Acta Crystallographica Section D

Volume 70 (2014)

Supporting information for article:

Structure-specificity relationships in Abp, a GH27 β-L-arabinopyranosidase from *Geobacillus stearothermophilus*

Shifra Lansky, Rachel Salama, Hodaya V. Solomon, Hadar Feinberg, Hassan Belrhali, Yuval Shoham and Gil Shoham

S1. Crystallization experiments

Best Abp-WT crystals were obtained by the hanging-drop technique, using a protein concentration of 4-6 µg/µl and a reservoir solution of 1.8M NH_4SO_4, 0.1M citrate buffer, pH 4.8. These crystals were used for a number of X-ray diffraction data measurement sessions with synchrotron radiation (BM14, ESRF, Grenoble, France). Two of these diffraction data sets were used for the structure determination of Abp-WT, a medium resolution set of 2.90 Å and a high resolution set of 2.28 Å, both belonging to the primitive orthorhombic space group $P2_12_12_1$.

S2. Content of Abp asymmetric unit

A structural superposition of the eight independent monomeric Abp chains demonstrates an almost perfect overlap between any pair of them, with r.m.s.d. values in the range of 0.09-0.28Å for the different chain combinations, as based on 429 of the 448 Cαatoms. Thus, despite some fluctuations and conformational differences in the side chains, especially those on the surface of the protein, the eight monomeric protein chains of Abp-WT are practically identical in their overall three dimensional structure. This conclusion further validates the biological relevance of the current crystal structure, as independent protein monomers that are situated in different intermolecular environments in the crystal demonstrate very similar three-dimensional structures.

The citrate and sulfate molecules present in the final model probably originated from the ammonium sulfate salt and the sodium citrate buffer present in the crystallization solution of Abp-WT (Lansky *et al.*, 2013b), while the glycerol molecules most likely originated from the cryo-protecting solution (containing 20% glycerol) in which the crystal was soaked for about 2060 seconds prior to the data measurement (Lansky *et al.*, 2013b). Except for two specific glycerol molecules, which are present in the active site of the protein (see main text), all the glycerol molecules included in the crystallographic model appear on the surface of the protein. Although relatively many in number, these "non-physiological" molecules that are included in the crystallographic model should not affect significantly the overall conformation of the Abp protein since they are relatively small and since they occupy only those surface positions which are likely to host water molecules in biological aqueous solutions. We therefore expect that the crystallographic Abp model presented here should not be significantly different from the actual structure of the protein as present in a physiological environment, where those extra glycerol, sulfate and citrate molecules are probably absent.

S3. Preparation attempts of the Abp-Δ358-448 deletion mutant

As shown in the main text, the Abp protein monomer is built of two distinct domains, an N-terminal and a C-terminal. Since the function of the C-terminal domain is yet unclear, an attempt was made to remove it in order to test the resulting protein in terms of activity and specificity. For this purpose, the *abp* gene without the all-βC-terminal domain (***Abp*-Δ*358-448***) was cloned into pET9d *via* PCR. The primers used for the cloning were as follows:
5'-CTCAAATCCATGGCTCATCATCATCATCATCATTCCTCTTATTCTAATAAGG -3' and
5'-ATATATGGATCCTCAATTTTGATGAACGTGTAATACTTC -3'. Unfortunately, all our efforts to produce a soluble protein of this deletion mutant were so far unsuccessful, as all of the protein appeared in the form of inclusion bodies. Additional attempts to obtain this and related deletion mutants of Abp are currently underway.

S4. SAXS analysis of the tetramer

SAXS data were measured for Abp-WT, confirming it is a tetramer formed of chains A,B,C,D (Figure 1, S2). This result was further validated by examining all alternative tetrameric combinations possible among the eight monomers constituting the crystallographic asymmetric unit. Of these alternatives, only the tetramer built of chains ABEF seemed to deserve further checking. A comparison with CRYSOL was therefore performed for the tetramer composed of chains ABEF, similar to the calculation made for the ABCD tetramer (see main text). However, the fit to the SAXS data obtained for the ABEF tetramer was significantly worse compared to the corresponding ABCD tetramer (**Figure S2c**), further confirming that the tetramer assembly that best represents the Abp assembly in solution is that of the ABCD crystallographic tetramer. The other relevant tetramer in the asymmetric unit is therefore built of the EFGH monomers.

S5. Abp interfaces

Interface I (*within* the dimers) involves an averaged buried interface area of 788.5Å^2, an averaged $\Delta^i G$ of -8.5 kcal/mol and a $\Delta^i G$ P-value of 0.117 (all calculated by the PISA server, Krissinel & Henrick, 2007). **Interface II** (*between* the dimers) similarly involves an averaged buried interface area of 767.4Å^2, an averaged $\Delta^i G$ of -3.6 kcal/mol and a $\Delta^i G$ P-value of 0.397. This difference in energy gain suggests that the "basic dimers" of the tetramer (and probably those formed first) are indeed the ones built of chains A&C and B&D, and not the ones formed by chains A&B and C&D.

S6. Comparison of active site of Abp with 3CC1

Looking at the residues involved in substrate binding at subsite -1 (see section *3.3.* in the main text), Asp197, Asp255, Trp32, Asp66, Ile67, His126 and Lys195 of Abp are situated in similar positions to the

corresponding residues in 3CC1, which are Asp183, Asp241, Trp18, Asp52, Ile53, His112 and Lys181, respectively. This comparison also confirms that in 3CC1 the catalytic residues are Asp183 and Asp241, where Asp183 serves as the nucleophile and Asp241 serves as the acid/base.

Sequence alignment between Abp and 3CC1 shows that the putative residues involved in the +1 subsite in Abp, Asp254 and Trp161, are analogous to residues Asp240 and Trp147 in 3CC1. The two Trp residues are situated in very similar positions in both proteins, but the positions of the two Asp residues differ significantly. This is not very surprising, however, considering the semi-conservation of Asp254 (see section 3.3. in the main text), as observed among GH27 proteins.

Most of the residues involved in the active sites are identical for both proteins, except for residues Ser231, Pro232, His126 and Ile67 in Abp, which are replaced by Cys224, Trp226, Tyr140 and Glu99 in SaAraP27A, respectively. Nevertheless, these residues are situated in similar positions of the active site, and seem to bind to similar functional groups of a bound arabinose molecule at the -1 subsite, as seen for the complexes of the two enzymes with an L-arabinose product (**Figure 9c**). Another similarity between the two proteins is that in both structures a glycerol molecule was captured in the putative +1 binding subsite. However, the specific residues involved in the binding of this glycerol molecule differ for the two proteins. In Abp these residues include Asp254, Asp255 and Trp161, while in SaAraP27A these residues include Tyr152, Trp226 and Tyr250, residues which are non-conserved among GH27 enzymes (Ichinose *et al.*, 2009). Assuming that the +1 subsite has been identified correctly, these differences raise the question of why subsite -1 is relatively conserved, while subsite +1 is not. The answer may relate to slightly different natural substrates, since although both enzymes remove the same β-L-arabinose moieties from arabinose-containing natural oligosaccharides, the other parts of the substrates may be different, in a similar manner to the case of 3CC1, as discussed in the main text.

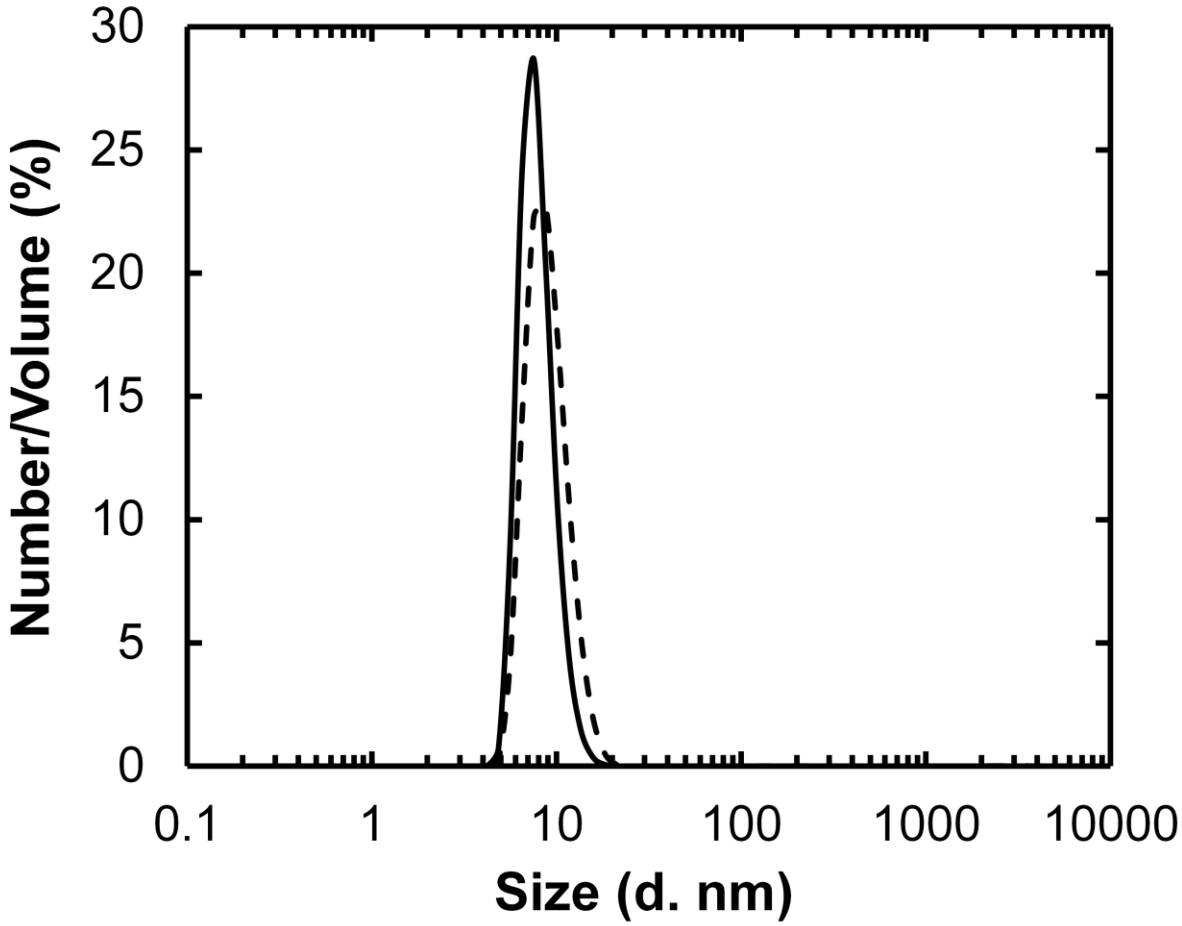
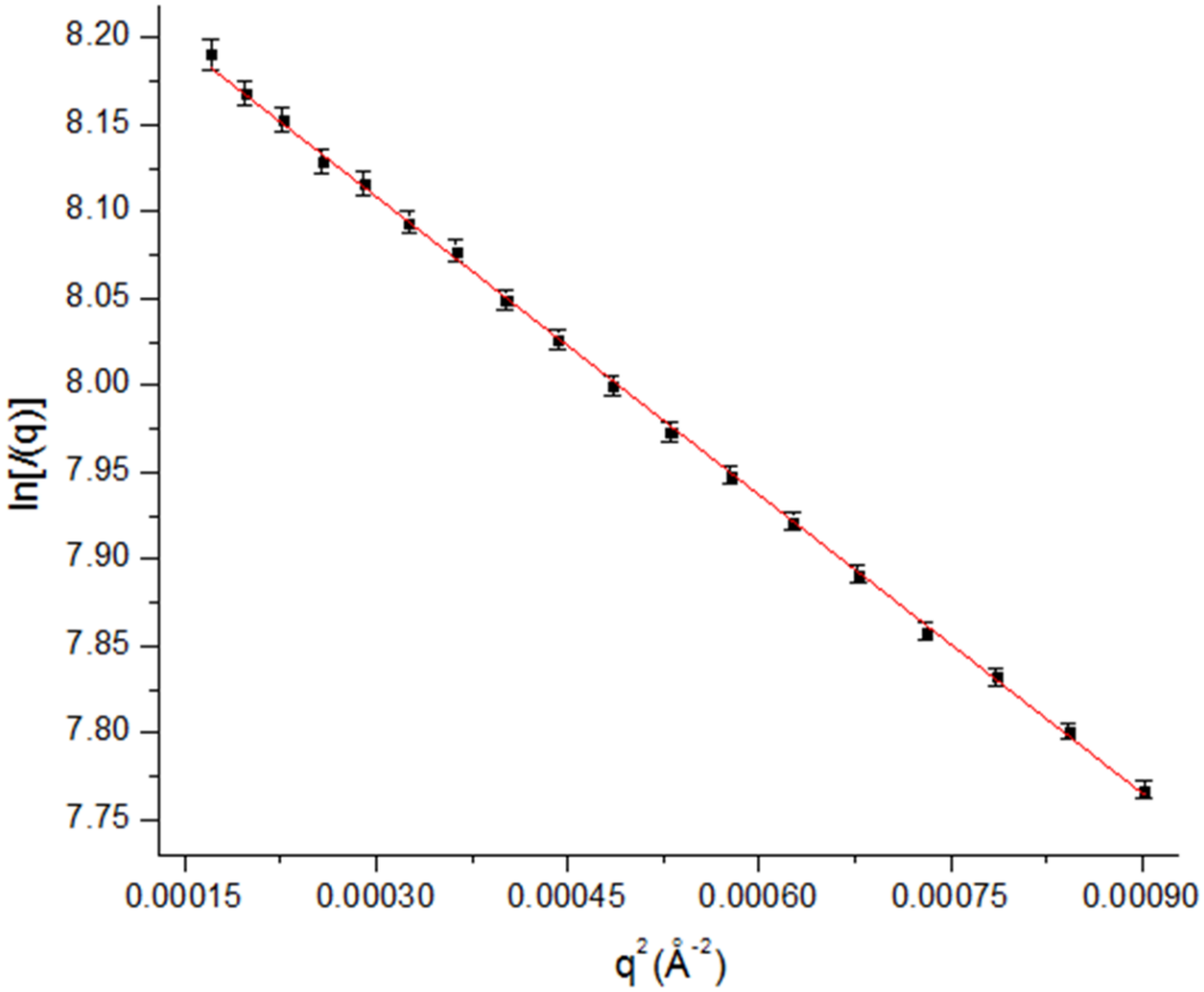
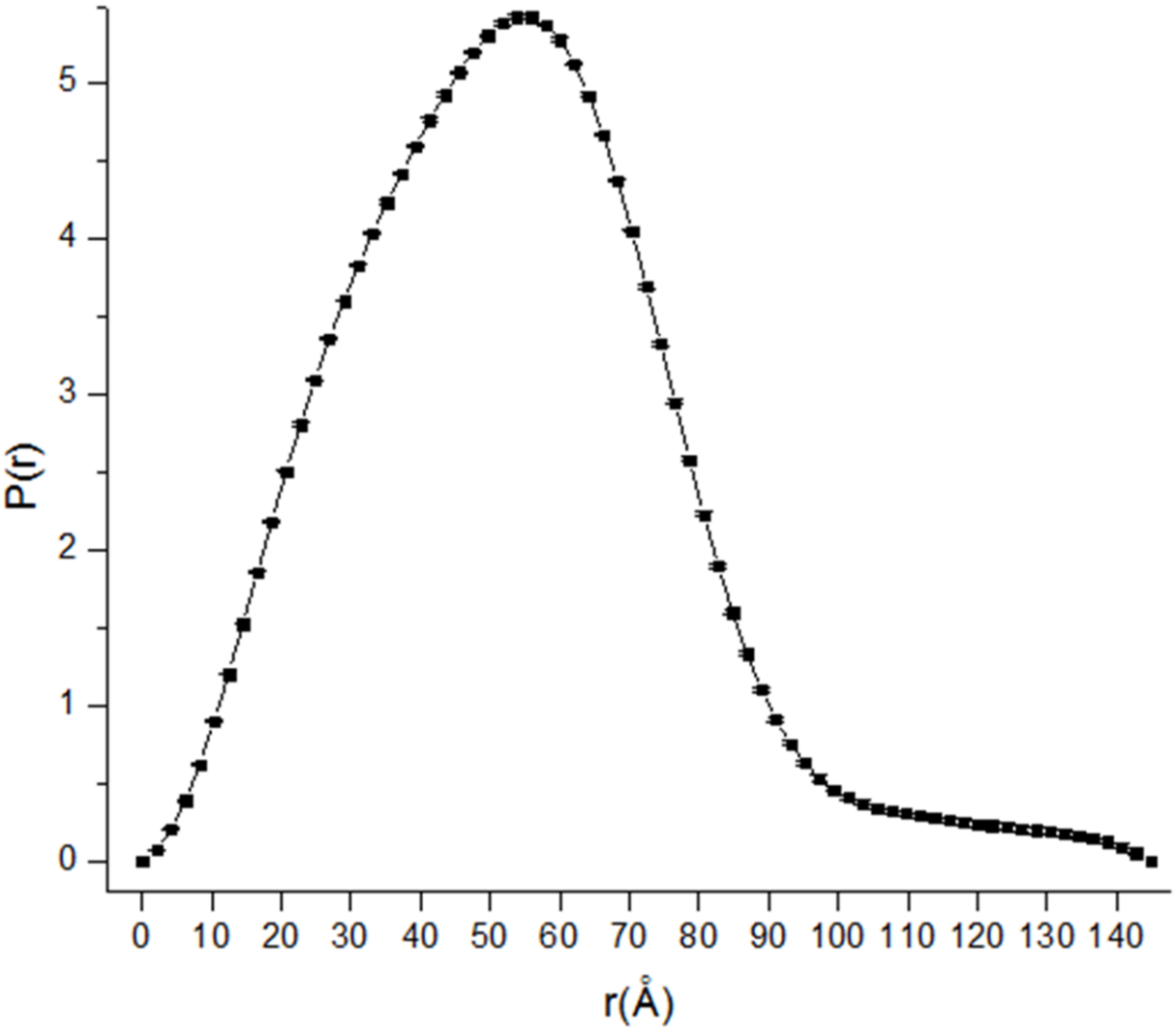


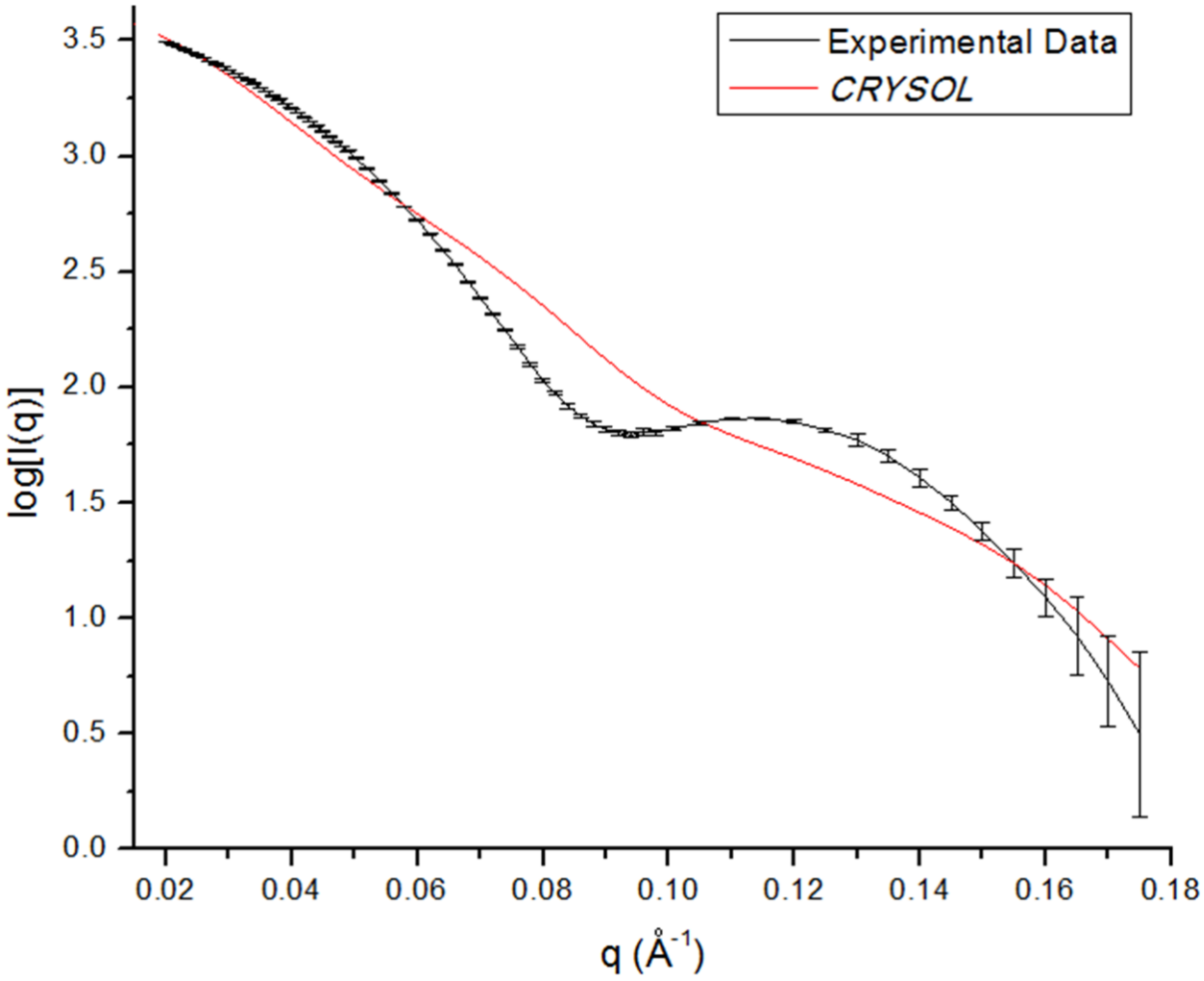
Figure S1 Dynamic light-scattering (DLS) curves of the hydrodynamic diameter distributions (in nm) of the Abp-WT protein in solution, as estimated by number (solid line) and by volume (broken line).


8.20
8.15
8.10
8.05
8.00
7.95
7.90
7.85
7.80
7.75
ln[I(q)]
0.00015
0.00030
0.00045
0.00060
0.00075
0.00090
q² (Å⁻²)

(a)


P(r)
0
1
2
3
4
5
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
r(Å)

(b)



(c)

Figure S2 **(a).** Guinier plot for Abp-WT, showing a good linear fit (adj. R square=0.9995) within the range of qRg=0.58-1.28, indicating no aggregation of the protein sample. (**b).** Pair distance distribution function P(r) for Abp-WT, D_{max}=145Å. (**c).** An attempt to fit the experimental SAXS results with the "wrong" tetramer of Abp composed of chains ABEF. The scattering curve for the Abp-WT protein is shown in black, while the simulated scattering curve from the tetramer composed of chains ABEF, calculated by CRYSOL, is shown in red. The corresponding χ value is 25.7.

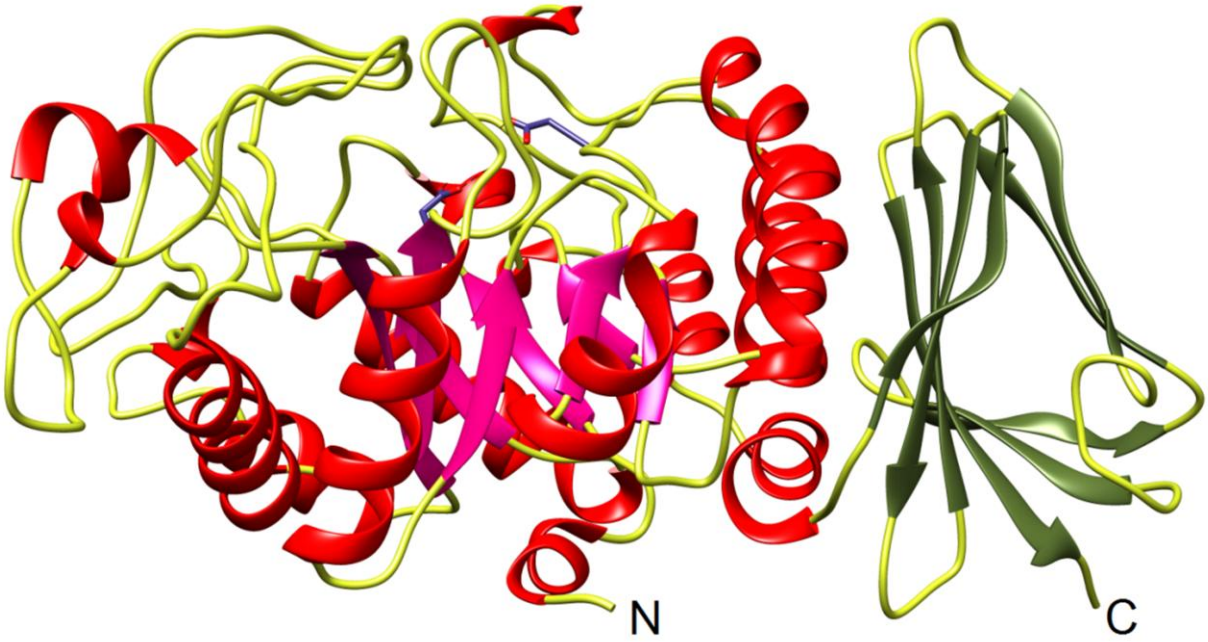


Figure S3 Abp monomer. A view perpendicular to Figure 2, demonstrating the two domains of the protein and the short polypeptide "hinge" that connects them.

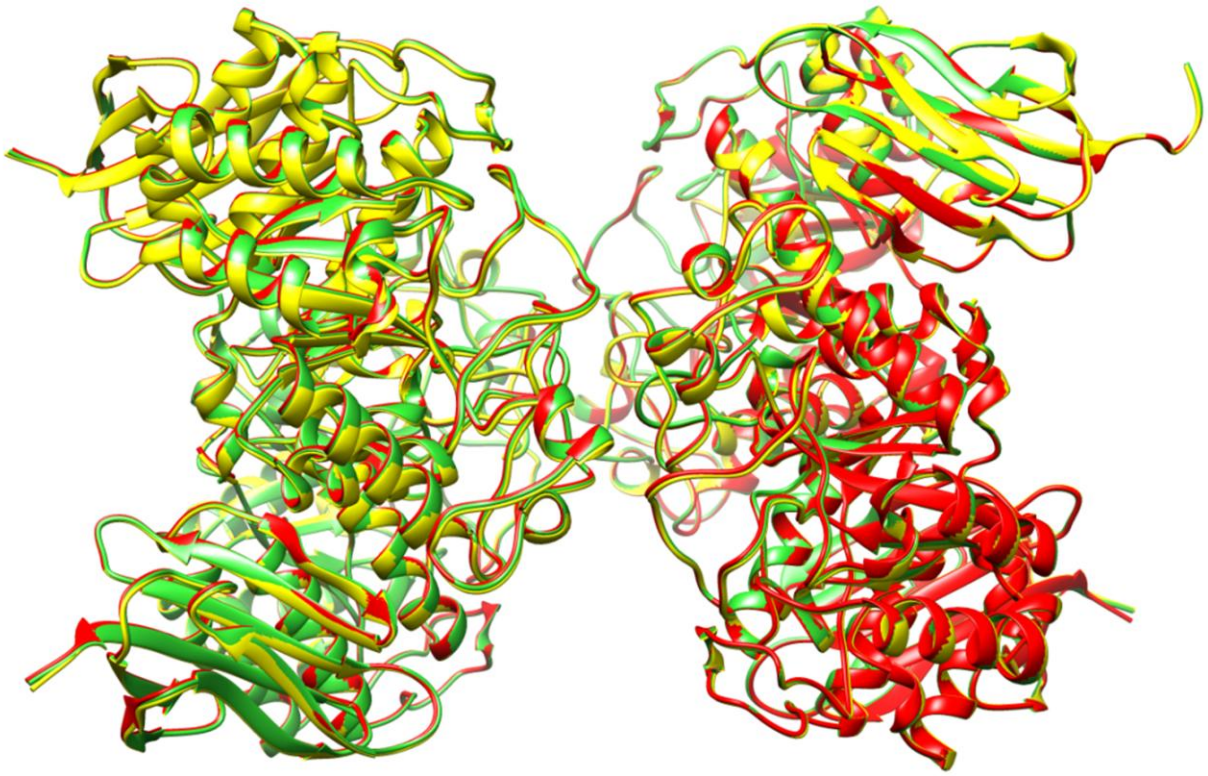


Figure S4 A superposition of the Abp crystallographic tetramers obtained for Abp-WT (green), Abp-D197A (yellow) and Abp-D197A-ARB (red), demonstrating a very good fit.

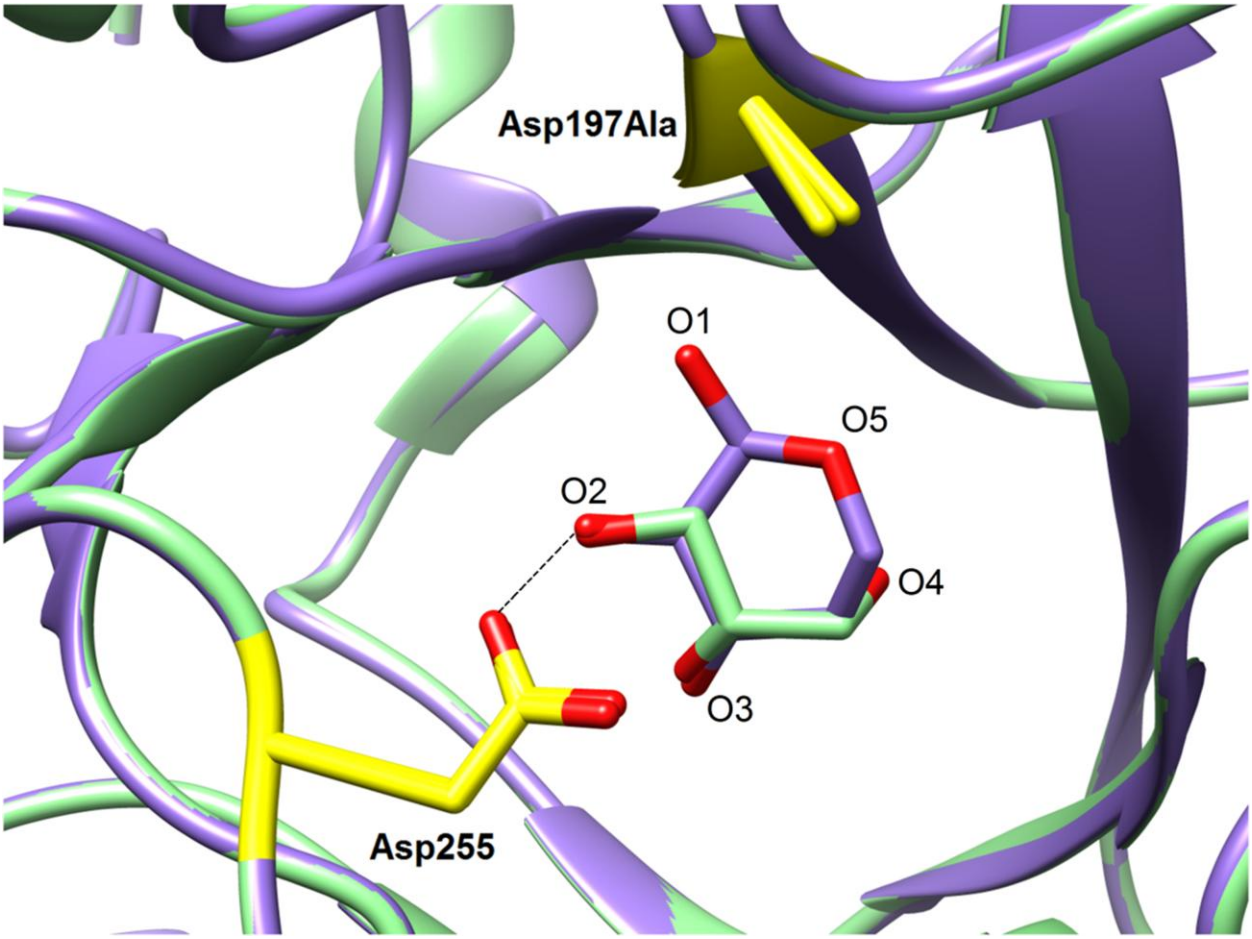


Figure S5 Superposition of the active site of Abp-D197A (in green) on the active site of Abp-D197A-ARB (in purple), showing the respective glycerol and arabinose molecules trapped in subsite -1 of the active site. The relatively good overlap between these bound molecules suggests that the glycerol molecule mimics the binding of half the arabinose molecule in the case of Abp-D197A and Abp-WT. Such mimicking is specifically relevant for oxygens O2, O3, O4 and carbons C2, C3, C4 of the β-L-arabinose. The catalytic residues are shown in yellow.

Figure S6 Relative conservation of the Abp amino-acid sequence, as calculated by the Consurf server (Ashkenazy *et al.*, 2010).