Supplemental Material

Molecular basis of unexpected specificity of ABC transporter-associated substrate-binding protein DppA from *Helicobacter pylori*

Mohammad M. Rahman, Mayra A. Machuca, Mohammad F. Khan, Christopher K. Barlow, Ralf B. Schittenhelm and Anna Roujeinikova

Supplemental Figures

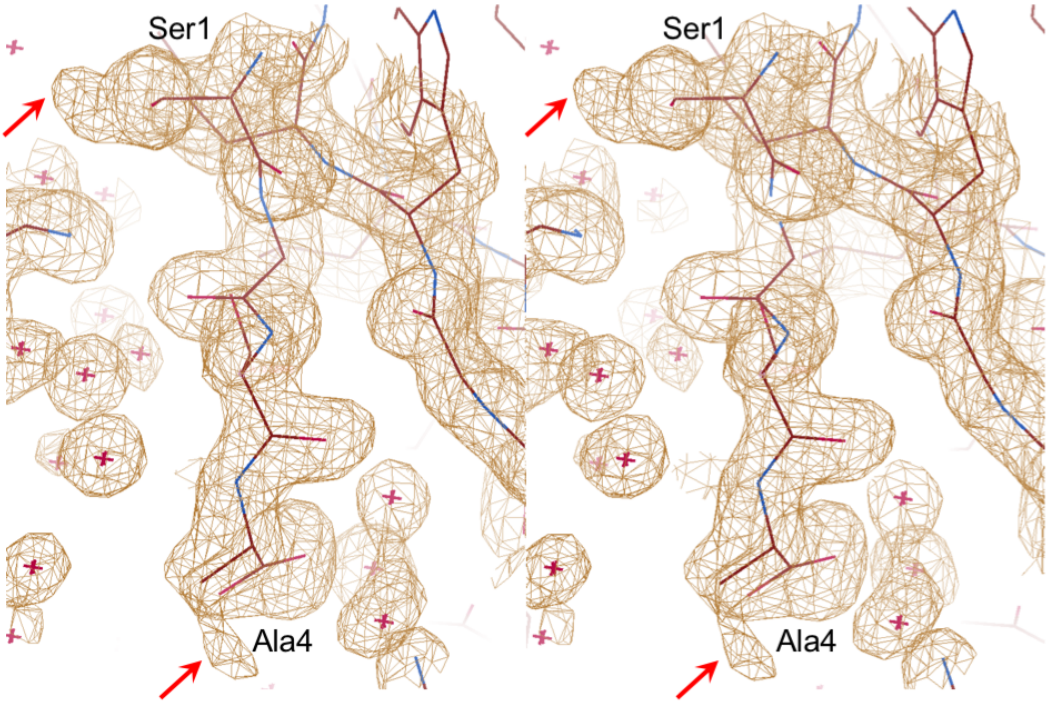

Figure S1. Stereoview of the $2mF_o$-DF_c composite omit map showing electron density in the ligand-binding pocket of HpDppA crystallized with no exogenous peptides added. The map is contoured at 1σ. The co-purified ligand was modelled as a peptide with amino acid sequence STSA. However, it is apparent that while the density for the backbone of the peptide is clearly defined, the density for the side chains is ambiguous. The arrows point to the regions that are not fully consistent with the STSA sequence.

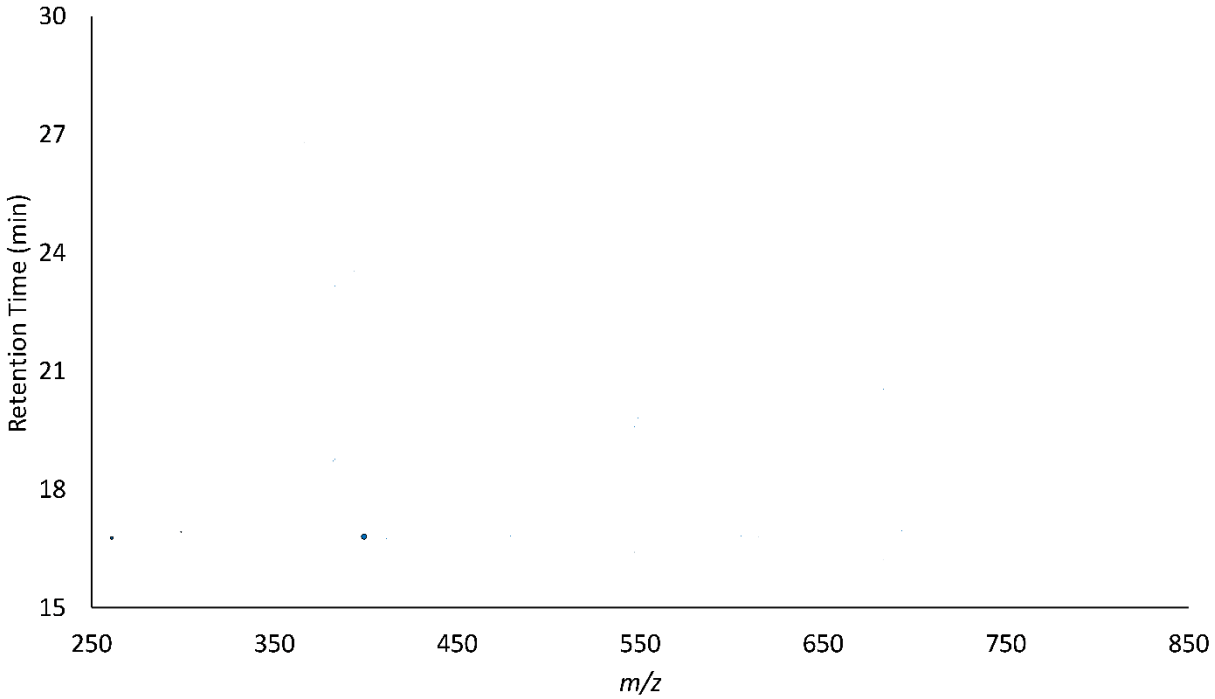

Figure S2. An overlay of the unique chromatographic features identified in the YckK sample (compared with the HpDppA sample) by untargeted feature extraction. Note that this figure is prepared in the same style and to the same bubble scale as Fig. 4A in the main text so that they may be directly compared.

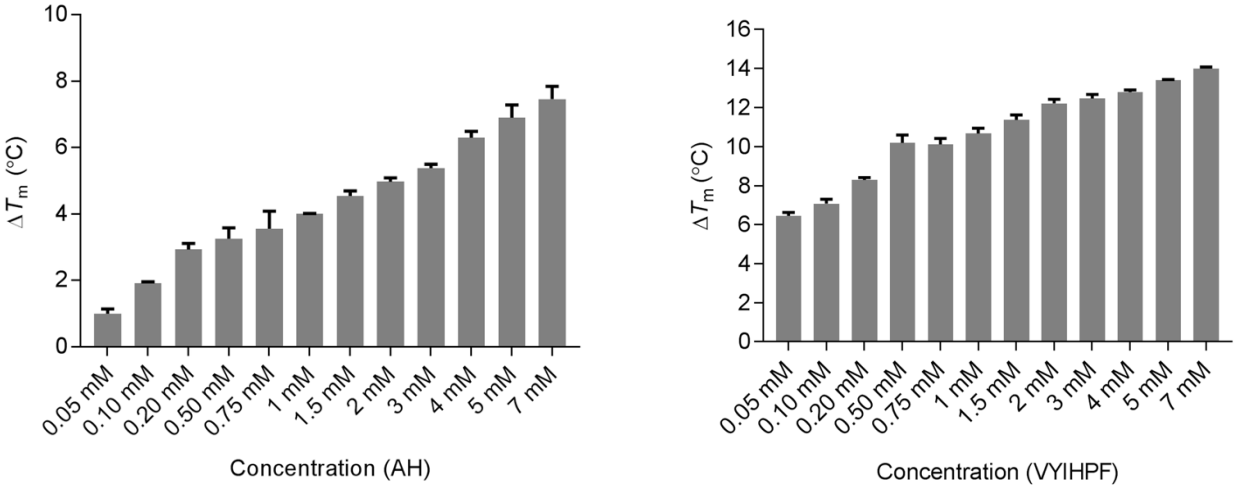

Figure S3. Thermal shift assay results showing that peptides AH (left) and VYIHPF (right) stabilize HpDppA in a concentration-dependent manner.

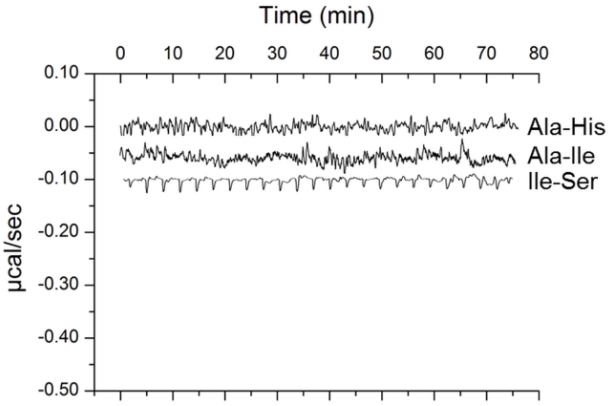

Figure S4. Raw ITC titration data for AH, AI and IS.

Supplemental Materials and Methods

Amino acid and codon-optimized nucleotide sequence of the recombinant HpDppA construct.

The amino acid and nucleotide sequences derived from the expression vector are underlined.

Amino Acid Sequence

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTSENPNATLNPSKENISVKEQKRFGGVLVFARG
ADGSSMDPALVTDGESYVATGNIYDTLVQFKYGTTEIEPALATSWDISPDGLVYTFHLRKGVYF
HQTKYWNKKVEFSAKDVLFSFERQMDKAKRYYSPGAKSYKYWEGMGMSHIIKSIEALDDYTIR
FTLNGPEAPFLANLGMDFLSILSKDYADYLEQNNKKDELAKKPVGTGPFKFFLWNKDEKIILLKN
QDYWGPKAYLDKVVVRTIPNSSTRALALRTGEIMLMTGPNLNEVEQLEKLPNIVVDKSAGLLAS
WLSLNTQKKYFNNPLVRLAINHAINVDDYIKVIYEGFAQKMVNPFPPTIWGYNYNIKPYEYDLK
KAKELLKQAGYPNGFKTTIFTTSTRNPKGAVFIQASLAKIGIDVKIEVYEWGAYLKRTGLGEHEM
AFAGWMADIADPDNFLYTLWSKQAASAIPTQNGSFYKSDAFSDLLIKAKRVSDQKEREALYLKA
QEIIHKDAPYVPLAYPYSVVPHLSKVKGYKTTGVSVNRFFKVYLEK

Nucleotide Sequence

ATGCATCATCACCATCACCATGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACG
GAAAACCTGTATTTTCAGGGAATTGATCCCTTCACCAGCGAGAACCCGAACGCGACCCTGAA
CCCGAGCAAGGAGAACATCAGCGTTAAGGAACAGAAACGTTTCGGTGGCGTTCTGGTGTTT
GCGCGTGGTGCGGATGGTAGCAGCATGGACCCGGCGCTGGTTACCGACGGTGAAAGCTACG
TGGCGACCGGCAACATTTATGATACCCTGGTGCAATTCAAGTATGGCACCACCGAGATTGAA
CCGGCGCTGGCGACCAGCTGGGACATTAGCCCGGATGGTCTGGTTTATACCTTCCACCTGCG
TAAGGGCGTGTACTTTCACCAGACCAAATATTGGAACAAGAAAGTTGAGTTTAGCGCGAAA
GACGTGCTGTTCAGCTTTGAACGTCAAATGGATAAGGCGAAACGTTACTATTCTCCGGGTGC
GAAGAGCTACAAATATTGGGAGGGTATGGGCATGAGCCACATCATTAAGAGCATCGAAGCG
CTGGACGATTACACCATTCGTTTCACCCTGAACGGTCCGGAGGCGCCGTTCCTGGCGAACCT
GGGTATGGACTTTCTGAGCATCCTGAGCAAAGACTACGCGGATTATCTGGAGCAGAACAAC
AAGAAAGATGAACTGGCGAAGAAACCGGTTGGCACCGGCCCGTTTAAGTTCTTTCTGTGGA
ACAAGGACGAAAAAATCATTCTGCTGAAAAACCAAGACTACTGGGGTCCGAAGGCGTATCT
GGATAAAGTGGTTGTGCGTACCATTCCGAACAGCAGCACCCGTGCGCTGGCGCTGCGTACCG
GCGAGATTATGCTGATGACCGGCCCGAACCTGAACGAGGTGGAACAGCTGGAAAAGCTGCC

GAACATCGTTGTGGATAAAAGCGCGGGCCTGCTGGCGAGCTGGCTGAGCCTGAACACCCAA
AAGAAATACTTCAACAACCCGCTGGTTCGTCTGGCGATCAACCACGCGATTAACGTTGACGA
TTACATCAAAGTGATTTATGAGGGTTTCGCGCAGAAAATGGTGAACCCGTTTCCGCCGACCA
TCTGGGGCTACAACTATAACATTAAGCCGTACGAGTATGACCTGAAGAAAGCGAAGGAACT
GCTGAAACAGGCGGGTTACCCGAACGGCTTCAAGACCACCATCTTTACCACCAGCACCCGTA
ACCCGAAGGGTGCGGTTTTCATCCAAGCGAGCCTGGCGAAGATCGGTATTGATGTTAAAATT
GAGGTGTACGAATGGGGCGCGTATCTGAAACGTACCGGTCTGGGCGAGCACGAAATGGCGT
TCGCGGGTTGGATGGCGGACATTGCGGACCCGGATAACTTTCTGTATACCCTGTGGAGCAAA
CAGGCGGCGAGCGCGATCCCGACCCAAAACGGCAGCTTCTATAAAAGCGACGCGTTTAGCG
ATCTGCTGATTAAGGCGAAACGTGTTAGCGACCAGAAGGAGCGTGAAGCGCTGTACCTGAA
GGCGCAAGAGATCATTCACAAAGATGCGCCGTATGTTCCGCTGGCGTACCCGTATAGCGTTG
TGCCGCACCTGAGCAAGGTGAAAGGTTACAAAACCACCGGCGTTAGCGTGAACCGTTTCTTT
AAGGTGTATCTGGAAAAATAA

Mass spectrometry analysis

The chromatographic features were identified using two approaches – untargeted feature extraction in an unbiased fashion, and *de novo* peptide sequencing of the MS/MS data.

Untargeted Feature Extraction

Each data file (datafile.raw) was centroided and converted to mzXML format using MSConvert (1) using the command: msconvert datafile.raw --filter "peakPicking true 1-" –mzXML. The resulting .mzXML files were then imported into MZmine 2.26 (2). Each file is then processed individually by sequentially applying the following modules/conditions as follows:

Mass detection module with the scans parameter set to MS level: 1, mass detector set to centroid with a noise level 1.0E5. Chromatogram builder module with scans parameter set to MS level: 1, Min time span (min) = 0.5, Min height = 1.0E7, m/z tolerance = 0.002 m/z or 2.0 ppm. Duplicate peak filter module with m/z tolerance = 0.002 m/z or 2.0 ppm, RT tolerance = 5.0 absolute (min), require same identification = false. Isotopic peak greater module with m/z tolerance = 0.001 m/z or 1.0 ppm, retention time tolerance 0.2 absolute (min), monotonic shape = false, Maximum charge = 2, representative isotope = lowest m/z. Chromatogram deconvolution module with local minimum search algorithm with parameters; chromatographic threshold = 10%, Search minimum in RT range (min) 0.20, Minimum relative height = 10%,

Minimum absolute height = 1.0E3, Min ratio of peak top/edge = 2, Peak duration range (min) = 0.50 – 50.00. Peak filter module with FWHM = 0.00 – 5.00 (all other parameters set to false). Peak list rows filter module with minimum peaks in a row = 1, all other parameters set to false.

The resulting peak lists from all samples in the experiment were then aligned using the RANSAC aligner module using the following conditions: m/z tolerance = 0.001 m/z or 1.0 ppm, RT tolerance = 1.0 absolute (min), RT tolerance after correction = 0.5 absolute (min), RANSAC iterations = 1000, Minimum number of points = 50.0%, Threshold value = 0.2, Linear model = false, require same charge state = false. The combined peak list was then further processed using the same RT and m/z range gap filler module with m/z tolerance = 0.002 m/z or 5.0 ppm and then the duplicate peak filter module with m/z tolerance = 0.001 m/z or 1.0 ppm, RT tolerance = 1.0 absolute (min), require same identification = false. The resulting peak list was then exported as a .csv file which was opened in Excel for further visualization and processing.

Specifically we produced bubble plots in which m/z is given on the x-axis, retention time (min) on the y-axis and the bubble size is proportional to the chromatographic peak area. In order to identify those features which are uniquely associated with the protein of interest rather than small molecule background we excluded all features with peak areas less than ten times that of a corresponding control sample.

From the corresponding bubble plot we selected five of the most significant features and extracted the MS/MS spectra corresponding to the same parent m/z and retention time. It was possible to assign these MS/MS spectra to short peptide sequences manually.

***De novo* peptide sequencing using PEAKS**

The .mzXML files generated following conversion with MSCovert (described above) were imported into Peaks Studio 64-bit v 7.5 (Bioinformatics Solutions Inc.) and processed using the default setting for Orbitrap (orbi-orbi). Initial data refinement was then performed with the following parameters: Precursor options = corrected, charge options = no correction, Filter Quality > 0.65, process = true and default = true. This was followed by de novo sequencing with the following instrument settings: Parent mass error tolerance = 10.0 ppm, fragment mass error tolerance = 0.01 Da, enzyme = none, report # peptides = 3, data refinement dependencies = 7. *De*

novo peptide sequencing resulted in the identification of 309 (DppA) and 11 (YckK) unique peptides at an average local confidence (ALC) greater than 50%.

Extraction of the chromatographic features corresponding to the peptides identified by PEAKS

Extracted ion chromatograms for all putatively identified peptides from the PEAKS analysis were extracted in MZmine 2.26 for each sample. The following modules/parameters were used: After importing the .mzXML files into MZmine 2.26 the targeted peak detection module was used with the parameters, intensity tolerance = 50%, noise level = 1.0E3, m/z tolerance = 0.001 m/z or 1.0 ppm and retention time tolerance = 1.0 absolute (min). The peak list file with the series of m/z and RTs to search was generated from the PEAKS data by taking the parent ion m/z and RT from all MS/MS scans for which a sequence had been assigned. Duplicate peak filter module with m/z tolerance = 0.02 m/z or 10.0 ppm, RT tolerance = 5.0 absolute (min), require same identification = false. Peak filter module with FWHM = 0.0 to 1.0, all other parameters = false. Peak list rows filter module with minimum peaks in a row = 1, all other parameters = false. Once this had been performed for each sample, the resulting peak lists were aligned using the RANSAC aligner module using the following conditions: m/z tolerance = 0.001 m/z or 1.0 ppm, RT tolerance = 1.0 absolute (min), RT tolerance after correction = 0.5 absolute (min), RANSAC iterations = 1000, Minimum number of points = 50.0%, Threshold value = 0.2, Linear model = false, require same charge state = false. The combined peak list was then further processed using the same RT and m/z range gap filler module with m/z tolerance = 0.002 m/z or 5.0 ppm and then the duplicate peak filter module with m/z tolerance = 0.005 m/z or 5.0 ppm, RT tolerance = 5.0 absolute (min), require same identification = true. The resulting peak list was then exported as a .csv file which was opened in Excel for further visualization and processing.

References

1. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak M-Y, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J,

Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P. 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nature Biotechnol 30:918.

2. Pluskal T, Castillo S, Villar-Briones A, Oresic M. 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11:395.