wwPDB X-ray Structure Validation Summary Report (i)

Feb 11, 2024 – 09:57 AM EST

PDB ID	:	3CSQ
Title	:	Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage phi29 tail
Authors	:	Xiang, Y.; Rossmann, M.G.
Deposited on	:	2008-04-10
Resolution	:	1.80 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

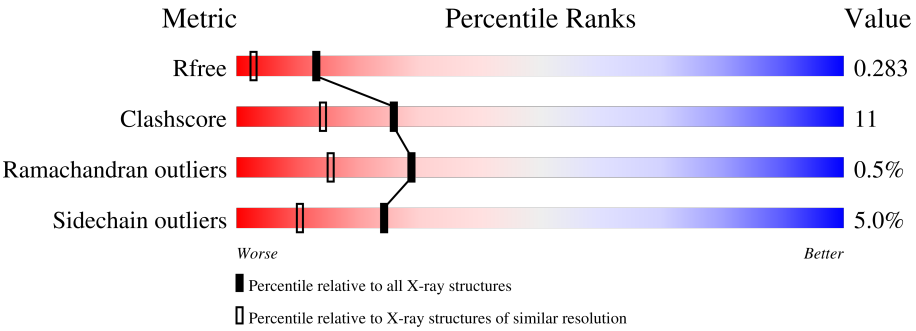
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 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 1.80 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}	130704	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)

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%

Mol	Chain	Length	Quality of chain
1	A	334	77% 19%
1	B	334	78% 17%
1	C	334	78% 16%
1	D	334	78% 16%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 10684 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 Morphogenesis protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	324	Total	C	N	O	S	0	3	0
			2614	1664	435	497	18			
1	B	324	Total	C	N	O	S	0	0	0
			2597	1652	433	494	18			
1	C	324	Total	C	N	O	S	0	4	0
			2616	1665	436	497	18			
1	D	326	Total	C	N	O	S	0	4	0
			2633	1676	443	496	18			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	89	ASN	ASP	SEE REMARK 999	UNP P15132
B	89	ASN	ASP	SEE REMARK 999	UNP P15132
C	89	ASN	ASP	SEE REMARK 999	UNP P15132
D	89	ASN	ASP	SEE REMARK 999	UNP P15132

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		
2	B	1	Total	Zn	0	0
			1	1		
2	C	1	Total	Zn	0	0
			1	1		
2	D	1	Total	Zn	0	0
			1	1		

- Molecule 3 is water.

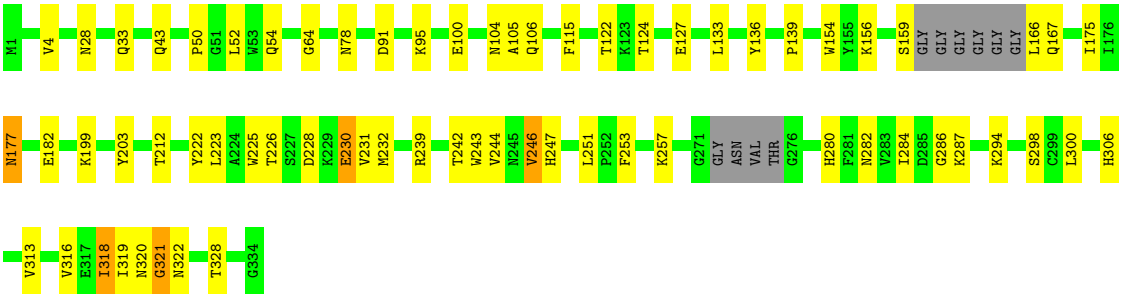
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	49	Total O 49 49	0	0
3	B	54	Total O 54 54	0	0
3	C	58	Total O 62 62	0	4
3	D	55	Total O 55 55	0	0

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. 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. 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: Morphogenesis protein 1

Chain A:

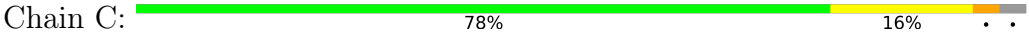


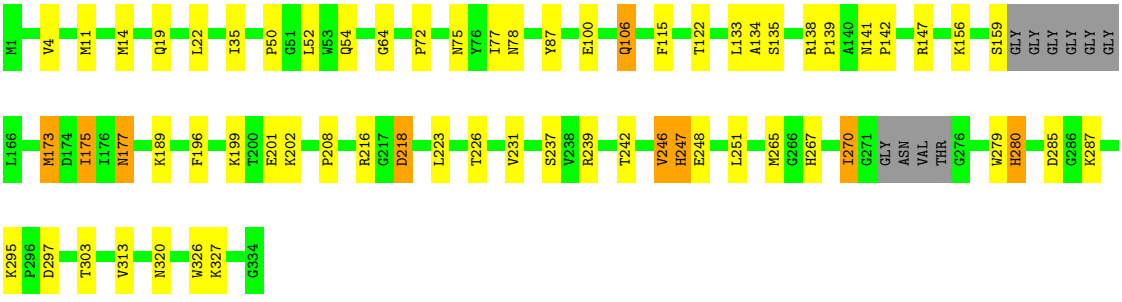
• Molecule 1: Morphogenesis protein 1

Chain B:

• Molecule 1: Morphogenesis protein 1

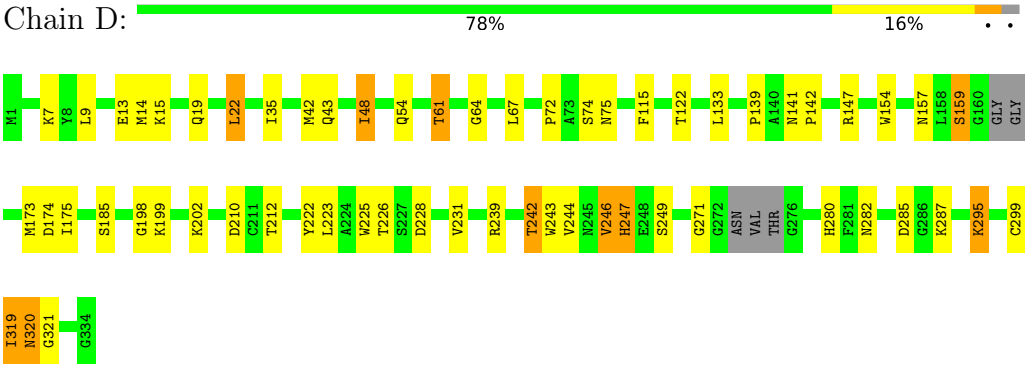
Chain C:





• Molecule 1: Morphogenesis protein 1

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	53.84Å 133.96Å 85.72Å 90.00° 89.99° 90.00°	Depositor
Resolution (Å)	19.90 – 1.80 19.70 – 1.80	Depositor EDS
% Data completeness (in resolution range)	97.6 (19.90-1.80) 97.5 (19.70-1.80)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	8.84 (at 1.80Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.238 , 0.282 0.239 , 0.283	Depositor DCC
R_{free} test set	5487 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	15.4	Xtriage
Anisotropy	0.012	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 23.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	0.450 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	10684	wwPDB-VP
Average B, all atoms (Å^2)	18.0	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 48.74 % 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 8.2421e-05.*

[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

5.1 Standard geometry

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

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	0.69	0/2694	0.76	1/3656 (0.0%)
1	B	0.72	0/2668	0.80	3/3622 (0.1%)
1	C	0.68	0/2699	0.74	0/3663
1	D	0.70	0/2716	0.74	0/3682
All	All	0.70	0/10777	0.76	4/14623 (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	B	0	1
1	D	0	1
All	All	0	2

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	321	GLY	N-CA-C	-9.52	89.31	113.10
1	A	321	GLY	N-CA-C	-6.14	97.74	113.10
1	B	320	ASN	C-N-CA	5.70	134.26	122.30
1	B	239	ARG	NE-CZ-NH2	5.11	122.85	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	320	ASN	Peptide
1	D	320	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	2614	0	2492	46	0
1	B	2597	0	2467	57	0
1	C	2616	0	2497	54	0
1	D	2633	0	2525	64	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	49	0	0	5	0
3	B	54	0	0	2	0
3	C	62	0	0	6	0
3	D	55	0	0	5	0
All	All	10684	0	9981	215	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 11.

The worst 5 of 215 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:174:ASP:OD2	1:D:175:ILE:HD12	1.46	1.15
1:C:246:VAL:HG22	1:C:280:HIS:HB3	1.35	1.07
1:B:42:MET:HE3	1:B:67:LEU:HB3	1.40	1.04
1:B:42:MET:CE	1:B:67:LEU:HB3	1.87	1.03
1:D:42:MET:HE2	1:D:48:ILE:HG23	1.39	1.02

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	321/334 (96%)	305 (95%)	15 (5%)	1 (0%)	41	27
1	B	318/334 (95%)	303 (95%)	13 (4%)	2 (1%)	25	12
1	C	322/334 (96%)	311 (97%)	9 (3%)	2 (1%)	25	12
1	D	324/334 (97%)	305 (94%)	17 (5%)	2 (1%)	25	12
All	All	1285/1336 (96%)	1224 (95%)	54 (4%)	7 (0%)	29	15

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	295	LYS
1	D	247	HIS
1	A	247	HIS
1	B	58	GLU
1	C	247	HIS

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	283/283 (100%)	268 (95%)	15 (5%)	22	9
1	B	280/283 (99%)	268 (96%)	12 (4%)	29	14
1	C	284/283 (100%)	268 (94%)	16 (6%)	21	8
1	D	284/283 (100%)	270 (95%)	14 (5%)	25	11
All	All	1131/1132 (100%)	1074 (95%)	57 (5%)	24	10

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	106	GLN
1	D	318	ILE
1	C	218	ASP
1	D	287	LYS
1	D	185	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	177	ASN
1	D	103	ASN
1	D	19	GLN
1	D	54	GLN
1	D	167	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 are monoatomic - 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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.