SI Appendix

Reversible inhibition of the ClpP protease via an N-terminal conformational switch

Siavash Vahidi[1,2,3,4+], Zev A. Ripstein[2,4+], Massimiliano Bonomi[5], Tairan Yuwen[1,2,3,4], Mark F. Mabanglo[2], Jordan B. Juravsky[4], Kamran Rizzolo[2], Algirdas Velyvis[1,2,3,4], Walid A. Houry[2,3], Michele Vendruscolo[5], John L. Rubinstein[2,4,6*], 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, The Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada
[5]Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, United Kingdom
[6]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, M5G 1L7, Canada

Correspondence to:

Lewis E. Kay (kay@pound.med.utoronto.ca)
John L. Rubinstein (john.rubinstein@utoronto.ca)

Short title: NMR and cryo-EM of ClpP

Keywords: ClpP, conformational dynamics, allostery, methyl-TROSY NMR, Cryo-EM

www.pnas.org/cgi/doi/10.1073/pnas.1805125115

Plasmids and constructs The ClpP gene from *S. aureus* (Uniprot entry: A6QF76), and *N. meningitidis* (Uniprot entry: Q9JZ38) were synthesized by GenScript (Piscataway, NJ, USA) and cloned into the NdeI and BamHI sites of pET24a+ (Novagen, Madison, WI, USA). The *E. coli* ClpP gene (Uniprot entry: P0A6G7 - excluding the 14 residue pro-peptide; available in-house) was cloned into a pET9a+ expression vector. Point mutations were introduced using Quikchange site-directed mutagenesis (Agilent, Santa Clara, CA, USA). Cleavable N-terminal His_6-SUMO, and non-cleavable C-terminal His_6 tags were added via Gibson assembly.

Protein expression and purification Transformed Codon+ *E. coli* BL21(DE3) 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. Production of [U-^{2}H; Ileδ1-$^{13}CH_3$; Leu,Val-$^{13}CH_3$/$^{12}CD_3$; Met-$^{13}CH_3$]-labeled protein samples was achieved by addition of the following precursors 1 hour prior to induction: 50 mg L^{-1} α-ketobutyrate, 80 mg L^{-1} α-ketoisovalerate, and 100 mg L^{-1} $^{13}CH_3$ Met. Samples with ProR $^{13}CH_3$-labeled Leu/Val residues [$^{13}CH_3$, Val/Leu-γ1/δ1(proR)] were prepared via the addition of 170 mg L^{-1} 2-hydroxy-2-methyl-D_3-3-oxobutanoate-4-^{13}C instead of α-ketoisovalerate. Expression was allowed to proceed for *ca.* 18 hr at 25 ˚C. For samples to be used in cryo-EM analysis and functional assays, cells were grown in Luria broth (LB) media at 37 ˚C and induced at OD_{600}=0.8 with 0.1 mM IPTG. Expression was allowed to proceed for 4 hours at 37 ˚C. Proteins were purified using Ni-affinity chromatography, followed by removal of the SUMO tag using Ulp1 protease and, subsequently, passed over a Ni column to remove the cleaved SUMO tag and other impurities. Finally, the flow-through from the reverse Ni purification was concentrated using an Amicon Ultra-15 50K MWCO (Millipore) concentrator and subjected to gel filtration on a HiLoad 16/60 Superdex 200 column (GE). Note that all ClpP samples with mutations in the N terminal hydrophobic cluster eluted as tetradecameric complexes, with very similar elution volumes. Constructs used for NMR resonance assignment contained a non-cleavable C-terminal His$_6$ tag and were purified similarly. These samples yielded HMQC spectra that were superimposable with those obtained from tag-free ClpP, and offered the advantage of faster purification. Protein concentrations were determined spectrophotometrically (GdnCl-denatured protein) using extinction coefficients obtained from ExPASy's ProtParam web-based tool (https://web.expasy.org/protparam/) (WT SaClpP $\mathcal{E}_{280}$=8940 M^{-1} cm^{-1}). Prior to NMR measurements the samples were buffer exchanged into Buffer A containing 100 mM KCl, 1 mM EDTA, 5mM MgCl$_2$, 50 mM imidazole at pH* 7.0 prepared in 99.9% D$_2$O.

Preparation of samples for PRE measurements A Q130C mutant of SaClpP was purified as described above with the addition of 5 mM DDT in all steps. DDT was removed by buffer exchanging into 100 mM KCl, 50 mM Tris pH 8.0 using an Amicon Ultra-15 50K MWCO (Millipore) concentrator. A DMSO stock solution of 200 mM TEMPO (N-(1-oxyl-2,2,6,6-tetramethyl-4-piperidinyl)maleimide) was added in 10× molar excess to a 0.3 mM solution of ClpP (total volume ~2.5 mL) and allowed to react overnight at room temperature. The reaction was terminated by buffer exchange into Buffer A prepared in D$_2$O. The reduced state of the spin label was prepared by incubating the samples with 100 mM L-ascorbic acid overnight, followed by buffer exchange into Buffer A.

MMTS labeling EcClpP and NmClpP with Cys mutations at positions 5 and 10, respectively, were purified as described above and subsequently incubated with 5 mM DTT for 1 hr. DDT was removed by buffer exchange with a degassed 100 mM KCl, 50 mM Tris pH 8.5 solution using an Amicon Ultra-15 50K MWCO (Millipore) concentrator. A DMSO stock solution of 100 mM ^{13}C-methyl-methanethiosulfonate (^{13}C-MMTS) was added in 20% molar excess to a solution of the protein (total volume ~2.5 mL) and allowed to react overnight at room temperature. The reactions were terminated by buffer exchange into Buffer A prepared in D_2O. ESI-MS was used to verify the incorporation of the MMTS moiety.

Preparation of mixed ClpP complexes ClpP complexes containing mixtures of WT:V7A or WT:S98A protomers were prepared following a procedure described previously (1). Briefly, ClpP molecules (pure WT and mutant) were mixed to achieve the desired ratio of WT and mutant protomers and concentrated to ~0.5 mL, followed by the addition of the unfolding buffer containing 100 mM KCl, 50 mM imidazole, 6 M GdnCl, and 10 mM DTT 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 to a final [GdnCl] of 300 μM. This refolded mixture was concentrated to ~1 mL using an Amicon Ultra-15 50K MWCO (Millipore) centrifugal concentrator and then applied to a HiLoad 16/60 Superdex 200 gel filtration column. Virtually no protein was found in the void peak and it eluted similarly to pure WT SaClpP. Fractions were pooled and buffer exchanged into Buffer A prepared in D_2O, followed by NMR spectroscopy. It should be noted that identical NMR spectra and catalytic activities were measured for samples that were first unfolded in GdnCl and subsequently refolded as for samples that were not manipulated in this way.

Methyl group assignment Over 95% of the Ile (20/22), Leu (17/17), Met (6/6), and Val (8/8) correlations in spectra of ClpP, 40 ˚C, were assigned by a combined mutagenesis and NOE-based approach, as described previously (2). Because multiple peaks were observed for many of the methyl groups each of the 53 ILVM-residues was mutated one at a time, generally Ile to Val and Met to Leu. Methyl-TROSY based 3D NOE experiments that record chemical shifts as ^{13}C[i]-NOE-^{13}C[j]-^{1}H[j] were measured with mixing times of 250 ms on [U-^{2}H; Ileδ1-$^{13}CH_3$; Leu/Val-$^{13}CH_3/^{12}CD_3$, Met-$^{13}CH_3$]-labeled samples of WT SaClpP, S98A SaClpP and V7A SaClpP (1.5 mM subunit concentration) to assist in the assignment process.

PRE measurements Relaxation rates for the slowly relaxing component of methyl ^{1}H transverse magnetization (R_{2s}^{H}) were quantified on a per-methyl group basis using the pulse scheme of Tugarinov and Kay (3) for samples where the spin-label was either in a paramagnetic (oxidized) or diamagnetic (reduced) state. R_{2s}^{H} values were obtained from exponential fits of peak intensities, *I*, as a function of relaxation delay, *T*, using the equation

$$I = I_0 \exp(-R_{2s}^{H} T) \ . \qquad \text{[S1]}$$

Subsequently PRE rates, Γ_2^{H}, were calculated according to

$$\Gamma_2^{H} = R_{2s,para}^{H} - R_{2s,dia}^{H} \qquad \text{[S2]}$$

and converted to distances, *r*, using the relation

$$r = \left[\frac{\frac{1}{15} S(S+1)\gamma^2 g^2 \beta^2}{\Gamma_2^{H}} \left(4\tau_c + \frac{3\tau_c}{1+{\omega_h}^2 {\tau_c}^2} \right) \right]^{\frac{1}{6}} \qquad \text{[S3]}$$

where τ_c is the correlation time for the overall tumbling of the protein complex, ω_h is the proton Larmor frequency, γ is the ^{1}H gyromagnetic ratio, g and β are the electronic *g*-factor, and Bohr magneton, respectively, and S=1/2 (4). This formula provides an upper bound for the estimated distance in the sense that a value of 1 is assumed for the order parameter that describes the amplitude of fluctuations of the bond vector connecting the electron and methyl ^{1}H spin of interest. A value of 156 ns was used for τ_c, as described below.

Uncertainties in PRE rates, Γ_2^H, were obtained from errors in $R^H_{2s,para}$ and $R^H_{2s,dia}$ values (generated from a Monte Carlo analysis using 1000 realizations of the decay data sets), $\sigma_{R^H_{2s,para}}$ and $\sigma_{R^H_{2s,dia}}$, respectively, and calculated as $\sigma_{\Gamma_2^H} = \sqrt{\sigma^2_{R^H_{2s,para}} + \sigma^2_{R^H_{2s,dia}}}$. The lower and upper bounds of distances between the paramagnetic site and each methyl ^{1}H spin were calculated from $\Gamma_2^H + \sigma_{\Gamma_2^H}$ and $\Gamma_2^H - \sigma_{\Gamma_2^H}$ respectively, using Eq. [S3], and plotted in Figure 2C (vertical error bars).

Measurement of order parameters S^2_{axis} Rapid (ps-ns) time-scale dynamics of ILVM-methyl containing residues were quantified using a ^{1}H triple-quantum-based relaxation-violated coherence transfer scheme, described previously (5). Relaxation delays, during which differential ^{1}H relaxation occurs, of 0.2, 0.4, 0.7, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 12.0, and 14.0 ms were used. A correlation time of 156 ns for the assumed isotropic tumbling of ClpP, 40 °C, was chosen by setting S^2_{axis}=1 for the methyl group with the largest η value (Eq. [S4]). Intensity ratios of peaks in triple-quantum and single-quantum data sets, I_a/I_b, were fit to

$$\left|\frac{I_a}{I_b}\right| = \frac{3}{4}\frac{\eta \tanh(\sqrt{\eta^2+\delta^2}T)}{\sqrt{\eta^2+\delta^2} - \delta \tanh(\sqrt{\eta^2+\delta^2}T)} \qquad \text{[S4]}$$

$$\eta = \frac{R^{\mathrm{H}}_{2f} - R^{\mathrm{H}}_{2s}}{2} \approx \frac{9}{10}\left(\frac{\mu_0}{4\pi}\right)^2 [P_2(\cos\theta_{axis,HH})]^2 \frac{S^2_{axis}\gamma_H^4\hbar^2\tau_c}{r^6_{HH}}$$

where η is the ^{1}H-^{1}H dipolar cross-correlated relaxation rate, δ is a fitting term that takes into account contributions from external ^{1}H spins, R^H_{2f} and R^H_{2s} are relaxation rates for fast and slowly relaxing components of ^{1}H magnetization, respectively, τ_c is the assumed isotropic molecular tumbling time, μ_0 is the vacuum permittivity constant, γ_H is the gyromagnetic ratio of a proton spin, r_HH is the distance between pairs of methyl protons (1.813 Å), and $\theta_\mathrm{axis,HH}$ (= 90°) is the angle between the methyl 3-fold axis and a vector connecting a pair of methyl ^{1}H spins. Uncertainties in η, σ_η, were estimated from a Monte Carlo analysis taking into account the uncertainties in I_a/I_b for each residue, calculated as $\sigma_{I_a/I_b} = \sqrt{\left\langle [(I_a/I_b)_{fit} - (I_a/I_b)]^2 \right\rangle}$, where the angular brackets denote the average over all the measured points in the profile. σ_η was calculated as the standard deviation of all obtained η values from the Monte Carlo run.

Estimation of an upper bound for exchange between conformers A 2D ^{1}H-^{13}C heteronuclear magnetization exchange experiment was recorded in order to estimate the exchange between ClpP conformers, exploiting the large multiplicity of peaks observed for residues in the protein (6). Exchange delay periods of 200, 300, 400, 500 and 600 ms were used, 50 ˚C, however cross-peaks were not observed. Following an approach described previously (7), and using a simple two-site exchange model, upper bounds for k_{ex} were estimated assuming that the absence of cross-peaks implies that their intensities are less than twice the average spectral noise level. Upper bounds were estimated to be 0.06 and 0.09 s^{-1} for Ile-4 and Ile-8, respectively.

Calculation of protomer distributions in particles generated from mixing experiments In what follows we consider the mixing of p_{WT} WT and (1- p_{WT}) mutant protomers, where p_{WT} is the fractional population of WT protomers that are mixed. The protomers are unfolded prior to mixing and subsequently allowed to refold to produce 14-mer complexes whose composition can be readily calculated, assuming random mixing, according to the relation,

$$C_k^{14} p_{WT}^k (1-p_{WT})^{14-k} \quad [\mathrm{S5}]$$

where k is the number of WT protomers in the complex and $C_k^{14} = \frac{14!}{k!(14-k)!}$. The k WT protomers can be distributed between the two heptameric rings of the 14-mer in a number of different ways that, in turn, can also be calculated. This is readily done by first focusing on each of the 7-mer rings individually and noting that the probability of a ring containing l WT protomers is

$$C_l^7 p_{WT}^l (1-p_{WT})^{7-l} \ . \quad [\mathrm{S6}]$$

Suppose that the two rings of the 14-mer contain l and j WT rings, such that $l+j=k$. Then the fraction of 14-mers with k WT protomers in total and l and j WT subunits in each 7-mer is given by

$$2C_l^7 p_{WT}^l (1-p_{WT})^{7-l} C_j^7 p_{WT}^j (1-p_{WT})^{7-j} = 2C_l^7 C_j^7 p_{WT}^k (1-p_{WT})^{14-k} \quad [\mathrm{S7}]$$

where the factor of 2 takes into account that there are two ways of distributing the l and j WT subunits in the pair of rings (l in 1 ring and j in the other), so long as $l \neq j$ (if $l=j$ the factor of 2 is removed). It is straightforward to show that by adding the probabilities corresponding to all of the permutations for which $l+j=k$ the probability given in Eq. [S5] is obtained.

Degradation assays The catalytic activities of a number of ClpP constructs were measured using the fluorogenic substrates, succinyl-L-leucyl-L-tyrosyl-AMC (LY-AMC) ($\lambda_{ex}/\lambda_{em}$: 355/460 nm) and α-casein-FITC ($\lambda_{ex}/\lambda_{em}$: 490/525 nm). An EnSpire (Perkin Elmer) 96-well microplate reader was used to monitor the assay for 120 minutes in 30-second increments, with shaking between each measurement. Each well contained 1 μM ClpP (monomer concentration) and 200 μM LY-

AMC or 15 μM α-casein-FITC incubated at 40 °C in Buffer A in a volume of 200 μL. ADEP-7 (prepared in house) was added to select wells at 2.5 μM. In the case of SaClpP all activities, established on the basis of initial rates of substrate degradation, are reported relative to apo WT SaClpP (LY-AMC) or ADEP-bound WT SaClpP (α-casein) samples. Activities are based on initial rates, extracted and analyzed using a python script written in-house. Standard errors are calculated from repeating each reaction a total of eight times within two independent sample preparations.

Electron cryomicroscopy All purified SaClpP constructs were concentrated to ~30 mg/mL in Buffer A. Immediately before freezing, 0.025 % (wt/vol) IGEPAL CA-630 (Sigma-Aldrich) was added to each protein solution to increase the number of particles adopting side-views on the EM grid. The addition of IGEPAL CA-630 reduced the particle density in the holes ~8 to 10 fold, requiring higher protein concentrations (~30 mg/mL) than is usual in cryoEM. A total of 2.5 μL of sample was applied to nanofabricated holey gold grids (8, 9), with a hole size of ~1 μm and blotted using a modified FEI Vitribot for 4.5 s before plunge freezing in a liquid ethane/propane mixture (ratio of ~3:2) at liquid nitrogen temperature.

For the WT SaClpP, Δ3 SaClpP, and V7A SaClpP + ADEP datasets, movies consisting of 30 frames at 2 frames/second were collected with defocuses ranging from 1.2 to 2.9 μm using a FEI Tecnai F20 microscope operating at 200 kV and equipped with a Gatan K2 Summit DDD camera. Movies were collected in counting mode at a nominal magnification of 25,000× corresponding to a calibrated pixel size of 1.45 Å and with an exposure of 5 electrons/pixel/s, giving a total exposure of 35 electrons/Å^2. For the WT SaClpP dataset 104 movies were collected and processed, while for the V7A SaClpP + ADEP and ΔN SaClpP datasets 75 movies were collected and processed.

The V7A SaClpP structure was calculated from movies consisting of 44 frames, obtained over a 60 second exposure with defocuses ranging from 1.0 to 3.9 µm recorded with a FEI Titan Krios G3 microscope operating at 300 kV and equipped with a FEI Falcon III DDD camera. Movies were collected in counting mode 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. 1837 movies were collected using the microscope's *EPU* software.

EM image analysis For all datasets, whole frame alignment was performed with *alignframes_lmbfgs* (10) and the resulting averages of frames were used for contrast transfer function (CTF) determination with *CTFFIND4* (11) and automated particle selection with *RELION* (12). Particle coordinates were used to extract particles in 160-pixel boxes for the K2 Summit data and 192-pixel boxes for the Falcon III data from the unaligned movies, while performing individual particle alignment and exposure weighting with *alignparts_lmbfgs* (10).

Particle image stacks were imported into *cryoSPARC* (13) for further analysis. For the WT SaClpP dataset consisting of 77789 particle images, *ab initio* map calculation was performed with two classes, one of which yielded the familiar symmetric barrel, while the other remained uninterpretable. The symmetric barrel was then used as a "good" reference and four copies of the un-interpretable map were used as "junk" references in a subsequent round of heterogeneous refinement designed to remove particles that were not contributing to the consensus structure. The 37218 particle images from the well-ordered class were then refined using D7 symmetry, resulting in a map at 3.6 Å resolution. Further attempts at classification with 3 to 5 classes did not improve the resolution or increase the interpretability of the apical loops. Omitting symmetry during map refinement resulted in a map at 4.2 Å resolution, however this map did not show marked asymmetry or improvement of the density in the apical region. The map determined without symmetry was used for subsequent MD calculations. Additional focused classification was performed, concentrating on the N-terminus of a single protomer, using masking and signal subtraction (12), however this procedure did not result in the identification of additional conformations. Map resolutions were assessed using the "gold standard" method of keeping the two halves of the datasets entirely independent and Fourier Shell Correlation curves were corrected for the effects of masking.

For the V7A SaClpP dataset of 878240 particle images, *ab initio* map calculation was performed with two classes, resulting in a 3D class with clear C2 symmetry from a broken-barrel structure. As with the WT SaClpP data, heterogeneous refinement using "good" and "junk" starting references was used to remove images that did not contribute high-resolution information, repeated twice to yield a class of 324522 particle images that were subsequently refined with C2 symmetry to 3.7 Å resolution.

The V7A SaClpP + ADEP dataset, comprising 76614 particle images, was treated in the same manner as for the WT SaClpP dataset, as the dominant class from *ab initio* map calculation was observed to be the D7 symmetric barrel conformation. Refinement of 37133 selected particle images led to a map at 5.6 Å resolution. For the Δ3 SaClpP dataset 48894 particle images were subjected to 2D classified with 50 classes in *cryoSPARC*.

Model building into EM density for V7A SaClpP Previous crystal structures of the extended (PDB 3V5E) (14) and compact forms (PDB 3QWD) (15) of ClpP were rigidly fitted into the EM density with *UCSF Chimera* (16). Six out of the seven subunits in each half fit the extended form well, while the remaining subunit (with the bent handle α-helix) was a better fit with the compact form. The protomers with an extended conformation in the 3.7 Å map were of sufficient quality to distinguish sidechains, while the protomer with a bent handle α-helix was only sufficient to build a poly-alanine model. These rigidly fit protomers were then subjected to successive rounds of manual model building in *Coot* (17) and real space refinement with geometry restraints in *Phenix* (18), resulting in a final model with an *EMringer* score (19) of 2.62 and a *molprobity score* (20) of 1.85 (Table 1).

Restrained molecular dynamics simulations Metainference (21) was used to produce an ensemble of models consistent both with the cryo-EM map of WT SaClpP and physico-chemical information about the system (22), the latter provided by the AMBER99SB-ILDN atomistic force field (23) and the GBSA implicit water model (24). An initial model was built from the crystal structure of SaClpP in the active, extended conformation (PDB code 3V5E (14)). The N-terminal residues 1-3 and 10-15 were added to each chain using MODELLER v9.17 (25). The initial model was first energy-minimized and then equilibrated at 300K for 5 ns. Subsequently, molecular dynamics simulations were carried out on a metainference ensemble of 24 replicas using a time step of 2 fs together with LINCS constraints (26). Van der Waals and Coulomb interactions were cut off at 2.0 nm and neighbor lists for all long-range interactions were similarly cut off at 2.0 nm and updated every 10 molecular dynamics steps. Simulations were carried out using a non-periodic cell in the canonical ensemble at 300K, enforced by the Bussi-Donadio-Parrinello thermostat (27). The cryo-EM map of WT SaClpP was fit with a Gaussian Mixture Model (GMM) using a divide-and-conquer approach (28). The GMM was composed of

15993 Gaussians and resulted in a cross-correlation with the experimental map equal to 0.98. A total aggregated simulation time of 0.8 μs was used, with each of the 24 replicas initialized from a random conformation extracted from the equilibration run. All simulations were performed with GROMACS 4.5.7 (29) and the PLUMED-ISDB module (30) of the PLUMED library (31). Methyl chemical shifts were back-calculated from the metainference ensemble using SHIFTX2 (32).

To assess the extent that an individual structural model or an ensemble of models is consistent with the cryo-EM density map of WT SaClpP, we calculated the local cross-correlation between the experimental density map and the map predicted from the model(s). The *gmconvert* (33) utility was used to predict a density map from a structural model. The local cross-correlation was calculated using Chimera (16) on a sliding window corresponding to a cubic box with dimensions of 5 voxels (7.25 Å). This analysis was carried out for three cases: (A) a model constructed from the X-ray structure (PDB code 3V5E (14)) after adding missing residues, as described above; (B) a cryo-EM-refined model constructed from the X-ray structure and optimized with the same force field used to model the metainference ensemble; (C) the metainference ensemble. In the latter case, the predicted map was calculated as the average of the maps generated using 11943 models randomly extracted from the metainference ensemble (see Fig. S8).

Supporting References

1. 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.

2. Sprangers R & Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445:618-622.
3. Tugarinov V & Kay LE (2006) Relaxation Rates of Degenerate 1H Transitions in Methyl Groups of Proteins as Reporters of Side-Chain Dynamics. *J Am Chem Soc* 128(22):7299-7308.
4. Battiste JL & Wagner G (2000) Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data. *Biochemistry* 39(18):5355-5365.
5. Sun H, Kay LE, & Tugarinov V (2011) An Optimized Relaxation-Based Coherence Transfer NMR Experiment for the Measurement of Side-Chain Order in Methyl-Protonated, Highly Deuterated Proteins. *J Phys Chem B* 115(49):14878-14884.
6. Religa TL, Sprangers R, & Kay LE (2010) Dynamic Regulation of Archaeal Proteasome Gate Opening As Studied by TROSY NMR. *Science* 328(5974):98-102.
7. Huang R*, et al.* (2016) Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study. *Proc Natl Acad Sci USA* 113(29):E4190-E4199.
8. 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.
9. Russo CJ & Passmore LA (2014) Ultrastable gold substrates for electron cryomicroscopy. *Science* 346(6215):1377-1380.
10. Rubinstein JL & Brubaker MA (2015) Alignment of cryo-EM movies of individual particles by optimization of image translations. *J Struct Biol* 192(2):188-195.
11. Rohou A & Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 192(2):216-221.
12. Scheres SHW (2015) Semi-automated selection of cryo-EM particles in RELION-1.3. *J Struct Biol* 189(2):114-122.
13. Punjani A, Rubinstein JL, Fleet DJ, & Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14(3):290-+.
14. Gersch M, List A, Groll M, & Sieber SA (2012) Insights into Structural Network Responsible for Oligomerization and Activity of Bacterial Virulence Regulator Caseinolytic Protease P (ClpP) Protein. *J Biol Chem* 287(12):9484-9494.
15. Geiger SR, Bottcher T, Sieber SA, & Cramer P (2011) A Conformational Switch Underlies ClpP Protease Function. *Angew Chem Int Ed* 50:5749-5752.
16. Pettersen EF*, et al.* (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comp Chem* 25(13):1605-1612.
17. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr Sect D-Biol Crystallogr* 66:486-501.
18. Afonine PV, Headd JJ, Terwilliger TC, & Adams PD (2013) PHENIX News. *Computational Crystallography Newsletter* 4:43-44.
19. Barad BA*, et al.* (2015) EMRinger: side chain directed model and map validation for 3D cryo-electron microscopy. *Nature Methods* 12(10):943-946.
20. Chen VB*, et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D-Biol Crystallogr* 66:12-21.

21. Bonomi M, Camilloni C, Cavalli A, & Vendruscolo M (2016) Metainference: A Bayesian inference method for heterogeneous systems. *Sci Adv* 2(1):e1501177.
22. Bonomi M, Pellarin R, & Vendruscolo M (2017) Simultaneous Determination of Protein Structure and Dynamics Using Cryo-Electron Microscopy. *bioRxiv* doi: 10.1101/219972.
23. Lindorff-Larsen K*, et al*. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78(8):1950-1958.
24. Still WC, Tempczyk A, Hawley RC, & Hendrickson T (1990) Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J Am Chem Soc* 112(16):6127-6129.
25. Sali A & Blundell TL (1993) Comparative Protein Modeling by Satisfaction of Spatial Restraints. *J Mol Biol* 234(3):779-815.
26. Hess B (2008) P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theo Comp* 4(1):116-122.
27. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1):0141011-0141017.
28. Hanot S*, et al*. (2017) Bayesian multi-scale modeling of macromolecular structures based on cryo-electron microscopy density maps. *bioRxiv* doi: 10.1101/113951.
29. Hess B, Kutzner C, van der Spoel D, & Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* 4:435-447.
30. Bonomi M & Camilloni C (2017) Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics* 33(24):3999-4000.
31. Tribello GA, Bonomi M, Branduardi D, Camilloni C, & Bussi G (2014) PLUMED 2: New feathers for an old bird. *Comput Phys Commun* 185(2):604-613.
32. Han B, Liu Y, Ginzinger SW, & Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* 50(1):43.
33. Kawabata T (2008) Multiple Subunit Fitting into a Low-Resolution Density Map of a Macromolecular Complex Using a Gaussian Mixture Model. *Biophys J* 95(10):4643-4658.
34. Cardone G, Heymann JB, & Steven AC (2013) One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *J Struct Biol* 184(2):226-236.

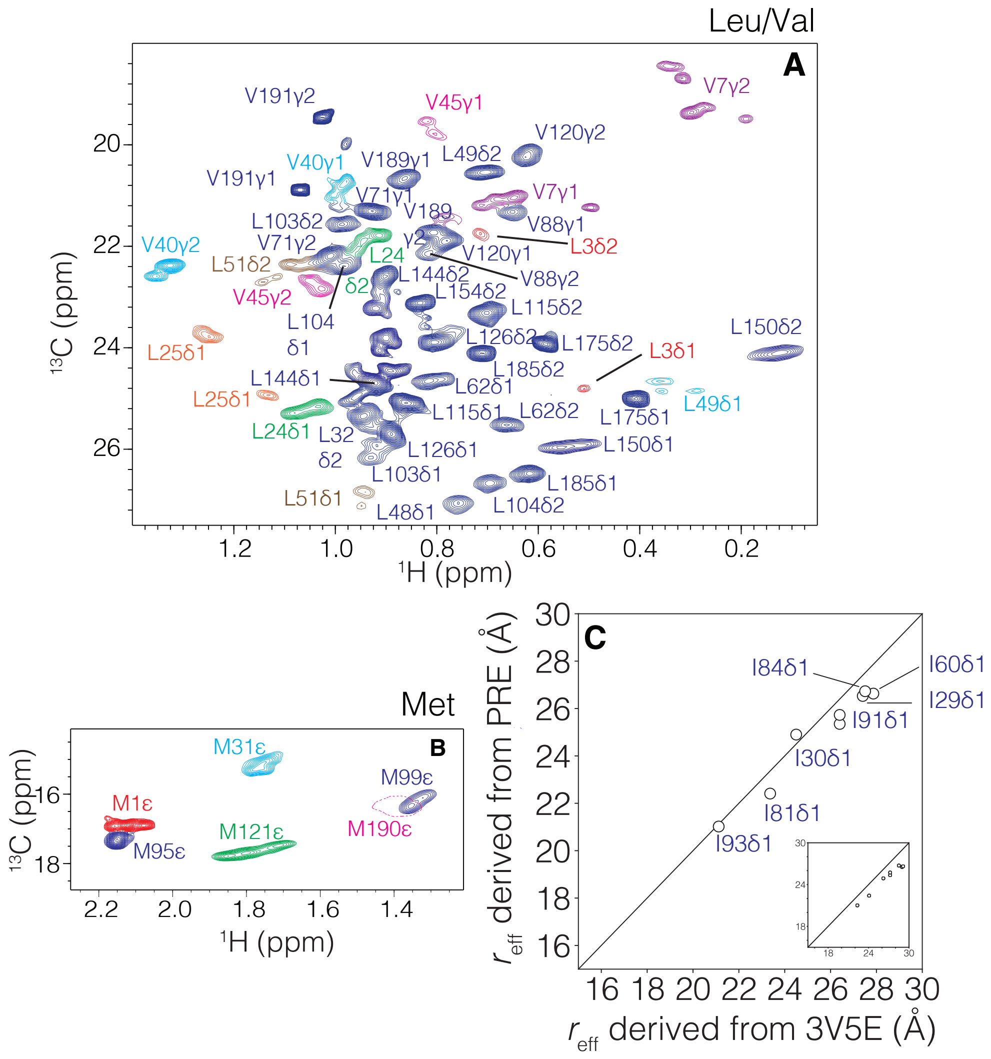

Figure S1. Conformational heterogeneity of SaClpP as observed by methyl-TROSY NMR. (A) Leu/Val and (B) Met region of the ^{1}H-^{13}C correlation map of [U-^{2}H Ileδ1-^{13}CH$_3$; Leu/Val-^{13}CH$_3$/^{12}CD$_3$; Met-^{13}CH$_3$]-labeled WT SaClpP, 40 ˚C, 18.8 T. Methyl chemical shift assignments are included. (C) Comparison of measured distances between a spin label attached to position 130 of WT SaClpP and methyl groups from well-ordered regions of the protein as derived from PRE measurements (vertical axis) with the corresponding distances calculated from an X-ray structure (horizontal axis) (PDB 3V5E (14)). Effective distances in the crystal structure are calculated as $r_{eff}^{-1} = \sqrt[6]{\sum_{i=1}^{14} r_i^{-6}}$, where r_i is the distance from the oxygen atom of the piperidinyloxy moiety of the spin label (main figure) or Cβ of Gln-130 (shown as inset) on chain i to the Cδ position of an Ile residue of any given protomer. The line y=x is shown as a guide. The good agreement observed here provides confidence in distance measurements between the spin-label and the N-terminal residues reported in Figure 2C. Each data point in the plot corresponds to an Ile residue, as annotated.

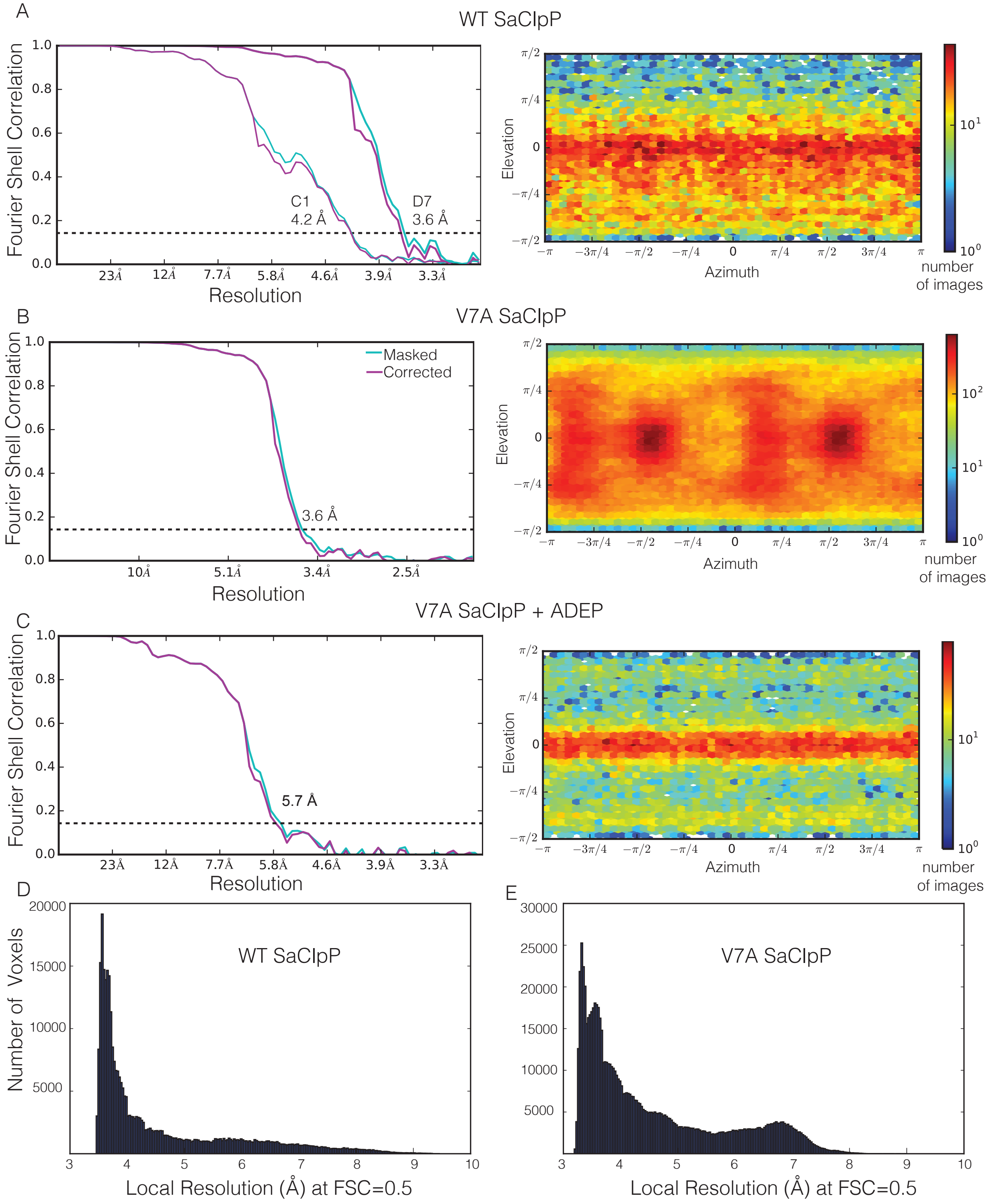

Figure S2. (A) (Left) Fourier shell correlation (FSC) for WT SaClpP after gold-standard refinement (resolutions reported at FSC=0.143). Curves for both C1 and D7 reconstructions are shown. FSC curves were calculated after masking solvent from the background of the map (cyan curve) and corrected for the effects of masking using phase randomization (purple curve). (Right) Euler angle distribution plot for the final refinement of the WT map. (B) FSC curve and Euler angle distribution for V7A SaClpP. (C) FSC curve and Euler angle distribution for V7A SaClpP with added ADEP. (D) Local resolution of WT SaClpP map estimated by a local windowed FSC method using an implementation of the *locres* (34) algorithm in *cryoSPARC* (13). The majority of the map is close to the global resolution estimate, with a long tail of voxels at lower resolutions. (E) Local resolution estimates for V7A SaClpP. The bimodal distribution reflects the lower resolution of the flexible protomers with bent handle helices.

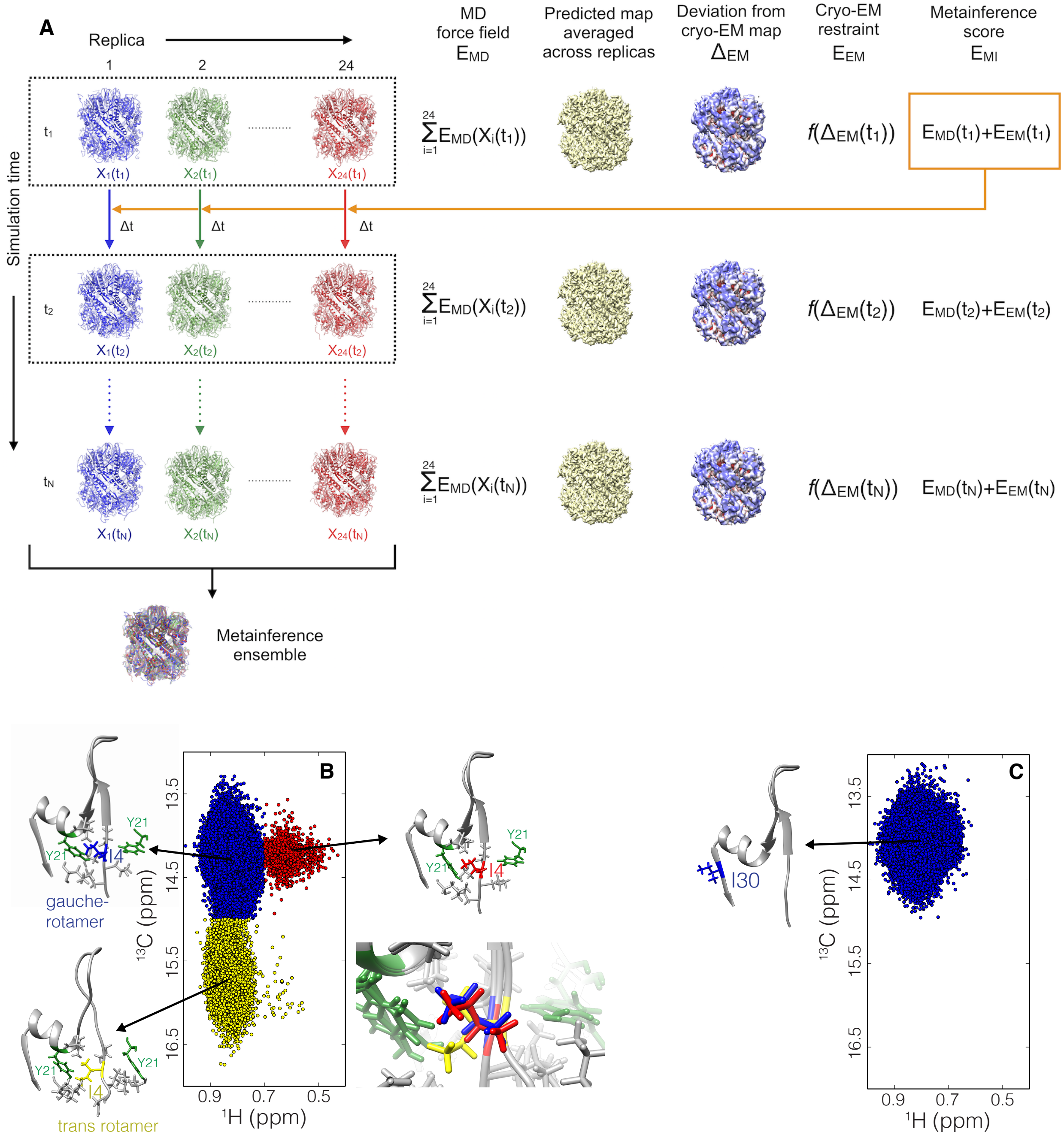

Figure S3. Metainference dynamics analysis of ClpP establishes that the N-domain is heterogeneous. (A) Schematic of the approach used. To generate an ensemble of models consistent with the cryo-EM map of WT SaClpP we used metainference (21), a Bayesian framework to model conformational ensembles by integrating prior information on a system with noisy, ensemble-averaged experimental data. Metainference models a heterogeneous system and all sources of error in the input information by simulating *N* replicas of the system (24 in this study) using molecular dynamics (MD). The generation of models is guided by the metainference score, defined as $E_{MI}(\{X_i\}) = -k_B T \log p(\{X_i\})$, where $\{X_i\}$ are the replica conformations, k_B is Boltzmann's constant, T the temperature of the system (300 K in this study), and p the metainference posterior probability, defined for cryo-EM data (22). The metainference score can be decomposed into two terms $E_{MI}(\{X_i\}) = \sum_{i=1}^{N} E_{MD}(X_i) + E_{EM}(\{X_i\})$. Here E_{MD} is the score based on standard molecular dynamics simulations (AMBER99SB-ILDN with the GBSA implicit water in this study) and E_{EM} is the cryo-EM score that, together, are used to generate forces acting on each of the ensemble replicas to restrain the predicted map averaged over all replicas to be close to the experimental map. The metainference simulation proceeds as follows. First, an initial model was built from the crystal structure of SaClpP in the active, extended conformation (PDB code 3V5E (14)) and equilibrated at 300 K for 5 ns. From this equilibration run, 24 structures were randomly extracted and used as initial replica conformations. The initial velocities were randomly extracted from the Maxwell-Boltzmann distribution at 300 K. At each MD step, the MD force field E_{MD} was calculated for each individual replica. A predictor of cryo-EM density maps (22) was then used to generate a synthetic map from the conformation of each replica. The deviation between the experimental map and the synthetic map averaged across all replicas was used to calculate the cryo-EM score, E_{EM}. In the figure the deviation is visualized by the local cross-correlation (lcc) between the predicted and experimental density maps, color-coded as red (lcc ~0.7), white (~0.85), and blue (~1). The total metainference score $E_{MI}(\{X_i\})$ was used to evolve the positions and velocities of all replicas using the velocity verlet integrator and a time step of 2 fs. The procedure was re-iterated for a total aggregated simulation time of 0.8 μs. (B,C) The collection of all replica conformations generated by the procedure described above defines the metainference ensemble and was used to back-calculate ^{1}H and ^{13}C chemical shifts, as predicted by SHIFTX2 (32) for (B) Ile-4δ1 and (C) Ile-30δ1. Select conformations from each chemical shift cluster are shown to highlight the structural differences.

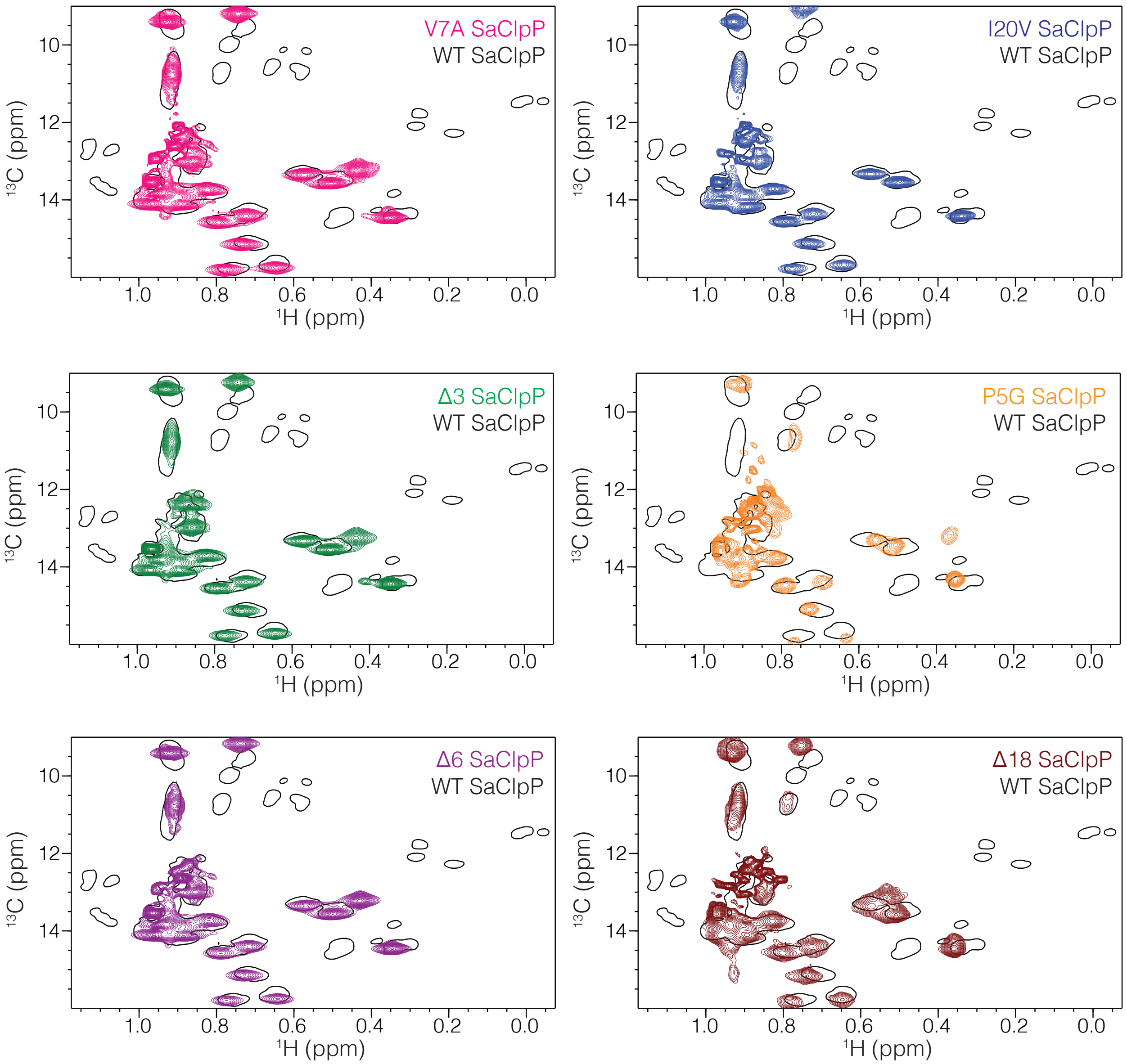

Figure S4. N-terminal mutants of SaClpP are structurally similar. Overlay of the Ile regions of ^{1}H-^{13}C correlation maps of ILVM-labeled WT (single-contour, black) and N-terminal mutant (multi-contour, colored) SaClpP, 40 ˚C, 18.8 T. HMQC spectra of all N-terminal mutants are nearly indistinguishable from that of V7A SaClpP, indicating that the N domain is unfolded in each case. The Δ3, Δ6 and Δ18 constructs are deletion mutants lacking residues Asn-2 – Ile-4 (Δ3), Asn-2 – Val-7 (Δ6) and Asn-2 – Asp-19 (Δ18).

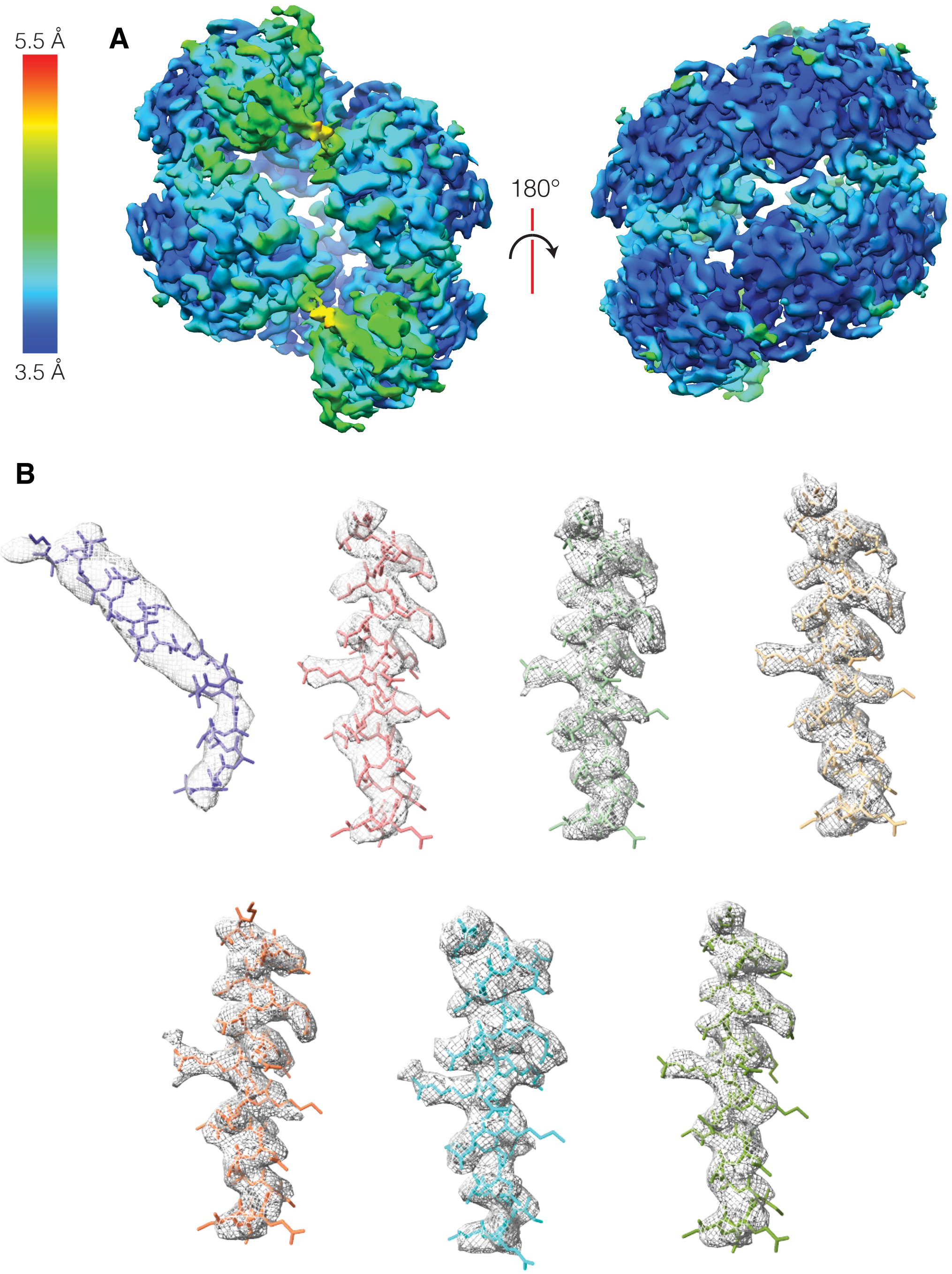

Figure S5. (A) Local resolution map of V7A SaClpP. The majority of protomers lie close to the global resolution estimate of 3.6 Å, with the exception of the protomers forming the unique interface abutting the side pores. The apparent flexibility of these protomers leads to resolution estimates closer to 4 to 5 Å. (B) Model-in map fits for the handle-helix of protomers in descending order (*i.e.*, 7 protomers of a ring starting from the top protomer of A). For the bent handle helix in the top protomer, only a poly-alanine model could be fit into the 4 to 5 Å resolution density.

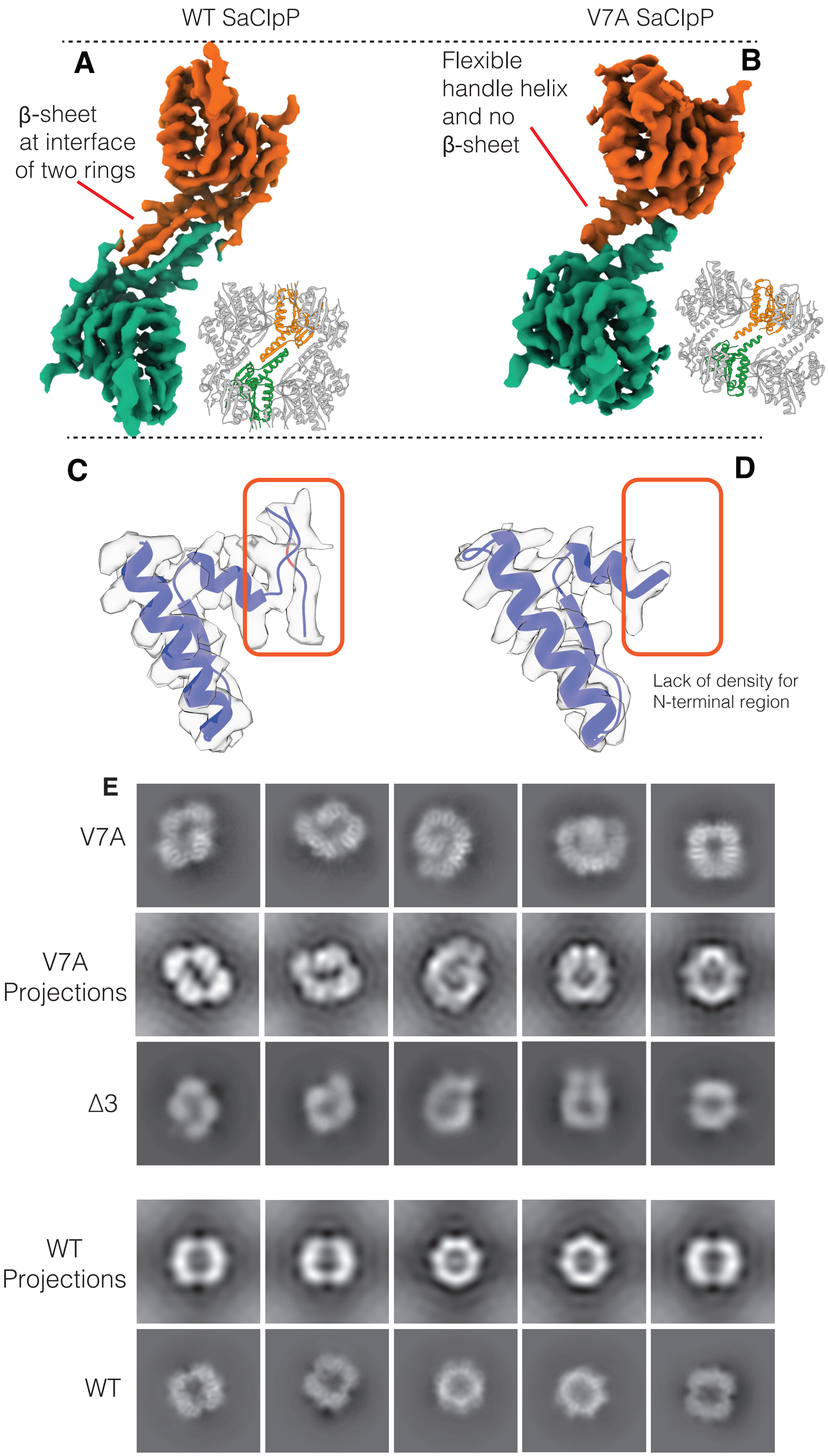
WT SaClpP
V7A SaClpP
A
β-sheet at interface of two rings
Flexible handle helix and no β-sheet
B
C
D
Lack of density for N-terminal region
E
V7A
V7A Projections
Δ3
WT Projections
WT

Figure S6. Structural differences at the ring-ring interface of WT and V7A SaClpP. Regions from cryo-EM density maps focusing on select protomers from (A,C) WT SaClpP and (B,D) V7A SaClpP. ClpP full structures are shown to highlight the protomers that are depicted. Key structural differences are localized to (i) the handle region (A,B) that connects the heptameric rings as well as to (ii) the N-terminal domains (C,D). In V7A SaClpP the handle α-helix appears more flexible and density for the handle β-sheet was not observed for the handle. The dashed lines are drawn to highlight the contraction in the height of the V7A protomers relative to WT SaClpP. (E) 2D class averages of V7A, Δ3, and WT SaClpP along with 2D projections from the 3D maps of V7A and WT SaClpP. Projections have been low-pass filtered to 20 Å for comparison to Δ3 class averages. The similarities of the V7A and Δ3 2D classes, the Δ3 2D classes and the V7A projections, as well as the close correspondence of NMR spectra for these constructs indicates that Δ3 also forms the split-ring structure (as is almost certainly the case for the other mutants considered, Fig. S4).

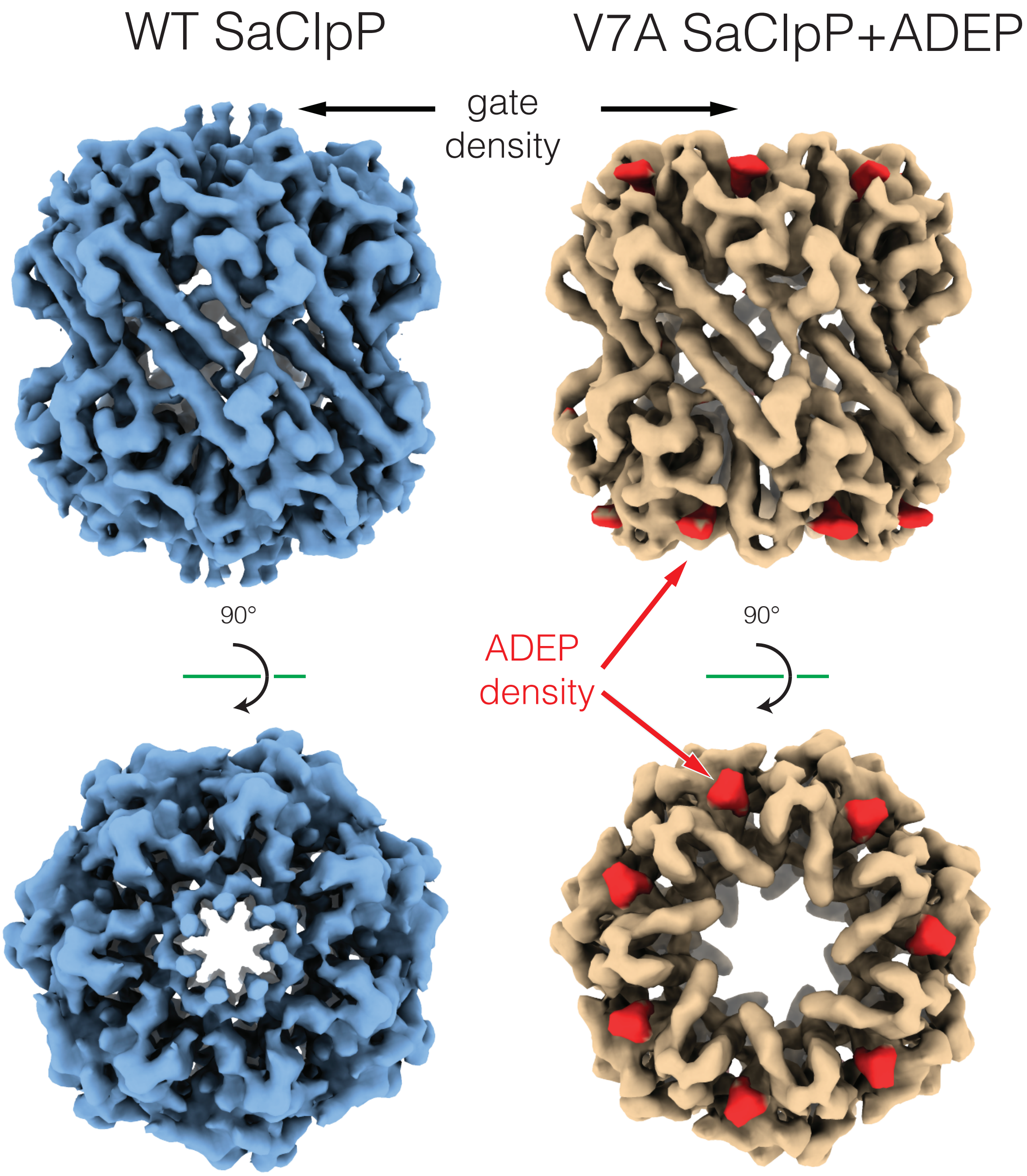

Figure S7. ADEP binding to V7A SaClpP repopulates the extended form. Top and side views of cryo-EM density maps of WT SaClpP and V7A SaClpP bound to ADEP. The WT map was filtered to match the lower 6.4 Å resolution of the V7A SaClpP+ADEP map. Density corresponding to ADEP is shown in red. Overall, the structure of ADEP-bound V7A SaClpP resembles the active, extended form with the exception of the N-terminal residues, for which no density is detected.

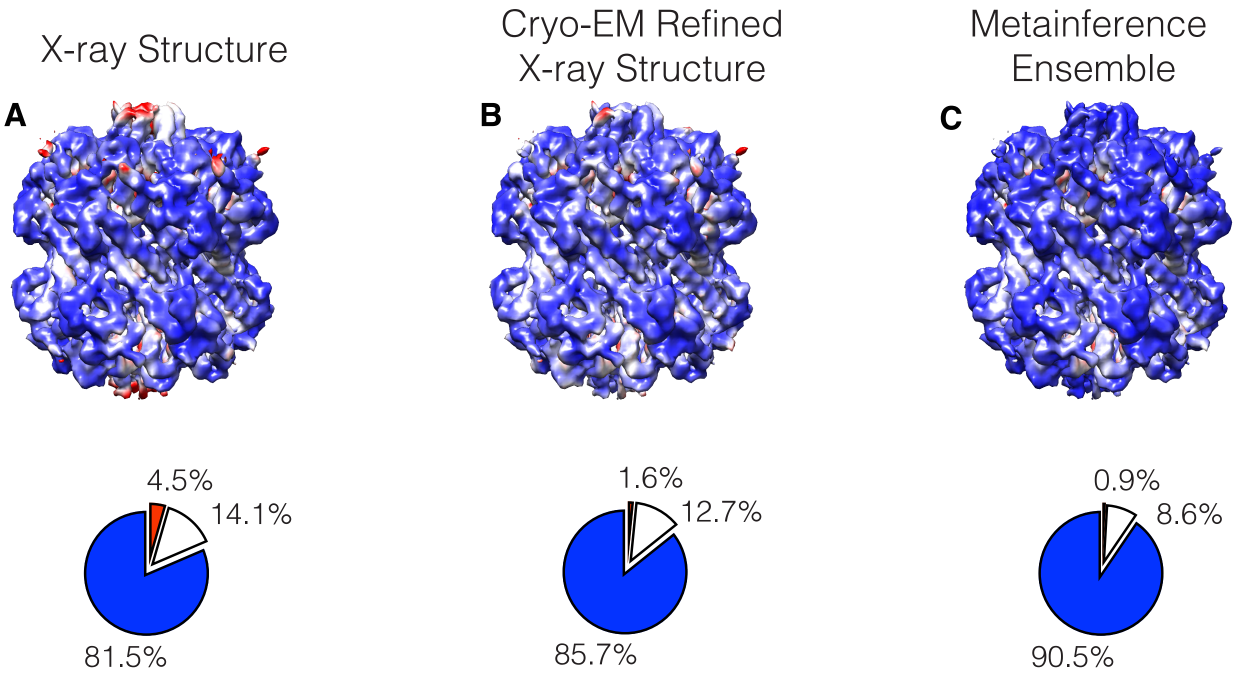

Figure S8. The metainference ensemble of structures better explain cryo-EM density than a single conformation. As described in SI Materials and Methods, a cross-correlation analysis was carried out between the experimental cryo-EM density map and density maps generated from (A) an X-ray model (PDB code 3V5E (14)) after adding missing residues; (B) a model from cryo-EM; and (C) the metainference ensemble. In each of these three cases, the experimental density map is visualized at a threshold equal to 0.25 and colored according to the local cross-correlation (*lcc)* with the predicted map. Coloring ranges from red (*lcc*<=0.7), to white (*lcc*=0.85), to blue (*lcc*=1.0). To better quantify the agreement with the experimental map, we calculated the fraction of voxels with density values between 0.2 and 0.3 and with *lcc*<=0.7, 0.7<*lcc*<=0.85, and *lcc*>0.85. These fractions are visualized in the pie charts and colored in red, white, and blue, respectively.

Table 1. Cryo-EM data acquisition, processing, and atomic model statistics for V7A SaClpP

Data Collection	**V7A SaClpP**
Electron Microscope	FEI Titan Krios
Camera	Falcon3EC
Voltage	300 kV
Nominal Magnification	75,000×
Calibrated physical pixel size	1.06 Å
Total exposure	43 electrons/Å^2
Exposure rate	0.8 electrons/pixel/s
Number of frames	44
Total exposure time	60 seconds
Defocus range	1.0-3.9 µm
Image Processing	
Motion correction software	*alignparts_lmbfgs*
CTF estimation software	CTFFIND4
Particle selection software	RELION 1.4
Micrographs used	1837
Particles selected	878240
3D map classification and refinement software	cryoSPARC
Particles contributing to final map	324522
Applied symmetry	C2
Applied B-factor	-217 Å^2
Global resolution (FSC = 0.143)	3.7 Å
Model Building	
Modeling software	*Coot, Phenix*
Number of residues built	2314
RMS (bonds)	0.0101
RMS (angles)	1.18
Ramachandaran outliers	0.00 %
Ramachandran favoured	96.1 %
Clashscore	11.63
MolProbity score	1.85
EMRinger score	2.62