W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Feb 14, 2017 – 05:20 pm GMT

PDB ID	:	1TX6
Title	:	trypsin:BBI complex
Authors	:	Song, H.K.; Park, E.Y.; Kim, J.A.; Kim, H.W.; Kim, Y.S.
Deposited on	:	2004-07-02
Resolution	:	2.20 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

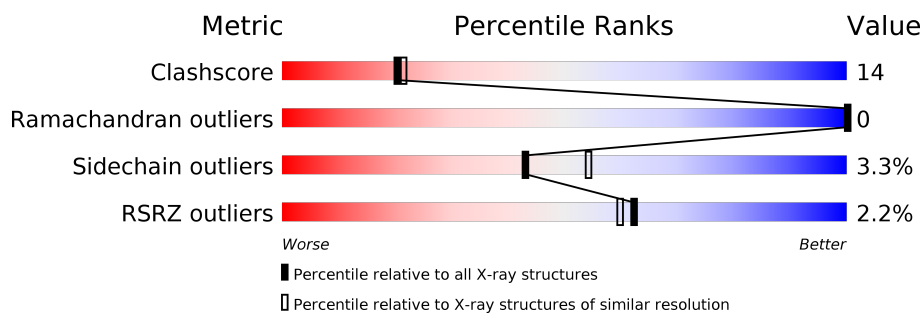|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

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 2.20 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	Percentile Ranks	Value
Clashscore		14
Ramachandran outliers		0
Sidechain outliers		3.3%
RSRZ outliers		2.2%

Worse — *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)
RSRZ outliers	101464	4033 (2.20-2.20)

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. 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	A	223	74% / 24% •
1	B	223	% / 75% / 23% •
1	C	223	77% / 23%
1	D	223	% / 82% / 17% •
2	I	125	10% / 63% / 23% • 12%
2	J	125	6% / 64% / 26% • 9%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	CA	A	1300	-	-	-	X

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 8769 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 Trypsin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	223	Total	C	N	O	S	0	0	0
			1642	1020	287	321	14			
1	B	223	Total	C	N	O	S	0	0	0
			1642	1020	287	321	14			
1	C	223	Total	C	N	O	S	0	0	0
			1642	1020	287	321	14			
1	D	223	Total	C	N	O	S	14	0	0
			1642	1020	287	321	14			

- Molecule 2 is a protein called Bowman-Birk type trypsin inhibitor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	110	Total	C	N	O	S	0	0	0
			834	509	148	155	22			
2	J	114	Total	C	N	O	S	0	0	0
			871	530	158	161	22			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	42	SER	-	SEE REMARK 999	UNP P12940
J	42	SER	-	SEE REMARK 999	UNP P12940

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Ca	0	0
			1	1		
3	A	1	Total	Ca	0	0
			1	1		
3	D	1	Total	Ca	0	0
			1	1		

Continued on next page...

Continued from previous page...

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

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	102	Total 102	O 102	0	0
4	B	108	Total 108	O 108	0	0
4	C	83	Total 83	O 83	0	0
4	D	124	Total 124	O 124	0	0
4	I	31	Total 31	O 31	0	0
4	J	44	Total 44	O 44	0	0

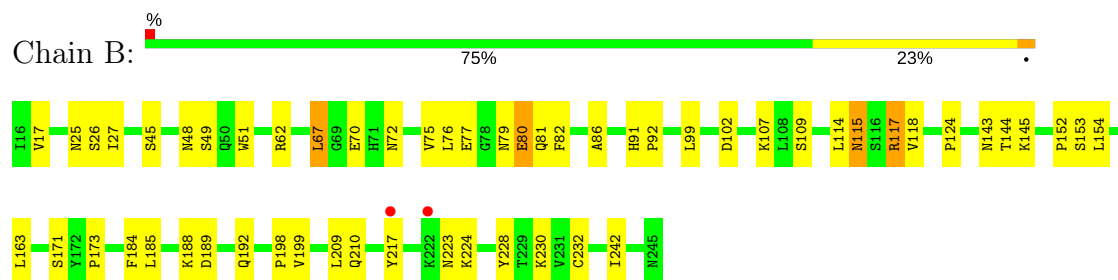# 3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.
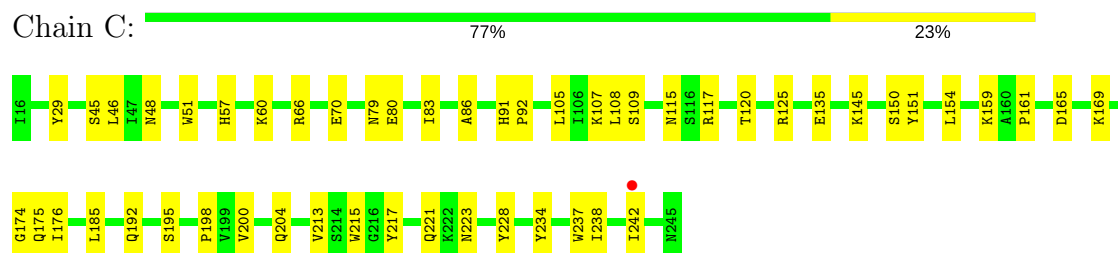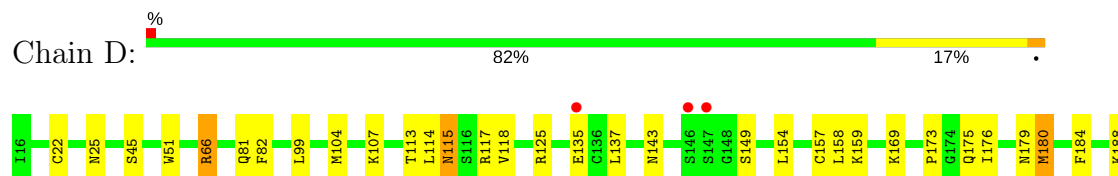
- Molecule 1: Trypsin

Chain A:

- Molecule 1: Trypsin

Chain B:

- Molecule 1: Trypsin

Chain C:

- Molecule 1: Trypsin

Chain D:

P198 Q204 I212 V213 S214 W215 G216 Y217 K224 Y228 T229 N245

• Molecule 2: Bowman-Birk type trypsin inhibitor

Chain I:

10%

63% 23% • 12%

ALA GLY LYS LYS R5 P6 W7 K8 D11 T16 R17 S18 I19 T24 E31 C32 PRO LYS THR C36 K37 S38 D45 P46 S47 R48 R49 I50 C51 Q54 Y55 I62 C63 W66 R83 V88 K89 K90 C91 A92 P93 T94 C95 K96 T97 C98 L99 PRO SER ARG

SER ARG PRO SER ARG R108 I111 D112 P119 P120 R121 P124 R125

• Molecule 2: Bowman-Birk type trypsin inhibitor

Chain J:

6%

64% 26% • 9%

ALA GLY LYS LYS R5 K8 V14 C15 T16 R17 V29 F30 E31 C32 PRO LYS T35 C36 K37 G40 P41 S42 M43 G44 D45 P46 S47 R48 R49 I50 C51 Q52 I62 C63 R64 P65 W66 C69 T75 P79 P80 V88 K89 K96 T97 S101 ARG SER ARG

PRO SER R107 R108 F115 G116 R121 C122 T123 P124 R125

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	67.21Å 88.54Å 203.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.20 44.17 – 2.20	Depositor EDS
% Data completeness (in resolution range)	(Not available) (30.00-2.20) 91.5 (44.17-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.09 (at 2.20Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.223 , 0.274 0.242 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	30.0	Xtriage
Anisotropy	0.646	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 63.2	EDS
L-test for twinning[2]	$< |L| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	8769	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 25.33 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.2354e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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	A	0.35	0/1674	0.64	0/2273
1	B	0.33	0/1674	0.63	0/2273
1	C	0.33	0/1674	0.63	0/2273
1	D	0.37	0/1674	0.66	0/2273
2	I	0.29	0/854	0.68	0/1158
2	J	0.35	0/892	0.72	0/1209
All	All	0.34	0/8442	0.65	0/11459

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	1642	0	1595	50	0
1	B	1642	0	1595	53	0
1	C	1642	0	1595	37	0
1	D	1642	0	1595	34	0
2	I	834	0	786	29	0
2	J	871	0	829	39	0
3	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	102	0	0	3	0
4	B	108	0	0	4	0
4	C	83	0	0	2	0
4	D	124	0	0	1	0
4	I	31	0	0	5	0
4	J	44	0	0	1	0
All	All	8769	0	7995	228	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 14.

The worst 5 of 228 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:222:LYS:H	1:A:222:LYS:HE2	1.23	1.02
1:C:125:ARG:H	1:C:204:GLN:HE21	1.10	1.00
1:A:125:ARG:H	1:A:204:GLN:HE21	1.07	0.93
1:D:81:GLN:HE22	1:D:113:THR:H	1.08	0.93
1:D:125:ARG:H	1:D:204:GLN:NE2	1.74	0.85

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	A	221/223 (99%)	214 (97%)	7 (3%)	0	100	100
1	B	221/223 (99%)	209 (95%)	12 (5%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	221/223 (99%)	212 (96%)	9 (4%)	0	100	100
1	D	221/223 (99%)	212 (96%)	9 (4%)	0	100	100
2	I	104/125 (83%)	98 (94%)	6 (6%)	0	100	100
2	J	108/125 (86%)	99 (92%)	9 (8%)	0	100	100
All	All	1096/1142 (96%)	1044 (95%)	52 (5%)	0	100	100

There are no Ramachandran outliers to report.

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	183/183 (100%)	179 (98%)	4 (2%)	57	70
1	B	183/183 (100%)	178 (97%)	5 (3%)	50	62
1	C	183/183 (100%)	176 (96%)	7 (4%)	38	47
1	D	183/183 (100%)	175 (96%)	8 (4%)	33	40
2	I	101/115 (88%)	96 (95%)	5 (5%)	28	34
2	J	106/115 (92%)	104 (98%)	2 (2%)	62	76
All	All	939/962 (98%)	908 (97%)	31 (3%)	43	54

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	175	GLN
1	D	66	ARG
2	I	121	ARG
1	C	221	GLN
1	D	104	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 40 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	25	ASN
1	C	101	ASN
1	D	210	GLN
1	C	84	ASN
1	C	175	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 carbohydrates in this entry.

5.6   Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 4 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 ⓘ

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, 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	A	223/223 (100%)	-0.25	0	100	100	20, 31, 45, 58	0
1	B	223/223 (100%)	-0.04	2 (0%)	84	83	21, 36, 53, 64	0
1	C	223/223 (100%)	-0.12	1 (0%)	92	91	22, 36, 55, 66	0
1	D	223/223 (100%)	-0.23	3 (1%)	77	75	18, 28, 47, 66	14 (6%)
2	I	110/125 (88%)	0.59	12 (10%)	6	5	26, 51, 77, 79	0
2	J	114/125 (91%)	0.30	7 (6%)	22	21	24, 43, 80, 95	0
All	All	1116/1142 (97%)	-0.04	25 (2%)	62	60	18, 35, 64, 95	14 (1%)

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	I	89	LYS	5.1
2	I	32	CYS	5.0
2	J	44	GLY	4.4
2	J	43	MET	4.3
1	D	147	SER	4.0

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, 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	LLDF	B-factors(Å^2)	Q<0.9
3	CA	A	1300	1/1	0.71	0.23	3.87	76,76,76,76	0
3	CA	D	4300	1/1	0.91	0.12	-0.14	49,49,49,49	0
3	CA	C	3300	1/1	0.82	0.13	-0.51	72,72,72,72	0
3	CA	B	2300	1/1	0.88	0.11	-1.55	60,60,60,60	0

6.5 Other polymers ⓘ

There are no such residues in this entry.