



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 05:36 AM GMT

PDB ID : 2REJ  
Title : ABC-transporter choline binding protein in unliganded semi-closed conformation  
Authors : Oswald, C.; Smits, S.H.J.; Hoeing, M.; Sohn-Boeser, L.; Le Rudulier, D.; Schmitt, L.; Bremer, E.  
Deposited on : 2007-09-26  
Resolution : 2.60 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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

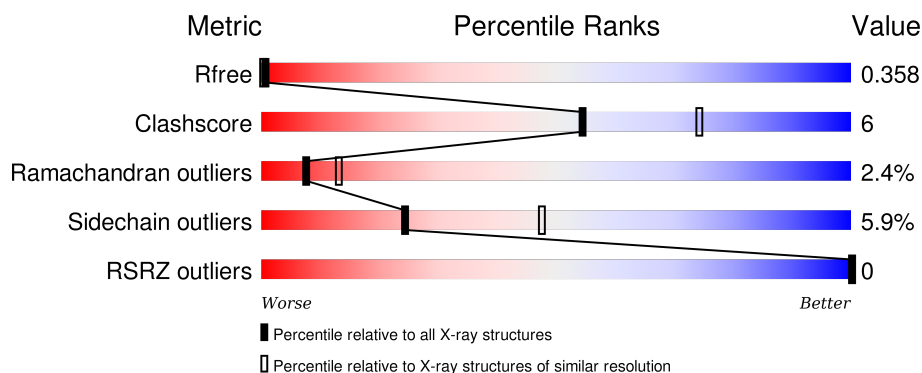
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*



The reported resolution of this entry is 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	298	 79% 13% . . .
1	B	298	 79% 16% . .

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 4352 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 PUTATIVE GLYCINE BETAINES ABC TRANSPORTER PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	288	Total	C	N	O	S	0	0	0
			2176	1371	357	439	9			
1	B	288	Total	C	N	O	S	0	0	0
			2176	1371	357	439	9			

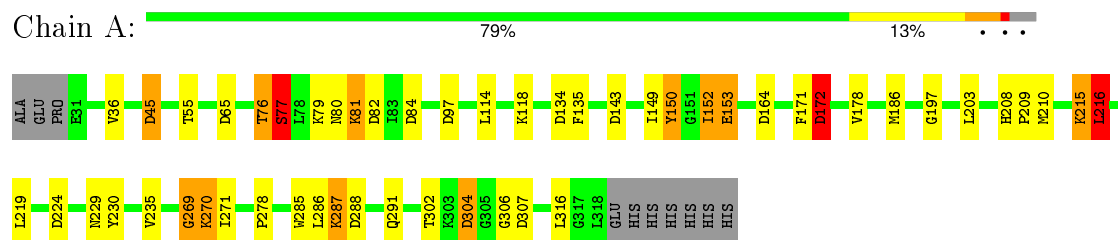
There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	251	ASP	GLY	engineered	UNP Q92N37
A	319	GLU	-	expression tag	UNP Q92N37
A	320	HIS	-	expression tag	UNP Q92N37
A	321	HIS	-	expression tag	UNP Q92N37
A	322	HIS	-	expression tag	UNP Q92N37
A	323	HIS	-	expression tag	UNP Q92N37
A	324	HIS	-	expression tag	UNP Q92N37
A	325	HIS	-	expression tag	UNP Q92N37
B	251	ASP	GLY	engineered	UNP Q92N37
B	319	GLU	-	expression tag	UNP Q92N37
B	320	HIS	-	expression tag	UNP Q92N37
B	321	HIS	-	expression tag	UNP Q92N37
B	322	HIS	-	expression tag	UNP Q92N37
B	323	HIS	-	expression tag	UNP Q92N37
B	324	HIS	-	expression tag	UNP Q92N37
B	325	HIS	-	expression tag	UNP Q92N37

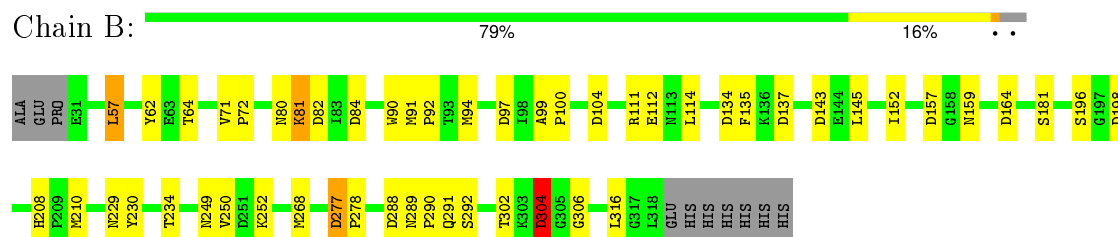
### 3 Residue-property plots

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: PUTATIVE GLYCINE BETAIN E ABC TRANSPORTER PROTEIN



#### • Molecule 1: PUTATIVE GLYCINE BETAIN E ABC TRANSPORTER PROTEIN



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	34.10 Å   232.50 Å   47.30 Å 90.00°   90.20°   90.00°	Depositor
Resolution (Å)	20.00 – 2.60 19.47 – 2.60	Depositor EDS
% Data completeness (in resolution range)	98.9 (20.00-2.60) 98.2 (19.47-2.60)	Depositor EDS
$R_{merge}$	0.22	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.03 (at 2.59 Å)	Xtriage
Refinement program	REFMAC 5.2.0003	Depositor
R, $R_{free}$	0.296 , 0.359 0.297 , 0.358	Depositor DCC
$R_{free}$ test set	1158 reflections (5.48%)	DCC
Wilson B-factor (Å <sup>2</sup> )	34.2	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 77.5	EDS
Estimated twinning fraction	0.318 for h,-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.24$	Xtriage
Outliers	2 of 22281 reflections (0.009%)	Xtriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	4352	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.51% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

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	0.44	0/2217	0.86	13/3009 (0.4%)
1	B	0.42	0/2217	0.76	13/3009 (0.4%)
All	All	0.43	0/4434	0.81	26/6018 (0.4%)

There are no bond length outliers.

All (26) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	45	ASP	CB-CG-OD2	5.64	123.37	118.30
1	B	143	ASP	CB-CG-OD2	5.62	123.36	118.30
1	B	82	ASP	CB-CG-OD2	5.57	123.31	118.30
1	B	198	ASP	CB-CG-OD2	5.39	123.15	118.30
1	B	164	ASP	CB-CG-OD2	5.34	123.11	118.30
1	A	82	ASP	CB-CG-OD2	5.33	123.10	118.30
1	B	104	ASP	CB-CG-OD2	5.33	123.10	118.30
1	A	84	ASP	CB-CG-OD2	5.30	123.07	118.30
1	A	224	ASP	CB-CG-OD2	5.25	123.02	118.30
1	B	277	ASP	CB-CG-OD2	5.22	123.00	118.30
1	B	134	ASP	CB-CG-OD2	5.19	122.97	118.30
1	A	288	ASP	CB-CG-OD2	5.18	122.96	118.30
1	B	137	ASP	CB-CG-OD2	5.18	122.96	118.30
1	B	304	ASP	CB-CG-OD2	5.16	122.95	118.30
1	A	134	ASP	CB-CG-OD2	5.16	122.94	118.30
1	B	288	ASP	CB-CG-OD2	5.15	122.93	118.30
1	A	97	ASP	CB-CG-OD2	5.13	122.92	118.30
1	A	304	ASP	CB-CG-OD2	5.13	122.91	118.30
1	A	307	ASP	CB-CG-OD2	5.12	122.91	118.30
1	A	143	ASP	CB-CG-OD2	5.11	122.90	118.30
1	B	157	ASP	CB-CG-OD2	5.10	122.89	118.30
1	A	65	ASP	CB-CG-OD2	5.03	122.83	118.30
1	B	84	ASP	CB-CG-OD2	5.03	122.82	118.30
1	A	172	ASP	CB-CG-OD2	5.02	122.82	118.30
1	A	164	ASP	CB-CG-OD2	5.01	122.81	118.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	97	ASP	CB-CG-OD2	5.00	122.80	118.30

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	2176	0	2122	35	0
1	B	2176	0	2122	20	0
All	All	4352	0	4244	55	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.

All (55) 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:269:GLY:HA3	1:A:270:LYS:CB	1.75	1.14
1:A:269:GLY:HA3	1:A:270:LYS:HB2	1.17	1.08
1:A:149:ILE:HA	1:A:150:TYR:HB2	1.39	1.02
1:A:208:HIS:HD2	1:A:210:MET:HB2	1.31	0.94
1:A:215:LYS:CB	1:A:216:LEU:HB2	1.99	0.93
1:B:111:ARG:HG2	1:B:112:GLU:H	1.44	0.82
1:A:215:LYS:HB2	1:A:216:LEU:HB2	1.61	0.82
1:A:215:LYS:CA	1:A:216:LEU:HB2	2.14	0.76
1:B:302:THR:HG22	1:B:306:GLY:H	1.54	0.73
1:B:111:ARG:HG2	1:B:112:GLU:N	2.04	0.72
1:A:302:THR:HG22	1:A:306:GLY:H	1.54	0.71
1:A:149:ILE:O	1:A:178:VAL:HA	1.94	0.68
1:B:208:HIS:HD2	1:B:210:MET:HB2	1.58	0.68
1:A:269:GLY:CA	1:A:270:LYS:CB	2.61	0.67
1:A:269:GLY:CA	1:A:270:LYS:HB2	2.10	0.67
1:B:57:LEU:HD23	1:B:62:TYR:HB2	1.78	0.65
1:A:80:ASN:N	1:A:81:LYS:HA	2.10	0.65

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:208:HIS:CD2	1:A:210:MET:HB2	2.23	0.63
1:A:171:PHE:N	1:A:172:ASP:HA	2.14	0.63
1:A:229:ASN:O	1:A:230:TYR:HB2	1.97	0.62
1:A:269:GLY:HA3	1:A:270:LYS:HB3	1.77	0.62
1:B:90:TRP:O	1:B:94:MET:HB2	1.99	0.62
1:A:149:ILE:CA	1:A:150:TYR:HB2	2.23	0.61
1:A:76:THR:CA	1:A:77:SER:HB2	2.30	0.60
1:A:215:LYS:HB2	1:A:216:LEU:CB	2.31	0.60
1:B:291:GLN:HG2	1:B:292:SER:HA	1.83	0.59
1:B:111:ARG:CG	1:B:112:GLU:H	2.15	0.59
1:B:80:ASN:O	1:B:81:LYS:HB2	2.02	0.58
1:B:208:HIS:CD2	1:B:210:MET:HB2	2.39	0.56
1:B:91:MET:HB3	1:B:92:PRO:HA	1.88	0.54
1:A:76:THR:CB	1:A:77:SER:HB2	2.39	0.53
1:A:215:LYS:HA	1:A:216:LEU:HB2	1.90	0.52
1:A:286:LEU:H	1:A:287:LYS:HB3	1.73	0.52
1:A:271:ILE:HD11	1:A:278:PRO:HG3	1.91	0.51
1:A:149:ILE:HA	1:A:150:TYR:CB	2.28	0.51
1:B:71:VAL:HB	1:B:72:PRO:HD3	1.92	0.51
1:A:186:MET:HE1	1:A:203:LEU:H	1.75	0.51
1:A:118:LYS:HE3	1:A:230:TYR:HB3	1.92	0.49
1:A:152:ILE:HG23	1:A:153:GLU:N	2.27	0.49
1:A:152:ILE:HG23	1:A:153:GLU:H	1.78	0.48
1:B:229:ASN:O	1:B:230:TYR:HB2	2.13	0.48
1:B:152:ILE:H	1:B:159:ASN:HD21	1.62	0.47
1:A:76:THR:N	1:A:77:SER:HB2	2.30	0.46
1:A:76:THR:O	1:A:79:LYS:HB3	2.17	0.44
1:A:76:THR:H	1:A:77:SER:HB2	1.83	0.44
1:A:76:THR:H	1:A:77:SER:CB	2.31	0.43
1:B:99:ALA:HB3	1:B:100:PRO:HD3	2.01	0.43
1:A:208:HIS:CD2	1:A:210:MET:H	2.37	0.43
1:A:80:ASN:H	1:A:81:LYS:HA	1.79	0.42
1:A:76:THR:N	1:A:77:SER:CB	2.82	0.42
1:B:302:THR:HG23	1:B:304:ASP:H	1.85	0.42
1:B:249:ASN:O	1:B:252:LYS:HB3	2.20	0.42
1:B:277:ASP:HA	1:B:278:PRO:HD3	1.96	0.41
1:B:289:ASN:C	1:B:291:GLN:H	2.23	0.41
1:B:302:THR:HG22	1:B:306:GLY:N	2.31	0.40

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	286/298 (96%)	245 (86%)	31 (11%)	10 (4%)	4	6
1	B	286/298 (96%)	263 (92%)	19 (7%)	4 (1%)	14	28
All	All	572/596 (96%)	508 (89%)	50 (9%)	14 (2%)	7	13

All (14) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	77	SER
1	A	150	TYR
1	A	216	LEU
1	B	81	LYS
1	A	269	GLY
1	A	270	LYS
1	A	285	TRP
1	A	291	GLN
1	A	287	LYS
1	B	181	SER
1	B	196	SER
1	A	209	PRO
1	A	197	GLY
1	B	290	PRO

### 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	230/240 (96%)	213 (93%)	17 (7%)	17	34
1	B	230/240 (96%)	220 (96%)	10 (4%)	35	64
All	All	460/480 (96%)	433 (94%)	27 (6%)	24	47

All (27) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	36	VAL
1	A	45	ASP
1	A	55	THR
1	A	76	THR
1	A	77	SER
1	A	81	LYS
1	A	114	LEU
1	A	135	PHE
1	A	152	ILE
1	A	153	GLU
1	A	172	ASP
1	A	215	LYS
1	A	216	LEU
1	A	219	LEU
1	A	235	VAL
1	A	304	ASP
1	A	316	LEU
1	B	57	LEU
1	B	64	THR
1	B	114	LEU
1	B	135	PHE
1	B	145	LEU
1	B	234	THR
1	B	250	VAL
1	B	268	MET
1	B	304	ASP
1	B	316	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	141	HIS
1	A	159	ASN
1	A	208	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	273	ASN
1	B	159	ASN
1	B	189	GLN
1	B	208	HIS

### 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](#)

There are no ligands in this entry.

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	288/298 (96%)	-0.34	0 100 100	21, 30, 40, 43	0
1	B	288/298 (96%)	-0.26	0 100 100	21, 32, 40, 44	0
All	All	576/596 (96%)	-0.30	0 100 100	21, 31, 40, 44	0

There are no RSRZ outliers to report.

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.