



# wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 02:55 AM GMT

PDB ID : 2X09  
Title : INHIBITION OF THE EXO-BETA-D-GLUCOSAMINIDASECSXA BY A  
GLUCOSAMINE-CONFIGUREDCASTANOSPERMINE AND AN AMINO-  
AUSTRALINE ANALOGUE  
Authors : Pluvinae, B.; Ghinet, M.G.; Brzezinski, R.; Boraston, A.B.; Stubbs, K.A.  
Deposited on : 2009-12-07  
Resolution : 2.40 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

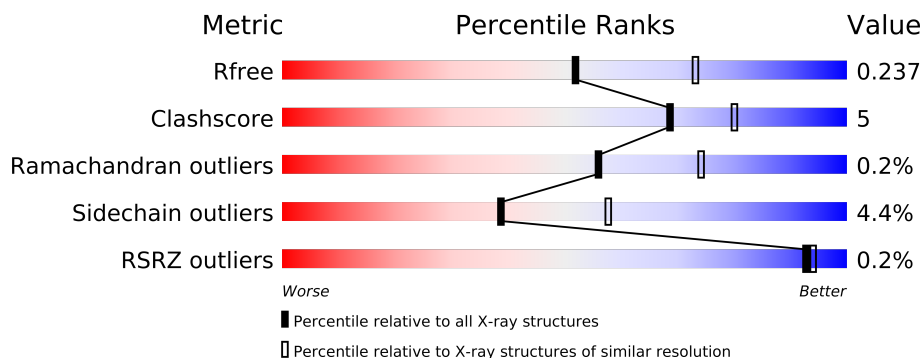
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.40 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))
$R_{free}$	66092	2207 (2.40-2.40)
Clashscore	79885	2789 (2.40-2.40)
Ramachandran outliers	78287	2736 (2.40-2.40)
Sidechain outliers	78261	2737 (2.40-2.40)
RSRZ outliers	66119	2210 (2.40-2.40)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	1032	
1	B	1032	

## 2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 14500 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

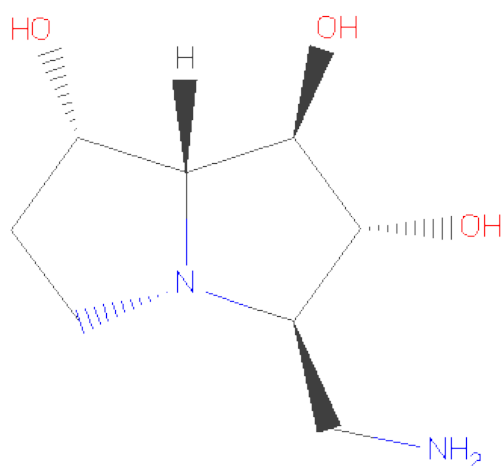
- Molecule 1 is a protein called EXO-BETA-D-GLUCOSAMINIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	851	Total	C	N	O	S	15	1	1
			6518	4095	1127	1279	17			
1	B	851	Total	C	N	O	S	9	0	1
			6512	4090	1126	1279	17			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	750	ASN	TRP	CONFLICT	UNP Q56F26
B	750	ASN	TRP	CONFLICT	UNP Q56F26

- Molecule 2 is AMINO-AUSTRALINE (three-letter code: X09) (formula: C<sub>8</sub>H<sub>16</sub>N<sub>2</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			13	8	2	3		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	B	1	Total	C	N	O	0	0
			13	8	2	3		

- Molecule 3 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Cd	0	0
			1	1		
3	A	2	Total	Cd	0	0
			2	2		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	733	Total	O	0	0
			733	733		
4	B	708	Total	O	0	0
			708	708		



T839	T865	T866	L877	G888	N890	T891	T895	D899	GLY	SER	GLY	PRO	GLY	PRO	SER	ASP	PRO	VAL	ASP	THR	TYR	ARG	GLN	ALA	GLU	ASP	ALA	THR	ILE	ILE	ASN	GLY	SER	ASN	HIS	ALA	GLY	TYR	GLY	PHE	VAL	ASN	TYR	ASP	ASN	VAL	ALA	GLY	SER
									VAL	GLY	ARG	TYR	ALA	ASN	GLY	THR	THR	SER	SER	ARG	PRO	LEU	ASP	PHE	SER	VAL	ASN	GLY	SER	ILE	SER	ALA	GLY	VAL	PHE	THR	THR	PRO	ALA	THR	THR	LYS	THR	VAL	ARG				

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	86.75Å 121.95Å 92.05Å 90.00° 90.65° 90.00°	Depositor
Resolution (Å)	20.00 – 2.40 19.99 – 2.40	Depositor EDS
% Data completeness (in resolution range)	97.5 (20.00-2.40) 96.7 (19.99-2.40)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.06 (at 2.41Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, $R_{free}$	0.176 , 0.240 0.175 , 0.237	Depositor DCC
$R_{free}$ test set	3652 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	19.9	Xtriage
Anisotropy	0.182	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 13.7	EDS
Estimated twinning fraction	0.028 for h,-k,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 72247 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	14500	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.96% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 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: X09, CD

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	0.61	4/6684 (0.1%)	0.74	10/9119 (0.1%)
1	B	0.84	4/6675 (0.1%)	0.77	10/9108 (0.1%)
All	All	0.73	8/13359 (0.1%)	0.75	20/18227 (0.1%)

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	297	ARG	CD-NE	-34.45	0.87	1.46
1	B	530	LYS	CD-CE	24.63	2.12	1.51
1	B	758	LYS	CD-CE	-22.07	0.96	1.51
1	B	573	LYS	CE-NZ	-16.59	1.07	1.49
1	A	127	LEU	CG-CD1	-15.65	0.94	1.51

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	297	ARG	CG-CD-NE	23.39	160.91	111.80
1	A	127	LEU	CB-CG-CD1	23.14	150.35	111.00
1	B	297	ARG	CD-NE-CZ	14.76	144.27	123.60
1	A	127	LEU	CB-CG-CD2	-14.21	86.84	111.00
1	B	573	LYS	CD-CE-NZ	11.70	138.61	111.70

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens



added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	6518	0	6298	58	0
1	B	6512	0	6285	63	2
2	A	13	0	16	5	0
2	B	13	0	16	4	0
3	A	2	0	0	0	2
3	B	1	0	0	0	0
4	A	733	0	0	2	0
4	B	708	0	0	11	0
All	All	14500	0	12615	123	2

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 5.

The worst 5 of 123 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:109:MET:CE	1:B:368:PHE:HE1	1.68	1.06
1:B:201:TRP:HE1	1:B:212:ASN:HD21	1.04	0.98
1:A:201:TRP:HE1	1:A:212:ASN:HD21	1.10	0.96
1:B:179:THR:CA	4:B:2149:HOH:O	2.14	0.95
1:B:109:MET:CE	1:B:368:PHE:CE1	2.50	0.94

All (2) 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	Distance(Å)	Clash(Å)
1:B:246:ASP:OD1	3:A:1902:CD:CD[2.557]	1.94	0.26
1:B:244:HIS:NE2	3:A:1902:CD:CD[2.557]	2.10	0.10

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	850/1032 (82%)	821 (97%)	27 (3%)	2 (0%)	56 74
1	B	849/1032 (82%)	819 (96%)	28 (3%)	2 (0%)	56 74
All	All	1699/2064 (82%)	1640 (96%)	55 (3%)	4 (0%)	56 74

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	541	GLU
1	A	202	ILE
1	B	202	ILE
1	B	541	GLU

### 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	700/834 (84%)	674 (96%)	26 (4%)	45 66
1	B	699/834 (84%)	664 (95%)	35 (5%)	34 51
All	All	1399/1668 (84%)	1338 (96%)	61 (4%)	39 58

5 of 61 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	79	SER
1	B	237	LYS
1	B	830	ARG
1	B	131	ASP
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	194	ASN
1	B	796	ASN
1	B	128	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	176	GLN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 5 ligands modelled in this entry, 3 are monoatomic - leaving 2 for Mogul analysis.

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	X09	A	1900	-	14,14,14	0.72	0	21,21,21	1.80	3 (14%)
2	X09	B	1900	-	14,14,14	0.57	0	21,21,21	2.01	6 (28%)

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	X09	A	1900	-	-	0/1/31/31	0/0/2/2
2	X09	B	1900	-	-	0/1/31/31	0/0/2/2

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1900	X09	CAJ-CAI-CAK	-5.07	94.15	102.70
2	B	1900	X09	CAI-CAJ-CAL	-4.76	97.36	102.73
2	B	1900	X09	CAJ-CAI-CAK	-4.74	94.71	102.70
2	A	1900	X09	CAI-CAJ-CAL	-4.43	97.73	102.73
2	B	1900	X09	CAI-CAK-NAM	-3.00	100.59	104.55

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

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

### 6.1 Protein, DNA and RNA chains

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	851/1032 (82%)	-0.55	2 (0%) 93 94	6, 16, 27, 40	7 (0%)
1	B	851/1032 (82%)	-0.59	1 (0%) 93 95	4, 15, 26, 37	4 (0%)
All	All	1702/2064 (82%)	-0.57	3 (0%) 93 94	4, 15, 26, 40	11 (0%)

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	130	ASP	3.1
1	A	130	ASP	2.9
1	A	131	ASP	2.3

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

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

### 6.3 Carbohydrates

There are no carbohydrates in this entry.

### 6.4 Ligands

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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CD	B	1901	1/1	0.18	0.52	66,66,66,66	0
2	X09	A	1900	13/13	0.12	0.37	13,16,17,19	0
2	X09	B	1900	13/13	0.10	0.34	12,14,15,15	0
3	CD	A	1901	1/1	0.03	-2.59	11,11,11,11	0
3	CD	A	1902	1/1	0.03	-4.16	9,9,9,9	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.