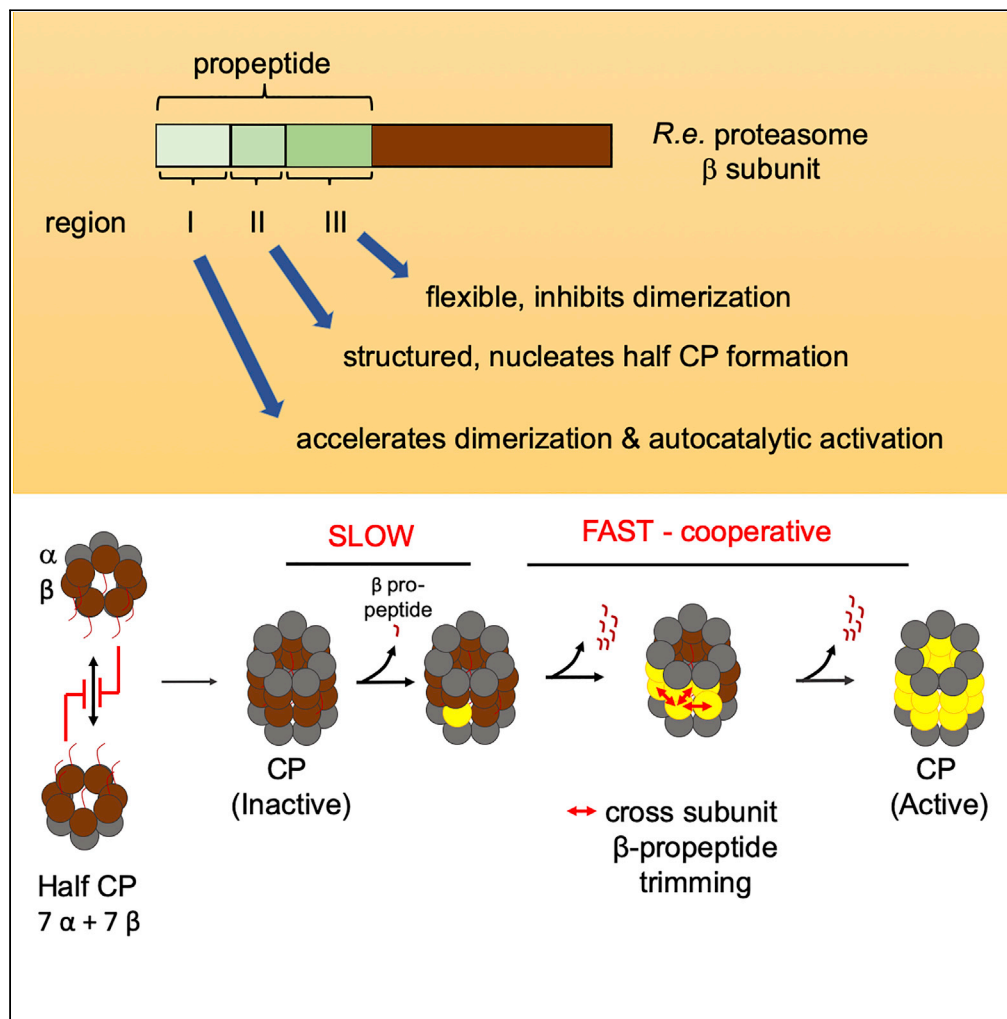


Article

Cooperativity in Proteasome Core Particle Maturation



Anjana Suppahia,
Pushpa Itagi,
Alicia Burris, ...,
Wonpil Im, Eric J.
Deeds, Jeroen
Roelofs

deeds@ucla.edu (E.J.D.)
jroelofs@kumc.edu (J.R.)

HIGHLIGHTS

A flexible region in β propeptide regulates proteasome core particle (CP) formation

Identified pre-holo-complex distinguishes dimerization from autocatalytic activation

Cooperativity found in proteasome core particle (CP) activation

Bacterial CP displayed propeptide cross-cleavage



Article

Cooperativity in Proteasome Core Particle Maturation

Anjana Suppahia,^{1,2} Pushpa Itagi,^{3,4} Alicia Burris,^{1,2} Faith Mi Ge Kim,² Alexander Vontz,² Anupama Kante,^{4,5} Seonghoon Kim,⁶ Wonpil Im,^{6,7,8} Eric J. Deeds,^{3,4,9,*} and Jeroen Roelofs^{1,2,10,*}

SUMMARY

Proteasomes are multi-subunit protease complexes found in all domains of life. The maturation of the core particle (CP), which harbors the active sites, involves dimerization of two half CPs (HPs) and an autocatalytic cleavage that removes β propeptides. How these steps are regulated remains poorly understood. Here, we used the *Rhodococcus erythropolis* CP to dissect this process *in vitro*. Our data show that propeptides regulate the dimerization of HPs through flexible loops we identified. Furthermore, N-terminal truncations of the propeptides accelerated HP dimerization and decelerated CP auto-activation. We identified cooperativity in autocatalysis and found that the propeptide can be partially cleaved by adjacent active sites, potentially aiding an otherwise strictly autocatalytic mechanism. We propose that cross-processing during bacterial CP maturation is the underlying mechanism leading to the observed cooperativity of activation. Our work suggests that the bacterial β propeptide plays an unexpected and complex role in regulating dimerization and autocatalytic activation.

INTRODUCTION

Protein degradation is an essential cellular process required to maintain homeostasis and to allow the cell to react efficiently to changing environmental conditions. The proteasome, one of the major proteases, is found ubiquitously in eukaryotes, archaea, and some bacterial orders like Actinomycetales and Nitrospirales (reviewed in [Becker and Darwin, 2017](#)). At its center is the structurally conserved core particle (CP) complex that consists of four heptameric rings stacked to form a hollow, cylindrical protease complex. In eukaryotes, the rings consist of seven distinct α and β subunits. However, archaeal and eubacterial genomes mostly encode only one paralog of the α and β subunit each and thus have homo-heptameric α and β rings (with an overall $\alpha_7 \beta_7 \alpha_7 \beta_7$ stoichiometry) ([Becker and Darwin, 2017](#)).

The formation of CP involves the dimerization of two half CPs (the $\alpha_7 \beta_7$ "Half Proteasome", here referred to as HP), and two distinct assembly pathways for the HP have been described. One, common in eukaryotes and archaea, starts with the formation of an α subunit ring. This ring then serves as a docking site for the β subunits ([Hirano et al., 2008](#); [Zwickl et al., 1994](#)). The second, found mainly in bacteria, starts with dimerization of α and β subunits, which then rapidly combine to form HP ([Zühl et al., 1997a, 1997b](#)). The stability of the α - α versus α - β interactions, a feature that correlates with the buried surface area in the interaction, seems to be the distinguishing factor for these pathways ([Hu et al., 2006](#); [Kwon et al., 2004](#); [Panfair et al., 2015](#); [Zühl et al., 1997a, 1997b](#)). Regardless of the pathway, the catalytically active β subunits are synthesized in an inactive form, with an N-terminal propeptide sequence. Dimerization of the HP coincides with a proteolytic processing of some (eukaryotes) or all (archaea and bacteria) of the β subunits to form a proteolytically active CP ([Becker and Darwin, 2017](#); [Budenholzer et al., 2017](#); [Kunjappu and Hochstrasser, 2014](#); [Sharon et al., 2007](#)).

The assembly process in eukaryotes is more complex since the eukaryotic CP consists of fourteen different polypeptides. Perhaps owing to this increased complexity, eukaryotic CP assembly involves at least five CP-dedicated chaperones. Some of these chaperones have orthologs in archaea, whereas none have been identified in bacteria, and prokaryote-derived CPs have been reconstituted without the need for chaperones ([Kusmierczyk et al., 2011](#); [Sharon et al., 2007](#); [Zühl et al., 1997a](#)). Apart from separate chaperone molecules, the propeptides of β subunits have been shown to function as "intrinsic chaperones." For example, in yeast, the propeptide of Doa3 (i.e., β_5) is required for its incorporation into the CP ([Chen and](#)

¹Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA

²Molecular, Cellular, and Developmental Biology Program, Division of Biology, Kansas State University, 338 Ackert Hall, Manhattan, KS 66506, USA

³Center for Computational Biology, University of Kansas, 2030 Becker Drive, Lawrence, KS 66047, USA

⁴Institute for Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles, CA 99024, USA

⁵Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047, USA

⁶Department of Biological Sciences, Lehigh University, Bethlehem, PA 18105, USA

⁷Department of Bioengineering, Lehigh University, Bethlehem, PA 18105, USA

⁸Department of Chemistry, Lehigh University, Bethlehem, PA 18105, USA

⁹Department of Integrative Biology and Physiology, University of California Los Angeles, Los Angeles, CA 99024, USA

¹⁰Lead Contact

*Correspondence: deeds@ucla.edu (E.J.D.), jroelofs@kumc.edu (J.R.)

<https://doi.org/10.1016/j.isci.2020.101090>



Hochstrasser, 1996). Similarly, propeptides in actinobacteria are also essential for their proper folding and incorporation into higher-order complexes during assembly (Zühl et al., 1997a). The bacterial propeptides bind to the α subunits and assist in the formation of $\alpha\beta$ heterodimers during HP assembly. The lack of dedicated chaperones in bacteria suggest that parallels between the functions of eukaryotic chaperones and bacterial propeptides exist. For example, the eukaryotic chaperone Ump1 (a.k.a. POMP in humans) is an intrinsically disordered protein that remains associated with immature HPs, preventing premature dimerization (Ramos et al., 1998). This chaperone is degraded along with other propeptides as the first substrate of an assembled and active CP (Ramos et al., 1998). This is in many ways similar to the bacterial propeptide, which is partially disordered, regulates dimerization, and is cleaved during CP maturation (Kwon et al., 2004; Zühl et al., 1997a).

A better understanding of bacterial propeptides will thus help to elucidate several fundamental evolutionarily conserved aspects of CP assembly and reveal the critical core functions of the propeptides themselves. Moreover, this may also reveal differences in the assembly process that could be exploited in the development of drugs. For example, specifically targeting the *M. tuberculosis* CP is thought to be of therapeutic value (Lin et al., 2009; Totaro et al., 2017). This actinomycete bacterium is the causative agent of tuberculosis (Tb), a major disease with ~9 million new cases each year and about 1.5 million deaths (Zumla et al., 2015). Furthermore, recent data suggest that anti-PD-1 drugs used and tested against a variety of cancers are associated with higher abundance of Tb (Barber et al., 2019). In *M. tuberculosis*, the prokaryotic ubiquitin-like protein (Pup)-proteasome system, a bacterial conjugating system with parallels to the Ubiquitin-Proteasome System of eukaryotes, is important for virulence of this pathogen (Cerdeira-Maira et al., 2010; Darwin et al., 2003; Gandotra et al., 2007). Thus, a mechanistic understanding of the function of bacterial CP assembly and the role of propeptides might reveal ways how we can interfere with CP assembly, which could be developed in therapeutic targets for Tb treatment.

In this work, we used a bioinformatics approach to define three distinct regions in the bacterial β subunit propeptides based on their conservation patterns. Using *in vitro* reconstitutions and molecular dynamics studies, we identified a role for the N-terminal region of the propeptide in regulating the speed of HP dimerization and the autocatalytic activation of the CP. Based on these data, we propose a mechanism for the activation of the CP that involves cooperativity in the processing of propeptides between β subunits present in CP.

RESULTS

Bacterial Propeptide Can Be Divided into Three Evolutionarily Conserved Regions

The assembly of the eukaryotic CP involves five dedicated chaperones and seven unique α and β subunits. The genomes of bacteria and archaea normally encode one α and β subunit each (Maupin-Furrow et al., 2006), eliminating the need for a specific order of subunits within the rings as well as for the rings relative to each other (Murata et al., 2009). Consistent with this lower complexity, the archaeal and bacterial CPs assemble more readily and without the need for specific chaperones *in vitro* or in *E. coli* (Becker and Darwin, 2017), whereas human CP has only been heterologously expressed recently with the need of chaperones (Toste Rego and da Fonseca, 2019). Considering the absence of chaperones in bacteria and the reported role for the propeptides of the β subunits in assembly, we hypothesized that some of the functional roles of eukaryotic assembly chaperones could be performed by the propeptides of the eubacterial β subunits. To assess this, we focused on the propeptide of *Rhodococcus erythropolis* (*R.e.*) 20S β subunit Prc β 1 (proteasome component β 1). *R.e.* has two paralogs of α and β each and it has been repeatedly demonstrated that a functional CP forms with only one α and β present (Sharon et al., 2007; Zühl et al., 1997a, 1997b). Furthermore, it is amenable to *in vitro* reconstitution and there is detailed structural and biochemical information, allowing us to interpret our results in a structural context (Figure S1) (Kwon et al., 2004; Witt et al., 2006; Zühl et al., 1997a).

To assess potential conservation of properties among bacterial propeptides, we performed a multiple sequence alignment of β subunits from different bacterial species that showed $\geq 61\%$ or more sequence identity to the *R.e.* β subunit (Prc β 1) sequence (256 unique sequences in total). We chose this cutoff to include *M. tuberculosis* and all more closely related sequences. We defined three distinct regions (named I to III) in the propeptide (Figure 1A). Region II is largely identical to the previously described “central box”; it has an average length of 16 amino acids and corresponds to residues from -42 to -27 in the *R.e.* Prc β 1 sequence (propeptide residues are labeled with negative numbers, with residue -1 being just N-terminal

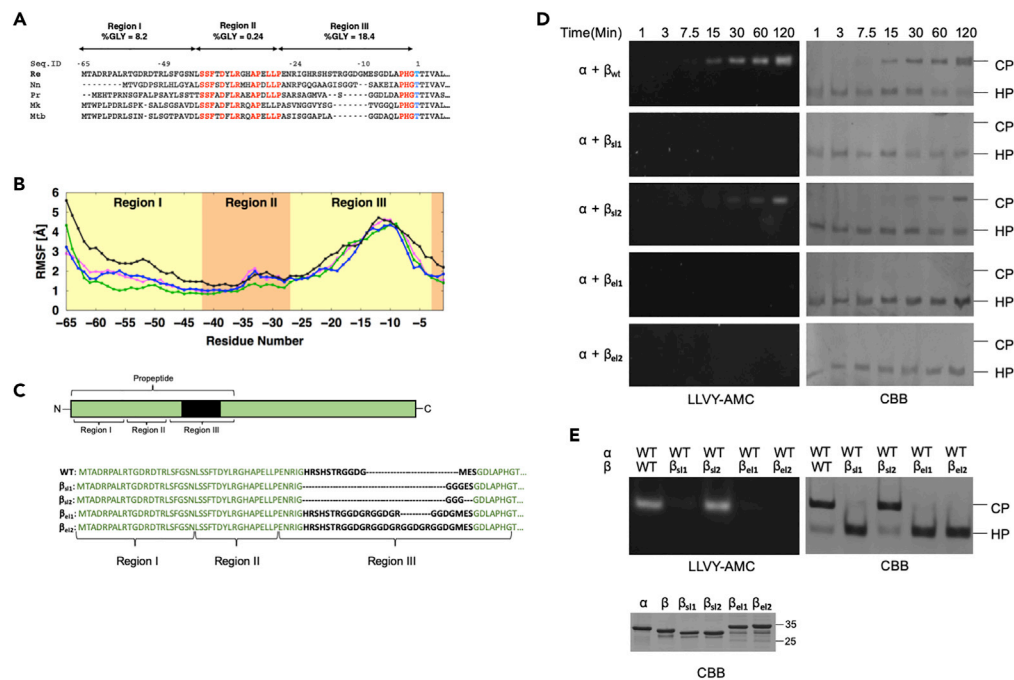


Figure 1. Region III of *Re* β Subunit Is Flexible Loop that Controls HP Dimerization

(A) Representative alignment of the N-terminal sequence derived from a multiple sequence alignment of 256 bacterial proteasome β subunits. Sequences shown are derived from *Rhodococcus erythropolis* (*Re*), *Nocardia niwae* (*Nn*), *Prausserella rugosa* (*Pr*), *Mycobacterium kansasii* (*Mk*), and *Mycobacterium tuberculosis* (*Mtb*). Conserved residues are highlighted in red and the active site threonine in blue. The three regions recognized in the propeptide are depicted by arrows with the percentage of glycine residues indicated.

(B) Root-mean-square fluctuations (RMSF) calculated from molecular dynamics simulations of the *Re* propeptide for the three replicates (magenta, blue, and green) and ANTON 2 (black) system are plotted for each amino acid. The background color indicates ordered region (orange) and disordered region (yellow) of the propeptide.

(C) Graphical representation of the β subunit of *Re* with the flexible loop indicated in black. The N-terminal sequence, including the flexible loop (black), is shown for wild-type and mutants β_{s11} , β_{s12} , β_{e11} , and β_{e12} .

(D) Equimolar amounts of α and indicated flexible loop mutants were reconstituted *in vitro* at 30°C for indicated time points. Samples were separated on native-PAGE visualized by in gel LLVY-AMC assay and CBB staining.

(E) Indicated α and β mutants were reconstituted overnight and analyzed as in (D) (top). CBB-stained SDS-PAGE of indicated subunits used for reconstitution (bottom).

of the propeptide cleavage site). Region II has a well-defined crystal structure and functions in ring formation by allowing a single β subunit to interact with two α subunits (Kwon et al., 2004). It is highly conserved among bacterial species at the sequence level (Figure 1A). Region I has an average length of 18 amino acids (residues -65 to -43 in *Re*.) and is not strongly conserved across bacterial taxa. Region III has an average length of 17 amino acids (-26 to -1 in *Re*.) Like region I, region III did not show any obvious sequence conservation. However, our analysis indicates region III is highly enriched in glycine residues with 18.4% glycine, compared with 0.24% for region II and 8.2% for region I. The latter is close to the average glycine composition for globular proteins (~8% [Creighton, 1983]).

The enrichment of glycine in region III is highly unlikely to have arisen purely by chance ($p = 3.9 \times 10^{-142}$, hypergeometric test), indicating that there is likely some evolutionary pressure to maintain it. This suggests that flexibility of this region may be important for its function. Nevertheless, biochemical and structural analyses to date have not identified any clear function for either region I or III.

Propeptide Region III Regulates HP Dimerization

Glycine residues generally disrupt α helices and β sheets and are more common in loops between secondary structural elements and as flexible linkers between protein domains or regions (Imai and Mitaku, 2005; Levitt, 1978). Therefore, we postulated that there has been an evolutionary pressure to generate a flexible,

disordered loop between the active site and the structured region II of the propeptide. To gain insight into this flexibility we used molecular dynamics (MD) simulations of the HP from *R.e*. We performed two distinct sets of simulations of the HP structure (generated from PDB:1Q5R). One set used nanoscale molecular dynamics (NAMMD) to simulate three separate replicate systems, each running independently for 100 ns per replicate. The other set used the Anton 2 machine (Shaw et al., 2014) to simulate the same HP structure for a total of 2 μ s. We characterized and plotted the flexibility of the propeptide in these simulations by calculating root-mean-square fluctuation (RMSF) for the backbone atoms (C, O, N, and C α) in the HP simulations (Figure 1B). RMSF is a metric used for determining the atomic mobility per residue averaged over time. As expected, region III showed considerable flexibility and a high RMSF, consistent with a highly disordered loop (Figure 1B). Interestingly, in the MD simulations, part of region III resided for an extended time outside of the HP (for example, see Video S1). The presence of propeptide density outside the HP could potentially cause a steric clash with a second HP, which would need to occupy this space during dimerization. Propeptide density observed outside of the barrel in cryo-EM structures of the *M. tuberculosis* HP also points to a similar role (Li et al., 2010).

To understand the role of region III, we designed mutants that altered the length of the flexible loop (Figure 1C). Two mutants, β_{s11} and β_{s12} , were created with shorter flexible loops based on the minimal theoretical length needed to span the distance between region II and the remainder of the β subunit in the crystal structure (Kwon et al., 2004). Two other mutants, β_{e11} and β_{e12} , were created with extended flexible loops with the idea that these loops would spend more time in the space outside the HP and thus be more effective at sterically blocking HP dimerization. We attempted to maintain (putative) flexibility by including glycine in the modified loops. After heterologous expression in *E. coli*, purified mutant forms were reconstituted with α subunits. Native gel analyses showed that all were able to quickly and efficiently form the HP, indicating the flexible loop in region III is dispensable for HP formation and that region II is likely properly folded and positioned to allow the association of β and α subunits (Figure 1D) (Kwon et al., 2004). However, all four loop mutants showed strongly reduced efficiency in dimerization with less (β_{s12}) or no (β_{s11} , β_{e11} , and β_{e12}) detectable active CP being formed after 120 min. This suggests that there is some evolutionary optimization in the length and/or composition of this region (Figure 1D). Upon overnight incubation, we detected a very small amount of active CPs for the β_{s11} , β_{e11} , and β_{e12} mutants (Figure 1E). The observed enzymatic activity correlated with the amount of full CP formed. This suggests that these three mutants are severely compromised in HP dimerization. To test the ability of these mutants to form full CP, we mixed them with a form of the WT β where the active site threonine was substituted to alanine (β_{TtoA}) rendering this mutant catalytically dead but with a WT propeptide sequence. Here, any CP activity observed upon mixing the two forms of β must be derived from the flexible loop mutants. This would also indicate that (1) those mutants are successfully incorporated into the CP and (2) they retain the capacity for autocatalytic cleavage of the propeptides. When β_{s11} was reconstituted with α and β_{TtoA} , we observed a substantial amount of active CP with a 1:1 ratio of $\beta_{s11}:\beta_{TtoA}$ (Figure S2A). Similar experiments with the extended loop mutants (β_{e11} and β_{e12}) showed very different results (Figure S2B). The 1:1 ratio showed very little CP. Counterintuitively, when reducing the relative amount of β_{e11} or β_{e12} mutant compared with the inactive β_{TtoA} (1:7) we saw increased amounts of CP and activity, indicating these mutants dramatically inhibit dimerization. Thus, although the levels of β subunits with WT active site residues (i.e., β_{e11} or β_{e12}) were lower, the CPs showed increased activity, which indicates that β_{e11} or β_{e12} mutants undergo efficient autocatalytic processing when they are successfully incorporated into CPs. In sum, region III affects the HP dimerization, but it has no apparent role in HP formation or in the autocatalytic processing of the β propeptide during CP maturation.

In the reconstitution experiments with an equimolar mixture of β_{e11} and β_{TtoA} , the HPs that form will have a distribution in the number of β_{e11} subunits versus β_{TtoA} , which should be centered on the 1:1 ratio if both are incorporated with the same efficiency into HPs. Based on our model, the extended region III would interfere with dimerization. Hence, the HPs that failed to dimerize should be those that are enriched in β s with an extended loop (β_{e11} or β_{e12}). On the other hand, HPs with lower levels of, for example, β_{e11} relative to β_{TtoA} should be preferentially incorporated into CPs. To test this, we performed a reconstitution using β_{e11} in combination with $\beta_{\Delta 23 TtoA}$ (which is the N-terminal deletion mutant in the β_{TtoA} , see Figure 2A). The truncation in β_{TtoA} allowed us to determine the ratios of the two different β s in CP and HP bands by using native SDS-PAGE 2D gel analyses (Figure S2C). The quantification of the relative abundance of $\beta_{\Delta 23 TtoA}$ to β_{e11} or β_{e12} mutant in HP and CP showed that indeed the HPs that failed to form CPs were enriched in the mutant with extended flexible loops (β_{e11} and β_{e12}) (Figure S2D).

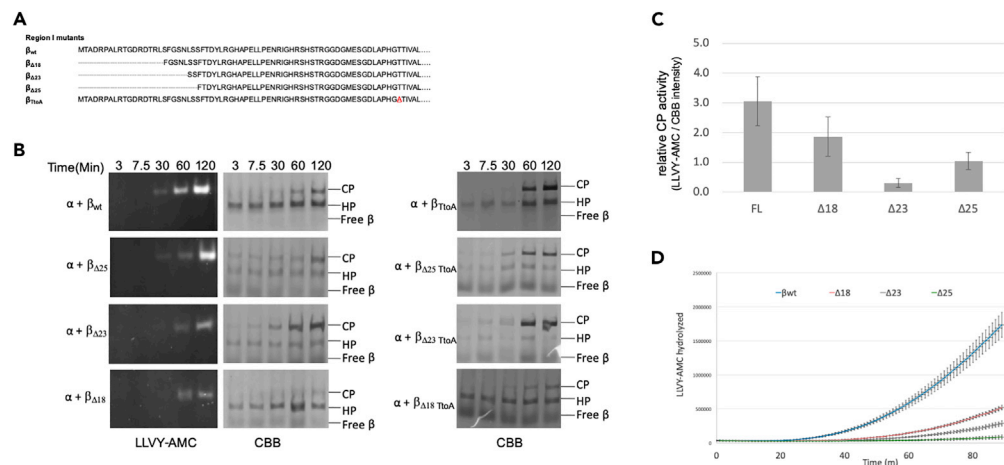


Figure 2. Region I of R.e. β Propeptide Inhibits HP Dimerization and Promotes Autocatalytic Activation

(A) Alignment of N-terminal sequence of the region I in wild-type and mutants of β .

(B) Equimolar amounts of wild-type or mutant proteasome subunits were reconstituted at 30°C for indicated time periods and analyzed as in (1D). Data indicate that the truncations exhibited slower maturation as there is less CP activity despite the more rapid formation of CP. LLVY-AMC activity assays for T10A mutants are not shown as these mutants lack hydrolytic activity.

(C) Quantification of LLVY-AMC and CBB for region I mutants carrying active site T in Figure 2B. The ratio of AMC fluorescence signal (reflects the activity) and the amount of assembled CP (as CBB scanned intensity) for the different forms of β subunits is shown here. Data here are the average of three independent experiments and are shown with standard error of mean (SEM, n = 3).

(D) Equimolar amount of α was reconstituted with indicated β subunits in the presence of LLVY-AMC. Accumulation of fluorescent AMC over time was monitored at 30°C. Data are the average of seven independent experiments and are shown with standard error of mean (SEM, n = 7).

Role of Region I in Autocatalytic Processing

In vitro reconstitution assays have shown that β subunits with large truncations from the N-terminal region I had defects in HP assembly. These mutants accumulated as either assembly intermediates or free subunits (Sharon et al., 2007). This indicates that the poorly conserved region I might be important in assembly. It has been noted before that poorly conserved regions can play important roles in proteasome assembly with $\beta 5$ in eukaryotes (Li et al., 2016). To assess the role of region I, we generated three N-terminal truncation mutants in β or β_{T10A} (Figure 2A). Reconstitution of these mutants showed a faster formation of CP (Figure 2B). This effect was most pronounced in the $\Delta 23$ and $\Delta 25$ truncations, where CP could be observed with CBB-stained native PAGE gels within 3 min of starting the reconstitution (Figure 2B, middle and right panels). Although there was a more rapid rate of dimerization, we surprisingly saw a delayed or reduced proteolytic activity (Figure 2B, left panels). We also quantified these gels and determined the ratio between 7-Amino-4-Methylcoumarin (AMC) fluorescent signal (which reflects activity) and the amount of assembled CP (as Coomassie Brilliant Blue (CBB) scanned intensity) (Figure 2C). With wild-type CP this number is higher than that of mutants, indicating wild-type CP has already gained full activity. For the N-terminal truncations this number is much lower, indicating the same amount of CP produced less AMC. This is consistent with the idea of slower maturation in the region I mutants. Consistently, in solution, real-time LLVY-AMC hydrolysis was dramatically reduced in all three region I mutants compared with β_{wt} (Figure 2D). This is particularly interesting as generally the dimerization of HPs appears to be tightly coupled with the activation of the CP. For example, despite extensive efforts, pre-holocomplexes (i.e., CP containing all unprocessed propeptides) have not been observed (Witt et al., 2006), unless mutants were used to block propeptide processing completely, such as K33A (Kwon et al., 2004) or T1A (Figure S1C).

The slow activation of the CP for region I mutants suggests that the N-terminal region plays a role in efficient autocatalytic cleavage of the propeptide from β . To test this, we analyzed reconstitutions on 2D gels where native PAGE was followed by SDS-PAGE, allowing us to determine the molecular weight (MW) of HP and CP components (Figure 3A). For wild-type α and β , two bands of different MW were visible in the area where HP migrates, namely, full-length tagged α (31 kDa) and full-length tagged β (32 kDa). For the CP, we

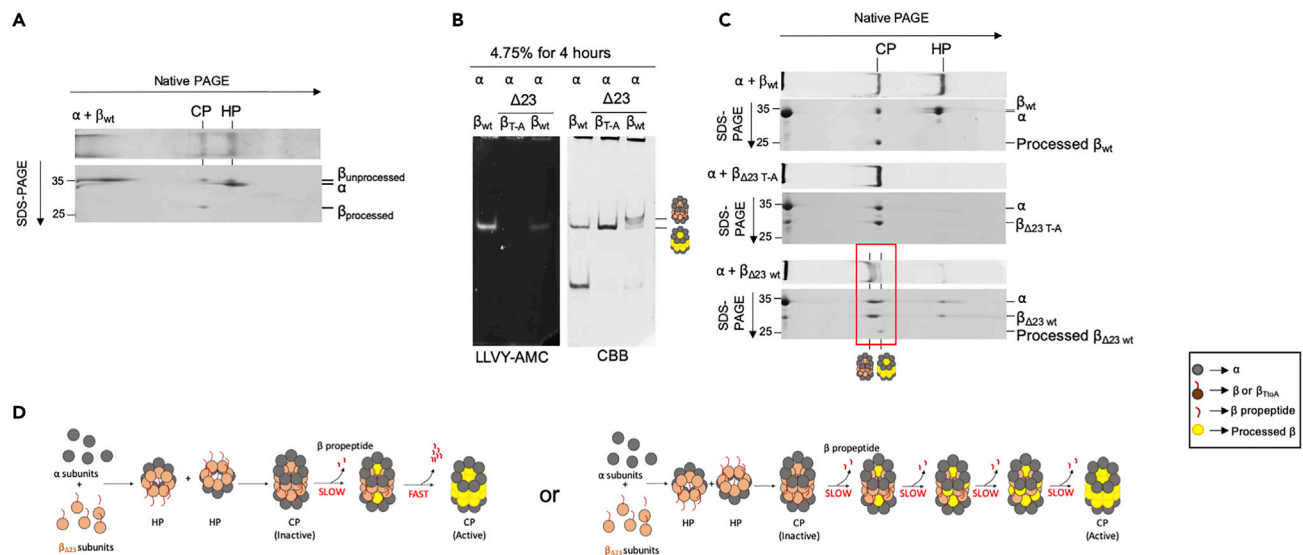


Figure 3. Cooperativity in the Activation of CP.

(A) Reconstitution of α with β_{wt} was evaluated by 2D native SDS PAGE. Top panel shows CBB stain of one lane from a native PAGE gel. Lower panel shows 2D-PAGE analyses where the native PAGE lane was separated on a second dimension of SDS-PAGE. The second dimension can reveal the MW of the polypeptides found in native PAGE CP and HP bands.

(B and C) (B) Reconstitution with indicated subunits (2 h, 30°C) were separated on native-PAGE (4.75% gel for 4 h) and visualized for hydrolytic activity using LLVY-AMC (left) and protein using CBB (right). (C) Lanes from (B) were also excised prior to visualization and separated on a second dimension of SDS-PAGE and described in (A).

(D) Slow and independent autocatalysis of CP, as indicated in the model of the right, should result in the presence of many intermediate forms of partial CP maturation. Since this is not observed, we propose a model that incorporates cooperativity in the autocatalytic processing of the propeptides resulting in an accumulation of either immature or mature CP (Left).

again saw full-length α but observed a shift in MW of β to ~ 25 kDa, consistent with the molecular weight of β after the propeptide has been removed. When we used the β_{T10A} mutant in our reconstitutions, both HP and CP were composed of full-length α and β (Figure S3). Reconstitution of α with $\beta_{\Delta 23}$ resolved under optimal conditions on native-PAGE showed three distinguishable migrating CP bands (Figure 3B). The top two bands migrated very closely and were not distinguishable on 2D gels, suggesting they are identical in composition and are either a staining artifact or two different conformations of the same complex. Only the faster migrating species had proteolytic activity, indicating this was matured CP, whereas the upper bands were pre-holoenzyme. Pre-holoenzyme has never been observed with autocatalytically competent mutants. The 2D gel electrophoresis of these samples (Figure 3C, lower right panel) showed two distinct spots for β associated with these two CP forms, which are distinct in both dimensions. The slower migrating CP species contained a spot of ~ 28 kDa corresponding to unprocessed β (confirmed by mass spectrometry to be 28,545 Da). The faster, active CP species shows a spot of ~ 24 kDa corresponding to processed β (24,030 Da by mass spectrometry). α Subunits are visible as a single broad spot that encompasses both forms of CP. Thus, truncation of the region I results in two distinct CP populations, one without proteolytic activity and exclusively composed of β with propeptides and one with proteolytically active CP where all β 's had undergone autocatalytic processing.

Cooperativity in Maturation

The existence of these two forms of CP is unexpected when considering reports that the maturation of CP in bacteria, archaea, and eukaryotes depends on the autocatalytic processing of the β propeptide for each active site (Baumeister et al., 1998). If all sites undergo proteolytic processing independently, then under conditions of slow maturation as we identified here, one would expect to see an array of intermediate species where some β 's are processed while others are not (Figure 3C, model on the right). However, we see a sharp delineation with CP populations either all containing propeptide or none at all (Figure 3B). This suggests the existence of a cooperative phenomenon during the maturation process. One possibility is that all active sites undergo autocatalytic processing simultaneously, e.g., as a result of simultaneous global conformational changes. Another explanation is that activation is a two-step process involving a slow initial

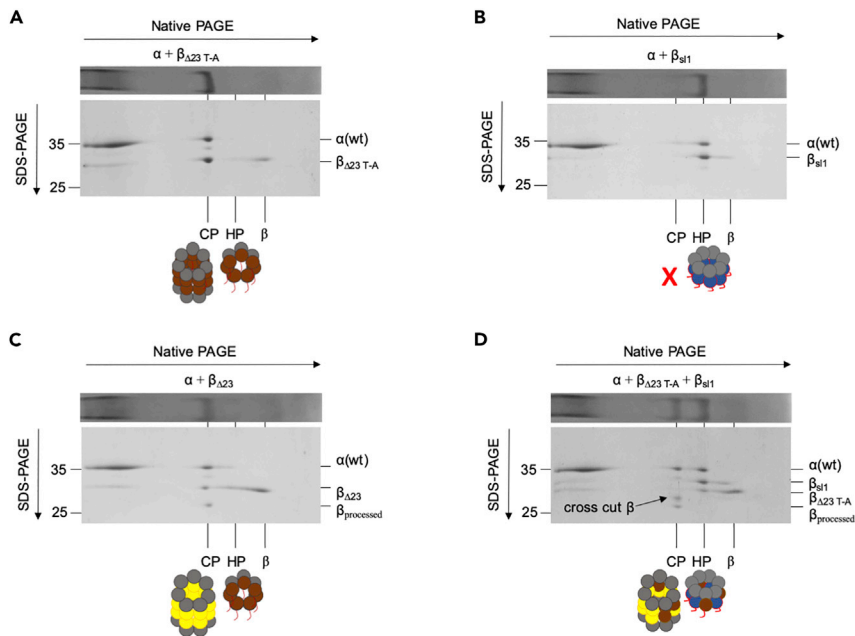


Figure 4. Active β Subunits Can Cleave the Propeptide of a Neighboring β Subunit

Four panels show 2D-PAGE analysis as described in (3A) for α reconstituted with $\beta_{\Delta 23 \text{ T-toA}}$ (A), β_{s11} (B), $\beta_{\Delta 23}$ (C), or an equimolar mixture of $\beta_{\Delta 23 \text{ T-toA}}$ and β_{s11} (D). All subunits were reconstituted at $4 \mu\text{M}$ concentrations, except in (D) where each β subunit was present at $2 \mu\text{M}$ for a total of $4 \mu\text{M}$ β . (A) and (B) serve as controls to show cleavage (indicated by arrow) observed in (D) requires the combination of β s. (C) was used to determine the size of completely processed β subunit present in an active CP.

step where one β subunit undergoes autocatalytic processing. This slow initial step would then be followed by a very rapid activation of all other active sites through an unknown mechanism that depends on the presence of one active β subunit (Figure 3D, model on the left).

Interestingly, it has been shown in eukaryotes that β subunits can cut the propeptides of neighboring β subunits. Indeed, since only three of the five β subunits with cleaved propeptides actually have functional active sites, eukaryotic β 's must and are able to cleave propeptides from other subunits, this cross-cutting can happen within a ring as well as between the two β rings (Chen and Hochstrasser, 1996; Groll et al., 1999; Schmidtke et al., 1996; Seemuller et al., 1996). However, the functional significance of this cutting is not well understood. We hypothesized that, if such cross-subunit cutting is conserved in bacterial CPs, it provides a potential mechanism for the cooperativity that we observed. Specifically, we could envision that an initial autocatalytic activation of one β subunit allows this active site to cut the neighboring propeptides. These cross-cuts cannot be at the site where autocatalytic cleavage occurs (owing to the distance between active sites, which is more than 24 \AA) but could happen within the propeptide sequence. The flexible region III seems like the most likely place for this to occur. A cleavage here would disconnect the structured part of the propeptide, which is physically constrained by binding to α subunits, from the part that needs to be properly positioned for autocatalytic cleavage. Cross-cutting would thus eliminate physical constraints that might slow activation. In this scenario, a single initial activation event would lead to rapid subsequent cleavage of all the propeptides in the CP.

Cleavage of Propeptides by Neighboring Beta Subunits

To test if the cross-cutting by β subunits is conserved, we combined equal amounts of $\beta_{\Delta 23 \text{ T-toA}}$ and β_{s11} and reconstituted with α (Figure 4). We used $\beta_{\Delta 23 \text{ T-toA}}$ as it cannot undergo autocatalytic cleavage and the truncation enables us to distinguish it on 2D SDS PAGE by size (Figures 4A and 4C). β_{s11} fails to form CP when reconstituted by itself with α for 2 h (Figure 4B). Thus, combining these two forms of β ensured that all active CP was derived from a heterogeneous mixture of both β forms (Figure 4D). Our analyses showed that the active CP was composed of α , completely processed β , and a novel band, which we refer to as "cross-cut β "

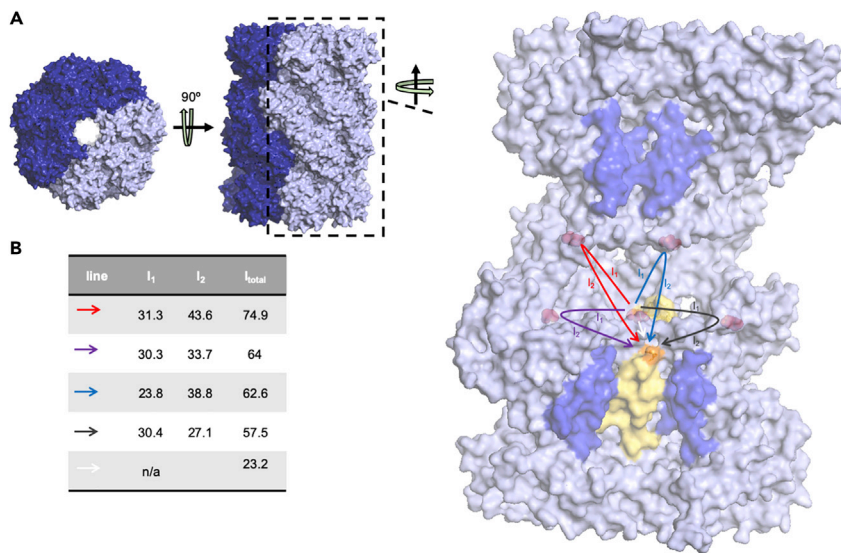


Figure 5. Distance between Active Sites Relative to Cleavage Sites

Space filled rendering of the structure of *R.e.* CP (PDB:1Q5R), visualized from two angles and an enlarged view of a subset of residues. The later provides an interior view exposing CP active site threonines (red) and the structured elements of several neighboring propeptides (dark blue and one in yellow). The orange residue indicates arginine –24, which is the last structured amino acid of region II (yellow). The flexible region III spans from the orange R to the active site in the center. The white arrow depicts the straight distance, but the observed cleavage by neighboring active sites requires region III to engage with neighboring active sites and thus span the distances (l_1) between the central active site and the different neighboring active site. l_2 indicates the distance from the neighboring active site T to R⁻²⁴. Total required distance l_{total} is indicated with colored arrows.

(B) Table of distances between active site and structured propeptide when propeptide would engage with specific neighboring sites as described in (A).

(Figure 4D arrow). The partially processed β subunits remained present even after extended incubation period (data not shown). The cross-cut β must represent $\beta_{\Delta 23 \text{ TtoA}}$ subunits that have been processed by a neighboring β_{s11} present in the CP. The neighboring site would be unable to remove the complete propeptide; hence, we see a higher MW band as compared with fully processed β subunits. The occurrence of this processing of $\beta_{\Delta 23 \text{ TtoA}}$ by β_{s11} shows that cross-processing by neighboring β subunits is evolutionarily conserved between bacteria and eukaryotes and provides a potential mechanism for the cooperativity we observed in the activation process.

To accurately determine the size of “cross-cut β ” and the exact site in propeptide region where the cleavage occurred during the cross-cutting described above, we determined the mass of proteins present in the reconstituted samples using MALDI-TOF (Figure S4A). As expected, we obtained peaks corresponding to the molecular weight of the subunits used for reconstitution and a strong peak corresponding to the molecular weight of β with completely processed propeptide (25,195.88 Da). Instead of obtaining a single peak for the partially processed band, we saw four peaks each separated in molecular weight by the equivalent of approximately two amino acids. Based on the molecular weight of these peaks, we determined the likely sites of cross-cutting (Figures S4B and S4c). We obtained a similar result when we repeated the reconstitution and analyses using β_{s11} and β_{TtoA} (data not shown). Interestingly, our analysis indicates that the cleavage we observed occurs in the flexible linker region III (Figure 5A). The identified cleavage sites are consistent with distances between active sites within the crystal structures (Figures 5B and 5C). In all, these data show that, in the CP, the propeptide can be cross-cut in the flexible region III by a neighboring β subunit.

Propeptide Cleavage across HPs

The distances between active site threonines within a β ring are comparable with the distances between active site threonines present across two different HPs. This would suggest that cross-cutting of a

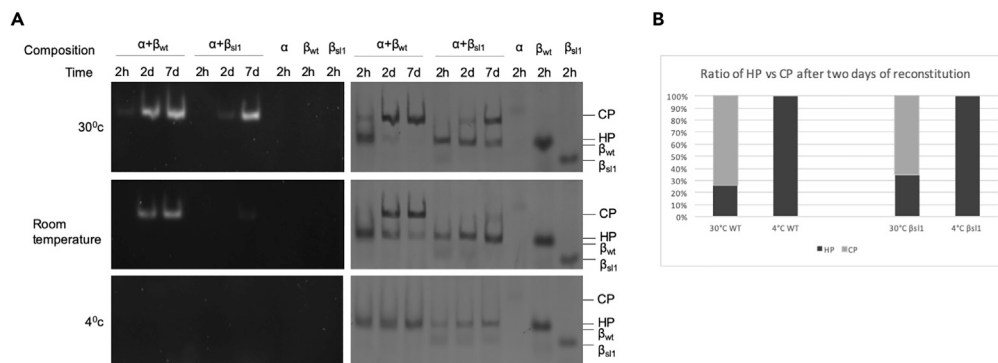


Figure 6. Assembly of Proteasomes Can Be Controlled by Reconstitution at Different Temperatures

(A) Reconstitution of indicated subunits was incubated at different temperatures (30°C, room temperature [$\sim 21^\circ\text{C}$], or 4°C) for 2 h, 2 days, or 7 days. Samples were separated by native PAGE and visualized for peptidase activity (using LLVY-AMC), followed by protein staining using CBB.

(B) Quantification of the ratio of HP versus CPs at 30°C and 4°C. The concentrations of HP (black) and CP (gray) were determined from quantitative analysis of native gel bands using GeneTools (Version 4.03.05.0) from SynGene.

propeptide can occur between two adjoining HPs, across the dimer interface. To test this, we developed an assay where we could form stable HPs and control their dimerization process, by conducting the reconstitution at different temperatures (30°C, room temperature [$\sim 21^\circ\text{C}$], and 4°C) over different times (Figure 6A). At 30°C, we observed wild-type HPs and active CPs as early as 2 h followed by a progressive increase in active CP and nearly complete absence of HPs after 2 days of reconstitution (Figure 6B). β_{s11} has a comparatively slower dimerization rate with active CPs forming after 2 days. The reconstitution at room temperature showed similar but slower dimerization rates. Interestingly, the reconstitution done at 4°C showed no dimerization even after an incubation of 7 days. We took advantage of this inability to dimerize at 4°C to generate HPs containing specific β mutants, HP $_{\beta s11}$ (HP containing α and β_{s11} ; does not dimerize with itself) and HP $_{\beta_{\Delta 23} T_{toA}}$ (HP containing α and $\beta_{\Delta 23} T_{toA}$; does not form active CPs). Using these in reconstitution mixtures at 30°C allowed the analyses of inter-HP interactions during the assembly process (Figure 7A). Reconstituted HP $_{\beta s11}$ with HP $_{\beta_{\Delta 23} T_{toA}}$ at 30°C was analyzed by native-PAGE over a 2-h period (Figure 7B). We saw an increased peptidase activity over time, indicating the dimerization of two different HPs. 2D analysis of reconstitution showed the presence of the “cross-cut β ” form as seen previously (Figure 7C). MALDI-TOF analyses confirmed the partially processed β 's of similar MWs as seen previously (data not shown). Thus, our data show that β subunits can be processed across the HP dimer interface as well within a ring.

Utilizing this new two-step reconstitution assay, we also re-evaluated dimerization rates of HPs formed by other region I mutants and obtained same results as seen in Figure 2B (Figure S5A). Similarly, consistent with what was seen in Figure 1D, when we used β_{el1} or β_{el2} to first make HPs and tested how they dimerize with themselves, we observed a failure to dimerize (Figure S5B). However, on reconstitution of HP $_{\beta_{el1}}$ with HPs that were formed with $\beta_{\Delta 23} T_{toA}$ (HP $_{\beta_{\Delta 23} T_{toA}}$), we observed the formation of active CP. Since HP $_{\beta_{el1}}$ or HP $_{\beta_{el2}}$ does not dimerize by itself and CP from $\beta_{\Delta 23} T_{toA}$ is inactive, the CPs that showed activity must be derived from HP containing β_{el1} or β_{el2} that dimerized with HP containing $\beta_{\Delta 23} T_{toA}$. The formation of active CP here is surprising because HPs that are formed by mixing of β_{el1} (or β_{el2}) and $\beta_{\Delta 23} T_{toA}$ in a 1:1 ratio failed to dimerize (Figure S5B, lane 1 and 4). This difference suggests that the single extended loop by itself can be accommodated in dimerization but dimerization is inhibited when extended loops are present on opposing HPs that meet. This is consistent with our model that the extended flexible region III can regulate dimerization by populating states outside of the HP where it can sterically clash with a dimerization partner.

DISCUSSION

The assembly of the CP starts with the formation of HPs. This is followed by two key events, the dimerization of HPs and the processing of propeptides to form active CP. In eukaryotes the dimerization is controlled both positively and negatively by some of the propeptides as well as specific assembly chaperones (Barraut et al., 2012; Ehlinger and Walters, 2013; Kock et al., 2015; Park et al., 2013; Roelofs et al., 2009; Satoh et al., 2014; Singh et al., 2014; Wani et al., 2015). As far as we know, bacterial genomes do not encode

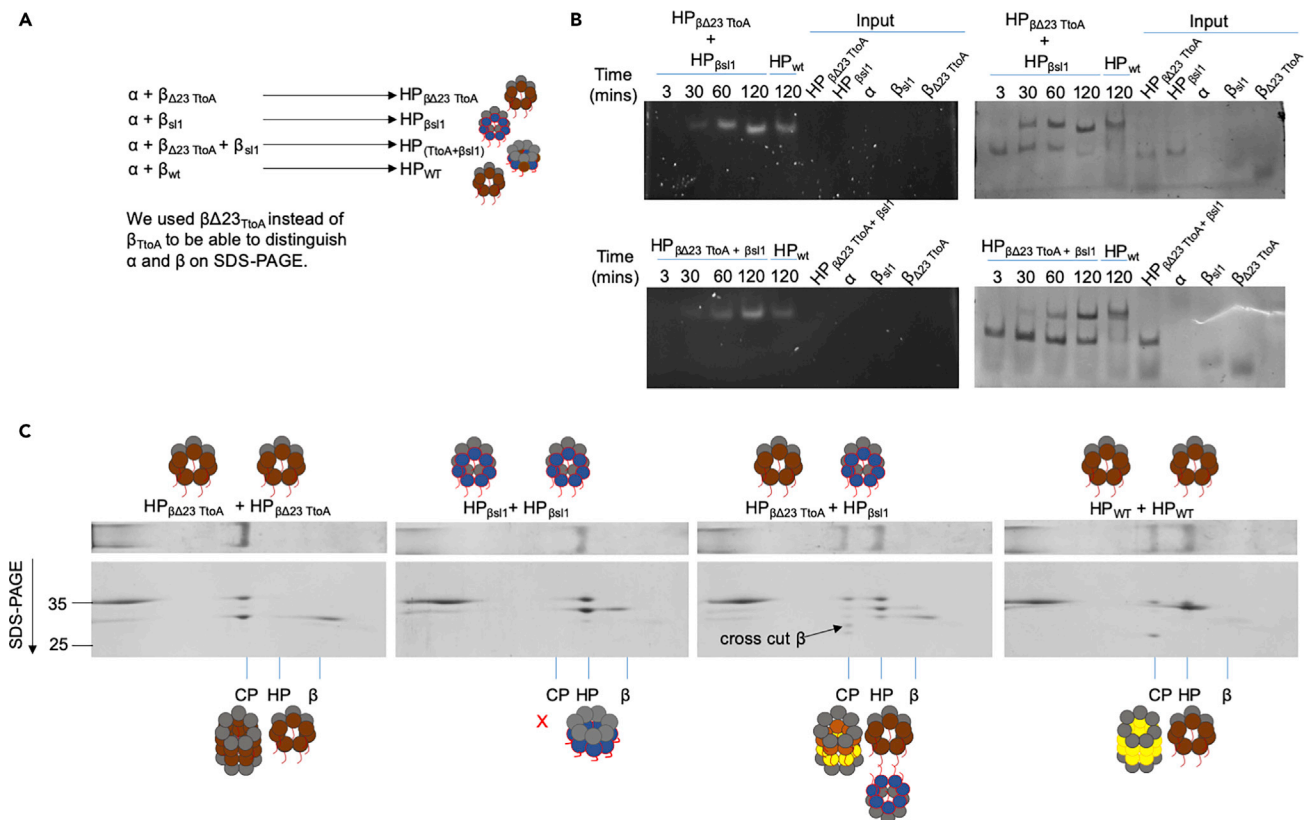


Figure 7. Active β Subunits Can Cleave the Propeptide of a β Subunit across the Dimerizing Interface

(A) Schematic representation of the HPs formed by reconstituting α with indicated β mutants at 4°C. Reconstitutions were incubated overnight at 4°C to maximize HP and minimize non-incorporated subunits.

(B) The dimerization of HPs was studied by mixing preformed HPs of indicated type and incubating samples over a 2-h time course at 30°C. CP formation and activation were monitored after native PAGE by peptidase activity assays using LLVY-AMC and subsequently CBB staining of gels.

(C) 2D-PAGE analysis of a 2-h reconstitution mixture consisting of indicated HPs was done by first separating on native gel followed by SDS-PAGE and are stained with CBB. Arrow indicates cross-cut of $\beta_{\Delta 23 \text{ TtoA}}$ across the dimerizing interface by active β present in the other HP.

assembly chaperones, but we report here that a flexible loop in the propeptide of *R.e.* β subunits also has a major impact on dimerization. We identified mutants that show rapid dimerization and delayed autocatalytic processing, showing for the first time that these two steps are distinct and allowing us to show the existence of the pre-holoenzyme as an intermediate in a complex capable of autocatalysis. Thus, dimerization by itself is not the trigger for autocatalysis. Our data indicate there is a cooperative process that coordinates autocatalysis, potentially involving propeptide trimming of neighboring β subunits by an initially autocatalytically activated β subunit.

Effect on HP Dimerization

Our data showed that region III (see Figure 1) of the *R.e.* β propeptide can negatively regulate the dimerization of HPs. This is a property strikingly similar to the role of several proteasome chaperones, as they sterically block steps in assembly (Barrault et al., 2012; Ehlinger and Walters, 2013; Kock et al., 2015; Park et al., 2013; Roelofs et al., 2009; Satoh et al., 2014; Singh et al., 2014; Wani et al., 2015). Particularly interesting is Ump1, a eukaryotic chaperone that assists CP assembly by inhibiting the premature dimerization of HPs (Li et al., 2007; Ramos et al., 1998). Our MD simulation data suggest that the residues in region III provide for a highly flexible loop that extends out of the HPs on the dimerization interface. Manipulating the length of this loop dramatically reduced HP dimerization rates when tested *in vitro*, whereas there was no apparent effect on HP formation. The inability to form CP, even after overnight incubation, for most mutants suggests a conservation of optimal size or charge on the flexible disordered loop to allow for proper repositioning of the propeptides during assembly. Although both the longer (e11 and e12) and shorter (s11 and s12) β mutants were compromised in dimerization as compared with wild-type, the

reconstitution experiments done using a mixture of these mutants with β_{TtoA} showed that the mutants with shorter flexible loops were fundamentally different from those with extended flexible loops. The smaller flexible loop mutant β_{s11} , for instance, exhibited a severely reduced ability to dimerize by itself but assembled efficiently when reconstituted with a normal-length region III as in β_{TtoA} . This might indicate a role for specific flexible loop interactions in bringing the dimerization interfaces together. Interestingly, we found that there is little to no inhibitory effect on the activation step in this mutant. On the other hand, the extended flexible loops of β_{e11} and β_{e12} showed little dimerization when mixed in a 1:1 ratio with β_{TtoA} , and decreasing the extended loop mutants in the ratio resulted in more CP activity. Consistently, we observed that the stalled HP was enriched in mutants with the extended flexible loop, suggesting that these mutants inhibit dimerization when they extend out from the HP at the dimerization surface. Interestingly, some recent data suggest the eukaryotic propeptide can also extend out from the HP structure, since the yeast $\beta 5$ propeptide can be cross-linked to $\beta 4$ lysine 28, which faces outside of the HP (Kock et al., 2015). These results thus strengthen our model of conservation of the length of flexible loop in β subunits and the function of beta propeptides in regulating dimerization rates.

Competent Pre-holoc complexes

Studies in *R.e.* mutants have suggested a mechanism where the assembled proteasomes mature into active proteasomes following a switch (Witt et al., 2006). Here, the interactions between two helices of the dimerizing HPs (H3 and H4) drive the positioning of specific loops (S2-S3) of β subunits to act as an activation switch by allowing the cleavage of propeptides from β subunits. When we introduced N-terminal truncations in the β subunit propeptide Region I, we unexpectedly observed a slower maturation of the CP that cannot be readily explained based on our current understanding of CP assembly. These findings highlight a positive role for region I in stimulating the autocatalytic processing of the β propeptide. These mutants also allowed us to capture, for the first time, the pre-holoproteasome complex containing β subunits that are fully competent for autocatalysis. This complex showed a slower migration on native gels compared with the CP formed by catalytically dead β subunits (T1A or K33A [Kwon et al., 2004; Witt et al., 2006]), suggesting that it is a distinct structure different from previously observed and crystalized structures. Early studies postulated the autocatalytic cleavage of β to be a rate-limiting step (Zühl et al., 1997a, 1997b), which would imply that the pre-holoproteasome is a relatively abundant intermediate. However, this intermediate has always remained elusive when using autocatalytically competent forms of β (Sharon et al., 2007; Witt et al., 2006), and propeptide-containing CP crystal structures all from autocatalytically defective mutants are very similar to structures of the active CP.

Based on the crystal structure, region I is expected to extend in the direction away from the loop region mentioned above and toward the gate region of the α ring. A similar direction has been observed for the $\beta 5$ propeptide in yeast, which can be cross-linked to $\alpha 6$, and thus this propeptide seems to be oriented toward the α ring as well (Kock et al., 2015). Nevertheless, deletion of the N-terminal portion of the $\beta 5$ propeptide also caused maturation defects (Li et al., 2016). Considering the orientation of this propeptide region, any mechanism that allows the N-terminal portion of the propeptide to stimulate maturation would involve long-distance allosteric changes. Such allosteric pathways have been proposed for cylindrical proteases like CPs (Huber et al., 2016; Kleijnen et al., 2007; Shi and Kay, 2014; Wani et al., 2015). Alternatively, we could envision that the extra mass from the propeptide in the tight cavity of the CP might destabilize the propeptide structures that have formed in HP, thereby facilitating the proper orientation of the propeptide for autocatalytic processing. In support of this idea, the proper positioning of propeptide residues in the β active site pocket is crucial for propeptide processing (Ditzel et al., 1998; Li et al., 2016).

Cooperativity

Some of our mutants showed a slow autocatalytic activation. If each of the 14 active sites in one CP underwent activation independently of others, intermediate forms of the CP (e.g., with 50% processed and 50% unprocessed active sites) would be expected (see model in Figure 3D). Our inability to detect these and our observation that these CPs are either completely active or completely inactive suggest that a specific mechanism is responsible for the (almost) simultaneous processing of the propeptides by all β subunits. A possible underlying mechanism that could explain these observations is a two-step process. Here, an initial slow step involves the autocatalytic cleavage of a propeptide from one of the fourteen β subunits in CP. This triggers a second step, where the newly created active site initiates hydrolysis of the propeptides of neighboring β subunits. This hydrolysis cannot be at the active site threonine but instead would be upstream in the propeptide. This is consistent with our observation that one β can cut the flexible loop of a neighboring β propeptide either within the ring or across the HP

dimer interface. The cross-processing of propeptides could stimulate autocatalytic cleavage by the β subunit with the truncated propeptide if the truncation removed physical constraints that limited the proper positioning required for autocatalysis of the propeptides (Ditzel et al., 1998; Huber et al., 2016), thus triggering a rapid phase of complete cleavage and activation.

Evolutionary Perspective

The short-length propeptides found in archaea are not required during assembly. Nevertheless, the longer propeptides found in bacterial species like *R.e.* used in this paper have been proven indispensable for assembly. Similarly, in eukaryotes, five of the β subunits retain their propeptides. One proposed function of the propeptides is to protect the active site threonine from N-terminal acetylation (Chen and Hochstrasser, 1996); however, this does not explain two of the eukaryotic propeptides, as they are found on β subunits that have lost a functional active site. Clearly, the presence of these propeptides in eukaryotes has been conserved in evolution. Here we report a functional conservation between bacterial and eukaryotic CP assembly: in both, the dimerization of HPs is tightly controlled. Our observation that the propeptides influence the transition from pre-holoproteasomes to active holoproteasomes is intriguing. This transition is likely to involve a conformational change of the CP, as both molecular species can be separated on native gel. In yeast, there is also a conformational change that triggers maturation and the exchange of immature CP binding from the chaperone Pba1/2 to the Regulatory Particle (Kock et al., 2015; Wani et al., 2015). How this switch and maturation are connected remains poorly understood and has been proposed to involve the degradation of Ump1. Our data suggest the propeptides themselves can be another important factor. It may thus be that the propeptides on the catalytically inactive $\beta 6$ and $\beta 7$ subunits in eukaryotes are retained in part because they play a role in this conformational switch.

As mentioned in the [Introduction](#), *Mycobacterium tuberculosis* (*M.tb.*), the causative agent of tuberculosis (Tb), is a close relative of *R.e.* Tb remains a global health threat, and in particular the emergence of antibiotic-resistant strains of *M.tb.* necessitates the development of new drugs to treat this disease. The *M.tb.* CP is essential for pathogenicity and is a validated drug target (Lin et al., 2009; Totaro et al., 2017). Thus, the development of drugs that interfere with *M.tb.* CP function or formation have the potential to be potent drugs against this disease. A key requirement for a successful drug that targets the *M.tb.* CP is the ability to target the bacterial CP but not host proteasomes. The evolutionary perspective presented above suggests that there are unique aspects of bacterial CP assembly that could be leveraged in the development of therapeutics for Tb. In any case, a full understanding of the evolutionary conservation of CP assembly mechanisms between humans and bacteria will be critical to further exploration of CP biogenesis as a drug target.

Limitations of the Study

- A limitation of the current work is that our assays assume that there is no further processing of the propeptides during electrophoresis at 4°, or, if there is, the processing mechanisms in the gel are not substantially different from those that occur in solution. We believe this assumption is reasonable as we observe species where no CP propeptides have been processed at all, whereas the same assay at 30° would lead to full propeptide processing.
- We used specific mutants that allowed us to expose the existence of cooperativity. However, owing to technical limitations in the time resolution we have not been able to show this occurrence in a wild-type background.
- The occurrence of cross-processing of propeptides provides an intriguing mechanism for the observed cooperativity and could also provide a rationale for the need and existence of propeptides in non-catalytic β subunits in eukaryotes. However, a complete causal relationship remains to be established.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jeroen Roelofs (jroelofs@kumc.edu).

Materials Availability

All unique/stable reagents generated in this study will be made available on request, but we may require a payment and/or a completed Materials Transfer Agreement if there is potential for commercial application.

DATA AND CODE AVAILABILITY

This study did not generate datasets or code.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101090>.

ACKNOWLEDGMENTS

We thank the Roelofs lab members for their support and valuable discussions and Dr. Tomich and Susan K. Whitaker at biotechnology core facility in the Department of Biochemistry and Molecular Biophysics at the Kansas state University for the MALDI-TOF analyses and expert technical advice. This work was supported by grants from the National Institutes of Health (National Institute of General Medical Science: K-INBRE program P20GM103418, R15GM112142, and R01GM118660 to J.R.) and National Science Foundation (MCB award 1412262 to E.D.).

AUTHOR CONTRIBUTIONS

Conceptualization, E.J.D. and J.R.; Methodology, A.S., P.I., S.K., and W.I.; Investigation, A.S., P.I., A.B., F.M.G.K., A.V., P.I., and A.K.; Writing - Original Draft, A.S. and J.R.; Writing - Review & Editing, A.S., P.I., A.B., W.I., E.J.D., and J.R.; Funding Acquisition, W.I., E.J.D., and J.R.; Supervision, A.S., W.I., E.J.D., and J.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 20, 2020

Revised: April 7, 2020

Accepted: April 16, 2020

Published: May 22, 2020

REFERENCES

- Barber, D.L., Sakai, S., Kudchadkar, R.R., Fling, S.P., Day, T.A., Vergara, J.A., Ashkin, D., Cheng, J.H., Lundgren, L.M., Raabe, V.N., et al. (2019). Tuberculosis following PD-1 blockade for cancer immunotherapy. *Sci. Transl. Med.* *11*, 1–7.eaat2702.
- Barrault, M.B., Richet, N., Godard, C., Murciano, B., Le Tallec, B., Rousseau, E., Legrand, P., Charbonnier, J.B., Le Du, M.H., Guerois, R., et al. (2012). Dual functions of the Hsm3 protein in chaperoning and scaffolding regulatory particle subunits during the proteasome assembly. *Proc. Natl. Acad. Sci. U S A* *109*, E1001–E1010.
- Baumeister, W., Walz, J., Zuhl, F., and Seemuller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell* *92*, 367–380.
- Becker, S.H., and Darwin, K.H. (2017). Bacterial proteasomes: mechanistic and functional insights. *Microbiol. Mol. Biol. Rev.* *81*, 1–20.e00036-16.
- Budenholzer, L., Cheng, C.L., Li, Y., and Hochstrasser, M. (2017). Proteasome structure and assembly. *J. Mol. Biol.* *429*, 3500–3524.
- Cerda-Maira, F.A., Pearce, M.J., Fuortes, M., Bishai, W.R., Hubbard, S.R., and Darwin, K.H. (2010). Molecular analysis of the prokaryotic ubiquitin-like protein (Pup) conjugation pathway in *Mycobacterium tuberculosis*. *Mol. Microbiol.* *77*, 1123–1135.
- Chen, P., and Hochstrasser, M. (1996). Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* *86*, 961–972.
- Creighton, T.A. (1983). *Proteins: Structures and Molecular Properties* (W.H. Freeman).
- Darwin, K.H., Ehrt, S., Gutierrez-Ramos, J.C., Weich, N., and Nathan, C.F. (2003). The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. *Science* *302*, 1963–1966.
- Ditzel, L., Huber, R., Mann, K., Heinemeyer, W., Wolf, D.H., and Groll, M. (1998). Conformational constraints for protein self-cleavage in the proteasome. *J. Mol. Biol.* *279*, 1187–1191.
- Ehlinger, A., and Walters, K.J. (2013). Structural insights into proteasome activation by the 19S regulatory particle. *Biochemistry* *52*, 3618–3628.
- Gandotra, S., Schnappinger, D., Monteleone, M., Hillen, W., and Ehrt, S. (2007). In vivo gene silencing identifies the *Mycobacterium tuberculosis* proteasome as essential for the bacteria to persist in mice. *Nat. Med.* *13*, 1515–1520.
- Groll, M., Heinemeyer, W., Jager, S., Ullrich, T., Bochtler, M., Wolf, D.H., and Huber, R. (1999). The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. *Proc. Natl. Acad. Sci. U S A* *96*, 10976–10983.
- Hirano, Y., Kaneko, T., Okamoto, K., Bai, M., Yashiroda, H., Furuyama, K., Kato, K., Tanaka, K., and Murata, S. (2008). Dissecting beta-ring assembly pathway of the mammalian 20S proteasome. *EMBO J.* *27*, 2204–2213.
- Hu, G., Lin, G., Wang, M., Dick, L., Xu, R.M., Nathan, C., and Li, H. (2006). Structure of the *Mycobacterium tuberculosis* proteasome and mechanism of inhibition by a peptidyl boronate. *Mol. Microbiol.* *59*, 1417–1428.
- Huber, E.M., Heinemeyer, W., Li, X., Arendt, C.S., Hochstrasser, M., and Groll, M. (2016). A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nat. Commun.* *7*, 10900.
- Imai, K., and Mitaku, S. (2005). Mechanisms of secondary structure breakers in soluble proteins. *Biophysics (Nagoya-shi)* *1*, 55–65.

- Kleijnen, M.F., Roelofs, J., Park, S., Hathaway, N.A., Glickman, M., King, R.W., and Finley, D. (2007). Stability of the proteasome can be regulated allosterically through engagement of its proteolytic active sites. *Nat. Struct. Mol. Biol.* **14**, 1180–1188.
- Kock, M., Nunes, M.M., Hemann, M., Kube, S., Dohmen, R.J., Herzog, F., Ramos, P.C., and Wendler, P. (2015). Proteasome assembly from 15S precursors involves major conformational changes and recycling of the Pba1-Pba2 chaperone. *Nat. Commun.* **6**, 6123.
- Kunjappu, M.J., and Hochstrasser, M. (2014). Assembly of the 20S proteasome. *Biochim. Biophys. Acta* **1843**, 2–12.
- Kusmierczyk, A.R., Kunjappu, M.J., Kim, R.Y., and Hochstrasser, M. (2011). A conserved 20S proteasome assembly factor requires a C-terminal HbYX motif for proteasomal precursor binding. *Nat. Struct. Mol. Biol.* **18**, 622–629.
- Kwon, Y.D., Nagy, I., Adams, P.D., Baumeister, W., and Jap, B.K. (2004). Crystal structures of the *Rhodococcus* proteasome with and without its pro-peptides: implications for the role of the pro-peptide in proteasome assembly. *J. Mol. Biol.* **335**, 233–245.
- Levitt, M. (1978). Conformational preferences of amino acids in globular proteins. *Biochemistry* **17**, 4277–4285.
- Li, D., Li, H., Wang, T., Pan, H., Lin, G., and Li, H. (2010). Structural basis for the assembly and gate closure mechanisms of the *Mycobacterium tuberculosis* 20S proteasome. *EMBO J.* **29**, 2037–2047.
- Li, X., Kusmierczyk, A.R., Wong, P., Emili, A., and Hochstrasser, M. (2007). beta-Subunit appendages promote 20S proteasome assembly by overcoming an Ump1-dependent checkpoint. *EMBO J.* **26**, 2339–2349.
- Li, X., Li, Y., Arendt, C.S., and Hochstrasser, M. (2016). Distinct elements in the proteasomal beta5 subunit propeptide required for autocatalytic processing and proteasome assembly. *J. Biol. Chem.* **291**, 1991–2003.
- Lin, G., Li, D., de Carvalho, L.P., Deng, H., Tao, H., Vogt, G., Wu, K., Schneider, J., Chidawanyika, T., Warren, J.D., et al. (2009). Inhibitors selective for mycobacterial versus human proteasomes. *Nature* **461**, 621–626.
- Maupin-Furlow, J.A., Humbard, M.A., Kirkland, P.A., Li, W., Reuter, C.J., Wright, A.J., and Zhou, G. (2006). Proteasomes from structure to function: perspectives from Archaea. *Curr. Top. Dev. Biol.* **75**, 125–169.
- Murata, S., Yashiroda, H., and Tanaka, K. (2009). Molecular mechanisms of proteasome assembly. *Nat. Rev. Mol. Cell Biol.* **10**, 104–115.
- Panfair, D., Ramamurthy, A., and Kusmierczyk, A.R. (2015). Alpha-ring independent assembly of the 20S proteasome. *Sci. Rep.* **5**, 13130.
- Park, S., Li, X., Kim, H.M., Singh, C.R., Tian, G., Hoyt, M.A., Lovell, S., Battaile, K.P., Zolkiewski, M., Coffino, P., et al. (2013). Reconfiguration of the proteasome during chaperone-mediated assembly. *Nature* **497**, 512–516.
- Ramos, P.C., Hockendorff, J., Johnson, E.S., Varshavsky, A., and Dohmen, R.J. (1998). Ump1p is required for proper maturation of the 20S proteasome and becomes its substrate upon completion of the assembly. *Cell* **92**, 489–499.
- Roelofs, J., Park, S., Haas, W., Tian, G., McAllister, F.E., Huo, Y., Lee, B.H., Zhang, F., Shi, Y., Gygi, S.P., et al. (2009). Chaperone-mediated pathway of proteasome regulatory particle assembly. *Nature* **459**, 861–865.
- Satoh, T., Saeki, Y., Hiromoto, T., Wang, Y.H., Uekusa, Y., Yagi, H., Yoshihara, H., Yagi-Utsumi, M., Mizushima, T., Tanaka, K., et al. (2014). Structural basis for proteasome formation controlled by an assembly chaperone nas2. *Structure* **22**, 731–743.
- Schmidtke, G., Kraft, R., Kostka, S., Henklein, P., Frommel, C., Lowe, J., Huber, R., Kloetzel, P.M., and Schmidt, M. (1996). Analysis of mammalian 20S proteasome biogenesis: the maturation of beta-subunits is an ordered two-step mechanism involving autocatalysis. *EMBO J.* **15**, 6887–6898.
- Seemuller, E., Lupas, A., and Baumeister, W. (1996). Autocatalytic processing of the 20S proteasome. *Nature* **382**, 468–471.
- Sharon, M., Witt, S., Glasmacher, E., Baumeister, W., and Robinson, C.V. (2007). Mass spectrometry reveals the missing links in the assembly pathway of the bacterial 20 S proteasome. *J. Biol. Chem.* **282**, 18448–18457.
- Shaw, D.E., Grossman, J.P., Bank, J.A., Batson, B., Butts, J.A., Chao, J.C., Deneroff, M.M., Dror, R.O., Even, A., Fenton, C.H., et al. (2014). Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 41–53.
- Shi, L., and Kay, L.E. (2014). Tracing an allosteric pathway regulating the activity of the HslV protease. *Proc. Natl. Acad. Sci. U S A* **111**, 2140–2145.
- Singh, C.R., Lovell, S., Mehzabeen, N., Chowdhury, W.Q., Geanes, E.S., Battaile, K.P., and Roelofs, J. (2014). 1.15 Å resolution structure of the proteasome-assembly chaperone Nas2 PDZ domain. *Acta Crystallogr. F Struct. Biol. Commun.* **70**, 418–423.
- Toste Rego, A., and da Fonseca, P.C.A. (2019). Characterization of fully recombinant human 20S and 20S-PA200 proteasome complexes. *Mol. Cell* **76**, 138–147 e135.
- Totaro, K.A., Barthelme, D., Simpson, P.T., Jiang, X., Lin, G., Nathan, C.F., Sauer, R.T., and Sello, J.K. (2017). Rational design of selective and bioactive inhibitors of the *Mycobacterium tuberculosis* proteasome. *ACS Infect. Dis.* **3**, 176–181.
- Wani, P.S., Rowland, M.A., Ondracek, A., Deeds, E.J., and Roelofs, J. (2015). Maturation of the proteasome core particle induces an affinity switch that controls regulatory particle association. *Nat. Commun.* **6**, 6384.
- Witt, S., Kwon, Y.D., Sharon, M., Felderer, K., Beuttler, M., Robinson, C.V., Baumeister, W., and Jap, B.K. (2006). Proteasome assembly triggers a switch required for active-site maturation. *Structure* **14**, 1179–1188.
- Zühl, F., Seemuller, E., Golbik, R., and Baumeister, W. (1997a). Dissecting the assembly pathway of the 20S proteasome. *FEBS Lett.* **418**, 189–194.
- Zühl, F., Tamura, T., Dolenc, I., Cejka, Z., Nagy, I., De Mot, R., and Baumeister, W. (1997b). Subunit topology of the *Rhodococcus* proteasome. *FEBS Lett.* **400**, 83–90.
- Zumla, A., George, A., Sharma, V., Herbert, R.H.N., Oxley, A., and Oliver, M. (2015). The WHO 2014 Global tuberculosis report—further to go. *Lancet Glob. Health* **3**, e10–e12.
- Zwickl, P., Kleinz, J., and Baumeister, W. (1994). Critical elements in proteasome assembly. *Nat. Struct. Biol.* **1**, 765–770.

iScience, Volume 23

Supplemental Information

Cooperativity in Proteasome Core

Particle Maturation

Anjana Suppahia, Pushpa Itagi, Alicia Burris, Faith Mi Ge Kim, Alexander Vontz, Anupama Kante, Seonghoon Kim, Wonpil Im, Eric J. Deeds, and Jeroen Roelofs

Supplementary figures:

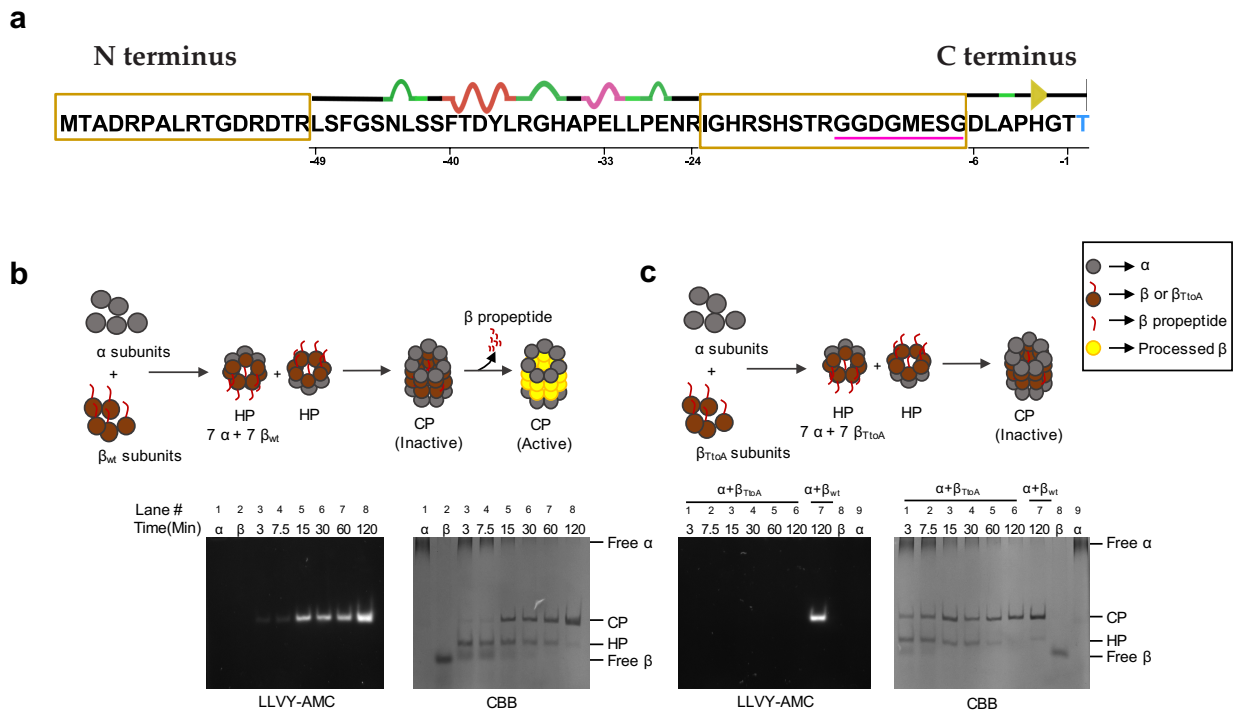


Figure S1. *In-vitro* assembly of the bacterial 20S core particle, Related to Figure 1.

- a.** Sequence view of the propeptide sequence of β subunit of *Rhodococcus erythropolis* (*R.e.*) with indicated structure elements adapted from PDB (Berman et al., 2000). The residues with missing electron density are highlighted in yellow boxes and are highly mobile. The flexible region of the propeptide (residues -7 to -14) is underlined in magenta and is a part of the region III that is seen at the half proteasomes interface in the simulations. The active site T (threonine) is shown in blue. This sequence view was created using the PDB (ID:1Q5R) sequence information from the crystal structure of *Rhodococcus erythropolis*. The secondary structure assignment is based on DSSP algorithm (Kabsch and Sander, 1983); the red structure is an alpha helix, pink structure represents a 3_{10} helix, green structures represent bends and turns and black lines indicate unassigned secondary structures.
- b.** Graphical representation of the *in-vitro* reconstitution experiment using purified α and β_{wt} . Equimolar amount of His-tagged α and β_{wt} proteasome subunits were incubated at 30 °C for indicated time-points and samples were separated on Native PAGE (bottom). The gels were stained for peptidase activity using the fluorogenic peptide suc-LLVY-amc and by Coomassie Brilliant Blue (CBB) to visualize all the protein complexes on the gel. The free α subunits in lane 1 tends to form

aggregates and run on top of gel, whereas β subunits by themselves migrated furthest in the gel and are visible as distinct band (lane 2). The slower migrating band in lane 3 represents the HP that formed rapidly after reconstitution as it lacks LLVY-AMC hydrolytic activity. As the reconstitution progresses, a slower migrating band representing CP appears above HP. The appearance of this band coincides with peptidase activity indicating this is active mature CP. With time, HP disappears and we see more of active CP.

- c. Same as b, except an inactive version of β_{TtoA} was used for reconstitution. Mutation of active site T to A renders the mutant unable to remove the propeptide and form active CP complex, also indicated by the absence of activity upon LLVY-AMC assay. Reconstitution mixture containing α and β_{wt} was used as a positive control for peptidase activity upon LLVY-amc assay.

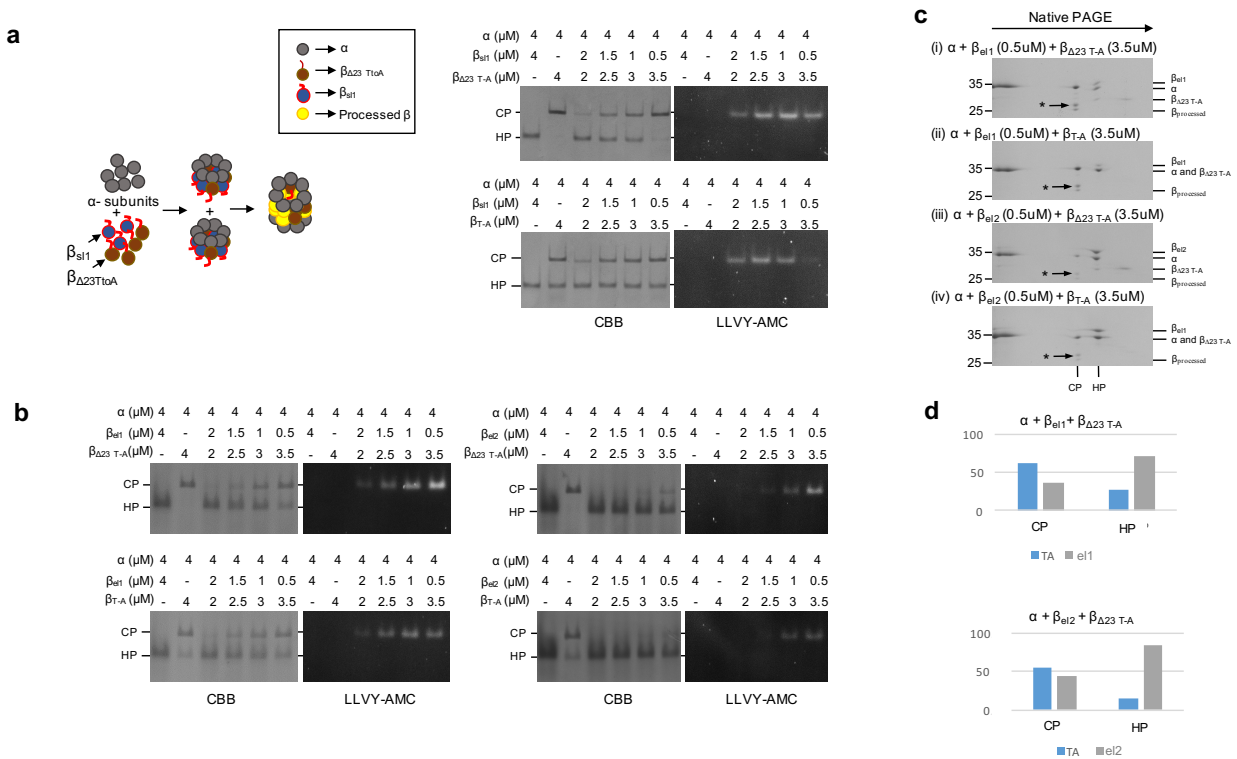


Figure S2. Incorporation of Region III mutants into mature CP upon mixing with β containing complete propeptide, Related to Figure 1.

- Graphical representation of the assay (left), LLVY-AMC assay and CBB stain of native-PAGE containing samples obtained after reconstituting α with β_{s11} and β_{T10A} at indicated molar concentrations for two hours (right).
- Reconstitution of β_{el1} and β_{el2} with $\beta_{\Delta23 T10A}$ at the indicated molar concentrations, analyzed by LLVY-AMC assay and CBB stain.
- 2D-PAGE analysis of indicated labelled samples reconstituted at 30°C for 2 hours. The samples were separated on native-PAGE, excised and separated on a second dimension using SDS-PAGE. Arrow indicates the $\beta_{\Delta23 T10A}$ that has been trimmed by other active site. The size is larger in MW then that of proteolytic active β forms, because it lacks the autocatalytic capacity to cleave at the propeptide-protein interface.
- Quantification of the relative abundance of indicated β subunits in HP and CP seen in set (i) and (iii) of fig S2c. Data indicate that the HPs that do not dimerize are enriched in β with extended loop, while HP with lower levels of β with extended loop preferably dimerize into CP.

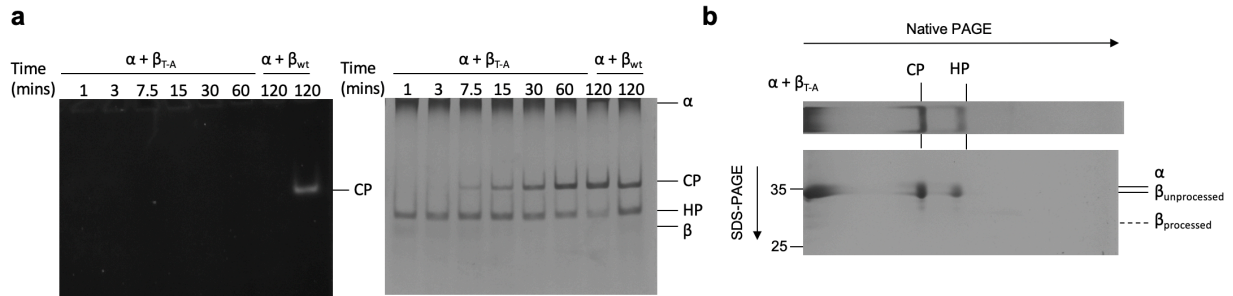


Figure S3. 2-D PAGE analysis of mixture of α and β_{TtoA} , Related to Figure 3.

- a. Time-course analysis to monitor the assembly of α and β_{TtoA} reconstitution mixture on 1D native gel.
- b. Sample from 1-hour reconstitution was also separated on a second dimension by using SDS-PAGE, as described before. Since β_{TtoA} is an inactive form of β , we do not see any band representing $\beta_{processed}$ in CP.

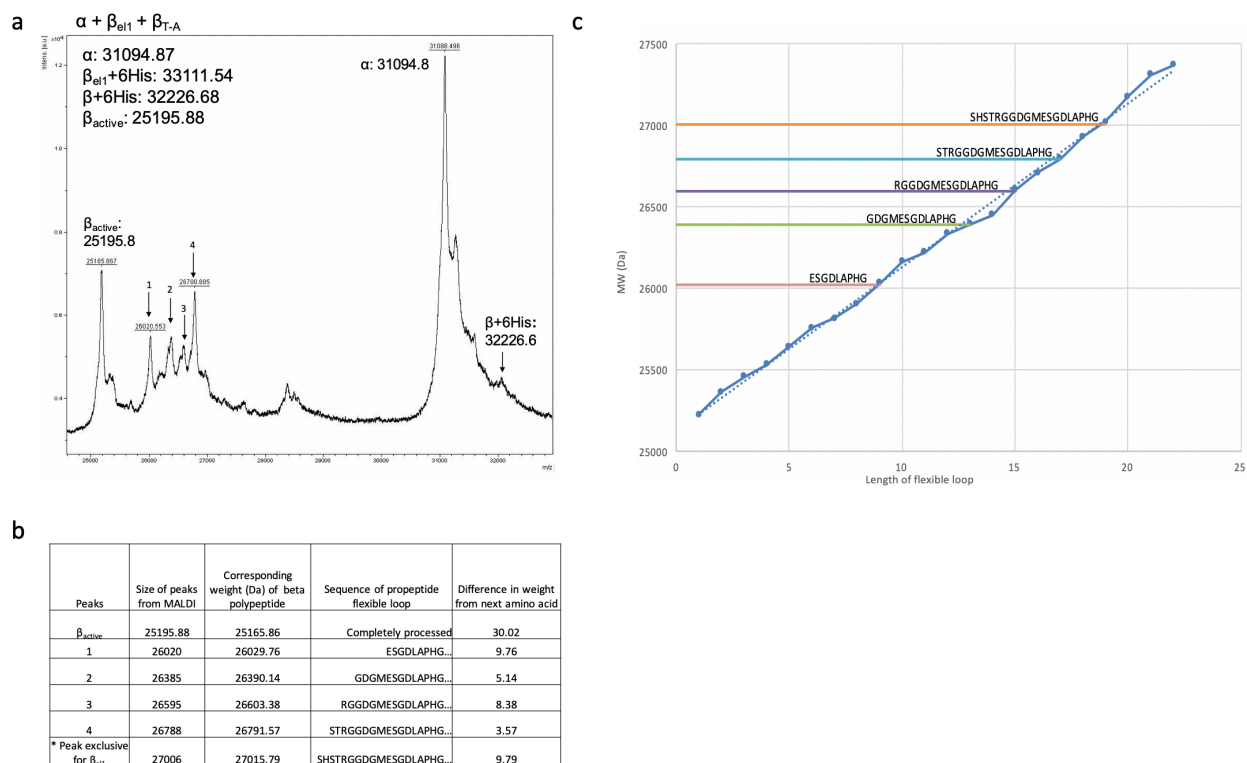


Figure S4. Analysis of the partially processed band formed during CP activation, Related to Figure 5.

- To estimate the size of partially processed band (Fig 4d), we performed MALDI analysis of the reconstitution mixture containing 4 μ M of α , 2 μ M of β_{el1} and 2 μ M of β_{TtoA} . Three out of four numbered peaks were also obtained upon analysis of reconstitution mixture of α with β_{sl1} and $\beta_{\Delta 23 TtoA}$ (data not shown).
- Analysis of masses identified in (a). Comparison of the molecular weight of the peaks obtained by MALDI with the predicted molecular weight of β polypeptide truncated at different amino acid residues as shown in (c) where molecular weight v/s the length of flexible loop was plotted (counted backwards from -1 position). Peaks unique to the reconstitution mixture consisting of either β_{el1} or β_{sl1} are indicated in the table.

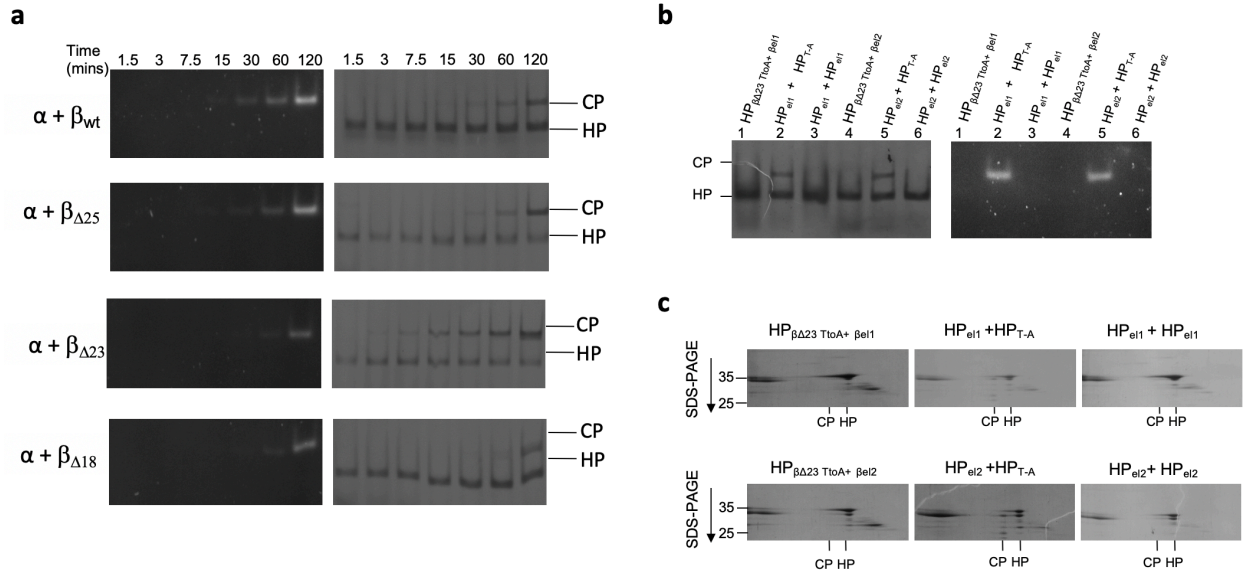


Figure S5. Delay in propeptides autocatalysis for N-terminal β truncations is also observed when preassembled HPs are allowed to dimerize at 30 °C, Related to Figure 7.

- Equimolar amounts of α with indicated labelled Region I truncations were reconstituted at 4 °C overnight to allow for the formation of HPs. Next, HPs were reconstituted in 1:1 ratio at 30 °C and analyzed at indicated time points using peptidase activity assay and CBB staining following native-PAGE separation.
- Similar to (a), dimerization using HPs formed by Region III mutants β_{el1} and β_{el2} was analyzed on native-PAGE
- 2D-PAGE analysis of indicated lanes from (b) was done as described before.

Supplemental table 1. plasmids and primers used in this study, Related to Figure 1 and 2.

Plasmid	Template	Primers	Protein expressed	Abbreviation
pJR659	pET22B	FP (5' GCGATGTATACATATGACCGCGGATCGTCCG 3') RP (5'AAAATATATA CTCGAGACGCGCGCTGCCGCC 3')	N-(PrcB1 ORF)-6xHis-C	β_{wt}
pJR662	pTBSG	FP (5'-TACTTCCAATCCAATGCGATGACCATGCGTACTACG-3') RP (5'-TTATCCACTTCCAATGTTAGTCCTGAGTGTCCGGCGG-3')	pTBSG-PrcA	α
pJR681	pJR659	FP: pRL437 (5'-GCCGCACGGCgcgACCATTGTGGCGCTGAC-3') RP: pRL438 (5'-GCCAGGTCACCGCTCTCCATG-3')	N-(PrcB1 _{T-A} ORF)-6xHis-C	β_{T-A}
pJR770	pJR659	FP: pRL519 (5'-gaaagcgggTACCTGGCGCCGCACGGC-3') RP: pRL520 (5'-gccaccgccACCGATACGGTTTTCCGGCAGCAG-3')	N-(PrcB1 _{sl1} ORF)-6xHis-C	β_{sl1}
pJR771	pJR659	FP: pRL521 (5'-ggcgggTACCTGGCGCCGCACGGC-3') RP: pRL522 (5'-accgccACCGATACGGTTTTCCGGCAGCAG-3')	N-(PrcB1 _{sl2} ORF)-6xHis-C	β_{sl2}
pJR772	pJR659	FP: pRL523 (5'-ggcgggTACCTGGCGCCGCACGGC-3') RP: pRL524 (5'-gtcacctccgacgaccATCGCCACCACGGGTGCT-3')	N-(PrcB1 _{el1} ORF)-6xHis-C	β_{el1}
pJR773	pJR659	FP: pRL525 (5'-cgggatggacgcggtggagatGGCATGGAGAGCGGTGAC-3') RP: pRL526 (5'-ccacggccgctcacctccgacgaccATCGCCACCACGGGTGCT-3')	N-(PrcB1 _{el2} ORF)-6xHis-C	β_{el2}
pJR775	pJR779	FP: pRL463 (5'-GCCGCACGGCACCACCATTGTGGCGCTGAC-3') RP: pRL438 (5'-GCCAGGTCACCGCTCTCCATG -3')	N-(Δ 18 PrcB1 ORF)-6xHis-C	$\beta_{\Delta 18}$
pJR776	pJR780	FP: pRL463 (5'-GCCGCACGGCACCACCATTGTGGCGCTGAC-3') RP: pRL438 (5'-GCCAGGTCACCGCTCTCCATG -3')	N-(Δ 23 PrcB1 ORF)-6xHis-C	$\beta_{\Delta 23}$
pJR777	pJR781	FP: pRL463 (5'-GCCGCACGGCACCACCATTGTGGCGCTGAC-3') RP: pRL438 (5'-GCCAGGTCACCGCTCTCCATG -3')	N-(Δ 25 PrcB1 ORF)-6xHis-C	$\beta_{\Delta 25}$
pJR779	pJR681	FP: pRL441 (5'- TTTGGCAGCAATCTGAGCAGC-3') RP: pRL449 (5'-CATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGG-3')	N-(Δ 18 PrcB1 _{TtoA} ORF)-6xHis-C	$\beta_{\Delta 18 T-A}$
pJR780	pJR681	FP: pRL442 (5'- AGCAGCTTTACCGACTACC-3') RP: pRL449 (5'-CATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGG-3')	N-(Δ 23 PrcB1 _{TtoA} ORF)-6xHis-C	$\beta_{\Delta 23 T-A}$
pJR781	pJR681	FP: pRL443 (5'- TTTACCGACTACCTGCGTGGTC-3') RP: pRL449 (5'-CATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGG-3')	N-(Δ 25 PrcB1 _{TtoA} ORF)-6xHis-C	$\beta_{\Delta 25 T-A}$

Transparent Methods

Plasmids: *Rhodococcus erythropolis* PrcA gene was amplified using pT7-7 α 1 (a generous gift from Wolfgang Baumeister) as template. The PCR amplified fragment was cloned in pTBSG plasmid (a generous gift from Philip Gao, Protein Production Group, University of Kansas) using ligation independent cloning to generate pTBSG α 1, (pJR662). PrcB gene was similarly amplified from pT7-7 β 1 plasmid (a generous gift from Wolfgang Baumeister). The PCR amplified fragment was cloned in Nde1 and Xho1 digested pET22B plasmid (a gift from Roberto DeGuzman, University of Kansas) resulting in generation of pET22B β 1, (pJR659). Plasmids used in this paper are listed in Supplementary Table 1. The accuracy of the plasmids generated was confirmed by sanger sequencing.

Protein expression and purification: For expression of recombinant proteins, a plasmid carrying the gene of interest was transformed into competent Rossetta cells (Novagen, BL-21 DE3 derived strain). The transformed cells were inoculated in 3 ml Luria-Bertani (LB) broth supplemented with 100 μ g/mL ampicillin and 34 μ g/ml chloramphenicol and incubated under shaking conditions at 37 °C overnight. This culture was used to inoculate 100 ml of the same media. At OD₆₀₀ of 0.6 IPTG was added to a final concentration of 0.1 mM IPTG and the culture was incubated at 30 °C for 3 hours under continuous shaking. Cells were collected by centrifugation and pellets were stored at -80 °C. For purification of the His-tagged proteins, the pellet was re-suspended in lysis buffer (5 mM Imidazole, 20 mM Tris-HCl [pH8], 100 mM NaCl, 0.75 mM EDTA) supplemented with protease inhibitors (Roche proteasome inhibitor cocktail) and cells were then lysed by French Press at 900 psi. The total lysate was cleared by centrifugation (SS34 rotor, 10,000 rpm, 20 minutes, 4 °C). The cleared lysate was then incubated with 100 μ l Roche complete His-Tag purification resin for 1 hour at 4 °C under rotation. The resin was collected by centrifugation (500 g, 4 minutes) and re-suspended in 5 ml wash buffer (15 mM Imidazole in PBS buffer (0.1 % Triton X-100, 0.02 % Sodium Azide)). The wash step was repeated 3 times. The His-tagged proteins were eluted step wise by resuspending the resin in 150 μ l of elution buffer (50 mM Tris-HCl [pH7.5], 100 mM NaCl, 1 mM EDTA, 0.25 mM DTT) with increasing concentrations of Imidazole (60 mM imidazole, 120mM, 250mM, 400mM and 500mM). Fractions containing the protein of interest were pooled and dialyzed against dialysis buffer (50 mM Tris-HCl [pH7.5], 100 mM NaCl, 1 mM EDTA, 0.25 mM DTT).

In vitro reconstitution assays: The reconstitution of the purified proteins and specified concentrations was done in the buffer containing 1 M Tris-HCl [pH7.5], 1 M MgCl₂, 0.5 mM ATP. The total reaction volume for assays was 25 μ l. Reconstitutions were done at

30 °C unless stated otherwise. The reconstitution of subunits to form HPs was done at 4 °C.

Native gels and two dimensional (2D) assays: To analyze reconstituted samples by Native-PAGE , native gel loading buffer was added to samples (50 mM Tris-HCl [pH 7.4], 50% glycerol, 60 ng/ml 1-xylene cyanol). Samples were separated on a 3.6% native gel and analyzed for the activity using in-gel using the substrate suc-LLVY-AMC as described previously (Elsasser et al., 2005; Roelofs et al., 2018). Next, gels were then stained with Coomassie Brilliant Blue. For 2D-PAGE analysis, after an initial native-PAGE, the lane containing the sample of interest was excised from gel, incubated with 1X SDS-Sample buffer for 10 minutes, and loaded on a second dimension to separate the proteins based on their size using SDS-PAGE (Roelofs et al., 2018). After electrophoresis the gels were stained using Coomassie Brilliant Blue.

MALDI-ToF analyses: Matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) on a Bruker Ultraflex II machine was used to determine masses of proteins. The protein solution was spotted with an equal amount of sinapic acid (Sigma-Aldrich, St. Louis, MO) matrix solution, and the samples were run in linear mode to determine protein mass.

Multiple sequence alignment of the bacterial β propeptide sequences: The β subunit of *Rhodococcus erythropolis* was used as an input query sequence to search the non-redundant NCBI protein sequence database for all bacterial species using the default parameters of BLASTp (Altschul et al., 1990; Sievers et al., 2011). The 256 sequences used for the multiple sequence alignment were selected based on sequence identity (more than 60% identity), best E-values, and manually curation (to avoid duplicates, incorrectly annotated sequences and hypothetical proteins). Next, to identify regions of similarity we performed a multiple sequence alignment (MSA) using the Clustal Omega package with the default parameters (Sievers et al., 2011). The resulting alignment, together with structural data from *R.e.* (PDB:1Q5R) and *Mtb* (PDB ID : 3MKA), allowed us to define three distinct regions in the propeptides. For *R.e.* Prc β 1 these were: i) Region I: From -65th to -43rd N terminal residues. ii) Region II : From -42nd to -27th residues. iii) Region III : From -26th to -4th residues (refer to the table in the Supplementary Material for more detail on these three regions). Calculation of average glycine content was performed by collecting all sequences from a given region (e.g. Region III) across all species, counting the number of glycine residues, and dividing by the total number of residues in that region across all species. In Region III, very few species deviated from the general observation of glycine enrichment (see Supplementary Material). Note that the conserved “PHG” motif at the C-terminal end of Region III was excluded from this analysis, since this final G residue is likely involved in promoting cleavage and thus would not be relevant to understanding the evolution of flexibility in Region III. Statistical

significance was calculated using a hypergeometric test, with the null hypothesis that the observed number of glycines in Region III would be obtained by placing residues at random in Regions I, II and III.

Molecular modelling and simulations: The systems were generated using the CHARMM-GUI (Jo et al., 2008) input generator by using the initial coordinates from the crystal structure of the mutant *R.e.* CP that is catalytically dead and thus retains the propeptide (PDB ID :1Q5R). HP models were generated by only selecting a HP from the starting CP structure. The missing electron density residues (residue IDs from -65 to -50 and -24 to -7) were modeled using the Galaxy-Fill tool as implemented directly in CHARMM-GUI (Coutsias et al., 2004). The proteins were solvated with water molecules in a periodic water box with 10 Å buffering distance between the protein surface and the box, using the TIP3P explicit water model (Jorgensen, 1983). Counterions of 0.1 M NaCl were added to neutralize the system.

The NAMD 2 (Phillips et al., 2005) program with the CHARMM C36m (Huang et al., 2017) force field was used to initiate all-atom Molecular dynamics simulations of the HP structure. Systems were equilibrated for 100ps using NVT (constant particle number, volume, and temperature) dynamics at 303.3 K without any restraints. The simulation systems measured about $143 \times 143 \times 143 \text{ \AA}^3$ with total $\sim 280,000$ atoms. For production runs we used NPT (constant particle number, pressure, and temperature) dynamics with temperature and pressure held at 303.3 K and 1 bar, respectively. In the production runs, three replicates were simulated for 100ns using a local cluster and one replicate was run on the Anton 2 machine for 2 μs (Shaw et al., 2009). For the 100 ns runs we used a 2 fs time step and trajectories were saved every 2 fs. All the equilibration and production runs were performed using the default values based on the CHARMM-GUI input scripts (Brooks et al., 2009; Jo et al., 2008; Lee et al., 2016). For the Anton2 production run, the NPT ensemble was used with pressure and temperature maintained at 1bar and 303.3 K respectively, and the time step was 2 fs. Trajectories were saved every 240 ps. All analysis of the resulting trajectories was performed using the CHARMM simulation package and VMD.

Root Mean Square fluctuations (RMSF) of the propeptide in MD simulations: RMSF is a metric that is used to measure the fluctuation in the position of an atom or group of atoms. To calculate it, we first aligned each frame the starting structure, in order to ignore motions that arise from translation or rotation of the entire structure during the simulation. After alignment, RMSF is calculated as the standard deviation in position about the mean position, where the mean is taken by averaging the position of the residue across all (aligned) frames. For every replicate the reported RMSF values are averaged for all the

backbone atoms of a given residue (C, O, N and C α) to calculate the residue-based RMSF. This residue-based RMSF is then averaged across all seven β propeptides to generate the RMSF for a given replicate.

Supplementary section for the Multiple Sequence Alignment

Using a BLASTP search algorithm with *R.e.* β sequence as the input yielded about 1000 hits with at least 60% sequence identity. This cutoff was chosen to include the β sequence from *M. tuberculosis*. Results were then filtered to remove redundant sequences, sequencing errors or misannotated sequences. This resulted in 256 distinct and non-redundant propeptide sequences. These were used in a Multiple sequence Alignment (MSA), using Clustal Omega with default parameters, which led to the identification of the glycine-rich Region III. Since MSA does not provide statistical estimates (Pearson, 2013), we conducted a hypergeometric test. This statistical test allowed us assess the significance of glycine enrichment in Region III and obtain a p-value that would indicate the chance that the glycine enrichment is not due to evolutionary pressure but resulted from random chance. This p-value is computing using the following formula:

$$p - value = P(x \geq b) = \sum_{k=b}^{\min(K,n)} \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

where N is the total number amino acids in 256 sequences (13108) and n represents the number of glycines in all three regions (1200). K represents the number of amino acids in Region III (4380) and b denotes the total number of glycines in Region III (810). This p-value represents the chance of observing glycine enrichment of 17.8% or more in Region III under a null hypothesis where the residues are assigned to Regions I, II and III completely at random. The calculated p-value was 3.94×10^{-142} indicating strong evidence against null hypothesis. Thus, its highly likely that glycine enrichment of Region III is caused by evolutionary pressure and highly unlikely that it occurred by chance.

Supplemental references

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235-242.

Brooks, B.R., Brooks, C.L., 3rd, Mackerell, A.D., Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., *et al.* (2009). CHARMM: the biomolecular simulation program. *J Comput Chem* 30, 1545-1614.

Coutsias, E.A., Seok, C., Jacobson, M.P., and Dill, K.A. (2004). A kinematic view of loop closure. *J Comput Chem* 25, 510-528.

Elsasser, S., Schmidt, M., and Finley, D. (2005). Characterization of the Proteasome Using Native Gel Electrophoresis. In *Ubiquitin and Protein Degradation, Part A*, pp. 353-363.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmuller, H., and Mackerell, A.D., Jr. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14, 71-73.

Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* 29, 1859-1865.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926-935.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-2637.

Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y., *et al.* (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* 12, 405-413.

Pearson, W.R. (2013). An introduction to sequence similarity ("homology") searching. *Curr Protoc Bioinformatics Chapter 3, Unit3 1*.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem 26*, 1781-1802.

Roelofs, J., Suppahia, A., Waite, K.A., and Park, S. (2018). Native Gel Approaches in Studying Proteasome Assembly and Chaperones. *Methods Mol Biol 1844*, 237-260.

Shaw, D.E., Dror, R.O., Salmon, J.K., Grossman, J., Mackenzie, K.M., Bank, J.A., Young, C., Deneroff, M.M., Batson, B., and Bowers, K.J. (2009). Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the conference on high performance computing networking, storage and analysis (ACM)*, p. 39.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol 7*, 539.