



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 10:32 AM GMT

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 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

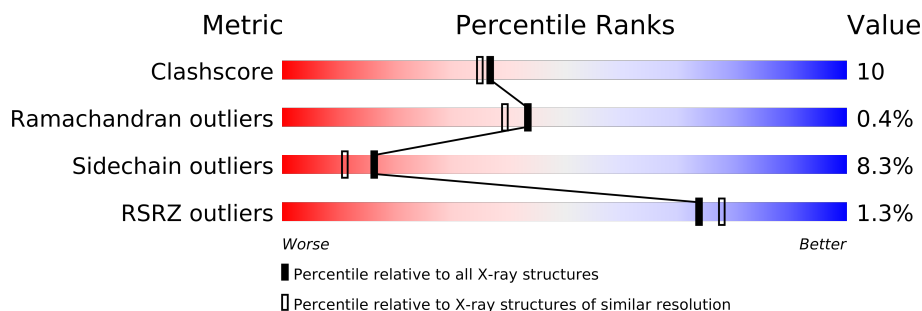
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

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	79885	3649 (2.10-2.10)
Ramachandran outliers	78287	3610 (2.10-2.10)
Sidechain outliers	78261	3611 (2.10-2.10)
RSRZ outliers	66119	3013 (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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	E	223	
2	I	22	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2340 atoms, of which 562 are hydrogens and 0 are deuterium.

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
1	E	223	Total	C	H	N	O	S	0	0	0
			2014	1012	385	279	324	14			

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

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

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
3	E	1	Total	Ca	0	0
			1	1		

- Molecule 4 is water.

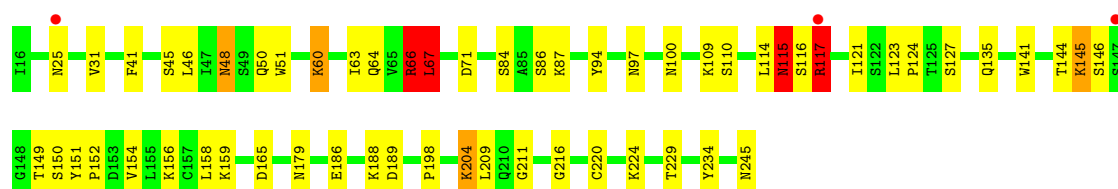
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	E	81	Total	H	O	0	0
			243	162	81		
4	I	2	Total	H	O	0	0
			6	4	2		

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

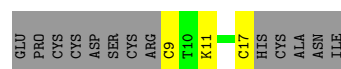
#### • Molecule 1: TRYPSIN

Chain E: 



#### • Molecule 2: BOWMAN-BIRK TYPE TRYPSIN INHIBITOR

Chain I: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	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
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.13 (at 2.16Å)	Xtriage
Refinement program	X-PLOR	Depositor
R, $R_{free}$	0.190 , (Not available) 0.183 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	7.7	Xtriage
Anisotropy	0.490	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 23.4	EDS
Estimated twinning fraction	0.049 for k,h,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Outliers	0 of 13304 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	2340	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	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.*

<sup>1</sup>Intensities estimated from amplitudes.

## 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	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( $^{\circ}$ )	Ideal( $^{\circ}$ )
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 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1629	385	1206	32	0
2	I	65	11	52	2	0
3	E	1	0	0	0	0
4	E	81	162	0	7	0
4	I	2	4	0	2	0
All	All	1778	562	1258	34	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 10.

All (34) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:I:9:CYS:N	4:I:24:HOH:O	2.29	0.66
1:E:48:ASN:HD22	1:E:50:GLN:H	1.45	0.65
1:E:115:ASN:ND2	1:E:117:ARG:HG2	2.15	0.61
1:E:45:SER:OG	1:E:198:PRO:HB3	2.06	0.56
1:E:100:ASN:HD21	1:E:179:ASN:HD22	1.52	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	2.06	0.54
1:E:64:GLN:HE21	1:E:66:ARG:HH11	1.58	0.52
1:E:71:ASP:O	1:E:154:VAL:HA	2.13	0.47
1:E:48:ASN:HD21	1:E:51:TRP:HD1	1.62	0.47
1:E:158:LEU:HD11	1:E:188:LYS:HB3	1.96	0.47
1:E:64:GLN:NE2	1:E:66:ARG:HH11	2.13	0.47
1:E:84:SER:HB2	1:E:109:LYS:HE3	1.97	0.47
4:E:248:HOH:H2	2:I:11:LYS:HZ3	1.64	0.45
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	Distance(Å)	Clash(Å)
1:E:117:ARG:HE	1:E:117:ARG:H	1.64	0.44
1:E:46:LEU:HD21	1:E:114:LEU:HD21	1.98	0.44
1:E:220:CYS:H	4:E:299:HOH:H2	1.67	0.43
1:E:156:LYS:HB2	4:E:269:HOH:H2	1.81	0.43
1:E:41:PHE:HZ	1:E:63:ILE:HG12	1.84	0.43
1:E:141:TRP:H	4:E:280:HOH:H2	1.65	0.43
1:E:84:SER:H	4:E:290:HOH:H2	1.65	0.42
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:216:GLY:HA3	4:I:23:HOH:H1	1.84	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

### 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	E	221/223 (99%)	216 (98%)	4 (2%)	1 (0%)	38	33
2	I	7/22 (32%)	6 (86%)	1 (14%)	0	100	100
All	All	228/245 (93%)	222 (97%)	5 (2%)	1 (0%)	43	39

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	115	ASN



### 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	E	184/184 (100%)	168 (91%)	16 (9%)	15	10
2	I	9/21 (43%)	9 (100%)	0	100	100
All	All	193/205 (94%)	177 (92%)	16 (8%)	16	11

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 (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	30	GLN
1	E	48	ASN
1	E	64	GLN
1	E	100	ASN
1	E	101	ASN
1	E	115	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 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.

## 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	E	223/223 (100%)	-0.54	3 (1%) 74 78	2, 6, 23, 54	0
2	I	9/22 (40%)	-0.38	0 100 100	6, 10, 18, 26	0
All	All	232/245 (94%)	-0.53	3 (1%) 74 78	2, 6, 26, 54	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	147	SER	2.2
1	E	25	ASN	2.1
1	E	117	ARG	2.0

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

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

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CA	E	0	1/1	0.08	-	9,9,9,9	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.