wwPDB X-ray Structure Validation Summary Report (i)

Feb 3, 2024 – 08:24 PM EST

PDB ID	:	1MSP
Title	:	MAJOR SPERM PROTEIN, ALPHA ISOFORM (RECOMBINANT), PH 4.6
Authors	:	Bullock, T.L.; Roberts, T.M.; Stewart, M.
Deposited on	:	1996-05-20
Resolution	:	2.50 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 (i)) were used in the production of this report:

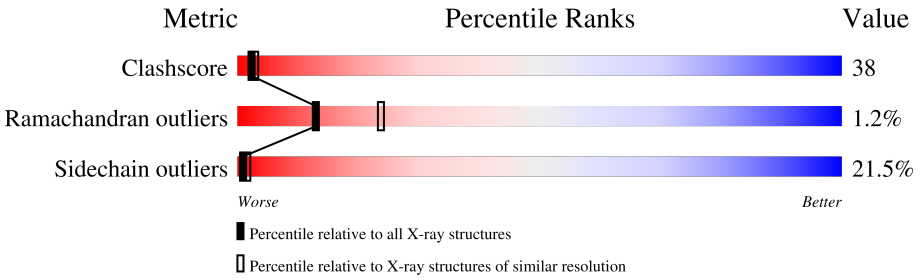
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.50 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	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	126	44% 40% 10%
1	B	126	37% 40% 17%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1931 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 SPERM PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	124	Total	C	N	O	S	0	0	0
			948	599	163	182	4			
1	B	122	Total	C	N	O	S	0	0	0
			909	573	157	176	3			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	40	Total	O	0	0
			40	40		
2	B	34	Total	O	0	0
			34	34		

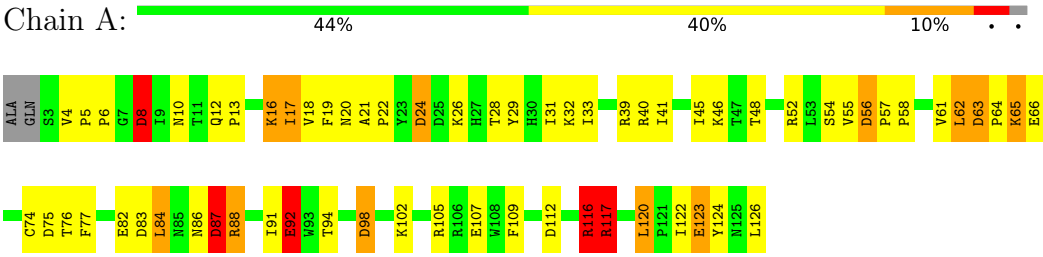
3 Residue-property plots (i)

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.

Note EDS was not executed.

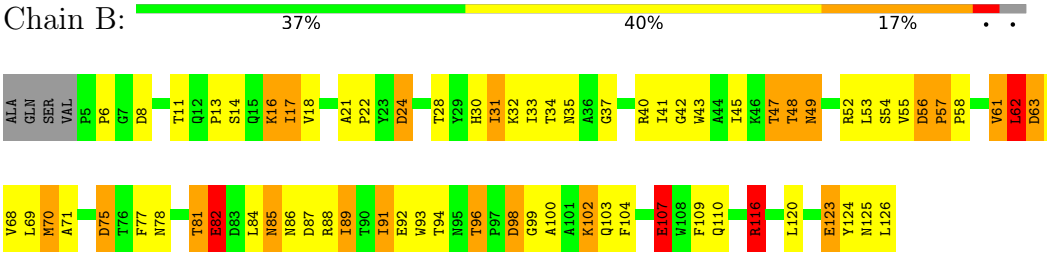
- Molecule 1: MAJOR SPERM PROTEIN

Chain A:



- Molecule 1: MAJOR SPERM PROTEIN

Chain B:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	215.36Å 38.48Å 32.46Å 90.00° 93.11° 90.00°	Depositor
Resolution (Å)	20.00 – 2.50	Depositor
% Data completeness (in resolution range)	(Not available) (20.00-2.50)	Depositor
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	TNT	Depositor
R, R_{free}	(Not available) , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1931	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

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	1.00	5/970 (0.5%)	1.60	28/1321 (2.1%)
1	B	0.93	4/931 (0.4%)	1.53	16/1268 (1.3%)
All	All	0.97	9/1901 (0.5%)	1.57	44/2589 (1.7%)

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	1	0

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	107	GLU	CD-OE2	6.54	1.32	1.25
1	A	92	GLU	CD-OE2	6.50	1.32	1.25
1	A	82	GLU	CD-OE2	6.38	1.32	1.25
1	B	66	GLU	CD-OE2	6.26	1.32	1.25
1	B	123	GLU	CD-OE2	5.81	1.32	1.25

The worst 5 of 44 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	112	ASP	CB-CG-OD2	-9.79	109.49	118.30
1	B	116	ARG	NE-CZ-NH1	9.20	124.90	120.30
1	A	24	ASP	CB-CG-OD2	-8.06	111.04	118.30
1	A	112	ASP	CB-CG-OD1	7.79	125.31	118.30
1	A	116	ARG	NE-CZ-NH2	-7.59	116.50	120.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	117	ARG	CA

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	948	0	894	61	0
1	B	909	0	841	81	2
2	A	40	0	0	4	0
2	B	34	0	0	6	2
All	All	1931	0	1735	138	2

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 38.

The worst 5 of 138 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:116:ARG:HG2	1:A:116:ARG:HH11	1.32	0.92
1:A:94:THR:HG22	1:A:109:PHE:HE1	1.38	0.88
1:B:47:THR:CG2	1:B:49:ASN:H	1.95	0.78
1:B:33:ILE:HD11	1:B:70:MET:SD	2.24	0.77
1:B:31:ILE:O	1:B:69:LEU:HD12	1.85	0.77

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	Interatomic distance (Å)	Clash overlap (Å)
1:B:110:GLN:CG	2:B:158:HOH:O[4_548]	1.43	0.77
1:B:110:GLN:CB	2:B:158:HOH:O[4_548]	1.83	0.37

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	122/126 (97%)	116 (95%)	6 (5%)	0	100 100
1	B	120/126 (95%)	110 (92%)	7 (6%)	3 (2%)	5 8
All	All	242/252 (96%)	226 (93%)	13 (5%)	3 (1%)	13 24

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	107	GLU
1	B	102	LYS
1	B	98	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	96/109 (88%)	79 (82%)	17 (18%)	2 3
1	B	90/109 (83%)	67 (74%)	23 (26%)	0 1
All	All	186/218 (85%)	146 (78%)	40 (22%)	1 1

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	62	LEU
1	B	91	ILE
1	B	70	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	82	GLU
1	B	96	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	30	HIS
1	A	110	GLN
1	B	30	HIS

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)

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

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.