Full wwPDB X-ray Structure Validation Report (i)

May 28, 2020 – 08:27 pm BST

PDB ID	:	1SMF
Title	:	Studies on an artificial trypsin inhibitor peptide derived from the mung bean inhibitor
Authors	:	Huang, Q.; Li, Y.; Zhang, S.; Liu, S.; Tang, Y.; Qi, C.
Deposited on	:	1992-10-24
Resolution	:	2.10 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

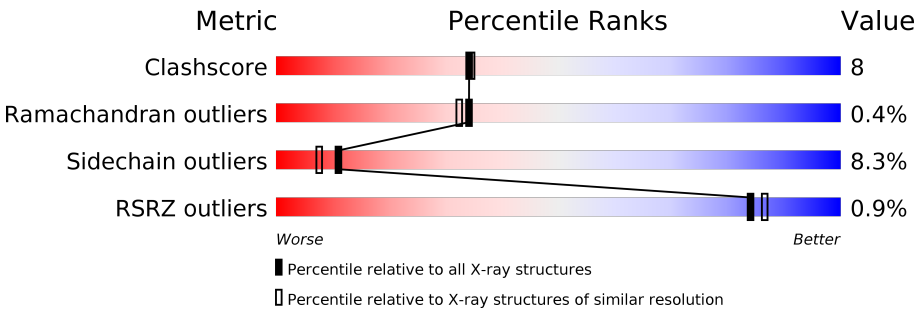
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 >=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% 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	E	223	% 76% 21% • •
2	I	22	32% 9% 59%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2340 atoms, of which 562 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 TRYPSIN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	E	223	2014	1012	385	279	324	14	0	0	0

- Molecule 2 is a protein called BOWMAN-BIRK TYPE TRYPSIN INHIBITOR.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	I	9	76	40	11	10	13	2	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	8	ARG	ASP	CONFLICT	UNP P01062

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
3	E	1	1	1	0	0

- Molecule 4 is water.

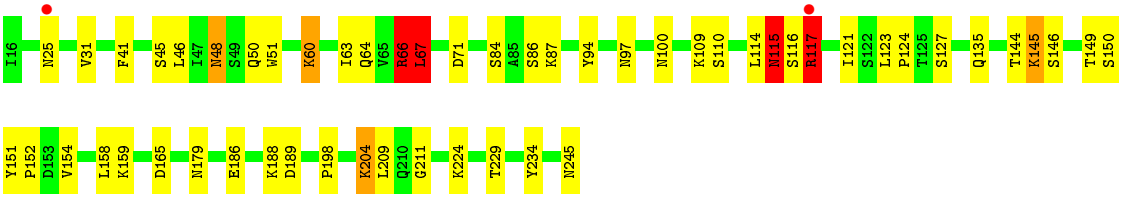
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	H	O		
4	E	81	243	162	81	0	0
			Total	H	O		
4	I	2	6	4	2	0	0

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.

- Molecule 1: TRYPSIN

Chain E:



- Molecule 2: BOWMAN-BIRK TYPE TRYPSIN INHIBITOR

Chain I:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	63.23Å 63.62Å 69.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	(Not available) – 2.10 30.51 – 2.15	Depositor EDS
% Data completeness (in resolution range)	(Not available) ((Not available)-2.10) 84.2 (30.51-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.13 (at 2.16Å)	Xtriage
Refinement program	X-PLOR	Depositor
R, R_{free}	0.190 , (Not available) 0.186 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	7.7	Xtriage
Anisotropy	0.490	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 14.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.049 for k,h,-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	2340	wwPDB-VP
Average B, all atoms (Å^2)	7.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.91% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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	E	0.71	0/1660	1.34	11/2250 (0.5%)
2	I	0.70	0/66	2.51	1/89 (1.1%)
All	All	0.71	0/1726	1.40	12/2339 (0.5%)

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	E	0	2

There are no bond length outliers.

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	I	17	CYS	CA-C-O	18.93	159.85	120.10
1	E	66	ARG	NE-CZ-NH2	-9.57	115.51	120.30
1	E	60	LYS	CB-CG-CD	-7.68	91.62	111.60
1	E	66	ARG	NH1-CZ-NH2	-7.29	111.38	119.40
1	E	67	LEU	CA-CB-CG	7.09	131.61	115.30
1	E	66	ARG	NE-CZ-NH1	6.91	123.75	120.30
1	E	25	ASN	CA-CB-CG	6.29	127.23	113.40
1	E	110	SER	N-CA-CB	-5.58	102.13	110.50
1	E	189	ASP	CB-CG-OD1	5.32	123.08	118.30
1	E	60	LYS	CA-CB-CG	5.28	125.02	113.40
1	E	234	TYR	CB-CG-CD2	-5.21	117.88	121.00
1	E	94	TYR	CB-CG-CD2	-5.04	117.98	121.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	117	ARG	Sidechain
1	E	66	ARG	Sidechain

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	E	1629	385	1588	27	0
2	I	65	11	63	1	0
3	E	1	0	0	0	0
4	E	81	162	0	2	0
4	I	2	4	0	1	0
All	All	1778	562	1651	28	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 8.

All (28) 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:E:64:GLN:HE22	1:E:66:ARG:HH11	1.32	0.78
1:E:48:ASN:HD22	1:E:50:GLN:H	1.38	0.69
2:I:9:CYS:N	4:I:24:HOH:O	2.29	0.66
1:E:115:ASN:ND2	1:E:117:ARG:HG2	2.15	0.61
1:E:100:ASN:HD21	1:E:179:ASN:HD22	1.47	0.61
1:E:45:SER:OG	1:E:198:PRO:HB3	2.06	0.56
1:E:64:GLN:NE2	1:E:66:ARG:HH11	2.01	0.56
1:E:60:LYS:NZ	4:E:296:HOH:O	2.40	0.54
1:E:48:ASN:ND2	1:E:50:GLN:H	2.05	0.50
1:E:71:ASP:O	1:E:154:VAL:HA	2.13	0.47
1:E:117:ARG:HE	1:E:117:ARG:H	1.62	0.47
1:E:158:LEU:HD11	1:E:188:LYS:HB3	1.96	0.47
1:E:84:SER:HB2	1:E:109:LYS:HE3	1.97	0.47
1:E:149:THR:HG22	1:E:151:TYR:CE2	2.51	0.45
1:E:145:LYS:HD3	1:E:150:SER:HB3	1.98	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:46:LEU:HD21	1:E:114:LEU:HD21	1.98	0.44
1:E:48:ASN:HD21	1:E:51:TRP:HD1	1.64	0.43
1:E:41:PHE:HZ	1:E:63:ILE:HG12	1.84	0.43
1:E:124:PRO:HA	1:E:204:LYS:HG2	2.01	0.42
1:E:31:VAL:HG12	1:E:67:LEU:HD13	2.00	0.42
1:E:124:PRO:HD3	1:E:209:LEU:O	2.19	0.42
1:E:211:GLY:HA2	1:E:229:THR:O	2.20	0.42
1:E:87:LYS:HE2	1:E:245:ASN:OD1	2.20	0.41
1:E:144:THR:HG23	1:E:152:PRO:HD3	2.02	0.41
1:E:186:GLU:HG2	4:E:302:HOH:O	2.19	0.41
1:E:121:ILE:HA	1:E:121:ILE:HD12	1.89	0.40
1:E:224:LYS:HD3	1:E:224:LYS:HA	2.00	0.40
1:E:144:THR:C	1:E:145:LYS:HD2	2.41	0.40

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	E	221/223 (99%)	216 (98%)	4 (2%)	1 (0%)	29 26
2	I	7/22 (32%)	6 (86%)	1 (14%)	0	100 100
All	All	228/245 (93%)	222 (97%)	5 (2%)	1 (0%)	34 32

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	115	ASN

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	E	184/184 (100%)	168 (91%)	16 (9%)	10 7
2	I	9/21 (43%)	9 (100%)	0	100 100
All	All	193/205 (94%)	177 (92%)	16 (8%)	11 7

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	48	ASN
1	E	66	ARG
1	E	67	LEU
1	E	86	SER
1	E	97	ASN
1	E	115	ASN
1	E	116	SER
1	E	117	ARG
1	E	123	LEU
1	E	127	SER
1	E	135	GLN
1	E	145	LYS
1	E	146	SER
1	E	159	LYS
1	E	165	ASP
1	E	204	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	25	ASN
1	E	30	GLN
1	E	48	ASN
1	E	64	GLN
1	E	100	ASN
1	E	101	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	115	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)

Of 1 ligands modelled in this entry, 1 is 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 (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, 95th 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^2)	Q<0.9
1	E	223/223 (100%)	-0.62	2 (0%) 84 86	2, 6, 23, 54	0
2	I	9/22 (40%)	-0.48	0 100 100	6, 10, 18, 26	0
All	All	232/245 (94%)	-0.62	2 (0%) 84 86	2, 6, 26, 54	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	25	ASN	2.2
1	E	117	ARG	2.1

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, 95th 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(Å^2)	Q<0.9
3	CA	E	0	1/1	0.99	0.09	9,9,9,9	0

6.5 Other polymers (i)

There are no such residues in this entry.