INTRAMEMBRANE-CLEAVING PROTEASES (I-CLiPs)

PROTEASES IN BIOLOGY AND DISEASE

INTRAMEMBRANE-CLEAVING PROTEASES (I-CLiPs)

Edited by

NIGEL M. HOOPER
University of Leeds,
United Kingdom

and

UWE LENDECKEL
Otto-von-Guericke Universität,
Magdeburg, Germany

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-6310-7 (HB)
ISBN 978-1-4020-6311-4 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

CONTENTS

PREFACE

In recent years a growing number of proteases have been identified that catalyse peptide bond hydrolysis in the plane of the cellular membrane. These so-called 'intramembrane-cleaving proteases' (I-CLiPs) are involved in a diverse range of cellular processes, including cell regulation, signalling and protein processing. Some I-CLiPs play critical roles in diseases such as Alzheimer's and viral infection. The aim of this book is to provide an update on this emerging group of unusual but important proteases for both the specialist and those with a broader interest in proteases.

The book begins with a chapter by Rob Rawson and Wei-ping Li on the first I-CLiP to have been recognised as such, the Site-2 protease. This protease, which has the prototypical zinc binding His-Glu-Xaa-Xaa-His motif, plays a crucial role in mammalian lipid metabolism and the unfolded protein response, and therefore is key to both normal and disease processes. Although it is now 10 years since Site-2 protease was discovered, several questions about this protease remain unanswered and these are highlighted.

In Chapter 2, Todd Golde and colleagues introduce the signal peptide peptidases (SPPs), focusing on SPP and SPPL3. These multipass membrane proteins with critical aspartic acid residues in their active site, function as proteases without the need for additional cofactors. Only a few endogenous substrates for SPP have been identified and its biological role is largely unknown. The possibility that the SPPs may be novel antiviral drug targets in humans and represent a novel drug target for major human pathogens, such as malaria, is discussed. This family of GXGD-type intramembrane aspartate proteases, is further expounded upon in Chapter 3 by Harald Steiner and Christian Haass. They focus on the role that SPP plays to clear the ER membrane of signal peptides of secretory proteins, and the role that SPPL2a and b may have in cleaving tumour necrosis factor-α to release an intracellular domain that triggers interleukin-12 signaling.

The intramembrane serine proteases, the rhomboids, are the attention of Chapter 4 by Sinisa Urban. These proteases play key roles in a range of cell communication events, including tyrosine kinase signalling during animal development and quorum sensing during bacterial growth. The first high resolution crystal structure of a rhomboid protease has recently been reported providing new insights into the

structure and mechanism of action of these proteases. In Chapter 5, Elke Pratje continues with a description of the rhomboid family members, Rbd2 and Pcp1, in yeast. Rbd2 is associated with the Golgi, but its function and substrates are unknown, while Pcp1 is located in the inner mitochondrial membrane where it catalyses the second step in the proteolytic processing of cytochrome c peroxidase. PcP1 also affects the morphology of mitochondria by acting on the dynamin-related GTPase, MgM1.

The final two chapters discuss the role of the presenilin/γ-secretase complex in the proteolytic processing of the Alzheimer's amyloid precursor protein and the developmental protein Notch, respectively. In Chapter 6 Michael Wolfe describes the role of the amyloid-β peptide in Alzheimer's disease and the key part played by the γ-secretase complex in this process. The γ-secretase complex consists of four different integral membrane proteins, the presenilins, nicastrin, Aph-1 and Pen-2. Two critical aspartic acid residues in the presenilins constitute the active site of this I-CLiP. The role of the subunits in the maturation of the complex and in the recognition of substrates is discussed, along with the potential for inhibitors and allosteric modulators of γ-secretase activity as potential Alzheimer's disease therapeutics. In Chapter 7, Raphael Kopan and colleagues describe how the convergence of previously independent fields of research led to deciphering the proteolytic mechanism for Notch activation and the role of γ-secretase in its regulated intramembrane proteolysis.

The study of I-CLiPs has emerged as an exciting research area in cell biology, and we trust that this volume in the *Proteases in Biology and Disease* series will prove to be a timely and valuable source of information on these proteases. Finally, we would like to thank all the authors for their scholarly contributions.

Nigel M. Hooper and Uwe Lendeckel

CONTRIBUTORS

Robert B. Rawson and Wei-ping Li Departments of Molecular Genetics and Cell Biology, University of Texas Southwestern Medical Center, Dallas TX, USA

Todd E. Golde, Criag Zwizinski, and Andrew Nyborg Department of Neuroscience, Mayo Clinic College of Medicine, Mayo Clinic Jacksonville, Jacksonville, Florida, USA

Harald Steiner and Christian Haass Laboratory for Alzheimer's and Parkinson's Disease Research, Department of Biochemistry, Adolf-Butenandt-Institute, Ludwig-Maximilians-University, Munich, Germany

Sinisa Urban Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Elke Pratje University of Hamburg, Biozentrum Klein Flottbek und Botanischer Garten, Ohnhorststr. 18, D-22609 Hamburg, Germany

Michael S. Wolfe Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Ma. Xenia G. Ilagan, Dilip Chandu, and Raphael Kopan Department of Molecular Biology and Pharmacology, Washington University School of Medicine, St. Louis, Missouri, USA 63110

CHAPTER 1

THE SITE-2 PROTEASE AT TEN

ROBERT B. RAWSON[1] AND WEI-PING LI[2]
[1]*Departments of Molecular Genetics and* [2]*Cell Biology, University of Texas Southwestern Medical Center, Dallas TX, USA*

Abstract: The site-2 protease (S2P) is a highly hydrophobic integral membrane protean required for cleavage of various mambrane-bound transcription factors within a membrane-spanning helix. S2P was the first intramembrane-cleaving protease to be recognized but more has been learned about other such proteins. Fundamental questions about the role and function of S2P remain unanswered. S2P plays a crucial role in mammalian lipid metabolism and the unfolded protein response. Thus, finding the answers has implications for our understanding of human health and disease. Recent advances with rhomboid proteins and gamma secretase indicate that the technical challenges to getting the answers can be overcome

Keywords: ATF-6, S2P, SREBP, Regulated intramembrane proteolysis (Rip)

1. INTRODUCTION

A decade after its discovery, the Site-2 protease (S2P) continues to resist efforts to shed light on some of the most basic questions concerning this enigmatic protein. Despite being the first intramembrane-cleaving protease (I-CLiP; (Wolfe *et al.*, 1999)) to be recognized as such, more has been learned about other, more recently-discovered I-CLiPs.

The evidence that S2P is indeed a protease, for example, rests solely on extensive genetic analysis (Rawson *et al.*, 1997; Zelenski *et al.*, 1999) and on its primary amino acid sequence that has an histidine-glutamate-(any amino acid)$_2$-histidine (His-Glu-*X*-*X*-His) motif. These residues are characteristic of the active site of many metalloproteases (Rawlings and Barrett, 1995). Direct evidence of proteolytic activity, such as cleavage of a substrate *in vitro* by purified S2P, are still lacking. Notwithstanding these caveats, it seems certain that S2P is indeed a protease.

With the available data and in the absence of any evidence to the contrary, we proceed on this widely-held assumption. Indeed, S2P is included in the MEROPS database of proteolytic enzymes ((Rawlings *et al.*, 2006); clan MM, family M50A,

N. M. Hooper and U. Lendeckel (eds.), Intramembrane-Cleaving Proteases (I-CLiPs), 1–15.

peptidase M50.001) and proteins highly similar in sequence to S2P are found in nearly every genome that has been sequenced. The widespread occurrence of S2P family members suggests that an S2P protein was present in the last common ancestor of current life forms (Kinch *et al.*, 2006). Here we consider what *has* been learned about S2P, and what questions remain unanswered.

2. BACKGROUND

S2P was discovered as part of the effort to understand the global regulation of lipid metabolism in mammalian cells (Brown and Goldstein, 1999). One of the major systems regulating lipid metabolism is the sterol regulatory element binding protein (SREBP) pathway. The SREBPs are transcriptional activators of genes needed for lipid synthesis and uptake (*e.g.* fatty acid synthase (FAS) and the low density lipoprotein (LDL) receptor gene, respectively). In vertebrates, there are two distinct genes that encode three different isoforms, SREBP-1a, -1c, and -2 (Hua *et al.*, 1993; Yokoyama *et al.*, 1993). When cellular need for lipid rises, SREBPs are activated.

2.1. Membrane-bound Transcription Factors

An unusual feature of the SREBPs, given that they are transcription factors, is that they are made as precursors that are integral membrane proteins of the endoplasmic reticulum (ER), owing to the presence of two membrane-spanning helices. The precursor adopts a hairpin configuration such that both the amino- and carboxy-terminal domains are cytoplasmic (Hua *et al.*, 1995). Transcriptionally active SREBP is the amino-terminal fragment of the precursor. This fragment, lacking the membrane anchors, is free to enter the nucleus and bind to the sterol regulatory elements (SREs) in the promoters of target genes, resulting in their increased transcription.

The observed production of an active amino-terminal fragment from a membrane-bound precursor indicated that some protease (or proteases) cleaved the precursor to release active SREBP. Further, Wang *et al.* demonstrated that cleavage of the precursor was regulated by sterols (Wang *et al.*, 1994). This revealed part of the mechanism by which sterols regulate their own synthesis in mammalian cells: when sterols are in short supply, SREBPs are cleaved and the transcription of the genes of sterol synthesis and uptake is increased; when sufficient sterols are present, no cleavage occurs, and there is no increase in transcription or synthesis.

Initially, some hypothesized that this was the whole story. Gasic proposed that SREBP, by virtue of its two membrane-spanning helices, sensed the sterol content of the ER membrane directly, perhaps undergoing a conformational change that rendered it susceptible to proteolysis. This then would release the transcriptionally active amino terminus (Gasic, 1994). Further work, however, revealed a much more complex and intriguing mechanism of SREBP activation.

2.2. Two-step Cleavage

For example, Hua *et al.* found that amino acid sequences on each side of the first transmembrane domain of SREBP are required in order for regulated cleavage to occur (Hua *et al.*, 1996). On the luminal side, the arginine residue at amino acid 519, located in the solvent-accessible loop separating the two membrane-spanning helices, was required for cleavage. On the cytoplasmic side of the membrane, the motif Asp-Arg-Ser-Arg in the juxtamembrane region was also required (Hua *et al.*, 1996). When mutated to Ala-Ser, normal cleavage of SREBP 2 cannot occur.

Proteases which cleave extracellular domains of proteins in the juxtamembrane region had been identified previously, such as matrix metalloproteinase-3 (MMP3) that cleaves heparin-binding EGF-like growth factor (Suzuki *et al.*, 1997). None of the substrates identified, however, also required sequences on the opposite side of the membrane in order for cleavage to take place. Thus, either the enzyme that cleaved SREBP was highly unusual (for example, spanning the membrane with two active sites on different sides, or possessing a single site that traversed the membrane), or two separate proteases were needed to release the amino terminal domain of SREBP. These unanticipated findings made identification of the cleavage machinery all the more important.

2.3. Isolating S2P

In order to identify S2P, workers in the laboratory of Mike Brown and Joe Goldstein continued with a complementation approach that had been initiated in the laboratory of T.Y. Chang at Dartmouth (Hasan *et al.*, 1994). Chang and co-workers isolated mutant Chinese hamster ovary cell lines that failed to upregulate the transcription of the genes of cholesterol synthesis and uptake in the face of increased demand. These mutant cells therefore required medium supplemented with cholesterol and unsaturated fatty acid (oleate) in order to grow (Hasan *et al.*, 1994). One of these lines, designated M19, was selected for complementation cloning.

In order to circumvent potential pitfalls of complementation using cDNAs, high molecular weight genomic DNA from human cells was used to transfect the M19 cells (Fig. 1). Those cells that regained the ability to grow in unsupplemented medium had received a functional copy of the gene they lacked from the transfected human DNA. Of course, those cells had received other fragments of human DNA in addition to the one that restored the mutant function. Some of these irrelevant human sequences also were stably integrated into the genomes of the rescued cells.

Genomic DNA from the complemented mutant cells was used once more to transfect mutant M19 cells (Fig. 1). This served to partially ‘purify’ the human DNA sequences encoding the rescuing activity. All complemented cells must have in common the rescuing DNA. On the other hand, the extraneous human DNA sequences that were fortuitously integrated into the genomes of the rescued cells would vary from line to line. Each successive round of transfection reduced the likelihood that any given stretch of irrelevant, non-rescuing, human DNA would be

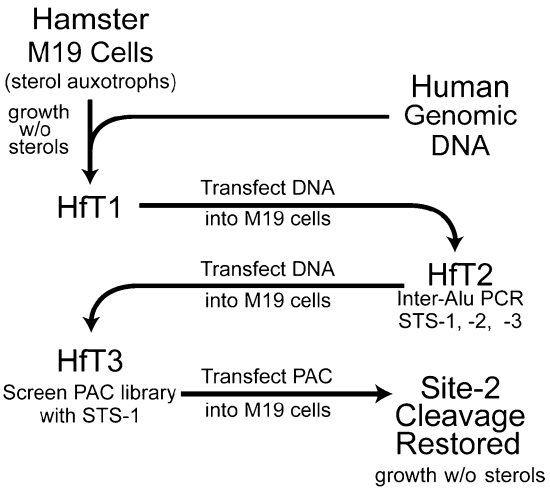

Figure 1. Steps in cloning S2P. Mutant **M19** Chinese hamster ovary cells fail to upregulate the genes of cholesterol and unsaturated fatty acid synthesis and require free cholesterol added to the medium in ethanolic solution in order to grow (Hasan and Chang, 1994). M19 cells were transfected with high molecular weight genomic DNA prepared from human cells. The transfected cells were then cultured in medium without added free cholesterol. Under these conditions, the mutant cells die while cells with wild-type function survive. One surviving clone, designated **HfT1**, was selected for further study. Genomic DNA from rescued HfT1 cells was transfected onto M19 cells and the selection repeated. The resulting clone was designated **HfT2** (Hasan *et al.*, 1994). The process was repeated using genomic DNA prepared from HfT2 cells to yield the **HfT3** cells. Inter-Alu PCR was used to amplify human sequences from HfT2 cells and unique human sequences were identified, designated **STS-1**, **-2**, and **-3** (Rawson *et al.*, 1997). Primers specific for STS-1, -2, and -3 were used to screen the rescued mutant cells. STS-1 was common to all rescued cells. STS-1 was used to screen a BAC library and the resulting human genomic DNA clone was transfected into M19 cells. The growth of the PAC-transfected mutant cells in the absence of added cholesterol confirmed that all sequences necessary for rescue were included within the PAC clone. Sequencing and BLAST searches lead to the identification of a human cDNA that also rescued mutant cells, S2P (Rawson *et al.*, 1997)

stably integrated in the rescued mutants. This made possible the identification of the rescuing gene.

The Alu family of ~300 base-pair dispersed middle repetitive sequences is abundant in the human genome (Schmid and Jelinek, 1982) but is not found in rodents. Thus Alu sequences serve as markers of human DNA. We employed the technique of inter-Alu PCR (Nelson *et al.*, 1989) to isolate probes for unique human DNA sequences present in the rescued mutant cells (Rawson *et al.*, 1997). One of these probes, STS-1, identified sequences unique to the human DNA found in each generation of the rescued mutant cells (Rawson *et al.*, 1997). This probe was used to screen a library of human genomic DNA cloned into a bacterial artificial chromosome (BAC) vector. The ~ 100 kb BAC clone thus isolated also rescued when transfected onto mutant M19 cells (Fig. 1). This confirmed that this clone encoded all sequences required for rescue of the mutant cells.

Sequence analysis of this clone lead to the identification of a human X chromosome gene (and the consequent isolation of its cDNA) that encoded a previously unknown protein. While its function was also unknown, the predicted amino acid sequence contained an His-Glu-*X*-*X*-His motif, which suggested that the unknown protein was a metalloproteinase.

2.4. A Protease?

As mentioned above, site-directed mutagenesis and transfection studies of the cDNA (including rescue assays - the restoration of SREBP cleavage to cells lacking S2P) demonstrated that both the histidines and the glutamate were necessary in order to restore cleavage of SREBP to mutant cells (Rawson *et al.*, 1997). This evidence strongly supported the conclusion that this protein was a protease that cleaved SREBPs and thus it was designated S2P (site-2 protease). Unusually for a protease, S2P is extremely hydrophobic; 43% of its residues are non-polar amino acids (Ala, Ile, Leu, Phe, Trp and Val). In fact, S2P and its homologues are the most hydrophobic proteases yet described. Its primary structure includes numerous sequences predicted to be membrane-embedded helices (Zelenski *et al.*, 1999). Biochemical studies confirmed that S2P is an integral membrane protein (Rawson *et al.*, 1997). Despite substantial effort, the difficulties encountered in reconstituting the activity of a membrane-imbedded enzyme against its membrane-imbedded substrate have not been overcome for S2P. No assay employing purified S2P or purified substrate has been reported. Thus, direct demonstration of proteolytic activity is lacking for S2P and it is unknown whether it requires any cofactors in order to function. This situation may not remain for long.

Successful efforts for the development of *in vitro* assays have been reported for Rhomboid family intramembrane proteases (Lemberg *et al.*, 2005; Urban and Wolfe, 2005) and gamma secretase (Fraering *et al.*, 2004; Hoke *et al.*, 2005). Thus it is possible to reconstitute the activity of at least some intramembrane proteases *in vitro,* and future efforts with S2P may prove availing.

3. STRUCTURE AND FUNCTION

3.1. Membrane Topology of S2P

In the absence of in vitro assays, much of what is known about S2P comes from studies in mammalian cells. Zelenski *et al.* determined the topology of S2P with respect to the membrane by a combination of epitope tagging and protease protection assays as well as access of portions of S2P to the glycosylation machinery in the lumen of the ER (Zelenski *et al.*, 1999). The proposed topological model supported by experimental evidence (Fig. 2) differs from that derived from algorithmic analysis (Lewis and Thomas, 1999).

S2P contains three regions of significant hydrophilicity: (1) residues 21–70 (numbered according to the human sequence); (2) residues 108–141, comprised

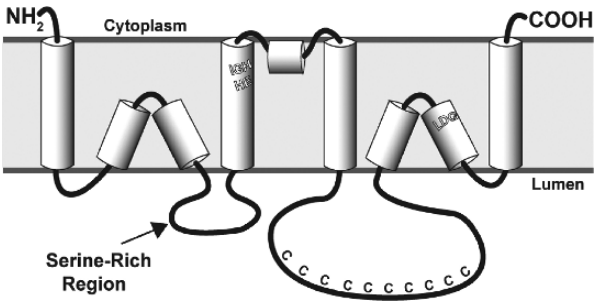

Figure 2. Membrane topology of S2P. This cartoon represents a model of the proposed membrane topology of S2P. The cytoplasmically-disposed amino- and carboxy-termini are indicated. The solvent-accessible loops (*e.g.* Ser-rich region, Cys-rich region) are in the lumen. The hydrophobic sequences are most likely embedded in the bilayer (grey box). The proposed location of the active site residues (HEIGH, LDG) are shown. The disposition of hydrophobic sequences within the bilayer is inferred from the location of the solvent-accesible sequences. Cylinders represent probable alpha helices. This model is based on data from (Kinch *et al.*, 2006; Zelenski *et al.*, 1999)

of multiple Ser residues of undetermined function; there are 26 Ser in this region of human S2P but varying numbers in other species, and (3) residues 258–446, including 12 Cys residues conserved in all S2P orthologues. These hydrophilic regions reside in the luminal space (Fig. 2). The amino- and carboxy-termini face the cytoplasm. Several of the predicted membrane-spanning helices do not cross the bilayer; the hydrophilic regions on either side of them can be glycosylated and thus must be disposed in the lumen (Zelenski *et al.*, 1999). If sequences on either side are luminal, then the hydrophobic sequence cannot span the bilayer. Instead, these sequences must 'dip' into the membrane with the peptide chain reemerging on the same side of the membrane as it entered.

Other intramembrane proteases are broadly similar. For example, the recently-solved crystal structure of the Rhomboid protein GlpG from *E. coli* shows that it also has helices that lie within the plane of the bilayer without crossing it as well as having additional membrane-associated sequences (Wang *et al.*, 2006; Wu *et al.*, 2006). Wang *et al.* and Wu *et al.* suggest that various of these may serve a gating function to regulate substrate access to the active site. It will be interesting to see if similar structural features are present in S2P, bearing in mind that the two protein families are completely unrelated by sequence.

The available data afford one more inference about the topology of S2P: if Asp_{467} is the third residue coordinating the zinc atom of the active site, then the carboxy-terminal region of S2P must bend back to the amino-terminal portion. This would enable membrane embedded regions of the protein that are distant from one another in the linear sequence of the protein to come together to form the active site.

3.2. Other Substrates

ATF-6α and β activate the transcription of genes of the unfolded protein response (Yoshida *et al.*, 1998). These membrane-bound transcription factors also need to be cleaved in order to release a transcriptionally active fragment from the precursor (Haze *et al.*, 1999; Yoshida *et al.*, 2000). Just as for the SREBPs, this processing requires the sequential action of S1P and S2P (Ye *et al.*, 2000b). Recently, Murakami *et al.* reported evidence that a protein related to ATF-6, which they refer to as old astrocyte specifically induced substance (OASIS; also known as cAMP responsive element binding protein 3-like 1), also requires S1P- and S2P- mediated proteolysis for its transcriptional activity (Murakami *et al.*, 2006).

They identified a potentially helix-destabilizing Pro residue (see below) at position 391 within the membrane-spanning helix, as well as a site that conforms to the S1P Arg-*X*-*X*-Leu consensus (Arg_{421}-Ser-Leu-Leu). These motifs are conserved in OASIS homologues from other organisms, including zebrafish (RBR, unpublished observations). It seems probable that OASIS is indeed a substrate of S1P and S2P. However, the compelling evidence that mutant cells that lack S1P and S2P fail to process OASIS has not been produced and explanations for these data other than those offered by Murakami *et al.* are possible.

3.3. Mechanism

3.3.1. Substrate selectivity

All known substrates of S2P are type II membrane-spanning domains (amino-terminus cytoplasmic, carboxy-terminus extra-cytoplasmic (e.g. the lumen of the ER)). This includes the SREBPs (once cleavage has taken place at site-1) and activating transcription factors (ATF) -6α and –β (see below). Restricted substrate orientation seems to be a general feature of intramembrane cleaving proteases. The presenilins, which are aspartyl proteases, cleave type I membrane-spanning helices while the other class of aspartyl I-CLiPs, the signal peptide peptidase family, cleaves type II (Weihofen *et al.*, 2002). Rhomboid proteases seem to be restricted to type I substrates.

An exception to this general specificity for substrate topology within the S2P family is SpoIV FB. This protease from *Bacillus subtilis* cleaves a type I substrate. Its predicted topology is opposite that of mammalian S2P and its active site is therefore disposed in the opposite sense relative to the cytoplasm (Rudner *et al.*, 1999). This may explain the alternative orientation of its substrates. A similar situation applies to the membrane topology of signal peptide peptidase (type II substrates) and presenilins (type I substrates) whose active sites are disposed in the opposite sense relative to the membrane (Friedmann *et al.*, 2004).

3.3.2. Active site

In the absence of in vitro data, clues to the mechanism of intramembrane proteolysis by S2P are indirect. In well-studied metalloproteases such as thermolysin,