wwPDB X-ray Structure Validation Summary Report (i)

Sep 18, 2023 – 02:20 PM EDT

PDB ID	:	1SBT
Title	:	ATOMIC COORDINATES FOR SUBTILISIN BPN (OR NOVO)
Authors	:	Alden, R.A.; Birktoft, J.J.; Kraut, J.; Robertus, J.D.; Wright, C.S.
Deposited on	:	1972-08-11
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:

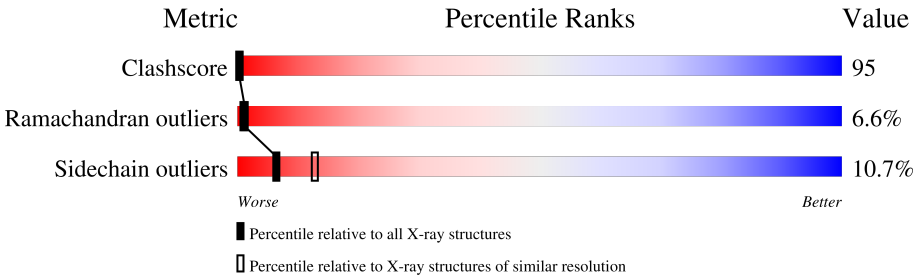
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 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	275	5% 33% 52% 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1955 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 BPN'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	275	Total	C	N	O	S	0	0	0
			1938	1204	335	394	5			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	56	PRO	ASN	conflict	UNP P00782
A	57	ASN	PRO	conflict	UNP P00782
A	61	ASP	ASN	conflict	UNP P00782
A	88	SER	ALA	conflict	UNP P00782
A	89	ALA	SER	conflict	UNP P00782
A	98	ASP	ALA	conflict	UNP P00782
A	99	ALA	ASP	conflict	UNP P00782
A	158	SER	THR	conflict	UNP P00782
A	159	THR	SER	conflict	UNP P00782
A	251	GLN	GLU	conflict	UNP P00782

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	17	Total	O	0	0
			17	17		

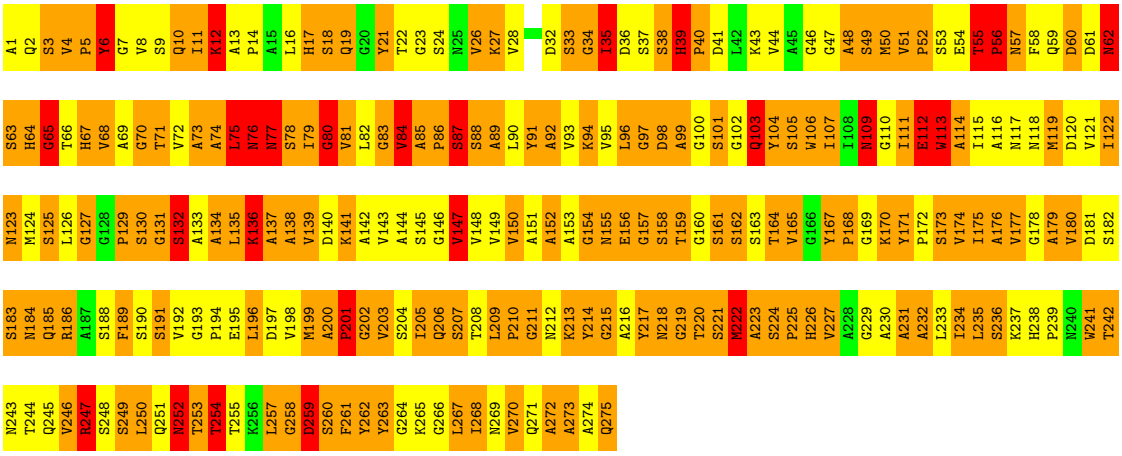
3 Residue-property plots (i)

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: SUBTILISIN BPN'

Chain A:



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	66.70Å 54.40Å 62.90Å 90.00° 91.90° 90.00°	Depositor
Resolution (Å)	(Not available) – 2.50	Depositor
% Data completeness (in resolution range)	(Not available) ((Not available)-2.50)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	unknown	Depositor
R, R_{free}	(Not available) , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1955	wwPDB-VP
Average B, all atoms (Å^2)	0.0	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	13.30	367/1976 (18.6%)	3.24	289/2697 (10.7%)

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	0	19

The worst 5 of 367 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	76	ASN	CG-OD1	561.54	13.59	1.24
1	A	76	ASN	CB-CG	45.28	2.55	1.51
1	A	112	GLU	CG-CD	-19.62	1.22	1.51
1	A	101	SER	CB-OG	-18.06	1.18	1.42
1	A	168	PRO	N-CD	16.35	1.70	1.47

The worst 5 of 289 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	186	ARG	NE-CZ-NH2	19.12	129.86	120.30
1	A	232	ALA	O-C-N	18.35	152.05	122.70
1	A	261	PHE	CB-CG-CD1	-17.97	108.22	120.80
1	A	76	ASN	OD1-CG-ND2	-16.36	84.26	121.90
1	A	213	LYS	CB-CA-C	-14.79	80.81	110.40

There are no chirality outliers.

5 of 19 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	10	GLN	Mainchain
1	A	39	HIS	Sidechain
1	A	44	VAL	Mainchain
1	A	51	VAL	Mainchain
1	A	56	PRO	Mainchain

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	1938	0	1875	364	15
2	A	17	0	0	2	0
All	All	1955	0	1875	364	15

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 95.

The worst 5 of 364 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:60:ASP:CB	1:A:60:ASP:CA	1.75	1.58
1:A:84:VAL:CA	1:A:84:VAL:CB	1.76	1.54
1:A:112:GLU:CB	1:A:112:GLU:CG	1.78	1.54
1:A:201:PRO:C	1:A:201:PRO:CA	1.75	1.53
1:A:111:ILE:CA	1:A:111:ILE:CB	1.84	1.51

The worst 5 of 15 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:A:6:TYR:CG	1:A:161:SER:OG[4_545]	0.90	1.30
1:A:18:SER:CB	1:A:162:SER:OG[3_545]	1.11	1.09
1:A:6:TYR:CB	1:A:161:SER:OG[4_545]	1.26	0.94
1:A:185:GLN:OE1	1:A:213:LYS:CD[4_555]	1.37	0.83
1:A:183:SER:OG	1:A:271:GLN:OE1[2_655]	1.51	0.69

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%)	212 (78%)	43 (16%)	18 (7%)	1 1

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	12	LYS
1	A	35	ILE
1	A	55	THR
1	A	69	ALA
1	A	75	LEU

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	205/205 (100%)	183 (89%)	22 (11%)	6 13

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

Mol	Chain	Res	Type
1	A	243	ASN
1	A	251	GLN
1	A	250	LEU
1	A	252	ASN
1	A	109	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	77	ASN
1	A	238	HIS
1	A	245	GLN
1	A	57	ASN
1	A	19	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	19

The worst 5 of 19 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	79:ILE	C	80:GLY	N	1.62

Continued on next page...

Continued from previous page...

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	265:LYS	C	266:GLY	N	1.20
1	A	129:PRO	C	130:SER	N	1.19
1	A	215:GLY	C	216:ALA	N	1.19
1	A	254:THR	C	255:THR	N	1.19

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.