



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:07 AM GMT

PDB ID : 2VZV  
Title : SUBSTRATE COMPLEX OF AMYCOLATOPSIS ORIENTALIS EXO-CHITOSANASE CSXA E541A WITH CHITOSAN  
Authors : Lammerts Van Bueren, A.; Ghinet, M.G.; Gregg, K.; Fleury, A.; Brzezinski, R.; Boraston, A.B.  
Deposited on : 2008-08-05  
Resolution : 2.70 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](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

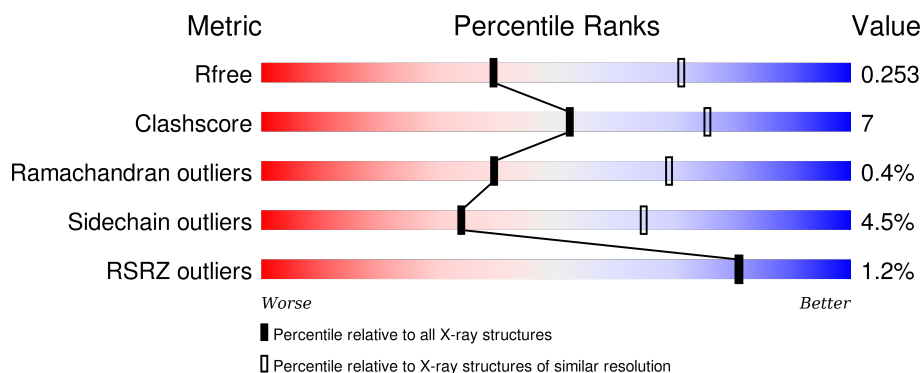
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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	1032	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between; align-items: center;"> <span>%</span> <div style="width: 100%; height: 10px; background-color: green;"></div> <span>70%</span> <span>13%</span> <span>•</span> <span>17%</span> </div> </div>
1	B	1032	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between; align-items: center;"> <span>%</span> <div style="width: 100%; height: 10px; background-color: green;"></div> <span>69%</span> <span>12%</span> <span>•</span> <span>17%</span> </div> </div>

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	GCS	A	1900	-	-	-	X
2	GCS	B	1900	-	-	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 13565 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 EXO-BETA-D-GLUCOSAMINIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	858	Total	C	N	O	S	4	0	1
			6548	4113	1133	1285	17			
1	B	858	Total	C	N	O	S	4	0	1
			6548	4113	1133	1285	17			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	541	ALA	GLU	ENGINEERED MUTATION	UNP Q56F26
A	750	ASN	TRP	CONFLICT	UNP Q56F26
B	541	ALA	GLU	ENGINEERED MUTATION	UNP Q56F26
B	750	ASN	TRP	CONFLICT	UNP Q56F26

- Molecule 2 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	2	Total	C	N	O	0	0
			23	12	2	9		
2	B	2	Total	C	N	O	0	0
			23	12	2	9		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	212	Total	O	0	0
			212	212		
3	B	211	Total	O	0	0
			211	211		



Lys	ALA	Q894	Y680	L490	D296
VAL	GLY	T895	S681	L491	R297
ALA	TYR	A898	R685	P495	H310
THR	ASP	D899	S688	S502	D315
THR	VAL	GLY	G699	P503	T323
ALA	VAL	GLY	L700	L504	P324
GLY	ARG	PRO	T701		
GLY	TYR	GLY		K511	
PRO	ALA	PRO	T704	M512	
ASN	ASN	SER			R335
VAL	GLY	ASP	L709	D517	D336
VAL	GLY	PRO			V337
ASP	THR	THR	V736	P521	
Lys	THR	VAL	F737		N352
ILE	THR	ASP	A738	K527	G353
THR	SER	TYR	V739	S523	K354
LEU	ARG	GLN		Q529	P355
	PRO	ALA			L356
	LEU	GLU	T745	M536	
	ASP	ASP			L369
	PHE	ALA	T753	D552	
	SER	THR	D754		A375
	VAL	ILE		P570	
	ASN	VAL	K758		K379
	GLY	GLN	V764	R577	
	SER	GLY			T390
	ILE	ALA	T769	F583	V391
	SER	VAL			R392
	ALA	GLU			
	GLY	SER	L774	Y597	G395
	SER	ASN	H775	G598	
	VAL	HIS	W776	A599	
	ALA	ALA		D605	D400
	PHE	GLY	L792		
	GLY	TYR			L412
	SER	THR	L798	R608	T413
	THR	GLY			M414
	GLY	THR	V803	D516	P415
	THR	GLY			G416
	TRP	PHE	A807	R627	H417
	PRO	VAL	N808		E418
	ALA	ASN	S809	D631	
	TRP	TYR		S632	N428
	THR	ASP	P832		G429
	THR	ASN			E430
	Lys	VAL	V836	M643	E431
	THR	ALA			
	VAL	GLY	R846	M648	A452
	ARG	SER			E453
	VAL	SER	A857	H652	R454
	THR	VAL	V858		
	LEU	GLU	S859	F656	F464
	ALA	TRP	L860		H465
	ALA	THR	W861	Y659	
	GLY	VAL		M660	E477
	VAL	VAL	T865		
	ASN	VAL		H677	M464
	Lys	PRO	L877	M678	E469
	ILE	SER	O879		

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	86.86Å 121.41Å 184.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	185.70 – 2.70 19.97 – 2.70	Depositor EDS
% Data completeness (in resolution range)	91.3 (185.70-2.70) 91.6 (19.97-2.70)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.33 (at 2.71Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.182 , 0.255 0.184 , 0.253	Depositor DCC
$R_{free}$ test set	2545 reflections (5.40%)	DCC
Wilson B-factor (Å <sup>2</sup> )	20.5	Xtriage
Anisotropy	0.152	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 24.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Outliers	0 of 49770 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	13565	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 16.86% 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 [i](#)

### 5.1 Standard geometry [i](#)

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

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	0.54	3/6711 (0.0%)	0.66	5/9158 (0.1%)
1	B	0.52	3/6711 (0.0%)	0.66	8/9158 (0.1%)
All	All	0.53	6/13422 (0.0%)	0.66	13/18316 (0.1%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	430	GLU	CB-CG	16.66	1.83	1.52
1	B	430	GLU	CB-CG	10.96	1.73	1.52
1	B	246	ASP	CB-CG	7.18	1.66	1.51
1	B	296	ASP	CB-CG	6.59	1.65	1.51
1	A	246	ASP	CB-CG	6.17	1.64	1.51

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	246	ASP	CB-CG-OD1	15.77	132.49	118.30
1	B	246	ASP	CB-CG-OD1	15.13	131.92	118.30
1	B	296	ASP	CB-CG-OD1	14.89	131.71	118.30
1	A	296	ASP	CB-CG-OD2	11.98	129.08	118.30
1	A	430	GLU	CA-CB-CG	-11.08	89.02	113.40

There are no chirality outliers.

There are no planarity outliers.

### 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	6548	0	6327	81	9
1	B	6548	0	6327	92	9
2	A	23	0	23	0	0
2	B	23	0	23	2	0
3	A	212	0	0	9	0
3	B	211	0	0	17	0
All	All	13565	0	12700	173	9

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 7.

The worst 5 of 173 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:B:201:TRP:HE1	1:B:212:ASN:HD21	1.00	0.93
1:B:310:HIS:CE1	3:B:2079:HOH:O	2.23	0.90
1:B:577:ARG:HG2	1:B:583:PHE:O	1.75	0.85
1:B:244:HIS:HE2	1:B:246:ASP:CG	1.82	0.80
1:B:115:ALA:O	1:B:118:SER:HB2	1.82	0.79

The worst 5 of 9 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:246:ASP:OD2	1:B:296:ASP:OD1[3_554]	1.58	0.62
1:A:296:ASP:OD2	1:B:246:ASP:OD1[3_554]	1.61	0.59
1:A:296:ASP:OD2	1:B:246:ASP:OD2[3_554]	1.61	0.59
1:A:246:ASP:OD1	1:B:296:ASP:OD1[3_554]	1.75	0.45
1:A:246:ASP:OD2	1:B:296:ASP:CG[3_554]	1.82	0.38

## 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	856/1032 (83%)	811 (95%)	42 (5%)	3 (0%)	39	69
1	B	856/1032 (83%)	810 (95%)	42 (5%)	4 (0%)	34	63
All	All	1712/2064 (83%)	1621 (95%)	84 (5%)	7 (0%)	39	69

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	758	LYS
1	B	202	ILE
1	A	202	ILE
1	A	205	ALA
1	B	205	ALA

### 5.3.2 Protein sidechains ⓘ

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	701/833 (84%)	672 (96%)	29 (4%)	37	69
1	B	701/833 (84%)	667 (95%)	34 (5%)	31	61
All	All	1402/1666 (84%)	1339 (96%)	63 (4%)	34	65

5 of 63 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	895	THR
1	B	237	LYS
1	B	808	ASN
1	B	118	SER
1	B	144	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	796	ASN
1	B	178	HIS
1	B	796	ASN
1	B	128	ASN
1	B	169	HIS

### 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 ⓘ

4 carbohydrates 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	GCS	A	1899	2	11,11,12	0.57	0	13,15,17	1.51	2 (15%)
2	GCS	A	1900	2	12,12,12	1.66	4 (33%)	15,17,17	1.39	3 (20%)
2	GCS	B	1899	2	11,11,12	0.77	0	13,15,17	1.98	2 (15%)
2	GCS	B	1900	2	12,12,12	1.37	1 (8%)	15,17,17	1.50	2 (13%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GCS	A	1899	2	-	0/2/19/22	0/1/1/1
2	GCS	A	1900	2	-	0/2/22/22	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GCS	B	1899	2	-	0/2/19/22	0/1/1/1
2	GCS	B	1900	2	-	0/2/22/22	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1900	GCS	C6-C5	2.08	1.59	1.51
2	A	1900	GCS	C4-C5	2.12	1.57	1.53
2	A	1900	GCS	O5-C5	2.25	1.50	1.44
2	B	1900	GCS	O4-C4	2.71	1.49	1.43
2	A	1900	GCS	O4-C4	2.72	1.49	1.43

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1899	GCS	O5-C5-C6	-2.68	101.55	107.35
2	B	1899	GCS	C4-C3-C2	-2.30	108.28	111.39
2	B	1900	GCS	C6-C5-C4	-2.05	107.95	113.02
2	A	1900	GCS	O5-C5-C4	2.21	113.83	109.68
2	A	1900	GCS	O1-C1-C2	2.43	114.55	109.02

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	B	1900	GCS	2	0

## 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 ⓘ

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	858/1032 (83%)	-0.42	11 (1%) 79 79	3, 15, 27, 66	1 (0%)
1	B	858/1032 (83%)	-0.41	10 (1%) 81 81	3, 15, 25, 68	1 (0%)
All	All	1716/2064 (83%)	-0.42	21 (1%) 81 81	3, 15, 26, 68	2 (0%)

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	43	SER	7.8
1	A	47	ALA	6.3
1	A	44	VAL	6.0
1	A	45	GLY	5.9
1	B	46	ALA	5.6

### 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](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	GCS	B	1900	12/12	0.90	0.20	2.72	28,33,34,35	0
2	GCS	A	1900	12/12	0.88	0.20	2.68	26,28,28,29	0
2	GCS	B	1899	11/12	0.94	0.15	1.56	19,22,23,24	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	GCS	A	1899	11/12	0.96	0.15	1.01	15,19,21,22	0

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.