W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Sep 6, 2017 – 06:33 PM EDT

PDB ID	:	5K82
Title	:	Crystal Structure of a Primate APOBEC3G N-Terminal Domain
Authors	:	Xiao, X.; Li, S.-X.; Yang, H.; Chen, X.S.
Deposited on	:	unknown
Resolution	:	2.91 Å(reported)

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

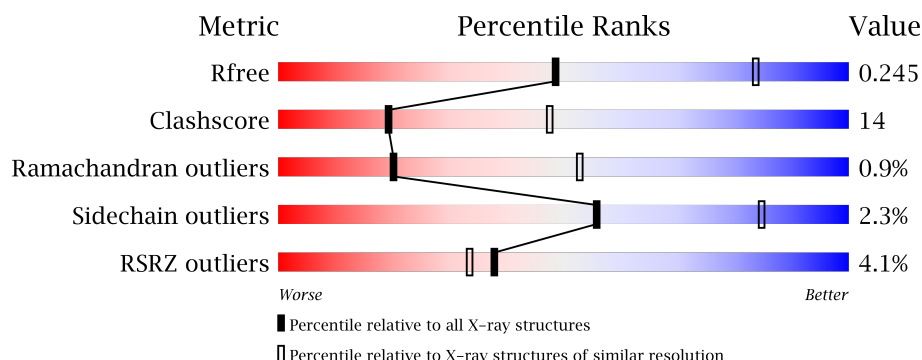|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20029824

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 2.91 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
Rfree		0.245
Clashscore		14
Ramachandran outliers		0.9%
Sidechain outliers		2.3%
RSRZ outliers		4.1%

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))
R_{free}	100719	1813 (2.94-2.90)
Clashscore	112137	2045 (2.94-2.90)
Ramachandran outliers	110173	1997 (2.94-2.90)
Sidechain outliers	110143	1999 (2.94-2.90)
RSRZ outliers	101464	1825 (2.94-2.90)

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	196	4% 65% 28% · ·
1	B	196	4% 66% 29% · ·
1	C	196	4% 66% 26% · · ·
1	D	196	4% 69% 27% · ·

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
2	ZN	C	401	-	-	-	X

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6342 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 Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G,Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	188	Total	C	N	O	S	0	0	0
			1579	1029	275	265	10			
1	B	188	Total	C	N	O	S	0	0	0
			1578	1029	275	264	10			
1	C	188	Total	C	N	O	S	0	0	0
			1578	1029	275	264	10			
1	D	191	Total	C	N	O	S	0	0	0
			1603	1045	279	268	11			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP M1GSK9
A	-3	PRO	-	expression tag	UNP M1GSK9
A	-2	ALA	-	expression tag	UNP M1GSK9
A	-1	GLY	-	expression tag	UNP M1GSK9
A	0	SER	-	expression tag	UNP M1GSK9
A	143	ALA	-	linker	UNP M1GSK9
A	144	GLU	-	linker	UNP M1GSK9
A	145	ALA	-	linker	UNP M1GSK9
A	146	GLY	-	linker	UNP M1GSK9
B	-4	GLY	-	expression tag	UNP M1GSK9
B	-3	PRO	-	expression tag	UNP M1GSK9
B	-2	ALA	-	expression tag	UNP M1GSK9
B	-1	GLY	-	expression tag	UNP M1GSK9
B	0	SER	-	expression tag	UNP M1GSK9
B	143	ALA	-	linker	UNP M1GSK9
B	144	GLU	-	linker	UNP M1GSK9
B	145	ALA	-	linker	UNP M1GSK9
B	146	GLY	-	linker	UNP M1GSK9
C	-4	GLY	-	expression tag	UNP M1GSK9
C	-3	PRO	-	expression tag	UNP M1GSK9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-2	ALA	-	expression tag	UNP M1GSK9
C	-1	GLY	-	expression tag	UNP M1GSK9
C	0	SER	-	expression tag	UNP M1GSK9
C	143	ALA	-	linker	UNP M1GSK9
C	144	GLU	-	linker	UNP M1GSK9
C	145	ALA	-	linker	UNP M1GSK9
C	146	GLY	-	linker	UNP M1GSK9
D	-4	GLY	-	expression tag	UNP M1GSK9
D	-3	PRO	-	expression tag	UNP M1GSK9
D	-2	ALA	-	expression tag	UNP M1GSK9
D	-1	GLY	-	expression tag	UNP M1GSK9
D	0	SER	-	expression tag	UNP M1GSK9
D	143	ALA	-	linker	UNP M1GSK9
D	144	GLU	-	linker	UNP M1GSK9
D	145	ALA	-	linker	UNP M1GSK9
D	146	GLY	-	linker	UNP M1GSK9

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total 1	Zn 1	0	0
2	A	1	Total 1	Zn 1	0	0
2	D	1	Total 1	Zn 1	0	0
2	C	1	Total 1	Zn 1	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.
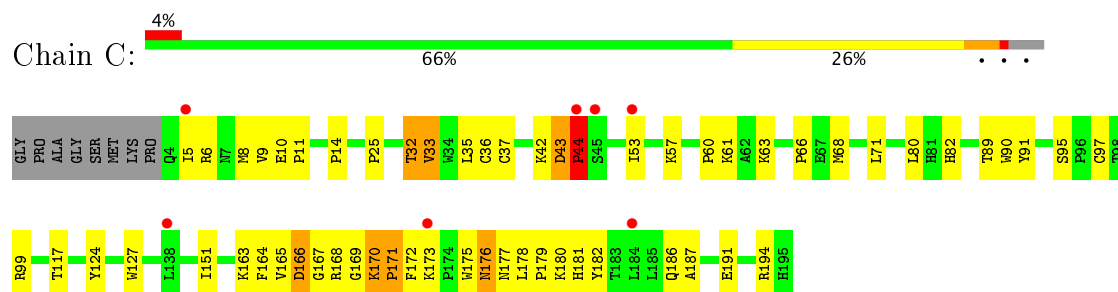
• Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G,Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G

• Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G,Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G
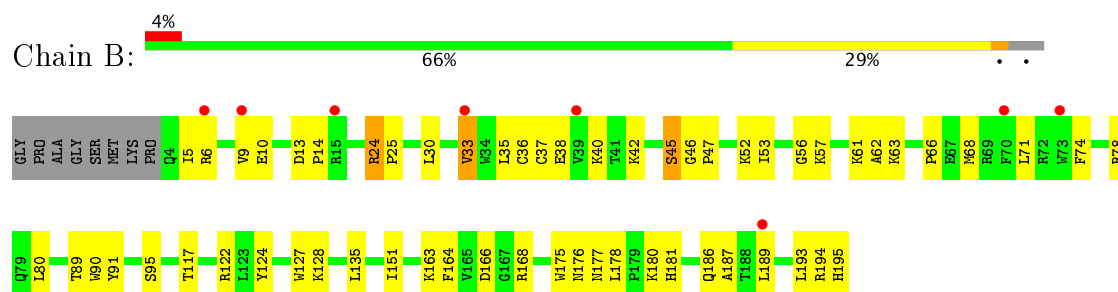
• Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G,Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G
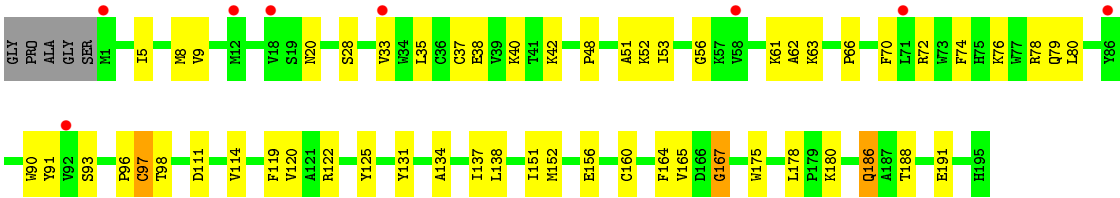

• Molecule 1: Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G,Apolipoprotein B mRNA editing enzyme, catalytic peptide-like 3G

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	152.40Å 152.40Å 79.78Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.30 – 2.91 49.88 – 2.91	Depositor EDS
% Data completeness (in resolution range)	98.4 (42.30-2.91) 91.3 (49.88-2.91)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.11 (at 2.91Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.224 , 0.251 0.221 , 0.245	Depositor DCC
R_{free} test set	1837 reflections (4.45%)	DCC
Wilson B-factor (Å^2)	71.1	Xtriage
Anisotropy	0.772	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 51.2	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.022 for -h,-k,l 0.477 for h,-h-k,-l 0.026 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6342	wwPDB-VP
Average B, all atoms (Å^2)	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.34% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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: ZN

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	0.57	0/1634	0.80	3/2219 (0.1%)
1	B	0.56	1/1633 (0.1%)	0.83	4/2219 (0.2%)
1	C	0.76	3/1633 (0.2%)	0.85	7/2219 (0.3%)
1	D	0.53	0/1659	0.71	0/2252
All	All	0.61	4/6559 (0.1%)	0.80	14/8909 (0.2%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	44	PRO	N-CD	18.93	1.74	1.47
1	B	37	CYS	CB-SG	-5.14	1.73	1.81
1	C	171	PRO	N-CD	5.06	1.54	1.47
1	C	37	CYS	CB-SG	-5.06	1.73	1.81

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	45	SER	N-CA-C	14.13	149.15	111.00
1	A	145	ALA	CB-CA-C	-13.10	90.46	110.10
1	C	43	ASP	C-N-CD	-10.83	96.77	120.60
1	B	45	SER	CB-CA-C	-10.00	91.09	110.10
1	C	44	PRO	N-CA-CB	9.94	115.23	103.30

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	1579	0	1556	59	0
1	B	1578	0	1556	42	0
1	C	1578	0	1555	49	0
1	D	1603	0	1588	41	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
All	All	6342	0	6255	182	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 182 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:C:44:PRO:N	1:C:44:PRO:CD	1.74	1.35
1:C:171:PRO:HD2	1:C:173:LYS:NZ	1.50	1.22
1:C:43:ASP:C	1:C:44:PRO:CD	2.33	0.98
1:C:171:PRO:HD2	1:C:173:LYS:HZ2	1.16	0.91
1:A:101:ALA:HB1	1:A:138:LEU:HD21	1.50	0.90

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	186/196 (95%)	176 (95%)	8 (4%)	2 (1%)	17	47
1	B	186/196 (95%)	176 (95%)	8 (4%)	2 (1%)	17	47
1	C	186/196 (95%)	175 (94%)	9 (5%)	2 (1%)	17	47
1	D	189/196 (96%)	183 (97%)	5 (3%)	1 (0%)	32	67
All	All	747/784 (95%)	710 (95%)	30 (4%)	7 (1%)	20	53

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	44	PRO
1	B	45	SER
1	D	167	GLY
1	A	29	GLY
1	A	167	GLY

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	170/175 (97%)	168 (99%)	2 (1%)	75	93
1	B	170/175 (97%)	166 (98%)	4 (2%)	54	84
1	C	170/175 (97%)	163 (96%)	7 (4%)	35	69
1	D	173/175 (99%)	170 (98%)	3 (2%)	66	89
All	All	683/700 (98%)	667 (98%)	16 (2%)	56	85

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	32	THR
1	C	33	VAL
1	C	176	ASN
1	C	6	ARG
1	D	28	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

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 (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	188/196 (95%)	0.69	8 (4%)	36	32	50, 94, 126, 141	0
1	B	188/196 (95%)	0.74	8 (4%)	36	32	52, 85, 123, 148	0
1	C	188/196 (95%)	0.74	7 (3%)	42	37	51, 84, 126, 150	0
1	D	191/196 (97%)	0.69	8 (4%)	37	32	48, 93, 130, 170	0
All	All	755/784 (96%)	0.71	31 (4%)	38	33	48, 88, 127, 170	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	44	PRO	4.5
1	C	5	ILE	4.5
1	D	1	MET	4.5
1	B	15	ARG	3.7
1	C	45	SER	3.3

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 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
2	ZN	C	401	1/1	0.98	0.32	<mark>3.55</mark>	73,73,73,73	0
2	ZN	B	401	1/1	0.99	0.31	0.68	78,78,78,78	0
2	ZN	A	401	1/1	0.95	0.31	-	92,92,92,92	0
2	ZN	D	401	1/1	0.96	0.28	-	92,92,92,92	0

6.5 Other polymers (i)

There are no such residues in this entry.