SI Appendix

An allosteric switch regulates *Mycobacterium tuberculosis* ClpP1P2 protease function as established by cryo-EM and methyl-TROSY NMR

Siavash Vahidi[1,2,3,4+], Zev A. Ripstein[2,4+], Jordan B. Juravsky[4], Enrico Rennella[1,2,3,4], Alfred L. Goldberg[5], Anthony K. Mittermaier[6], John L. Rubinstein[2,4,7], Lewis E. Kay[1,2,3,4*]

[+] These authors have contributed equally
[1]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S 1A8, Canada
[2]Department of Biochemistry, University of Toronto, Toronto, Ontario, M5S 1A8, Canada
[3]Department of Chemistry, University of Toronto, Toronto, Ontario, M5S 1A8, Canada
[4] Program in Molecular Medicine, Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada
[5]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02138, USA
[6]Department of Chemistry, McGill University, 801 Sherbrooke Street West, Montreal, Quebec, H3A 0B8, Canada
[7]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, M5G 1L7, Canada

*Correspondence to:

Lewis E. Kay (kay@pound.med.utoronto.ca)

Short title: NMR and cryo-EM of *M. tuberculosis* ClpP1P2

Key words: Protein homeostasis, Allostery, Intra-/Inter-ring cooperativity, ClpP handle region

Data deposition: Electron microscopy density maps have been deposited in the Electron Microscopy Databank (accession nos. EMD-21197, EMD-21198, EMD-21199). Atomic models have been deposited in the Protein Databank (PDB ID 6VGK, 6VGN, 6VGQ).

www.pnas.org/cgi/doi/10.1073/pnas.1921630117

Materials and Methods

Plasmids and constructs The *clpP1* (Uniprot: P9WPC5) and *clpP2* (Uniprot: P9WPC3) genes from *Mycobacterium tuberculosis* and the *clpP* gene from *S. aureus* (Uniprot: A6QF76) were synthesized by GenScript (Piscataway, NJ, USA) and cloned into the NdeI and BamHI sites of pET24a+ (Novagen, Madison, WI, USA). A cleavable N-terminal His_6-SUMO tag was introduced into all constructs via Gibson assembly. Point mutations were introduced using Quikchange site-directed mutagenesis (Agilent, Santa Clara, CA, USA).

Protein expression and purification MtClpP1, MtClpP2, and SaClpP were expressed heterologously and purified as detailed in previous work (1). Briefly, transformed BL21(DE3) Δ*clpP::cat E. coli* cells were grown in minimal M9 D_2O media supplemented with $^{15}NH_4Cl$ and d_7-glucose as the sole nitrogen and carbon sources, respectively. Cells were grown at 37 ˚C and protein overexpression was induced with 0.1 mM IPTG at OD_{600}=1.0 and allowed to proceed for ~18 hours at 18 ˚C. [U-^{2}H; Ileδ1-$^{13}CH_3$; Leu,Val-$^{13}CH_3/^{12}CD_3$; Met-$^{13}CH_3$]-labeled (referred to a $^{13}CH_3$ ILVM-labeled in text) and Val/Leu-γ1/δ1(proR) samples were produced as described previously (1). For the production of samples for cryo-EM analysis, cells were grown in Lysogeny Broth (LB) media at 37 ˚C and induced at OD_{600}=1 with 0.1 mM IPTG. Expression was allowed to proceed for 18 hours at 25 ˚C. Proteins were first purified using Ni-affinity chromatography. The Ni lysis/wash buffer contained 50 mM Tris, 300 mM KCl, 20 mM imidazole, and 10% glycerol adjusted to pH 7.0. The imidazole concentration was increased to 500 mM in the Ni elution buffer. Because MtClpP2 tends to aggregate at high imidazole concentrations, elution from the NiNTA column was collected in a Falcon tube containing ~25 mL of Ni lysis/wash buffer to immediately dilute the high imidazole concentration in the Ni elution buffer. This step was followed by cleavage of the SUMO tag using Ulp1 protease. This

mixture was concentrated using an Amicon Ultra-15 50K MWCO concentrator and subjected to size exclusion chromatography (SEC) on a Superdex 200 Increase 10/300 (GE Healthcare) column in SEC Buffer containing 50 mM imidazole, 100 mM KCl, 5 mM DTT adjusted to pH 7.0. Protein concentrations were determined spectrophotometrically (GdnCl-denatured protein) using extinction coefficients obtained from ExPASy's ProtParam web-based tool (https://web.expasy.org/protparam/). For NMR measurements and degradation assays, the samples were buffer exchanged into NMR Buffer containing 50 mM imidazole, 100 mM KCl, 1 mM TCEP adjusted to $pH^{measured}$ 7.0 prepared in 99.9% D_2O.

Preparation of mixed MtClpP1P2 complexes for NMR MtClpP1 rings containing mixtures of WT and S98A protomers were prepared following a procedure described previously (1, 2). Briefly, pure [U-^{2}H] WT and ILVM-labeled S98A MtClpP1 heptamers were purified separately and mixed to achieve a 95%:5% WT:S98A ratio. This protein mixture was concentrated to ~0.5 mL and unfolded and diluted with the addition of unfolding buffer containing 100 mM KCl, 50 mM imidazole, 6 M GdnCl, and 10 mM DTT at pH 7.0 to a final protein concentration of 500 μM. The complexes were reconstituted by drop-wise addition into a refolding buffer containing 100 mM KCl, 50 mM imidazole, 1 M arginine, 10 mM DTT, and 15% glycerol at pH 7.0 to a final GdnCl concentration of 300 μM. The refolded mixture was concentrated to ~1 mL using an Amicon Ultra-15 50K MWCO (Millipore) concentrator and then applied to a Superdex 200 Increase 10/300 (GE) column in SEC buffer. The resultant MtClpP1 rings eluted in a manner identical to that of pure WT MtClpP1 and there was no protein in the void volume. MtClpP1 fractions were pooled and mixed with a two-fold excess of [U-^{2}H] WT MtClpP2, and allowed to react overnight with a ten-fold excess of benzyloxycarbonyl-Gly-Leu-Phe-chloromethyl ketone (abbreviated GLF-CMK throughout the manuscript) (New England Peptide Inc, Gardner, MA,

USA) in the presence of 4 mM Bz-LL. Following verification of complete GLF-CMK modification of MtClpP1 active sites by ESI-MS, Bz-LL was removed by buffer exchanging the mixture into NMR Buffer.

NMR Spectroscopy All NMR measurements were performed at 18.8 T and 40 °C, using a Bruker AVANCE III HD spectrometer equipped with a cryogenically cooled x,y,z pulsed-field gradient triple-resonance probe. ^{1}H-^{13}C correlation spectra were recorded as HMQC datasets, exploiting a methyl-TROSY effect that is particularly beneficial for applications to high molecular weight proteins (3). Spectra were processed using the *NMRPipe* suite of programs (4), analyzed using scripts written in-house, and visualized using *Ccpnmr* (5).

Methyl group assignment Over 90% of the Ile (17/17), Leu (14/18), Met (8/8), and Val (7/7) correlations in spectra of MtClpP1, 40 ˚C, were assigned by a combined mutagenesis and NOE-based approach, as described previously (1, 6). ^{1}H-^{13}C HMQC spectra were recorded of the following MtClpP1 assignment mutants: M75L, M81L, M122L, I29V, I30V, I40V, I60V, I77V, I88V, I120V, I136V, I189V, V36I, V82I, V129I, V145I, V186I, V195I, L14M, L16M, L25M, L44M, L50M, L83M, L103M, and L126M. A methyl-TROSY based 3D NOE experiment (7) that records chemical shifts as ^{13}C[i]-NOE-^{13}C[j]-^{1}H[j] was measured on an ILVM-labeled sample of apo WT MtClpP1 with an NOE mixing time of 250 ms. To extend assignments to the *T* and *R* states of MtClpP1P2, NOE experiments were also recorded on samples that contained 1 mM ILVM-labeled MtClpP1 mixed with 1.2 mM [U-^{2}H] MtClpP2 in the presence or absence of 8 mM Bz-LL. The side chains of Leu in Bz-LL were uniformly deuterated to remove T1 noise. Stereospecific assignments of methyl groups of Leu and Val residues was obtained using the labeling approach described previously (8).

NMR data fitting Binding constants for the MtClpP1 and MtClpP2 interaction (with and without 8 mM Bz-LL) were obtained by titrating [U-^{2}H] MtClpP2 into an ILVM-labeled MtClpP1 sample (50 μM in monomer concentration). The decrease in the intensities of cross-peaks derived from the unbound state and the concomitant increase in cross-peak intensities from the bound-state can be quantified and subsequently fit as a function of MtClpP2 concentrations to the following expressions,

$$\text{Fraction of bound MtClpP1} = \frac{(P_1 + P_2 + K_d) - \sqrt{(P_1 + P_2 + K_d)^2 - 4P_1P_2}}{2P_1}$$

$$\text{Fraction of free MtClpP1} = 1 - \frac{(P_1 + P_2 + K_d) - \sqrt{(P_1 + P_2 + K_d)^2 - 4P_1P_2}}{2P_1}$$

where P_1 and P_2 are the total concentrations of MtClpP1 and MtClpP2 at each titration step, respectively, and K_d is dissociation constant. Peak intensities were extracted using the *NMRPipe* suite of programs (4). Intensities for a given residue obtained in this manner were subsequently normalized by the maximum peak intensity in the titration series, obtained either from the initial spectrum (no added MtClpP2) for fraction free MtClpP1 or from the final spectrum (maximum addition of MtClpP2) for fraction bound MtClpP1. This normalization procedure removes residue specific contributions to peak intensities, such as those derived from different transverse relaxation rates, so that a single effective binding curve can be generated by combining all the residues, as has been done in this work, both for the MtClpP2 and Bz-LL titrations. The analysis of peak intensities described here, valid when the exchange rate between conformers is much less than the chemical shift differences between the probes of each interconverting state (slow exchange), is to be distinguished from that associated with the case of fast exchange, where the titration of peak positions as a function of added ligand is quantified to obtain binding affinities. Accurate binding constants can be obtained by both methods so long as the exchange regime is

truly in the slow or fast limit; if this is not the case it is necessary to take into account the exchange kinetics (9). In this regard, we have attempted to measure exchange rates between *R* and *T* states at the midpoint of the Bz-LL titration (Fig. 4) using magnetization exchange based experiments, but exchange cross-peaks were not observed, indicating that in this case the exchange is indeed in the slow regime, and normalized intensities provide an accurate measure of fractional populations of each state.

In the analysis of the MtClpP2 titration data 18 data points (9 points from each of the two profiles derived from intensities of the free and the bound peaks) were fit to the above equations to extract K_d using a script (available upon request) written in Python 3.8 that used the lmfit (v. 0.9.14) package (10). The error in K_d was estimated using a Monte Carlo approach (11) whereby random errors in both protein concentrations (*x* axis of Fig. 1E and F) and relative populations of free and bound states (*y* axis of Fig. 1E and F) were added to the best-fit model to produce 1000 synthetic data sets that were subsequently fit as per the experimental data. Errors in protein concentration were estimated from 5 repeat measurements; errors in the relative populations were calculated as the root-mean-squared deviation between experimental points and those generated with the best-fit model. The values derived from Monte Carlo repeats were fit to a normal distribution function to yield expectation values and standard deviations σ. Final errors are reported as 2σ in the extracted values (95% confidence interval).

An extended MWC model that includes competitive binding of Bz-LL and substrate We begin by initially considering a binding model for a simplified system comprising a pair of rings, as for MtClpP1P2, but where each ring in turn contains of only a single binding site for activator (denoted as *X*) or substrate (*Y*). As described in the text, we assume that binding of *X* and *Y* is competitive (*i.e.*, both bind to the same site) and that each ring is in the same state, either *R* or *T*.

We distinguish rings 1 and 2 by subscripts 1, 2, so that R_1R_2 denotes a complex where rings 1 and 2 are both in the *R* state, for example. A number of equilibria immediately follow,

$$
\begin{aligned}
&R_1R_2 \rightleftarrows T_1T_2 && L = \frac{[T_1T_2]}{[R_1R_2]} \\
&R_1R_2 + X \rightleftarrows R_1XR_2 && K_{R1X} \\
&R_1R_2 + X \rightleftarrows R_1R_2X && K_{R2X} \\
&R_1R_2 + 2X \rightleftarrows R_1XR_2X && K_{R1X}K_{R2X} \\
&R_1R_2 + Y \rightleftarrows R_1YR_2 && K_{R1Y} \\
&R_1R_2 + Y \rightleftarrows R_1R_2Y && K_{R2Y} \\
&R_1R_2 + 2Y \rightleftarrows R_1YR_2Y && K_{R1Y}K_{R2Y} \\
&R_1R_2 + X + Y \rightleftarrows R_1XR_2Y && K_{R1X}K_{R2Y} \\
&R_1R_2 + X + Y \rightleftarrows R_1YR_2X && K_{R1Y}K_{R2X}
\end{aligned}
\qquad \text{[S1]}
$$

A similar set of equations exists for binding of *X* and *Y* to T_1T_2 that are obtained by replacing each *R* with *T* in Eq. [S1]. We now define a binding polynomial *Q* that is equal to the sum of the concentrations of all protein components,

$$
\begin{aligned}
Q = {} & [R_1R_2] + [R_1XR_2] + [R_1R_2X] + [R_1XR_2X] + [R_1YR_2] + [R_1R_2Y] + [R_1YR_2Y] + [R_1XR_2Y] + [R_1YR_2X] \\
& + [T_1T_2] + [T_1XT_2] + [T_1T_2X] + [T_1XT_2X] + [T_1YT_2] + [T_1T_2Y] + [T_1YT_2Y] + [T_1XT_2Y] + [T_1YT_2X] \\
= {} & [R_1R_2]\{1 + K_{R1X}[X] + K_{R2X}[X] + K_{R1X}K_{R2X}[X]^2 + K_{R1Y}[Y] + K_{R2Y}[Y] + K_{R1Y}K_{R2Y}[Y]^2 + \\
& \quad + K_{R1X}K_{R2Y}[X][Y] + K_{R1Y}K_{R2X}[X][Y]\} \qquad [S2] \\
& + [T_1T_2]\{1 + K_{T1X}[X] + K_{T2X}[X] + K_{T1X}K_{T2X}[X]^2 + K_{T1Y}[Y] + K_{T2Y}[Y] + K_{T1Y}K_{T2Y}[Y]^2 + \\
& \quad + K_{T1X}K_{T2Y}[X][Y] + K_{T1Y}K_{T2X}[X][Y]\}
\end{aligned}
$$

where R_1XR_2 is a complex with ligand *X* (Bz-LL) bound in ring 1, R_1XR_2Y is a complex with ligands *X* and Y (substrate) bound to rings 1 and 2, respectively, and so forth. Eq [S2] can be simplified using the first entry of Eq. [S1] to give,

$$Q = [R_1R_2]\{(1+K_{R1X}[X]+K_{R1Y}[Y])(1+K_{R2X}[X]+K_{R2Y}[Y]) + L(1+K_{T1X}[X]+K_{T1Y}[Y])(1+K_{T2X}[X]+K_{T2Y}[Y])\} \quad [S3]$$

The fraction of conformers where both rings are in the R state, f_R, is given by the sum of the first 9 terms of Eq. [S2], $([R_1R_2]+[R_1XR_2]+...+[R_1YR_2X])$ divided by the total concentration of protein, Q,

$$\frac{(1+K_{R1X}[X]+K_{R1Y}[Y])(1+K_{R2X}[X]+K_{R2Y}[Y])}{Q'}, \qquad Q' = \frac{Q}{[R_1R_2]} \quad [S4]$$

and similarly the fraction of conformers where both rings are in the T configuration, f_T, is

$$\frac{L(1+K_{T1X}[X]+K_{T1Y}[Y])(1+K_{T2X}[X]+K_{T2Y}[Y])}{Q'} \quad [S5]$$

Assuming that the activity of the complex (A) is proportional to the average number of substrate molecules bound to R_1, as discussed in the text, it follows that

$$A = V_o\frac{1\cdot[R_1YR_2]+1\cdot[R_1YR_2X]+1\cdot[R_1YR_2Y]}{Q} = \frac{V_oK_{R1Y}[Y](1+K_{R2X}[X]+K_{R2Y}[Y])}{Q'} \quad [S6]$$

where the constant of proportionality V_o is the activity measured under saturating amounts of Y (limit when $K_{R1Y}[Y]>>1$), when $[X]=0$ and assuming all molecules are in the active R state. The power of the binding polynomial approach (12) that has been adopted here is that expressions like Eq [S6] can be readily derived directly from Q (or Q') Consider the case where Q' is given as

$$Q' = Q'_R + LQ'_T = Q'_{R1}Q'_{R2} + LQ'_{T1}Q'_{T2} \quad [S7]$$

where Q'_{A1} and Q'_{A2} contain terms related only to the binding of *X* or *Y* to rings 1 and 2, respectively, and $A \in \{R, T\}$, corresponding to the *R* or *T* state of a ring. In the case of the example here

$$\begin{aligned} Q'_{R1} &= 1 + K_{R1X}[X] + K_{R1Y}[Y] \\ Q'_{R2} &= 1 + K_{R2X}[X] + K_{R2Y}[Y] \\ Q'_{T1} &= 1 + K_{T1X}[X] + K_{T1Y}[Y] \\ Q'_{T2} &= 1 + K_{T2X}[X] + K_{T2Y}[Y] \end{aligned} \qquad \text{[S8]}$$

and the average number of substrate molecules bound to the *R* state of a complex, R_1R_2, (both rings are in the *R* state) is given by (12)

$$\begin{aligned} \frac{[Y]}{Q'}\frac{\partial Q'_R}{\partial Y} &= \frac{[Y]}{Q'}\{(1 + K_{R2X}[X] + K_{R2Y}[Y])K_{R1Y} + (1 + K_{R1X}[X] + K_{R1Y}[Y])K_{R2Y}\} \\ &= \frac{[Y]}{Q'}\{Q'_{R2}K_{R1Y} + Q'_{R1}K_{R2Y}\} \end{aligned} \qquad \text{[S9]}$$

while the average number of substrate molecules bound to R_1 can be derived from the relation $\frac{[Y]}{Q'}Q'_{R2}\frac{\partial Q'_{R1}}{\partial [Y]}$ to give the first term of Eq. [S9] which is the expression in Eq. [S6], neglecting V_o.

The results of this section can be easily generalized to the case where each ring is made up of *n* protomers (*n*=7 for MtClpP1P2), rather than 1 considered to this point. This is achieved simply by noting that the binding of *X* and *Y* to each of the *n* protomers is independent so that the terms Q'_{A1} and Q'_{A2} in the expressions above are to be replaced with $(Q'_{A1})^7$ and $(Q'_{A2})^7$, respectively, as described in reference (12).

In this case,

$$\begin{aligned} Q' = \{&(1 + K_{R1X}[X] + K_{R1Y}[Y])^n(1 + K_{R2X}[X] + K_{R2Y}[Y])^n \\ &+ L(1 + K_{T1X}[X] + K_{T1Y}[Y])^n(1 + K_{T2X}[X] + K_{T2Y}[Y])^n\} \end{aligned} \qquad \text{[S10]}$$

and f_R and f_T are given by

$$f_R = \frac{(1+K_{R1X}[X]+K_{R1Y}[Y])^n(1+K_{R2X}[X]+K_{R2Y}[Y])^n}{Q'}, \quad [S11]$$

and

$$f_T = \frac{L(1+K_{T1X}[X]+K_{T1Y}[Y])^n(1+K_{T2X}[X]+K_{T2Y}[Y])^n}{Q'}, \quad [S12]$$

respectively. Similarly, A can be expressed as

$$A = \frac{nV_oK_{R1Y}[Y](1+K_{R1X}[X]+K_{R1Y}[Y])^{n-1}(1+K_{R2X}[X]+K_{R2Y}[Y])^n}{Q'} \quad [S13]$$

where V_o is the maximal activity per subunit. As described in the text, we were not able to fit our functional assays using Eq. [S13]. However, the data were well fit when V_o was allowed to vary with [Y], that is $V_o = V_o([Y])$, which takes into account the decrease in A at high concentrations of substrate (substrate inhibition). Thus, in the fits illustrated in Fig. 5C, $V_o([Y])$ is floated for each [Y] (each panel). The plot of Fig. 5E shows $V_o[Y]$ *vs* [Y] (circles) and a fit to the data using the function $V_o([Y]) = \frac{V_{\max}}{1+K_I^N[Y]^N}$ (solid line), where V_{max} is the maximum possible enzyme activity per subunit, assuming therefore that (i) all molecules are in the R state, (ii) [X]=0, and (iii) the absence of substrate inhibition ($K_I = 0$), K_I is the association constant for the binding of each of N molecules of Y to inhibitory sites on MtClpP1P2 and N is the Hill coefficient for this process,

$$ClpP1P2 + NY \rightleftarrows ClpP1P2 - Y_N \qquad K_I^N. \quad [S14]$$

Eqs. [S11], [S12] and [S13] (where $V_o = V_o([Y])$ have been used to obtain the fits shown in Fig. 4B and C of the main text, with a scaling factor included to weight the NMR data relative to the data from the functional assays to account for the larger amount of activity data and activity

values that can peak at approximately 30. Thus, $\chi^2_{red,Total} = f\chi^2_{red,NMR} + \chi^2_{red,Assays}$, where f was set to 14 (purely empirically). Indeed, we have tested f values from 0.1-25 and these have very little effect on the extracted parameters with $-RT\ln(z)$ (simplest model, see text) ranging from -4.5 kcal/mol – -4.6 kcal/mol; -3.4 kcal/mol – -3.6 kcal/mol; -7.2 kcal/mol – -7.6 kcal/mol for z = K_{R1X}, K_{R1Y}, and L.

The average number of activator (Bz-LL = X) molecules that are bound to MtClpP1P2 can be calculated as follows,

$$\begin{aligned}\frac{[X]}{Q'}\frac{\partial Q'}{\partial X} = \frac{n[X]}{Q'}\{&(1+K_{R1X}[X]+K_{R1Y}[Y])^n(1+K_{R2X}[X]+K_{R2Y}[Y])^{n-1}K_{R2X}\\ &+(1+K_{R1X}[X]+K_{R1Y}[Y])^{n-1}K_{R1X}(1+K_{R2X}[X]+K_{R2Y}[Y])^n\\ &+L[(1+K_{T1X}[X]+K_{T1Y}[Y])^n(1+K_{T2X}[X]+K_{T2Y}[Y])^{n-1}K_{T2X}\\ &+(1+K_{T1X}[X]+K_{T1Y}[Y])^{n-1}K_{T1X}(1+K_{T2X}[X]+K_{T2Y}[Y])^n]\}\end{aligned} \quad [S15]$$

where Q' is given by Eq [S10]. In a similar manner, the average number of activator (Bz-LL = X) molecules that are bound to the R configuration of the complex (i.e., both rings are in the R state) is calculated as,

$$\begin{aligned}\frac{[X]}{Q'}\frac{\partial Q'_R}{\partial X} = \frac{n[X]}{Q'}\{&(1+K_{R1X}[X]+K_{R1Y}[Y])^n(1+K_{R2X}[X]+K_{R2Y}[Y])^{n-1}K_{R2X}\\ &+(1+K_{R1X}[X]+K_{R1Y}[Y])^{n-1}K_{R1X}(1+K_{R2X}[X]+K_{R2Y}[Y])^n\}\end{aligned} \quad . \quad [S16]$$

Finally, the average number of activator molecules that are bound to the MtClpP1 ring in the R configuration of the complex (green solid curves in Fig. 4B and C) is

$$\frac{[X]}{Q'}\frac{Q'_{R2}\partial Q'_{R1}}{\partial X} = \frac{n[X]}{Q'}\{(1+K_{R1X}[X]+K_{R1Y}[Y])^{n-1}K_{R1X}(1+K_{R2X}[X]+K_{R2Y}[Y])^n\}$$

[S17]

A description of the fits, including the number of fitting parameters, is provided in subsequent sections.

Activity Assays

Peptidase rate measurements The peptidase activity of MtClpP1P2 was measured at 40 ˚C with Acetyl-L-Pro-L-Lys-L-Met bearing a C-terminal fluorogenic 7-amino-4-methylcoumarin group (abbreviated PKM-AMC throughout the manuscript) as substrate. The reaction was followed with a Synergy Neo2 96-well microplate reader by taking a measurement every 21 seconds for 60 minutes at λ_{ex}: 355 nm, λ_{em}: 460 nm. For all activity response measurements the concentration (protomer) of MtClpP1 was 1 μM while the concentration of MtClpP2 (protomer) was 20 μM, ensuring that the fraction of MtClpP1 in complex with MtClpP2 does not change substantially as [Bz-LL] is varied (note the difference in K_d values when Bz-LL is added, Fig. 1E and F). To ensure maximum similarity with the NMR titrations and facilitate subsequent data fitting, all functional assays, except for those shown in Fig. 5, were performed using deuterated enzyme in 100% D_2O-based NMR buffer. Activities are derived from initial rates extracted and analyzed using a Python script written in-house (see below). Error bars correspond to one standard deviation derived from three repeat measurements.

Combined analysis of NMR intensities and activity assays as a function of Bz-LL concentration (Fig. 4B-F) Peak intensities in ^{1}H-^{13}C HMQC spectra of MtClpP1P2, with ILVM-labelled MtClpP1 or MtClpP2, were quantified as a function of 11 Bz-LL concentrations (Fig. 4B) and activity profiles of MtClpP1P2 measured from initial rates of fluorescence change due to hydrolysis of the substrate PKM-AMC quantified with 11 Bz-LL concentrations and 8 substrate concentrations. The NMR and activity data, a total of 132 data points, were jointly fit to a modified MWC model (Fig. 4D, described in SI) using Eqs. [S11]-[S13] with $V_o([Y])$ floated for each concentration of *Y*. This was accomplished using a protocol in which initial guesses for

parameters were derived by a grid search that explored values for the association constants between 10^{-4} M^{-1} and 10^{4} M^{-1}, between 10^{-2} A.U. and 10^{6} A.U. for $V_o[(Y)]$ (A.U. = μM PKM min^{-1}/μM MtClpP1P2) and between 10^{-2} and 10^{6} for L. Each dimension was searched in 8 steps that were linear on the logarithmic scale. Once starting parameters were identified by this procedure a least squares fit of the data was performed using an in-house package written in Python 3.8 using a Levenberg-Marquardt search engine that is available within the lmfit (v. 0.9.14) package (10). The 8 extracted values of $V_o([Y])$ were subsequently fit to a Hill-model of substrate inhibition described above to obtain the association constant for substrate binding to the inhibitory sites in the complex. For the simplified model displayed in Fig. 4D (red box), a total of 11 parameters were fit, including K_{R1X}, K_{R1Y}, L, and eight $V_o([Y])$ values. When the data were fit to the full model (Fig. D), illustrated in Fig. S8, a total of 17 parameters were used, including all eight association constants, L, and eight $V_o([Y])$ values. Errors in the fitted parameters were estimated using a Monte Carlo approach (11) whereby random errors, calculated as the root-mean-squared deviation between experimental points and those generated with the best-fit model, were added to the best-fit model to produce 1500 synthetic data sets which were fit as per the experimental data. Initial guesses for the Monte Carlo runs were generated randomly. The values derived from Monte Carlo repeats were converted to a histogram, which was subsequently fit to a normal distribution function to yield expectation values and standard deviations σ. Final errors are reported as 2σ in the extracted values (95% confidence interval). All scripts are available upon request.

Cryo-EM

Sample preparation for cryo-EM All purified proteins were concentrated to ~20-30 mg/mL in buffer. Immediately before freezing, samples were mixed with 0.025 % (wt/vol) IGEPAL CA-

630 (Sigma-Aldrich) to increase the proportion of protein complexes adopting side views on the grid. 2.5 μL of the sample mixtures were applied to nanofabricated holey gold grids (13–15) with a hole size of ~1-2 μm. Grids were blotted on both sides using a FEI Vitrobot mark III for 15 seconds at 4 °C and ~100% relative humidity before freezing in a liquid ethane/propane mixture (16).

Electron microscopy All MtClpP1P2 complexes were imaged with a Thermo Fisher Scientific Titan Krios G3 microscope operating at 300 kV and equipped with a FEI Falcon III DDD camera. Structures were calculated from counting mode movies consisting of 30 frames, obtained over a 60 second exposure with defocuses ranging from 0.7 to 2.0 μm. Movies were at a nominal magnification of 75000×, corresponding to a calibrated pixel size of 1.06 Å and with an exposure of 0.8 electrons/pixel/s, giving a total exposure of 43 electrons/Å^2. For apo, ADEP-bound, and GLF-CMK modified MtClpP1P1 2092, 725, and 1645 movies were collected respectively, using the microscope's *EPU* software.

EM image analysis Whole frame alignment was performed in *cryoSPARC* v2 (17) with the resulting averages of frames used for contrast transfer function (CTF) determination (18). Templates for particle selection were generated by 2D classification of manually selected particles. Particle images were extracted in 184×184-pixel boxes, and individual particle alignment and exposure weighting was performed within *cryoSPARC* v2 (17).

Atomic model building and refinement To model MtClpP1P2 (S10), a single subunit of each of MtClpP1 and MtClpP2 of the Bz-LL bound crystal structure (PDBID: 5DZK) (19) was fit into the EM density map as rigid bodies using *UCSF Chimera* (20). For apo MtClpP1P2, *Rosetta* (21) was used to minimize the structure with C7 symmetry enforced, with iterative backbone rebuilding. The best scoring models were visually inspected, and the best fitting model was used

for further analysis. For ADEP bound MtClpP1P2 a single pair of protomers, one from MtClpP1 and one from MtClpP2 (apo structure) were used as a starting model for further refinement. The N-terminal domains were built in *Coot* (22), and the entire model relaxed with *Rosetta* enforcing C7 symmetry. The top scoring models were then used for further analysis. ADEP was modelled based on PDBID 6CFD, in *Coot,* and real space refined in *PHENIX* using ligand restraints built in *PHENIX* elbow (23). For GLF-CMK bound MtClpP1P2, MtClpP1 and MtClpP2 protomers from the Bz-LL bound crystal structure (PDBID: 5DZK) were rigidly docked into the density in *UCSF Chimera*. Restraints for the GLF-CMK ligand were generated in *PHENIX* elbow followed by real space refinement in *PHENIX*. Validation reports (Table 1) were prepared in *PHENIX.*

Models were evaluated with *Molprobity* (24) and *EMRinger* (25). Figures were generated in *UCSF Chimera* (20) and *UCSF ChimeraX* (26), and colors chosen with *ColorBrewer* (27).

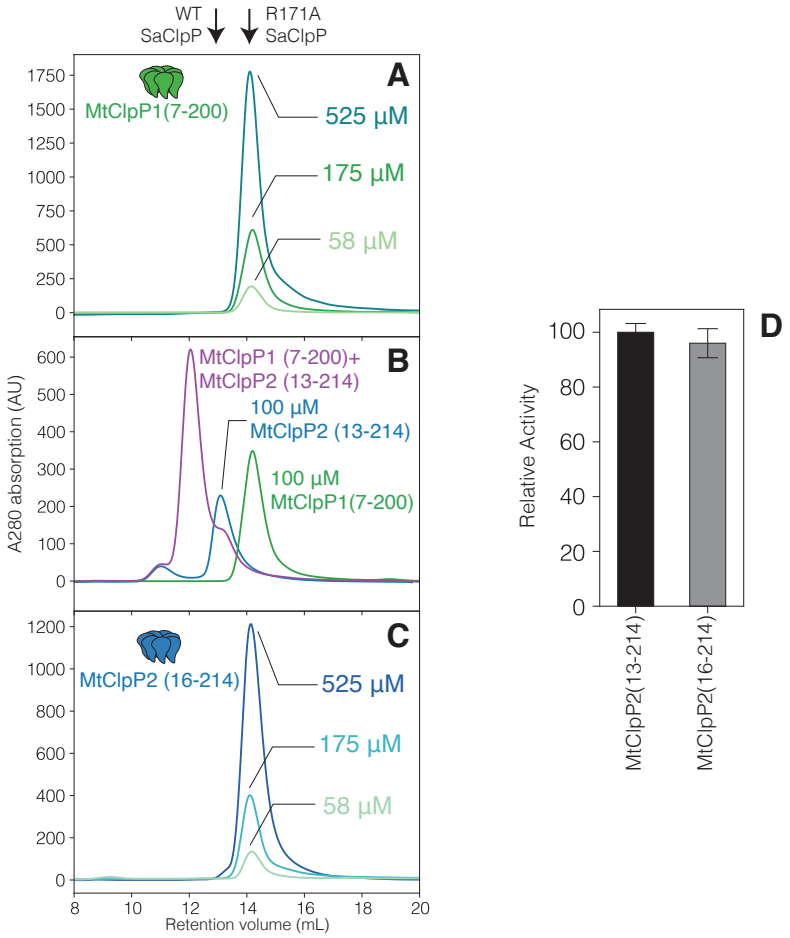


Figure S1. Effect of protein concentration on oligomeric state of (A) MtClpP1 (residues 7-200) as established by SEC. (B) SEC profile of MtClpP2 with native propeptide processing (residues 13-214) measured in isolation (blue trace). Mixing with equimolar concentration of MtClpP1 (residues 7-200) (green trace) leads to the formation of MtClpP1P2 complexes (purple trace). (C) As in (A) but for MtClpP2 (residues 16-214), with concentrations as indicated. In all panels 0.5 mL of protein at the denoted concentration (monomer) was injected. (D) Peptidase activity assays on a pair of mixtures containing MtClpP2 (residues 13-214) (black bar) or MtClpP2 (residues 16-214) (grey bar) performed in the presence of MtClpP1 (residues 7-200), 4 mM Bz-LL, with 250 µM Suc-LY-AMC used as substrate. In both cases the (monomer) concentration of each of MtClpP1 and MtClpP2 is 1 µM. Error bars correspond to one standard deviation based on three measurements.

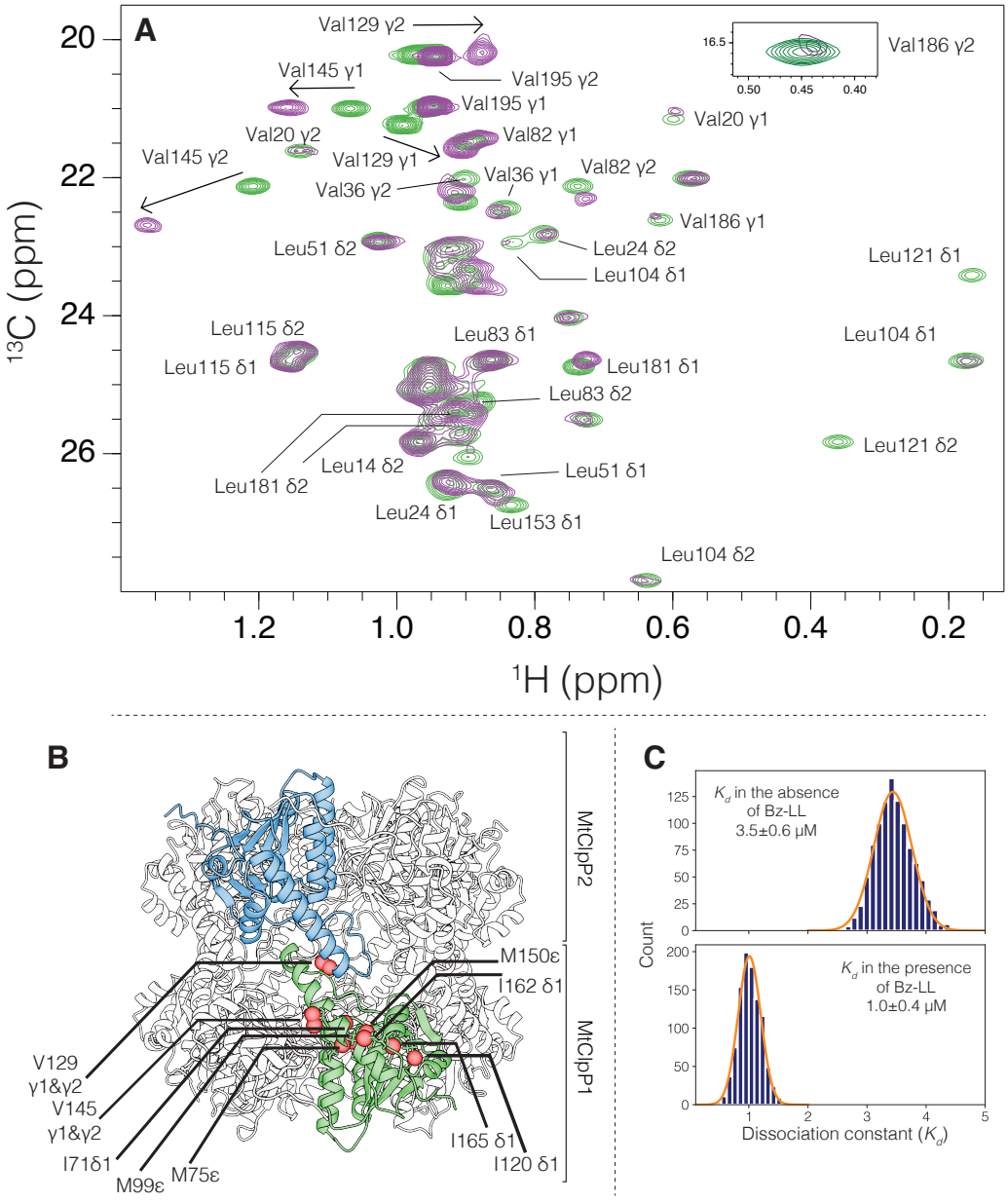


Figure S2. (A) Overlay of the Leu and Val regions of ^{1}H-^{13}C HMQC correlation maps of ILVM-labeled MtClpP1 (green contour) and a mixture containing 50 μM ILVM-labeled MtClpP1 and 100 μM [U-^{2}H] MtClpP2 (purple contours), recorded at 40 °C, 18.8 T, with stereospecific assignments as indicated. (B) Methyl groups in MtClpP1 (red circles; MtClpP1 is ILVM-labeled) showing chemical shift changes as a result of the addition of [U-^{2}H] MtClpP2 are mapped onto the structure of apo MtClpP1P2. (C) The dissociation constants for the binding of MtClpP1 and

MtClpP2, measured at 40 °C in the absence (top panel) and presence (bottom panel) of Bz-LL are listed, along with the distribution of values based on a Monte Carlo analysis (11) that included errors in both peak intensities and protein concentrations. The distributions are fitted to a Gaussian function (orange curve) and the errors reported as twice the standard deviation in K_d. These distributions show that the difference in measured affinities is outside the experimental error.

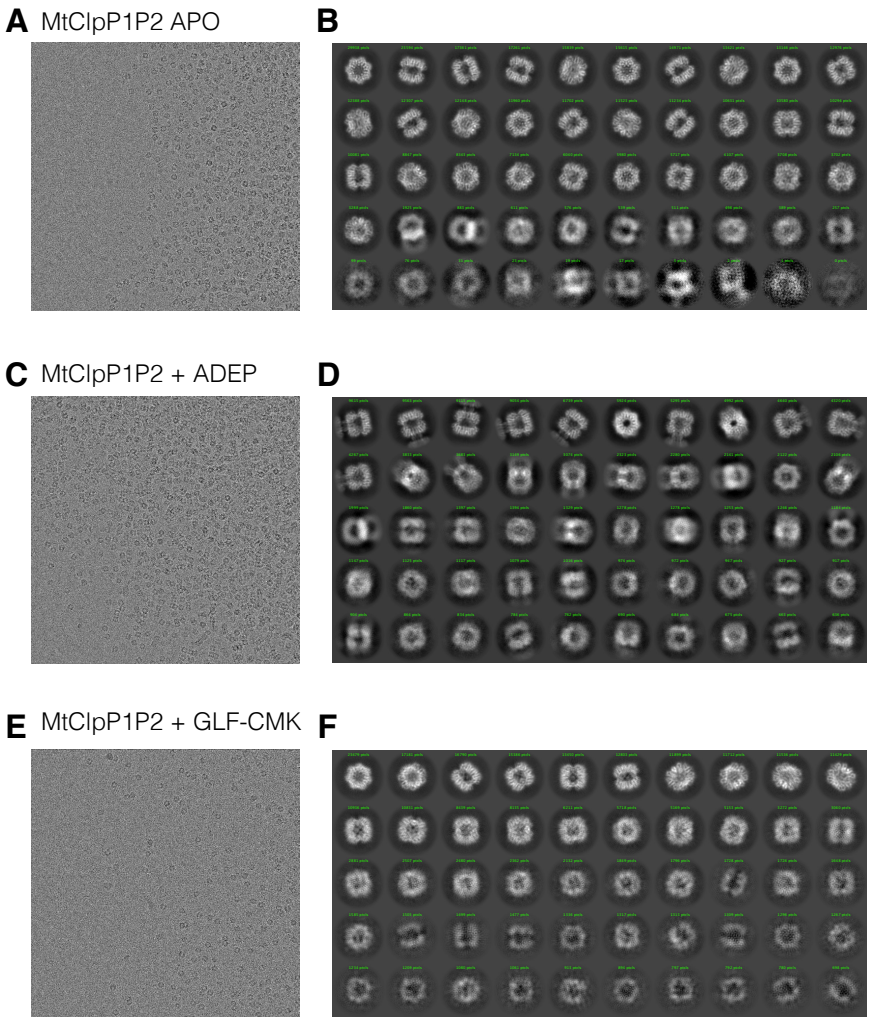


Figure S3. Representative electron micrographs and 2D classes of the MtClpP1P2 complexes in the (A and B) apo, (C and D) ADEP-bound, and (E and F) GLF-CMK bound forms of the enzyme. The number of particles used in each class average is indicted.

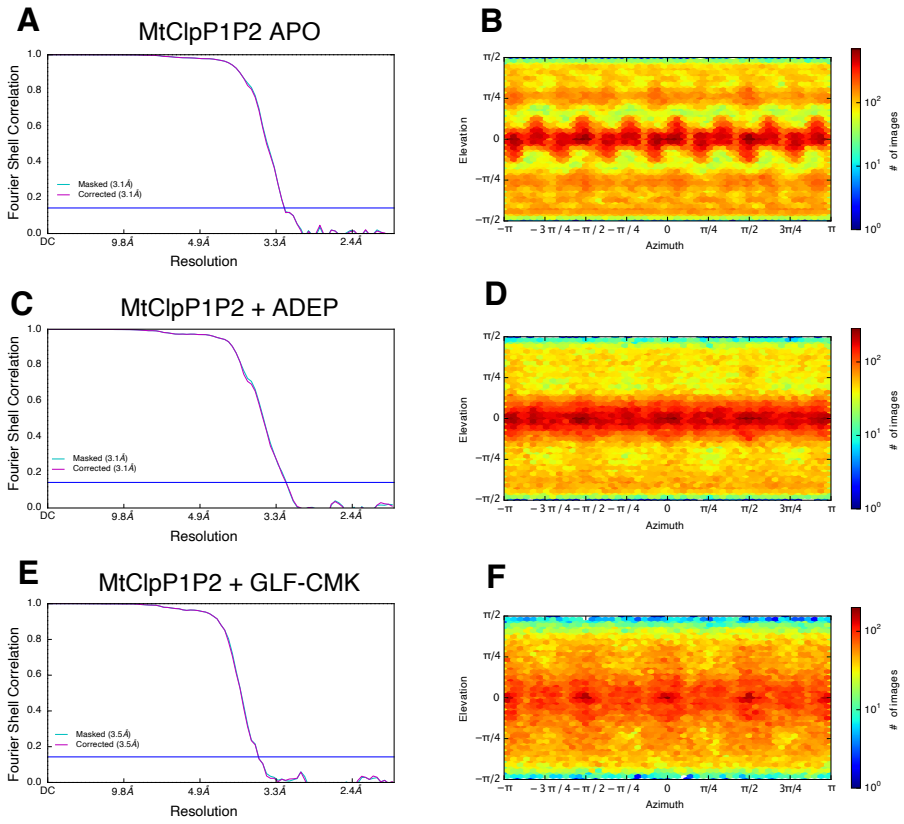


Figure S4. Fourier Shell Correlation (FSC) as a function of spatial resolution and orientation plots for MtClpP1P2 complexes in the (A and B) Apo, (C and D) ADEP-bound, and (E and F) GLF-CMK bound forms. Resolution values reported are for FSC=0.143.

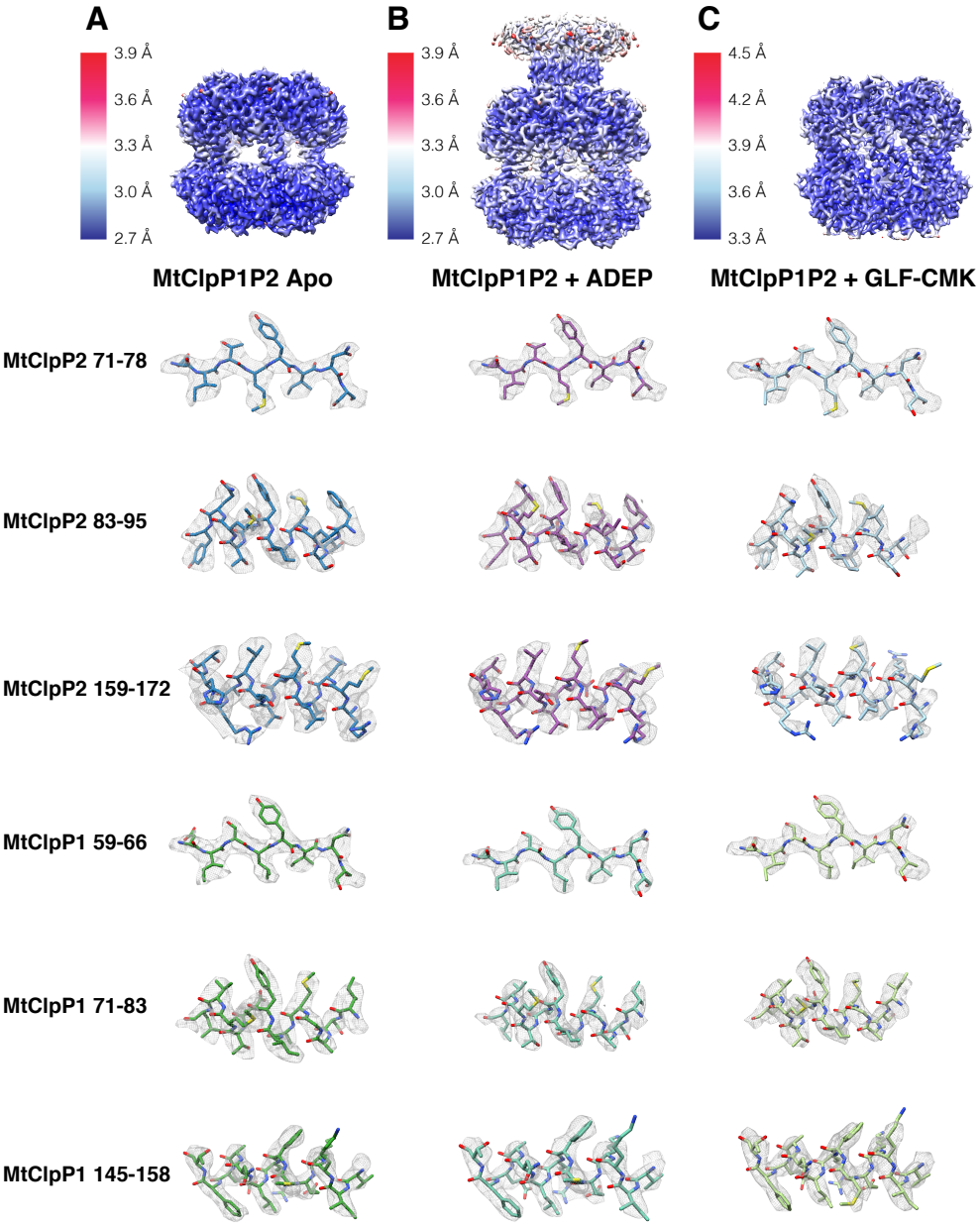


Figure S5. Local resolution maps of MtClpP1P2 complexes in the (A) Apo, (B) ADEP-bound, and (C) GLF-CMK bound forms. Note that in the ADEP-bound state particles are 'dimers' of tetradecamers linked via the gates of the P2 ring. Example regions of models built into the experimental cryo-EM maps are also shown (D). The additional density above the MtClpP2 N-terminal gates in (B) corresponds to a second MtClpP2 ring that results from dimerization of the complex upon addition of ADEP (see Fig. S10).

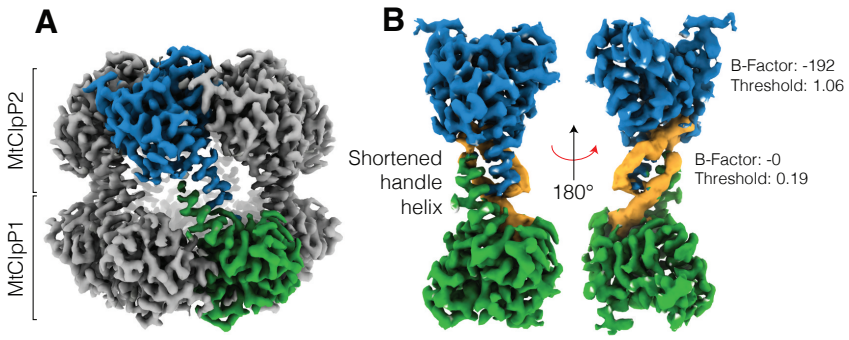


Figure S6. (A) Density map for the apo MtClpP1P2 complex. Single MtClpP1 and MtClpP2 protomers in the complex are coloured green and blue, respectively. (B) Density for a single protomer pair is shown along with an unsharpened map for the flexible handle region.

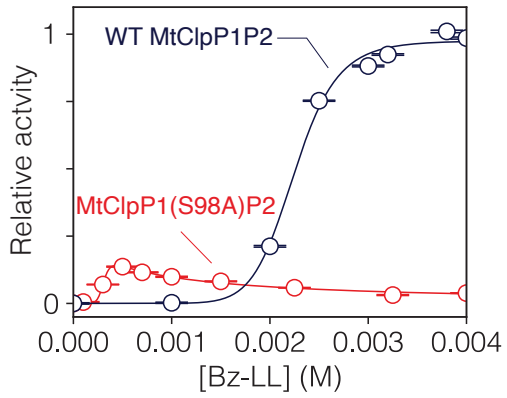


Figure S7. Activity response curves measured as a function of Bz-LL concentration using 250 µM PKM-AMC as substrate for WT MtClpP1P2 (black curve) and MtClpP1(S98A)P2 (red curve). The shift in the shape of the activity curve to lower Bz-LL concentrations for MtClpP1(S98A)P2 is due to the S98A mutation in MtClpP1 (see Discussion).

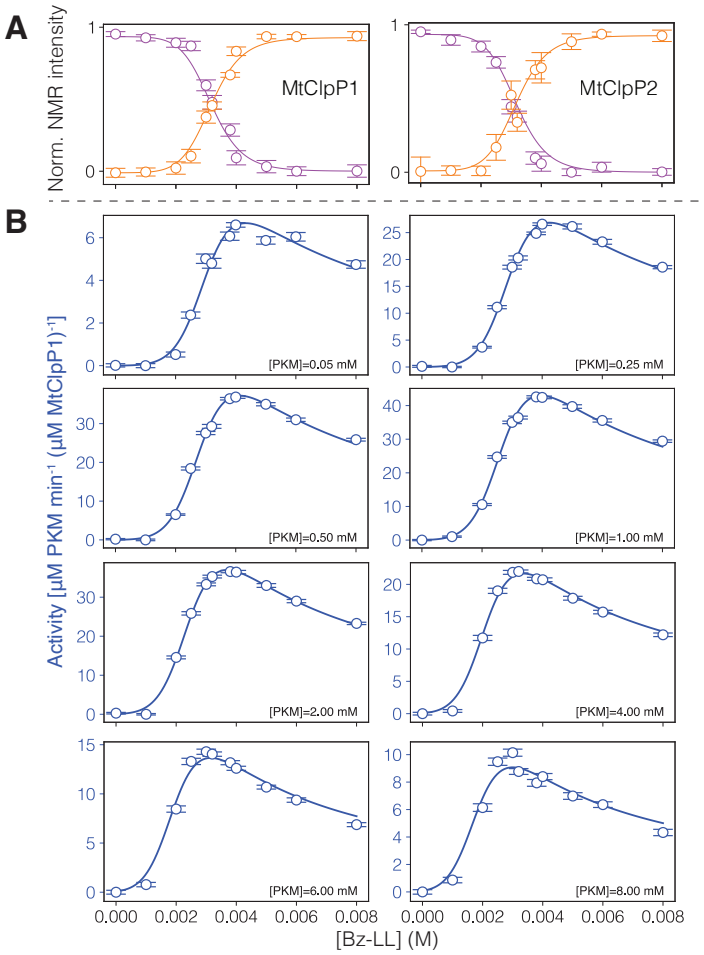


Figure S8. Combined fits of (A) NMR peak intensities and (B) MtClpP1P2 activity as a function of the concentration of Bz-LL. The modified MWC model illustrated in Fig. 4D has been used which includes binding of Bz-LL and substrate to the *T* and *R* states of MtClpP1 and MtClpP2 of the complex.

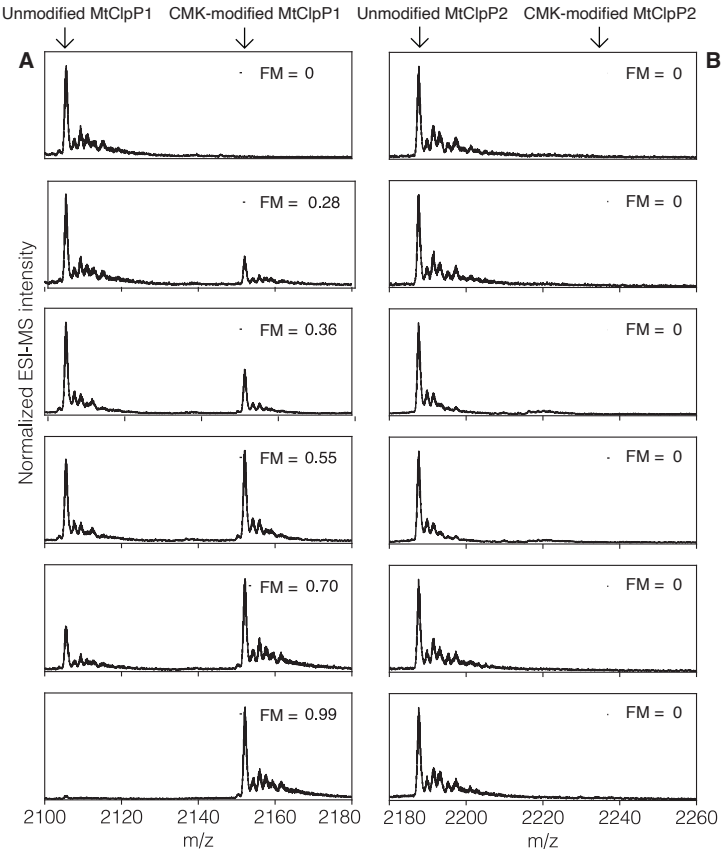
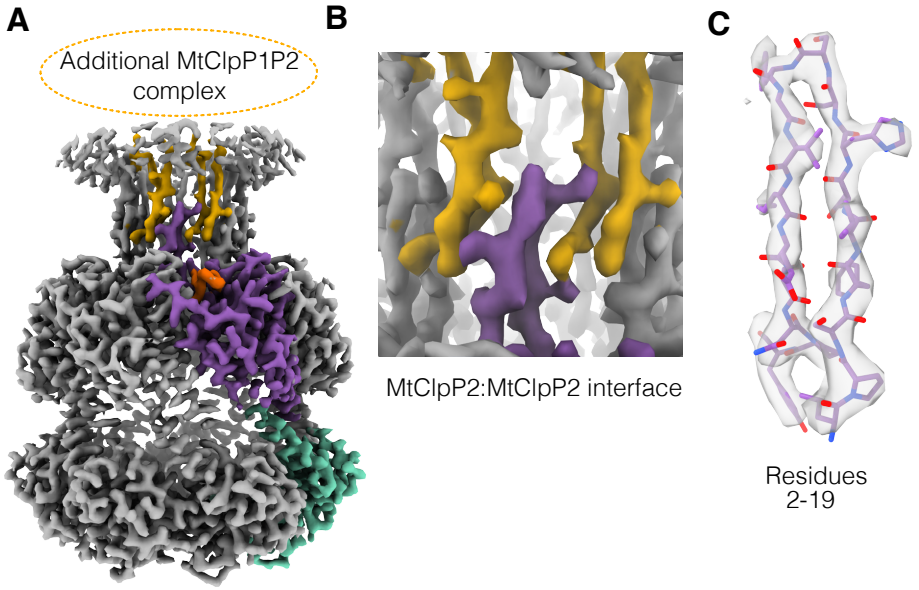


Figure S9. Intact protein mass spectra of each ring of MtClpP1P2 partially modified by GLF-CMK. The 10+ charge state for each of (A) MtClpP1 and (B) MtClpP2 along with expected positions of the unmodified and GLF-CMK modified protein peaks are indicated with arrows. The fraction of active sites modified (FM) is indicated for each panel.



D

Ligand	Ligand binding site	MtClpP2 gate conformation
Apo	N/A	Disordered
ADEP	P2 Head domain	Beta-Hairpin
GLF CMK	MtClpP1 active site	Disordered
BZ-LL	MtClpP1 and MtClpP2 active site	Semi-ordered (bottom half of hairpin)

Figure S10. ADEP bound MtClpP2 gates and density. (A) Density map for the MtClpP1P2 complex bound to ADEP. Additional density was present above the MtClpP2 N-terminal gates corresponding to a second MtClpP2 ring that results from dimerization of the complex upon addition of ADEP. Single MtClpP1 and MtClpP2 protomers are coloured teal and purple, respectively. Gate density for two protomers of the second P2 ring are coloured yellow. (B) Magnified view of the MtClpP2:MtClpP2 interface mediated by the N-terminal gates. (C) Model in map fit for a single N-terminal gate. (D) Table summarizing gate conformations in all the states observed for MtClpP1P2.

Table 1. Cryo-EM data acquisition, processing, atomic model statistics, and map/model depositions.

A. Cryo-EM data acquisition and image processing.

Data Collection			
Electron Microscope	Titan Krios		
Camera	Falcon 3EC		
Voltage (kV)	300		
Nominal Magnification	75,000		
Calibrated physical pixel size (Å)	1.06		
Total exposure (e/Å^2)	42.7		
Exposure rate (e/pixel/s)	0.8		
Number of frames	30		
Defocus range (μm)	0.7 to 2.0		
Image Processing			
	MtClpP1P2 APO	MtClpP1P2 + ADEP	MtClpP1P2 + GLF-CMK
Motion correction software	*cryoSPARC v2*	*cryoSPARC v2*	*cryoSPARC v2*
CTF estimation software	*cryoSPARC v2*	*cryoSPARC v2*	*cryoSPARC v2*
Particle selection software	*cryoSPARC v2*	*cryoSPARC v2*	*cryoSPARC v2*
Micrographs used	2,092	725	1645
Particle images selected	612,408	366,129	257,060
3D map classification and refinement software	*cryoSPARC v2*	*cryoSPARC v2*	*cryoSPARC v2*

B. Map and model statistics.

EM maps	MtClpP1P2 APO	MtClpP1P2 + ADEP	MtClpP1P2 + GLF-CMK
Particle images contributing to maps	373,064	192,430	143,748
Applied symmetry	C7	C7	C7
Applied B-factor (Å^2)	-191.8	-168.6	-217.4
Global resolution (FSC = 0.143, Å)	3.1	3.1	3.5
Model Building			
Modeling software	Coot, Phenix, Rosetta		
Number of residues	2,513	2,618	2,506
RMS bond length (Å)	0.0196	0.0202	0.0049
RMS bond angle (°)	1.84	1.70	1.08
Ramachandaran outliers (%)	0.85	0.54	0
Ramachandran favoured (%)	95.77	96.22	98.31
Rotamer outliers	0	0	0
C-beta deviations	0	0	0
Clashscore	0.26	2.66	1.16
MolProbity score	0.89	1.31	0.83
EMRinger score	4.3	3.7	2.41
Ligand	N/A	ADEP-7	GLF-CMK

C. Residues excluded in atomic models.

ClpX Protomer	MtClpP1P2 APO	MtClpP1P2 + ADEP	MtClpP1P2 + GLF-CMK
ClpP1	1-15,193-200	1-15,193-200	1-15,193-200
ClpP2	1-30, 211-214	1-15, 211-214	1-31, 211-214

D. Deposited maps and associated coordinate files.

Maps	EMDB code	Associated PDB ID
MtClpP1P2 APO	EMD-21197	6VGK
MtClpP1P2 + ADEP	EMD-21198	6VGN
MtClpP1P2 + GLF-CMK	EMD-21199	6VGQ

Reference:

1. Vahidi S, et al. (2018) Reversible inhibition of the ClpP protease via an N-terminal conformational switch. *Proc Natl Acad Sci USA* 115(28):E6447–E6456.
2. Huang R, Perez F, Kay LE (2017) Probing the cooperativity of Thermoplasma acidophilum proteasome core particle gating by NMR spectroscopy. *Proc Natl Acad Sci USA* 114(46):E9846–E9854.
3. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-Correlated Relaxation Enhanced 1H−13C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *J Am Chem Soc* 125(34):10420–10428.
4. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
5. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59(4):687–696.
6. Tugarinov V, Kay LE (2004) An Isotope Labeling Strategy for Methyl TROSY Spectroscopy. *J Biomol NMR* 28(2):165–172.
7. Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S

proteasome by NMR. *Nature* 445:618–622.

8. Gans P, et al. (2010) Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angew Chem Int Ed* 49(11):1958–1962.
9. Tugarinov V, Kay LE (2003) Quantitative NMR Studies of High Molecular Weight Proteins: Application to Domain Orientation and Ligand Binding in the 723 Residue Enzyme Malate Synthase G. *J Mol Biol* 327(5):1121–1133.
10. Newville M, Stensitzki T, Allen DB, Ingargiola A (2014) LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. doi:10.5281/zenodo.11813.
11. Frolkovič P (1990) Numerical recipes: The art of scientific computing. *Acta Appl Math* 19(3):297–299.
12. Dill K, Bromberg S (2003) *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology* (Garland Science, New York).
13. Marr CR, Benlekbir S, Rubinstein JL (2014) Fabrication of carbon films with ~500nm holes for cryo-EM with a direct detector device. *J Struct Biol* 185(1):42–47.
14. Russo CJ, Passmore LA (2014) Ultrastable gold subtrates for electron cryomicroscopy. *Science (80-)* 346(6215):1377–1380.
15. Meyerson JR, et al. (2014) Self-assembled monolayers improve protein distribution on holey carbon cryo-EM supports. *Sci Rep* 4:7084.
16. Tivol WF, Briegel A, Jensen GJ (2008) An improved cryogen for plunge freezing. *Microsc Microanal* 14(5):375–379.
17. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14(3):290-+.

18. Rohou A, Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 192(2):216–221.

19. Li M, et al. (2016) Structure and Functional Properties of the Active Form of the Proteolytic Complex, ClpP1P2, from Mycobacterium tuberculosis. *J Biol Chem* 291(14):7465–7476.

20. Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comp Chem* 25(13):1605–1612.

21. Wang RYR, et al. (2015) De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat Methods* 12(4):335–338.

22. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D* 60(12 Part 1):2126–2132.

23. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D* 66(2):213–221.

24. Chen VB, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D* 66(1):12–21.

25. Barad BA, et al. (2015) EMRinger: side chain–directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* 12:943.

26. Goddard TD, et al. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 27(1):14–25.

27. Harrower M, Brewer CA (2003) ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *Cartogr J* 40(1):27–37.