



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 12:16 pm GMT

PDB ID : 2MSP  
Title : MAJOR SPERM PROTEIN, BETA ISOFORM, ENGINEERED C59S/T90C  
MUTANT, PUTATIVE SUBFILAMENT STRUCTURE, PH 8.5  
Authors : Bullock, T.L.; Mccoy, A.J.; Stewart, M.  
Deposited on : 1997-12-19  
Resolution : 3.30 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  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

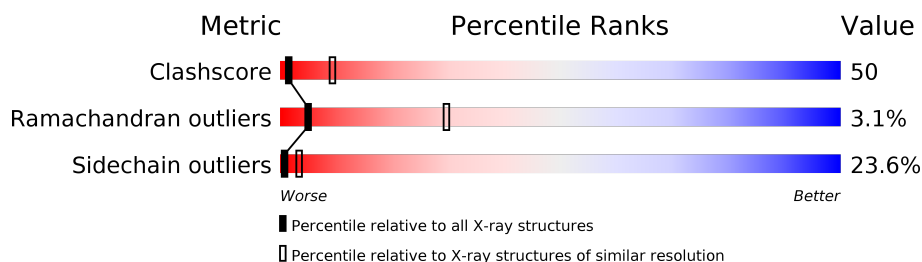
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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))
Clashscore	112137	1100 (3.36-3.24)
Ramachandran outliers	110173	1081 (3.36-3.24)
Sidechain outliers	110143	1080 (3.36-3.24)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	126	
1	B	126	
1	C	126	
1	D	126	
1	E	126	
1	F	126	
1	G	126	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	126	<div><div></div><div>30%</div><div>49%</div><div>17%</div><div></div></div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 7824 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	57	0	0
			978	610	178	185	5			
1	B	124	Total	C	N	O	S	57	0	0
			978	610	178	185	5			
1	C	124	Total	C	N	O	S	57	0	0
			978	610	178	185	5			
1	D	124	Total	C	N	O	S	57	0	0
			978	610	178	185	5			
1	E	124	Total	C	N	O	S	57	0	0
			978	610	178	185	5			
1	F	124	Total	C	N	O	S	57	0	0
			978	610	178	185	5			
1	G	124	Total	C	N	O	S	57	0	0
			978	610	178	185	5			
1	H	124	Total	C	N	O	S	51	0	0
			978	610	178	185	5			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	59	SER	CYS	ENGINEERED	UNP P27440
A	65	SER	LYS	CONFLICT	UNP P27440
A	67	LYS	SER	CONFLICT	UNP P27440
A	90	CYS	THR	ENGINEERED	UNP P27440
B	59	SER	CYS	ENGINEERED	UNP P27440
B	65	SER	LYS	CONFLICT	UNP P27440
B	67	LYS	SER	CONFLICT	UNP P27440
B	90	CYS	THR	ENGINEERED	UNP P27440
C	59	SER	CYS	ENGINEERED	UNP P27440
C	65	SER	LYS	CONFLICT	UNP P27440
C	67	LYS	SER	CONFLICT	UNP P27440
C	90	CYS	THR	ENGINEERED	UNP P27440
D	59	SER	CYS	ENGINEERED	UNP P27440

*Continued on next page...*

*Continued from previous page...*

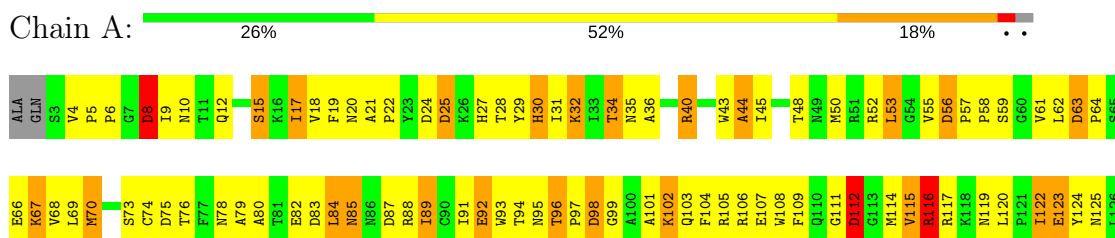
Chain	Residue	Modelled	Actual	Comment	Reference
D	65	SER	LYS	CONFLICT	UNP P27440
D	67	LYS	SER	CONFLICT	UNP P27440
D	90	CYS	THR	ENGINEERED	UNP P27440
E	59	SER	CYS	ENGINEERED	UNP P27440
E	65	SER	LYS	CONFLICT	UNP P27440
E	67	LYS	SER	CONFLICT	UNP P27440
E	90	CYS	THR	ENGINEERED	UNP P27440
F	59	SER	CYS	ENGINEERED	UNP P27440
F	65	SER	LYS	CONFLICT	UNP P27440
F	67	LYS	SER	CONFLICT	UNP P27440
F	90	CYS	THR	ENGINEERED	UNP P27440
G	59	SER	CYS	ENGINEERED	UNP P27440
G	65	SER	LYS	CONFLICT	UNP P27440
G	67	LYS	SER	CONFLICT	UNP P27440
G	90	CYS	THR	ENGINEERED	UNP P27440
H	59	SER	CYS	ENGINEERED	UNP P27440
H	65	SER	LYS	CONFLICT	UNP P27440
H	67	LYS	SER	CONFLICT	UNP P27440
H	90	CYS	THR	ENGINEERED	UNP P27440

### 3 Residue-property plots [i](#)

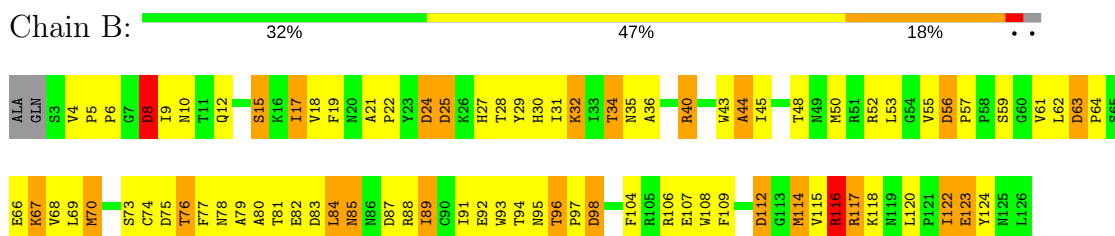
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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.

Note EDS was not executed.

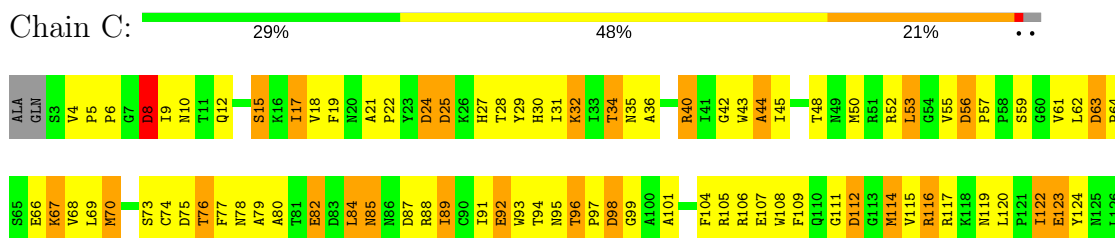
#### • Molecule 1: MAJOR SPERM PROTEIN



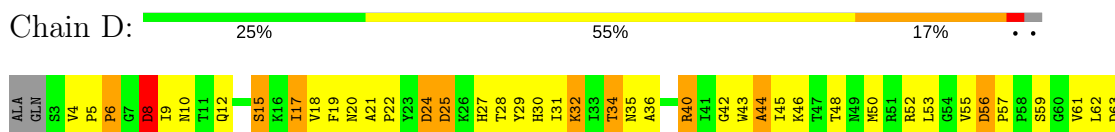
#### • Molecule 1: MAJOR SPERM PROTEIN

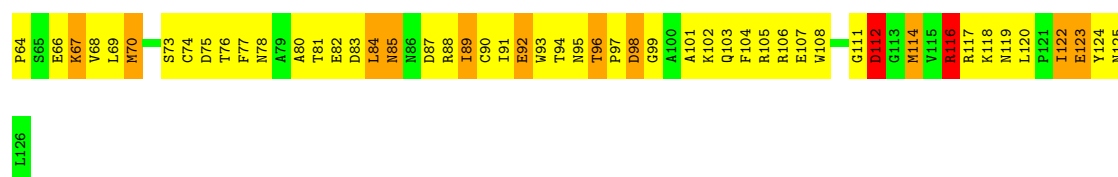


#### • Molecule 1: MAJOR SPERM PROTEIN



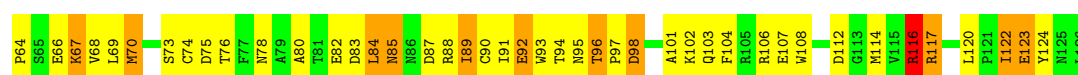
#### • Molecule 1: MAJOR SPERM PROTEIN





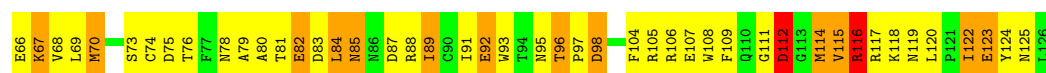
• Molecule 1: MAJOR SPERM PROTEIN

Chain E: 32% 47% 18% ..



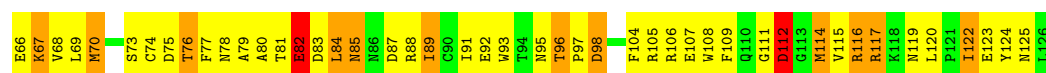
• Molecule 1: MAJOR SPERM PROTEIN

Chain F: 29% 48% 18% ..



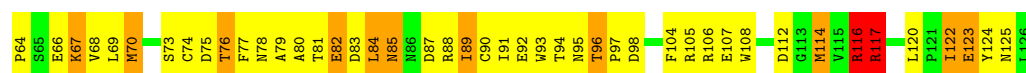
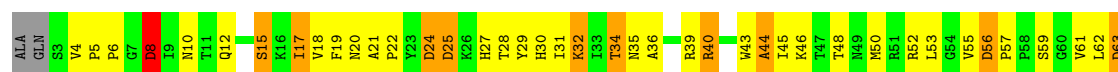
• Molecule 1: MAJOR SPERM PROTEIN

Chain G: 29% 50% 17% ..



• Molecule 1: MAJOR SPERM PROTEIN

Chain H: 30% 49% 17% ..



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.66Å 79.66Å 463.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	10.00 – 3.30	Depositor
% Data completeness (in resolution range)	100.0 (10.00-3.30)	Depositor
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	TNT V. 5-D	Depositor
R, $R_{free}$	0.218 , 0.260	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	7824	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	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	1.30	5/1001 (0.5%)	1.34	22/1358 (1.6%)
1	B	1.31	5/1001 (0.5%)	1.36	21/1358 (1.5%)
1	C	1.30	6/1001 (0.6%)	1.32	19/1358 (1.4%)
1	D	1.31	5/1001 (0.5%)	1.33	19/1358 (1.4%)
1	E	1.30	5/1001 (0.5%)	1.36	20/1358 (1.5%)
1	F	1.29	5/1001 (0.5%)	1.34	22/1358 (1.6%)
1	G	1.29	6/1001 (0.6%)	1.32	20/1358 (1.5%)
1	H	1.28	5/1001 (0.5%)	1.35	22/1358 (1.6%)
All	All	1.30	42/8008 (0.5%)	1.34	165/10864 (1.5%)

The worst 5 of 42 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	82	GLU	CD-OE2	7.36	1.33	1.25
1	D	66	GLU	CD-OE2	7.22	1.33	1.25
1	D	82	GLU	CD-OE2	7.03	1.33	1.25
1	A	82	GLU	CD-OE2	6.91	1.33	1.25
1	B	82	GLU	CD-OE2	6.87	1.33	1.25

The worst 5 of 165 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	112	ASP	CB-CG-OD2	-9.63	109.63	118.30
1	H	112	ASP	CB-CG-OD1	8.09	125.58	118.30
1	B	112	ASP	CB-CG-OD1	8.02	125.52	118.30
1	H	112	ASP	CB-CG-OD2	-7.97	111.13	118.30
1	H	117	ARG	NE-CZ-NH2	-7.66	116.47	120.30

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	978	0	954	96	0
1	B	978	0	954	87	0
1	C	978	0	954	93	0
1	D	978	0	954	91	0
1	E	978	0	954	88	0
1	F	978	0	954	95	0
1	G	978	0	954	93	0
1	H	978	0	954	96	0
All	All	7824	0	7632	716	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 50.

The worst 5 of 716 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:C:21:ALA:HB1	1:C:22:PRO:HA	1.47	0.96
1:F:21:ALA:HB1	1:F:22:PRO:HA	1.47	0.95
1:G:21:ALA:HB1	1:G:22:PRO:HA	1.48	0.95
1:E:85:ASN:HD22	1:E:85:ASN:H	1.16	0.94
1:A:85:ASN:H	1:A:85:ASN:HD22	1.13	0.94

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	122/126 (97%)	103 (84%)	14 (12%)	5 (4%)	3	22
1	B	122/126 (97%)	103 (84%)	15 (12%)	4 (3%)	4	28
1	C	122/126 (97%)	102 (84%)	16 (13%)	4 (3%)	4	28
1	D	122/126 (97%)	103 (84%)	15 (12%)	4 (3%)	4	28
1	E	122/126 (97%)	104 (85%)	15 (12%)	3 (2%)	6	34
1	F	122/126 (97%)	102 (84%)	16 (13%)	4 (3%)	4	28
1	G	122/126 (97%)	103 (84%)	15 (12%)	4 (3%)	4	28
1	H	122/126 (97%)	103 (84%)	17 (14%)	2 (2%)	11	43
All	All	976/1008 (97%)	823 (84%)	123 (13%)	30 (3%)	5	30

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	79	ALA
1	A	98	ASP
1	B	98	ASP
1	C	79	ALA
1	C	98	ASP

### 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	106/107 (99%)	81 (76%)	25 (24%)	1	3
1	B	106/107 (99%)	82 (77%)	24 (23%)	1	4
1	C	106/107 (99%)	80 (76%)	26 (24%)	1	3
1	D	106/107 (99%)	80 (76%)	26 (24%)	1	3
1	E	106/107 (99%)	81 (76%)	25 (24%)	1	3
1	F	106/107 (99%)	81 (76%)	25 (24%)	1	3
1	G	106/107 (99%)	81 (76%)	25 (24%)	1	3
1	H	106/107 (99%)	82 (77%)	24 (23%)	1	4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	848/856 (99%)	648 (76%)	200 (24%)	<b>1</b> <b>3</b>

5 of 200 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	81	THR
1	E	50	MET
1	H	40	ARG
1	D	88	ARG
1	E	8	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	85	ASN
1	D	119	ASN
1	G	27	HIS
1	D	27	HIS
1	G	85	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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.