Full wwPDB X-ray Structure Validation Report (i)

Oct 11, 2023 – 05:25 PM EDT

PDB ID	:	7RG8
Title	:	Crystal Structure of a Stable Heparanase Mutant
Authors	:	Whitefield, C.; Hong, N.S.; Jackson, C.J.
Deposited on	:	2021-07-14
Resolution	:	1.30 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*
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:

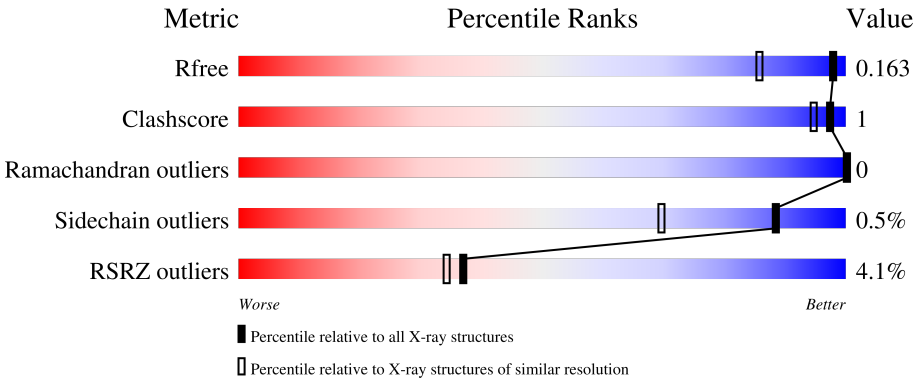
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 1.30 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	1058 (1.30-1.30)
Clashscore	141614	1101 (1.30-1.30)
Ramachandran outliers	138981	1058 (1.30-1.30)
Sidechain outliers	138945	1058 (1.30-1.30)
RSRZ outliers	127900	1029 (1.30-1.30)

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	387	4% 98%
2	B	92	2% 75% 5% 20%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7729 atoms, of which 3707 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 Heparanase 50 kDa subunit.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	385	6147	1971	3086	526	554	10	0	7	0

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	157	MET	-	initiating methionine	UNP Q9Y251
A	178	LYS	ASN	engineered mutation	UNP Q9Y251
A	195	SER	ALA	engineered mutation	UNP Q9Y251
A	197	GLY	LEU	engineered mutation	UNP Q9Y251
A	212	ALA	SER	engineered mutation	UNP Q9Y251
A	219	ASP	SER	engineered mutation	UNP Q9Y251
A	230	ARG	LEU	engineered mutation	UNP Q9Y251
A	234	GLY	ASP	engineered mutation	UNP Q9Y251
A	244	LYS	GLU	engineered mutation	UNP Q9Y251
A	248	HIS	GLN	engineered mutation	UNP Q9Y251
A	273	GLY	ARG	engineered mutation	UNP Q9Y251
A	292	ALA	SER	engineered mutation	UNP Q9Y251
A	307	LEU	LYS	engineered mutation	UNP Q9Y251
A	318	THR	ILE	engineered mutation	UNP Q9Y251
A	322	GLN	SER	engineered mutation	UNP Q9Y251
A	327	LEU	PHE	engineered mutation	UNP Q9Y251
A	354	GLY	LEU	engineered mutation	UNP Q9Y251
A	426	GLN	SER	engineered mutation	UNP Q9Y251
A	427	ASP	LYS	engineered mutation	UNP Q9Y251
A	477	GLN	LYS	engineered mutation	UNP Q9Y251
A	483	HIS	LEU	engineered mutation	UNP Q9Y251
A	486	ASP	HIS	engineered mutation	UNP Q9Y251
A	498	GLN	LEU	engineered mutation	UNP Q9Y251
A	512	LYS	MET	engineered mutation	UNP Q9Y251
A	513	PRO	GLU	engineered mutation	UNP Q9Y251
A	530	ALA	SER	engineered mutation	UNP Q9Y251
A	540	PRO	ALA	engineered mutation	UNP Q9Y251

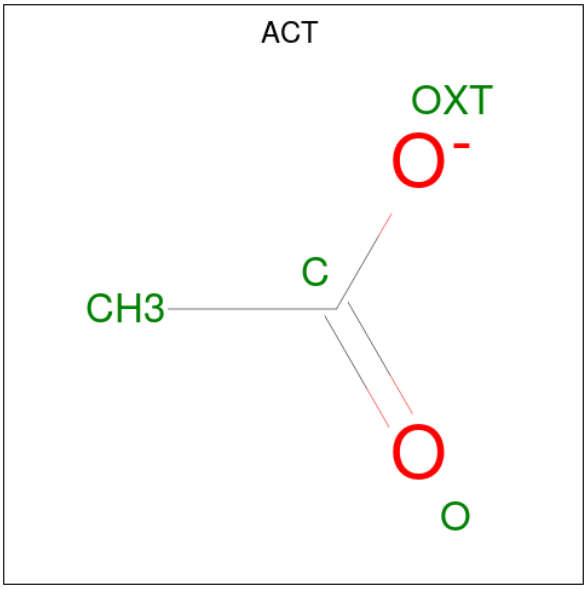
- Molecule 2 is a protein called Heparanase 8 kDa subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	H	N	O			
2	B	74	1217	394	618	96	109	0	3	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	18	MET	-	initiating methionine	UNP Q9Y251
B	19	GLY	-	expression tag	UNP Q9Y251
B	20	SER	-	expression tag	UNP Q9Y251
B	21	SER	-	expression tag	UNP Q9Y251
B	22	HIS	-	expression tag	UNP Q9Y251
B	23	HIS	-	expression tag	UNP Q9Y251
B	24	HIS	-	expression tag	UNP Q9Y251
B	25	HIS	-	expression tag	UNP Q9Y251
B	26	HIS	-	expression tag	UNP Q9Y251
B	27	HIS	-	expression tag	UNP Q9Y251
B	28	SER	-	expression tag	UNP Q9Y251
B	29	GLN	-	expression tag	UNP Q9Y251
B	30	ASP	-	expression tag	UNP Q9Y251
B	31	PRO	-	expression tag	UNP Q9Y251
B	32	ASN	-	expression tag	UNP Q9Y251
B	33	SER	-	expression tag	UNP Q9Y251
B	34	SER	-	expression tag	UNP Q9Y251
B	35	SER	-	expression tag	UNP Q9Y251

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			7	2	3	2		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total	Na	0	0
			3	3		
4	B	1	Total	Na	0	0
			1	1		

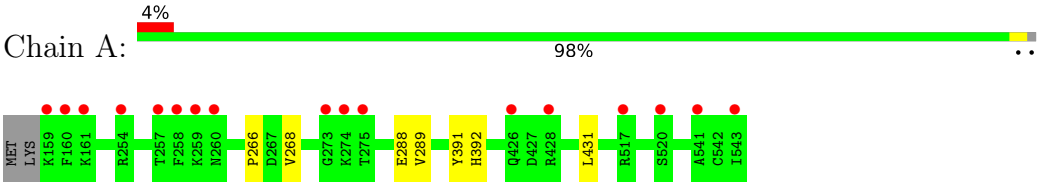
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	290	Total	O	0	0
			290	290		
5	B	64	Total	O	0	0
			64	64		

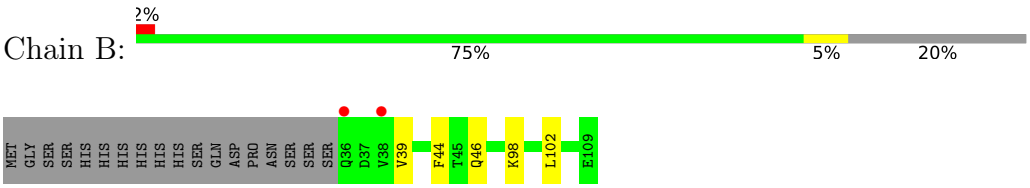
3 Residue-property plots (i)

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: Heparanase 50 kDa subunit



- Molecule 2: Heparanase 8 kDa subunit



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	59.78Å 76.09Å 124.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.01 – 1.30 53.89 – 1.30	Depositor EDS
% Data completeness (in resolution range)	99.8 (47.01-1.30) 99.9 (53.89-1.30)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.60 (at 1.30Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.143 , 0.165 0.142 , 0.163	Depositor DCC
R_{free} test set	6944 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	17.0	Xtriage
Anisotropy	0.572	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 47.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	7729	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.55% 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

5.1 Standard geometry

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

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	0.40	0/3171	0.65	0/4291
2	B	0.40	0/623	0.65	0/846
All	All	0.40	0/3794	0.65	0/5137

There are no bond length outliers.

There are no bond angle outliers.

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	3061	3086	3058	4	0
2	B	599	618	610	4	0
3	A	4	3	3	0	0
4	A	3	0	0	0	0
4	B	1	0	0	0	0
5	A	290	0	0	0	0
5	B	64	0	0	0	0
All	All	4022	3707	3671	6	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 1.

All (6) 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:266:PRO:HG2	1:A:268:VAL:HG13	1.88	0.55
1:A:431:LEU:HD22	2:B:39:VAL:HG21	1.95	0.49
1:A:431:LEU:HD23	2:B:39:VAL:HG11	1.97	0.47
2:B:98:LYS:HG3	2:B:102:LEU:HG	1.97	0.46
1:A:288[A]:GLU:HG2	1:A:289:VAL:HG13	1.99	0.45
2:B:44:PHE:CE1	2:B:46:GLN:HG2	2.55	0.42

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	390/387 (101%)	384 (98%)	6 (2%)	0	100	100
2	B	75/92 (82%)	75 (100%)	0	0	100	100
All	All	465/479 (97%)	459 (99%)	6 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	337/332 (102%)	335 (99%)	2 (1%)	86	65

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	B	69/83 (83%)	69 (100%)	0	100 100
All	All	406/415 (98%)	404 (100%)	2 (0%)	88 69

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	391	TYR
1	A	392	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
3	ACT	A	601	-	3,3,3	1.03	0	3,3,3	1.15	0

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

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^{th} 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	385/387 (99%)	-0.00	17 (4%)	34	32	15, 22, 43, 66	0
2	B	74/92 (80%)	-0.04	2 (2%)	54	52	15, 20, 36, 59	0
All	All	459/479 (95%)	-0.01	19 (4%)	37	34	15, 22, 43, 66	0

All (19) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	543	ILE	9.0
1	A	259	LYS	8.6
2	B	36	GLN	6.0
1	A	160	PHE	5.8
1	A	260	ASN	5.4
1	A	426	GLN	4.6
1	A	159	LYS	3.8
1	A	254	ARG	3.6
1	A	275	THR	3.4
1	A	257	THR	3.4
1	A	428	ARG	3.2
1	A	161	LYS	2.9
1	A	274	LYS	2.9
1	A	258	PHE	2.7
2	B	38	VAL	2.7
1	A	541	ALA	2.3
1	A	520	SER	2.2
1	A	273	GLY	2.2
1	A	517	ARG	2.1

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 (i)

There are no monosaccharides in this entry.

6.4 Ligands (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. The B-factors column lists the minimum, median, 95^{th} 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
3	ACT	A	601	4/4	0.99	0.06	19,21,26,26	0
4	NA	A	604	1/1	0.99	0.05	26,26,26,26	0
4	NA	A	603	1/1	1.00	0.06	23,23,23,23	0
4	NA	A	602	1/1	1.00	0.20	16,16,16,16	0
4	NA	B	201	1/1	1.00	0.04	25,25,25,25	0

6.5 Other polymers (i)

There are no such residues in this entry.