wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 07:47 PM GMT

PDB ID	:	1H6K
Title	:	NUCLEAR CAP BINDING COMPLEX
Authors	:	Mazza, C.; Ohno, M.; Segref, A.; Mattaj, I.W.; Cusack, S.
Deposited on	:	2001-06-18
Resolution	:	2.00 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:

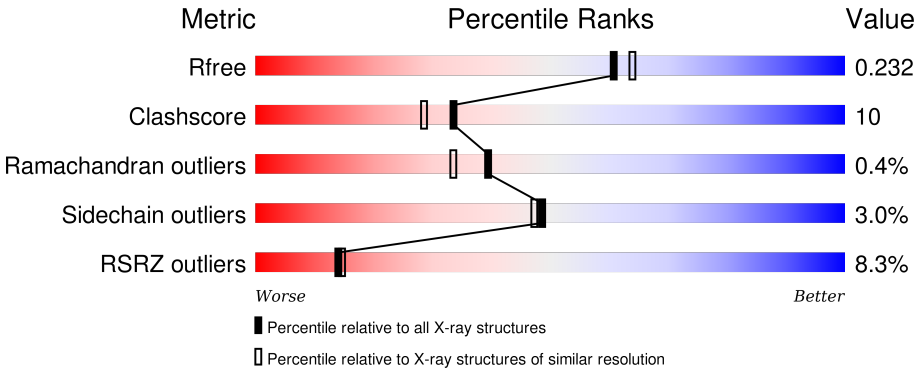
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.00 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.



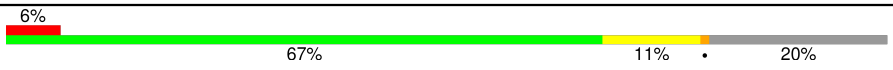
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

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	757	6% 77% 18%
1	B	757	8% 77% 17%
1	C	757	10% 77% 17%
2	X	98	4% 64% 13% 21%
2	Y	98	5% 64% 12% 22%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	Z	98	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 21515 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 CBP80.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	728	Total	C	N	O	S	0	0	0
			5959	3842	1004	1075	38			
1	B	729	Total	C	N	O	S	0	0	0
			5968	3846	1004	1080	38			
1	C	733	Total	C	N	O	S	0	0	0
			5998	3866	1009	1085	38			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	479	SER	ALA	ENGINEERED MUTATION	UNP Q09161
B	479	SER	ALA	ENGINEERED MUTATION	UNP Q09161
C	479	SER	ALA	ENGINEERED MUTATION	UNP Q09161

- Molecule 2 is a protein called 20 KDA NUCLEAR CAP BINDING PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	77	Total	C	N	O	S	0	0	0
			625	396	102	121	6			
2	Y	76	Total	C	N	O	S	0	0	0
			621	394	101	120	6			
2	Z	78	Total	C	N	O	S	0	0	0
			634	402	104	122	6			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	527	Total	O	0	0
			527	527		
3	B	535	Total	O	0	0
			535	535		

Continued on next page...

Continued from previous page...

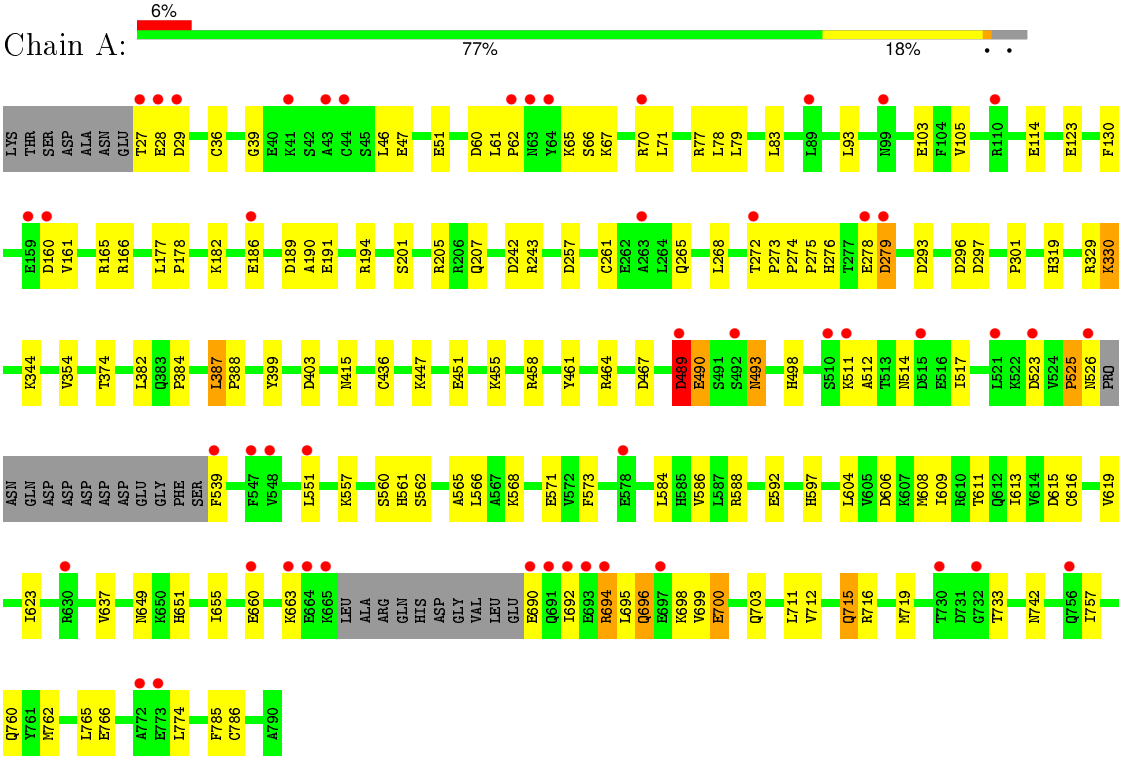
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	435	Total O 435 435	0	0
3	X	79	Total O 79 79	0	0
3	Y	63	Total O 63 63	0	0
3	Z	71	Total O 71 71	0	0

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.

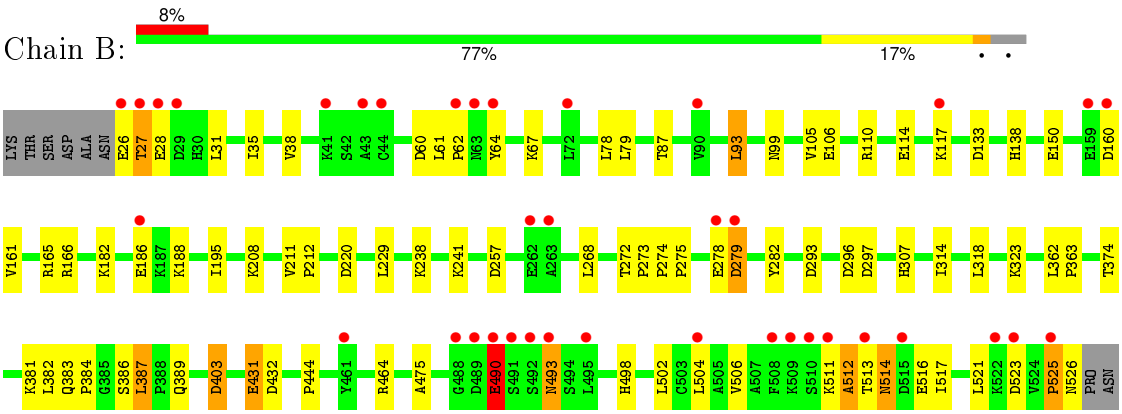
- Molecule 1: CBP80

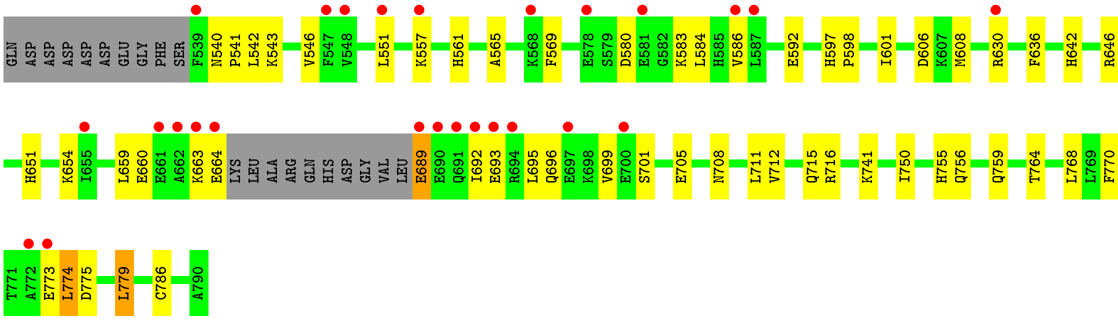
Chain A:



- Molecule 1: CBP80

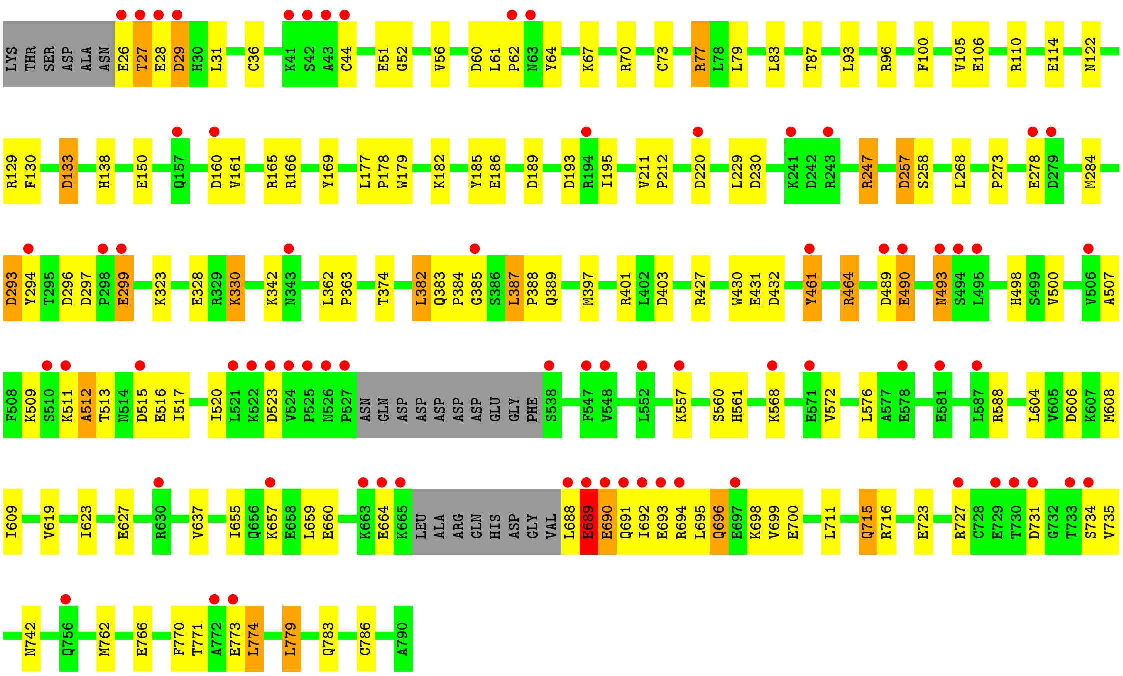
Chain B:





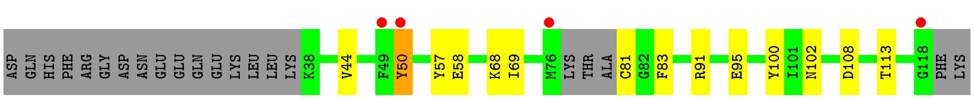
• Molecule 1: CBP80

Chain C:



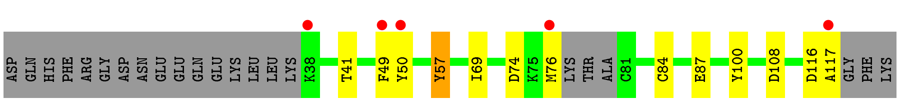
• Molecule 2: 20 KDA NUCLEAR CAP BINDING PROTEIN

Chain X:



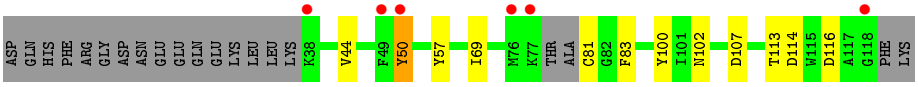
• Molecule 2: 20 KDA NUCLEAR CAP BINDING PROTEIN

Chain Y:



• Molecule 2: 20 KDA NUCLEAR CAP BINDING PROTEIN

Chain Z:


ASP
GLN
HIS
PHE
ARG
GLY
ASP
ASN
GLU
GLU
GLN
GLU
LYS
LEU
LEU
LYS
K38
V44
F49
Y50
Y57
I69
M76
K77
THR
ALA
C81
G82
F83
Y100
I101
N102
D107
T113
D114
W115
D116
A117
G118
PHE
LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	75.53Å 161.48Å 303.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00 19.90 – 1.90	Depositor EDS
% Data completeness (in resolution range)	83.0 (20.00-2.00) 75.6 (19.90-1.90)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.0	Depositor
R, R_{free}	0.203 , 0.234 0.202 , 0.232	Depositor DCC
R_{free} test set	2157 reflections (1.05%)	DCC
Wilson B-factor (Å^2)	32.7	Xtriage
Anisotropy	0.509	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 42.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 220200 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	21515	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 2.03% 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

5.1 Standard geometry

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	0.40	0/6111	0.64	16/8291 (0.2%)
1	B	0.39	0/6120	0.64	15/8304 (0.2%)
1	C	0.38	0/6151	0.64	18/8346 (0.2%)
2	X	0.46	0/635	0.79	1/850 (0.1%)
2	Y	0.43	0/631	0.81	2/845 (0.2%)
2	Z	0.44	0/644	0.82	2/861 (0.2%)
All	All	0.40	0/20292	0.66	54/27497 (0.2%)

There are no bond length outliers.

The worst 5 of 54 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	403	ASP	CB-CG-OD2	6.39	124.05	118.30
1	A	293	ASP	CB-CG-OD2	6.38	124.05	118.30
2	Y	108	ASP	CB-CG-OD2	6.12	123.81	118.30
2	Z	114	ASP	CB-CG-OD2	6.12	123.81	118.30
1	B	293	ASP	CB-CG-OD2	6.07	123.77	118.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	5959	0	5944	127	0
1	B	5968	0	5943	117	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	5998	0	5979	136	0
2	X	625	0	602	12	0
2	Y	621	0	599	9	0
2	Z	634	0	615	8	0
3	A	527	0	0	44	0
3	B	535	0	0	36	0
3	C	435	0	0	44	0
3	X	79	0	0	5	0
3	Y	63	0	0	4	0
3	Z	71	0	0	1	0
All	All	21515	0	19682	404	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 10.

The worst 5 of 404 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:36:CYS:HB3	3:C:2002:HOH:O	1.31	1.25
1:B:689:GLU:HB3	3:B:2475:HOH:O	1.30	1.23
1:C:186:GLU:HB3	3:C:2102:HOH:O	1.43	1.19
1:C:773:GLU:HB2	3:C:2421:HOH:O	1.46	1.16
1:A:526:ASN:HA	3:A:2404:HOH:O	1.47	1.13

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	722/757 (95%)	704 (98%)	15 (2%)	3 (0%)	39	33
1	B	723/757 (96%)	702 (97%)	18 (2%)	3 (0%)	39	33

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	C	727/757 (96%)	709 (98%)	14 (2%)	4 (1%)	30 22
2	X	73/98 (74%)	72 (99%)	1 (1%)	0	100 100
2	Y	72/98 (74%)	72 (100%)	0	0	100 100
2	Z	74/98 (76%)	74 (100%)	0	0	100 100
All	All	2391/2565 (93%)	2333 (98%)	48 (2%)	10 (0%)	39 33

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	490	GLU
1	B	490	GLU
1	B	525	PRO
1	C	490	GLU
1	C	512	ALA

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	669/694 (96%)	650 (97%)	19 (3%)	51 50
1	B	670/694 (96%)	651 (97%)	19 (3%)	51 50
1	C	674/694 (97%)	650 (96%)	24 (4%)	42 39
2	X	67/86 (78%)	66 (98%)	1 (2%)	72 75
2	Y	67/86 (78%)	65 (97%)	2 (3%)	48 47
2	Z	68/86 (79%)	67 (98%)	1 (2%)	72 75
All	All	2215/2340 (95%)	2149 (97%)	66 (3%)	48 47

5 of 66 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	514	ASN
1	B	774	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	774	LEU
1	B	584	LEU
1	B	693	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	561	HIS
1	B	706	GLN
1	C	708	ASN
1	B	656	GLN
1	B	708	ASN

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

There are no ligands in this entry.

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, 95^{th} 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	728/757 (96%)	0.22	49 (6%)	21	22	22, 35, 67, 81	0
1	B	729/757 (96%)	0.25	64 (8%)	12	13	22, 36, 70, 82	0
1	C	733/757 (96%)	0.35	72 (9%)	10	10	24, 39, 69, 81	0
2	X	77/98 (78%)	0.15	4 (5%)	31	33	27, 34, 47, 54	0
2	Y	76/98 (77%)	0.19	5 (6%)	22	22	28, 35, 51, 61	0
2	Z	78/98 (79%)	0.10	6 (7%)	16	17	27, 35, 52, 66	0
All	All	2421/2565 (94%)	0.26	200 (8%)	14	15	22, 37, 67, 82	0

The worst 5 of 200 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	688	LEU	7.3
1	C	690	GLU	6.6
1	A	691	GLN	6.4
1	B	27	THR	6.4
1	A	526	ASN	6.1

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

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.