wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:08 AM GMT

PDB ID	:	2PKS
Title	:	Thrombin in complex with inhibitor
Authors	:	Xue, Y.
Deposited on	:	2007-04-18
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
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:

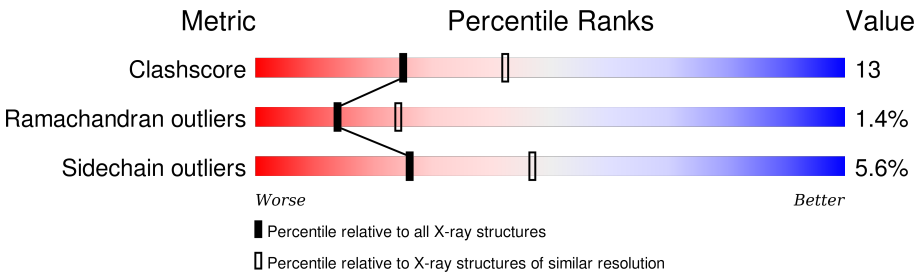
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (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 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	27	78% 19% •
2	B	147	67% 29% 5%
3	C	102	73% 25% •
4	D	10	90% 10%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 2401 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 Thrombin light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	27	222	140	36	45	1	0	0	0

- Molecule 2 is a protein called Thrombin heavy chain fragment.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	147	1200	767	212	214	7	0	0	0

- Molecule 3 is a protein called Thrombin heavy chain fragment.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	102	815	516	145	147	7	0	0	0

- Molecule 4 is a protein called Hirudin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	10	90	56	10	23	1	0	0	0

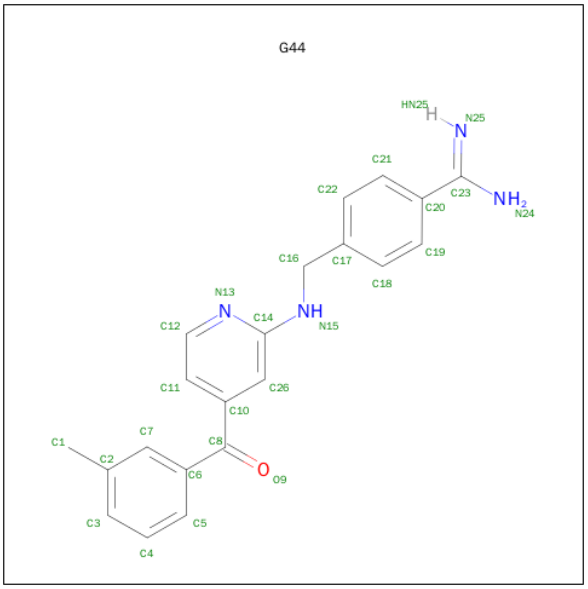
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	361	GLY	GLU	CONFLICT	UNP P28504

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
5	C	1	1	1	0	0

- Molecule 6 is 4-({[4-(3-METHYLBENZOYL)PYRIDIN-2-YL]AMINO}METHYL)BENZENECARBOXIMIDAMIDE (three-letter code: G44) (formula: $C_{21}H_{20}N_4O$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
6	C	1	26	21	4	1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	4	4	4	0	0
			Total	O		
7	B	23	23	23	0	0
			Total	O		
7	C	20	20	20	0	0

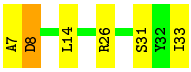
3 Residue-property plots (i)

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

Note EDS was not executed.

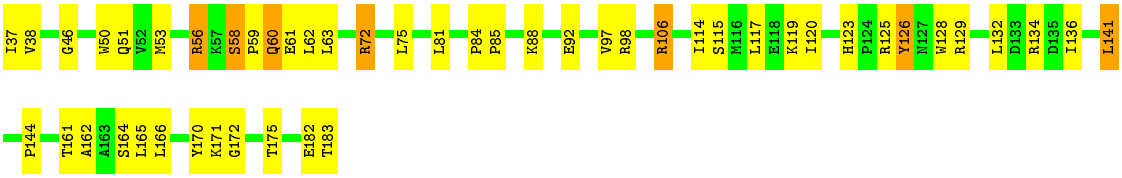
- Molecule 1: Thrombin light chain

Chain A:



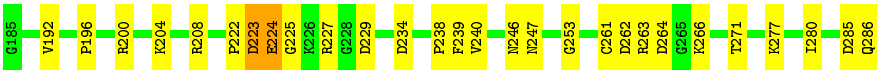
- Molecule 2: Thrombin heavy chain fragment

Chain B:



- Molecule 3: Thrombin heavy chain fragment

Chain C:



- Molecule 4: Hirudin

Chain D:

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	69.94Å 71.84Å 70.78Å 90.00° 100.23° 90.00°	Depositor
Resolution (Å)	69.67 – 2.50	Depositor
% Data completeness (in resolution range)	87.7 (69.67-2.50)	Depositor
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.185 , 0.259	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2401	wwPDB-VP
Average B, all atoms ($Å^2$)	41.0	wwPDB-VP

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: NA, G44, TYS

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	0.87	0/224	0.92	0/298
2	B	0.86	0/1230	0.90	2/1664 (0.1%)
3	C	0.96	0/836	0.90	1/1126 (0.1%)
4	D	0.95	0/74	0.82	0/96
All	All	0.90	0/2364	0.90	3/3184 (0.1%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	106	ARG	NE-CZ-NH2	-5.60	117.50	120.30
3	C	229	ASP	CB-CG-OD2	5.29	123.06	118.30
2	B	75	LEU	CA-CB-CG	5.29	127.46	115.30

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	222	0	222	5	0
2	B	1200	0	1203	33	0
3	C	815	0	789	28	0
4	D	90	0	70	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	1	0	0	0	0
6	C	26	0	20	3	0
7	A	4	0	0	3	0
7	B	23	0	0	1	0
7	C	20	0	0	7	0
All	All	2401	0	2304	59	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 13.

The worst 5 of 59 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)
6:C:101:G44:H18	7:C:405:HOH:O	1.61	0.99
2:B:182:GLU:O	2:B:183:THR:HG23	1.76	0.85
2:B:53:MET:HG3	2:B:62:LEU:HD12	1.66	0.78
2:B:59:PRO:O	2:B:61:GLU:HG2	1.88	0.74
3:C:247:ASN:HB3	7:C:419:HOH:O	1.86	0.73

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	25/27 (93%)	23 (92%)	2 (8%)	0	100 100
2	B	145/147 (99%)	129 (89%)	14 (10%)	2 (1%)	14 24
3	C	100/102 (98%)	89 (89%)	10 (10%)	1 (1%)	19 34
4	D	7/10 (70%)	4 (57%)	2 (29%)	1 (14%)	0 0
All	All	277/286 (97%)	245 (88%)	28 (10%)	4 (1%)	14 24

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	58	SER
3	C	224	GLU
4	D	360	PRO
2	B	60	GLN

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	A	25/25 (100%)	23 (92%)	2 (8%)	15	28
2	B	130/130 (100%)	121 (93%)	9 (7%)	19	35
3	C	88/88 (100%)	85 (97%)	3 (3%)	44	72
4	D	8/8 (100%)	8 (100%)	0	100	100
All	All	251/251 (100%)	237 (94%)	14 (6%)	26	47

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	88	LYS
2	B	106	ARG
3	C	223	ASP
2	B	72	ARG
2	B	141	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	51	GLN
3	C	246	ASN
3	C	251	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	TYS	D	363	-	15,16,17	1.45	2 (13%)	16,22,24	1.21	1 (6%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	TYS	D	363	-	-	0/9/11/13	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	363	TYS	OH-S	-4.32	1.55	1.63
4	D	363	TYS	OH-CZ	-2.16	1.39	1.42

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	363	TYS	OH-CZ-CE2	2.57	123.81	118.74

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
6	G44	C	101	-	28,28,28	1.74	6 (21%)	35,38,38	3.12	15 (42%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	G44	C	101	-	-	0/17/17/17	0/3/3/3

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	C	101	G44	C7-C6	2.25	1.42	1.39
6	C	101	G44	C26-C14	2.33	1.45	1.39
6	C	101	G44	C16-N15	2.72	1.52	1.45
6	C	101	G44	C23-N25	2.91	1.39	1.28
6	C	101	G44	C26-C10	3.07	1.44	1.39

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	C	101	G44	C26-C14-N13	-6.47	115.82	123.06
6	C	101	G44	C19-C20-C23	-5.72	113.43	120.67
6	C	101	G44	C11-C12-N13	-3.93	119.41	123.90
6	C	101	G44	C16-C17-C18	-3.34	113.46	120.90

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	C	101	G44	C10-C8-C6	-2.83	115.14	120.26

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	C	101	G44	3	0

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)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.