wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:11 PM GMT

PDB ID	:	1SGI
Title	:	Crystal structure of the anticoagulant slow form of thrombin
Authors	:	Pineda, A.O.; Carrell, C.J.; Bush, L.A.; Prasad, S.; Caccia, S.; Chen, Z.W.; Mathews, F.S.; Di Cera, E.
Deposited on	:	2004-02-23
Resolution	:	2.30 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 (1)) were used in the production of this report:

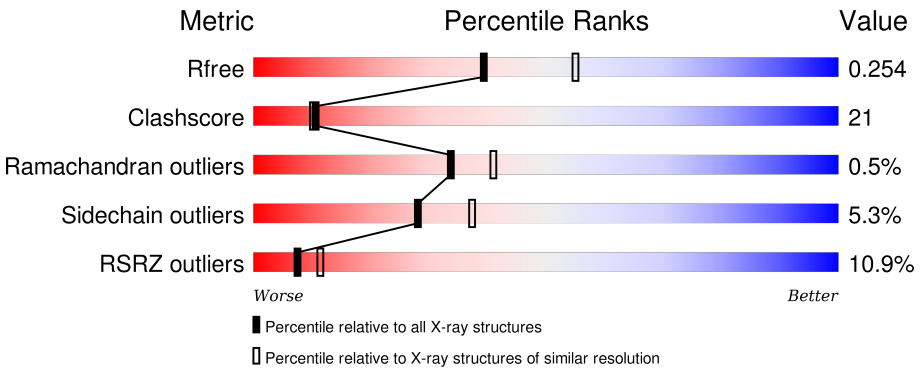
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	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 2.30 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}	91344	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (2.30-2.30)

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	36	8%; 47%, 36%, 6%, 11%
1	D	36	11%; 47%, 28%, •, 22%
2	B	259	12%; 62%, 33%, •, •
2	E	259	8%; 63%, 31%, •, •

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	NAG	B	500	-	-	X	-
3	NAG	E	400	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4875 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	32	252	156	41	54	1	0	0	0
1	D	28	230	144	37	48	1	0	0	0

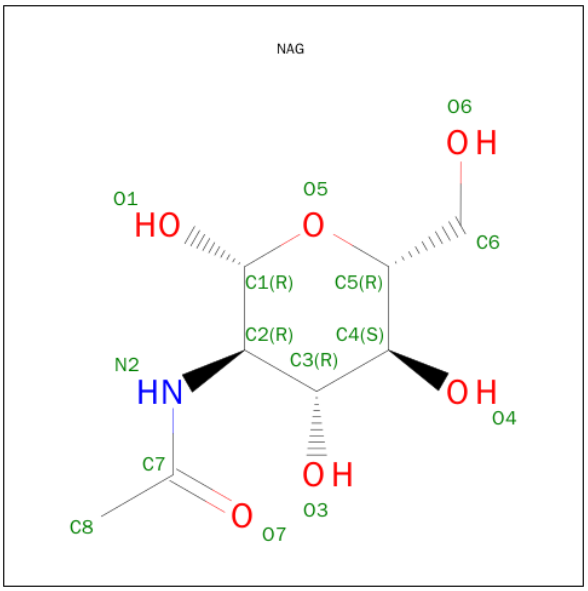
- Molecule 2 is a protein called thrombin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	251	2023	1291	356	362	14	0	0	0
2	E	250	2019	1289	355	361	14	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	77A	ALA	ARG	ENGINEERED	UNP P00734
E	77A	ALA	ARG	ENGINEERED	UNP P00734

- Molecule 3 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	E	1	14	8	1	5	0	0
			Total	C	N	O		
3	B	1	14	8	1	5	0	0

- Molecule 4 is water.

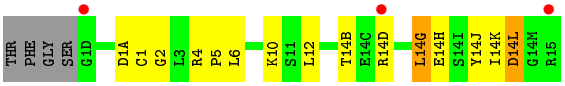
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	18	18	18	0	0
			Total	O		
4	B	130	130	130	0	0
			Total	O		
4	D	20	20	20	0	0
			Total	O		
4	E	155	155	155	0	0

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.

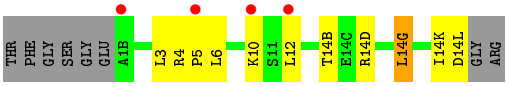
- Molecule 1: thrombin

Chain A:



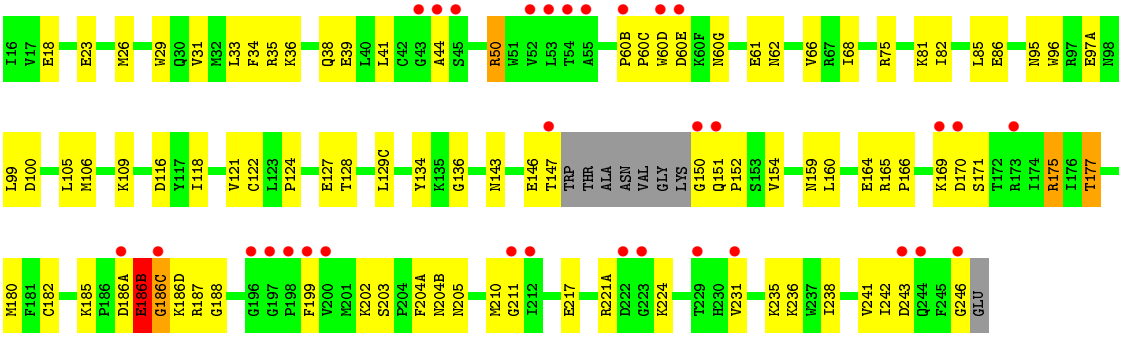
- Molecule 1: thrombin

Chain D:



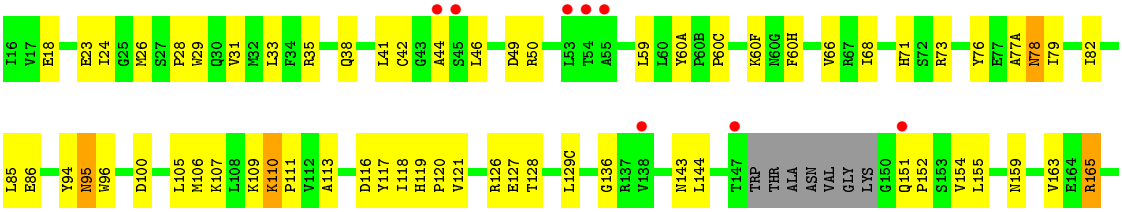
- Molecule 2: thrombin

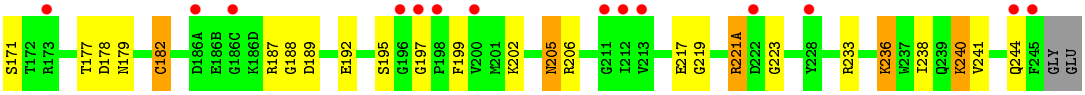
Chain B:



- Molecule 2: thrombin

Chain E:




S171
T172
R173
T177
D178
N179
C182
D186A
E186B
G186C
K186D
R187
G188
D189
E192
S195
G196
G197
P198
F199
V200
M201
K202
N205
R206
G211
I212
V213
E217
G219
R221A
D222
G223
Y228
R233
K236
W237
I238
Q239
K240
V241
Q244
F245
GLY
GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	61.85Å 68.07Å 162.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	23.52 – 2.30 30.93 – 2.30	Depositor EDS
% Data completeness (in resolution range)	83.4 (23.52-2.30) 83.3 (30.93-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.07 (at 2.29Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.201 , 0.251 0.205 , 0.254	Depositor DCC
R_{free} test set	1787 reflections (6.85%)	DCC
Wilson B-factor (Å^2)	36.2	Xtriage
Anisotropy	0.032	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 58.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 26998 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4875	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.76% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: NAG

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	0.46	0/254	0.81	0/338
1	D	0.48	0/232	0.71	0/309
2	B	0.34	0/2075	0.62	0/2804
2	E	0.35	0/2071	0.61	0/2799
All	All	0.36	0/4632	0.63	0/6250

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	252	0	243	17	0
1	D	230	0	229	15	0
2	B	2023	0	1996	75	0
2	E	2019	0	1991	87	0
3	B	14	0	13	7	0
3	E	14	0	12	11	0
4	A	18	0	0	2	0
4	B	130	0	0	8	3
4	D	20	0	0	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	155	0	0	16	3
All	All	4875	0	4484	192	3

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 21.

The worst 5 of 192 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)
3:E:400:NAG:C4	3:E:400:NAG:C5	1.81	1.52
3:E:400:NAG:C4	3:E:400:NAG:O4	1.64	1.44
1:D:14(D):ARG:HB3	4:D:795:HOH:O	1.28	1.32
3:E:400:NAG:O5	3:E:400:NAG:C5	1.78	1.31
3:E:400:NAG:C3	3:E:400:NAG:O3	1.81	1.29

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:746:HOH:O	4:E:555:HOH:O[2_664]	0.86	1.34
4:B:757:HOH:O	4:E:653:HOH:O[4_465]	1.22	0.98
4:B:621:HOH:O	4:E:520:HOH:O[4_555]	1.65	0.55

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	30/36 (83%)	24 (80%)	6 (20%)	0	100 100
1	D	26/36 (72%)	23 (88%)	3 (12%)	0	100 100
2	B	247/259 (95%)	228 (92%)	16 (6%)	3 (1%)	16 16

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	E	246/259 (95%)	229 (93%)	17 (7%)	0	100	100
All	All	549/590 (93%)	504 (92%)	42 (8%)	3 (0%)	34	41

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	186(C)	GLY
2	B	186(B)	GLU
2	B	170	ASP

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	27/31 (87%)	24 (89%)	3 (11%)	8	8
1	D	26/31 (84%)	25 (96%)	1 (4%)	40	54
2	B	218/224 (97%)	208 (95%)	10 (5%)	33	44
2	E	218/224 (97%)	206 (94%)	12 (6%)	27	36
All	All	489/510 (96%)	463 (95%)	26 (5%)	28	37

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	186(B)	GLU
2	E	33	LEU
2	E	236	LYS
2	B	243	ASP
1	D	14(G)	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	151	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	205	ASN
2	E	143	ASN
2	B	95	ASN
2	E	95	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)

2 ligands are 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	$\#\|Z\| > 2$
3	NAG	B	500	-	14,14,15	5.35	9 (64%)	15,19,21	4.27	8 (53%)
3	NAG	E	400	2	14,14,15	8.54	11 (78%)	15,19,21	6.17	11 (73%)

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
3	NAG	B	500	-	-	0/6/23/26	0/1/1/1
3	NAG	E	400	2	-	0/6/23/26	0/1/1/1

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	400	NAG	C1-C2	-8.08	1.41	1.52
3	E	400	NAG	O6-C6	-5.25	1.19	1.42
3	E	400	NAG	O5-C1	-3.73	1.37	1.43
3	B	500	NAG	O3-C3	-2.44	1.37	1.43
3	B	500	NAG	O7-C7	2.79	1.29	1.23

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	E	400	NAG	O3-C3-C4	-13.32	80.35	110.34
3	E	400	NAG	O7-C7-N2	-10.35	100.76	121.86
3	B	500	NAG	C2-N2-C7	-8.64	111.94	123.04
3	E	400	NAG	C3-C4-C5	-7.51	97.11	110.20
3	E	400	NAG	C1-O5-C5	-7.20	103.11	112.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 18 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	500	NAG	7	0
3	E	400	NAG	11	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)

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	32/36 (88%)	0.71	3 (9%)	11	16	27, 39, 69, 97	0
1	D	28/36 (77%)	0.67	4 (14%)	4	6	35, 43, 60, 69	0
2	B	251/259 (96%)	0.62	32 (12%)	5	8	20, 39, 71, 87	0
2	E	250/259 (96%)	0.47	22 (8%)	12	18	18, 36, 60, 74	0
All	All	561/590 (95%)	0.56	61 (10%)	7	11	18, 38, 66, 97	0

The worst 5 of 61 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	15	ARG	12.0
2	B	150	GLY	4.9
2	B	173	ARG	4.7
1	A	1(D)	GLY	4.7
2	B	222	ASP	4.6

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. 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	NAG	E	400	14/15	0.82	0.29	-	68,73,74,75	0
3	NAG	B	500	14/15	0.80	0.26	-	66,73,75,75	0

6.5 Other polymers (i)

There are no such residues in this entry.