

Electronic Supplementary Information

Trapping and structural characterisation of a covalent intermediate in vitamin B₆ biosynthesis catalysed by the Pdx1 PLP synthase

Matthew J. Rodrigues^{1,2}, Nitai Giri³, Antoine Royant^{4,5}, Yang Zhang⁶, Rachel Bolton^{1,2}, Gwyndaf Evans², Steve E. Ealick⁶, Tadhg Begley³, Ivo Tews¹

¹Biological Sciences, Institute for Life Sciences, University of Southampton, Southampton, SO17 1BJ, UK. ²Diamond Light Source, Harwell Science and Innovation Campus, Didcot, OX11 0DE, UK. ³Department of Chemistry, Texas A&M University, College Station, TX 77843, USA.

⁴Université Grenoble Alpes, CNRS, CEA, IBS, CS 10090, 38044 Grenoble Cedex 9, France.

⁵European Synchrotron Radiation Facility, CS 40220, 38043 Grenoble Cedex 9, France.

⁶Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA

Protein expression and purification:

Pdx1.3 of *Arabidopsis thaliana* was cloned as described previously¹. The K166R mutation was introduced using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies), high purity salt free primers were supplied by Eurofins MWG.

Expression and purification protocols were similar to the ones published previously². Briefly, BL21 (DE3) cells were transformed with native and Pdx1.3^{K166R} plasmid as required and grown in LB medium to an optical density of 0.6 at 37 °C. Protein expression was induced by addition of 60 ml 25% (w/v) lactose per L culture, and cells were grown for a further 16 h at 30 °C before harvesting. Cells were lysed by sonication, followed by centrifugation at 140,000 × g for 1 hour. Cleared supernatants were applied to 1 ml Immobilised Metal Affinity Chromatography HiTrap columns (GE Healthcare) pre-equilibrated with lysis buffer (50 mM Tris-Cl, pH 7.5, 500 mM sodium chloride, 10 mM imidazole, 2% glycerol); wash buffer additionally contained 50 mM imidazole; elution buffer additionally contained 500 mM imidazole and 5% glycerol. Eluted proteins were applied to size-exclusion chromatography using a Superdex 26/60 column (GE Healthcare) equilibrated with gel filtration buffer (20 mM Tris-Cl pH 8.0, 200 mM KCl). Proteins were concentrated using Vivaspin 20 centrifugal concentrators with a 30 kDa molecular weight cut off (Sartorius).

UV-vis absorption spectroscopy of the I₃₂₀ and I₃₃₃ intermediates:

Pdx1.3 (20 μM) was incubated at RT with 10 mM ribose 5-phosphate (R5P) for 15 minutes, followed by addition of 300 mM ammonium sulphate and incubation for an additional 45 minutes². An identical procedure was used with Pdx1.3^{K166R}. To detect pyridoxal 5-phosphate (PLP) formation, the above reaction was performed with Pdx1.3 and 20 mM glyceraldehyde 3-phosphate (G3P) was subsequently added, incubating the reaction mixture for an additional 1 hour^{3, 4}.

UV-vis absorption spectra were collected in a quartz cuvette using a Hitachi U-3010 spectrophotometer. The changeover between the deuterium and halogen lamps in this instrument causes a small discontinuity in the spectra at 370 nm, visible in Figure 1. The effects of Rayleigh scattering, due to protein aggregation after adding ammonium sulphate, were corrected for by fitting a Rayleigh function to subtract the modelled baseline^{5, 6}. The spectra were smoothed using a Savitzky-Golay⁷ algorithm with a 21 nm window and second order polynomial and scaled to each other by their absorbance at 280 nm. The spectra were processed and plotted using SciPy⁸ and Matplotlib⁹ in a custom Python script.

Preparation of protein complexes for mass spectrometry:

To generate apo-protein, bound R5P had to be removed from the enzymes. For this, 10 μM native Pdx1.3 or Pdx1.3^{K166R} was incubated with 100 mM ammonium sulphate for 1 hour before adding 10 mM G3P and incubating for an additional 3 hours^{3, 4, 10}. Proteins were then repurified by size-exclusion chromatography (as above) to remove biosynthesised PLP.

To generate the R5P adduct, 10 μM native Pdx1.3 or Pdx1.3^{K166R} were incubated with 5 mM R5P for 30 minutes². To generate the I₃₂₀ or I₃₃₃ complexes, 10 mM native Pdx1.3 or Pdx1.3^{K166R} were incubated with 5 mM R5P for 15 minutes before adding 100 mM ammonium sulphate, incubating the proteins for a further 2 hours².

Trypsin digestion:

Protein samples (20 mL of a 10 mM solution in 200 mM KCl, 50 mM Tris pH 8.0) were mixed with the denaturation buffer (10 mL of a solution containing 6M guanidine-HCl in 25 mM ammonium bicarbonate pH 8.0). DTT (2 mL of 100 mM solution in 25 mM ammonium bicarbonate pH 8.0) was added and the samples were incubated at room temperature for 1 hour. Iodoacetamide (10 mL of 200 mM solution in 25 mM ammonium bicarbonate pH 8.0) was added and the samples were incubated, in the dark, at room temperature for 1 hour. Finally, 58 mL of 25 mM ammonium bicarbonate pH 8.0 and 1 mg of trypsin were added, and the reaction mixture was incubated at 37 °C overnight.

Mass spectrometry:

Prior to LC-MS analysis, the trypsin digested samples were purified using C₁₈ ZipTips (Merck Millipore). The samples were then dried and resuspended in 0.1% aqueous formic acid. LC-ESI-TOF-MS/MS was performed using a Dionex ultimate 3000 nLC system followed by a Thermo Fisher Orbitrap Fusion mass spectrometer using an ESI source in positive mode. LC-MS was performed using an Acclaim PepMap C₁₈ LC column.

LC conditions:

Buffer A: 0.1% aqueous formic acid

Buffer B: 0.1% formic acid in acetonitrile

LC method:

0 min-98% A, 5 min-98% A, 37 min- 45% A, 40 min- 90% A, 45 min-90% A, 47 min-98%A, 100 min-98%A

Results are shown for Pdx1.3 (Fig. S1), Pdx1.3^{K166R} (Fig S2), Pdx1.3, after addition of R5P (Fig S3), Pdx1.3^{K166R} after addition of R5P (Fig. S4), Pdx1.3, after addition of R5P and (NH₄)₂SO₄ (Fig S5), Pdx1.3^{K166R} after addition of R5P and (NH₄)₂SO₄ (Fig. S6).

Crystallographic analysis:

Pdx1.3^{K166R} was crystallised at a concentration of 6.5 mg ml⁻¹ at 18 °C in sitting drops composed of 1 µl protein and 0.5 µl mother liquor (28.1% PEG 1000 (w/v) and 100 mM HEPES (pH 7.0)). Crystals were cryoprotected in mother liquor containing an additional 20% glycerol (v/v).

For the Pdx1.3^{K166R}:I₃₃₃ complex, Pdx1.3^{K166R} crystals in 1.5 µl crystallisation drops were incubated after addition of 0.6 ml substrate buffer consisting of 28.6% PEG 1000 (w/v), 100 mM HEPES (pH 7.5), 10 mM ribose 5-phosphate, and 800 mM ammonium chloride. After 2 hours, crystals were cryoprotected in mother liquor containing an additional 20% glycerol (v/v).

Crystallographic datasets were indexed and integrated using DIALS¹¹ and scaled using AIMLESS¹². Molecular replacement was performed with MOLREP¹³ using the native Pdx1.3-I₃₂₀ coordinates (PDB: 5LNU, ²) with solvent and ligand atoms removed from the model. The electron density map shown in Figure 3b was calculated using the program phenix.composite_omit_map with simulated annealing, and the Lys98-I333 residues omitted from map calculation¹⁴. Iterative model building and refinement was performed using COOT¹⁵, PHENIX¹⁴ and REFMAC5¹⁶. Ligand restraints were generated using JLigand¹⁷. Protein structures were deposited with the PDB under accession codes 7NHE for the Pdx1.3^{K166R}:I333 complex and 7NHF for Pdx1.3^{K166R}.

***In crystallo* UV-vis absorption spectroscopy:**

UV-vis spectra of flash cooled protein crystals were collected offline at the ESRF at the *icOS* Lab; the experimental setup was described previously¹⁸. Data collection on crystals is limited by the experimental setup (discussed in ¹⁸), as light is scattered away from the detection path¹⁹. The effects of light scattering at the crystal-solvent and crystal-air interfaces were corrected for by background subtraction (background proportional to 1/λ⁴). UV-Vis spectra of Pdx1.3:I₃₂₀ and Pdx1.3^{K166R}:I₃₃₃ were scaled to each other by matching the maximum optical density in each spectrum. The spectra were processed and plotted using SciPy⁸ and Matplotlib⁹ in custom Python scripts and smoothed using a Savitzky-Golay⁷ algorithm with a 21 nm window and second order polynomial (using `scipy.signal.savgol`).

Supplementary text

Review of previously published intermediate complex structures and assignment of I333

1. Previously published intermediate complex structures.

It is usually very difficult to experimentally characterise individual snapshots of the reaction coordinate in crystal structures. *Arabidopsis thaliana* proteins have successfully been used in two studies^{2, 20}. Of note, there are three alleles of the *pdx1* gene in *Arabidopsis*, called *pdx1.1*, *pdx1.2*, *pdx1.3*, where Pdx1.2 is an inactive pseudo-enzyme^{4, 21}. A series of Pdx1.3 intermediate complexes was published by us and others, as follows: Pdx1.3 (5K3V,²⁰), R5P (5LNS,²), I320 (5K2Z,²⁰), I320 single crystal (5LNU,²); I320 multi crystal (5LNV,²); I320-G3P (5LNW,²) and PLP (5LNR,²).

Additionally, a structure of the paralog variant protein Pdx1.1^{K166R} was published and named preI320 complex (5LNT,²). While it was desirable to produce this structure as part of the Pdx1.3 series, we were unable to determine the structure of a similar complex for the Pdx1.3^{K166R} variant, and thus vital information for the Pdx1.3 paralog was missing. This knowledge gap is closed with the data presented here.

2. Structures in this communication

We present the structure of an intermediate of the Pdx1.3 K166R variant (7NHE), obtained using a comparable protocol as for preparation of the preI320 Pdx1.1 K166R complex (5LNR,²). Comparison of unit cell dimensions shows that Pdx1.1 and Pdx1.3 crystal systems are very similar, and crystals diffracted to similar resolution. We also determined the structure of the Pdx1.3 K166R protein (7NHF) to understand the effect of the K166R exchange in 3D structure. The data presented here complete the Pdx1.3 series of structures, in a single system.

3. Data interpretation and structure assignment

To determine the structure of the Pdx1.3^{K166R} intermediate that formed after R5P and (NH₄)₂SO₄ addition we first tested the intermediate structure of the Pdx1.1^{K166R} exchange variant prepared using similar conditions². However, it became quickly apparent from notable difference density remaining that this structure was incorrect.

Using chemical reasoning, we tested three different structures in refinement, as shown in Figure S7. The three structures are tautomers that fit the m/s data presented in Table 1 and Figure S6. A further indication was taken from the spectral analysis, predicting conjugation. The resulting interpretation for the intermediate in Pdx1.3^{K166R} (7NHE) is consistent with all data presented here, but differs from the previously published Pdx1.1^{K166R}:R5P:(NH₄)₂SO₄ complex (5LNR,²), as shown in Figure S8.

4. Comparison of all crystal structures

We sought to better understand the differences in intermediate structures and noted further notable differences found in the beta strand and loop region supporting the local structure of K166/R166. We previously observed a peptide flip of 166/167 in the preI320 Pdx1.1 complex (5LNR,²), which was also observed in the earlier I320 complexes (5LNU, 5LNV,²), and additionally was observed in work of others (5K2Z,²⁰ and 4WY0,²²). This peptide flip was not observed in the I333 Pdx1.3 complex presented here (7NHE), nor was it not observed in Pdx1.3 K166R protein presented here (7NHF), Figure S9. The peptide flip may well be a requirement for the lysine-swing of K166 in preparation for I320 formation, and may be linked to ammonia transfer, as suggested earlier by us² and others²⁰. The present work suggests that the peptide flip may not be required for I333 formation.

Table S1. Crystallographic analysis

	Pdx1.3^{K166R} PDB:7NHF	Pdx1.3^{K166R}:I₃₃₃ PDB:7NHE
Data collection	ESRF ID23-1	ESRF ID23-1
Space group	R3	R3
Unit cell a=b, c (Å)	178.0, 115.1	178.0, 115.1
Unit cell α , β , γ (°)	90, 90, 120	90, 90, 120
Resolution (Å) *	38.38-2.35 (2.42-2.35)	92.23-2.23 (2.28-2.23)
R _{merge} *	0.120 (1.144)	0.104 (0.886)
CC1/2 *	0.987 (0.649)	0.990 (0.517)
I / σ I *	4.6 (1.3)	7.2 (2.4)
Completeness (%) *	98.7 (94.4)	99.7 (99.8)
Multiplicity *	2.8 (2.5)	3.9 (3.8)
Wilson B (Å ²)	38.09	34.2
Refinement		
Resolution (Å)	38.38-2.35	92.23-2.23
No. Reflections	55,906 (4,366)	66,508 (4,475)
R _{work} / R _{free} (%)	21.95/24.90	0.169/0.212
No. Atoms		
Protein	7,917	8,168
Ligand/ion	20	125
Water	39	447
B-factors		
Protein (Å ²)	57.8	48.7
Ligand/ion (Å ²)	84.3	62.7
Water (Å ²)	41.2	46.3
Ramachandran		
Allowed (# / %)	997 / 95.0	1016 / 97.9
Generally allowed (# / %)	51 / 4.9	21 / 2.0
Disallowed (# / %)	1 / 0.1	1, 0.1
R.m.s. deviations		
Bond lengths (Å)	0.008	0.006
Bond angles (°)	1.516	0.821

* Numbers given in parentheses are for the highest resolution shell

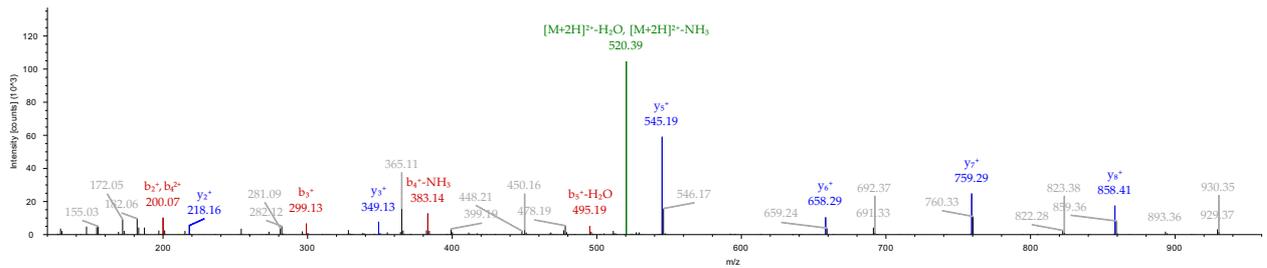


Figure S1. Pdx1.3: $m/z = 529.3074$ ($z = +2$), observed $[M+H]^+ = 1057.6076$ matches to the peptide QAVTIPVMAK₉₈ with a predicted $[M+H]^+ = 1057.6074$.

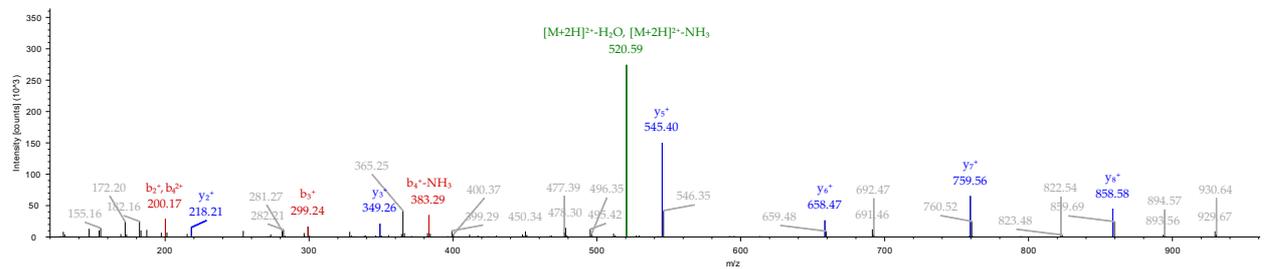


Figure S2. Pdx1.3^{K166R}: $m/z = 529.3079$ ($z = +2$), observed $[M+H]^+ = 1057.6085$ matches to the peptide QAVTIPVMAK₉₈ with a predicted $[M+H]^+ = 1057.6074$.

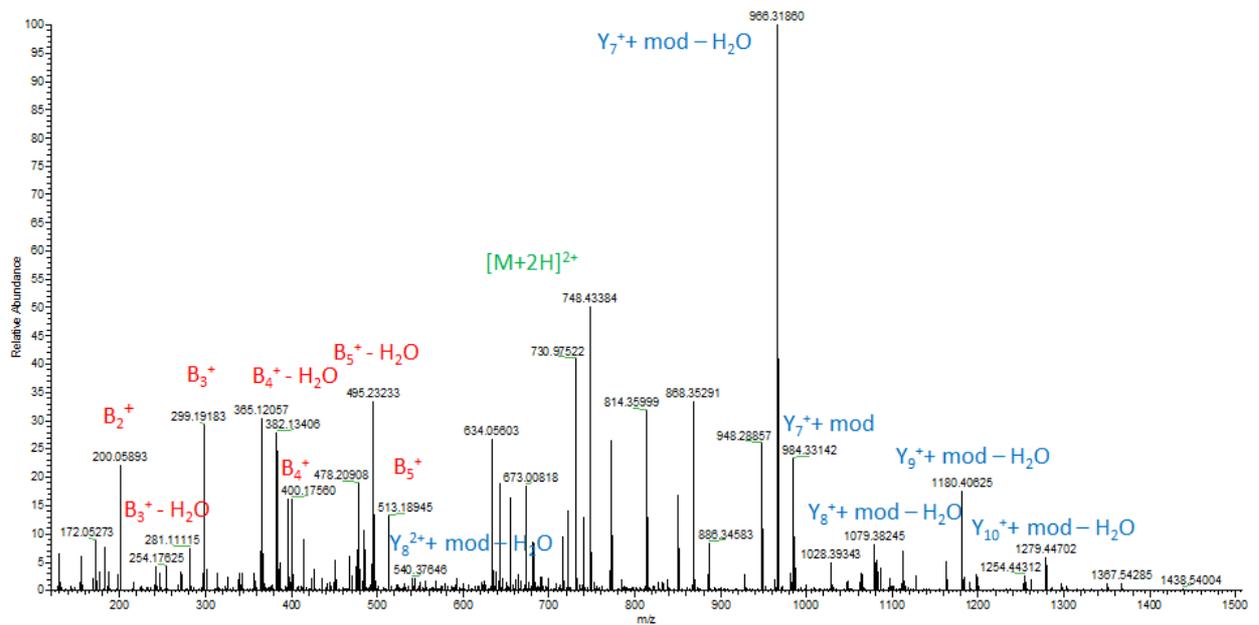


Figure S3. Pdx1.3, after addition of R5P: $m/z = 748.8812$ ($z = +2$), observed $[M+H]^+ = 1496.7546$. Subtracting the peptide QAVTIPVMAK₉₈AR with a predicted $[M+H]^+ = 1284.7457$ yields $1496.7546 - 1284.7457 = 212.0089$; this modification is interpreted as structure 1.

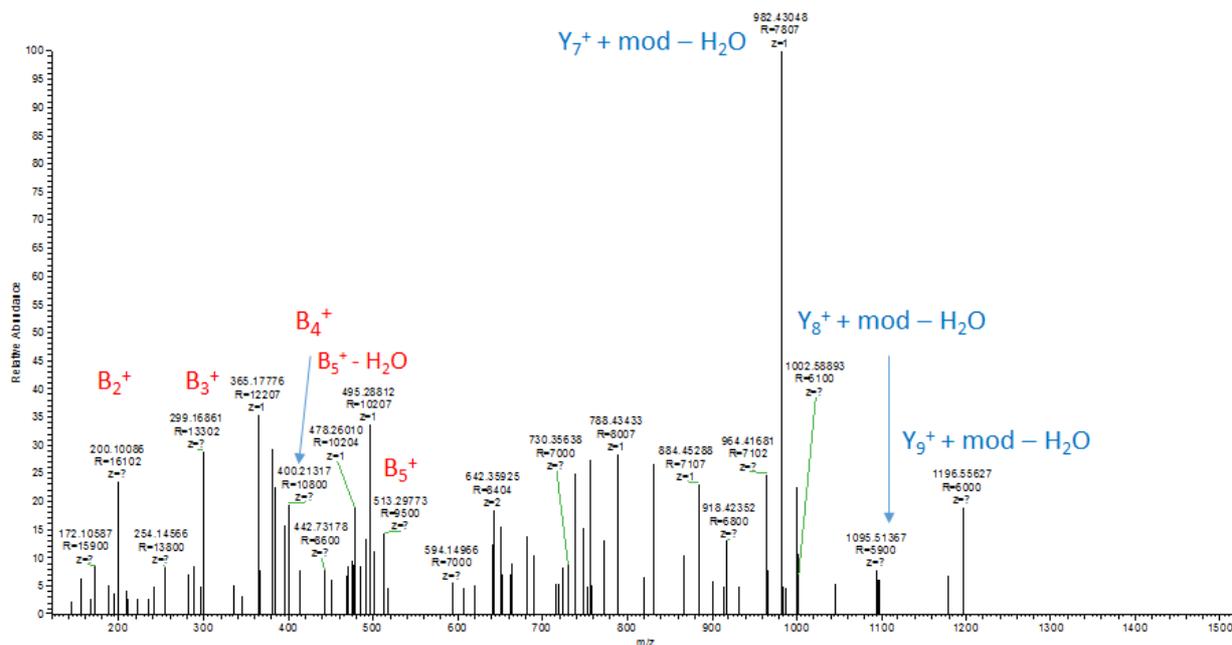


Figure S4. Pdx1.3^{K166R} after addition of R5P: $m/z = 756.8718$ ($z = +2$), observed $[M+H]^+ = 1512.7436$. Subtracting QAVTIPVM(O)AK₉₈AR with a predicted $[M+H]^+ = 1300.7406$ yields $1512.7436 - 1300.7406 = 212.003$; this modification is interpreted as structure **1**. NB. M(O) indicates an additional oxygen in the Met residue.

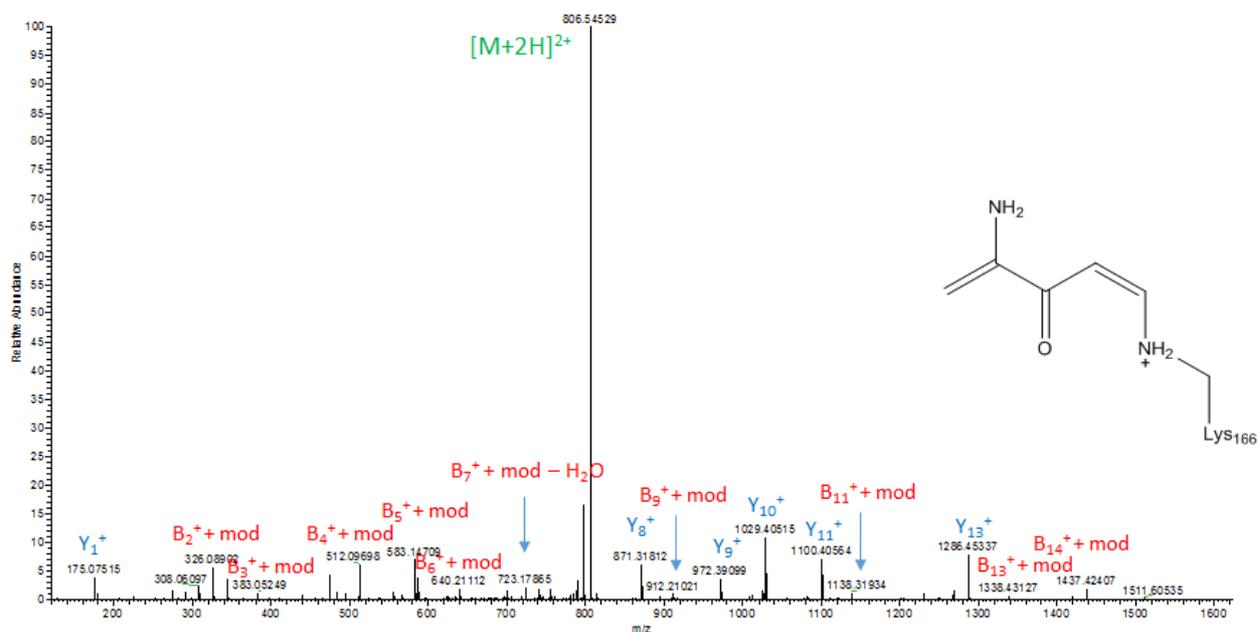


Figure S5. Pdx1.3, after addition of R5P and $(NH_4)_2SO_4$: $m/z = 806.4202$ ($z = +2$), observed $[M+H]^+ = 1611.8326$. Subtracting TK₁₆₆GEAGTGNIEAVR with a predicted $[M+H]^+ = 1515.8125$ yields $1611.8326 - 1515.8125 = 96.0201$ which is interpreted as structure **12** shown here.

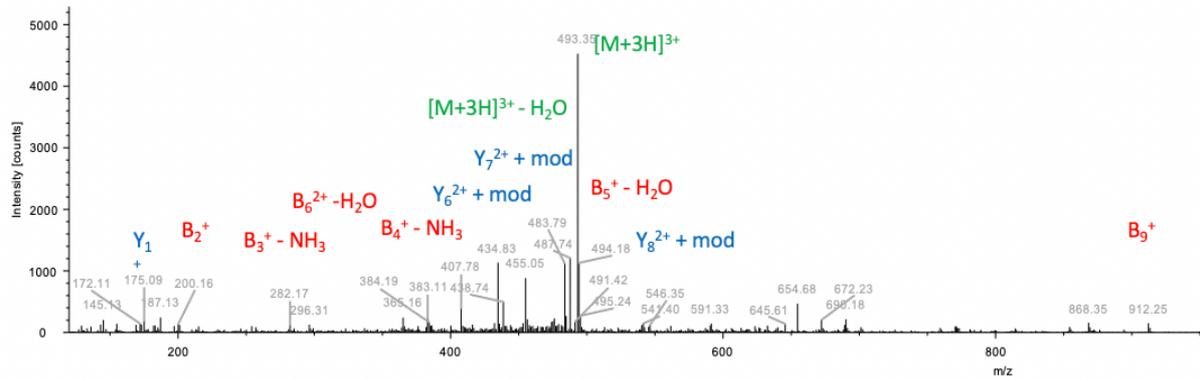


Figure S6. Pdx1.3^{K166R} after addition of R5P and (NH₄)₂SO₄: m/z = 493.5902 (z = +3), observed [M+H]⁺ = 1478.7561. Subtracting QAVTIPVMAK₉₈AR with a predicted [M+H]⁺ = 1284.7457 yields 1478.7561 – 1284.7457 = 194.0104; this modification is interpreted as structure 9.

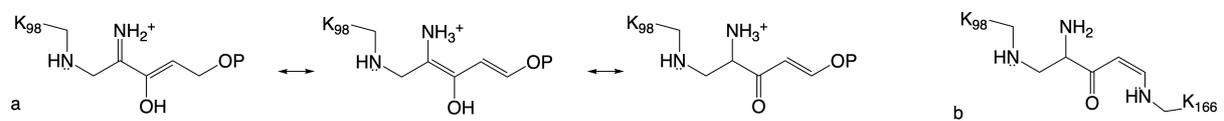


Figure S7. (a) I333 intermediate structures tested in refinement; (b) proposed I320 intermediate structure².

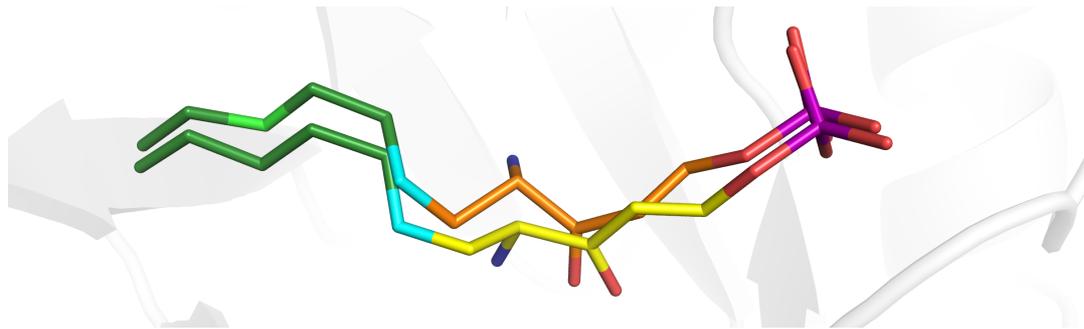


Figure S8. The Pdx1.3^{K166R}:I333 intermediate structure in 7NHE (this study, yellow) vs. the Pdx1.1^{K166R}:pre-I320 complex 5LNT (NCB, orange).

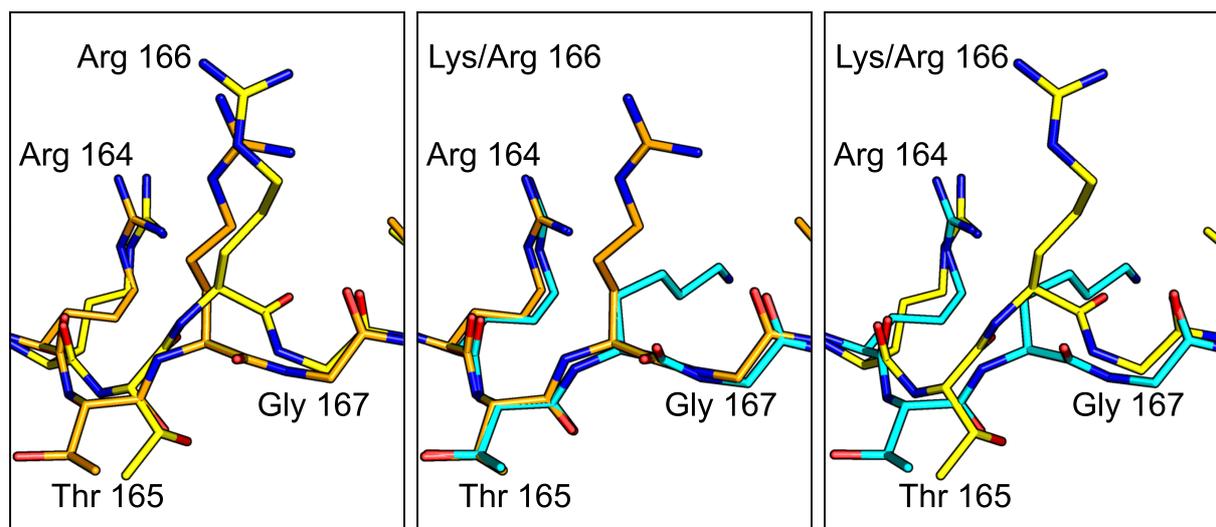


Figure S9. Pdx1.3:I333 (7NHE, this study, yellow), Pdx1.1:preI320 (5LNT, orange,²) and Pdx1.3:I320 (5LNU, blue,²)

References:

1. M. Tambasco-Studart, O. Titiz, T. Raschle, G. Forster, N. Amrhein and T. B. Fitzpatrick, *Proc Natl Acad Sci U S A*, 2005, **102**, 13687-13692.
2. M. J. Rodrigues, V. Windeisen, Y. Zhang, G. Guédez, S. Weber, M. Strohmeier, J. W. Hanes, A. Royant, G. Evans, I. Sinning, S. E. Ealick, T. P. Begley and I. Tews, *Nat Chem Biol*, 2017, **13**, 290-294.
3. K. E. Burns, Y. Xiang, C. L. Kinsland, F. W. McLafferty and T. P. Begley, *J Am Chem Soc*, 2005, **127**, 3682-3683.
4. T. Raschle, N. Amrhein and T. B. Fitzpatrick, *J Biol Chem*, 2005, **280**, 32291-32300.
5. A. J. Cox, A. J. DeWeerd and J. Linden, *American Journal of Physics*, 2002, **70**, 620-625.
6. J. Z. Porterfield and A. Zlotnick, *Virology*, 2010, **407**, 281-288.
7. A. Savitzky and J. Golay, *Anal Chem*, 1964, **36**, 1627-1239.
8. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa and P. van Mulbregt, *Nature Methods*, 2020, **17**, 261-272.
9. J. D. Hunter, *Computing in Science & Engineering*, 2007, **9**, 90-95.
10. J. W. Hanes, I. Keresztes and T. P. Begley, *Nature Chemical Biology*, 2008, **4**, 425.
11. G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter and G. Evans, *Acta Crystallographica Section D Structural Biology*, 2018, **74**, 85-97.
12. P. R. Evans and G. N. Murshudov, *Acta crystallographica. Section D, Biological crystallography*, 2013, **69**, 1204-1214.
13. A. Vagin and A. Teplyakov, *Journal of Applied Crystallography*, 1997, **30**, 1022-1025.
14. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams and P. D. Adams, *Acta Crystallogr D Struct Biol*, 2019, **75**, 861-877.
15. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta crystallographica. Section D, Biological crystallography*, 2010, **66**, 486-501.
16. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallographica Section D*, 1997, **53**, 240-255.
17. A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin and G. N. Murshudov, *Acta crystallographica. Section D, Biological crystallography*, 2012, **68**, 431-440.
18. D. von Stetten, T. Giraud, P. Carpentier, F. Sever, M. Terrien, F. Dobias, D. H. Juers, D. Flot, C. Mueller-Dieckmann, G. A. Leonard, D. de Sanctis and A. Royant, *Acta Crystallogr D Biol Crystallogr*, 2015, **71**, 15-26.
19. F. S. Dworkowski, M. A. Hough, G. Pompidor and M. R. Fuchs, *Acta Crystallogr D Biol Crystallogr*, 2015, **71**, 27-35.
20. G. C. Robinson, M. Kaufmann, C. Roux and T. B. Fitzpatrick, *Proc Natl Acad Sci U S A*, 2016, **113**, E5821-E5829.
21. G. C. Robinson, M. Kaufmann, C. Roux, J. Martinez-Font, M. Hothorn, S. Thore and T. B. Fitzpatrick, *Acta Crystallogr D Struct Biol*, 2019, **75**, 400-415.
22. A. M. Smith, W. C. Brown, E. Harms and J. L. Smith, *J Biol Chem*, 2015, **290**, 5226-5239.