wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 12:21 PM GMT

PDB ID	:	3QUJ
Title	:	Crystal structure of the phosphonate binding protein, PhnD, from Escherichia coli
Authors	:	Alicea, I.; Schreiter, E.R.
Deposited on	:	2011-02-24
Resolution	:	2.20 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

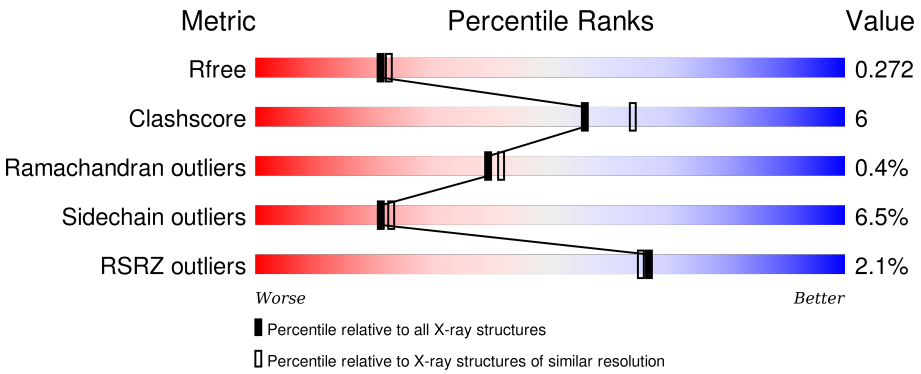
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.20 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}	91344	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	321	% 78% 15% • 6%
1	B	321	% 79% 14% • 5%
1	C	321	% 71% 16% • 12%
1	D	321	4% 73% 15% • 11%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9467 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 PhnD, subunit of alkylphosphonate ABC transporter.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	301	2359	1501	403	449	6	0	1	0
1	B	304	2386	1516	407	457	6	0	1	0
1	C	284	2241	1430	381	424	6	0	2	0
1	D	287	2260	1441	384	428	7	0	2	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-8	MSE	-	EXPRESSION TAG	UNP Q1R3F7
A	-7	HIS	-	EXPRESSION TAG	UNP Q1R3F7
A	-6	HIS	-	EXPRESSION TAG	UNP Q1R3F7
A	-5	HIS	-	EXPRESSION TAG	UNP Q1R3F7
A	-4	HIS	-	EXPRESSION TAG	UNP Q1R3F7
A	-3	HIS	-	EXPRESSION TAG	UNP Q1R3F7
A	-2	HIS	-	EXPRESSION TAG	UNP Q1R3F7
A	-1	GLY	-	EXPRESSION TAG	UNP Q1R3F7
A	0	SER	-	EXPRESSION TAG	UNP Q1R3F7
B	-8	MSE	-	EXPRESSION TAG	UNP Q1R3F7
B	-7	HIS	-	EXPRESSION TAG	UNP Q1R3F7
B	-6	HIS	-	EXPRESSION TAG	UNP Q1R3F7
B	-5	HIS	-	EXPRESSION TAG	UNP Q1R3F7
B	-4	HIS	-	EXPRESSION TAG	UNP Q1R3F7
B	-3	HIS	-	EXPRESSION TAG	UNP Q1R3F7
B	-2	HIS	-	EXPRESSION TAG	UNP Q1R3F7
B	-1	GLY	-	EXPRESSION TAG	UNP Q1R3F7
B	0	SER	-	EXPRESSION TAG	UNP Q1R3F7
C	-8	MSE	-	EXPRESSION TAG	UNP Q1R3F7
C	-7	HIS	-	EXPRESSION TAG	UNP Q1R3F7
C	-6	HIS	-	EXPRESSION TAG	UNP Q1R3F7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-5	HIS	-	EXPRESSION TAG	UNP Q1R3F7
C	-4	HIS	-	EXPRESSION TAG	UNP Q1R3F7
C	-3	HIS	-	EXPRESSION TAG	UNP Q1R3F7
C	-2	HIS	-	EXPRESSION TAG	UNP Q1R3F7
C	-1	GLY	-	EXPRESSION TAG	UNP Q1R3F7
C	0	SER	-	EXPRESSION TAG	UNP Q1R3F7
D	-8	MSE	-	EXPRESSION TAG	UNP Q1R3F7
D	-7	HIS	-	EXPRESSION TAG	UNP Q1R3F7
D	-6	HIS	-	EXPRESSION TAG	UNP Q1R3F7
D	-5	HIS	-	EXPRESSION TAG	UNP Q1R3F7
D	-4	HIS	-	EXPRESSION TAG	UNP Q1R3F7
D	-3	HIS	-	EXPRESSION TAG	UNP Q1R3F7
D	-2	HIS	-	EXPRESSION TAG	UNP Q1R3F7
D	-1	GLY	-	EXPRESSION TAG	UNP Q1R3F7
D	0	SER	-	EXPRESSION TAG	UNP Q1R3F7

- Molecule 2 is UNKNOWN LIGAND (three-letter code: UNL) (formula:).

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

- Molecule 3 is water.

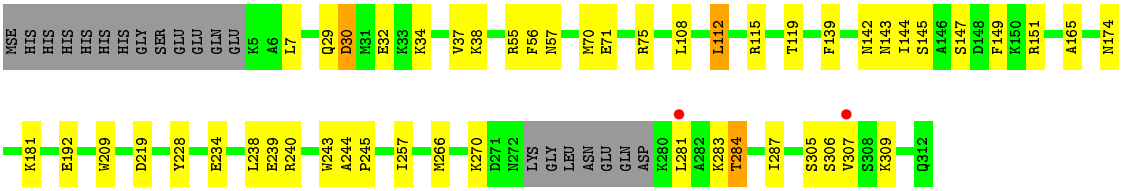
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	68	68	68	0	0
3	B	66	66	66	0	0
3	C	40	40	40	0	0
3	D	27	27	27	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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.

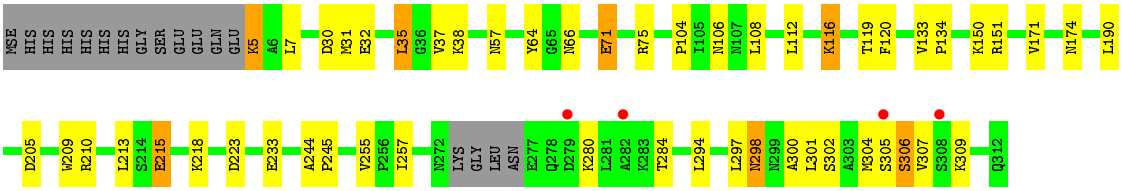
- Molecule 1: PhnD, subunit of alkylphosphonate ABC transporter

Chain A:



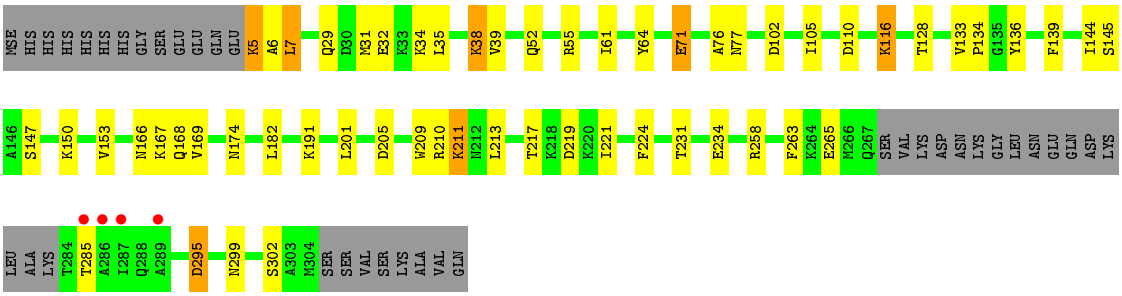
- Molecule 1: PhnD, subunit of alkylphosphonate ABC transporter

Chain B:



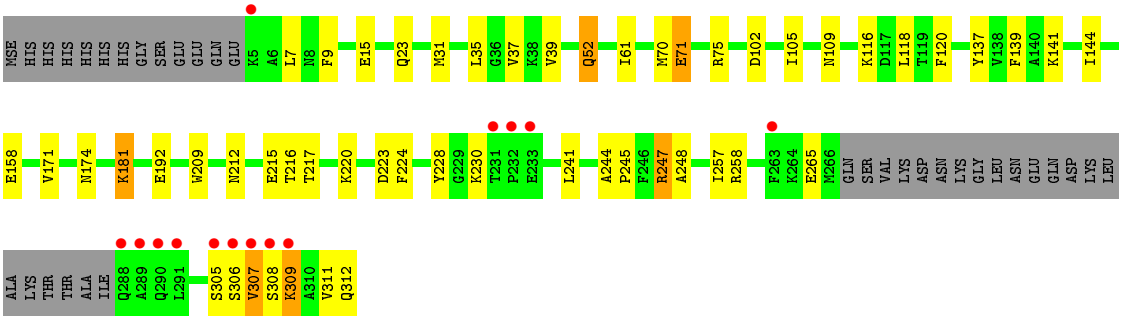
- Molecule 1: PhnD, subunit of alkylphosphonate ABC transporter

Chain C:



- Molecule 1: PhnD, subunit of alkylphosphonate ABC transporter

Chain D:


MSE
HIS
HIS
HIS
HIS
HIS
HIS
GLY
SER
GLU
GLU
GLN
GLU
K5
A6
L7
N8
F9
E15
Q23
N31
L35
G36
V37
K38
V39
Q52
I61
M70
E71
R75
D102
I105
N109
K116
D117
I118
T119
F120
Y137
V138
F139
A140
K141
I144
E158
V171
N174
K181
E192
W209
N212
E215
T216
T217
K220
D223
F224
Y228
G229
K230
T231
P232
E233
L241
A244
P245
F246
R247
A248
I257
R258
F263
K264
E265
M266
GLN
SER
VAL
LYS
ASP
ASN
LYS
GLY
LEU
ASN
GLU
GLN
ASP
LYS
LEU
ALA
LYS
THR
THR
ALA
ILE
Q288
A289
Q290
L291
S305
S306
V307
S308
K309
A310
V311
Q312

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	79.77Å 64.40Å 107.69Å 90.00° 93.52° 90.00°	Depositor
Resolution (Å)	28.90 – 2.20 28.93 – 2.20	Depositor EDS
% Data completeness (in resolution range)	98.8 (28.90-2.20) 98.8 (28.93-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.56 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, R_{free}	0.202 , 0.276 0.201 , 0.272	Depositor DCC
R_{free} test set	2798 reflections (5.36%)	DCC
Wilson B-factor (Å^2)	22.6	Xtriage
Anisotropy	0.626	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 40.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 54990 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9467	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 38.77 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.5096e-04.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1]Intensities estimated from amplitudes.

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

5 Model quality (i)

5.1 Standard geometry (i)

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

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	0.82	0/2402	0.79	0/3242
1	B	0.85	0/2429	0.79	1/3278 (0.0%)
1	C	0.81	0/2287	0.77	1/3090 (0.0%)
1	D	0.76	0/2306	0.75	0/3112
All	All	0.81	0/9424	0.78	2/12722 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	D	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	223	ASP	CB-CG-OD2	5.42	123.18	118.30
1	C	7	LEU	CA-CB-CG	5.08	126.98	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	212	ASN	Peptide

5.2 Too-close contacts (i)

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	2359	0	2370	23	0
1	B	2386	0	2390	26	0
1	C	2241	0	2242	29	0
1	D	2260	0	2264	39	0
2	A	5	0	0	0	0
2	B	5	0	0	1	0
2	C	5	0	0	0	0
2	D	5	0	0	0	0
3	A	68	0	0	0	0
3	B	66	0	0	1	0
3	C	40	0	0	0	0
3	D	27	0	0	0	0
All	All	9467	0	9266	115	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 6.

The worst 5 of 115 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:D:306:SER:O	1:D:307:VAL:HB	1.58	0.98
1:D:307:VAL:HG13	1:D:308:SER:HB2	1.49	0.94
1:B:32:GLU:OE2	1:B:38:LYS:HD2	1.84	0.78
1:D:70[A]:MSE:SE	1:D:258:ARG:HG3	2.35	0.76
1:D:105:ILE:HD11	1:D:118:LEU:HD11	1.66	0.76

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	298/321 (93%)	283 (95%)	14 (5%)	1 (0%)	46	50
1	B	301/321 (94%)	293 (97%)	7 (2%)	1 (0%)	46	50
1	C	282/321 (88%)	277 (98%)	5 (2%)	0	100	100
1	D	285/321 (89%)	264 (93%)	18 (6%)	3 (1%)	17	14
All	All	1166/1284 (91%)	1117 (96%)	44 (4%)	5 (0%)	39	42

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	307	VAL
1	D	265	GLU
1	D	309	LYS
1	A	305	SER
1	B	306	SER

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	256/266 (96%)	239 (93%)	17 (7%)	21	22
1	B	259/266 (97%)	245 (95%)	14 (5%)	27	31
1	C	242/266 (91%)	222 (92%)	20 (8%)	14	13
1	D	245/266 (92%)	228 (93%)	17 (7%)	19	20
All	All	1002/1064 (94%)	934 (93%)	68 (7%)	21	21

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	302	SER
1	C	116	LYS
1	D	217	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	5	LYS
1	C	34	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	29	GLN
1	D	52	GLN
1	C	212	ASN
1	B	298	ASN
1	C	52	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

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

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 are unknown - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	295/321 (91%)	-0.36	2 (0%)	89	88	9, 20, 39, 63	0
1	B	298/321 (92%)	-0.44	4 (1%)	79	78	7, 18, 42, 75	0
1	C	278/321 (86%)	-0.23	4 (1%)	78	77	10, 24, 44, 50	0
1	D	281/321 (87%)	-0.01	14 (4%)	32	32	9, 27, 51, 64	0
All	All	1152/1284 (89%)	-0.26	24 (2%)	67	65	7, 22, 46, 75	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	308	SER	4.5
1	C	286	ALA	4.5
1	D	231	THR	4.5
1	C	289	ALA	4.3
1	D	233	GLU	4.1

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
2	UNL	B	313	5/-	0.99	0.12	0.53	7,9,13,13	0
2	UNL	C	313	5/-	0.99	0.13	0.16	3,4,8,9	0
2	UNL	D	313	5/-	0.99	0.14	0.14	10,13,15,16	0
2	UNL	A	313	5/-	1.00	0.12	0.07	8,10,11,11	0

6.5 Other polymers (i)

There are no such residues in this entry.