TITLE: The tunable intrahexamer heteroassembly mechanism of pseudoenzyme/enzyme pair, PDX1.2/PDX1.3, with its molecular impact on vitamin B_6 regulation.

Authors

Irina Novikova[1]*, Mowei Zhou[1], Jared Shaw[1], Marcelina Parra[2], Hanjo Hellmann[2]* and James E. Evans[1,2]*

Affiliations

[1] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA

[2] School of Biological Sciences, Washington State University, Pullman, WA, 99164, USA

Corresponding Author E-mail: james.evans@pnnl.gov, irina.novikova@pnnl.gov,

Keywords

Vitamin B_6, PLP, PDX, cell-free expression, co-expression, Native Mass Spectrometry, cryo-EM, heterocomplex, electrostatics

Abstract

Plants boost the expression of pseudo-enzyme PDX1.2 under heat stress and during embryonic development. PDX1.2 positively regulates vitamin B_6 production by hetero-association with its active catalytic homologs such as PDX1.1 and PDX1.3. These heterologous interactions were found challenging to understand. For instance, the crystals of PDX1.2-PDX1.3 heterocomplexes were found to be statistically disordered and individual proteins could not be assigned. Using a combination of biochemical and structural tools, we find that the key to this phenomenon is the nature of PDX1.2 hetero-assembly with its catalytic counterparts. Using a cell-free protein synthesis approach, we were able to set up a precise control of co-expression where we systematically varied the ratios of co-produced proteins by tuning the ratios of input DNA. These were further analyzed by Native Mass Spectrometry, which elucidated that 6-8 hetero-complex species of dodecamers of variable stoichiometry are produced for each co-expression condition tested. This is in contrast to previous hypothesis of stacked inter-hexamer assembly mechanism. As proposed previously, our high-resolution Cryo-EM structure of pseudo-enzyme PDX1.2 closely mimics the fold of PDX1.3 and maintains all necessary protein-protein interactions between subunits. In PDX1.2, the altered catalytic site P1 appears perturbed in concordance with its lack of activity, while the P2 site appears largely unchanged. The most surprising finding is that we observe a complete switch in the surface electrostatics for PDX1.2. Based on the activity assays and its structure, we hypothesize that the change in electrostatic would have a significant impact on the neighboring P2 site of the PDX1.3 and influence the turnover efficiency at that site. These data suggest that pseudo-enzyme PDX1.2 rather acts as an electrostatic tuning module, that, in combination with its hetero-assembly mechanism based on random incorporation, imposes a perfect regulatory control of such important process.

Introduction

Vitamin B_6 is a central metabolite for all organisms. It functions as a cofactor for over a hundred enzymatic processes and also provides an important defense mechanism against oxidative stress[1-3]. Current knowledge indicates that pyridoxal 5'-phosphate (PLP) is the main active vitamin B_6 in the cell. Unlike bacteria, plants are able to synthesize PLP *de novo* via a very elegant and efficient mechanism executed by just two proteins - PDX1 and PDX2 (Pyridoxine Biosynthesis 1 and 2)[1,4]. These proteins form a large complex with up to 24 subunits (12 of each). The core of this complex consists of PDX1 proteins, which are the key synthases that assemble into a dodecameric structure, composed of two stacked hexamer rings[5]. This dodecamer serves as scaffold for binding additional 12 units of glutaminase PDX2. However, the nature of PDX1/PDX2 association is rather transient, and the PDX1 dodecamer is not always fully saturated with PDX2[6,7]. Phosphorylated triose and pentose sugars in combination with glutamine are used as substrates in the lysine-mediated production of PLP, which occurs within each PDX1/PDX2 pair of the complex[5]. Interestingly, the synthase activity of PDX1 proteins can be PDX2-independent if ammonium ions are supplemented[7].

In the model plant *Arabidopsis thaliana*, there are three homologs of PDX1, designated as PDX1.1, PDX1.2 and PDX1.3, and one homolog of PDX2[4]. PDX1.1, PDX1.3 and PDX2 are all catalytically active enzymes while PDX1.2 was shown to be inactive[4]. Despite the lack of synthase enzymatic activities for PDX1.2, studies showed that null mutants have an embryo lethal phenotype, and its expression is upregulated by heat and oxidative stress[8-10]. The effect of PDX1.2 on embryo development remains to be elucidated and is likely independent from the PLP production[1,8]. However, under stress conditions, the induced expression of *PDX1.2* coincides with a boost in vitamin B_6 synthesis, suggesting a positive regulatory role[9]. Previously, PDX1.2 was found to associate with its catalytic PDX1.1 and PDX1.3 proteins *in vivo* and forms complexes of high molecular weight similar to its catalytic homologs[11]. Furthermore, recombinant co-expression of PDX1.2 with either PDX1.1 or PDX1.3 showed that hetero-complexes are dodecameric in nature[9]. It was also determined that co-expression was required for the formation of such structures, and these hetero-complexes can't be reconstituted when proteins are individually expressed and then later combined.

It has been proven extremely challenging so far to work with these proteins. While the structure of PDX1.3 was solved by X-ray crystallography[5], PDX1.2 failed to crystallize on its own[12]. Current X-ray data on the PDX1.2-PDX1.3 hetero-complexes do exist but suffer from statistical disorder, where the individual contribution of the proteins cannot be dissected in such assembly[12]. Up to now, an interhexamer assembly mechanism was proposed as the likely cause of interaction where one hexameric ring of active PDX1.1 (or PDX1.3) stacks on top of a hexamer ring of inactive PDX1.2 to create a final hetero-dodecameric state[9,12] (Fig.1a).

Thus, a new approach was needed to resolve the mechanism controlling mixed PDX1 complex formation as well as the biological relevance of the interaction on PLP synthetase activity. Here, using an integrated biochemical and structural approach combining cell-free protein synthesis technology, Native Mass Spectrometry and Cryo-EM, we delve into the mechanism, stoichiometry and structure of these hetero-interactions.

Results

Cell-free co-expression of PDX1.2 and PDX1.3 proteins results in heterocomplexes of variable molecular weight. For protein production, we decided to employ cell-free protein expression pipeline using wheat germ extract[13-15] for several reasons. First, this "open-format' cell-free platform provides an opportunity for precise stoichiometry control of protein co-expression by varying the amounts of corresponding DNA templates. Second, the wheat germ protein extract being plant based also provides the closest translational environment for the production of these plant-derived proteins in terms of folding and post-translational modifications.

We initially prepared two plasmids constructs, which were used as DNA templates for cell-free protein synthesis of individual PDX1.2 and PDX1.3 proteins and co-expressed complexes. Both protein constructs were designed to carry a 3XFLAG purification tag on their N-terminus (Fig. S1). In the initial

set of cell-free translation experiments (Fig. 1), we supplemented all the translation reactions with fluorophore-labeled lysine-charged tRNA as recently described[15]. This procedure allows detection of newly-synthetized proteins in the crude mixture with no need for purification. For co-expression, different DNA template ratios (9:1, 3:1, 1:1, 1:3 and 1:9) were tested to establish that the molar ratio of expressed proteins can be controlled in a precise fashion. For instance, by in-gel quantification, the co-expression conditions 9:1, 3:1 and 1:1 correspond to protein product compositions of 3:1, 1:1 and 1:3 (Fig.1b left). The same samples were also analyzed under native conditions (Fig. 1b right). It is immediately evident that all co-expressed samples travel faster than individual PDX1.2 homomeric complexes but slower than PDX1.3 homomeric complexes. In addition, the bands for co-expressed samples appear significantly wider in size suggesting multiple species. The mobilities of these samples also appear to be dependent on the protein composition where the higher proportion of PDX1.3 in the co-expressed sample causes faster migration of the hetero-complex, and the opposite for PDX1.2.

As we employed a different expression platform more reminiscent of native conditions in comparison with the previous co-expression study in *E. coli*[9], we also wanted to verify whether cell-free co-expression was required for the formation of such complexes or if simple incubation of samples following translation and expression termination would result in intermixed complexes. Thus, PDX1.2 and PDX1.3 homomeric samples were combined and left to incubate for additional 48 hours and then re-run on Native PAGE along with all other samples (Fig. S2a). As clearly seen on the gel, no formation of intermixed complexes is observed since PDX1.2 and PDX1.3 proteins travel as independent species (Fig. S2a, last lane). Thus, co-expression was confirmed to be a critical parameter in the PDX1 hetero-complex formation, supporting the previous discoveries[9,12]. However, our overall findings do not support the proposed interhexamer assembly mechanism where a defined single band would be expected. It appears that PDX1.2 and PDX1.3 rather interact intramolecularly to form hetero-complexes of similar architecture as homo-assemblies, but with variable stoichiometry.

The same experiments were conducted using PDX1.1 (instead of PDX1.3), and the same pseudoezyme-enzyme assembly behavior was seen and suggests the same assembly mechanism for

PDX1.2 interaction with both PDX1.2 and PDX1.3 (Fig.S2b). Since PDX1.3 is the dominant homolog in *A. thaliana*[16,17] and was previously found to be more impacted by PDX1.2[9], we have limited our further structural survey to PDX1.2 and PDX1.3 only to simplify the variable space.

Post-translational modifications. We further scaled-up the protein synthesis on select complexes. Denaturing SDS and Native PAGE analysis was repeated on purified complexes and retention of their native properties (Fig. S3). Liquid chromatography – mass spectrometry (LCMS) peptide mapping of PDX1.2 and PDX1.3 homomeric samples showed the correct protein sequences as top hits with near complete sequence coverage of 98% and higher (Fig. S4). In the previous study, *E. coli* expressed plant PDX1.2 was found to carry an acetyl group on the alanine (A5)[9]. Such modification was not detected in the cell-free produced PDX1.2 and PDX1.3. Nevertheless, N-terminal acetylations were detected but were located on the first methionine residue, preceding the tag (Fig. S5). In all homomeric PDX1.2 and PDX1.3 samples along with all three co-expressed samples, roughly half of the protein population was acetylated (Fig. S6). These modifications, which are incorporated in the protein during its synthesis by ribosome-bound N-terminal acetyltransferases (NATs), are the most abundant, especially in higher eukaryotes[18]. In general, such acetyl modifications are associated with various protein degradation and stabilization mechanisms inside the cell[19]. Nevertheless, no preferential acetylation profile was found across the samples, which supports the idea that these modifications are likely irrelevant to PLP regulation itself. Finally, the measured masses for PDX1.2 and PDX1.3 were determined to be 37339.7 and 36405.5 Da, respectively, which is consistent with their theoretical predicted values of 37339.6 Da and 36405.3 Da (monoisotopic) (see the sequence information in Fig.S1).

Pseudo-enzyme PDX1.2 displays positive regulatory properties that can be tuned. While the kinetic parameters for individual PDX1.2 and PDX1.3 dodecamers (including co-expressed dodecamer at 0.44:1 stoichiometry ratio) were already reported[9], here, we intended to compare the relative activities across all samples in order to understand how the PDX1.2:PDX1.3 ratio in hetero-assembly affects their

enzymatic performance. We employed a diagnostic assay where PLP is used as a cofactor molecule for apo-enzyme chemical transformations, the products of which are detected by a colorimetric approach (Fig.2a). Under these assay conditions with saturating substrate amounts, the specific activity of active enzyme PDX1.3 was determined to be 793 µmoles/min/mg protein. As expected, pseudo-enzyme PDX1.2 was found to be inactive and displays a slightly inhibitory effect. When proteins are co-expressed at conditions 3:1 and 1:1, the presence of PDX1.2 heterocomplex positively impacts PLP synthesis with an increase of up to 120% in relative enzymatic activity. However, at 9:1, where 75% of the total protein is PDX1.2, the heterocomplex appears to display a negative regulatory effect as indicated by a decline in the relative activity to 67% that observed for homomeric PDX1.3. However, the measured activities represent a sum of PDX1.2 and PDX1.3. If these relative enzymatic activities are normalized based on the amount of catalytically active PDX1.3 in each co-expressed sample, calculated by in-gel quantification, the positive regulatory impact of PDX1.2 is observed for every condition. The obvious trend emerges where the higher ratio of PDX1.2 incorporation coincides with higher impact on PDX1.3 catalytic activity with the corresponding outputs such as: 3-fold increase for co-expression condition 9:1, 2.4-fold increase for 3:1 and 1.5-fold increase for 1:1.

Each co-expressed PDX sample generates a population of 6-8 species. Native MS measurements, where we electrosprayed the protein complexes under nondenaturing conditions in 200 mM ammonium acetate solution (pH 7), determined that homomeric PDX1.2 and PDX1.3 proteins form 449 kDa and 437 kDa dodecamers respectively (Fig. 2b). When co-expressed, each condition generated a population of 6-8 species of heterocomplexes of variable stoichiometry that were all dodecameric in their fold. For instance, at 9:1, PDX1.2 represents the major constituent in the hetero-dodecamers with stoichiometry values of 6:6, 7:5, 8:4, 9:3,10:2,11:1 and 12:0. An example of raw data, which includes the full mass spectrum and zoomed-in spectrum of the 12-mer, is in Fig. 2c. At the 3:1 co-expression condition, there is a shift to 3:9 -11:1 stoichiometry range, while, at 1:1, we are limited to 0:12-8:4 dodecamer pool (Fig. 2b). As the result, by varying DNA ratio (and thus protein ratios), the full range

from 0:12 to 12:0 has been sampled. The observed stoichiometry values are consistent with the amounts of proteins produced in each sample (see Fig.1b and Fig.2b).

Interestingly, significant amount of free monomer and dimer species were also observed for PDX1.3 in Native MS data (Fig. S7). In the co-expressed samples, the relative abundance of PDX1.3 monomers and dimers was also found to be positively correlated with the ratio of PDX1.3 in the hetero-complex. However, Native PAGE analysis of samples in the same ammonium acetate solution right before gas-induced dissociation did not show any disassembly (Fig. S8), suggesting that PDX1.3 is inherently sensitive to the gas-phase dissociation. Unlike PDX1.3, PDX1.2 displays a different behavior and appears stable under the same experimental conditions. Heterocomplexes containing higher ratios of PDX1.2 also display higher stability based on fewer monomer dissociation events.

Structure of PDX1.2 reveals an inverted electrostatic surface potential relative PDX1.3. As recently reported, PDX1.2 protein resists crystallization protocols while PDX1.3 homomers and PDX1.2/PDX1.3 heteromers are compatible with macromolecular crystallography[12]. Interestingly, unlike PDX1.3 homo-dodecamers, the crystals of PDX1.2/PDX1.3 heteromers were found to suffer from a statistical disorder, where the positions of individual proteins were not distinguished despite multiple and very exhaustive approaches used[12]. As stated by the authors, the X-ray data were found to be ambiguous, prohibiting the determination of assembly mechanism. In attempt to gain further understanding of such intriguing system, we first employed single particle Cryo-EM to resolve the structure of PDX1.2. Purified PDX1.2 samples were first vitrified on EM grids and imaged in super-resolution mode using Titan Krios (Fig. S9). A homogenous distribution of particles in various orientations are clearly visible in the thin layer of ice, and several rounds of 2D classification produced well-defined 2D classes where a stacked 2-ring architecture was clearly observed and consistent with the dodecameric fold, known for catalytically active PDX1.1 and PDX1.3. The final 3D volume was reconstructed at 3.2Å resolution, where the PDX1.2 homology model was then docked and real space refined (Fig.S10, Table S1).

Residues 29-288 were clearly resolved while N- (aas 1-28) and C-terminal (aas 289-313) regions appear rather flexible due to the lack of associated densities in the volume.

PDX1.3 implements an extremely efficient lysine relay mechanism where all catalytic action happens within a single domain (catalytic sites P1 and P2). Two key residues K98 and K166 trap the substrates, covalently tether the intermediates and then shuttle the product out[5,20]. We compared the obtained Cryo-EM structure of inactive PDX1.2 with several crystallographic structures of PDX1.3 available at different stages of such action: precursor binding stage PDX1.3-R5P (PDB:5lns), intermediate stage PDX1.3-I_{320} (PDB:5lnu), post-intermediate PDX1.3-I_{320}-G3P (PDB:5lnw) and final product stage where PLP is still covalently bound to K166 (PDB:5lnr) (Fig.3a). We found that the overall fold of PDX1.2 is very similar to PDX1.3, and there is no deviation of the structure at the edges or interfaces - presumably to preserve proper interactions with neighboring subunits. The best overall fit is with the X-ray structure of PDX1.3-PLP, which has P1 site unoccupied. This is consistent with the fact that there were no ligands supplemented in the PDX1.2 sample as it is a known (and shown in Fig. 2a) to be inactive. The **α8'** helix of PDX1.2 was found to fully trace the α8' helix of PDX1.3 in PDX1.3-PLP state (5lnr).

Major differences between PDX1.2 and PDX1.3 are observed in the area, surrounding the catalytic site P1. Specifically, α2' helix and **ß6-α6** loop are significantly altered in PDX1.2. The **α2'** helix, present in PDX1.3, is absent in PDX1.2 due to the amino acid truncation in that region and is substituted by a loop that points slightly outwards. The amino acid changes in **ß6-α6** loop, including an additional amino acid insertion, create a slight kink in its architecture towards the center of the dodecamer, resulting in the potential weakening/aborting of the phosphate binding at this site. Key lysine residues K98 and K166 which form covalent adducts with the intermediates of PLP formation in PDX1.3 are mutated to R100 and Q168 respectively in PDX1.2. While R100 is positioned similar to K98 and can potentially form the imine bond with R5P as well, Q168 in PDX1.2 (in place of K166 in PDX1.3) points towards the P1 site similar to the intermediate state of the PLP synthesis instead of pointing to P2 as

observed in the priming state PDX1.3-R5P state. The same positional occupancy of Q168 was assigned to the composite assignment of Q168 and K166 in the crystallized PDX1.2-PDX1.3 co-expressed complex where individual protein positions were not distinguished[12]. Such structural alterations in the P1 site in PDX1.2 are in agreement with the experimental data, showing the loss of catalytic potential. Unlike P1, the catalytic site P2 in PDX1.2 appears to be undisturbed. Key arginine residues R149, R156, R157 in PDX1.2 (corresponding to R147, R154, R155 in PDX1.3) are positioned properly for coordinating the phosphate anions at this site. The only aa difference at F143 (instead of H142 for PDX1.3) might weaken the water coordination at this site.

We also obtained a high-resolution Cryo-EM dataset for the co-complex 9:1 and referred herein as PDXcoexp (Fig.S12). Well-defined two-ring fold seems visually identical to the PDX1.2 dodecamer. Various heterogenous, non-uniform and local refinements did not sort out the potential subclasses, which is not surprising due to the diversity of co-expressed species generated, the similarity of PDX1.2 and PDX1.3 folds and the small size difference between their monomer molecular weights. As the result, PDXcoexp particle dataset was processed as a single 3D class, and non-uniform refinement yielded the final map. Both the PDX1.2 dodecamer or PDX1.3 dodecamer can be fitted independent with high validation scores (Fig. S13, Table S1) similar to the results reported recently for X-ray structure[12]. However, major structural differences across two proteins are defined by presence/absence of **α2'** helix and a **ß6-α6** loop kink. As such, the PDX1.2 protein was found to be a better fit in these two areas consistent with its more prominent composition in this co-complex. The comparison of the atomic model of PDX1.2, fitted in PDXcoexp cryoEM volume, with the PDX1.2 itself did not show any differences and thus did not provide any additional clues in its regulatory impact.

Based on the structural analysis of overall shape and the comparison of areas around the catalytic sites between PDX1.2 and PDXcoexp, there is no obvious benefit in keeping such a defective homolog. We did not find any additional interactions at the interfaces between subunits which might weaken or strengthen the overall complex (Fig. S14). Then why would plants keep functionally inactive PDX1.2 and

what is its impact? Hydrophobicity plots of PDX1.2 and PDX1.3 do not show any differences between the positions of polar and hydrophobic residues (Fig. S15). Interestingly, the most stunning differences are observed when one compares the electrostatic surface potentials between the two (Fig. 3b). The surface potential of outer and inner regions of the PDX1.2 dodecamer, which spans the P1 and P2 sites, looks inverted relative to PDX1.3, where negative charges are in place of the positively charged regions. The catalytic site P1 is buried inside the center of the individual PDX1.3 subunit, and it is therefore hard to imagine how the surface potential of PDX1.2 might potentially impact it. Instead, in concordance with the observed mechanism of assembly, we hypothesize that the most profound effect would be on the catalytic site P2. Site P2 is located at the edge of the PDX1.3 subunit and shares two subunit-subunit interfaces: side-by-side **I1'** within the hexamer ring and bottom-bottom **I3** at the ring-ring interface (Fig. 4a). To understand it better, we constructed several PDX1.2/PDX1.3 assembly models around the catalytic site P2 of PDX1.3 (Fig. 4b-d). Lateral PDX1.2 incorporation can create a local supercharging effect on **ß4-α4** loop, which could potentially act on the nitrogen atom of its pyridoxine ring (Fig.4b). This supercharging effect could promote imine bond breaking between PLP and K166, help in the translocation and assist with the exit of the final product. The supercharging of the same loop could also be possible when PDX1.2 shares a bottom-bottom interface (Fig. 4c). The positioning of PDX1.2 on the bottom also slightly weakens the charge state around the phosphate binding site in **α4**. As such, one can imagine that the both side and bottom positioning of PDX1.2 (Fig.4d) could explain enhanced PDX1.3 activity in the heterocomplex and also explain the benefit of having a pseudoenzyme that associates with active PDX1.3 in a random positional and stochiometric fashion.

Conclusions

Pseudoenzymes still represent a largely uncharted territory with much to learn about the additional layer of regulatory control they impose in nature. Mainly learned from select protein families

of kinases and phosphatases, many pseudoenzymes retain the structural fold of their catalytically active partners and rather function as protein interaction modules[21]. Structural studies were the key tools in exploring and clarifying the molecular mechanisms of such systems, and they uncovered their roles as allosteric regulators, scaffolds, molecular switches and substrate competitors[22-24]. To understand the impact and assembly of pseudoenzyme PDX1.2 with its catalytic homolog PDX1.3, we employed an integrated approach, which includes the cell-free expression, Native MS and Cryo-EM.

Our findings demonstrate that PDX1.2 co-assembly with its catalytic partner PDX1.3 is based on random subunit incorporation at different locations, driven by the molar ratios of individual components during co-expression conditions (Fig.5). The entire spectrum of all possible stoichiometry combinations (from 12:0 to 0:12) has been recorded, suggesting the lack of any preferences for a specific arrangement. This mechanism of assembly supports previously recorded statistical disorder and the inability of protein assignment in the X-ray crystal for the co-expressed samples[12]. The same explains the inability of identifying any unique 3D classes in our Cryo-EM data for PDXcoexp sample.

We also observe the inherent sensitivity of PDX1.3 to disassemble upon transition and activation in the gas-phase. The observed loss of its unfolded monomeric subunits in homo- and hetero-complexes is possible due to the differential charge surface properties PDX1.3 as compared to PDX1.2. Recent studies have shown that charge plays important roles in gas-phase disassembly of protein complexes[25,26]. This behavior is in agreement with the significant differences in computed electrostatic surface potentials between the two, as described above.

It has also been shown that surface charge-charge interactions can be redesigned to increase the protein stability[27]. As PDX1.2 expression is induced under heat stress, this process might serve as a heat adaption mechanism. While it is hard to account for all possible factors present in plant cells under stress conditions, the most interesting aspect is that PDX1.2 regulatory impact on the PDX1.3 activity is observed *in vivo* and *in vitro* and is tunable based on its stoichiometry arrangements. Our *in vitro* enzymatic assays suggest that the key regulation function of PDX1.2 is defined by the co-complex itself, otherwise no positive, as well as tunable, impact of PDX1.2 on PDX1.3 catalysis would be manifested.

The observed switch in the charge distribution is a fascinating discovery that PDX1.2 might have evolved and been repurposed as an allosteric activator. Extensive and comparative MD simulations will be needed in the future to verify the supercharging effect on site P2 for PDX1.3. At this point, the most current (however, still limited) understanding about allosteric regulation of pseudoenzymes spans around their stabilizations of the active sites for their active homologs[21,22], thus it is possible that such mechanism might be the key to explain the differential relative enzymatic activities between PDX1.3 and its co-complexes with PDX1.2.

This hypothesized strategy, in combination with the random incorporation assembly mechanism, empowers simple, tunable and gentle approach to accelerate vitamin B_6 production by active PDX1.3 when it is needed without adjusting the PDX1.3 gene expression.

Methods

Plasmid construction. Gene sequences of PDX1.2 and PDX1.3 were sourced as described previously[17]. The clones were amplified by PCR to include the purification tag and then sub-cloned using Gibson Assembly in the designated vector pEU for the cell-free protein synthesis[15,28]. All reagents for PCR and Gibson reactions such as the Q5 Hot Start High-Fidelity 2X Master Mix and Gibson Assembly Master Mix were acquired from New England Biolabs. DNA primers were from ThermoFisher Scientific. For expression of PDX1.2 and PDX1.3 proteins, two vectors were constructed: pEU_3XFLAG_PDX1.2 and pEU_3XFLAG_PDX1.3, where both genes were fused with the 3XFLAG purification tag on their N-terminus (Fig. S1). The sequences of all plasmids used in this study were verified by Sanger sequencing (MCLAB).

Cell-free expression/coexpression and purification of PDX proteins. Protein synthesis was carried out using Wheat Germ Protein Research Kits WEPRO7240 from CellFree Sciences. For co-expression, stock plasmids were prepared containing various amounts of pEU_3XFLAG_PDX1.2 and

pEU_3XFLAG_PDX1.3. For example, in the case of co-expression at 9:1 DNA template ratio: 9 μl of 1 μg/μl pEU_3XFLAG_PDX1.2 was mixed with 1 μl of 1 μg/μl pEU_3XFLAG_PDX1.3 before the transcription step. The MINI-scale protein synthesis in the presence of fluorophore-charged lysine tRNA was conducted at the following conditions: (i) for transcription, 1 μl of 1 μg/μl stock plasmid was mixed with 9 μl of Transcription premix solution (1x Transcription buffer, 2.5 mM NTP, 8 U/μl SP6 polymerase, 8 U/μl RNase inhibitor) and left incubated at 37℃ for 4 hours, (ii) for translation, 2 μl of the resulting mRNA was mixed with 1 μl of FluoroTect $Green_{Lys}$ (Promega) and 3 μl of the wheat germ extract and transferred under 50 μl of the translation buffer (1x SUB-AMIX SGC), placed in the 96-well half-area plate. The translations were carried out overnight at 15℃ away from light at the vibration-free setting on the Eppendorf ThermoMixer C.

The MAXI-scale protein synthesis was performed using the robotic system Protemist DT II from CellFree Sciences. All steps were done at 4℃ and at 140 rpm. The buffer exchange step of the crude extract (in order to remove DTT) was skipped. Total of 600 μl of ANTI-M2 affinity resin (50% suspension, pre-equilibrated in 1xTBS buffer, Sigma) was used per each 6 ml translation reaction, which accounts for translation buffer. Protein binding to the resin was carried out for 1 hour. The resin was subsequently washed three times with 2.5 ml of 1xTBS buffer for 5 min each time. The protein products were eluted twice with the Elution Buffer (1xTBS buffer, 100 μg/ml 3XFLAG peptide, 4 mM DTT) – 1 ml each time for 30 min. Proteins were further concentrated and buffer-exchanged to the Storage Buffer (1xTBS with 4 mM DTT) using 30 kDa Amicon Ultra-2 centrifugal filter units. The concentrated samples were flash-frozen in liquid nitrogen in 10 μl aliquots and then transferred to -80℃ for storage.

General PAGE electrophoresis conditions employed in this work followed standard procedures. Bioanalyzer runs were performed with the Agilent Bioanalyzer 2100 system using the P80 kit according to the manufacturer's guidelines.

Native Mass Spectrometry. Purified PDX complexes at 0.2-0.4 mg/mL concentrations (10 μL each in 1xTBS with 4 mM DTT) were buffer exchanged into 200 mM ammonium acetate using Zeba 75

µL microspin columns from Thermofisher following manufacturer's instructions. Proteins were loaded into pulled glass capillaries (GlassTip™, part number BG12692N20, Scientific Instrument Services, Inc., Ringoes, NJ) with gel loading pipet tips. A platinum wire connected to the electrospray voltage was inserted into the capillary and is in contact with the protein solution. 1 kV was applied at the capillary to sustain the nanoelectrospray on a Thermo Q-Exactive UHMR mass spectrometer. The source was kept at 200 °C, with S-lens set to 100%. In-source fragmentation of 100 V and in-source trapping of 50 V were used for optimum de-solvation without extensive gas-phase dissociation of the complexes. Even with in-source fragmentation/trapping voltages off, a significant amount of PDX1.3 1-mers was still observed, implying that the PDX1.3 monomers were not exclusively generated by the desolvation conditions. HCD gas flow was set to 2. For each final spectrum, 500 microscans were averaged. Mass calibration was performed using cesium iodide clusters. To obtain the mass distribution of the PDX dodecameric complexes, mass spectra were processed using UniDec[29]. Original data within m/z 8000-11000 were extracted with curved baseline subtraction, Gaussian smoothing 2.0, and binned every 3 m/z. Deconvolution was restricted to sampling mass every 10 Da with peak FWHM of 8 Th.

Activity assay. For quantitative PLP detection, we used Enzymatic Vitamin B6 assay from A/C diagnostics. From an initial set of experiments using active enzyme PDX1.3, we determined that 5-10 ug/ml was found optimal protein concentration for conducting this assay where the OD675 signal falls within the linear PLP detection range of 0-200 nmol/L (data not shown).

To compare relative enzymatic activities, all PDX samples (individual or co-expressed) were first diluted to 12.5 ug/ml concentration in TBS buffer. Then 5 ul of 12.5 ug/ml protein were combined with 5 ul of 2XAssay buffer (50 mM Tris, pH 8.0**,** 1 mM ribose 5-phosphate**,** 2 mM D-glyceraldehyde 3-phosphate**,** 20 mM ammonium sulfate) and incubated away from light for 30 min at 30℃ in the Eppendorf thermocycler. For the calibration curve, PLP (Sigma, Cat # 82870) was diluted in a serial manner to 400 nM, 200 nM, 100 nM, 50 nM, 25 nM concentration in TBS buffer, further combined with

an equal volume of 2XAssay buffer and incubated away from light for 30 min at 30℃ along with PDX samples for consistency.

To quantify the amount of PLP in the mixture, 2.5 ul of each sample were further added to the 37.5 ul of Working Binding Assay buffer (with apo-enzyme), provided by the assay kit from A/C diagnostics and placed in the Corning NBS 384-well microplate. The mixture was incubated at 37℃ for 30 min at 750 rpm on the Eppendorf Thermomixer C away from light. Then 20 ul of Working Assay buffer (from the kit) was added to the mixture, and incubation continued for additional 20 min. This was followed by the addition of 6.25 ul of Chromogen RI and 3.75 ul of Chromogen RII and 10 min incubation. The color development was measured at 675 nm using the Infinite M200 PRO microplate reader by Tecan. For accuracy, all samples and calibration standards were assayed in triplicate.

Cryo-EM data acquisition and processing. Quantifoil grids (658-300-AU, Ted Pella) were first glow-discharged at 15 mA for 1 min using PELCO easiGlow (Ted Pella). The grids were then transferred to the plunge-freezer Leica EM GP, brought to 85% humidity at 25℃. Then 3 μl of protein sample (0.2 mg/ml, in TBS buffer with 4 mM DTT) was pipetted on the carbon side and blotted for 3 s, plunge-frozen in the liquid ethane and transfer to liquid nitrogen. The frozen samples were imaged on a 300 keV FEI Titan Krios equipped with a Gatan Image Filter at Pacific Northwest Center for Cryo-EM at OHSU (https://pncc.labworks.org) and a Gatan K2 summit direct electron detector. The data was collected using automated acquisition software SerialEM[30]. For PDX1.2 sample, 8,106 -resolution movies with 60 frames each at a dosage of 1.5 electrons per $Å^2$ per frame and the binned pixel size of 0.253 were collected. For PDXcoexp, 6,904 super-resolution movies with 50 frames each at a dosage of 2 electrons per $Å^2$ per frame and the binned pixel size of 0.253 Å were acquired. All movies were further processed using cryoSPARC v2 software[31] and were corrected for full frame motion using MotionCor2[32] and CTF using CTFFIND4[33]. For PDX1.2 sample, a total of 787,982 particles were identified from template-based auto-picking algorithm. Three rounds of reference-free 2D classification narrowed the pool to 265,224 particles. These were used to generate the ab initio 3D map as a reference. Homogenous refinement in

cryoSPARC v2 with D6 symmetry produced 3.2 Å reconstruction. For PDXcoexp sample, a total of 510,660 particles were identified from template-based auto-picking algorithm. Three rounds of reference-free 2D classification narrowed the pool to 286,642 particles. These were used to generate the ab initio 3D map as a reference. Homogenous refinement in cryosparc v2 with D6 symmetry produced 3.2 Å reconstruction while non-uniform refinement (BETA) yielded the 3.16 Å map. Local resolution calculations were performed in Relion 3[34].

Atomic modeling. A homology model of PDX1.2 monomer was generated using HHPRED[35] and MODELLER[36] using comparative modeling against known structures for PDX from other organisms (2NV1, 5LNR, 4JDY, 2YZR and 3O07). The Cryo-EM map and the aligned initial model for a single subunit were imported to PHENIX[37], docked using *Phenix.DockInMap*, mapped out the density region surrounding the monomer unit via *Phenix.MapBox*. Several cycles of *Phenix.RealspaceRefine* has been carried out to refine the atomic model for the individual monomer. Map symmetry parameters were then applied on the real space refined monomer model using *Phenix.ApplyNCSoperators* to generate the 12-mer, which was further gone through additional round of *Phenix.RealspaceRefine*. Final refinements statistics and validation scores are presented in Table S1. The final models and maps were uploaded to the PDB and EMDB databases under the accession numbers: **6PCJ** and **EMD-20302** for PDX1.2; and **6PCN** and **EMD-20303** for PDXcoexp. Multiscale models, structural comparisons, structure-based alignments and computing of electrostatic potentials were conducted in Chimera[38]. Protein-protein interactions analysis was perfomed by PDBsum server[39].

Acknowledgements.

This work was supported by DOE-BER Mesoscale to Molecules Bioimaging Project FWP# 66382 and EMSL Strategic Science Area Projects 50188 and 50165. A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL

(grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

We thank Carrie Nicora, Karl Weitz, Anil Shukla, Ronald Moore, and Rui Zhao for helping with the LCMS experiments.

Author contributions.

JEE and IVN devised experiments for the study. IVN performed molecular biology, cell-free expression, cryoEM data processing and atomic modeling. MZ performed all Mass Spectrometry data collection and analysis. JS assisted with the Native MS setup. MP and HH provided clones of PDX and assisted with analysis of the resulting proteins. MP also performed a portion of cell-free experiments. IVN wrote initial manuscript draft but all authors contributed to writing the manuscript and approval of final version.

Competing Interests: The authors declare no competing interests or conflicts of interest.

References:

1. Parra, M., Stahl, S. & Hellmann, H. Vitamin B(6) and Its Role in Cell Metabolism and Physiology. Cells 7, doi:10.3390/cells7070084 (2018).
2. Mooney, S., Leuendorf, J. E., Hendrickson, C. & Hellmann, H. Vitamin B6: a long known compound of surprising complexity. Molecules 14, 329-351, doi:10.3390/molecules14010329 (2009).
3. Hsu, C. C. et al. Role of vitamin B6 status on antioxidant defenses, glutathione, and related enzyme activities in mice with homocysteine-induced oxidative stress. Food Nutr Res 59, 25702, doi:10.3402/fnr.v59.25702 (2015).
4. Tambasco-Studart, M. et al. Vitamin B6 biosynthesis in higher plants. Proc Natl Acad Sci U S A 102, 13687-13692, doi:10.1073/pnas.0506228102 (2005).
5. Rodrigues, M. J. et al. Lysine relay mechanism coordinates intermediate transfer in vitamin B6 biosynthesis. Nat Chem Biol 13, 290-294, doi:10.1038/nchembio.2273 (2017).
6. Guedez, G. et al. Assembly of the eukaryotic PLP-synthase complex from Plasmodium and activation of the Pdx1 enzyme. Structure 20, 172-184, doi:10.1016/j.str.2011.11.015 (2012).
7. Raschle, T., Amrhein, N. & Fitzpatrick, T. B. On the two components of pyridoxal 5'-phosphate synthase from Bacillus subtilis. J Biol Chem 280, 32291-32300, doi:10.1074/jbc.M501356200 (2005).
8. Leuendorf, J. E., Mooney, S. L., Chen, L. & Hellmann, H. A. Arabidopsis thaliana PDX1.2 is critical for embryo development and heat shock tolerance. Planta 240, 137-146, doi:10.1007/s00425-014-2069-3 (2014).
9. Moccand, C. et al. The pseudoenzyme PDX1.2 boosts vitamin B6 biosynthesis under heat and oxidative stress in Arabidopsis. J Biol Chem 289, 8203-8216, doi:10.1074/jbc.M113.540526 (2014).

10. Denslow, S. A., Rueschhoff, E. E. & Daub, M. E. Regulation of the Arabidopsis thaliana vitamin B6 biosynthesis genes by abiotic stress. Plant Physiol Biochem 45, 152-161, doi:10.1016/j.plaphy.2007.01.007 (2007).
11. Leuendorf, J. E., Osorio, S., Szewczyk, A., Fernie, A. R. & Hellmann, H. Complex assembly and metabolic profiling of Arabidopsis thaliana plants overexpressing vitamin B(6) biosynthesis proteins. Mol Plant 3, 890-903, doi:10.1093/mp/ssq041 (2010).
12. Robinson, G. C. et al. Crystal structure of the pseudoenzyme PDX1.2 in complex with its cognate enzyme PDX1.3: a total eclipse. Acta Crystallogr D Struct Biol 75, 400-415, doi:10.1107/S2059798319002912 (2019).
13. Madin, K., Sawasaki, T., Ogasawara, T. & Endo, Y. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. Proc Natl Acad Sci U S A 97, 559-564, doi:10.1073/pnas.97.2.559 (2000).
14. Sawasaki, T. et al. A bilayer cell-free protein synthesis system for high-throughput screening of gene products. FEBS Lett 514, 102-105 (2002).
15. Novikova, I. V. et al. Protein structural biology using cell-free platform from wheat germ. Adv Struct Chem Imaging 4, 13, doi:10.1186/s40679-018-0062-9 (2018).
16. Titiz, O. et al. PDX1 is essential for vitamin B6 biosynthesis, development and stress tolerance in Arabidopsis. Plant J 48, 933-946, doi:10.1111/j.1365-313X.2006.02928.x (2006).
17. Wagner, S. et al. Analysis of the Arabidopsis rsr4-1/pdx1-3 mutant reveals the critical function of the PDX1 protein family in metabolism, development, and vitamin B6 biosynthesis. Plant Cell 18, 1722-1735, doi:10.1105/tpc.105.036269 (2006).
18. Arnesen, T. Towards a functional understanding of protein N-terminal acetylation. PLoS Biol 9, e1001074, doi:10.1371/journal.pbio.1001074 (2011).
19. Hwang, C. S., Shemorry, A. & Varshavsky, A. N-terminal acetylation of cellular proteins creates specific degradation signals. Science 327, 973-977, doi:10.1126/science.1183147 (2010).

20. Robinson, G. C., Kaufmann, M., Roux, C. & Fitzpatrick, T. B. Structural definition of the lysine swing in Arabidopsis thaliana PDX1: Intermediate channeling facilitating vitamin B6 biosynthesis. Proc Natl Acad Sci U S A 113, E5821-E5829, doi:10.1073/pnas.1608125113 (2016).
21. Murphy, J. M., Mace, P. D. & Eyers, P. A. Live and let die: insights into pseudoenzyme mechanisms from structure. Curr Opin Struct Biol 47, 95-104, doi:10.1016/j.sbi.2017.07.004 (2017).
22. Yu, J. W., Jeffrey, P. D. & Shi, Y. Mechanism of procaspase-8 activation by c-FLIPL. Proc Natl Acad Sci U S A 106, 8169-8174, doi:10.1073/pnas.0812453106 (2009).
23. Murphy, J. M. et al. The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. Immunity 39, 443-453, doi:10.1016/j.immuni.2013.06.018 (2013).
24. Schimpl, M. et al. Human YKL-39 is a pseudo-chitinase with retained chitooligosaccharide-binding properties. Biochem J 446, 149-157, doi:10.1042/BJ20120377 (2012).
25. Leney, A. C. Subunit pI Can Influence Protein Complex Dissociation Characteristics. J Am Soc Mass Spectrom, doi:10.1007/s13361-019-02198-3 (2019).
26. Hall, Z., Hernandez, H., Marsh, J. A., Teichmann, S. A. & Robinson, C. V. The role of salt bridges, charge density, and subunit flexibility in determining disassembly routes of protein complexes. Structure 21, 1325-1337, doi:10.1016/j.str.2013.06.004 (2013).
27. Strickler, S. S. et al. Protein stability and surface electrostatics: a charged relationship. Biochemistry-Us 45, 2761-2766, doi:10.1021/bi0600143 (2006).
28. Sawasaki, T., Hasegawa, Y., Tsuchimochi, M., Kasahara, Y. & Endo, Y. Construction of an efficient expression vector for coupled transcription/translation in a wheat germ cell-free system. Nucleic Acids Symp Ser, 9-10 (2000).
29. Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem 87, 4370-4376, doi:10.1021/acs.analchem.5b00140 (2015).

30. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 152, 36-51, doi:10.1016/j.jsb.2005.07.007 (2005).
31. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290-296, doi:10.1038/nmeth.4169 (2017).
32. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331-332, doi:10.1038/nmeth.4193 (2017).
33. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol 192, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).
34. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7, doi:10.7554/eLife.42166 (2018).
35. Zimmermann, L. et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol 430, 2237-2243, doi:10.1016/j.jmb.2017.12.007 (2018).
36. Webb, B. & Sali, A. Protein Structure Modeling with MODELLER. Methods Mol Biol 1654, 39-54, doi:10.1007/978-1-4939-7231-9_4 (2017).
37. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. Methods 55, 94-106, doi:10.1016/j.ymeth.2011.07.005 (2011).
38. Pettersen, E. F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612, doi:10.1002/jcc.20084 (2004).
39. Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. Protein Sci 27, 129-134, doi:10.1002/pro.3289 (2018).
40. Park, J. et al. Informed-Proteomics: open-source software package for top-down proteomics. Nat Methods 14, 909-914, doi:10.1038/nmeth.4388 (2017).

Figure Legends:

Fig.1. The pseudoenzyme PDX1.2 assembly with its catalytic pier PDX1.3. **a**. The hypothesized mechanism of its assembly. **b**. PAGE analysis of synthetized PDX proteins in the crude wheat germ mixture. Newly-synthesized proteins are detected by fluorescence. Control sample is a negative control translation (no DNA template added). Co-expressed conditions are denoted as 9:1, 3:1, 1:1, 1:3 and 1:9, where these numbers correspond to DNA template ratio of PDX1.2 to PDX1.3. PAGE data include a denaturing SDS-PAGE gel on the left and Native PAGE on the right. F-tRNA stands for fluorophore-labeled lysine-charged tRNA. In the co-expressed samples, protein molar ratios, determined by in-gel quantifications, are shown in red.

Fig.2. Structural and enzymatic characterization of individual and co-expressed PDX proteins. **a**. Histograms showing the enzymatic activities. **b**. Deconvoluted mass spectra of native PDX dodecamer complexes. The peaks are annotated with the compositions of the PDX dodecamers following the format - "number of PDX1.2 subunits:PDX1.3 subunits". **c**. An example of raw Native MS data for the coexpression conditions 9:1.

Fig.3. The comparison of the cryoEM atomic model PDX1.2 versus the crystal structure of PDX1.3. **a**. The comparison of individual subunit of PDX1.2 versus four different crystal structures, corresponding to the different stages of PLP synthesis at active sites P1 and P2. The PDX1.2 monomer subunit is in yellow, PDX1.3 subunit is in magenta. Key residues in PDX1.3 catalysis, K98 and K166, are shown in light blue. Mutated residues in PDX1.2 include R100 and Q168 (in yellow). Substrate, intermediate products and PLP are colored by heteroatom. For structure-based sequence alignments and annotations, see Fig.S11. **b**. Coulombic electrostatic potential maps. The middle images show the sliced view to show the interior potential of the complex.

Fig.4. The proposed electrostatic reorganization around the catalytic site P2 in the co-complexes (**b**,**c**,**d**) in comparison with PDX1.3 (PDB:5lnr) (**a**). The potential impact of PDX1.2 on PDX1.3 activity can be through I1' and I3 interfaces. Due to D6 symmetry of the complex, these can be represented by trimer portion of the dodecamer. The trimer model was reconstructed by replacing one of the subunits of PDX1.3 (PDB:5lnr) with PDX1.2 monomeric unit (from our Cryo-EM, PDB:6pcj). The yellow molecule represents PLP and shows its location in the P2 active site. Note the significantly changed electrostatic potential on the upper left of the P2 site when PDX1.2 is lateral to PDX1.3 in the same hexamer ring.

Fig.5. Revised model of PDX1.2 and PDX1.3 co-assembly. Each subclass is annotated as 12:0, 11:1, 10:2, 9:3 and etc. where the numbers correspond to number of PDX1.2 subunits to PDX1.3 subunits. Only one representative combination per class is shown while the total number of possible combinations is variable. For example, subclass 10:2 encompasses 7 possible permutations for the D6 symmetry complex. Note: while a stacked hexamer association is still possible, these results indicate that it is not the only association possible, nor the major one.

Fig.1

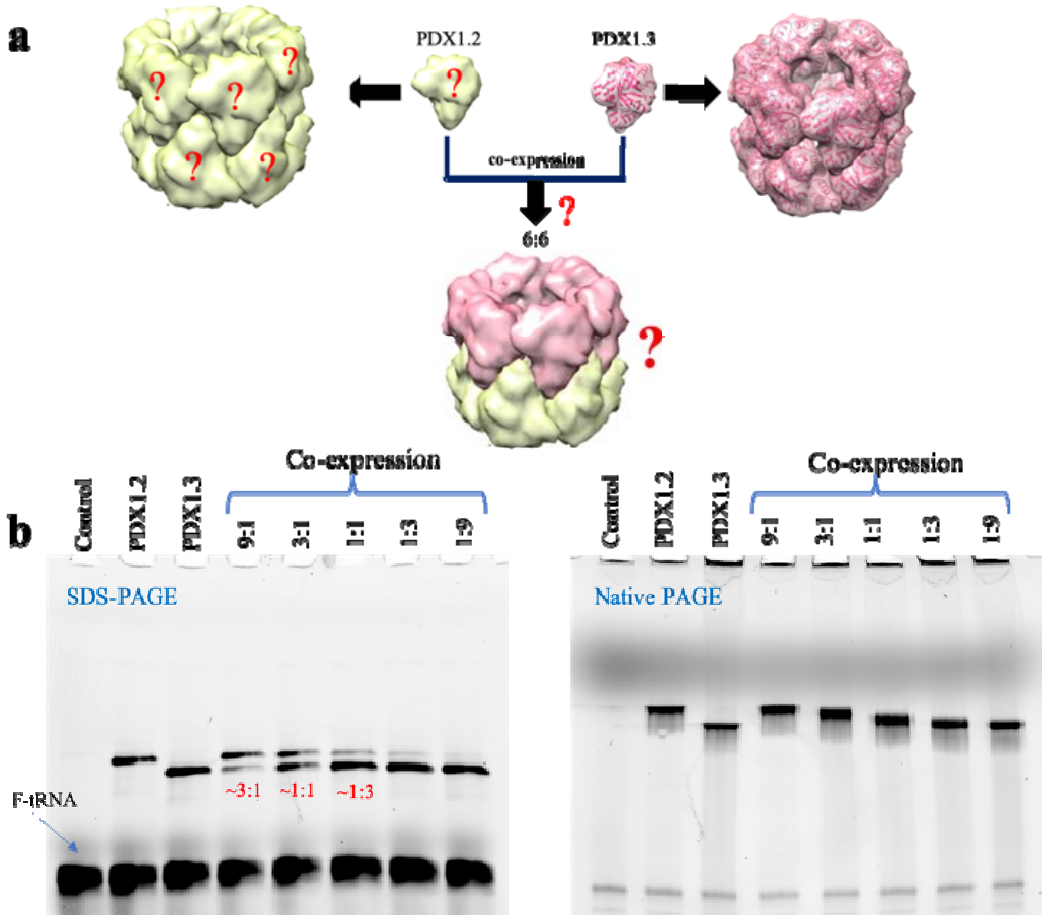

a
PDX1.2
PDX1.3
co-expression
6:6
b
Co-expression
Control
PDX1.2
PDX1.3
9:1
3:1
1:1
1:3
1:9
SDS-PAGE
F-tRNA
~3:1
~1:1
~1:3
Native PAGE

Fig.2.

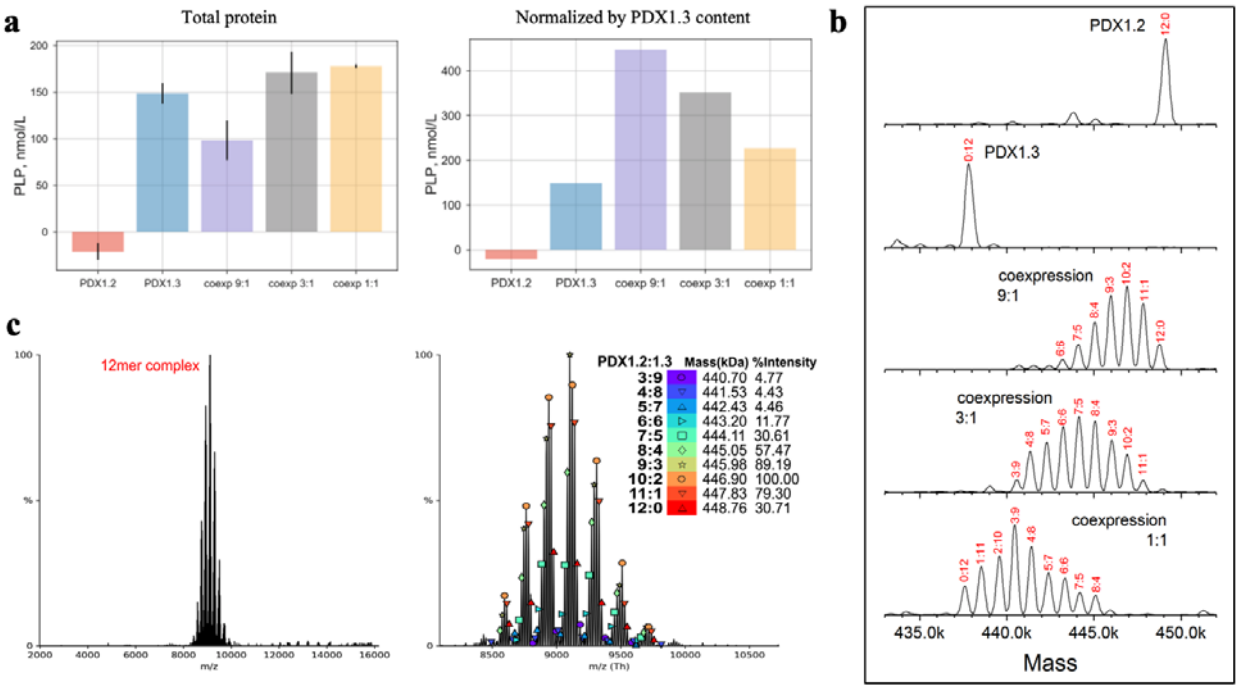

a
Total protein
Normalized by PDX1.3 content
PLP, nmol/L
PDX1.2
PDX1.3
coexp 9:1
coexp 3:1
coexp 1:1
b
PDX1.2
PDX1.3
coexpression 9:1
coexpression 3:1
coexpression 1:1
435.0k
440.0k
445.0k
450.0k
Mass
c
12mer complex
m/z
m/z (Th)
PDX1.2:1.3 Mass(kDa) %Intensity
3:9 440.70 4.77
4:8 441.53 4.43
5:7 442.43 4.46
6:6 443.20 11.77
7:5 444.11 30.61
8:4 445.05 57.47
9:3 445.98 89.19
10:2 446.90 100.00
11:1 447.83 79.30
12:0 448.76 30.71

Fig.3.

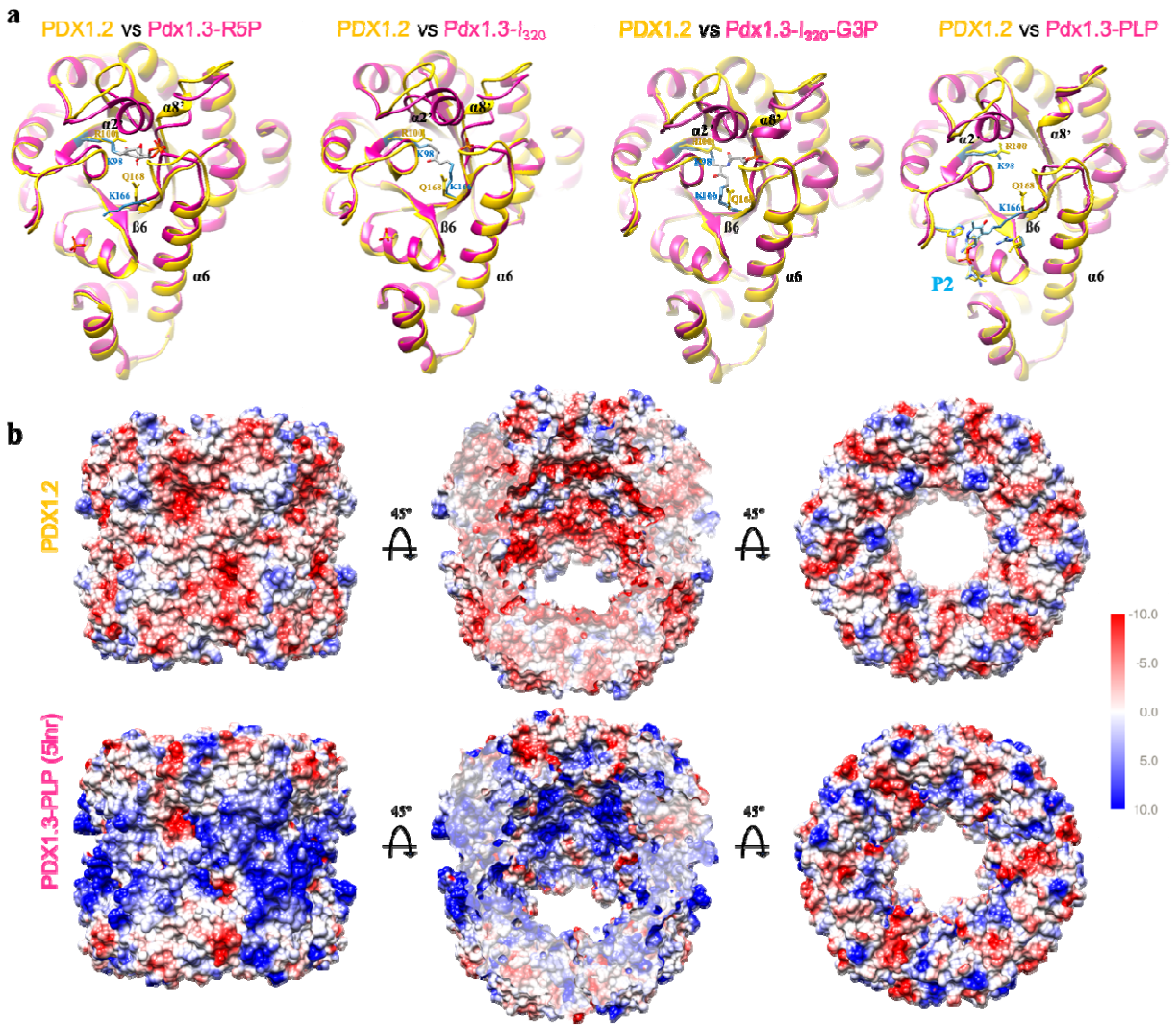

a
PDX1.2 vs Pdx1.3-R5P
PDX1.2 vs Pdx1.3-I320
PDX1.2 vs Pdx1.3-I320-G3P
PDX1.2 vs Pdx1.3-PLP
α2
α8'
R100
K98
Q168
K166
ß6
α6
P2
b
PDX1.2
PDX1.3-PLP (5lnr)
45°
-10.0
-5.0
0.0
5.0
10.0

Fig.4.

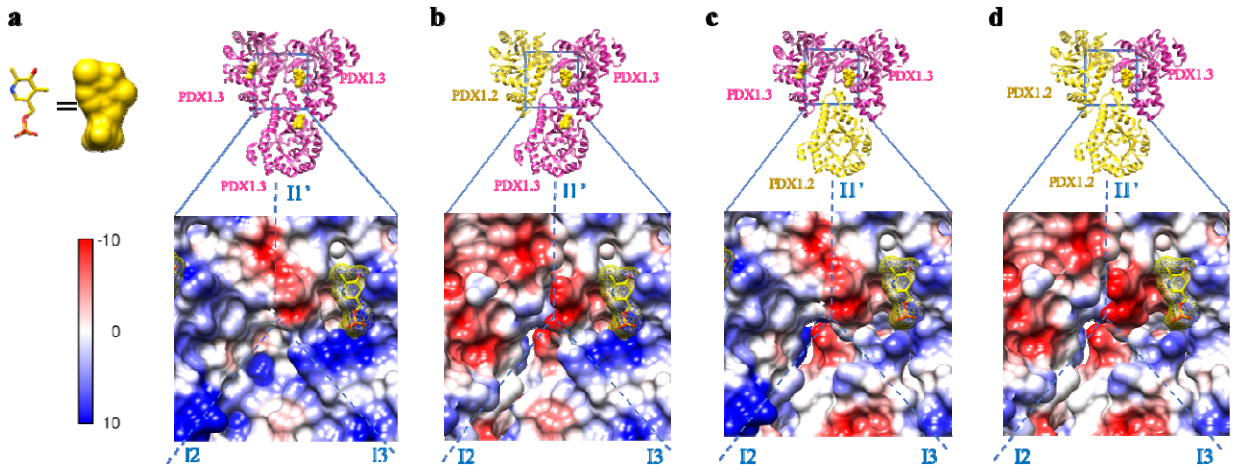
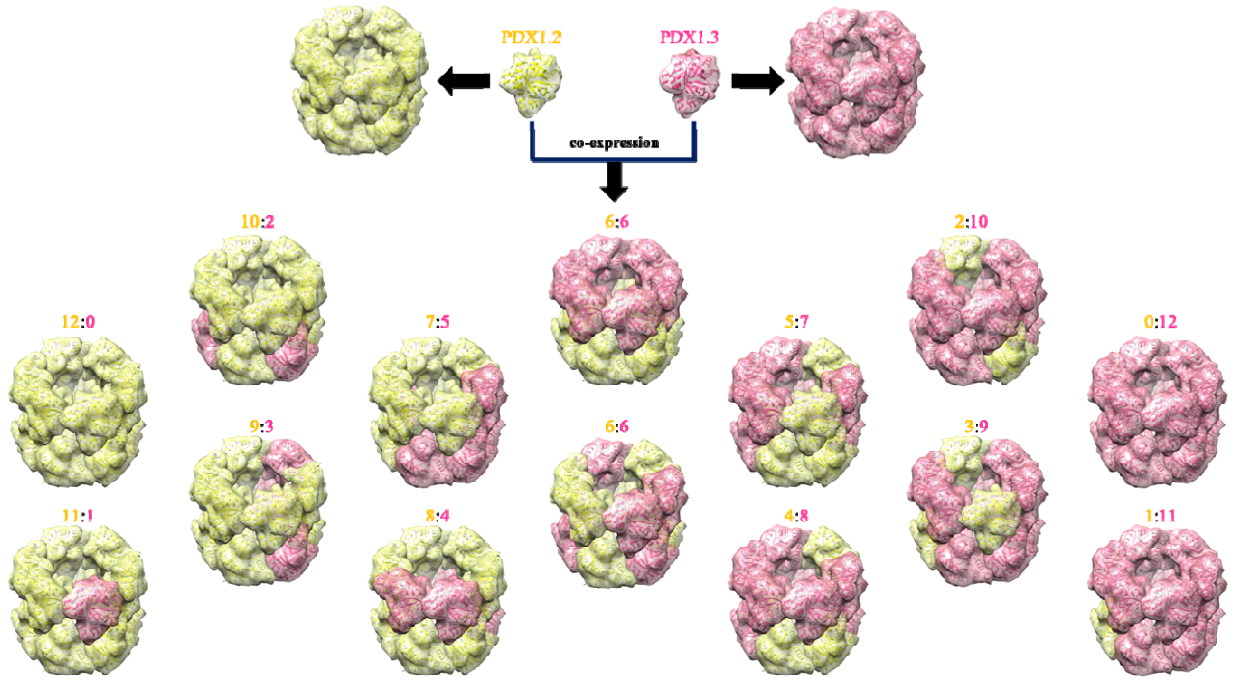


Fig.5.


PDX1.2
PDX1.3
co-expression
12:0
10:2
7:5
6:6
5:7
2:10
0:12
11:1
9:3
8:4
6:6
4:8
3:9
1:11