Supporting information

Structure and characterisation of CMP-Kdn synthetase from the haptophyte microalgae *Prymnesium parvum*

Claire Morley,[1] Alexandra J. Munro-Clark,[1] Ben A. Wagstaff,[1] Irina Ivanova,[3] Evgenia Dubinskaya,[1] Ava Rostock,[1] Mary E. Ortmayer,[1] Colin W. Levy,[1] Robert A. Field[1,2,3]*

Table of Contents

Additional sequence information

Sequences for PpNeuA-WT was from Wagstaff et al., 2018.[1] Sequences used in this study for protein expression. Underlined bases are overhangs added to the protein sequences for In-Fusion cloning into the pOPINF vector. Underlined and bold bases encode the single site mutation.

Codon optimised PpNeuA-R196L

AA GTT CTG TTT CAG GGC CCG ACC GTT TGG CAT CCG GTA CCT GAG GTA CGT
ATT GTA GCG GTA ATT CCG GCA CGT GGT GGC AGC GTT TCG ATT CCC CGG AAA
AAC ATT AAG CCT CTG GCG GGC CGC CCG CTG ATC GAT TGG GTC ATC AAA CCG
GCG CTG CAC TGC GGG ATT TTT ACC GAT GTA TAC GTG AGC ACC GAC GAT GAT
GCT ATC GCG AGC GTC GCT GAA AAA TGT GGC GCC AAA GTG CAT CGG CGT GAT
CCG GCC ACG GCG ACC GCT ACG GCC ACC ACC GAG TCT GCG CTG CTT GAC TTC
GCG CAG TCA CAC GGT GAC TTT GAC GTA CTG TGT CTT ATT CAA GCA ACC TCC
CCG TTT ATT ACC CCT CGC GAT CTG ATT AAC GGC TGG GAA TTA ATG CGC GCC
ATG GAA GCC GAT AGC CTC GTA ACC GCG GTG CGT GCG CAT CGC TTC CTT TGG
CAG GTT GAC AAA GAT ACA GGT CTT GCG AAA GCG AAA AAC TAT GAC CCA CTG
AAA CGC CCG CGC CGT CAG GAC TGG GAT GGG GAA CTG GTG GAG AAT GGC GCT
TTT TAC ATG ACC ACC AAA GCA TGC TTA GAG AAA CAT AAA TGT **CTC** CTC GGG
GAA AAG ATG GTC CTG CTG GAG ATG GAA GAG CAT ACG TTT ACT GAA CTG GAT
TCG TTA GTA GAC TGG CAG ATC GTG ACC AAT ATG ACC GAA AAT TAC GGT TAC
TGG TAA AGC TTT CTA GAC CAT

Codon optimised PpNeuA-R196A

AA GTT CTG TTT CAG GGC CCG ACC GTT TGG CAT CCG GTA CCT GAG GTA CGT
ATT GTA GCG GTA ATT CCG GCA CGT GGT GGC AGC GTT TCG ATT CCC CGG AAA
AAC ATT AAG CCT CTG GCG GGC CGC CCG CTG ATC GAT TGG GTC ATC AAA CCG
GCG CTG CAC TGC GGG ATT TTT ACC GAT GTA TAC GTG AGC ACC GAC GAT GAT
GCT ATC GCG AGC GTC GCT GAA AAA TGT GGC GCC AAA GTG CAT CGG CGT GAT
CCG GCC ACG GCG ACC GCT ACG GCC ACC ACC GAG TCT GCG CTG CTT GAC TTC
GCG CAG TCA CAC GGT GAC TTT GAC GTA CTG TGT CTT ATT CAA GCA ACC TCC
CCG TTT ATT ACC CCT CGC GAT CTG ATT AAC GGC TGG GAA TTA ATG CGC GCC
ATG GAA GCC GAT AGC CTC GTA ACC GCG GTG CGT GCG CAT CGC TTC CTT TGG
CAG GTT GAC AAA GAT ACA GGT CTT GCG AAA GCG AAA AAC TAT GAC CCA CTG
AAA CGC CCG CGC CGT CAG GAC TGG GAT GGG GAA CTG GTG GAG AAT GGC GCT
TTT TAC ATG ACC ACC AAA GCA TGC TTA GAG AAA CAT AAA TGT **GCG** CTC GGG
GAA AAG ATG GTC CTG CTG GAG ATG GAA GAG CAT ACG TTT ACT GAA CTG GAT
TCG TTA GTA GAC TGG CAG ATC GTG ACC AAT ATG ACC GAA AAT TAC GGT TAC
TGG TAA AGC TTT CTA GAC CAT

Additional Methods: Molecular Docking

To generate a model of PpNeuA in a closed conformation for docking studies a suitable homologue was selected from the protein data bank (PDB). The structure of Neisseria meningitidis CMP-sialic acid synthetase (NmCMAS, PDB ID: 6CKL) was chosen as the basis for the closed model of PpNeuA. Whilst the general protein architecture is maintained between PpNeuA and NmCMAS, the relative orientation of the two protein chains within the homodimer differs, this results in the "closed" conformation observed in NmCMAS. In addition to these differences in orientation, His157-Gln167 (PpNeuA Asn161-Gln171) within the dimerisation domain, adopt a more structured arrangement in NmCMAS containing an additional α-helix. It was necessary to model these changes onto PpNeuA prior to docking, this was achieved through the implementation of molecular dynamics in ICM-Pro.[2] Structure 6CKL was downloaded and chain C, which did not form part of the homodimer, deleted prior to superposition of the N-terminal globular domain of chain A with the equivalent domain of PpNeuA (RMSD 0.62Å). The conformation of NmCMAS chain B was used as a template and imposed onto the structure of PpNeuA, via molecular dynamics tool in ICM-Pro. The resultant PpNeuA model, reflecting both the orientation and secondary structural features observed in NmCMAS, formed the starting point for subsequent docking studies. ICM-Pro e-Dock scores of -32 and lower indicative of binding.[3] Docking of CTP resulted in an eDock score of –163. Docking of CMP-Kdn, in addition to the steps described above, required R196 to be set as a flexible side chain during the docking simulation. The flexibility resulted in a docked pose of CMP-Kdn that included interactions with R196 as detailed in Figure 4. The docking of CMP-Kdn yielded a radical convolution neural network (RTCNN) docking score of -23.6. Unlike eDock score, RTCNN does not have a threshold for binding, however, more negative scores are more indicative of binding. Attempts at docking CMP-Neu5Ac performed in the same manner as for CMP-Kdn, with a flexible R196 side chain, did not result in any plausible docked poses. To rationalise this a model of CMP-Neu5Ac was superimposed onto the docked pose of CMP-KDN revealing a likely steric clash arising between CMP-Neu5Ac and R196 as shown in Figure 4. A model of PpNeuA R196L was generated *in silico* using ICM-Pro and energy minimised.

Supplementary Figures

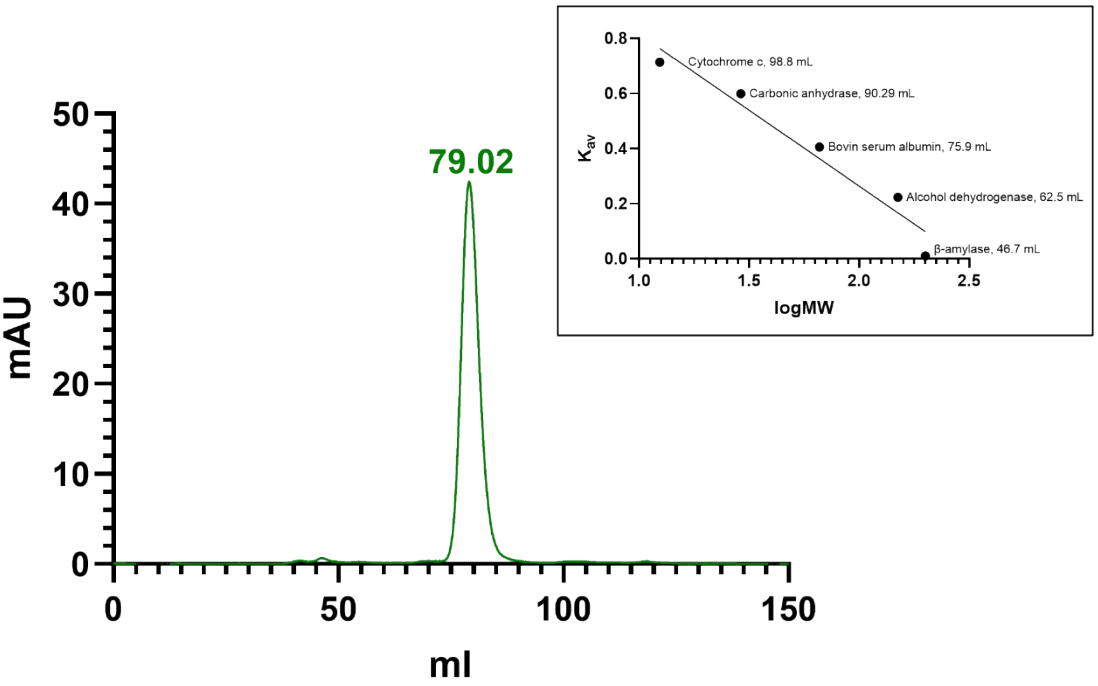

Figure S 1: Gel filtration profile of PpNeuA. Superdex 16/600 200 pg column was calibrated with Gel filtration markers kits for protein molecular weights 12,000 – 200,000 Da (Sigma). PpNeuA eluted as a single peak, which corresponded to 46.6 kDa. The monomeric unit is 26.2 kDa, hence, PpNeuA eluted as a dimer in solution. All runs were carried out in TBS at 0.5 mL min^{-1} flow rate.

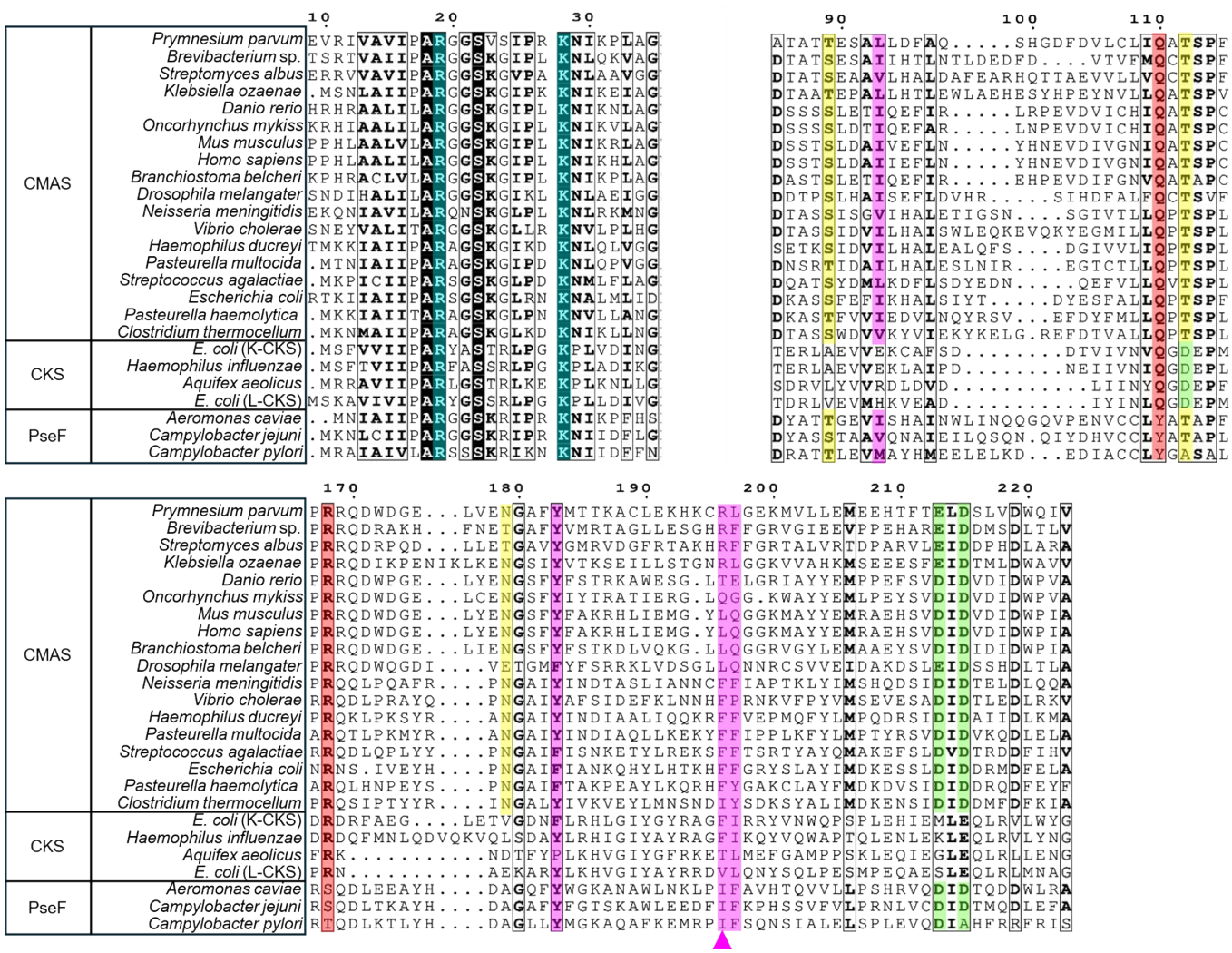

Figure S 2: Partial amino acid sequence alignments of CMAS, CKS and PseF enzymes. Highlighted with green are residues that are involved in binding Mg^{2+}, in cyan are residues that are involved in binding CTP, in yellow are residues that are implicated in binding sialic acid; in red are residues that bind both Mg^{2+} and sialic acid; and in magenta residues that are part of the hydrophobic binding pocket. The magenta triangle highlights the proposed residue that confers Kdn specificity. All enzymes bind metals, CMAS and PseF bind metals through the DxD motif, whilst CKS enzymes use a different metal binding motif, which is not shown in this sequence alignment.[4]

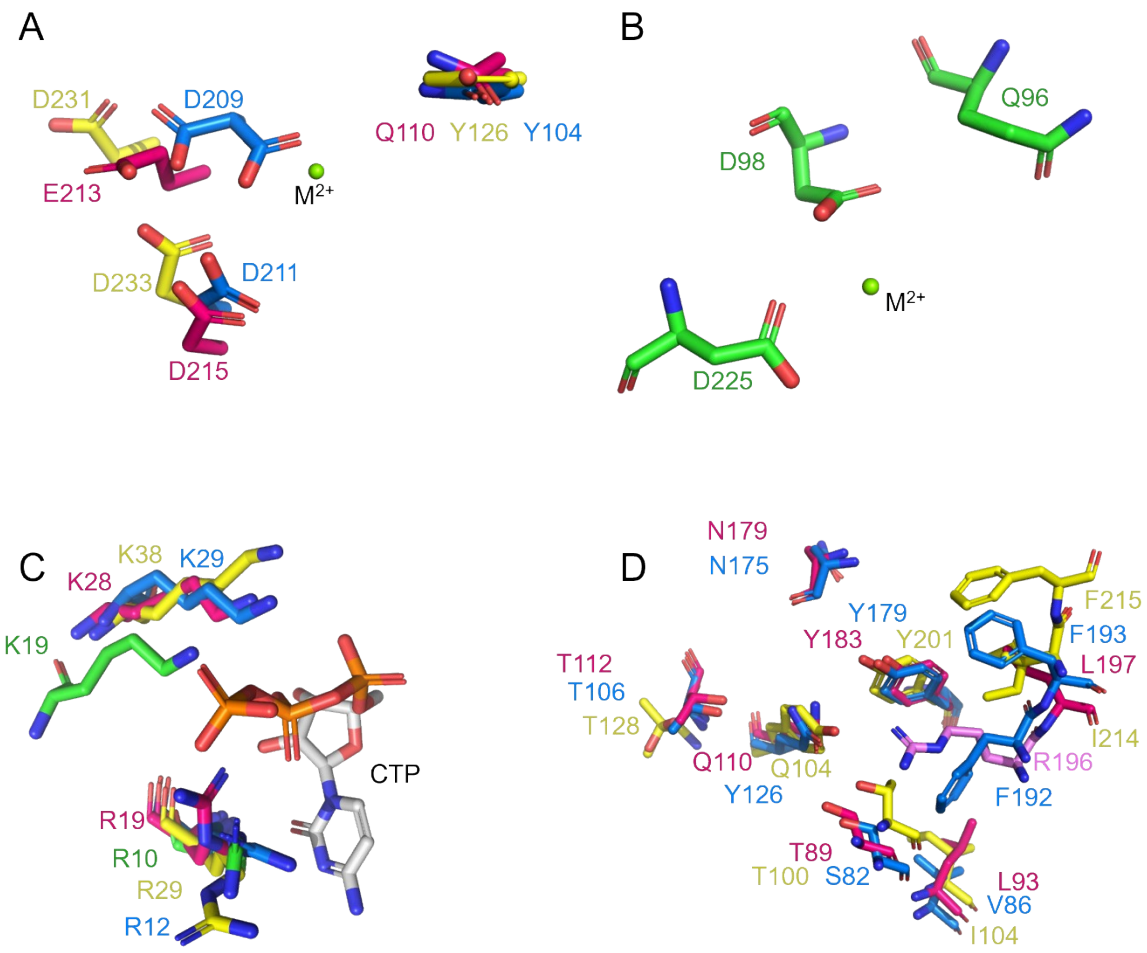

Figure S 3: Structural alignments of CMAS, CKS and PseF enzymes in open conformation. (A) Metal binding residues from CMAS, PseF and PpNeuA, (B) Metal binding residues from CKS, CKS has different metal binding residues than the other enzymes, (C) CTP binding residues, (D) Sialic acid binding residues in CMAS, PseF and PpNeuA. CMAS (PDB ID: 6CKJ) is in blue, CKS enzyme (PDB ID: 1GQ9) is in green, PseF enzyme (PDB ID: 9FTB) is in yellow and PpNeuA is in pink. The metal ion is shown in green, and CTP is shown in grey.

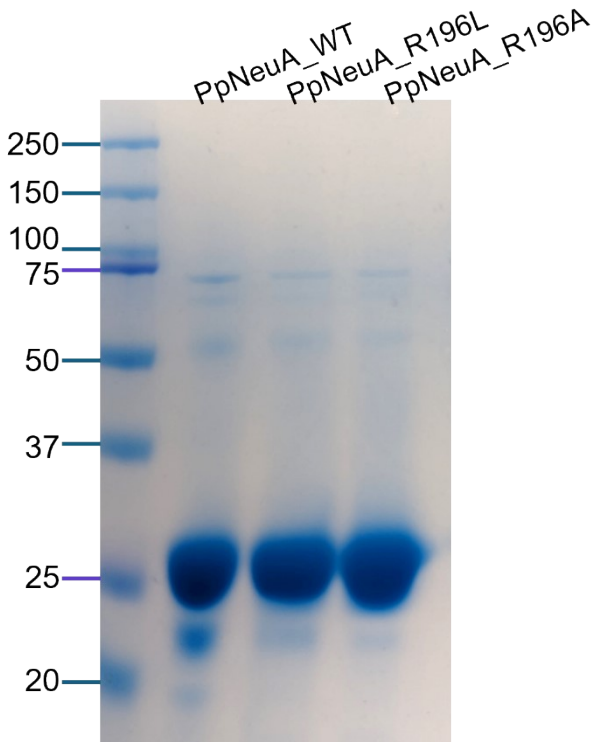

Figure S 4: SDS-PAGE of recombinant PpNeuA-WT, PpNeuA-R196L, PpNeuA-R196A. PpNeuA and mutants are 28 kDa.

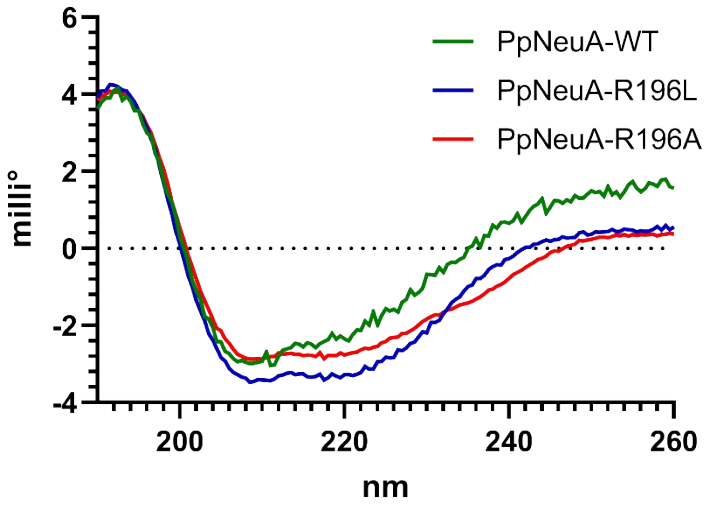

Figure S 5: Circular dichroism analysis of PpNeuA and mutant enzymes. All proteins were concentrated to 1 mg mL-1 and analysed in TBS at room temperature. The spectrum for all proteins were similar, indicating that the mutants had folded correctly.

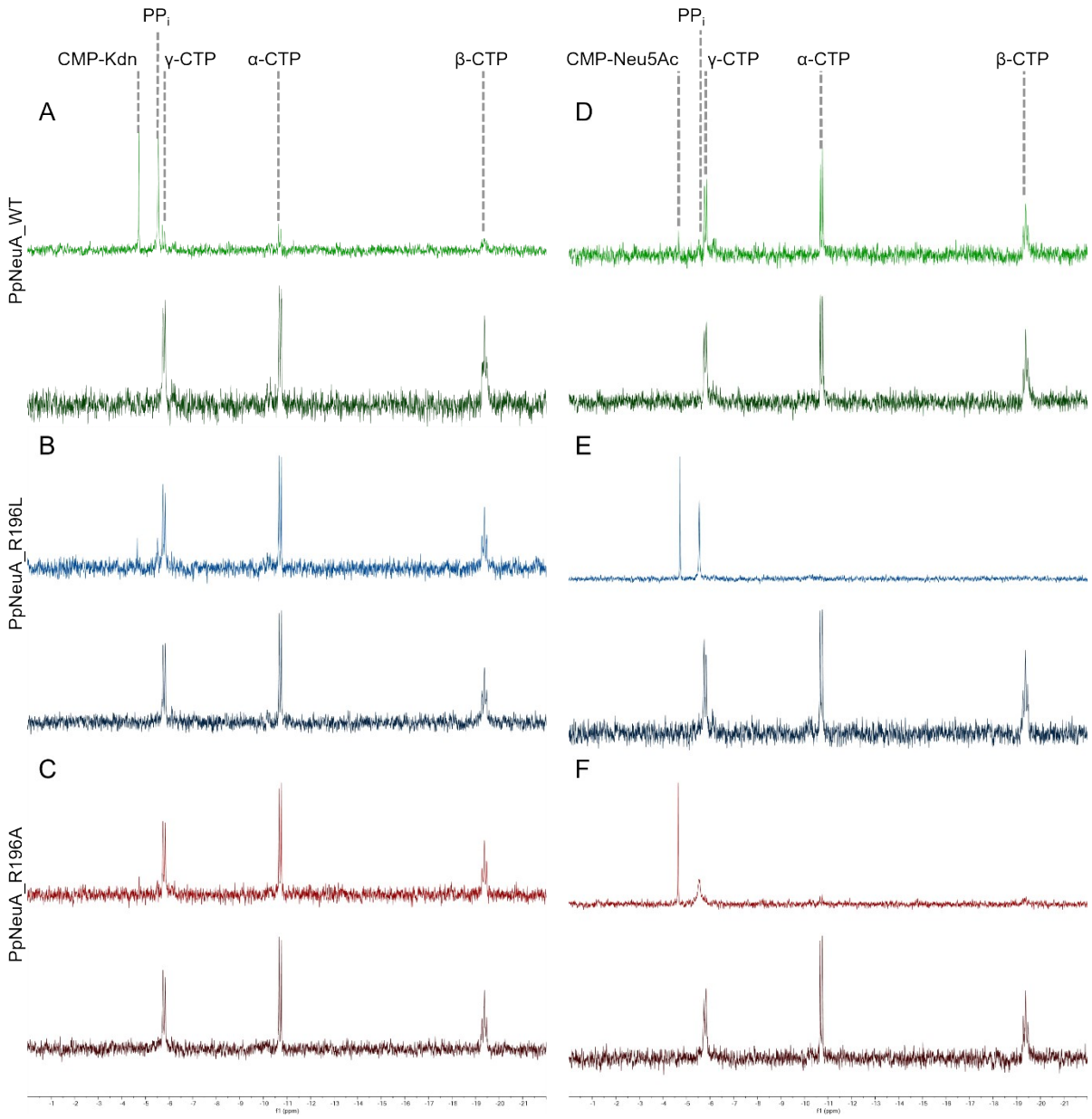

Figure S 6: ^{31}P NMR spectrum of reaction catalysed by PpNeuA and mutants. The reaction was initiated with the addition of 20 µg enzyme and monitored over 60 min. Reactions were carried out with Kdn (A, B, C) or Neu5Ac (D, E, F). Reactions with the enzymes less preferred substrates show some conversion.

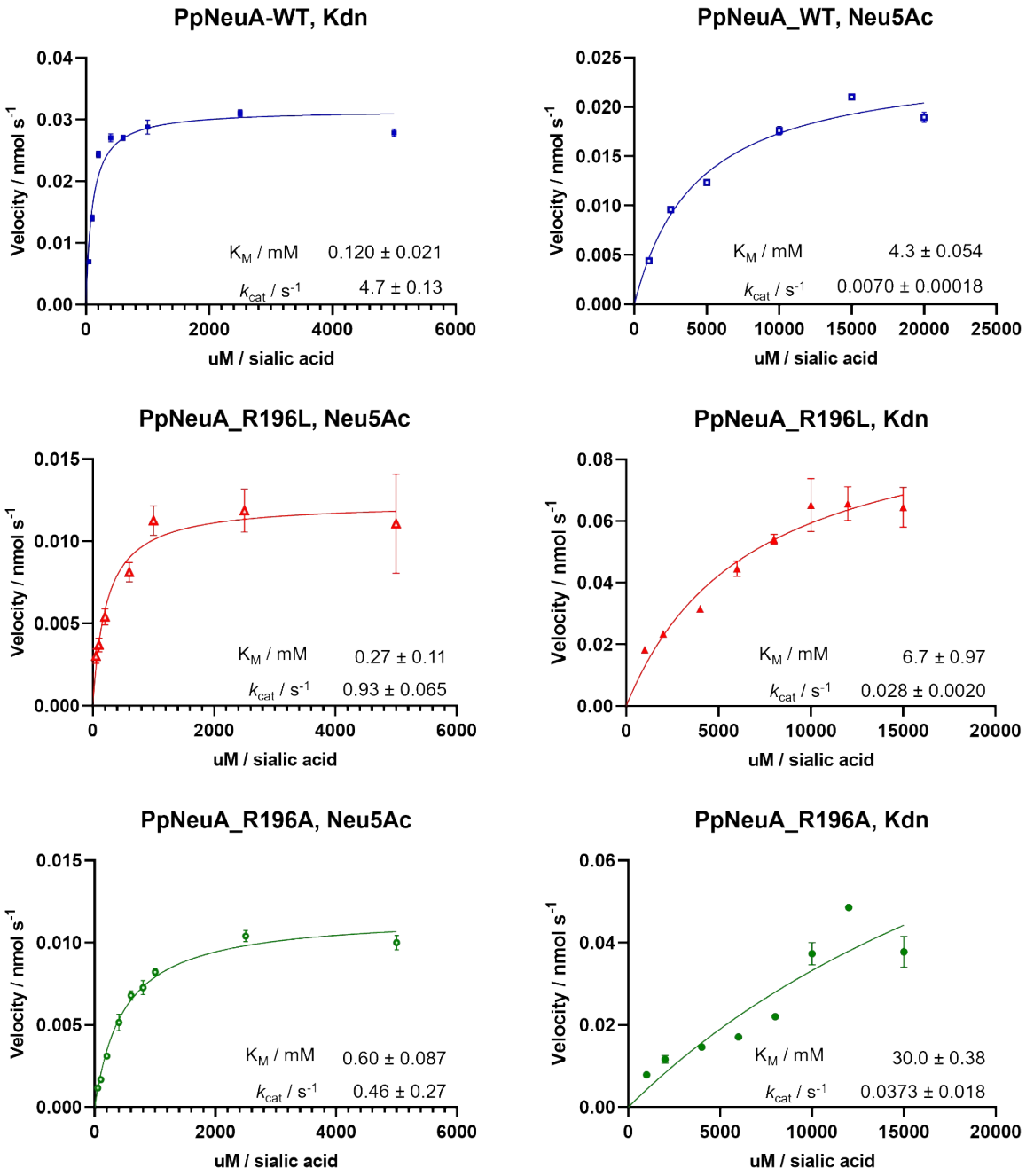

Figure S 7: Michaelis menten plots of PpNeuA and mutants with Neu5Ac and Kdn substrates. The reaction was initiated by the addition of 200 - 800 ng PpNeuA or mutants and followed by absorbance at 360 nm. . The initial rates of reaction were measured over a range of sialic acid (Neu5Ac, Kdn) concentrations with a fixed amount of CTP (0.5 mM). V_{max}, K_m and k_{cat} were calculated according to the Michaelis Menten equation using GraphPad Prism, standard deviation shown in error bars.

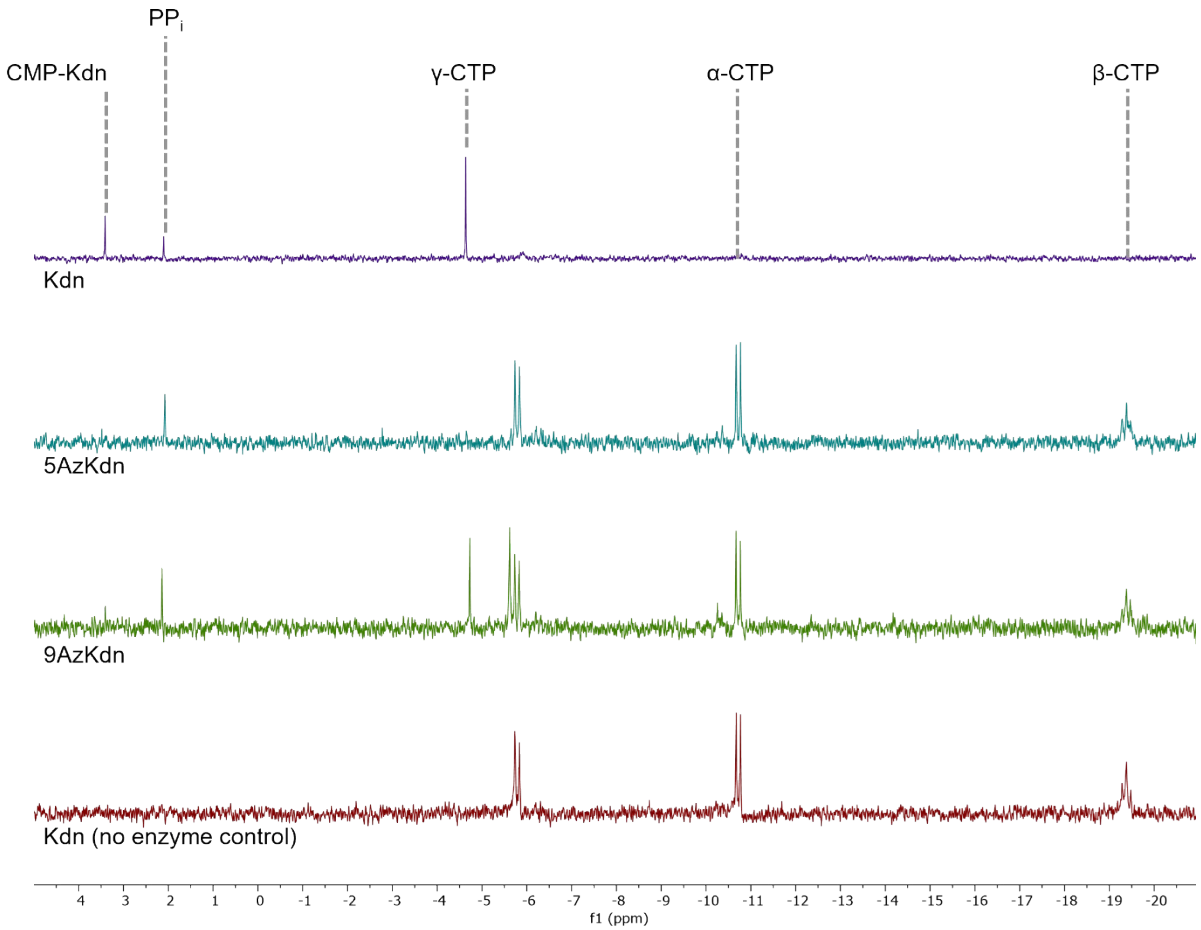

Figure S 8: ^{31}P NMR spectrum of reaction catalysed by PpNeuA with 5AzKdn and 9AzKdn. The reaction was initiated with the addition of 20 µg enzyme and monitored over 120 min.

Supplementary Tables

Table S 1: Data collection and refinement statistics, statistics in parentheses correspond to the outer resolution shell.

	PpNeuA PBD ID: 9T4Z
Wavelength	0.9763
Resolution range	51.61 - 1.8 (1.84 - 1.8)
Space group	P 1 21 1
Unit cell	56.432 47.782 87.85 90 102.12 90
Total reflections	359077 (17068)
Unique reflections	51291 (2747)
Multiplicity	7.0 (6.2)
Completeness (%)	97.63 (96.57)
Mean I/sigma(I)	7.92 (0.43)
Wilson B-factor	24.79
R-merge	0.1345 (2.436)
R-meas	0.1454 (2.659)
R-pim	0.05459 (1.045)
CC1/2	0.997 (0.289)
CC*	0.999 (0.67)
Reflections used in refinement	41768 (2756)
Reflections used for R-free	2047 (105)
R-work	0.1872 (0.2881)
R-free	0.2145 (0.3163)
Number of non-hydrogen atoms	3763
macromolecules	3519
ligands	0
solvent	244
Protein residues	445
RMS(bonds)	0.005
RMS(angles)	0.79
Ramachandran favored (%)	97.04
Ramachandran allowed (%)	2.51
Ramachandran outliers (%)	0.46
Rotamer outliers (%)	0.79
Clashscore	2.85
Average B-factor	34.51
macromolecules	34.42
solvent	35.83

Supplementary References

1. B. A. Wagstaff, M. Rejzek and R. A. Field, *J Biol Chem*, 2018, **293**, 16277-16290.
2. J. Fernandez-Recio, M. Totrov and R. Abagyan, *Proteins*, 2003, **52**, 113-117.
3. H. J. Zhong, L. J. Liu, C. M. Chong, L. Lu, M. Wang, D. S. Chan, P. W. Chan, S. M. Lee, D. L. Ma and C. H. Leung, *PLoS One*, 2014, **9**, e92905.
4. D. J. Heyes, C. Levy, P. Lafite, I. S. Roberts, M. Goldrick, A. V. Stachulski, S. B. Rossington, D. Stanford, S. E. Rigby, N. S. Scrutton and D. Leys, *J Biol Chem*, 2009, **284**, 35514-35523.