wwPDB X-ray Structure Validation Summary Report (i)

Aug 27, 2023 – 06:36 PM EDT

PDB ID	:	3HK6
Title	:	Crystal structure of murine thrombin mutant W215A/E217A (two molecules in the asymmetric unit)
Authors	:	Gandhi, P.S.; Page, M.J.; Chen, Z.; Bush-Pelc, L.; Di Cera, E.
Deposited on	:	2009-05-22
Resolution	:	3.20 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

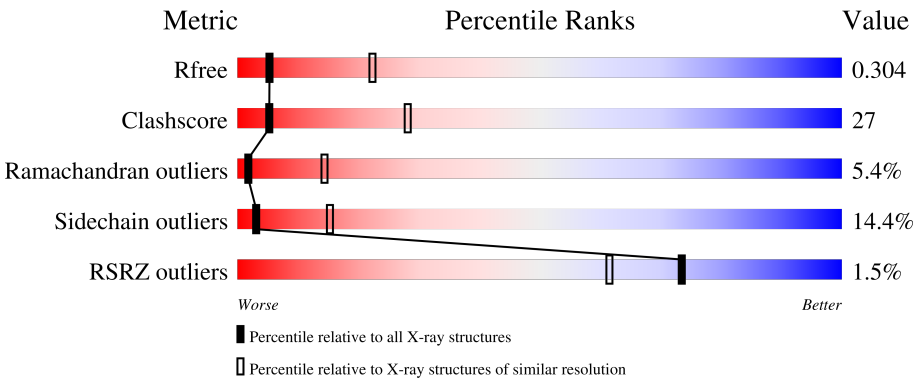
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.20 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}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	44	55% 34% 9% •
1	C	44	2% 50% 41% 7% •
2	B	258	% 43% 44% 12% •
2	D	258	2% 42% 48% 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4880 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	43	343	221	52	69	1	0	0	0
1	C	43	343	221	52	69	1	0	0	0

- Molecule 2 is a protein called Thrombin heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	258	2097	1342	377	366	12	0	0	0
2	D	258	2097	1342	377	366	12	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	215	ALA	TRP	engineered mutation	UNP P19221
B	217	ALA	GLU	engineered mutation	UNP P19221
D	215	ALA	TRP	engineered mutation	UNP P19221
D	217	ALA	GLU	engineered mutation	UNP P19221

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Thrombin light chain

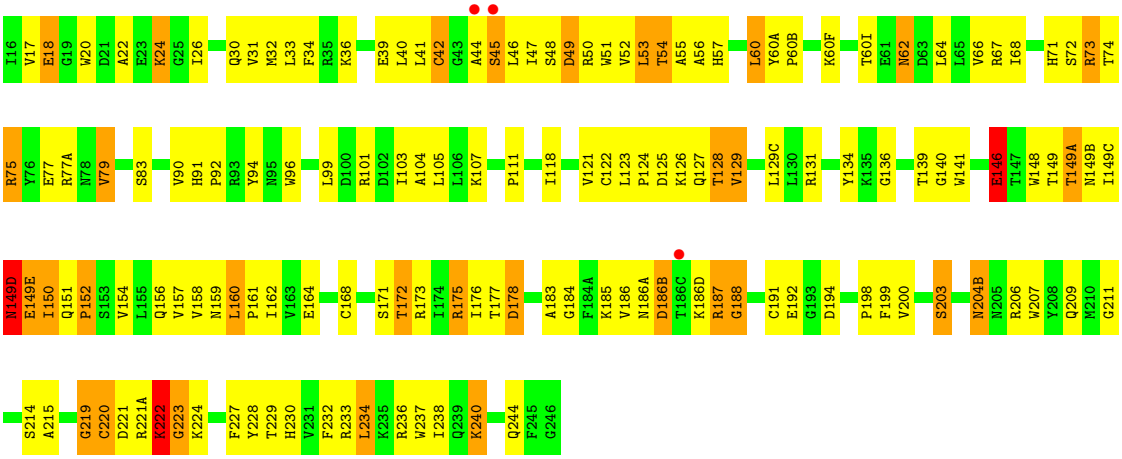
Chain A: 55% 34% 9%

• Molecule 1: Thrombin light chain

Chain C: 2% 50% 41% 7%

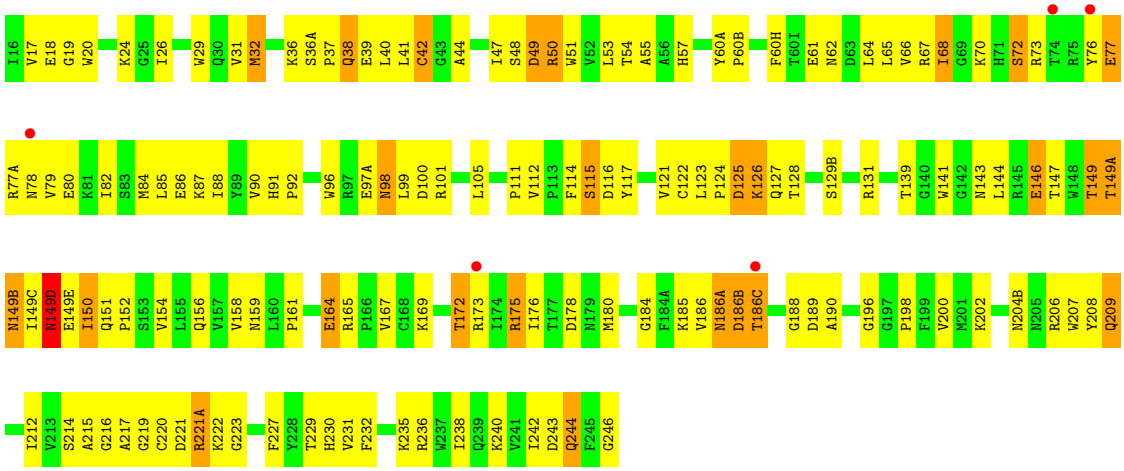
• Molecule 2: Thrombin heavy chain

Chain B: 1% 43% 44% 12%



• Molecule 2: Thrombin heavy chain

Chain D: 2% 42% 48% 10%


I16 V17 E18 G19 W20 K24 G25 I26 W29 Q30 V31 M32 K36 S36A P37 Q38 E39 L40 L41 C42 G43 A44 I47 S48 D49 R50 W51 V52 L53 T54 A55 A56 H57 Y60A P60B F60H T60I E61 N62 D63 L64 L65 V66 R67 I68 G69 K70 H71 S72 R73 T74 R75 Y76 E77
R77A N78 V79 E80 K81 I82 S83 M84 L85 E86 K87 I88 Y89 V90 H91 P92 W96 R97 E97A N98 L99 D100 R101 L105 P111 V112 P113 F114 S115 D116 Y117 V121 C122 L123 P124 D125 K126 Q127 T128 S129B R131 T139 G140 W141 G142 N143 L144 R145 E146 T147 W148 T149 T149A
M149B I149C N149D E149E I150 Q151 P152 S153 V154 L155 Q156 V157 V158 N159 L160 P161 E164 R165 P166 V167 C168 K169 T172 R173 I174 R175 I176 T177 D178 N179 M180 G184 F184A K185 V186 N186A D186B T186C G188 D189 A190 G196 G197 P198 F199 V200 M201 K202 N204B N205 R206 W207 Y208 Q209
I212 V213 S214 A215 G216 A217 G219 C220 D221 R221A K222 G223 F227 Y228 T229 H230 V231 F232 K235 R236 W237 I238 Q239 K240 V241 I242 D243 Q244 F245 G246

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	70.39Å 70.39Å 293.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 3.20 29.51 – 3.20	Depositor EDS
% Data completeness (in resolution range)	93.4 (40.00-3.20) 93.6 (29.51-3.20)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.71 (at 3.18Å)	Xtriage
Refinement program	REFMAC 5.5.0070	Depositor
R, R_{free}	0.222 , 0.314 0.214 , 0.304	Depositor DCC
R_{free} test set	583 reflections (4.80%)	wwPDB-VP
Wilson B-factor (Å^2)	64.1	Xtriage
Anisotropy	0.206	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 39.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	4880	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.37% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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	0.90	0/349	0.91	0/466
1	C	0.79	0/349	0.84	0/466
2	B	0.95	6/2152 (0.3%)	0.96	2/2913 (0.1%)
2	D	0.87	3/2152 (0.1%)	0.91	5/2913 (0.2%)
All	All	0.90	9/5002 (0.2%)	0.93	7/6758 (0.1%)

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
2	B	0	2
2	D	0	1
All	All	0	3

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	215	ALA	C-N	8.36	1.48	1.33
2	B	42	CYS	CB-SG	-7.12	1.70	1.82
2	D	42	CYS	CB-SG	-7.03	1.70	1.82
2	B	222	LYS	CD-CE	5.90	1.66	1.51
2	D	164	GLU	CG-CD	5.83	1.60	1.51

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	222	LYS	CD-CE-NZ	8.26	130.69	111.70
2	D	175	ARG	NE-CZ-NH1	6.85	123.72	120.30
2	B	149(E)	GLU	N-CA-C	-6.44	93.60	111.00
2	D	90	VAL	CB-CA-C	-6.06	99.88	111.40
2	D	149	THR	N-CA-C	-5.90	95.06	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	149(D)	ASN	Peptide
2	B	219	GLY	Mainchain
2	D	149(D)	ASN	Peptide

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	343	0	331	14	0
1	C	343	0	331	15	0
2	B	2097	0	2098	127	0
2	D	2097	0	2098	122	0
All	All	4880	0	4858	259	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 27.

The worst 5 of 259 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)
2:D:221(A):ARG:HH21	2:D:223:GLY:CA	1.59	1.16
2:D:47:ILE:HD11	2:D:51:TRP:HB2	1.15	1.13
2:D:221(A):ARG:NH2	2:D:223:GLY:HA2	1.66	1.08
2:B:68:ILE:HG22	2:B:118:ILE:HG12	1.35	1.06
2:D:221(A):ARG:NH2	2:D:223:GLY:CA	2.20	1.04

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	41/44 (93%)	25 (61%)	11 (27%)	5 (12%)	0 2
1	C	41/44 (93%)	33 (80%)	7 (17%)	1 (2%)	6 34
2	B	256/258 (99%)	203 (79%)	37 (14%)	16 (6%)	1 10
2	D	256/258 (99%)	203 (79%)	43 (17%)	10 (4%)	3 22
All	All	594/604 (98%)	464 (78%)	98 (16%)	32 (5%)	2 14

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	149(A)	THR
2	B	178	ASP
2	B	222	LYS
1	C	14(K)	ILE
2	D	186	VAL

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	37/39 (95%)	35 (95%)	2 (5%)	22 58
1	C	37/39 (95%)	31 (84%)	6 (16%)	2 11
2	B	227/227 (100%)	195 (86%)	32 (14%)	3 16
2	D	227/227 (100%)	191 (84%)	36 (16%)	2 12
All	All	528/532 (99%)	452 (86%)	76 (14%)	3 15

5 of 76 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	146	GLU
2	D	209	GLN
2	D	150	ILE
2	D	175	ARG
2	D	244	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	151	GLN
2	D	156	GLN
2	D	209	GLN
2	D	204(B)	ASN
2	B	204(B)	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	43/44 (97%)	-0.20	0	100	100	36, 46, 69, 77	0
1	C	43/44 (97%)	-0.25	1 (2%)	60	47	43, 56, 68, 76	0
2	B	258/258 (100%)	-0.27	3 (1%)	79	67	26, 40, 54, 66	0
2	D	258/258 (100%)	-0.19	5 (1%)	66	53	30, 44, 66, 80	0
All	All	602/604 (99%)	-0.23	9 (1%)	73	61	26, 44, 63, 80	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	186(C)	THR	5.4
1	C	14(L)	ASP	4.1
2	D	186(C)	THR	4.0
2	B	45	SER	2.4
2	D	173	ARG	2.3

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 monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.