STRUCTURAL BIOLOGY COMMUNICATIONS

Volume 72 (2016)

Supporting information for article:

Characterization and 1.57 Å resolution structure of the key fire blight phosphatase AmsI from *Erwinia amylovora*

Marco Salomone-Stagni, Francesco Musiani and Stefano Benini

S1. Supplementary data on AmsI structure analysis

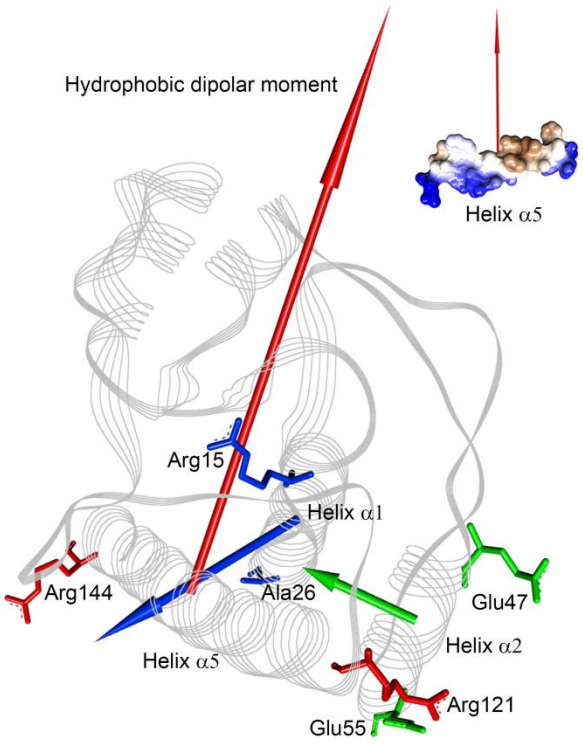

Figure S1 Wire-ribbon representation of AmsI structure with superposed the hydrophobic moments of the helices α_1 (blue arrow), α_2 (green arrow) and α_5 (red arrow). Following the same colour code of the arrows, the residues starting and ending each helix are shown. The figure was generated with the program Discovery Studio Visualizer 4.0, Accelrys Software Inc.

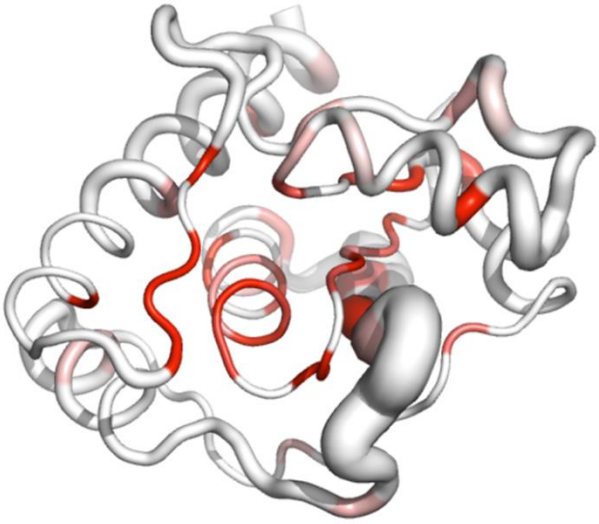

Figure S2 3D representation of the structural conservation of the low molecular weight protein tyrosine phosphatases presented in the main text alignment of Figure 1. The thicker portions refer to a higher positional variability in the space. The red portion identifies the most conserved residues. Figure obtained using the server EndScript 2.0.

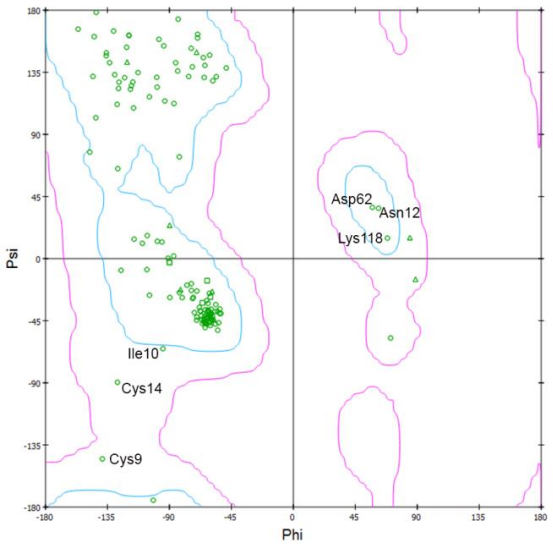

Figure S3 **S5**. Ramachandran plot of the AmsI structure. In evidence some outliers and left handed conformation residues. The figure was generated with the program Discovery Studio Visualizer 4.0, Accelrys Software Inc.

S1.1. SLS-QELS His tagged and cleaved AmsI

AmsI resulted entirely a monomer at the tested concentration. Sulfate ion is not promoting dimerization in solution. His tagged AmsI is a monomer in solution as well.

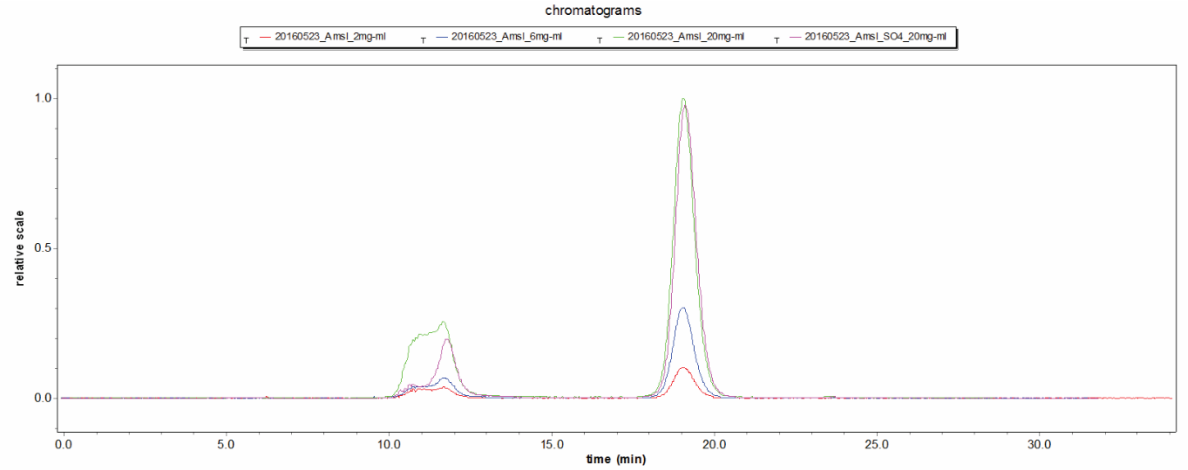

Figure S4 SEC-QELS chromatograms. SEC-QELS chromatogram comparison of AmsI at 2, 6, 20 mg/ml and of AmsI 20 mg/ml in the presence of 5mM SO_4.

Table S1 SEC-QELS results

POLYDISPERSITY	
MW/MN	1.006(0.7%)
MZ/MN	1.011(1%)
MOLAR MASS MOMENTS (G/MOL)	
MN	1.692e+4(0.5%)
MW	1.702e+4(0.5%)
MZ	1.711e+4(1%)
M(AVG)	1.691e+4(0.0%)
RMS RADIUS MOMENTS (NM)	
RN	4.5(128%)
RW	4.6(122%)
RZ	4.7(115%)
R(AVG)	n/a
HYDRODYNAMIC RADIUS MOMENTS (NM)	
RH(N)	2.0(4%)
RH(W)	2.0(4%)
RH(Z)	2.0(4%)
RH(AVG)	1.8(0.6%)
TRANSLATIONAL DIFFUSION MOMENTS (CM^2/SEC)	
DT(N)	1.33e-6(3%)
DT(W)	1.33e-6(3%)
DT(Z)	1.33e-6(3%)
DT(AVG)	1.34e-6(0.6%)

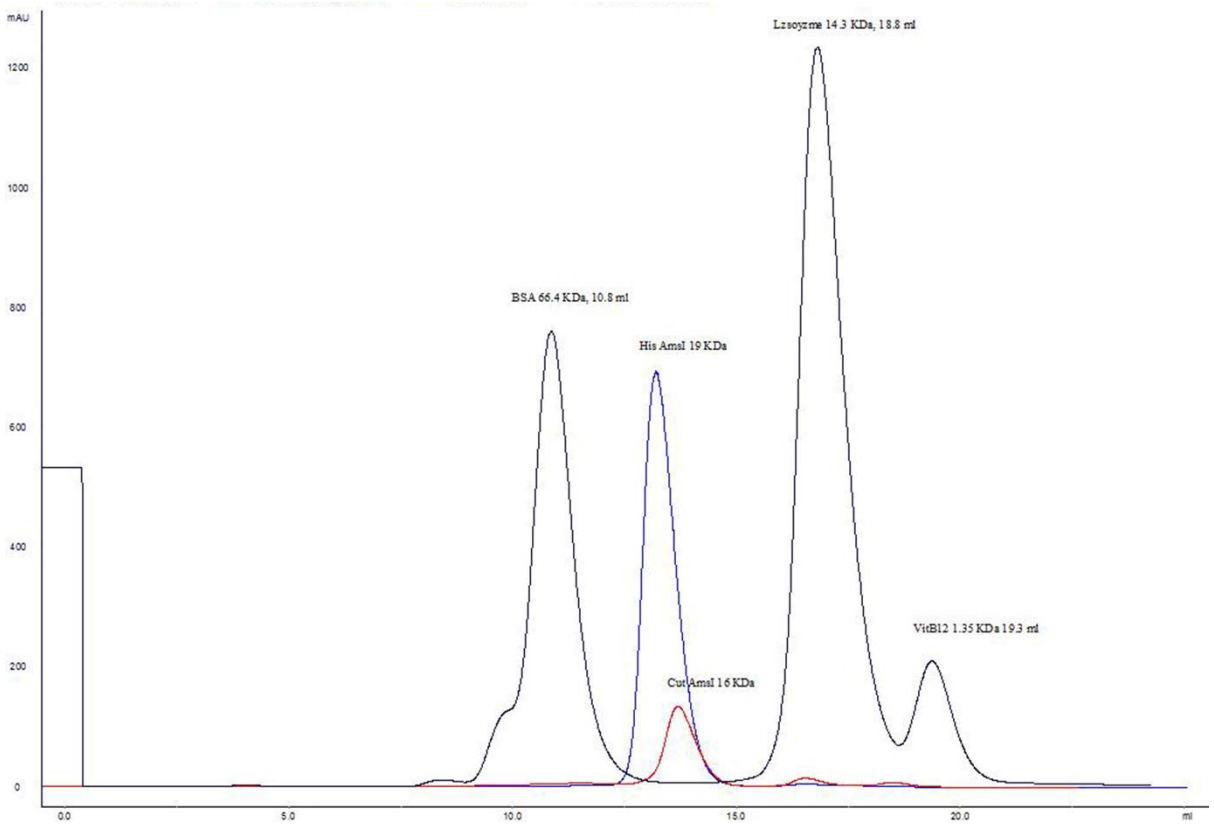

Figure S5 Chromatograms comparison between His tagged and cleaved AmsI.

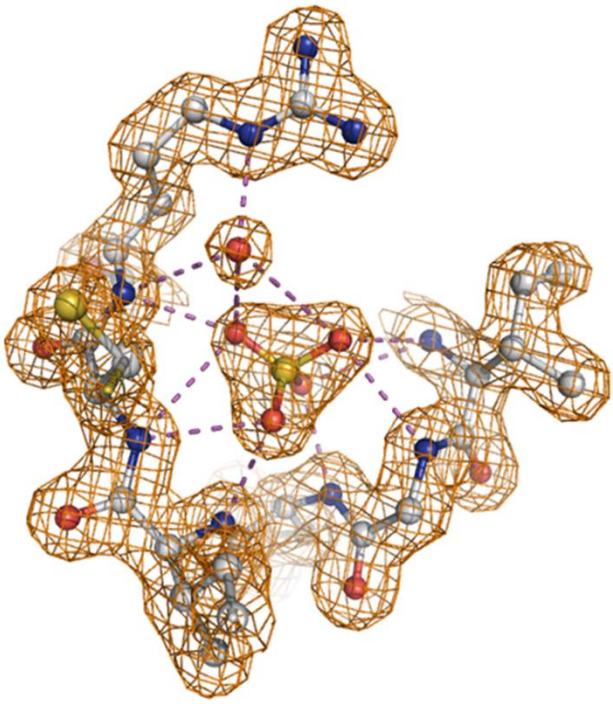

Figure S6 Ball-and-stick representation of AmsI's active site. The sulfate, the water molecule and the P-loop residues involved in a hydrogen bond network are coloured according to the elements. In orange is represented the electron density at a contour level of 1.5. The figure was generated with the program PyMol v0.99.

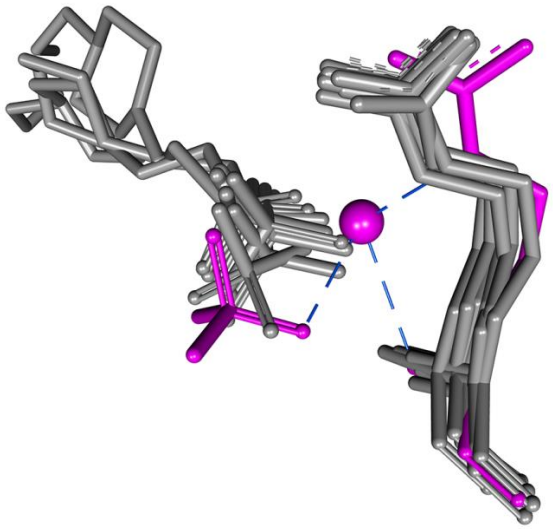

Figure S7 Superposition of the catalytic arginines and ligands (when present) of the LMW-PTP presented in the alignment of Figure 1, main text. The ligands were phosphate, or sulfate, or molecules with a sulfonic group, *i.e.* MES and HEPES. AmsI is represented in magenta. It is evident that the presence of the water splits apart the arginine and the ligand in respect to the other structures. The figure was generated with the program Discovery Studio Visualizer 4.0, Accelrys Software Inc.

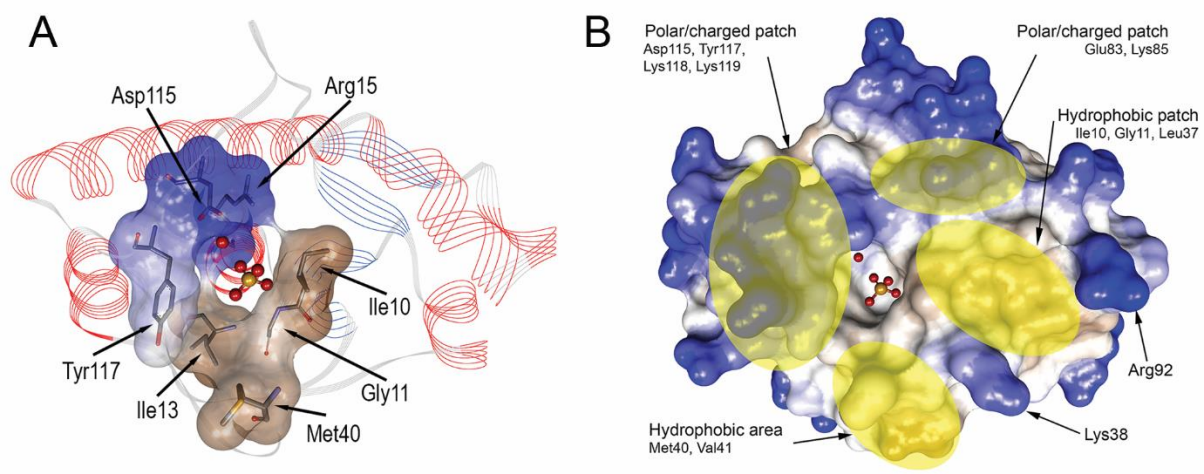

Figure S8 Solid surface representation (probe radius 1.4 Å) of the entrance into the active site of AmsI. The surface is coloured according to a hydrophobicity scale. From blue, to white, to brownish, represent polar to hydrophobic moieties. For clarity, the sulfate and the water in the active site are shown as ball-and-stick. **A**) The residues around the active site are also represented as sticks and coloured as follows: carbon, grey; oxygen, red; blue, nitrogen, yellow, sulphur. The rest of the protein is represented in wired ribbons and coloured according to the secondary structure (red, α-helices; light blue, β-strands). **B**) The patches identifying the hydrophobic and polar areas around the active site are highlighted with yellow ellipsoids and the involved residues specified. The figures were generated with the program Discovery Studio Visualizer 4.0, Accelrys Software Inc.

S2. Homology modelling of AmsA and experimental driven docking AmsI-AmsA

The model complex of the interaction between AmsI and the cognate kinase AmsA was calculated by the use of bioinformatics.

S2.1. Materials and methods

S2.1.1. Sequence alignments and AmsA homology modelling

The alignment between the sequence of *E. amylovora* AmsI and the homologue protein Wzb from *E. coli* strain K12, as well as the alignment of *E. amylovora* AmsA and Wzc from *E. coli* strain K12, were produced using the Promals3D server (see Fig. S8 and S9) (Pei *et al.*, 2008).

```
             1        10         20         30         40         50         60
             |        |          |          |          |          |          |
EcWzb      1 MFNNILVVCVGNICRSPTAERLLQRYHPELKVESAGLGALVGKGADPTAISVAAEHQLSL    60
EaAmsI     1 MVNSILVVCIGNICRSPTGERLLKAALPERKIASAGLKAMVGGSADETASIVANEHGVSL    60

            61        70         80         90         100        110        120
             |        |          |          |          |          |          |
EcWzb     61 EGHCARQISRRLCRNYDLILTMEKRHIERLCEMAPEMRGKVMLFGHWDNECEIPDPYRKS  120
EaAmsI    61 QDHVAQQLTADMCRDSDLILVMEKKHIDLVCRINPSVRGKTMLFGHWINQQEIADPYKKS  120

            121       130        140    147
             |        |          |      |
EcWzb    121 RETFAAVYTLLERSARQWAQALNAEQV  147
EaAmsI   121 RDAFEAVYGVLENAAQKWVNALSR---  144
```

Figure S9 *E. amylovora* AmsI and *E. coli* Wzb sequence alignment.

```
             450       460       470       480       490          500  505
              |         |         |         |         |            |    |
ChainE   449 FNRGIESPQVLEEHGISVYASIPLSEWQKARDSVKTIKGIKR---YKQSQLLAVGNPTDL  505
ChainG   449 FNRGIESPQVLEEHGISVYASIPLSEWQKARDSVKTIKGIKR---YKQSQLLAVGNPTDL  505
ChainL   449 FNRGIESPQVLEEHGISVYASIPLSEWQKARDSVKTIKGIKR---YKQSQLLAVGNPTDL  505
EaAmsA   449 FHHGIDNPEQLEELGLNVYASVPLSEWQRKKDQETLLKRKLDARTDPHNRLLALGNPTDL  508
              |         |         |         |         |         |      |
             450       460       470       480       490       500     508

            506 510       520       530       540       550       560  565
              |   |         |         |         |         |         |    |
ChainE   506 AIEAIRSLRTSLHFAMMQAQNNVLMMTGVSPSIGMTFVCANLAAVISQTNKRVLLIDCDM  565
ChainG   506 AIEAIRSLRTSLHFAMMQAQNNVLMMTGVSPSIGMTFVCANLAAVISQTNKRVLLIDCDM  565
ChainL   506 AIEAIRSLRTSLHFAMMQAQNNVLMMTGVSPSIGMTFVCANLAAVISQTNKRVLLIDCDM  565
EaAmsA   509 SIEAIRSLRTSLHFAMMDAQNNILMITGASPGIGKTFVCANLATLVAKTGEKVLLIDCDM  568
            / |         |         |         |         |         |       |
         509 510       520       530       540       550       560     568

            566 570       580       590       600       610       620  625
              |   |         |         |         |         |         |    |
ChainE   566 RKGYTHELLGTNNVNGLSEILIGQGDITTAAKPTSIAKFDLIPRGQVPPNPSELLMSERF  625
ChainG   566 RKGYTHELLGTNNVNGLSEILIGQGDITTAAKPTSIAKFDLIPRGQVPPNPSELLMSERF  625
ChainL   566 RKGYTHELLGTNNVNGLSEILIGQGDITTAAKPTSIAKFDLIPRGQVPPNPSELLMSERF  625
EaAmsA   569 RKGYTHELLGTNNVNGLSEILIGQGDITTAAKPTSIAKFDLIPRGQVPPNPSELLMSERF  628
            / |         |         |         |         |         |       |
         569 570       580       590       600       610       620     628

            626 530       640       650       660       670       680  685
              |   |         |         |         |         |         |    |
ChainE   626 AELVNWASKNYDLVLIDTPPILAVTDAAIVGRHVGTTLMVARYAVNTLKEVETSLSRFEQ  685
ChainG   626 AELVNWASKNYDLVLIDTPPILAVTDAAIVGRHVGTTLMVARYAVNTLKEVETSLSRFEQ  685
ChainL   626 AELVNWASKNYDLVLIDTPPILAVTDAAIVGRHVGTTLMVARYAVNTLKEVETSLSRFEQ  685
EaAmsA   629 AELVNWASKNYDLVLIDTPPILAVTDAAIVGRHVGTTLMVARYAVNTLKEVETSLSRFEQ  688
            / |         |         |         |         |         |       |
         629 630       640       650       660       670       680     688

            686 690       700         710       720
              |   |         |           |         |
ChainE   686 NGIPVKGVILNSIFRRASAYQDYG--YYEYEYKSDAK-  720
ChainG   686 NGIPVKGVILNSIFRRASAYQDYG--YYEYEYKSDAK-  720
ChainL   686 NGIPVKGVILNSIFRRASAYQDYG--YYEYEYKSDAK-  720
EaAmsA   689 NGIPVKGVILNSVVRKSANNYGYGYDYYDYSYQQGEKS  726
            / |         |         |         |     |
         689 690       700       710       720   726
```

Figure S10 Alignment among the cytosolic catalytic domains of *E. amylovora* AmsA and of *E. coli* Wzc chain E, G, L present in the PDB file 3LA6.

The AmsA/Wzc alignment was then used to calculate 100 structural models of the AmsA cytoplasmic catalytic domain using the Wzc crystal structure as template (PDB: 3LA6) (Bechet *et al.*, 2010) and the Modeller 9.14 software (Marti-Renom *et al.*, 2000). The crystal structure of Wzc is a homooctamer. All the Wzc monomers were used in the modelling. The best model was selected on the basis of the lowest value of the DOPE score in Modeller (Shen & Sali, 2006). The stereochemical quality of the model structure was established using ProCheck (Laskowski *et al.*, 1993). The results of this analysis (most favoured and additionally allowed residues in the Ramachandran plot 92.2% and 7.8%, respectively; overall average G-score: -0.06) confirm the reliability of the model structure.

S2.1.2. Data driven protein-protein docking

The AmsI structure was docked onto the AmsA structural model of the cytoplasmic catalytic domain using the data-driven docking program Haddock 2.2 as implemented in the Haddock Webserver (de Vries *et al.*, 2010). HADDOCK (High Ambiguity Driven biomolecular DOCKing) implements an approach that uses biochemical and/or biophysical interactions data to drive the docking process. The putative interacting residues of AmsI and AmsA were selected by considering the data on interacting residues in the homologue Wzb-Wzc complex. In particular, we used both the residues positions published by the group of Prof. Ghose in ref. (Temel *et al.*, 2013), and new partial data from the same group to identify putative interaction patches on the surface of the two proteins. The calculation was guided by defining the former residues as "active" in the protein-protein interaction. The docking algorithm rewards the complexes that have these active residues on the interaction interface. The solvent accessible residues found on the surface of the proteins and in contact with active residues were included in the calculation as "passive" residues. See table S2 for a complete list of the active and passive residues used in the docking calculation.

Table S2 Active and passive residues used to guide the docking calculation.

Protein	Active residues	Passive residues
AmsI	9, 10, 11, 12, 13, 15, 36, 37, 38, and 83.	14, 16, 39, 40, 41, 44, 46, 67, 84, 85, 86, 89, 93, 112, 113, 114, 115, and 117.
AmsA	457, 460, 461, 464, 465, 466, 467, 468, 508, 511, 512, 514, 515, 516, 517, 518, 519, 521, 522, 523, 525, 526, 527, 528, 530, 554, 649, 650, 677, 678, 681, 683, 685, 687, and 694.	449, 450, 455, 458, 462, 463, 469, 471, 501, 502, 505, 506, 529, 538, 539, 556, 557, 558, 559, 560, 632, 636, 639, 640, 641, 651, 652, 653, 663, 664, 665, 673, 675, 676, 680, 684, 688, 690, 692, 693, 695, 701, 709, 710, 712, 713, 714, 715, 716, 717, 718, 719, 720, and 721.

Haddock simulations are composed of three rounds: i) a rigid body energy minimization that produces 1000 putative docking complexes, ii) a semi-flexible simulated annealing carried out on the 200 best solutions calculated in the first round and found on the basis of the intermolecular energy, and iii) an explicit water refinement carried out on the same structures. The solutions were clustered using a root mean square deviation (RMSD) cut-off of 7.5 Å based on the pairwise backbone RMSD matrix, and further analysed for structural and functional congruence as described below. The best cluster of structures was identified on the basis of the HADDOCK Z-score, *i.e.* how many standard deviations from the average this cluster is located in terms of score. The best model was selected on the basis of the HADDOCK score among those in the best cluster of structures. The obtained molecular complex, as well as the molecular surfaces, were displayed and analysed using the DS Visualizer 4.0 (Accelrys Software, Inc.) and UCSF Chimera software (Yang *et al.*, 2012).

S2.2. Homology modelling and experimental driven molecular docking results

A reliable model of the cytosolic catalytic domain of AmsA was produced by homology modelling (Fig. S10).

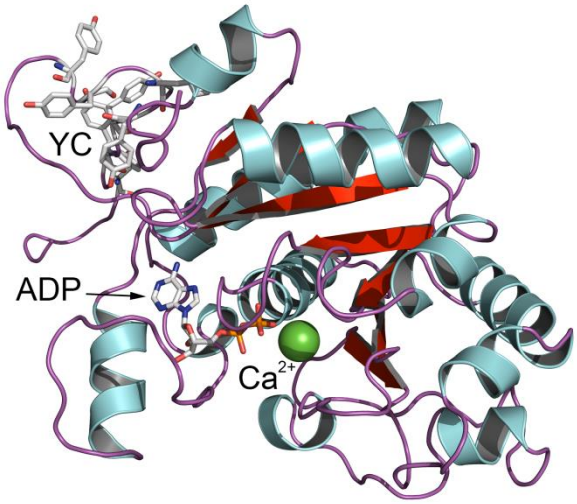

Figure S11 Cytosolic catalytic domain of AmsA. The protein is represented in cartoon representation (α-helices, light blue; β-strands, red; loops, purple). ADP, Ca^{+} and the tyrosines of the YC tail are represented as sticks and coloured as follows: carbon, grey; oxygen, red; blue, nitrogen. The image was generated with PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.

The AmsA model structure has a fold typical of BY-kinases composed by a central 5-strands parallel β-sheet surrounded by eleven α-helices. The AmsI structure and the AmsA model were then used to

perform experimental driven protein-protein docking based on the NMR data obtained by the group of Prof. Ghose in Temel *et al.* [44], and on new partial data from the same group. We calculated 200 model complexes that were grouped according to the RMSD in twelve clusters of structure. We then selected the cluster featuring the best HADDOCK Z-score, which also showed to have the highest biological significance (Fig. S11).

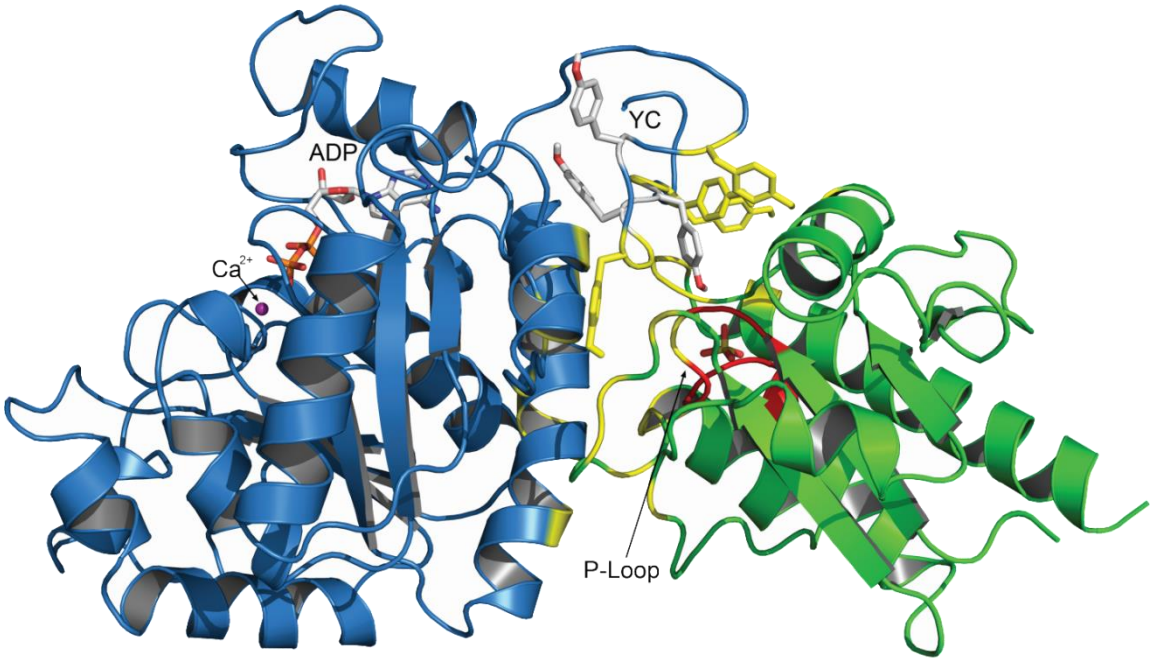

Figure S12 Results of the molecular docking calculations. The AmsI-AmsA molecular complex is shown as cartoon representation: AmsI, green; AmsA, blue. The position of the residues involved in the interaction is highlighted in yellow. AmsI P-loop is highlighted in red. Ligands and tyrosines within the YC are shown as stick representation and coloured as follows: carbon, grey; sulphur, yellow; calcium, green; oxygen, red; nitrogen, sky blue; hydrogen, white (the YC tyrosines involved in the interaction are colored in yellow). The image was generated with PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.

S2.3. Discussion about AmsI-AmsA molecular docking

The experimental driven molecular docking is based on the previously studied *E. coli* Wzc-Wzb system (Bechet *et al.*, 2010,Grangeasse *et al.*, 2003,Hagelueken *et al.*, 2009,Temel *et al.*, 2013). The results of Temel *et al.* (Temel *et al.*, 2013) suggested that the dephosphorylation is proximity mediated and that Wzb interaction determinants are present outside the active site. Our results are consistent with these observations. In fact, AmsA C-terminal YC is involved in the association with AmsI mainly through alkyl-Pi interactions supplied by the tail tyrosines and it is on a side of the docking surface, thus leaving the YC tail capable of rearrangements that would allow a proximity mediated dephosphorylation. Most of the docking energy is given by the N-terminal part of AmsA

catalytic domain where the bacterial kinases (BY) show the highest variability. In AmsI most of the amino acids involved are also not conserved. The residues that give the highest contribution to the interaction are Ser44, Glu47, Lys118 and Lys38. The latter protrudes from a hydrophobic patch (see Fig. S7B) and seems to work as a clasp during binding. Hence, common strategies would be used by the LMW-PTPs to regulate substrate selectivity. Specific positions on the interacting surface show high variability and determine the selectivity towards the substrate with no alteration of the common folding and architecture of the active site.