Acta Cryst D
BIOLOGICAL CRYSTALLOGRAPHY

Volume 71 (2015)

Supporting information for article:

The structure of *Aquifex aeolicus* FtsH in the ADP-bound state reveals a C_2-symmetric hexamer

Marina Vostrukhina, Alexander Popov, Elena Brunstein, Martin A. Lanz, Renato Baumgartner, Christoph Bieniossek, Magdalena Schacherl and Ulrich Baumann

S1. Evidence for refinement of Δ-TtFtsH against √F, not F or I

The structure of the *T. thermophilus* FtsH whole cytosolic region (Δ-*Tt*FtsH) was first published in 2006 (Suno *et al.*, 2006) and deposited in the PDB under entry number 2DHR. By that time the refinement statistics employing the program Refmac (Murshudov *et al.*, 1997) reported an R/R_{free} (%) of 30.1/34.2% at 3.9 Å resolution, including data between 15 and 3.9 Å with a cut-off of 3 sigma(F). The authors described the use of the program REFMAC, including TLS refinement, although no TLS groups can be found in the header of the PDB file of entry 2DHR. The structure factors deposited in the mmcif file of entry 2DHR are labeled as "refln.intensity_meas" and "refln.intensity_sigma", hinting at the deposition of experimental intensities, not amplitudes.

In 2012 a re-refined model (Refmac) was deposited by the same authors (PDB entry 4EIW) (Suno *et al.*, 2012) with a reported R/R_{free} (%) of 30.0/31.2 at 3.9 Å resolution including all data. The program used for refinement was Refmac version 5.6.0117. The deposited structure factors in the CIF file are labeled "refln.intensity_meas" and "refln.intensity_sigma", which are identical to those reported for entry 2DHR. Additionally, columns labeled as "refln.F_meas_au" and "refln.F_meas_sigma_au" are now found in the structure factor file of entry 4EIW, indicating the deposition of the structure factor amplitudes as well. These structure factor amplitudes are approximately the square root of the numbers of the values in the column labeled as intensities.

Assuming correctly labeled data, the deposited structure factors possess highly distorted intensity statistics as analyzed by phenix.xtriage from the Phenix program suite (Table S1). For acentric data, the second-order moment of the intensity distribution $<I^2>/<I>^2$ should be 2.0 for non-twinned or 1.5 for perfectly merohedrally twinned data. In the deposited structure factor file it is 1.236 (Table S1) and the mean value of $<E^2 - 1>$ is 0.387 (expected 0.736 for untwinned and 0.541 for perfectly twinned data). Similarly, the $|L|$ test (Padilla & Yeates, 2003) yields a value of 0.269 (expected 0.500 for untwinned and 0.375 for perfectly twinned data). Likewise, the cumulative intensity statistics shows very strong deviations from the expectation, revealing much fewer weak reflections than actually calculated for a perfect 1:1 twin (Table S1). Such strong deviations from the expected intensity statistics are more likely to arise from mislabeling of amplitudes as intensities rather than from twinning or other crystal pathologies. A simple re-declaration of the intensities as amplitudes cures most of the problems, as shown in Table S1, although the centric reflections still exhibit some significant deviation from the expected distribution.

In order to test whether the deposited model of entry 4EIW was refined against potentially mislabeled data, we re-refined it without any adjustments using Phenix (version 1833) or Refmac (version 5.8.0073). The refinement protocols included restrained positional and ADP refinement, TLS group refinement, and used NCS restraints and default bulk solvent scaling. TLS groups were taken

from the PDB file header. In Phenix.refine, optimization of the solvent mask and geometric and B-factor weights were also employed. For Refmac refinement, the test and working reflections were used as deposited. Because the Rfree flags are incomplete in entry 4EIW and 2DHR, for Phenix the test set had to be completed, resulting in minor differences in the number of test/working reflections compared to the original entry. The purpose of this re-refinement was solely the firm establishment of the R-factors at the beginning and end of the refinement cycle and did not serve for model improvement.

When refining just against the original deposited data the results displayed in Table S1 were obtained. Here, the initial R and R_{free} values are close to those reported in entry 4EIW.

Refinement against re-labeled data, i.e., declaring intensities as amplitudes, yielded initial R-factors in Refmac refinement that were very different from those reported by Suno *et al.* for entry 4EIW. Thus, it is likely that refinement of entry 4EIW had been carried out against the square root of the structure factor amplitudes (√F), and not against the amplitudes (F).

The electron density maps appear similar, although the one calculated from the relabeled structure factor amplitudes appears to exhibit much less model bias. In all maps, the 'lid helix' (residues 448–456) is virtually completely disordered in all six crystallographically independent copies. Simulated annealing omit maps show no significant density over the whole 'lid-helix' region using either deposited or relabeled data (Fig. S2). Omitting other parts, e.g., parts of the active-site helix α10, brings back some density in the $2F_O$-F_C and F_O-F_C maps (Fig. S2). Thus, we conclude that the so-called 'lid-helix' segment is disordered.

Table S1 Intensity and refinement statistics of re-refinement of entry 4EIW *

	Original data	Relabeled $I \rightarrow F$
Intensity statistics		
$<I^2>/<I>^2$ acentric (untwinned: 2.0; perfect twin: 1.5)	1.236	1.944
$<I^2>/<I>^2$ centric (untwinned: 3.0; perfect twin: 2.0)	1.350	2.394
$<\|E^2 - 1\|>$ acentric (untwinned: 0.736; perfect twin: 0.541)	0.387	0.691
$<\|L\|>$ / $<L^2>$ (untwinned: 0.500/0.333 perfect twin: 0.357/0.200)	0.269 / 0.108	0.457/0.284
N(Z) maximum deviation acentric/centric	0.248 / 0.368	0.039 / 0.170
<N(Z)(obs) – N(Z)(twin)> acentric/centric	-0.156 / -0.259	-0.003 / -0.123
Refinement REFMAC		
Resolution range (Å)	71.53 – 3.90	71.53 – 3.90
No. of reflections working/test set	37,156 / 1,960	37,267 / 1,967
Start R/R_{free}	0.309 / 0.322	0.357 / 0.372
End R/R_{free}	0.297 / 0.313	0.282 / 0.308
RMS bonds (Å)	0.011	0.012
RMS angles (deg.)	1.89	1.99
Ramachandran favored/forbidden (%)	83.5 / 3.3	83.2 / 3.3
Clashscore	29	28
Refinement PHENIX		
Resolution range (Å)	71.53 – 3.90	71.53 – 3.90
No. of reflections working/test set	37,265 / 1,967	37,236 / 1,967
Start R/R_{free}	0.289 / 0.307	0.298 / 0.320
End R/R_{free}	0.273 / 0.311	0.247 / 0.292
RMS bonds (Å)	0.005	0.006
RMS angles (deg.)	1.36	1.38
Ramachandran favored/forbidden (%)	83.5 / 3.3	83.2 / 3.3
Clashscore	29	28

*The reported R/R_{free} values are 30.0/31.2%.

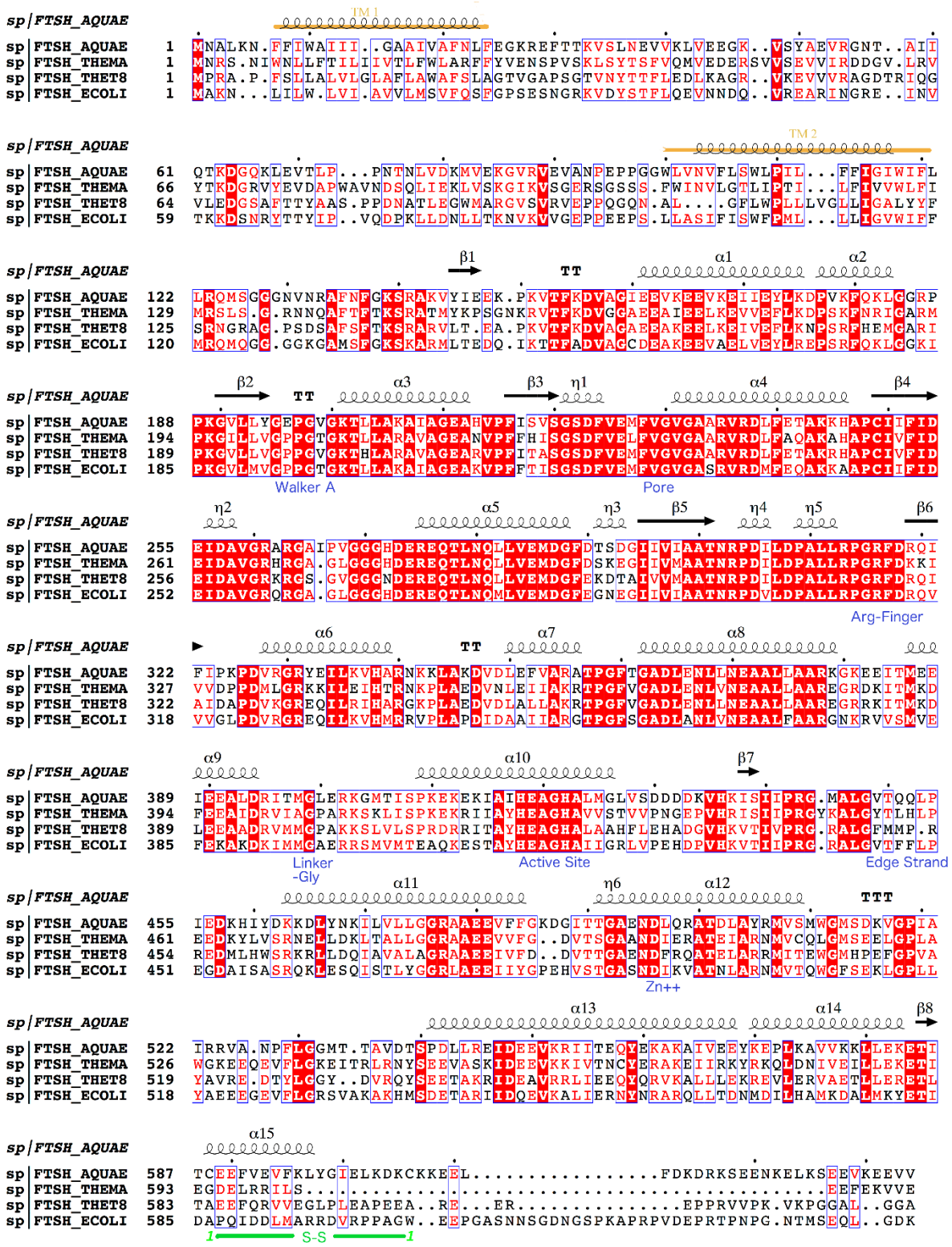

Figure S1 Sequence alignment of bacterial FtsH proteins. The sequences of FtsH from *A. aeolicus* (FTSH_AQUAE), *T. maritima* (FTSH_THEMA), *T. thermophiles* (FTSH_THET8) and *E. coli* (FTSH_COLI) are shown. Secondary structures are derived from the *A. aeolicus* structure of crystal form X2. Some key elements are annotated, e.g., 'edge' means the edge strand or active-site switch.

The green '1' denotes the cysteine residues linked by the disulfide bond. The figure was prepared with Espript (Gouet, 2003).

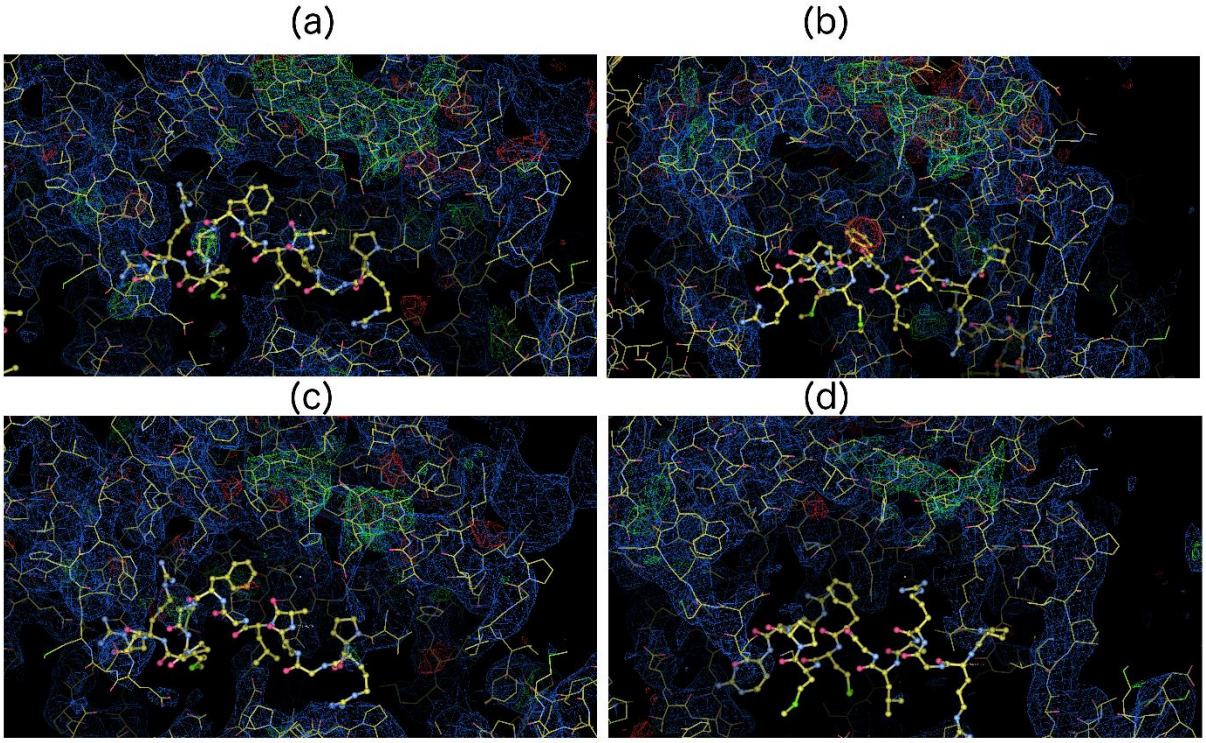

Figure S2 Simulated-annealing omit maps of the 'lid helix' in entry 4EIW. Shown are the SigmaA-weighted 2Fo-Fc (blue, 1.0 sigma contour level) and Fo-Fc electron density maps (green at +3.0 sigma and red at -3.0 sigma) of subunits A (left) and B (right). The other four subunits look very similar. Residues 442 to 454 (the lid helix, shown in ball-and-stick representation) and 417 to 428 (active site helix) were omitted from refinement and map calculation. (a) and (b): Refinement and map calculations using relabeled data, i.e. I → F. The lid helix does not show any density at these contour levels while the active site segment has strong positive difference density. (c) and (d): Refinement was carried out using the deposited data and omitting the same segments. The positive Fo-Fc density of the active-site helix is much weaker than in (a) and (b), while again the lid helix again has no density.