wwPDB X-ray Structure Validation Summary Report (i)

Jan 20, 2024 – 05:00 pm GMT

PDB ID	:	7NHF
Title	:	Crystal structure of Arabidopsis thaliana Pdx1K166R
Authors	:	Rodrigues, M.J.; Zhang, Y.; Bolton, R.; Evans, G.; Giri, N.; Royant, A.; Begley, T.; Ealick, S.E.; Tews, I.
Deposited on	:	2021-02-10
Resolution	:	2.35 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

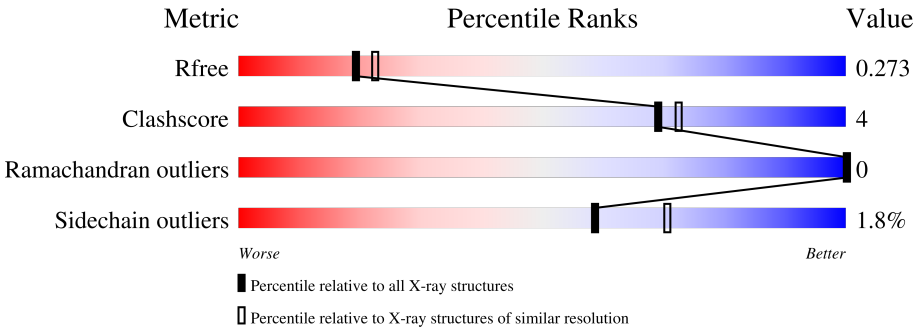
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.35 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	291	84% 8% 8%
1	B	291	85% 8% 7%
1	C	291	81% 11% 9%
1	D	291	83% 9% 8%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	PO4	D	301	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7975 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

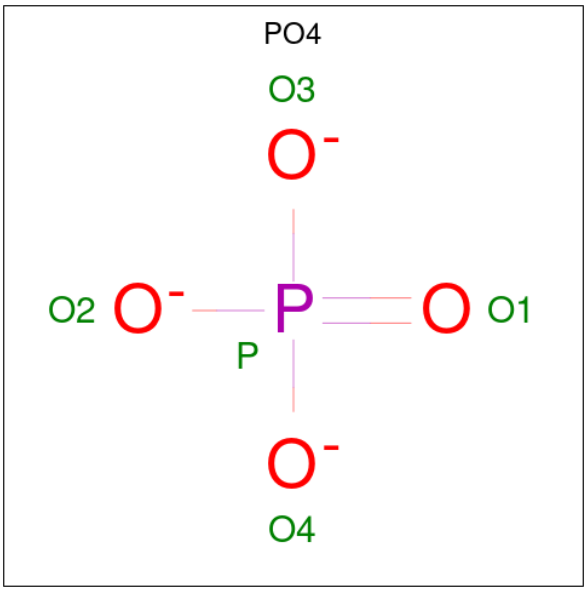
- Molecule 1 is a protein called Pyridoxal 5'-phosphate synthase subunit PDX1.3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	269	Total	C	N	O	S	0	2	0
			1998	1247	367	367	17			
1	B	270	Total	C	N	O	S	0	1	0
			1981	1239	362	364	16			
1	C	266	Total	C	N	O	S	0	1	0
			1965	1229	358	362	16			
1	D	267	Total	C	N	O	S	0	1	0
			1973	1233	361	362	17			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	166	ARG	LYS	engineered mutation	UNP Q8L940
B	166	ARG	LYS	engineered mutation	UNP Q8L940
C	166	ARG	LYS	engineered mutation	UNP Q8L940
D	166	ARG	LYS	engineered mutation	UNP Q8L940

- Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O P 5 4 1	0	0
2	B	1	Total O P 5 4 1	0	0
2	C	1	Total O P 5 4 1	0	0
2	D	1	Total O P 5 4 1	0	0

- Molecule 3 is water.

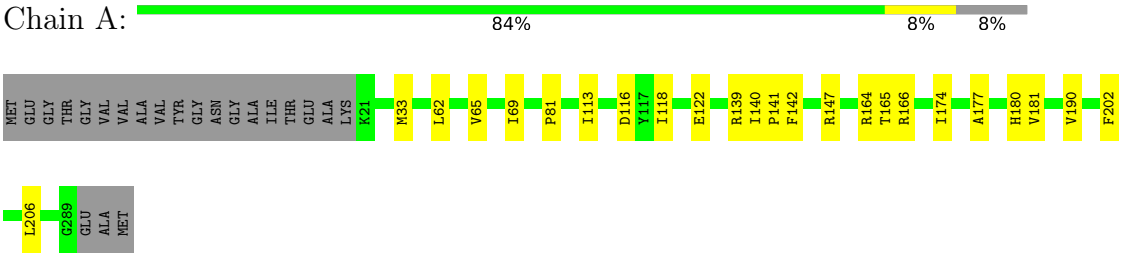
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	15	Total O 15 15	0	0
3	B	8	Total O 8 8	0	0
3	C	9	Total O 9 9	0	0
3	D	6	Total O 6 6	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

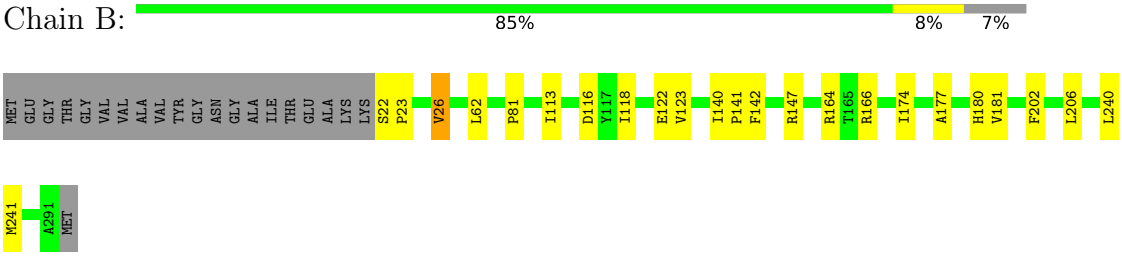
• Molecule 1: Pyridoxal 5'-phosphate synthase subunit PDX1.3

Chain A:



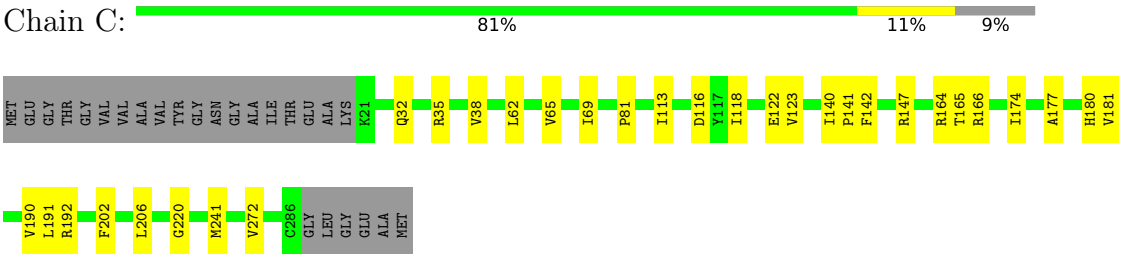
• Molecule 1: Pyridoxal 5'-phosphate synthase subunit PDX1.3

Chain B:



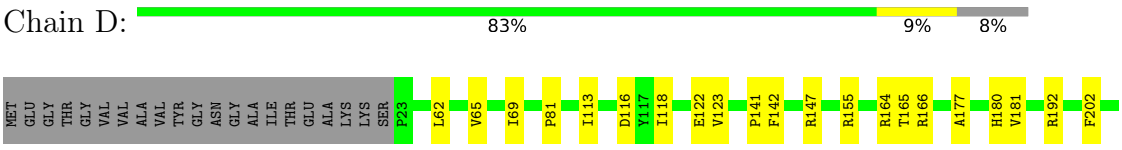
• Molecule 1: Pyridoxal 5'-phosphate synthase subunit PDX1.3

Chain C:



• Molecule 1: Pyridoxal 5'-phosphate synthase subunit PDX1.3

Chain D:




L206
Q218
L219
G220
S253
R266
G289
GLU
ALA
MET

4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	178.00Å 178.00Å 115.13Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.38 – 2.35 38.38 – 2.35	Depositor EDS
% Data completeness (in resolution range)	98.6 (38.38-2.35) 98.7 (38.38-2.35)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.85 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.220 , 0.249 0.252 , 0.273	Depositor DCC
R_{free} test set	2796 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	36.1	Xtriage
Anisotropy	0.806	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 26.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.000 for -1/3*h+1/3*k+4/3*l,-k,2/3*h+1/3*k+1/3*l 0.000 for -2/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+4/3*l,-1/3*h+1/3*k+1/3*l 0.000 for -h,1/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+1/3*l 0.000 for -1/3*h-2/3*k+4/3*l,-2/3*h-1/3*k-4/3*l,1/3*h-1/3*k-1/3*l 0.000 for -h,2/3*h+1/3*k+4/3*l,1/3*h+2/3*k-1/3*l 0.000 for 1/3*h+2/3*k-4/3*l,-k,-2/3*h-1/3*k-1/3*l 0.025 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	7975	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.25% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.71	0/2025	0.85	1/2738 (0.0%)
1	B	0.72	0/2008	0.85	1/2716 (0.0%)
1	C	0.70	0/1992	0.84	1/2696 (0.0%)
1	D	0.71	0/2000	0.86	1/2705 (0.0%)
All	All	0.71	0/8025	0.85	4/10855 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	164	ARG	CG-CD-NE	-5.32	100.64	111.80
1	C	164	ARG	CG-CD-NE	-5.21	100.85	111.80
1	B	164	ARG	CG-CD-NE	-5.08	101.13	111.80
1	D	164	ARG	CG-CD-NE	-5.04	101.22	111.80

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1998	0	1995	12	0
1	B	1981	0	1971	16	0
1	C	1965	0	1952	17	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	1973	0	1971	15	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
2	C	5	0	0	0	0
2	D	5	0	0	2	0
3	A	15	0	0	0	0
3	B	8	0	0	0	0
3	C	9	0	0	1	0
3	D	6	0	0	0	0
All	All	7975	0	7889	59	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

The worst 5 of 59 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:81:PRO:HB3	1:A:113:ILE:HD11	1.55	0.87
1:A:81:PRO:CB	1:A:113:ILE:HD11	2.26	0.66
1:C:81:PRO:HB3	1:C:113:ILE:HD11	1.76	0.65
1:B:22:SER:CB	1:B:23:PRO:CD	2.76	0.63
1:C:202:PHE:CE2	1:C:206:LEU:HD11	2.35	0.61

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	269/291 (92%)	261 (97%)	8 (3%)	0	100	100
1	B	269/291 (92%)	259 (96%)	10 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	265/291 (91%)	255 (96%)	10 (4%)	0	100	100
1	D	266/291 (91%)	259 (97%)	7 (3%)	0	100	100
All	All	1069/1164 (92%)	1034 (97%)	35 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	200/223 (90%)	196 (98%)	4 (2%)	55	66
1	B	195/223 (87%)	192 (98%)	3 (2%)	65	76
1	C	195/223 (87%)	191 (98%)	4 (2%)	53	65
1	D	197/223 (88%)	194 (98%)	3 (2%)	65	76
All	All	787/892 (88%)	773 (98%)	14 (2%)	59	70

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	35	ARG
1	C	62	LEU
1	D	218	GLN
1	D	62	LEU
1	D	123	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	218	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

4 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	PO4	D	301	-	4,4,4	0.62	0	6,6,6	0.54	0
2	PO4	C	301	-	4,4,4	0.89	0	6,6,6	0.47	0
2	PO4	B	301	-	4,4,4	1.29	1 (25%)	6,6,6	0.26	0
2	PO4	A	301	-	4,4,4	0.92	0	6,6,6	0.36	0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	301	PO4	P-O1	2.41	1.56	1.50

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	301	PO4	2	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.