



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 07:54 PM GMT

PDB ID : 1HQN  
Title : THE SELENOMETHIONINE DERIVATIVE OF P3, THE MAJOR COAT PROTEIN OF THE LIPID-CONTAINING BACTERIOPHAGE PRD1.  
Authors : Benson, S.D.; Bamford, J.K.H.; Bamford, D.H.; Burnett, R.M.  
Deposited on : 2000-12-18  
Resolution : 2.20 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

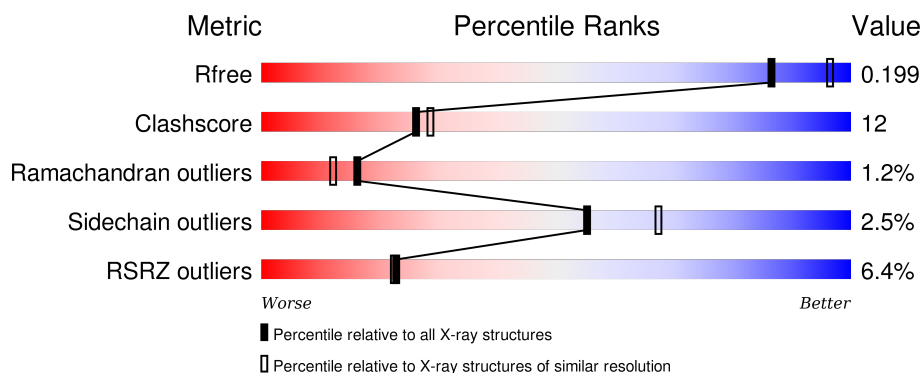
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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))
$R_{free}$	91344	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (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 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	394	<div> <div>6%</div> <div> <div></div> <div>73%</div> <div>19%</div> <div>• 6%</div> </div> </div>
1	B	394	<div> <div>6%</div> <div> <div></div> <div>76%</div> <div>15%</div> <div>• 6%</div> </div> </div>
1	C	394	<div> <div>6%</div> <div> <div></div> <div>76%</div> <div>18%</div> <div>• 6%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8974 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 MAJOR CAPSID PROTEIN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	369	Total	C	N	O	S	Se	0	0	0
			2842	1807	479	549	2	5			
1	B	370	Total	C	N	O	S	Se	0	0	0
			2832	1800	478	547	2	5			
1	C	372	Total	C	N	O	S	Se	0	0	0
			2863	1819	483	554	2	5			

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	21	MSE	MET	MODIFIED RESIDUE	UNP P22535
A	133	MSE	MET	MODIFIED RESIDUE	UNP P22535
A	145	MSE	MET	MODIFIED RESIDUE	UNP P22535
A	164	MSE	MET	MODIFIED RESIDUE	UNP P22535
A	375	MSE	MET	MODIFIED RESIDUE	UNP P22535
B	21	MSE	MET	MODIFIED RESIDUE	UNP P22535
B	133	MSE	MET	MODIFIED RESIDUE	UNP P22535
B	145	MSE	MET	MODIFIED RESIDUE	UNP P22535
B	164	MSE	MET	MODIFIED RESIDUE	UNP P22535
B	375	MSE	MET	MODIFIED RESIDUE	UNP P22535
C	21	MSE	MET	MODIFIED RESIDUE	UNP P22535
C	133	MSE	MET	MODIFIED RESIDUE	UNP P22535
C	145	MSE	MET	MODIFIED RESIDUE	UNP P22535
C	164	MSE	MET	MODIFIED RESIDUE	UNP P22535
C	375	MSE	MET	MODIFIED RESIDUE	UNP P22535

- Molecule 2 is water.

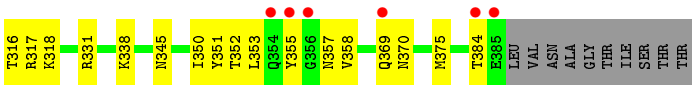
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	147	Total	O	0	0
			147	147		
2	B	167	Total	O	0	0
			167	167		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	123	Total 123	O 123	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	117.96Å 121.30Å 126.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.63 – 2.20 48.91 – 2.14	Depositor EDS
% Data completeness (in resolution range)	88.9 (40.63-2.20) 86.0 (48.91-2.14)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	6.50	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.29 (at 2.14Å)	Xtriage
Refinement program	CNS 0.9	Depositor
R, $R_{free}$	0.200 , 0.227 0.211 , 0.199	Depositor DCC
$R_{free}$ test set	103 reflections (0.11%)	DCC
Wilson B-factor (Å <sup>2</sup> )	22.1	Xtriage
Anisotropy	0.185	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 57.9	EDS
Estimated twinning fraction	0.011 for k,h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 145672 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	8974	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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](#)

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	0.37	0/2902	0.67	0/3962
1	B	0.38	0/2891	0.67	1/3948 (0.0%)
1	C	0.37	0/2923	0.67	0/3991
All	All	0.37	0/8716	0.67	1/11901 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	162	LEU	CA-CB-CG	5.37	127.64	115.30

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	2842	0	2733	67	0
1	B	2832	0	2716	66	0
1	C	2863	0	2753	79	0
2	A	147	0	0	7	0
2	B	167	0	0	1	0
2	C	123	0	0	1	0
All	All	8974	0	8202	205	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 12.

The worst 5 of 205 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:164:MSE:HE1	1:C:139:ILE:HD11	1.15	1.10
1:B:164:MSE:CE	1:C:139:ILE:HD11	1.90	1.01
1:A:310:THR:HG22	1:A:358:VAL:HG22	1.40	1.01
1:B:92:PRO:HG2	1:B:164:MSE:HE2	1.48	0.96
1:C:281:LEU:HD23	1:C:353:LEU:HD21	1.52	0.92

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	367/394 (93%)	343 (94%)	18 (5%)	6 (2%)	12	8
1	B	368/394 (93%)	345 (94%)	19 (5%)	4 (1%)	17	14
1	C	370/394 (94%)	341 (92%)	26 (7%)	3 (1%)	24	22
All	All	1105/1182 (94%)	1029 (93%)	63 (6%)	13 (1%)	16	12

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	353	LEU
1	C	280	ASN
1	C	314	SER
1	A	311	ALA
1	B	315	ASP

### 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	296 / 320 (92%)	289 (98%)	7 (2%)	57	69
1	B	293 / 320 (92%)	284 (97%)	9 (3%)	47	59
1	C	298 / 320 (93%)	292 (98%)	6 (2%)	63	76
All	All	887 / 960 (92%)	865 (98%)	22 (2%)	55	67

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	52	ARG
1	B	164	MSE
1	C	331	ARG
1	B	161	GLU
1	B	162	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	80	HIS
1	B	239	GLN
1	C	280	ASN
1	B	107	GLN
1	B	184	ASN

### 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 carbohydrates in this entry.

## 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 [i](#)

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	364/394 (92%)	-0.15	22 (6%) 25 25	10, 25, 55, 67	0
1	B	365/394 (92%)	-0.16	24 (6%) 22 21	10, 24, 59, 65	0
1	C	367/394 (93%)	-0.08	24 (6%) 22 22	12, 28, 59, 67	0
All	All	1096/1182 (92%)	-0.13	70 (6%) 23 22	10, 26, 58, 67	0

The worst 5 of 70 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	186	PRO	7.6
1	C	355	TYR	7.5
1	A	313	PHE	6.1
1	C	311	ALA	5.9
1	B	353	LEU	5.8

### 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](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.