WORLDWIDE PDB

PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report ⓘ

Mar 23, 2024 – 05:19 PM EDT

PDB ID	:	1A39
Title	:	HUMICOLA INSOLENS ENDOCELLULASE EGI S37W, P39W DOUBLE-MUTANT
Authors	:	Davies, G.J.; Ducros, V.; Lewis, R.J.; Borchert, T.V.; Schulein, M.
Deposited on	:	1998-01-28
Resolution	:	2.20 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.1
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

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 2.20 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.
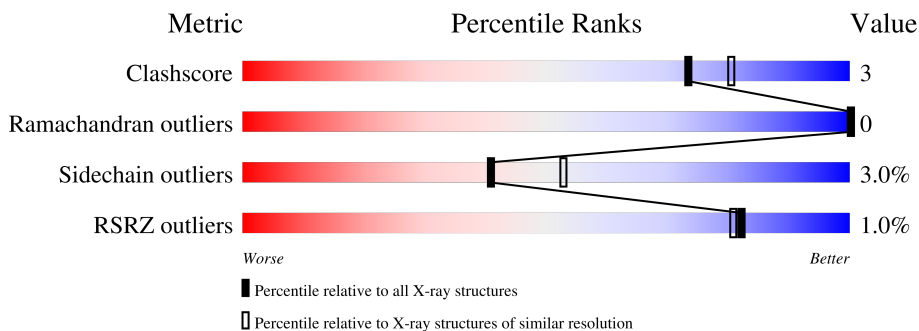
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 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% 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	402	

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 3479 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 ENDOGLUCANASE I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	402	Total 3138	C 1973	N 538	O 598	S 29	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	37	TRP	SER	conflict	UNP P56680
A	39	TRP	PRO	conflict	UNP P56680

- Molecule 2 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).
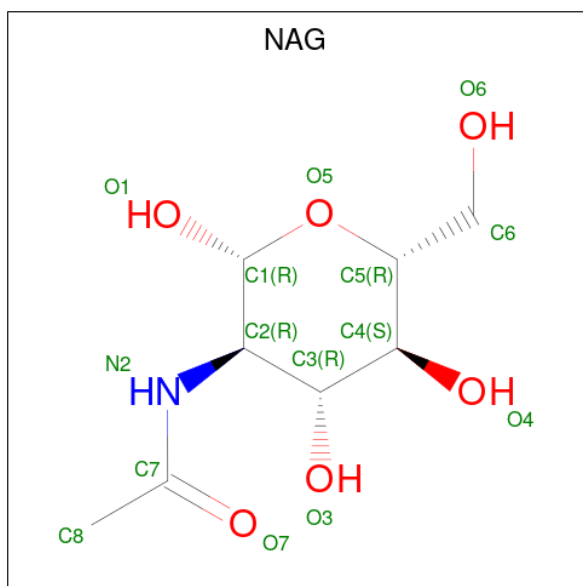
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 14	C 8	N 1	O 5	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	327	Total 327	O 327	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 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: ENDOGLUCANASE I
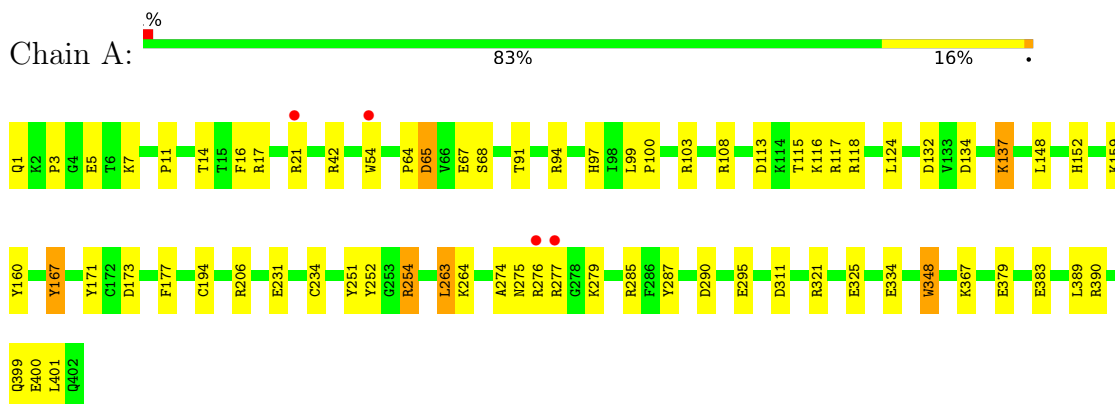
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	50.96Å 81.40Å 94.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.20 19.83 – 2.20	Depositor EDS
% Data completeness (in resolution range)	91.0 (20.00-2.20) 90.3 (19.83-2.20)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	4.37 (at 2.21Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.170 , 0.260 0.165 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	17.2	Xtriage
Anisotropy	0.327	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 57.8	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3479	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.90% 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 (i)

5.1 Standard geometry (i)

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

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	0.60	0/3215	1.50	42/4357 (1.0%)

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	A	0	1

There are no bond length outliers.

All (42) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	21	ARG	CD-NE-CZ	8.67	135.74	123.60
1	A	103	ARG	NE-CZ-NH1	8.55	124.57	120.30
1	A	1	PCA	C-N-CA	8.51	142.98	121.70
1	A	117	ARG	NE-CZ-NH1	8.44	124.52	120.30
1	A	285	ARG	NE-CZ-NH2	-7.77	116.42	120.30
1	A	1	PCA	O-C-N	-7.67	110.44	122.70
1	A	311	ASP	CB-CG-OD1	7.55	125.09	118.30
1	A	17	ARG	NE-CZ-NH2	7.50	124.05	120.30
1	A	251	TYR	CB-CG-CD2	-7.43	116.54	121.00
1	A	173	ASP	CB-CG-OD1	7.37	124.93	118.30
1	A	117	ARG	CD-NE-CZ	7.31	133.83	123.60
1	A	134	ASP	CB-CG-OD1	7.19	124.77	118.30
1	A	252	TYR	CB-CG-CD2	-7.14	116.72	121.00
1	A	152	HIS	N-CA-CB	-7.09	97.84	110.60
1	A	277	ARG	CA-C-N	7.05	130.30	116.20
1	A	21	ARG	NE-CZ-NH1	6.98	123.79	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	94	ARG	NE-CZ-NH2	-6.94	116.83	120.30
1	A	274	ALA	CA-C-N	6.56	131.63	117.20
1	A	254	ARG	NE-CZ-NH1	6.50	123.55	120.30
1	A	287	TYR	CB-CG-CD2	-6.49	117.11	121.00
1	A	132	ASP	CB-CG-OD1	6.46	124.12	118.30
1	A	251	TYR	CB-CG-CD1	6.25	124.75	121.00
1	A	276	ARG	CD-NE-CZ	-6.16	114.98	123.60
1	A	108	ARG	NE-CZ-NH1	6.15	123.38	120.30
1	A	118	ARG	NE-CZ-NH2	-6.11	117.25	120.30
1	A	321	ARG	NE-CZ-NH1	6.07	123.33	120.30
1	A	287	TYR	CB-CG-CD1	6.01	124.61	121.00
1	A	206	ARG	NE-CZ-NH2	5.96	123.28	120.30
1	A	171	TYR	CB-CG-CD1	5.77	124.46	121.00
1	A	108	ARG	CD-NE-CZ	5.69	131.56	123.60
1	A	277	ARG	CA-C-O	-5.66	108.21	120.10
1	A	252	TYR	CB-CG-CD1	5.65	124.39	121.00
1	A	167	TYR	CB-CG-CD1	5.56	124.33	121.00
1	A	42	ARG	CD-NE-CZ	5.41	131.18	123.60
1	A	132	ASP	CB-CG-OD2	-5.39	113.45	118.30
1	A	65	ASP	CB-CG-OD2	-5.31	113.53	118.30
1	A	231	GLU	OE1-CD-OE2	-5.30	116.94	123.30
1	A	67	GLU	OE1-CD-OE2	5.27	129.62	123.30
1	A	118	ARG	N-CA-CB	-5.13	101.37	110.60
1	A	167	TYR	CB-CG-CD2	-5.07	117.96	121.00
1	A	11	PRO	N-CA-CB	5.07	109.38	103.30
1	A	325	GLU	OE1-CD-OE2	5.06	129.37	123.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	5	GLU	Mainchain

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	3138	0	2965	20	0
2	A	14	0	13	0	0
3	A	327	0	0	5	0
All	All	3479	0	2978	20	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 3.

All (20) 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:115:THR:O	1:A:116:LYS:HB2	1.94	0.67
1:A:275:ASN:HB2	3:A:1120:HOH:O	1.99	0.63
1:A:116:LYS:HE2	3:A:924:HOH:O	2.06	0.55
1:A:194:CYS:HB3	1:A:234:CYS:SG	2.48	0.54
1:A:137:LYS:HD3	1:A:383:GLU:HG2	1.93	0.51
1:A:3:PRO:HB3	1:A:167:TYR:CZ	2.49	0.47
1:A:264:LYS:NZ	1:A:290:ASP:OD2	2.46	0.47
1:A:97:HIS:HE1	1:A:348:TRP:CD2	2.32	0.47
1:A:64:PRO:HD2	1:A:68:SER:HB2	1.96	0.47
1:A:379:GLU:HB2	3:A:1033:HOH:O	2.15	0.47
1:A:113:ASP:HB2	3:A:928:HOH:O	2.14	0.46
1:A:124:LEU:HD12	1:A:148:LEU:HD11	1.98	0.46
1:A:279:LYS:NZ	1:A:334:GLU:OE2	2.37	0.45
1:A:54:TRP:HZ2	1:A:177:PHE:CD1	2.37	0.42
1:A:159:LYS:NZ	1:A:160:TYR:OH	2.53	0.41
1:A:254:ARG:HB3	1:A:263:LEU:HD13	2.01	0.41
1:A:99:LEU:HB3	1:A:100:PRO:HD2	2.03	0.40
1:A:295:GLU:HG2	3:A:995:HOH:O	2.21	0.40
1:A:14:THR:HB	1:A:91:THR:HG22	2.03	0.40
1:A:16:PHE:HB2	1:A:390:ARG:HB3	2.03	0.40

There are no symmetry-related clashes.

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	400/402 (100%)	389 (97%)	11 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	334/334 (100%)	324 (97%)	10 (3%)	41	53

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	7	LYS
1	A	65	ASP
1	A	137	LYS
1	A	263	LEU
1	A	348	TRP
1	A	367	LYS
1	A	389	LEU
1	A	399	GLN
1	A	400	GLU
1	A	401	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	12	GLN
1	A	97	HIS
1	A	350	GLN
1	A	353	ASN
1	A	399	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	#$|Z| > 2$
1	PCA	A	1	1	7,8,9	1.06	0	9,10,12	1.45	2 (22%)

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
1	PCA	A	1	1	-	0/0/11/13	0/1/1/1

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	A	1	PCA	CB-CA-C	-2.64	109.07	112.70
1	A	1	PCA	CB-CA-N	2.05	109.18	103.30

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

1 ligand is 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	NAG	A	800	1	14,14,15	1.22	1 (7%)	17,19,21	1.35	4 (23%)

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	NAG	A	800	1	-	0/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	800	NAG	O7-C7	-3.86	1.14	1.23

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	800	NAG	C1-O5-C5	2.46	115.53	112.19
2	A	800	NAG	C4-C3-C2	-2.37	107.55	111.02
2	A	800	NAG	C2-N2-C7	2.31	126.19	122.90
2	A	800	NAG	O7-C7-C8	2.02	125.81	122.06

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 ⓘ

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, 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	401/402 (99%)	-0.56	4 (0%)	82	81	12, 20, 33, 50	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	54	TRP	4.9
1	A	21	ARG	4.2
1	A	276	ARG	4.1
1	A	277	ARG	3.1

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

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. 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	B-factors(Å^2)	Q<0.9
1	PCA	A	1	8/9	0.94	0.13	21,23,25,25	0

6.3   Carbohydrates ⓘ

There are no monosaccharides 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. 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	B-factors(Å^2)	Q<0.9
2	NAG	A	800	14/15	0.93	0.14	26,29,32,35	0

6.5 Other polymers ⓘ

There are no such residues in this entry.