Subcellular Biochemistry 66

David A. Dougan *Editor*

Regulated Proteolysis in Microorganisms

Springer

Regulated Proteolysis in Microorganisms

SUBCELLULAR BIOCHEMISTRY

Recent Volumes in this Series

Volume 52	**A Handbook of Transcription Factors** Edited by Tim Hughes
Volume 53	**Endotoxins: Stricture, Function and Recognition** Edited by Xiaoyuan Wang and Peter J. Quinn
Volume 54	**Conjugation and Deconjugation of Ubiquitin Family Modifiers** Edited by Marcus Groettrup
Volume 55	**Purinergic Regulation of Respiratory Diseases** Edited by Maryse Picher and Richard C. Boucher
Volume 56	**Water Soluble Vitamins** Edited by Olaf Stanger
Volume 57	**Aging Research in Yeast** Edited by Michael Breitenbach, Michal S. Jazwinski and Peter Laun
Volume 58	**Phosphoinositides I: Enzymes of Synthesis and Degradation** Edited by Tamas Balla, Matthias Wymann and John D. York
Volume 59	**Phosphoinositides II: The Diverse Biological Functions** Edited by Tamas Balla, Matthias Wymann and John D. York
Volume 60	**Adherens Junctions: From Molecular Mechanisms to Tissue Development and Disease** Edited by Tony Harris
Volume 61	**Epigenetics: Development and Disease** Edited by Kundu and Tapas Kumar
Volume 62	**The Eukaryotic Replisome: a Guide to Protein Structure and Function** Edited by MacNeill and Stuart
Volume 63	**vGPCR Signalling Complexes – Synthesis, Assembly, Trafficking and Specificity** Edited by Dupré J. Denis, Hébert E. Terence and Jockers Ralf
Volume 64	**Reprogramming Microbial Metabolic Pathways** Edited by Wang Xiaoyuan, Chen Jian and Quinn Peter
Volume 65	**Protein Aggregation and Fibrillogenesis in Cerebral and Systemic Amyloid Disease** Edited by Harris J. Robin
Volume 66	**Regulated Proteolysis in Microorganisms** Edited by David A. Dougan

For further volumes:
http://www.springer.com/series/6515

David A. Dougan
Editor

Regulated Proteolysis in Microorganisms

Springer

Editor
David A. Dougan
Department of Biochemistry
La Trobe Institute for Molecular Science
La Trobe University, Melbourne,
VIC, Australia

ISSN 0306-0225
ISBN 978-94-007-5939-8 ISBN 978-94-007-5940-4 (eBook)
DOI 10.1007/978-94-007-5940-4
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013934137

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

For Kaye and Matthew

Preface

All cells are composed of thousands of different proteins, each with a specific function. Collectively these proteins contribute to the proper function and maintenance of cells. As such it is not surprising, that regulating the integrity and concentration of each protein in the cell, not only under normal conditions but also under conditions of stress, is a fundamentally important biological process. For many years, it was believed that gene expression through regulated transcription and translation was primarily responsible for altering the abundance of individual proteins. Protein degradation was thought of only as a mechanism to recycle amino acids in a slow and somewhat non-selective manner. However, in the past 30 years, it has become evident that regulated protein degradation plays an important role in the cell's response to changing environmental conditions. Indeed in 2004 the world's attention was focussed on regulated proteolysis, when Aaron Ciechanover, Avram Hershko and Irwin Rose were awarded the Nobel Prize in Chemistry, for their fundamental discovery of Ubiquitin-mediated protein degradation. Although this research centred largely on regulated proteolysis in eukaryotes, it stimulated much research on related proteolytic systems in bacteria and other microorganisms. Indeed, during the past 10 years there have been numerous significant advances in this field.

The aim of this book is to highlight and compare the different proteolytic systems found in a selection of model and medically relevant microorganisms; from Gram-negative and Gram-positive bacteria (i.e. *Escherichia coli* and *Bacillus subtilis*, respectively), Archaea and *Saccharomyces cerevisiae*, to important pathogenic bacteria (i.e. *Mycobacterium tuberculosis*). The first chapter provides a general overview of the different proteolytic machines in *Escherichia coli*, focussing primarily on the mechanism of action of ClpAP and ClpXP (the two most extensively characterised AAA+ proteases) and the adaptor proteins that regulate substrate delivery to these machines. Chap. 2 takes an historical look at the first characterised, and most broadly conserved, ATP-dependent protease – Lon – and finishes with an elegant model for the allosteric-activation of protein degradation by this protease. Chap. 3 continues with a mechanistic analysis of the membrane bound ATP-dependent protease, FtsH. This chapter, also briefly examines the many

physiological roles this protease plays, primarily focussing on its role in the regulation of lipid synthesis. Many of these proteolytic machines also play important physiological roles during conditions of environmental or proteotoxic stress. The next four chapters focus on the physiological role of these machines in controlling a variety of stress response pathways in model and pathogenic strains of bacteria. The many and varied roles of regulatory proteolysis in the model Gram-positive bacterium, *B. subtilis*, are discussed in Chap. 4, while the two subsequent chapters (Chaps. 5 and 6) examine the importance of regulatory proteolysis in controlling distinct stress response pathways in *E. coli*. Chap. 5 describes the role these machines play in regulating the heat-shock response and the general stress response, while Chap. 6 centres on the role of proteolysis in controlling of the envelope stress response. Chap. 7 continues with the theme of regulatory proteolysis, focussing on its contribution to virulence in a number of pathogenic strains of bacteria. The next part (Chaps. 8 and 9) highlight the role of regulated protein degradation in *Saccharomyces cerevisiae*. Chap. 8 focuses on a single AAA+ protein, Cdc48 – as a key regulator of intracellular protein degradation in yeast. Cdc48 is not only an important regulator of a number of proteasome-mediated degradation pathways, including endoplasmic reticulum associated degradation (ERAD), but also plays a crucial role in autophagy and endolysosomal protein degradation. Chap. 9 highlights the contribution of the different AAA+ proteases to protein homeostasis in mitochondria, focussing primarily on the role of Lon, *i*-AAA and *m*-AAA in yeast but also touches on the role of ClpXP in the mitochondrion of higher eukaryotes. Finally, the novel "ubiquitin-like" protein modifications that were recently discovered in *Mycobacterium* sp. and Archaea are covered in the last two chapters (Chaps. 10 and 11, respectively). Both chapters discuss the current understanding of these types of protein modification and their possible link to proteasome-mediated degradation. In *Mycobacterium* sp., the process of protein modification has been termed pupylation as it involves the attachment of a novel prokaryotic ubiquitin-like protein (PUP) to a protein substrate. Chap. 10 provides a comprehensive biochemical description of pupylation, and includes a detailed structural analysis of several diverse components involved in this pathway, including the proteasome. Like *Mycobacterium* sp., Archaea also contain a functional proteasome and an "ubiquitin-like" protein modification system. However in contrast to bacteria (i.e. *Mycobacterium* tuberculosis) and Eukaryota, protein modification in Archaea involves the attachment of a novel protein known as small archaeal modifying protein (SAMP). The final chapter (Chap. 11) describes our current understanding of this modification process in Archaea, by SAMP (termed sampylation) and although the physiological role of this process is currently unclear, this chapter reflects on the possibility that sampylation is linked to regulatory proteolysis. Collectively, the book provides a comprehensive guide to regulatory proteolysis in distinct organisms. It illustrates the diverse mechanisms that AAA+ protease machines have evolved to selectivity recognise proteins for degradation in a spatial and temporal manner, while avoiding the unregulated degradation of the vast and concentrated pool of proteins in the cell.

As a final note, I would like to thank each of the authors, firstly for the quality of the chapters they have contributed, but also for their patience during the production

of this book. I would also like to sincerely thank the anonymous reviewers for their time, effort and invaluable expertise. I would also like to extend my thanks to Thijs van Vlijmen and Springer SBM for the opportunity to edit this book, it's been an incredible learning experience. My thanks also extend to all the members of my laboratory for their patience during the production of this book – undoubtedly, you will soon be wishing I was editing another one.

David A. Dougan

Contents

Contributors

Sarah E. Ades Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA

Sarah E. Barchinger Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA

Graduate program in BMMB, The Pennsylvania State University, University Park, PA, USA

Lone Brøndsted Department of Veterinary Disease Biology, Faculty of Life Sciences, University of Copenhagen, Frederiksberg C, Denmark

Alexander Buchberger Department of Biochemistry, Biocenter, University of Würzburg, Würzburg, Germany

K. Heran Darwin Department of Microbiology, New York University School of Medicine, New York, NY, USA

David A. Dougan Department of Biochemistry, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, VIC, Australia

Dorte Frees Department of Veterinary Disease Biology, Faculty of Life Sciences, University of Copenhagen, Frederiksberg C, Denmark

Eyal Gur Department of Life Sciences, Ben-Gurion University of the Negev

Department of Life Sciences, The National Institute for Biotechnology in the Negev, Beer-Sheva, Israel

Hanne Ingmer Department of Veterinary Disease Biology, Faculty of Life Sciences, University of Copenhagen, Frederiksberg C, Denmark

Huilin Li Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, NY, USA

Brookhaven National Laboratory Biology Department, Brookhaven National Laboratory, Upton, Brookhaven, NY, USA

Julie A. Maupin-Furlow Department of Microbiology and Cell Science, University of Florida, Gainesville, FL, USA

Dimce Micevski Department of Biochemistry, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, Australia

Noël Molière Institute for Microbiology, Leibniz Universität Hannover, Hannover, Germany

Teru Ogura Department of Molecular Cell Biology, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto, Japan

Takashi Okuno Department of Material and Biological Chemistry, Faculty of Science, Yamagata University, Yamagata, Japan

Ralf Ottofueling Department of Biochemistry, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, Australia

Marie I. Samanovic Department of Microbiology, New York University School of Medicine, New York, NY, USA

Kaye N. Truscott Department of Biochemistry, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, VIC, Australia

Kürşad Turgay Institute for Microbiology, Leibniz Universität Hannover, Hannover, Germany

Wolfgang Voos Institut für Biochemie und Molekularbiologie (IBMB), Universität Bonn, Bonn, Germany

Linda A. Ward Department of Biochemistry, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, VIC, Australia

Part I
AAA+ Proteolytic Machines

Chapter 1
Machines of Destruction – AAA+ Proteases and the Adaptors That Control Them

Eyal Gur, Ralf Ottofueling, and David A. Dougan

Abstract Bacteria are frequently exposed to changes in environmental conditions, such as fluctuations in temperature, pH or the availability of nutrients. These assaults can be detrimental to cell as they often result in a proteotoxic stress, which can cause the accumulation of unfolded proteins. In order to restore a productive folding environment in the cell, bacteria have evolved a network of proteins, known as the protein quality control (PQC) network, which is composed of both chaperones and AAA+ proteases. These AAA+ proteases form a major part of this PQC network, as they are responsible for the removal of unwanted and damaged proteins. They also play an important role in the turnover of specific regulatory or tagged proteins. In this review, we describe the general features of an AAA+ protease, and using two of the best-characterised AAA+ proteases in *Escherichia coli* (ClpAP and ClpXP) as a model for all AAA+ proteases, we provide a detailed mechanistic description of how these machines work. Specifically, the review examines the physiological role of these machines, as well as the substrates and the adaptor proteins that modulate their substrate specificity.

E. Gur (✉)
Life Sciences Department, Ben-Gurion University of the Negev,
Beer-Sheva 84105, Israel

The National Institute for Biotechnology in the Negev,
Beer-Sheva 84105, Israel
e-mail: gure@bgu.ac.il

R. Ottofueling • D.A. Dougan (✉)
Department for Biochemistry, La Trobe Institute for Molecular Science,
La Trobe University, Melbourne 3086, Australia
e-mail: d.dougan@latrobe.edu.au

D.A. Dougan (ed.), *Regulated Proteolysis in Microorganisms*, Subcellular Biochemistry 66,
DOI 10.1007/978-94-007-5940-4_1,

General Introduction

The bacterial cytosol is a complex mixture of macromolecules (proteins, DNA and RNA), which perform a variety of different functions. Given that proteins play a central role in many of these important cellular tasks, their correct maintenance within the cell is critical for cellular viability, not only under normal cellular conditions but also under conditions of stress. As such, a bacterial cell contains a network of molecular chaperones and proteases (often referred to as the protein quality control (PQC) network) dedicated to maintaining homeostasis of protein folding. Chaperones function to protect functional proteins against unfolding and to refold misfolded and aggregated species. The role of proteases is to remove unwanted and hopelessly damaged proteins.

In the bacterial cytosol, protein degradation is performed mainly by a number of different ATP-dependent proteolytic machines. In general these machines are composed of two components, a peptidase and an unfoldase. Invariably, the unfoldase is a member of the AAA+ (ATPase associated with diverse cellular activities) superfamily and as such these molecular machines are commonly referred to as AAA+ proteases [1]. In Gram-negative bacteria, such as *Escherichia coli* there are generally five different AAA+ proteases (ClpAP, ClpXP, HslUV, Lon (also refereed to as LonA) and FtsH). In contrast most Gram-positive bacteria, such as *Bacillus subtilis*, contain up to seven different AAA+ protease (ClpCP, ClpEP, ClpXP, HslUV (CodXW), LonA, LonB and FtsH). Interestingly, in bacteria belonging to the Actinobacteria and Nitrospira phyla (e.g. in *Mycobacterium tuberculosis* (*Mtb*)) one or more of these AAA+ proteolytic machines is replaced by the proteasome (for a detailed review of this AAA+ machine, and its physiological role in *Mtb* please refer to Darwin and colleagues [2]). Regardless of their origin, these machines can be divided into two broad groups; those that contain the unfoldase and peptidase components on separate polypeptides (e.g. ClpAP, ClpCP, ClpEP, ClpXP and HslUV (CodXW)), and those that contain both components on a single polypeptide (e.g. LonA, LonB and FtsH) (see Fig. 1.1).

This review will focus on the "two-component" proteolytic machines, primarily those from *E. coli* (e.g. ClpAP and ClpXP), with a brief comparison to the equivalent machines (e.g. ClpCP and ClpXP) in the model Gram-positive bacterium, *B. subtilis.* However, for an extensive review on regulatory proteolysis in *B. subtilis* please refer to [3]. Likewise, for a detailed review on the "single polypeptide" proteases, i.e. Lon and FtsH please refer to [4] and [5], respectively.

Structure and Function of the "ClpP Containing" Proteases (ClpAP, ClpXP and ClpCP)

As mentioned above, bacteria contain a wide variety of different proteolytic machines, of which ClpXP is certainly the best-studied AAA+ protease [1]. ClpXP is known to play a number of critical roles in a wide variety of bacterial species, from the control of different stress response pathways in Gram-positive and Gram-negative

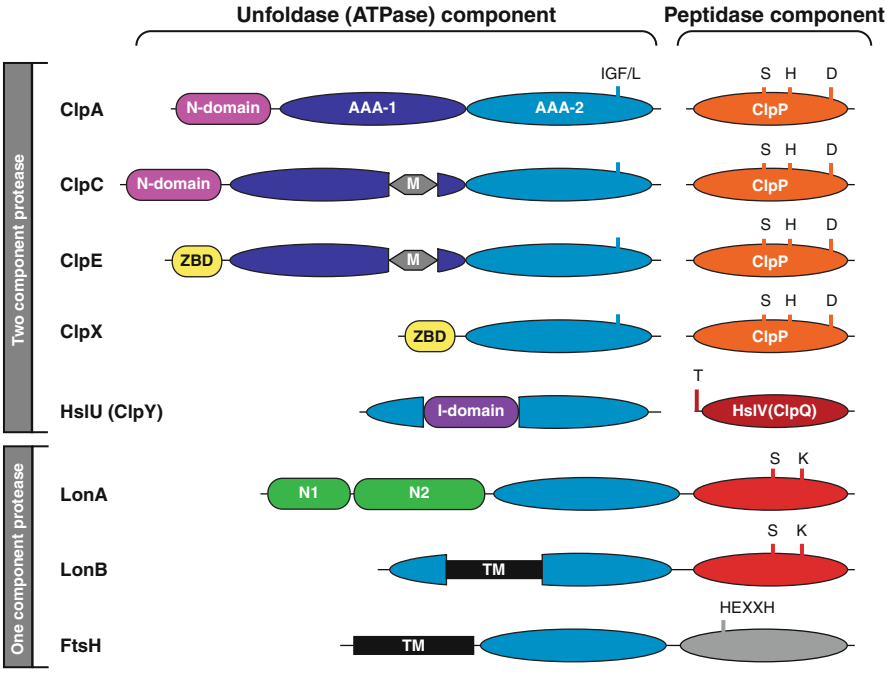

Fig. 1.1 Cartoon representation of the various different AAA+ proteases in bacteria. AAA+ proteases can be separated into two different groups. Two component proteases (e.g. ClpAP, ClpCP, ClpXP, ClpEP and HslUV) contain the unfoldase and peptidase components on separate polypeptides. One component proteases, contain the peptidase and the unfoldase on a single polypeptide (e.g. LonA, LonB and FtsH). The unfoldase component contains one or more AAA+ domains, responsible for ATP-dependent unfolding of the substrate. All unfoldase components also contain at least one accessory domains (e.g. ClpA and ClpC contain a conserved N-terminal domain (N-domain, *pink*), ClpC and ClpE contain a middle domain (M, *grey*), ClpE and ClpX contain a Zinc binding domain (*ZBD, yellow*), HslU contains an accessory domain inserted into the AAA+ domain (I-domain, *purple*), LonA contains two N-terminal domains unrelated to the N-domain of ClpA and ClpC (N1 and N2, *green*), while LonB and FtsH both contain a single transmembrane (TM) region), which serve various different functions (see main text for details). In the case of the ClpP-binding unfoldase components, the AAA-2 domain contains an IGF/L loop for interaction with ClpP. The protease components are responsible for cleavage of the unfolded substrate. In the case of ClpP, hydrolysis of the polypeptide is catalysed by the catalytic triad (S, H and D), while FtsH and HslV contain either a conserved HExxH motif or an N-terminal threonine (T) respectively, for peptide bond cleavage

bacteria (see [6, 7]) to the regulation of virulence through the degradation of key factors that control virulence (see [8]). ClpXP has also been shown to play an important role in regulating mitochondrial protein homeostasis (proteostasis) in eukaryotes such as worms [9, 10] and plants [11]. Surprisingly however, this proteolytic machine is absent from most fungi including, *Saccharomyces cereviseae* [12, 13]. For a detailed review of about the role of these AAA+ proteases in regulating mitochondrial function please refer to [14]. Although the AAA+ proteases ClpAP and

ClpCP are not as widely conserved as ClpXP, these proteases do, nevertheless, control a number of key proteolytic/regulatory pathways in Gram-negative and Gram-positive bacteria, respectively. Interestingly, ClpCP also appears to play an important role in proteostasis within the plastid of plants (for a recent review see [15, 16]).

Although these machines recognise a variety of different substrates and regulate a range of different physiological processes, each machine shares a common architecture and a similar mode of action. All form barrel-shaped complexes in which the oligomeric AAA+ unfoldase is concentrically aligned with the oligomeric protease component as is best illustrated by the crystal structure of the HslUV complex [17, 18]. Interestingly, the unfoldase component may be located at either or both ends of the peptidase component to form single-headed (1:1) or double-headed (2:1) complexes, respectively. For the ClpAP protease, the symmetric double-headed complexes have been shown to be most efficient at processing substrates [19]. Regardless of whether the complexes are single- or double-headed, both oligomeric components (i.e. the unfoldase and the peptidase) generally exhibit a six-fold symmetry throughout the entire complex. However in the Clp protease complexes (e.g. ClpAP, ClpCP and ClpXP) the machines display a unique symmetry mismatch between the unfoldase and the peptidase. While the AAA+ unfoldase component (i.e. ClpA, ClpC and ClpX) like most AAA+ proteins studied to date, form hexameric ring-shaped oligomers, the peptidase (i.e. ClpP) is composed of two heptameric rings [20]. The two heptameric rings of ClpP stack back-to-back, encapsulating the catalytic (active site) residues of ClpP within a barrel shaped tetradecamer. This symmetry mismatch poses some interesting questions. How do these two rings (the hexameric unfoldase and the heptameric peptidase) interact to form a functional complex, and how many subunits are required for a functional interaction. Regardless of whether the protease complex is symmetric or asymmetric, all AAA+ proteases undergo three basic steps in order to degrade a substrate protein (see Fig. 1.2). In the first step, the substrate is recognised by the unfoldase, although in some cases substrate recognition may be facilitated by an adaptor protein (see later). In bacteria, substrates are usually recognised via short sequence specific motifs (termed degrons), which are often located at the N- or C-terminus of the substrate protein. Following recognition, the substrate is then unfolded in an ATP-dependent fashion (Fig. 1.2, step 2). The unfolded substrate is then translocated into the associated peptidase, where the polypeptide chain is hydrolysed into small peptide fragments (~3–8 amino acids long), which have been proposed to egress through the holes in the sidewall of the peptidase, although this method of egress remains somewhat controversial (Fig. 1.2, step 3) [21, 22].

The Peptidase ClpP

The ClpP peptidase is synthesized as a zymogen, containing a N-terminal propeptide [23], which is autocatalytically cleaved upon oligomerization, resulting in the formation of a proteolytically active oligomer. ClpP is a serine protease, composed of a

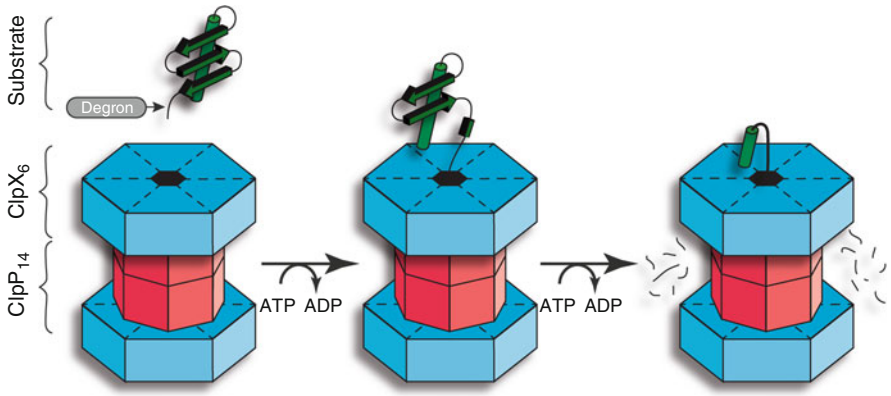

Fig. 1.2 Cartoon illustrating the main steps involved in substrate recognition and degradation by AAA+ proteases. The unfoldase (e.g. ClpX) forms a hexameric ring-shaped structure (*blue*) at one or both ends of the peptidase (e.g. ClpP), which forms two heptameric rings stacked back-to-back (*red*). The substrate (*green*) contains a degradation signal (*degron*) often located at the N- or C-terminus of the protein. The degron is recognised by the unfoldase and the substrate protein unfolded, in an ATP-dependent fashion, then translocated into the peptidase where the protein in cleaved into small peptide fragments, which diffuse through holes in the side-wall of the peptidase

Ser-His-Asp catalytic triad (Fig. 1.1), which exhibits chymotrypsin-like activity, that is, it cleaves peptide bonds mostly after non-polar residues [24, 25]. The active peptidase is a barrel-shaped oligomer composed of two heptameric rings, stacked back-to-back [20], that forms a degradation chamber in which the proteolytic active sites are sequestered away from cytosolic proteins (Fig. 1.3a). Each monomer of ClpP resembles a hatchet and consists of three subdomains: a handle, a globular head and a N-terminal loop. The heptameric ring is formed by the interaction of seven subunits through the head subdomain, and the tetradecamer is formed by the interaction of two heptameric rings through the handle subdomain (Fig. 1.3a). Entry into the catalytic chamber of this serine peptidase is restricted to a narrow entry portal (~10 Å) at both ends of the barrel-shaped complex. The N-terminal peptides of ClpP flank the axial pore and are proposed to act as a gate for entry into the proteolytic chamber. As a result of this narrow axial entry portal, folded proteins are excluded from entering the catalytic chamber, although small peptides and unfolded proteins can be degraded in an ATPase independent fashion, albeit unfolded proteins are degraded very slowly in the absence of the ATPase [26]. Importantly, the degradation of unfolded substrates can be accelerated by the addition of a cognate unfoldase (i.e. ClpX, ClpA or ClpC), which implies that entry into ClpP is gated and that this gated-entry can be activated by the unfoldase. Indeed, recent cryo-EM reconstructions have shown that binding of ClpA triggers a change in the N-terminal loops of ClpP, from a "down" conformation where they block entry to the catalytic chamber, to an "up" conformation which permits access to the chamber [27]. Consistent with a "gating" role for the N-terminal loops of ClpP, deletion of these loops was

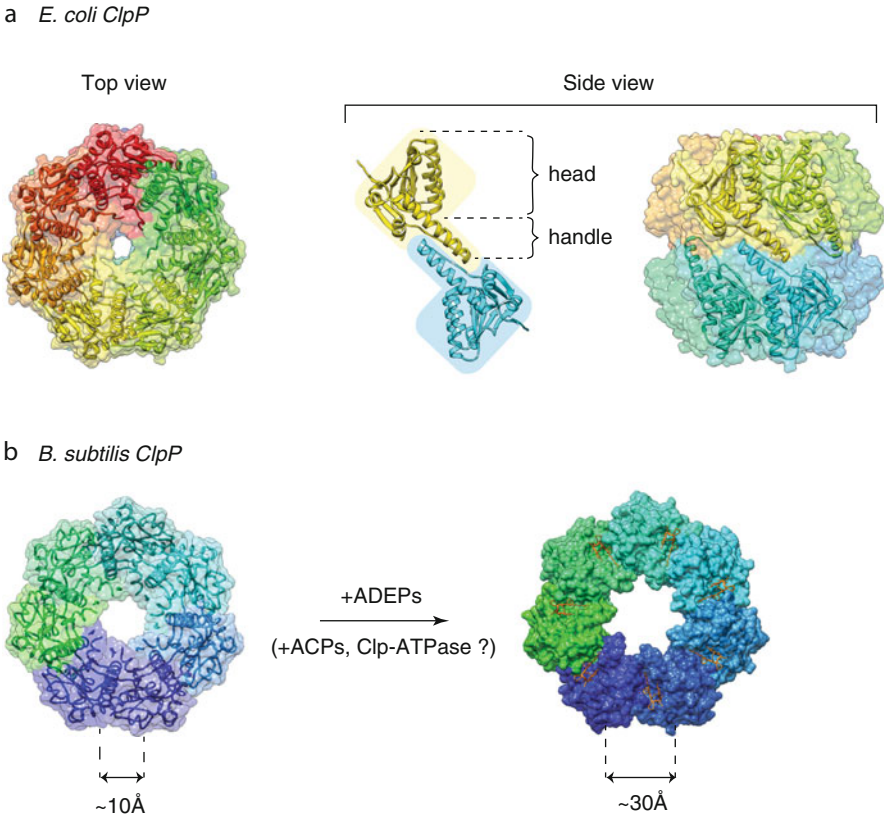

Fig. 1.3 Oligomeric structure of ClpP. (**a**) ClpP (PDB: 1TYF) forms two heptameric ring-shaped oligomers (*Top view*) stacked back-to-back (*Side view*) to create a barrel-shaped oligomer. Interactions between adjacent head subdomains drive oligomerisation of the seven-membered ring, while interactions between the handle subdomain of two heptamers are responsible for formation of the tetradecamer. (**b**) In the absence of the unfoldase, the entry portal into the catalytic chamber of ClpP (PDB: 3KTH) is narrow (~10 Å), in the presence of chemical activators of ClpP (i.e. ADEPs, ACPs and potentially the unfoldase), the entry portal into the catalytic chamber of ClpP (PDB: 3KTI) is opened (~30 Å)

shown to accelerate the degradation of short peptides [28]. The cognate AAA+ unfoldase also mediates the degradation of folded substrate proteins by actively unfolding and translocating the substrates through the axial pore and into the proteolytic chamber of ClpP. Indeed, it appears that the oligomeric structure of ClpP has been carefully designed to prevent widespread and indiscriminate degradation of cellular proteins by regulating substrate access to its proteolytic chamber. Consistent with this idea, several recent studies have identified a series of novel antibiotics (e.g. acyldepsipeptides (ADEPs) and ACPs) that activate ClpP (in the absence of its cognate unfoldase) for unregulated protein degradation [29–34]. This activation of ClpP

results in the unregulated degradation of nascent polypeptides and unfolded proteins in the cell [34], and in a recent study ADEP was shown to inhibit cell division of Gram positive bacteria, through the ClpP-mediated degradation of FtsZ, a key protein required for septum formation [35].

Based on a series of biochemical and structural studies, these chemical activators of ClpP dock into a hydrophobic pocket located on the surface of ClpP (Fig. 1.3b). Firstly, and most importantly, ADEP binding to this hydrophobic pocket results in opening of the ClpP pore (from ~10 Å in the absence of ADEP to ~21–27 Å in the presence of different forms of ADEP). This "gated-opening" of the ClpP pore, is proposed to be sufficient to allow entry of unfolded proteins into the proteolytic chamber of ClpP (where the catalytic residues are located) and possibly the primary reason for degradation of unfolded substrates. Interestingly, in the case of *B. subtilis* ClpP, ADEP not only triggers opening of the pore, but also triggers oligomerisation of ClpP from free "inactive" monomers to "active" tetradecamers [32], a step that is normally controlled by the cognate unfoldase, ClpC [36]. Similarly, ADEP activation of human ClpP for unregulated degradation is also likely to result from assembly of the ClpP tetradecamer [37] a process that normally requires the assistance of ClpX [38]. As a consequence, ADEP also appears to be a competitive inhibitor of unfoldase binding to ClpP, preventing the regulated degradation of substrates that would normally be delivered to ClpP by the unfoldase component [32]. As such, the ADEP-bound conformation of ClpP has been proposed to mimic the unfoldase-bound conformation of ClpP. Surprisingly, binding of ClpA to ClpP, as measured from sections of the ClpAP cryo-EM structure, appears to have little effect on the size of the ClpP pore (diameter ~12 Å) [27] and hence it has been suggested that the size of the pore may vary with translocation of different substrates [39]. Nevertheless, it remains to be seen, if an ordered arrangement of the N-terminal loops on ClpP (as observed in the *B. subtilis* ClpP-ADEP structure) or a disorder arrangement of the N-terminal loops of ClpP (as observed in the *E. coli* ClpP-ADEP complex) resembles the unfoldase bound complex.

The Unfoldase Components (ClpX/ClpA/ClpC)

In *E. coli*, ClpP forms proteolytic complexes with both ClpA and ClpX, while in *B. subtilis*, ClpP associates with three different unfoldases, ClpC, ClpX and ClpE [3]. Although the overall architecture of the different unfoldase components is similar, each unfoldase contains a unique organisation. While ClpA, ClpC and ClpE each contain two AAA+ domains, ClpX only contains a single AAA+ domain (Fig. 1.1). Regardless of the number of AAA+ domains present, each unfoldase contains one or more accessory domains. In the cases of ClpA and ClpX, a single accessory domain is located at the N-terminus of the protein, while both ClpC and ClpE contain two accessory domains, one at the N-terminus of the protein and the other located between the two AAA+ domains, termed the middle or M-domain (Fig. 1.1). In general, these accessory domains are required for the binding of substrates and/

or adaptor proteins. In the case of ClpA, the N-terminal domain is essential for docking of the adaptor protein ClpS [40–42] but also required for the recognition, and hence degradation of some substrates [43, 44]. Similarly, the N-terminal domain of *B. subtilis* ClpC is essential for the ClpP-mediated degradation of most substrates [45, 46]. However in this case, the N-domain is thought not to be directly involved in substrate recognition but rather plays a crucial role in binding adaptor proteins (i.e. MecA and McsB), which are required for ClpC oligomerisation and/or substrate delivery [36, 47, 48]. Interestingly in the case of *B. subtilis* ClpC, the second accessory domain (the M-domain) located between the two AAA+ domains, also plays an important role in the recognition of adaptor protein, however the details of substrate delivery by these adaptor proteins is currently unknown [36, 46–48]. For further details regarding the mechanism of action of ClpCP please refer to [3].

In the case of ClpX (and ClpE from Gram-positive bacteria) the N-terminal accessory domain (residues 1–60 in *E. coli* ClpX) is a C4-type Zinc binding domain (ZBD), which contains four Cysteine residues that coordinate a single Zn atom. In *E. coli* ClpX, this domain forms a very stable dimer [49], and is responsible for the recognition of several substrates (such as λO and MuA) but not SsrA-tagged proteins [50–52]. This domain is also essential for the recognition of the adaptor proteins, SspB [50, 52, 53] and UmuD [54], discussed in more detail later.

Given that *E. coli* ClpX is, by far the most extensively characterised Clp-ATPase, this section will focus primarily on the structure and function of ClpX. However, many of the features described here for the AAA+ domain of ClpX are likely to be generally applicable to most AAA+ proteases. At a structural level, the AAA+ domain (~200–250 a.a.) is composed of two subdomains – a large N-terminal subdomain, which forms an α/β wedge-shaped Rossman fold and a small C-terminal subdomain, which forms a α-helical lid across the nucleotide-binding site [55, 56]. ATP is bound in a cleft between the large and small subdomain of a single subunit and the large subdomain of the adjacent subunit. As such, these interactions provide much of the driving force for formation of the hexamer. To date, several highly conserved sequence motifs have been identified within the AAA+ domain, each of which is responsible for a specific function [57]. The Walker A motif (GXXXXGK [T/S], where X = any amino acid) is required for ATP binding and facilitates oligomerization of the protein into ring-shaped hexamers. The Walker B motif (hhhhDE, where h = any hydrophobic amino acid) is required for hydrolysis of bound ATP and hence drives conformational changes in the protein, mediating substrate binding and translocation. The central pore of the hexamer is comprised of several important motifs and loops (e.g. the pore-1 loop) involved in substrate binding [58–61]. The Sensor 1 and 2 motifs, together with the arginine fingers, are proposed to couple the nucleotide-bound state of the oligomer with conformational changes in the subdomains, which through movement of the substrate-binding loops, results in substrate unfolding and translocation [55, 58]. Despite the broad sequence conservation of AAA+ domains, individual AAA+ domains appear to serve different functions in proteins that contain two or more AAA+ domains (i.e. ClpA or ClpC) [62]. For example, the first AAA+ domain (D1) in ClpA is crucial for oligomerisation while the second AAA+ domain (D2) is primarily responsible for ATP hydrolysis [63]. Interestingly,

variants of ClpA lacking ATPase activity in either D1 or D2, are only able to process substrates with "intermediate" or "low" local stability respectively, suggesting that each domain can function independently, at least to a limited extent [64]. However, the ATPase activity of both domains is required for the efficient processing of substrates with "high" local stability [64] indicating that both domains work together to unfold and translocate substrates into ClpP.

As viewed from the top (or ClpP distal face) of the unfoldase, the ClpX hexamer can be divided into six units, each of which was composed of a small AAA+ subdomain from one subunit with a large AAA+ subdomain of the adjacent subunit [55, 56]. Recently, it was shown that the structures of all six of these units were highly superimposable [55] and hence it was proposed that each unit forms a functional rigid body (Fig. 1.4a, b). Despite the high degree of structural similarity between each rigid body unit, the overall shape of the ClpX hexamer is asymmetric, which suggests that the angle of the hinge between the rigid body units (i.e. the angle between the large and the small subdomains within a single subunit of ClpX) varies. This difference in the angle between the rigid body units results in a different ability of each subunit to bind nucleotide. Based on this description, each subunit within the ClpX hexamer can be classified into one of two groups; type 1 subunits, which are able to bind nucleotide (referred to as L, for "loadable"), and type 2 subunits, which are unable to bind nucleotide (referred to as U, for "unloadable"). In the crystal structure of ClpX, the hexamer is composed of four L (or type 1) subunits and two U (or type 2) subunits arranged in the following manner, L-L-U-L-L-U (Fig. 1.4c). Therefore, given that ATP binding and hydrolysis is expected to stabilise the L conformation, while the release of ADP is predicted to result in an transition from the L to the U conformation, it is proposed that the ATPase activity of ClpX will promote domain rotations within a subunit that will propagate around the hexamer and drive transition of the other subunits, in a chain reaction. These ATPase-induced conformational changes are proposed to form an integral part of the mechanism for substrate translocation by ClpX into ClpP (see later).

The Unfoldase-Peptidase Complex

Given that the AAA+ unfoldase component (i.e. ClpX, ClpA or ClpC) is hexameric and the associated peptidase (e.g. ClpP) is formed by two heptameric rings, the resulting proteolytic machines, ClpXP (ClpAP and ClpCP), exhibit an asymmetry between the two components. This asymmetry, although not unique in biology, poses several interesting questions. How do the two components interact with one-another? How many of these features per hexamer (i.e. how many subunits) are required for formation of a functional complex? Not surprisingly, the formation of the complex is transient, and efficient interaction of the two components is dependent on nucleotide-bound state of the unfoldase. Specifically, formation of the ClpXP complex is only supported by ATP, ATPγS (a slowly hydrolysable analogue of ATP) or a ClpX mutant that is defective in ATP hydrolysis [65]. In contrast, the complex

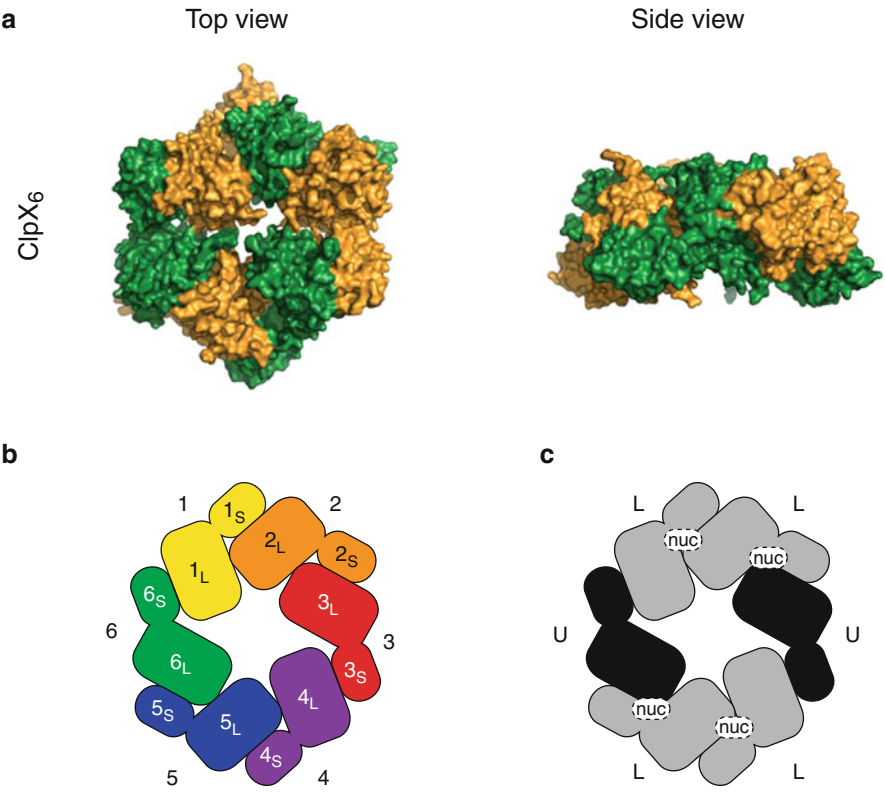

Fig. 1.4 Oligomeric structure of ClpX. In the presence of nucleotide, ClpX forms a hexameric *ring-shaped* oligomer. (**a**) Surface representation of the ClpX hexamer (PDB: 3HWS). (**b**) Cartoon, illustrating the asymmetric organisation of the ClpX hexamer. (**c**) The asymmetric organisation of the ClpX hexamer results from a differential binding of nucleotide (nuc) within the hexamer. Nucleotides are bound in a cleft formed by the large and small domain of one subunit and the large domain of the adjacent subunit. Depending on the orientation of the small and large domain within a subunit, a subunit can be classified into two types; loadable (L) which are able to bind nucleotide and unloadable (U) which are unable to bind nucleotide. The arrangement of these different subunit types, within the ring gives rise to an asymmetric appearance of the hexamer

dissociates in the presence of ADP or in the absence of nucleotide [66, 67]. This interaction, (i.e. between the two components), is mediated by two sets of contacts; one at the periphery of the interface and the other near the central pore. The peripheral contact occurs between a flexible loop on ClpX and a hydrophobic pocket on the surface of ClpP, and is important for a strong, nucleotide-independent interaction with ClpP. The flexible loop contains a conserved tripeptide motif ([L/I/V]-G-[F/L]) and as such is often referred to as the IGF/L-loop (Fig. 1.5a). This motif is unique to ClpP-binding unfoldases (i.e. ClpA, ClpC, ClpE and ClpX) and is essential for interaction with ClpP [68, 69]. Consistently, mutation of this motif dramatically reduces the affinity of ClpX to ClpP [67, 68]. The second contact

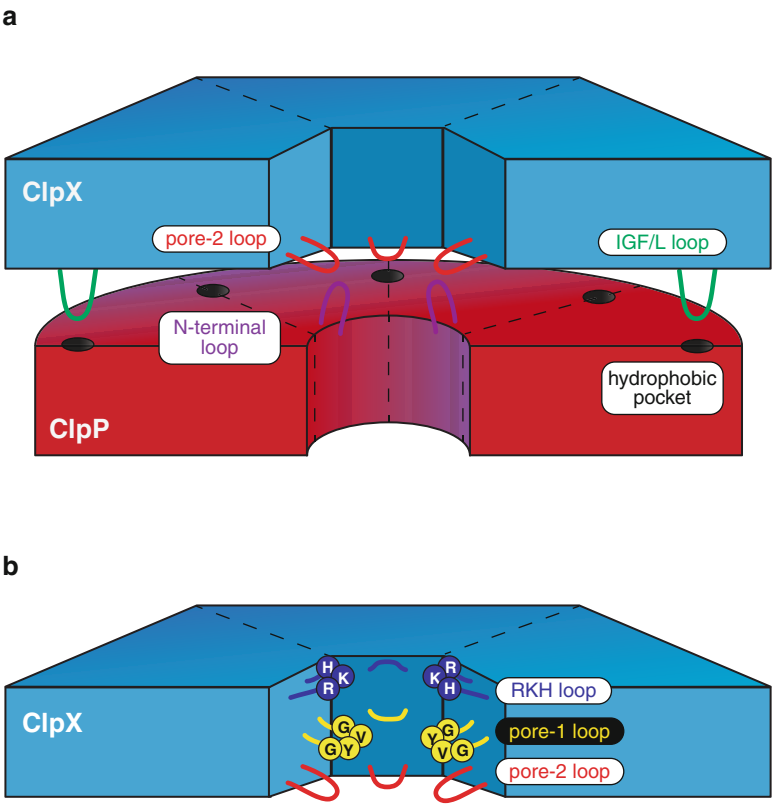

Fig. 1.5 ClpP-binding and substrate interaction is mediated by several loops and pockets. (**a**) Cut-away view of ClpX (*blue*), highlighting the important interactions that contribute to complex formation with ClpP (*red*). The IGF/L loops (*green*) on ClpX form a static interaction with the hydrophobic pocket on ClpP (*black*). ClpXP complex formation is modulated by the nucleotide state of ClpX, through a set of dynamic interactions, between pore-2 loops of ClpX (*red*) and the N-terminal loop of ClpP (*purple*). (**b**) The substrate is recognised and translocated through the pore via a set of conserved pore loops; RKH (*blue*), pore-1 (*yellow*) and pore-2 (*red*). These loops move up and down the pore of ClpX in a nucleotide-dependent fashion, thereby translocating the substrate into ClpP

occurs between two loops; one loop (termed the pore-2 loop) protrudes from the axial pore of ClpX, and interacts with the N-terminal loop of ClpP [21, 70, 71]. This interaction, between the two axial loops, appears to be highly dynamic and is dependent on the nucleotide-state of individual subunits of ClpX [71]. Although the ClpXP complex is asymmetric, both sets of loops (the IGF/L-loop, for docking into the hydrophobic pocket on ClpP and the two axial pore loops) appear to be flexible enough that contacts from each subunit of ClpX contribute to the interaction. Indeed loss of a single IGF-loop, within the ClpX hexamer, is sufficient to reduce ClpP binding and activity, while loss of more than one contact per hexamer completely abolishes ClpP binding [71].

Degradation Recognition Motifs (Degrons)

A bacterial cell is composed of thousands of different proteins, the concentration (or copy number) of which varies dramatically (from ~100 to 10^5 molecules per cell) [72]. Likewise, the concentration of each individual protein varies in response to changing environmental conditions or stress. As such, in order for the cell to maintain optimal function, not only under normal conditions but also under conditions of stress, the composition and active concentration of its proteins must be monitored and maintained. Hence it is important for the cell to specifically remove unwanted or damaged proteins from the cell when they are no longer required. To achieve this, bacterial proteases need to combine two seemingly incompatible properties, broad recognition of a range of different protein substrates, with a high degree of substrate specificity to prevent the recognition of properly folded or wanted cellular proteins.

A key feature of most, if not all, bacterial protein substrates is the presence of a specific amino acid motif, often referred to as a degradation tag or degron [73]. These degrons are generally located at the N- or C-terminus of the protein, although in some cases they are located internally. Although most degrons are intrinsic to the target protein, a handful of degrons (e.g. the SsrA tag and some N-end rule substrates) are not defined by the primary sequence of the protein, but rather are added (either co- or post-translationaly) to the protein [74, 75]. Often, intrinsic degrons are only revealed (for recognition by the protease) following exposure of the protein to stress (e.g. heat-shock) or processing by an endoprotease [76–79]. This conditional recognition of a protein substrate is ideally suited to the controlled degradation of a key regulatory protein, and forms the basis of controlling several stress response pathways in bacteria (see [6]). In some cases however, a degron may be constitutively exposed under normal conditions, in order to maintain low levels of the protein (e.g. SigmaS) [80].

Trans-translation and the SsrA-Tag: A Specific Protein Tagging System in Bacteria

Messenger RNA molecules normally contain a stop codon at the 3′ end of the transcript, which serves not only to signal the end of translation, but also triggers ribosome dissociation. In some cases however, as a result of truncation of the mRNA or errors during its transcription, the lack of a stop codon in the mRNA sequence caused "stalling" of protein synthesis [81–83]. To overcome this problem, bacteria possess a conserved mechanism, to restart translation and allow ribosome dissociation. This mechanism (illustrated in Fig. 1.6), often referred to as trans-translation, is sensed by an empty A-site and signalled by stalling of the translating ribosome [84]. This signal results in the recruitment of a specialised RNA molecule into the empty A-site of the ribosome. This RNA, encoded by *ssrA* (small stable RNA gene A) [85] has been termed a tmRNA as it functions both as a tRNA and as an mRNA [84, 86, 87]. The tRNA-like structure can be charged with alanine at its 3′ end, while an extended

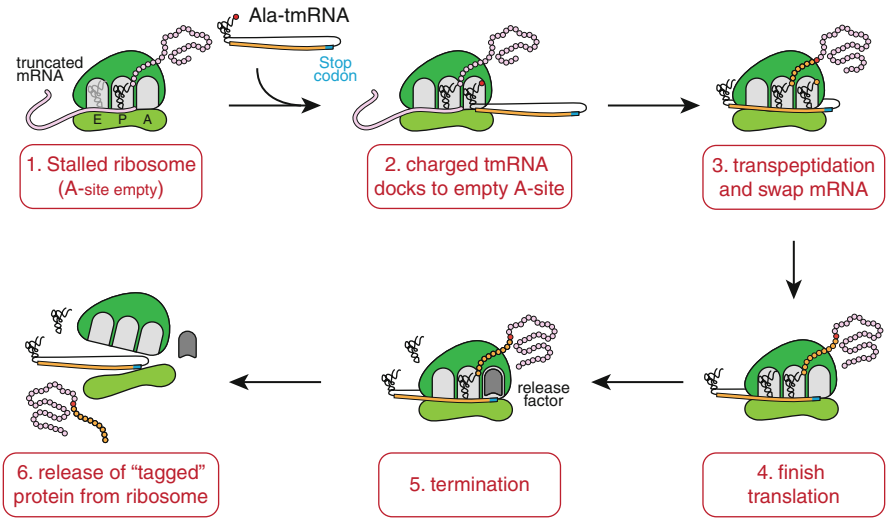

Fig. 1.6 Cartoon, illustrating the process of trans-translation. *1.* Truncated mRNA (lacking a stop codon) cause "stalling" of the ribosome. *2.* This "stalling" triggers binding of a tmRNA into the empty A-site of the ribosome. *3.* Following a transpeptidation reaction, the truncated mRNA is replace with the mRNA from the tmRNA and *4.* translation proceeds, resulting in *5.* correct termination of protein synthesis *6.* rescuing the ribosome and releasing the "tagged" protein for targeted degradation by ClpXP

loop within the same RNA molecule encodes a short open reading frame (ten amino acids in *E. coli*) that ends in a stop codon. Following docking of the charged tmRNA into the empty A-site, the alanine is transferred to the nascent polypeptide and the open reading frame (encoded by the mRNA portion of the tmRNA) is translated. Noteworthy, trans-translation results in the attachment of a short C-terminal extension (termed the SsrA tag) to the incompletely synthesised protein.

Importantly, given that SsrA-tagged proteins are produced from aberrant or incomplete mRNA, it is unlikely that they will be able to fold. For this reason, interaction of SsrA-tagged proteins with chaperones is wasteful, as attempts to refold trans-translation products would be futile. Rather, SsrA-tagged proteins are rapidly degraded by proteases. In *E. coli*, the SsrA tag is 11 amino acids long (AANDENYALAA) and substrates tagged with the sequence are recognised by ClpXP, ClpAP and FtsH [81, 88–90]. Despite the fact that the SsrA tag is recognised by several different proteases *in vitro*, the *in vivo* degradation of these substrates is almost exclusively performed by ClpXP [81, 91].

Nevertheless, this tag has been used extensively as a model degron to study the function of both ClpXP and ClpAP. As such, it has proved to be a powerful research tool to study the mechanism of protein recognition and degradation by AAA+ proteases. A major advantage of the SsrA tag, as a research tool to study protein degradation, is that any protein can be converted into a ClpXP (or ClpAP) substrate, simply through the attachment of the SsrA tag to its C-terminus. This has permitted

a detailed mechanistic analysis of protein degradation using a range of different substrates with a variety of unique or desired features (i.e. green fluorescent protein (GFP) or the I27 domain of the human titin) to examine unfolding [92–95]. Likewise, it has also served as an excellent tool to study the mechanism of adaptor-mediated substrate delivery (see below).

Other ClpX Recognition Motifs

Apart from the specific recognition of the SsrA-tag, ClpX is also involved in the recognition of several other proteins, including a number of proteins involved in various stress response pathways. In order to determine the complete substrate-binding repertoire of *E. coli* ClpX, a mutant version of ClpP was used to capture the physiological substrates of ClpXP *in vivo* [96]. Using this approach, ~100 putative ClpXP substrate proteins were identified [96, 97]. Following verification of several of these proteins (either by *in vitro* or *in vivo* degradation assays) five different ClpX "recognition" motifs were proposed [96]. Of the five different "recognition" motifs, two were located near the C-terminus of the protein and three near the N-terminus of the protein (Fig. 1.7). While both classes of C-terminal motifs (C-motif 1 and 2, Fig. 1.7) shared homology with known ClpXP substrates (i.e. the SsrA-tag and MuA, respectively), only a single N-terminal motif (N-motif 1, Fig. 1.7) had been observed previously (i.e. λO) [98].

Interestingly, the various degradation motifs appear to be recognised by different regions within the unfoldase. Some substrate classes (e.g. N-motif 1) strictly depend on interaction with the N-terminal domain, while other motif classes (e.g. C-motif 1, i.e. SsrA-tagged substrates) do not require this domain for direct recognition [50, 52, 69]. For example, λO (a replication protein of bacteriophage λ) carries an N-terminal degradation motif (N-motif 1, NH_2-TNTAKI), which is specifically recognised by the N-terminal domain of ClpX [52, 96, 99]. Indeed deletion of this domain (from ClpX) inhibits the ClpP-mediated degradation of λO [52], which is proposed to result from the low affinity of this class of substrate to the axial loops on ClpX. Tethering of this class of substrate, by the N-terminal domain, is likely to increase the effective concentration of the substrate, near the pore of ClpX. As a result, despite their low affinity to the pore loops, high affinity to the N-terminal domain promotes their engagement by the pore and, consequently, their efficient degradation. The N-terminal domain is also involved in the recognition of the adaptors proteins, SspB and UmuD, and substrate proteins such as MuA (C-motif 2, Fig. 1.7), which appear to share a conserved motif [50, 52, 54]. Importantly however, the adaptor proteins are not degraded by ClpXP, presumably because they are not recognised by the pore-1 motif of ClpX.

Other Degradation Tags

Currently, the substrate recognition motifs for ClpA are only poorly defined. The first ClpAP substrate to be identified was ClpA itself [100]. Interestingly, although the recognition motif within ClpA was originally proposed to be located at the