



# Full wwPDB X-ray Structure Validation Report ⓘ

May 28, 2020 – 08:26 pm BST

PDB ID : 1SCJ  
Title : CRYSTAL STRUCTURE OF SUBTILISIN-PROPEPTIDE COMPLEX  
Authors : Berman, H.M.; Jain, S.C.  
Deposited on : 1998-04-29  
Resolution : 2.00 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

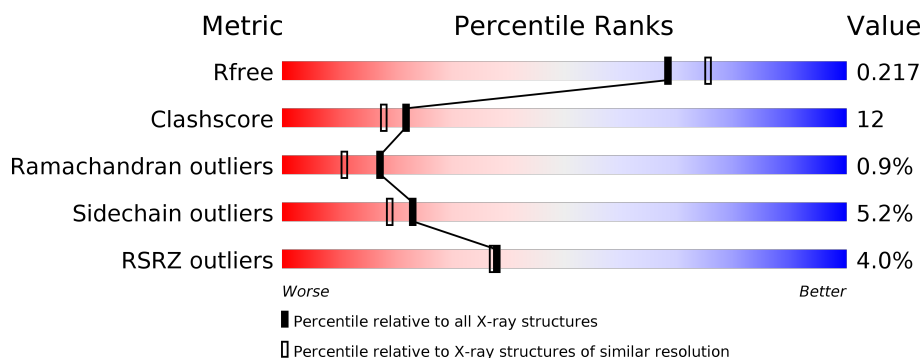
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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  $\leq 5\%$ . 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	275	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="width: 86%; height: 10px; background-color: green;"></div> <div style="width: 12%; height: 10px; background-color: yellow;"></div> <div style="width: 1%; height: 10px; background-color: orange;"></div> <div style="width: 1%; height: 10px; background-color: grey;"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <span>18%</span> <span>86%</span> <span>12%</span> <span>.</span> </div> </div>
2	B	71	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="width: 51%; height: 10px; background-color: green;"></div> <div style="width: 34%; height: 10px; background-color: yellow;"></div> <div style="width: 15%; height: 10px; background-color: orange;"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <span>18%</span> <span>51%</span> <span>34%</span> <span>15%</span> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2703 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 SUBTILISIN E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	275	Total	C	N	O	S	0	0	0
			1952	1211	336	400	5			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	221	CYS	SER	ENGINEERED MUTATION	UNP P04189

- Molecule 2 is a protein called SUBTILISIN E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	71	Total	C	N	O	S	0	0	0
			560	356	93	109	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	332	GLN	GLU	CONFLICT	UNP P04189
B	338	GLU	GLN	CONFLICT	UNP P04189

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	Ca	0	0
			2	2		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	168	Total	O	0	0
			168	168		

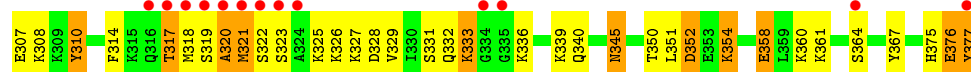
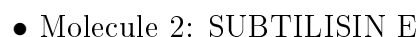
*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	21	Total	O	0	0
			21	21		



- Molecule 1: SUBTILISIN E



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.54Å 92.21Å 47.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00 39.85 – 1.81	Depositor EDS
% Data completeness (in resolution range)	87.0 (20.00-2.00) 73.5 (39.85-1.81)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.82 (at 1.81Å)	Xtriage
Refinement program	CNS 0.3	Depositor
R, $R_{free}$	0.173 , 0.220 0.171 , 0.217	Depositor DCC
$R_{free}$ test set	2211 reflections (9.90%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	13.8	Xtriage
Anisotropy	0.306	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 51.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2703	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.39% 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, 0.333 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:  
CA

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	0.62	0/1991	1.07	6/2719 (0.2%)
2	B	0.47	0/568	1.42	8/755 (1.1%)
All	All	0.59	0/2559	1.15	14/3474 (0.4%)

There are no bond length outliers.

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	358	GLU	OE1-CD-OE2	-9.35	112.08	123.30
1	A	249	ARG	NE-CZ-NH2	-7.64	116.48	120.30
2	B	352	ASP	CB-CG-OD2	-7.16	111.85	118.30
2	B	310	TYR	CB-CG-CD1	-6.84	116.90	121.00
1	A	221	CYS	CB-CA-C	-6.58	97.25	110.40
2	B	352	ASP	CB-CG-OD1	6.49	124.14	118.30
2	B	317	THR	CA-CB-CG2	6.31	121.24	112.40
1	A	58	TYR	CB-CG-CD2	6.01	124.61	121.00
2	B	367	TYR	CB-CG-CD1	5.84	124.50	121.00
1	A	249	ARG	NE-CZ-NH1	5.51	123.06	120.30
1	A	247	ARG	NE-CZ-NH2	-5.49	117.56	120.30
2	B	310	TYR	CB-CG-CD2	5.47	124.28	121.00
1	A	97	ASP	CB-CG-OD1	5.31	123.08	118.30
2	B	367	TYR	CB-CG-CD2	-5.30	117.82	121.00

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	1952	0	1902	31	0
2	B	560	0	573	36	0
3	A	2	0	0	0	0
4	A	168	0	0	1	0
4	B	21	0	0	2	0
All	All	2703	0	2475	59	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.

All (59) 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:155:ASN:HD21	1:A:220:THR:H	1.10	0.93
2:B:321:MET:HA	2:B:326:LYS:HE3	1.53	0.91
1:A:221:CYS:SG	1:A:222:MET:SD	2.80	0.80
1:A:155:ASN:ND2	1:A:220:THR:H	1.85	0.73
1:A:221:CYS:HB3	2:B:377:TYR:OXT	1.92	0.70
1:A:16:LEU:HD13	1:A:233:LEU:HB3	1.77	0.67
2:B:336:LYS:HB3	2:B:350:THR:OG1	1.93	0.67
2:B:308:LYS:HB2	2:B:310:TYR:CE2	2.31	0.65
1:A:238:HIS:HE1	1:A:274:ALA:O	1.79	0.64
2:B:325:LYS:O	2:B:329:VAL:HG23	1.98	0.63
2:B:314:PHE:CE2	2:B:326:LYS:HG2	2.33	0.63
1:A:45:ARG:HE	1:A:45:ARG:HA	1.65	0.61
2:B:318:MET:O	2:B:319:SER:HB3	2.02	0.60
2:B:327:LYS:HG3	2:B:340:GLN:NE2	2.18	0.58
2:B:320:ALA:HB2	2:B:329:VAL:HG21	1.86	0.56
1:A:2:GLN:HE22	1:A:76:ASN:ND2	2.04	0.56
2:B:329:VAL:O	2:B:333:LYS:HE3	2.06	0.55
2:B:327:LYS:O	2:B:331:SER:HB2	2.08	0.54
2:B:328:ASP:O	2:B:332:GLN:HG2	2.07	0.54
1:A:17:HIS:HE1	1:A:84:VAL:O	1.91	0.54
1:A:221:CYS:SG	1:A:222:MET:CE	2.96	0.53
1:A:221:CYS:SG	2:B:377:TYR:OXT	2.67	0.52

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:125:SER:HB3	1:A:221:CYS:CB	2.40	0.51
2:B:326:LYS:CE	2:B:345:ASN:ND2	2.73	0.51
1:A:221:CYS:HG	1:A:222:MET:CE	2.23	0.50
1:A:238:HIS:HD2	4:A:508:HOH:O	1.94	0.50
1:A:221:CYS:CB	2:B:377:TYR:OXT	2.59	0.50
1:A:16:LEU:HD12	1:A:270:VAL:HG13	1.94	0.49
2:B:358:GLU:OE1	2:B:361:LYS:NZ	2.46	0.48
1:A:16:LEU:CD1	1:A:270:VAL:HG22	2.42	0.48
1:A:2:GLN:HE22	1:A:76:ASN:HD22	1.61	0.48
2:B:360:LYS:HE3	4:B:534:HOH:O	2.13	0.48
2:B:326:LYS:CE	2:B:345:ASN:HD22	2.27	0.48
1:A:170:LYS:HG2	1:A:195:GLU:HG2	1.96	0.47
1:A:193:GLY:HA2	1:A:261:PHE:CE1	2.49	0.47
2:B:326:LYS:HE3	2:B:345:ASN:ND2	2.30	0.47
2:B:319:SER:OG	2:B:320:ALA:N	2.49	0.46
2:B:320:ALA:HB3	2:B:364:SER:HB2	1.98	0.45
1:A:205:ILE:O	1:A:216:ALA:HA	2.17	0.45
2:B:307:GLU:HA	2:B:351:LEU:O	2.15	0.45
2:B:314:PHE:CZ	2:B:326:LYS:HG2	2.52	0.44
2:B:352:ASP:OD2	2:B:354:LYS:HG3	2.18	0.44
1:A:126:LEU:HA	2:B:377:TYR:H	1.82	0.44
1:A:100:GLY:O	2:B:376:GLU:HG3	2.18	0.43
2:B:339:LYS:HD2	2:B:340:GLN:N	2.34	0.43
1:A:125:SER:HB3	1:A:221:CYS:HB2	1.99	0.43
1:A:26:VAL:HG11	1:A:232:ALA:HA	2.00	0.43
1:A:126:LEU:C	1:A:126:LEU:HD12	2.39	0.43
1:A:64:HIS:CD2	2:B:376:GLU:HB3	2.53	0.42
1:A:127:GLY:HA3	2:B:377:TYR:CE2	2.55	0.42
2:B:376:GLU:H	2:B:376:GLU:HG3	1.45	0.42
2:B:317:THR:O	2:B:319:SER:N	2.48	0.41
1:A:100:GLY:HA3	2:B:376:GLU:CG	2.50	0.41
2:B:321:MET:HB2	4:B:469:HOH:O	2.20	0.41
2:B:352:ASP:CG	2:B:354:LYS:HG3	2.40	0.41
1:A:13:ALA:N	1:A:14:PRO:CD	2.82	0.41
2:B:318:MET:O	2:B:364:SER:HA	2.21	0.41
1:A:16:LEU:HD11	1:A:270:VAL:HG22	2.03	0.41
2:B:375:HIS:O	2:B:376:GLU:C	2.57	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	273/275 (99%)	264 (97%)	8 (3%)	1 (0%)	34	30
2	B	69/71 (97%)	60 (87%)	7 (10%)	2 (3%)	4	1
All	All	342/346 (99%)	324 (95%)	15 (4%)	3 (1%)	17	11

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	320	ALA
1	A	63	SER
2	B	321	MET

### 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	208/208 (100%)	201 (97%)	7 (3%)	37	36
2	B	59/59 (100%)	52 (88%)	7 (12%)	5	3
All	All	267/267 (100%)	253 (95%)	14 (5%)	23	19

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	43	ASN
1	A	45	ARG
1	A	62	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	75	LEU
1	A	76	ASN
1	A	90	LEU
1	A	222	MET
2	B	322	SER
2	B	323	SER
2	B	333	LYS
2	B	345	ASN
2	B	354	LYS
2	B	376	GLU
2	B	377	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	17	HIS
1	A	59	GLN
1	A	76	ASN
1	A	109	ASN
1	A	155	ASN
1	A	206	GLN
1	A	238	HIS
1	A	271	GLN
2	B	345	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

Of 2 ligands modelled in this entry, 2 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

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	275/275 (100%)	-0.72	1 (0%) 92 92	9, 14, 20, 30	0
2	B	71/71 (100%)	1.14	13 (18%) 1 1	20, 34, 54, 58	0
All	All	346/346 (100%)	-0.34	14 (4%) 38 37	9, 15, 39, 58	0

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	319	SER	11.5
2	B	318	MET	10.9
2	B	320	ALA	7.4
2	B	317	THR	6.6
2	B	321	MET	6.2
2	B	377	TYR	5.8
2	B	324	ALA	5.8
2	B	316	GLN	4.4
2	B	323	SER	3.8
2	B	334	GLY	2.4
2	B	322	SER	2.3
1	A	221	CYS	2.2
2	B	335	GLY	2.2
2	B	364	SER	2.0

### 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](#)

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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	CA	A	382	1/1	0.94	0.14	40,40,40,40	0
3	CA	A	381	1/1	1.00	0.06	17,17,17,17	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.