFNTQFDTMDAQVQR/QQTTKGIWLA	351		24%	6%	9%	3%	67%	5%	o678-694i701-718o783
CAGL0L04466g	165	YQGNMGLLKTVFEVNLISLFGKR/GNDHKVLLLF	495		20%	5%	10%	5%	69%	6%	o678-694i701-718o783
CAGL0L04466g	323	TEDIANEAYAKFTEEYETEKR/INDKTELVS	124		40%	5%	5%	3%	55%	3%	o678-694i701-718o783
CAGL0L04466g	657	HSNRFAHILNETQKEKVQQFR/QINVTVLDC	195		20%	2%	8%	2%	72%	3%	o678-694i701-718o783
CAGL0L05434g	30	*SPKHDHEDRHAIKR/DVDVVTVEY	202	1	70%	15%	6%	2%	24%	14%	16/(no TMs)(326-GPI)o346
CAGL0L08096g	132	YCFHLSPTQTNLKTAAHNGYCRR/VETLFGSLDE	200		23%	4%	11%	2%	66%	3%	27/(no TMs)564
CAGL0L08448g	159	MIMSLNIFSNAGFGTKFIRRRR/NNAEVGVPSV	288		40%	5%	10%	1%	50%	6%	o12-28i40-58o64-93i123-146o173
CAGL0L10670g	207	STGSHTGSSSESQESDDTSKR/SLLKKRSLRE	102		51%	11%	8%	4%	41%	10%	21/(no TMs)(241-GPI)o263
CAGL0L11528g	53	STRLSEGILEYYDMFDSGAVVPR/EDFNTVCKTL	75		28%	1%	26%	15%	46%	16%	19/o275-291i319
CAGL0L12804g	166	GFIIPKSASGDAALKQLEAAVQR/AQTDKKNR	214		21%	6%	6%	4%	73%	10%	(no TMs)393
CAGL0L12804g	207	DSKGFDKLQEKERHALAVQKERR/SAMALARNEL	214		26%	7%	6%	3%	68%	10%	(no TMs)393
CAGL0L13134g	125	FANGYHSSDLFADSKYHYASDKR/DEAFHAFRS	68		26%	4%	26%	14%	48%	14%	o622-644i725
CAGL0L13134g	276	VHPEKVGAAEVNGLSASNLTSRR/SSTSSSPLQK	52		27%	11%	4%	3%	69%	9%	o622-644i725
CAGL0L13244g	352	LEKEATDFILQMVLDDIIVKKR/SEVFNQSDEN	396		22%	5%	12%	2%	66%	4%	19/o412-435i447
CAGL0L13266g	45	DINTLEQTSSESQEYDNFFKQSKR/PGNVKA	247		35%	7%	17%	12%	49%	12%	20/(no TMs)579
CAGL0L13266g	114	VAKLQLRGSISLAPNDPFFHHTR/DKEFSTEELG	70		29%	8%	25%	8%	46%	10%	20/(no TMs)579
CAGL0M00132g	324	VGIPCSSSQSIYGSSSSSSSSKR/LDCPSGLFFD	128		31%	16%	6%	3%	63%	15%	24/(no TMs)(892-GPI)o922
CAGL0M00132g	872	AIFRSATSRHDTYTTISAFFTPR/LSLQITTLVQ	44		34%	14%	5%	3%	60%	15%	24/(no TMs)(892-GPI)o922
CAGL0M01298g	128	VTVENFIRLLTDRWTEQPKSKR/LQTDENSNIF	434		42%	9%	9%	3%	49%	9%	22/o359-377i390
CAGL0M01386g	7	*VDKKKR/SFSETGIFKL	407		53%	17%	12%	2%	35%	18%	o1284-1303i1427
CAGL0M01386g	562	KRNSRITTLSENDEIVGDNDRKPR/SLSRKFLNRR	57		34%	15%	13%	2%	53%	16%	o1284-1303i1427
CAGL0M02211g	68	VQQLGHKYVSQYKANPQTVLPR/DHLFYPDQVH	59		58%	9%	12%	1%	31%	9%	23/(no TMs)415
CAGL0M04191g	44	ADLSYVKLDFDKYYGETFETAKR/GRSQADIRVN	51	1	62%	24%	7%	2%	31%	22%	19/(no TMs)(580-GPI)o601
CAGL0M04191g	56	YGETFETAKR/GRSQADIRVNKR/ANGYEEVQIT	364	0	16%	7%	6%	0%	78%	7%	19/(no TMs)(580-GPI)o601
CAGL0M05599g	100	HFRGPLKLSQFAVYYPASSNSKR/AVAEDCDSDV	295		39%	11%	7%	3%	54%	10%	18/(no TMs)446
CAGL0M05599g	120	SKRAVAEEDCDSVVAKPVHKHKKR/DVAVEVIEVT	848		40%	5%	7%	2%	53%	3%	18/(no TMs)446
CAGL0M05841g	44	DIDRNYFERSVAVTKYDGVRRR/EYFVPLKLAT	202		35%	15%	7%	2%	58%	16%	(no TMs)411
CAGL0M07205g	185	PVATAEEVELMHETFGAVPALKR/EPNPSHTASS	265		41%	16%	26%	9%	33%	19%	o6-23i88-105o241
CAGL0M08492g	62	TDYSSTFGIAVQPIATSSSVAKR/AVSQIGDGQV	338	1	56%	31%	4%	2%	40%	33%	20/(no TMs)335
CAGL0M08734g	34	LVAANKNIVEMITFRDENGLHRR/LAPEEYRGGV	43		49%	13%	6%	2%	45%	12%	(no TMs)253
CAGL0M08734g	56	RLAPEEYRGGVLDGEESSRLQKR/DLEMHPVDL	94		29%	10%	9%	2%	63%	9%	(no TMs)253
CAGL0M08734g	211	LLQSKQYSMQHSNGRGGDVLLKR/EGEKYCVASV	119		33%	9%	8%	3%	59%	7%	(no TMs)253

Anhang

CAGL0M10571g	172	KFAGPTLEASIKSKEKELRKLRR/EQKRQLDSGE	195		31%	6%	6%	3%	63%	9%	o200-221i242-262o274-293i305-324o380-397i418-442o462-482i544-563o569-587i599-617o619
CAGL0M12320g	284	VVVSNDILSISVQRRANLLKR/GISVASSSDS	76		28%	4%	12%	3%	60%	2%	20/o167-191i203-220o341-367i396-416o428-448i486-506o512-531i543-564o576-602i769
CAGL0M13453g	319	GICKFEMEAIIVDMVNTPQSLRR/GTNPIPRPLS	74		34%	6%	21%	9%	45%	14%	22/o582-602i646
CAGL0M13805g	30	*ANVHHEHKNKR/DVVTKTVHNQ	585	1	73%	13%	4%	1%	22%	13%	19/(no TMs)371
YAL034C-B	85	LIFNASYLPFLFTRTIKISKVER/QQGERSKMAM	56		20%	7%	12%	2%	69%	6%	23/(no TMs)117
YAL053W	291	ATNGVSNVVANKDVLISVQKR/AISMASSSDY	375		24%	5%	6%	2%	70%	6%	25/o348-374i403-424o430-455i493-513o519-538i550-571o583-609i783
YAL058W	461	GPTERKFMNSRLGNLQTTFHNER/ESPMPFDRII	53		22%	4%	13%	6%	65%	8%	30/(no TMs)502
YBR015C	316	FWRRTTHPLYDIAGIAVDKRR/VRNSRDDITP	203		23%	6%	14%	5%	63%	2%	(no TMs)597
YBR015C	321	THPLYDIAGIAVDKRRVRNSR/DDITPPAVYT	41		19%	6%	12%	2%	69%	4%	(no TMs)597
YBR023C	85	QEQMNHLDVLPSSSTGVNPNATRR/SGSLRSKGSV	92		62%	13%	8%	1%	30%	13%	o169-191i203-223o453-476i1029-1051o1057-1076i1083-1107o1165
YBR042C	96	VRITTENSSVPKGTFFLDLKKR/ILSHLKSNSV	407		20%	8%	10%	2%	70%	9%	o13-37i58-77o120-138i377-396o397
YBR054W	9	*SDYVELLKR/GGNEAIKINP	76		24%	3%	11%	4%	64%	7%	o35-56i63-82o117-135i142-162o168-193i200-219o239-260i344
YBR093C	174	LAQYGYMVENQTSFAVFTSNSKR/CHDTAQYFID	369		15%	6%	13%	3%	72%	4%	18/(no TMs)467
YBR161W	348	SSYKNKEKRRNPTREHYNSRGRK/LRKDSNIPYD	165		46%	19%	6%	2%	48%	18%	i7-24o200-220i274-299o376
YBR162C	134	KQSIESCKEGEAVSRHKHQHKKR/DVAVEYVQVT	483		35%	15%	7%	1%	58%	15%	24/(no TMs)455
YBR229C	591	YGLSVHEATYDAIKSIYSPDKR/PFLLTRAFFA	43		29%	4%	10%	3%	61%	3%	23/(no TMs)954
YBR229C	816	PLDKIPLFIEGGHIITMKDYRR/SSMLMKNDPY	294		22%	5%	8%	3%	70%	8%	23/(no TMs)954
YBR265W	60	VSRSEARLLDTCNEIRIEAHLRR/ETTDEGQVQH	395		14%	2%	13%	1%	73%	1%	o285-307i320
YBR290W	98	QAARHIGKIGRHFNLDRLEFKKR/TQQSSDIQQG	339		23%	4%	11%	2%	65%	4%	o179-200i212-230o281-301i321
YCL048W	432	QDGTIEKGYKDLTQEGEDSKKR/VISKYANSAN	96		35%	6%	8%	5%	57%	7%	26/(no TMs)(440-GPI)o463
YCR068W	139	TNPFDFEFELRRMPLLMKRMKER/DPEFIESYIY	42		31%	8%	24%	10%	45%	16%	i12-34o520
YCR089W	1039	SEMSTFSVSTQSLPLAFTSSEKR/STTSVSQWSN	99		26%	3%	8%	2%	66%	2%	23/(no TMs)(1588-GPI)o1609
YDL024C	175	LAKYGKLMENCTNFPIFTTNSKR/IYDTAQYFAE	369		18%	8%	15%	5%	67%	4%	21/(no TMs)468
YDL221W	106	PASSLSLSSFFVTVPFCAARSRR/GLTVALIAQS	307		31%	13%	11%	2%	58%	11%	(no TMs)183
YDL241W	102	DFYACGFLQKGTETNAGIHNTR/PSHSLAKFTI	62		20%	5%	16%	7%	65%	5%	(no TMs)(96-GPI)o123
YDR003W	13	*ILREQIDFLIHKKR/QDDNNNNGEA	412		24%	4%	11%	3%	65%	2%	o38-59i210
YDR107C	74	EEGNDVSGDKEHFLYSYDYNNKR/FHFCEPEHVE	306		23%	6%	6%	2%	71%	8%	21/o310-332i373-399o411-431i443-467o479-501i533-553o565-591i603-621o633-662i672
YDR230W	34	PASTFCFPLVLFTRTCTISSQRR/PEINSGYFVQ	72		22%	10%	19%	10%	60%	1%	(no TMs)115
YDR245W	101	SFQGGNSNDNKLSTGSSINSKR/YVPPYSKRSR	295		37%	8%	9%	1%	54%	8%	i50-67o393
YDR245W	109	DNKLSNTGSSINSKRYVPPYSKR/SRWSFVNQDP	142		40%	10%	18%	5%	42%	14%	i50-67o393
YDR245W	173	NKKAYAKRHGYALTIKDLTTSKR/YSHEYREGWQ	102		16%	4%	17%	4%	67%	1%	i50-67o393
YDR262W	44	*VVNLITFKDSNGKHLKKR/LAPEEIPPRL	85		23%	4%	22%	11%	54%	7%	27/(no TMs)(256-GPI)o272
YDR326C	9	*RDEATRKKR/SFSDGHFFKK	565		47%	8%	12%	1%	41%	9%	o1281-1302i1438
YDR331W	140	VTVENFIRLLTDRWTEDEHPKSKR/LLTDENSNIF	434		32%	5%	6%	3%	62%	3%	23/o378-397i411
YDR415C	217	DNGSGVTNMEALRLYTENFLKR/GFRPNNTVEF	254		21%	2%	10%	1%	70%	3%	19/(no TMs)374

Anhang

YDR420W	210	EEGTLAQTSSISETTRIAQMVTR/VSQISSITAA	59		36%	15%	6%	4%	58%	19%	22/o1486-1508i1802
YDR452W	448	NKEAYMNTVRDYYRKVWNKLER/VDEKNVENEK	86		27%	4%	12%	3%	60%	2%	i21-40o674
YEL028W	135	YMAKLGDATVNHTDLPFRNNKR/LTTQELNAFQ	319		20%	5%	10%	2%	69%	4%	22/(no TMs)153
YEL060C	86	HGKFHKKGRKQDKESPEFNKR/ASGSHGSAHE	323		25%	26%	8%	3%	67%	24%	19/(no TMs)635
YEL060C	188	TAQFNPDASKIIPNRYIIVFKR/GAPQEEIDFH	522		44%	23%	9%	4%	47%	19%	19/(no TMs)635
YER001W	519	LNECDNTLTTEEYIFKAFFHRR/QHQLEAGLFA	351		28%	6%	14%	3%	58%	7%	i12-36o762
YER001W	596	ALQGVYSGAIGDYVKKTDLNGKR/QEAAVEICSG	404		25%	5%	20%	8%	55%	6%	i12-36o762
YER005W	265	LAQLINTLPENTNDYENDDFSTR/NLNDPCMPRG	47		31%	4%	20%	10%	49%	13%	o501-517i630
YER019C-A	16	*AASVPPGGQRILQKRR/QAQSIKEKQA	118		32%	9%	7%	3%	61%	11%	o62-81i88
YER076C	74	MVEFYQITGLTANGTVPEFNKR/DATEYPNIIS	488		26%	8%	19%	5%	55%	7%	24/(no TMs)302
YFL027C	357	LWEEVLQITRKDATTASKKALKR/VSLNKYSALL	212		32%	9%	6%	1%	62%	10%	o429-445i497
YFL034W	62	GLSNQDDLTVMHTQAKEEVFKRR/EEDGTRTEDA	338		23%	4%	8%	2%	69%	6%	o461-479i542-565o585-606i1073
YFL038C	128	DRYATSTVLKLLVGNKCDLKDKR/VVEYDVAKEF	229		19%	3%	5%	3%	75%	6%	(noTMs)206
YFL042C	124	SPVSRSTTPTSPVSPSIIISHRR/EPSMGSKRRS	102		37%	6%	8%	4%	55%	2%	o636-654i674
YFL042C	137	SQPSIIISHRREPSMGSKRRSSRR/ISNATIAEIG	52		48%	19%	6%	3%	46%	17%	o636-654i674
YFL042C	162	NATIAEIGSPLQQVEKPDVEKTR/LTPSQMKEDN	41		57%	13%	5%	3%	38%	12%	o636-654i674
YFL042C	467	EGSQITTIIPNFTSIDGSSMTLKR/EYSYEKALHF	180		28%	10%	11%	5%	61%	5%	o636-654i674
YFL051C	33	*EACLPTNKR/EDGMNINFYE	173		12%	3%	13%	4%	75%	5%	24/(no TMs)160
YFL051C	99	LSIYSPPESTPTCVTYAVLKR/DEGDYDFCGP	927		28%	12%	15%	4%	57%	11%	24/(no TMs)160
YFL051C	116	YAVLKRDEDDGYDPCGPLYETKRR/DTEYCDPNTA	150		14%	5%	19%	2%	67%	6%	24/(no TMs)160
YGL028C	31	*ILPKREVVTTR/VHTASTNVV	62	1	62%	23%	4%	2%	34%	24%	21/(no TMs)542
YGL075C	268	FRRRLQRQTDITISFLKDQIRRR/GLNCSNDKVS	51		28%	3%	15%	3%	57%	3%	o305-327i387
YGL089C	100	KREAVADAWHLNLRPGQPMYKR/EANADAWHL	530		21%	2%	12%	6%	68%	5%	20/(no TMs)120
YGL125W	336	PWRKSLNPKRKNEEVRPIFWKRR/PYSYVARTSQ	92		28%	2%	12%	2%	60%	4%	(no TMs)600
YGL126W	220	NTFIHQMRWQGPLTTSAQCCR/FGGHWAGGHD	140		38%	9%	12%	3%	49%	10%	i12-31o51-66i87-105o236-254i326-349o355-371i380
YGL156W	74	HVKVWWYQVSFERGSSPVSPDKR/PSWKSIIERD	51		22%	9%	13%	14%	64%	22%	(no TMs)1083
YGL156W	547	GNVIPKLQVGITVDEFYDDILKR/TNQGHDLPW	237		32%	6%	17%	8%	51%	14%	(no TMs)1083
YGR141W	267	MSAHGGGSAYKFEATEKIKRLRR/VDGKLTNEVI	541		20%	1%	8%	1%	72%	2%	i15-37o467
YGR141W	416	SLCQRFKWWNFWKGPCARRYIKR/GEGLDAEKND	304		18%	0%	11%	3%	70%	3%	i15-37o467
YGR176W	95	HSRLWSKRLGRTQISRRLQTARR/LNPERSTAWS	52		31%	3%	15%	4%	53%	5%	26/(no TMs)115
YGR279C	29	*ANHEHKDKR/AVVTTTVQKQ	183	1	53%	34%	4%	1%	42%	35%	20/(no TMs)386
YHR003C	397	NKEISLTVNVLMTKEEQEIEHEKR/ILLDGEKLT	426		20%	6%	7%	4%	73%	10%	24/(no TMs)429
YHR009C	142	VIENYERLSKAYNLNVPPPKRR/PGYISNKFNI	113		26%	8%	16%	8%	58%	15%	i24-41o523
YHR009C	303	LKCEMDCVHSLKYLPSVVKNR/NSRGHAENPD	119		29%	8%	17%	7%	55%	11%	i24-41o523
YHR042W	25	GIDNTDFTVLAGLVAVLLYVKN/NSIKELMSD	403		22%	8%	13%	4%	65%	5%	(no TMs)691
YHR042W	522	VHYDLNGPRKLFANYKLPVHVR/SNFRLPSPNS	319		24%	8%	15%	5%	61%	10%	(no TMs)691
YHR045W	422	VKDAEVISTFPEKIPQLKKTNR/ITPSTEQLNK	56		26%	9%	7%	2%	67%	10%	20/(no TMs)560
YHR076W	60	IDSETNFSHIIGSCSQIISFSKR/TFYSSAKSGY	362		20%	8%	6%	4%	74%	12%	i17-39o374
YHR079C	34	*PLSSRSTRR/QIVEDEVAST	52		32%	5%	7%	3%	61%	7%	25/o534-555i897-914o1115
YHR079C	98	SSYPTPNLLNTADNRANKKRR/AANSISVPYL	242		24%	12%	4%	3%	71%	9%	25/o534-555i897-914o1115
YHR080C	50	RNFLNNTGENVSTTASAERFR/VGGNPDIPLS	476		35%	11%	16%	7%	48%	15%	o1204-1223i1345



YHR080C	147	NSNSVPSLSVQALEKEKLQSGKR/EGSSNQAEK	140		35%	12%	14%	4%	51%	16%	o1204-1223i1345
YHR080C	444	RLSFNSSLNGLTNNDEPEYEDREPR/EMSKKFLNRR	73		29%	6%	12%	4%	59%	3%	o1204-1223i1345
YHR101C	103	GLRKLDFLEFETEFVSLQRYIRR/SSTAIKFEKV	124		16%	3%	8%	2%	76%	4%	20/o276-295i335
YHR149C	31	LISVTSTVSVSDASYAYARLTRR/DDSDSSSSSA	52		24%	5%	10%	1%	65%	5%	o72-97i734
YHR188C	276	HYEMDARIEMVTHVPSALARGER/PIPKPLDGNT	59		32%	8%	24%	11%	44%	15%	20/o548-568i610
YHR204W	69	DNYLEHGYPYDEVKPI SCVPKRR/NFEDPTDQGT	141		21%	9%	21%	12%	59%	12%	21/(no TMs)796
YHR204W	511	DVIFSTEAHPMWLPQEVRSNYKR/NAKFNNNSVYS	196		31%	7%	8%	1%	60%	8%	21/(no TMs)796
YHR204W	764	ILPRHGSVQIKKHSPLVLT SNGTR/EDEFKMDGI	82		14%	4%	18%	13%	69%	12%	21/(no TMs)796
YIL005W	138	SGYIKSFTEPTRTKESLIAFARR/ESMDPNNLDT	231		27%	2%	10%	3%	63%	4%	28/o652-672i701
YIL014W	61	KLNNHASERTAVESSAFNWIEKR/QHQVRSENLM	164		31%	11%	8%	2%	61%	12%	31/(no TMs)630
YIL014W	387	NEQTFQYCIEMLNEVEPSAQERR/FIGSRLVFDS	122		23%	4%	8%	4%	69%	8%	31/(no TMs)630
YIL014W	557	EDDFGREPEYFKSRYGDISKLKR/IYDASLNVEG	763		23%	4%	8%	3%	69%	7%	31/(no TMs)630
YIL037C	400	NLSTLYPIPYQLTKLINQKIFKR/EPEKIDDKKV	408		27%	1%	15%	6%	58%	5%	o20-43i318-336o421-440i631-654o656
YIL077C	25	KEEEQQYQGNGKGMENELPFMKR/PWFKKAYENA	401		18%	6%	19%	17%	63%	22%	o67-86i98-119o320
YIL077C	190	LDGSGPSRWSMYFYITYQNPFERR/LPDPKVKLQQ	74		24%	8%	22%	15%	54%	18%	o67-86i98-119o320
YIL117C	8	*TVITIAKR/GLPKLTSTTS	169		21%	4%	12%	3%	67%	3%	o77-98i318
YIL123W	34	*HVDVHQEDAHQHKR/AVAYKYVYET	386		38%	7%	9%	2%	53%	7%	20/(no TMs)475
YIL136W	90	SLKDALLGVSQKAREEAPKVTKR/VISPEEDAQT	406		50%	16%	6%	4%	44%	12%	19/(no TMs)393
YIR007W	373	PEWKPDTCIWKLHGVWEIQNGKR/PVLLKPNYFS	323		21%	6%	7%	3%	72%	9%	(no TMs)764
YIR039C	51	FPVQKLADIINICTQDVTVFVKR/NEVLNNTVIN	653	1	65%	21%	4%	1%	32%	22%	22/(no TMs)537
YJL012C	399	TELTVMVREDNFDGVDRTHKNWRR/TDIGVDWPFK	259		40%	9%	13%	3%	46%	12%	(no TMs)648
YJL034W	309	KAFKKKHGIDVSDNNKALAKLKR/EAEKAKRALS	954	1	54%	34%	6%	1%	41%	33%	i12-29o682
YJL068C	47	TSMNVNIYLPKHYYAQDFPRNKR/IPTVFYLSGL	650		47%	8%	13%	3%	40%	11%	(no TMs)299
YJL079C	173	ASSSDSDLDFASSVLAHNKRR/ALHKDTPALS	277		27%	1%	8%	1%	65%	2%	20/(no TMs)299
YJL132W	307	NTNEAQAEELRLCANLPPVSQKR/INARPLVSSL	114		38%	8%	13%	3%	50%	10%	26/(no TMs)750
YJL158C	63	GAAEYTTTFGIAVQAITSSKAKR/DVISQIGDQG	543	1	60%	25%	5%	1%	34%	25%	22/(no TMs)(209-GPI)o227
YJL159W	71	VQPISTTSSASSAATTASSKAKR/AASQIGDQGV	434		31%	2%	6%	4%	63%	4%	24/(no TMs)(280-GPI)o310
YJL160C	61	TEYRTSFGLAVIPFTVTESVKRR/NVISQINDGQ	617		19%	6%	6%	3%	74%	8%	24/(no TMs)180
YJL216C	415	FKNRKIEQIKCVEGTGYAAIKR/DYGEDSEKMK	162		21%	7%	12%	2%	67%	6%	(noTMs)581
YJR038C	106	VKPTSASFGGIRLSQLTMIMERR/ATPTCQDPSL	158		21%	5%	12%	3%	67%	3%	o20-42i54-72o120
YJR075W	292	KPGHPVLREIISRIIETTLQRKR/DDQLNVNLRN	362		24%	4%	14%	4%	62%	4%	29/(no TMs)396
YJR126C	327	VTEENGLHYPYGPARNSYIRR/DSVNYFSDTP	155		24%	4%	16%	7%	60%	6%	i90-109o811
YJR126C	363	TTPGYPSKSDTEHMSPVGRVPR/IPSVMSARD	86		17%	6%	31%	9%	52%	9%	i90-109o811
YJR151C	652	LSATSTFTSAVSSSKAADFFTR/STVSAKSDVS	73		24%	3%	5%	3%	71%	4%	20/(no TMs)(1127-GPI)o1161
YJR151C	1077	VSCSAGACTTLTTVDAGISMFR/TGLSITQTTV	64		45%	12%	5%	3%	49%	15%	20/(no TMs)(1127-GPI)o1161
YKL128C	107	SGDEHGEWLDKLTPLGKDQVRR/TGSNVLLPMA	179		26%	1%	15%	5%	60%	4%	(no TMs)295
YKL163W	66	SFGIAIEAVATSASSVASSKAKR/AASQIGDQGV	434		39%	12%	5%	3%	56%	16%	19/(no TMs)325
YKL164C	62	DYSSTFGIAVEPIATTASSKAKR/AAAISQIGDG	434		22%	4%	6%	4%	72%	8%	19/(no TMs)341
YKL201C	661	MATLNI TELRDYITS DENKNHKKR/VPTDIDLKDL	462		29%	11%	8%	3%	63%	10%	i28-45o1178
YKL201C	1039	DKDPIIVYEDYAYAKLLEERKRR/EKKKKEEEEEK	90		19%	11%	6%	3%	75%	9%	i28-45o1178
YKL212W	218	TPIVLGLITRRSIFRAGTRYFRR/GVDDKGNVGN	279		22%	5%	8%	3%	69%	7%	o522-543i555-573o623

Anhang

YKL212W	467	DAVSVAISGTGALKTDFTRTGKR/TRLGAFNDFL	56		21%	4%	9%	5%	70%	8%	o522-543i555-573o623
YKL220C	55	LVTNACTRIFQKVTWEYTSKSKR/SSPVCSEYPA	434		11%	2%	13%	4%	76%	3%	21/o165-184i236-256o276-297i318-339o351-370i377-395o401-419i711
YKR013W	203	ASSTQSSSSDFSTSMVNEHNTKR/ALHKDTGSLT	355		23%	5%	6%	4%	71%	8%	19/(no TMs)329
YKR020W	20	*AEQISHKKSLRVSSLNKDRR/LLREFYNLE	93		21%	5%	11%	2%	68%	4%	(no TMs)164
YKR102W	131	VPVKRGVKLCSDNNTLSSKTEKR/ENDDCDQGAA	155		23%	5%	18%	7%	59%	4%	25/(no TMs)1169
YLR057W	188	QIKDIFMKSWNQEQLLLKSNLRR/ESTWPIDLID	265		51%	22%	8%	6%	40%	16%	(no TMs)849
YLR057W	668	NLDLVGIVKVSNYGKGGNDQAKR/NVLSKDGITE	434		28%	23%	10%	5%	62%	19%	(no TMs)849
YLR083C	249	YHDRGEGNYRVVGVIVPEVSIKR/SSPGTCETTG	88		20%	8%	13%	3%	67%	5%	23/o305-327i371-395o401-426i438-462o474-496i521-548o560-583i595-615o627-657i667
YLR084C	1052	NASSISSVSAIYTDIESNTLNKR/GINNVNNGSI	59		32%	3%	11%	2%	57%	2%	21/o1161-1188i1220
YLR120C	29	*IPAANKR/DDDSNSKFKV	317	1	53%	35%	16%	12%	32%	23%	22/(no TMs)(542-GPI)o569
YLR120C	66	KLYGDSLENVGSDDKPEVRLKLR/ADGYEIIIT	95	0	24%	17%	6%	2%	71%	18%	22/(no TMs)(542-GPI)o569
YLR120C	546	QTASFGNLTSTASATSTSSKR/NVGDHIVPSL	102	0	24%	18%	6%	2%	70%	19%	22/(no TMs)(542-GPI)o569
YLR121C	46	DGKYVKIPFTKKNNGDNGELSKR/SNGHEKFLA	54	1	74%	13%	5%	1%	22%	13%	21/(no TMs)(483-GPI)o508
YLR121C	474	TRSTTTKKTNTTTAKSTHKSQR/ALQRAATNSA	434	0	15%	6%	4%	2%	80%	6%	21/(no TMs)(483-GPI)o508
YLR155C	308	ATGSIVNEQLYEEYGPIVHSRR/TADGTVPDD	281		12%	2%	10%	1%	78%	1%	19/(no TMs)362
YLR157C	308	ATGSIVNEQLYEEYGPIVHSRR/TADGTVPDD	281		12%	2%	10%	1%	78%	1%	19/(no TMs)362
YLR158C	308	ATGSIVNEQLYEEYGPIVHSRR/TADGTVPDD	281		12%	2%	10%	1%	78%	1%	19/(no TMs)362
YLR160C	308	ATGSIVNEQLYEEYGPIVHSRR/TADGTVPDD	281		12%	2%	10%	1%	78%	1%	19/(no TMs)362
YLR207W	266	SDFESGLLGKGLNSVPSSTVRKR/TTRPDIGSPF	158		44%	13%	7%	2%	49%	13%	21/o768-789i833
YLR207W	269	ESGLLGKGLNSVPSSTVRKRTR/PDIGSPFIAQ	118		47%	3%	18%	5%	35%	4%	21/o768-789i833
YLR213C	127	EGNGPNKANTKVGVEGELNSKR/IIHYAKFLVT	295		23%	7%	6%	3%	71%	10%	21/(no TMs)422
YLR250W	138	MRLLDVDDNTKITKEEYLQFAKR/GNKFPDLGVG	181		22%	2%	12%	3%	66%	4%	21/(no TMs)234
YLR299W	441	LPKHIEEVLDFEWALKAVKSIKR/NSQDGNFKTL	88		21%	5%	11%	6%	68%	10%	i12-33o660
YLR299W	536	QFNKNSFELAPSIYNFPEPKR/PLSSTAPTIV	132		18%	5%	20%	5%	62%	5%	i12-33o660
YLR299W	653	SVVNAIRNVRGEWHAVSDYWRKR/GISSVY*...	181		15%	3%	8%	3%	77%	3%	i12-33o660
YLR300W	39	*VPARDPSSIQFVHEENKRR/YDYDHGSLG	346		32%	4%	8%	2%	60%	4%	20/(no TMs)448
YLR337C	282	VSPPSSEVPAGGLPFLAEINARR/SERGAVEGVS	94		55%	12%	9%	2%	36%	12%	(noTMs)817
YLR337C	285	PSSEVPAGGLPFLAEINARRSER/GAVEGVSTTK	54		39%	18%	13%	6%	48%	16%	(noTMs)817
YLR337C	485	SSSSAVTPGGPLPFLAEIQKRR/DRFVVGDDT	181		32%	13%	9%	5%	59%	17%	(noTMs)817
YLR353W	119	QFHDTQDNNIPSNNTVTRPTSLRR/DNEDTVPLRN	115		23%	4%	17%	9%	60%	12%	o516-536i576-600o603
YLR396C	582	YDRSILFHNYSSQPFFILSREPR/VSQTEDLIEQ	58		28%	3%	10%	7%	62%	10%	o640-657i691
YLR396C	624	ESIWVPGTITKINASIKSNRR/SIDGSGNFTFH	203		29%	4%	13%	7%	58%	11%	o640-657i691
YML012W	77	VGDGQKLNLFVRDSVGNERYRRR/DFAGDVRVAF	707		43%	5%	10%	1%	47%	5%	21/o182-201i211
YMR006C	61	DDTSLVRNASGLSTAETDWLKKR/DAYTKEALHS	64		40%	12%	11%	1%	48%	13%	20/(no TMs)(680-GPI)o706
YMR008C	59	DDINLVREASGLSDNETEWLKKR/DAYTKEALHS	64		31%	14%	9%	1%	60%	15%	23/(no TMs)(637-GPI)o664
YMR101C	84	VGPVPEHVSFIMDGNRRYAKSRR/LPVKKGHEAG	221		26%	3%	15%	4%	59%	3%	i32-51o274-292i343
YMR101C	222	SRNDMLHTIRDSVEDHLENKSPR/INIRKFTNKM	43		35%	7%	9%	2%	56%	6%	i32-51o274-292i343
YMR200W	205	PTETLNPTSSATSTDDPSSNKRR/SLRSLVRRSL	138		34%	4%	7%	4%	59%	6%	25/(no TMs)(235-GPI)o256
YMR200W	226	KRSLRSLVRRSLENRHKTNAIKR/QNTSFLTSNA	130		49%	8%	7%	2%	43%	6%	25/(no TMs)(235-GPI)o256

Anhang

YMR244W	46	*SPSIQMTENTNQDHHEHAKR/GGTCAFFNYD	440		60%	12%	13%	2%	28%	10%	26/(no TMs)355
YMR244W	328	RNGAKANIEVFDVGGSSSSKGR/ELNPLDVITT	593		54%	16%	16%	5%	30%	12%	26/(no TMs)355
YMR305C	28	*AVRHKHEKR/DVVTATVHAQ	533	1	54%	33%	7%	3%	38%	30%	19/(no TMs)389
YMR306W	124	IGFQNMRLHGKARQNRVKKMAKRR/GVSIKEQIKQ	141		39%	10%	9%	4%	52%	12%	o338-357i377-395o415-437i444-464o503-526i547-567o579-604i1215-1234o1264-1288i1358-1378o1384-1402i1473-1495o1515-1539i1551-1572o1584-1606i1649-1670o1676-1692i1713-1733o1785
YMR306W	482	FFAVRPLGGLFRPYLNKDKKHRR/YISSQTFTAS	346		38%	6%	11%	5%	51%	9%	o338-357i377-395o415-437i444-464o503-526i547-567o579-604i1215-1234o1264-1288i1358-1378o1384-1402i1473-1495o1515-1539i1551-1572o1584-1606i1649-1670o1676-1692i1713-1733o1785
YNL012W	182	SEEEIVTNAKKEYNDNDLKMKKR/QGGSLLITSSS	300		22%	3%	11%	3%	67%	2%	18/o374-396i631
YNL012W	217	YEQIEEIMNSIEEIPEDYMITKR/NLNPLARLKK	203		20%	5%	12%	2%	68%	3%	18/o374-396i631
YNL066W	44	. /AVAVTYVYET	386		35%	11%	6%	3%	59%	8%	23/(no TMs)420
YNL158W	41	*ESILYKVPHPNFKKPR/DSSTYARDVN	51		22%	6%	12%	3%	66%	3%	24/o168-187i198
YNL160W	175	PIVVTEDSAYAI PVANNKNATKR/GVLSVTSDKL	181		26%	2%	6%	2%	68%	4%	20/(no TMs)(322-GPI)o354
YNL238W	79	EHDVRGLPNHYVFSKELLKLGKR/SSLEELQGDN	59		36%	6%	20%	9%	44%	14%	24/o676-699i814
YNL238W	108	QGDNDHILSVHDLFPRNDFKFR/LPVPAPPMDS	84		29%	10%	8%	5%	63%	14%	24/o676-699i814
YNL238W	503	TEETLESVITISEKSLQDANFKR/IEHVTVTVDI	571		27%	7%	14%	10%	60%	16%	24/o676-699i814
YNL283C	266	STATVTSTPSSTSIGTSTHYTTR/VVTQSVVSQA	44		24%	7%	7%	4%	70%	11%	24/o322-347i503
YNL283C	287	TRVVTQSVVSQANQASTIFTR/TSVYATVSST	57		20%	5%	6%	4%	74%	9%	24/o322-347i503
YNL323W	51	EEDVDASEFEDEVKPVTRTKNRR/PKEDAFTQQR	149		44%	12%	18%	7%	38%	16%	o76-95i373-398o414
YNL327W	292	SISNTVELTYPDEGTSVRLGKR/DTSTTLASEL	74		19%	6%	9%	0%	72%	5%	21/(no TMs)(1019-GPI)o1041
YNR019W	26	KDLLENEQFLRIQKLNADAGKR/QSITVDDEGE	165		41%	6%	6%	3%	53%	4%	o218-239i259-281o293-312i324-343o404-422i434-459o479-499i567-586o592-610i622-640o642
YNR059W	279	PQELWFLDVKDMLNDEYLARFKR/FSNKWLAITF	933		22%	3%	6%	1%	72%	3%	31/(no TMs)580
YNR060W	33	*PPSKTSLINTHERR/SIYSCYVGLR	217		24%	8%	7%	1%	68%	9%	19/o157-180i229-249o269-290i311-332o344-364i371-388o394-412i719
YOL003C	248	RIYPESRSFQNKDAEEHLTKKR/PRFDELVNFP	48		31%	7%	23%	13%	46%	19%	o6-29i41-58o126-143i164-186o378
YOL003C	250	YPESRSFQNKDAEEHLTKRPR/FDELVNFPYD	64		31%	6%	22%	11%	48%	16%	o6-29i41-58o126-143i164-186o378
YOL031C	310	LISKILKADMYENDTNLILFKR/NAENWSSNLQ	105		15%	2%	13%	2%	72%	3%	19/(no TMs)421
YOL046C	165	FGDMGLYENACGDSLGTATRSER/SSSLNVDENE	54		28%	10%	5%	2%	66%	8%	(no TMs)(204-GPI)o224
YOL075C	352	LQLQAESYISNATEIQIQNMTTR/LPFWKQVTVL	50		28%	7%	11%	1%	62%	7%	o375-397i409-428o469-485i492-517o523-542i554-572o601-625i1038-1059o1121-1141i1148-1166o1178-1196i1208-1226o1266-1285i1294
YOL105C	170	STSYMPSTTSSLSSAQISSTTRR/TSTDMKSSEM	63		27%	2%	7%	3%	66%	3%	39/o381-405i556
YOL105C	346	TSVVTKAIVTSDQHQETIFVTR/TSVVERSEV	67		27%	8%	6%	3%	67%	11%	39/o381-405i556
YOL105C	352	AIVTSDQHQETIFVTRTSVVER/SSEVATATAA	54		28%	6%	5%	2%	67%	8%	39/o381-405i556
YOR041C	110	RSTSVVFEISSLIDSGTLLFSRR/EESLIPSFSS	288		15%	3%	30%	9%	55%	9%	(no TMs)143
YOR115C	86	GDVISSDSNVTSSENGINNMIKR/LKIEEHHTVD	137		28%	4%	5%	4%	66%	2%	(no TMs)268
YOR219C	162	SKLLPTKITRPKTSAGDSSLGKR/SFSIENVLNG	59		61%	6%	5%	3%	34%	5%	i120-139o931

Anhang

YOR238W	253	KNALSLFSSDWYATKDRLLTKKR/SRNPFRNTAP	48		33%	6%	23%	6%	44%	9%	(no TMs)303
YOR365C	313	ATSGSPTVVYRNKEVLSISVQKR/SLNSKIISAS	375		23%	8%	8%	2%	69%	9%	25/o203-222i234-259o374-396i417-436o456-478i517-537o543-561i573-595o607-634i703
YOR384W	26	*ASTKKR/TQWDQIAIDA	96		29%	9%	23%	10%	49%	7%	20/o162-184i205-225o237-254i266-288o350-368i375-392o398-417i694
YOR389W	377	IDYRYLVSGINRTYIAPDPNNRR/LLDEGMTWEK	203		21%	9%	10%	3%	69%	7%	24/(no TMs)624
YPL170W	46	GASNTNDSNKGSEPVVAGNFFPR/TLSKFNHDD	56		25%	4%	22%	10%	53%	8%	(no TMs)152
YPL187W	103	LDKREAEAWHWLQLKPGQPMYKR/EAEAEAWHWL	663	2	12%	5%	56%	21%	32%	16%	20/(no TMs)165
YPL187W	124	KREAEAEAWHWLQLKPGQPMYKR/EADAEAWHWL	663	2	12%	6%	50%	28%	38%	22%	20/(no TMs)165
YPL187W	145	KREADAEAWHWLQLKPGQPMYKR/EADAEAWHWL	663	2	12%	7%	49%	29%	39%	23%	20/(no TMs)165
YPL207W	155	EKAVNYPVTVDFTEVFRKPTKKR/SSTPKVFSKN	96		49%	14%	17%	5%	34%	17%	i7-39o810
YPR026W	449	GLAKMIALNSKGNYEKLLSSHKR/AWYDLYNDAF	160		26%	5%	12%	3%	62%	7%	i47-69o1211
YPR079W	44	HVYNGPGLSHEANEHRASHKQKR/TLANPDKPKS	482		50%	12%	5%	3%	45%	10%	i7-25o237-257i381
YPR149W	160	LMSVFNMISNGAFSGSFSKRRR/TGQVGVPTIS	208		65%	6%	7%	2%	28%	6%	o12-28i40-59o65-93i125-146o173

