



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 06:44 PM GMT

PDB ID : 2F91  
Title : 1.2A resolution structure of a crayfish trypsin complexed with a peptide inhibitor, SGTI  
Authors : Fodor, K.; Harmat, V.; Hetenyi, C.; Kardos, J.; Antal, J.; Perczel, A.; Patthy, A.; Katona, G.; Graf, L.  
Deposited on : 2005-12-05  
Resolution : 1.20 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>

---

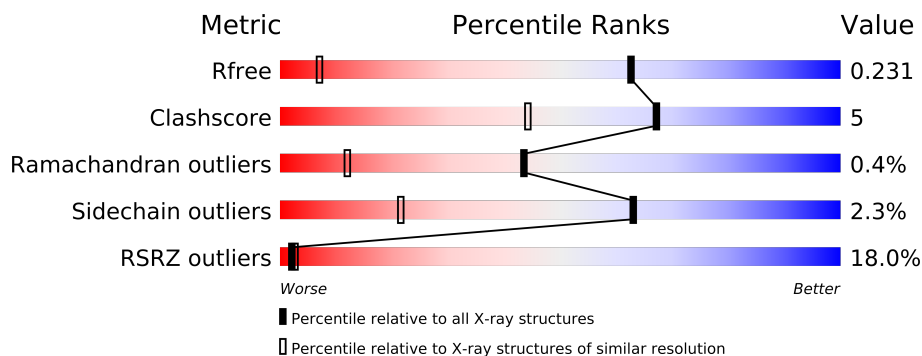
The following versions of software and data (see [references](#)) were used in the production of this report:

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 1.20 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))
$R_{free}$	66092	1038 (1.26-1.14)
Clashscore	79885	1158 (1.26-1.14)
Ramachandran outliers	78287	1106 (1.26-1.14)
Sidechain outliers	78261	1104 (1.26-1.14)
RSRZ outliers	66119	1038 (1.26-1.14)

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	A	237	
2	B	35	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4153 atoms, of which 1795 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 hepatopancreas trypsin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	237	Total	C	H	N	O	S	0	2	0
			3345	1108	1581	279	369	8			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	59	VAL	ALA	VARIANT	UNP Q52V24

- Molecule 2 is a protein called Serine protease inhibitor I/II.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	33	Total	C	H	N	O	S	12	1	0
			457	144	214	45	47	7			

- Molecule 3 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	6	Total	Cd	0	0
			6	6		

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Cl	0	0
			2	2		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	323	Total	O	0	0
			323	323		

*Continued on next page...*

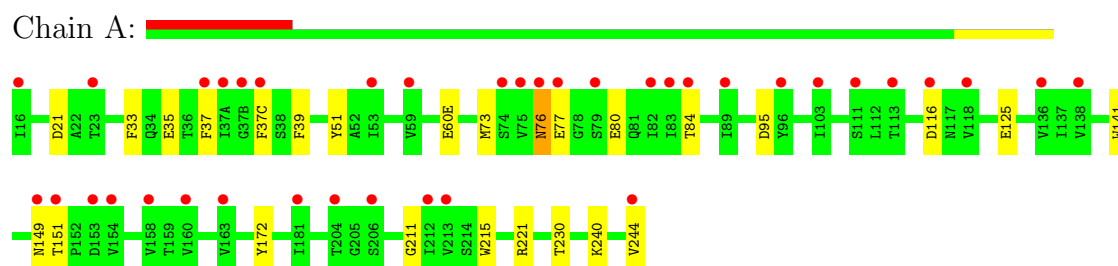
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	20	Total	O	0	0
			20	20		

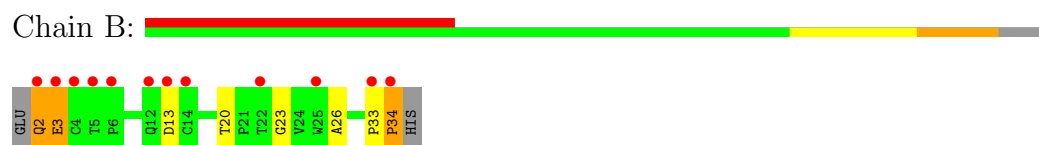
### 3 Residue-property plots

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 ( $\text{RSRZ} > 2$ ). 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: hepatopancreas trypsin



- Molecule 2: Serine protease inhibitor I/II



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	41.28Å 59.67Å 97.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.10 – 1.20 32.06 – 1.20	Depositor EDS
% Data completeness (in resolution range)	91.1 (32.10-1.20) 91.1 (32.06-1.20)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.31 (at 1.20Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, $R_{free}$	0.139 , 0.182 0.209 , 0.231	Depositor DCC
$R_{free}$ test set	3470 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	9.4	Xtriage
Anisotropy	0.079	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.50 , 57.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 69145 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4153	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.93% 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: CL, CD

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	0.67	0/1814	1.19	10/2477 (0.4%)
2	B	0.76	0/252	1.39	1/343 (0.3%)
All	All	0.68	0/2066	1.21	11/2820 (0.4%)

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	1

There are no bond length outliers.

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	77	GLU	OE1-CD-OE2	-10.76	110.39	123.30
1	A	77	GLU	CG-CD-OE1	8.08	134.46	118.30
2	B	2	GLN	C-N-CA	7.41	140.22	121.70
1	A	221	ARG	NE-CZ-NH2	-6.89	116.85	120.30
1	A	221	ARG	CD-NE-CZ	6.88	133.22	123.60
1	A	95	ASP	CB-CG-OD1	6.29	123.96	118.30
1	A	60(E)	GLU	OE1-CD-OE2	-6.14	115.93	123.30
1	A	21	ASP	CB-CG-OD1	5.80	123.52	118.30
1	A	116	ASP	CB-CG-OD2	-5.80	113.08	118.30
1	A	33	PHE	CB-CG-CD2	-5.18	117.17	120.80
1	A	77	GLU	CB-CG-CD	5.09	127.94	114.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	125	GLU	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	A	1764	1581	0	15	0
2	B	243	214	0	7	0
3	A	6	0	0	0	0
4	A	2	0	0	0	0
5	A	323	0	0	5	0
5	B	20	0	0	1	0
All	All	2358	1795	0	18	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 5.

All (18) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:244:VAL:HG23	5:A:575:HOH:O	1.86	0.74
1:A:51:TYR:OH	1:A:244:VAL:HG21	1.91	0.70
2:B:13:ASP:OD2	2:B:34:PRO:HG2	1.91	0.70
1:A:84:THR:HG23	5:A:700:HOH:O	1.93	0.68
1:A:35:GLU:HG3	1:A:37:PHE:CE2	2.35	0.61
1:A:244:VAL:HA	5:A:730:HOH:O	2.03	0.59
1:A:37(C):PHE:CE1	1:A:39:PHE:HB2	2.41	0.55
1:A:172:TYR:O	2:B:20:THR:HB	2.14	0.48
2:B:3:GLU:HB3	2:B:23:GLY:O	2.14	0.47
1:A:149:ASN:HB3	5:A:581:HOH:O	2.15	0.46
1:A:39:PHE:CD2	2:B:33:PRO:HG3	2.51	0.46
1:A:35:GLU:HB2	5:A:523:HOH:O	2.17	0.43
1:A:73:MET:HE2	1:A:141:TRP:CG	2.53	0.43
2:B:34:PRO:HA	5:B:343:HOH:O	2.18	0.43
1:A:76:ASN:HA	1:A:80:GLU:OE2	2.18	0.42
1:A:215:TRP:CE3	2:B:26:ALA:HB2	2.56	0.41
1:A:211:GLY:HA2	1:A:230:THR:O	2.21	0.41

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:39:PHE:CE2	2:B:33:PRO:HG3	2.56	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	237/237 (100%)	233 (98%)	4 (2%)	0	100	100
2	B	32/35 (91%)	30 (94%)	1 (3%)	1 (3%)	7	0
All	All	269/272 (99%)	263 (98%)	5 (2%)	1 (0%)	43	12

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	3	GLU

### 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	192/190 (101%)	189 (98%)	3 (2%)	75	40
2	B	30/31 (97%)	28 (93%)	2 (7%)	23	1
All	All	222/221 (100%)	217 (98%)	5 (2%)	63	21

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	76	ASN
1	A	151	THR
1	A	240	LYS
2	B	2	GLN
2	B	34	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	65	GLN
1	A	76	ASN
1	A	93	ASN
1	A	97	ASN
2	B	18	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 8 ligands modelled in this entry, 8 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.

## 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	A	237/237 (100%)	1.16	38 (16%) 3 3	7, 13, 24, 36	0
2	B	33/35 (94%)	2.59	12 (36%) 1 1	10, 17, 40, 47	2 (6%)
All	All	270/272 (99%)	1.34	50 (18%) 2 2	7, 13, 26, 47	2 (0%)

All (50) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	34	PRO	13.8
2	B	2	GLN	12.5
1	A	244	VAL	9.7
2	B	33	PRO	8.7
2	B	3	GLU	7.6
1	A	75	VAL	7.1
1	A	37(C)	PHE	7.1
1	A	76	ASN	7.1
1	A	37(B)	GLY	5.9
1	A	37(A)	ILE	5.2
1	A	204	THR	4.8
2	B	4	CYS	4.7
1	A	154	VAL	4.4
1	A	96	TYR	4.3
1	A	151	THR	4.3
1	A	77	GLU	3.7
2	B	5	THR	3.7
1	A	149	ASN	3.5
1	A	74	SER	3.5
1	A	158	VAL	3.3
1	A	153	ASP	3.2
1	A	113	THR	3.1
1	A	138	VAL	3.0
2	B	25	TRP	2.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	37	PHE	2.9
1	A	23	THR	2.8
1	A	118	VAL	2.7
2	B	14	CYS	2.7
1	A	89	ILE	2.7
1	A	16	ILE	2.7
2	B	13	ASP	2.7
2	B	22	THR	2.7
1	A	59	VAL	2.6
1	A	206	SER	2.6
2	B	12	GLN	2.6
1	A	82	ILE	2.5
1	A	136	VAL	2.5
1	A	213	VAL	2.4
1	A	84	THR	2.4
2	B	6	PRO	2.4
1	A	212	ILE	2.4
1	A	160	VAL	2.1
1	A	181	ILE	2.1
1	A	111	SER	2.1
1	A	163	VAL	2.1
1	A	53	ILE	2.1
1	A	103	ILE	2.1
1	A	79[A]	SER	2.0
1	A	116	ASP	2.0
1	A	83	ILE	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	CD	A	402	1/1	0.09	-	7,7,7,7	1
3	CD	A	403	1/1	0.04	-	13,13,13,13	1
3	CD	A	404	1/1	0.12	-	27,27,27,27	1
4	CL	A	408	1/1	0.16	-	11,11,11,11	1
3	CD	A	405	1/1	0.10	-	6,6,6,6	1
3	CD	A	401	1/1	0.10	-	18,18,18,18	0
4	CL	A	407	1/1	0.11	-	11,11,11,11	1
3	CD	A	406	1/1	0.04	-	40,40,40,40	1

## 6.5 Other polymers ⓘ

There are no such residues in this entry.