STRUCTURAL BIOLOGY COMMUNICATIONS

Volume 73 (2017)

Supporting information for article:

Crystal structure of the *Thermoplasma acidophilum* protein Ta1207

Ganesh Ramnath Pathare, István Nagy, Ágnes Hubert, Dennis R. Thomas and Andreas Bracher

S1. Cryo-electron microscopy

A 4 µl sample of the isolated Ta1207 complexes at an approximate concentration of 0.3 mg/ml was applied to custom ordered 200 mesh copper Quantifoil grids with 0.6 µm diameter holes spaced 1.0 µm apart. The specimen was blotted using Whatman #1 filter paper then flash-frozen in liquid ethane or an ethane-propane mixture cooled by liquid nitrogen. Images were acquired on a Tecnai F20 microscope (FEI, Eindhoven, Netherlands) operating at 120 kV, using a Gatan 656 cryo-holder (Gatan, Munich, DE). Images were acquired on a 4K FEI EAGLE (FEI, Eindhoven, Netherlands) charge-coupled device (CCD) camera at 111,100x magnification using SerialEM. The final pixel resolution was 1.35 Å/pixel. The images were collected with a defocus range (ΔF) from -0.8 and -2.8 µm with an electron dose of 15-20 $e^-/Å^2$. Images were screened for the presence of particles and good contrast transfer function (CTF) using the E2projectmanager from EMAN2. The defocus for each good image was determined and the image phases were corrected for the CTF using TOM_ctffindgui.

S2. EM analysis

The X-ray structure of this complex had been solved as a decamer and chromatography and light scattering data indicated the complex in solution was primarily a pentamer. Therefore, both a pentamer and decamer based on the Ta1207 crystal structure were used as starting models. These models were filtered to 30 Å resolution before generating reference projections of both. A gallery of false positives collected from class averages of numerous classified reference free alignments was included in the reference set to trap false positives and remove them from the dataset. We have found multi-reference alignment, including false positives as references, to be more effective than classification of the reference free aligned particles alone, as a greater number of false positives are removed than by classification alone. Images that matched the 10- or 5-fold reference were reconstructed and refined separately. False positives were not included in the reconstructions. After six rounds of refinement the assignment to either 5-fold, 10-fold or false positive becomes stable. The 10-fold class appeared to be incomplete complexes where some subset of subunits bound to the pentamer or there were neighboring particles. The three subsets of particles were subjected to classification to confirm the 10-fold, 5-fold or false positive nature of each class. The 10-fold particle class and the false positives were removed the datasets.

Image processing was performed with the SPIDER software package. Iterative multi-reference alignment of the images to projections of a reference map was used to align the images and generate particle orientations. The images underwent many rounds of refinement with the current reconstruction becoming the new reference.

The Fourier shell correlation was calculated to assess the resolution of the reconstructions. The SPIDER bp 32f Fourier based reconstruction algorithm was used for reconstruction. The final reconstructions were normalized with respect to one another by setting all maps to a constant standard deviation. The

maps were then Gaussian low pass filtered to 12 Å for display and further analysis. Negative values were set to zero before calculating differences between maps. Maps are displayed using CHIMERA with a threshold of 1 sigma and differences are displayed at a two sigma threshold based on the original map statistics.

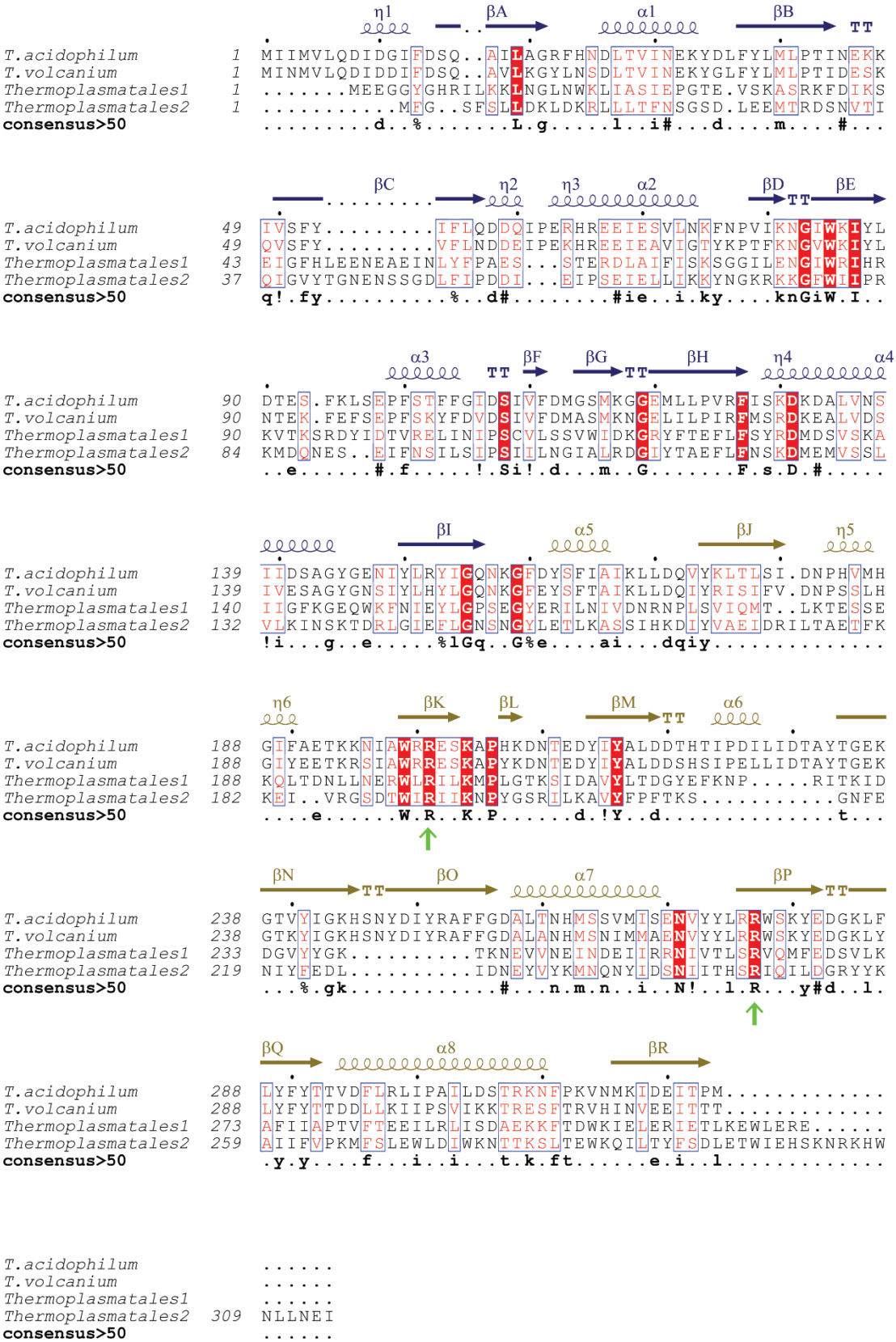

Figure S1 Multiple sequence alignment (MSA) of Ta1207 with the sequence homologs from *Thermoplasma volcanium* and two unknown *Thermoplasmatales* species from a mine drainage metagenome (Yelton *et al.*, 2013). Secondary structure elements are indicated above the sequences. The domain structure is indicated by blue and gold coloring of secondary structure. In the alignment, similar residues are shown in red and identical residues in white using bold lettering on red background. Blue frames indicate homologous regions. All sequence numbering is based on the open reading frames. The arginine clamp residues are indicated by green arrows. The Uniprot accession codes for the sequences are: Q9HIW9, *Thermoplasma acidophilum*; Q978Y5, *Thermoplasma volcanium*; T0N3P8, *Thermoplasmatales* archaeon Gpl; T0LRQ1, *Thermoplasmatales* archaeon E-plasma. The Figure was prepared with the program ESPript (Gouet et al., 1999).

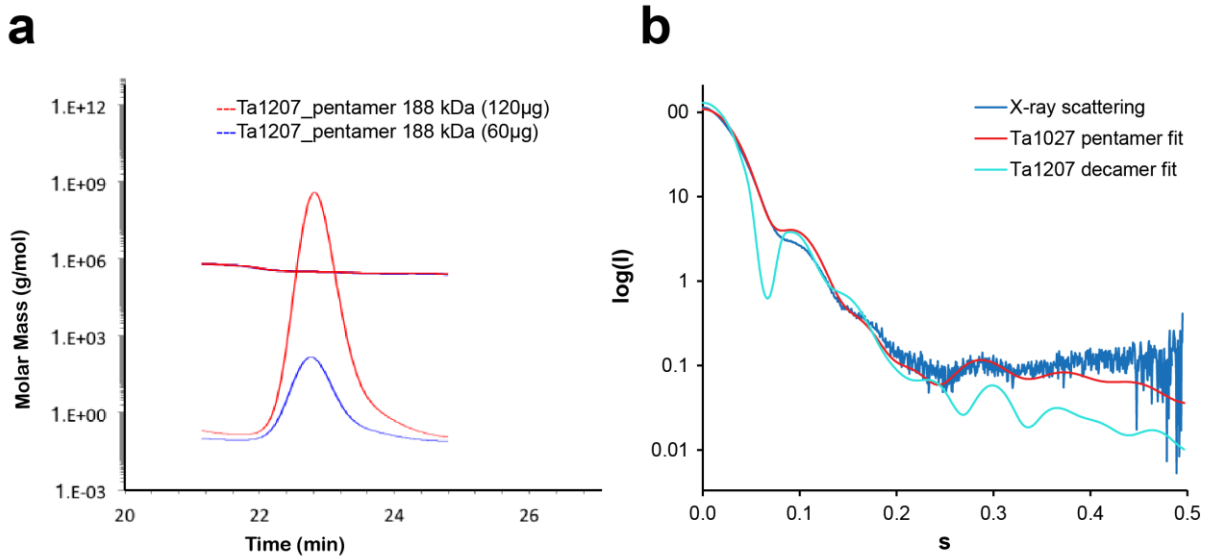

Figure S2 Solution structure of Ta1207: (a) Size exclusion chromatography-Multi-Angle Laser Light Scattering (SEC-MALS) analysis of Ta1207. Absorbance traces and molecular weights are shown. The molecular weight was determined from 60 µg (red) and 120 µg (blue) Ta1207 in 10 mM Tris pH 7.5, 100 mM NaCl and 50 mM KCl and 2 mM DTT. The molar mass and the radius of gyration (R_g) of Ta1207 was calculated using the ASTRA software (Wyatt technology). (b) Small Angle X-ray Scattering (SAXS) data of Ta1207. The experimental scattering curve is compared with the theoretical scattering curves of the crystallographic pentamer and decamer, respectively. SAXS measurements were performed at beamline BM29 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Protein samples at four different concentrations in 10 mM Tris pH 7.5, 100 mM NaCl and 50 mM KCl were exposed to X-rays for 1 s. Scattering data from ten repeats were averaged. Buffer background was subtracted. The protein scattering data were processed with Primus (Konarev et al., 2003, Vachette et al., 2003). Radii of gyration were determined using the Guinier approximation. Scattering curves were fitted to the crystallographic model with Crysol.

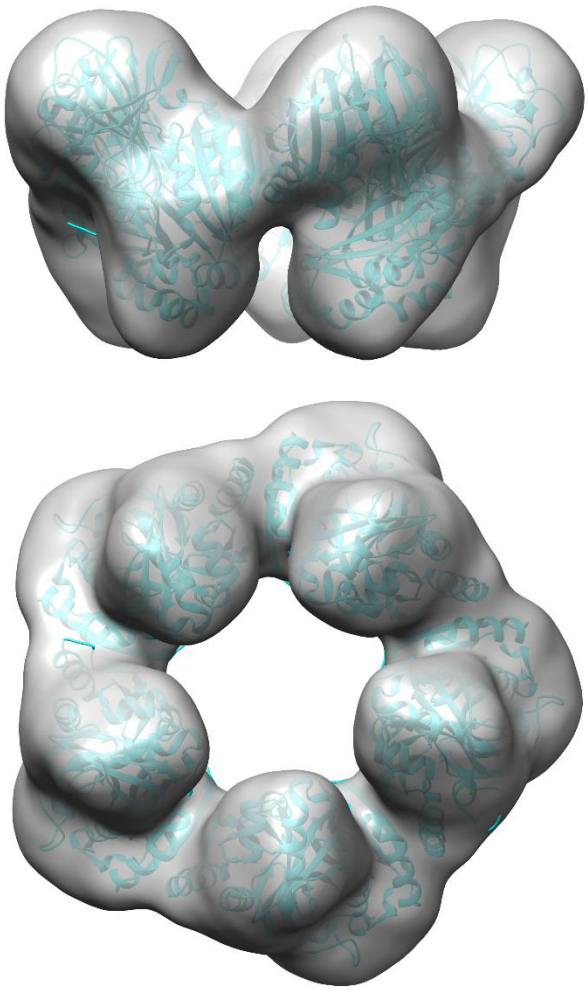

Figure S3 Fit of the crystallographic model to the cryoEM envelope of endogenous Ta1207. The Ta1207 pentamer is shown as ribbons in cyan.