wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 11:09 PM GMT

PDB ID	:	5AJD
Title	:	Not1 C-terminal domain in complex with Not4
Authors	:	Bhaskar, V.; Basquin, J.; Conti, E.
Deposited on	:	2015-02-23
Resolution	:	3.62 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:

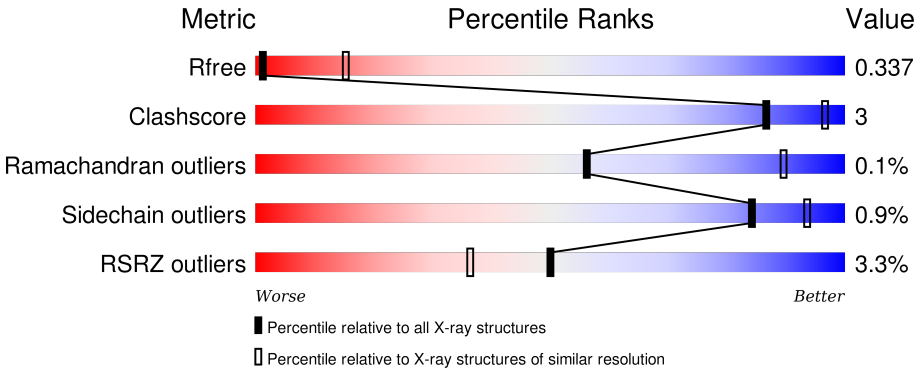
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 3.62 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.



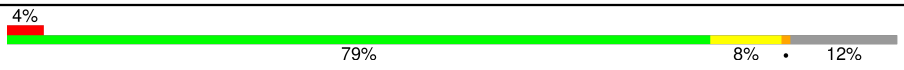
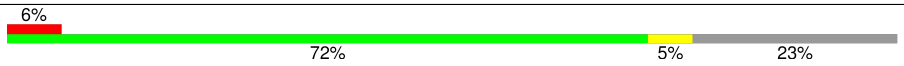
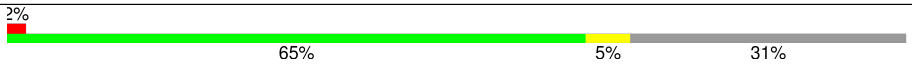
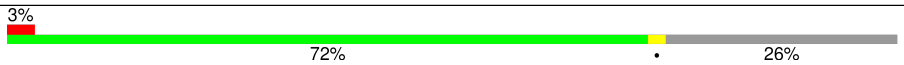
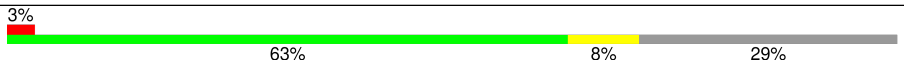
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1093 (3.80-3.44)
Clashscore	102246	1043 (3.78-3.46)
Ramachandran outliers	100387	1003 (3.78-3.46)
Sidechain outliers	100360	1003 (3.78-3.46)
RSRZ outliers	91569	1100 (3.80-3.44)

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	556	83% 5% 12%
1	C	556	3% 81% 6% 12%
1	E	556	2% 83% 6% 11%
1	G	556	4% 81% 7% 12%
1	I	556	3% 80% 8% 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	K	556	
2	B	65	
2	D	65	
2	F	65	
2	H	65	
2	J	65	
2	L	65	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 47509 atoms, of which 22962 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 CDC39P.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	489	Total	C	H	N	O	S	0	0	0
			7349	2487	3584	602	665	11			
1	C	487	Total	C	H	N	O	S	0	0	0
			7279	2461	3541	593	673	11			
1	E	496	Total	C	H	N	O	S	0	0	0
			7357	2498	3563	607	679	10			
1	G	489	Total	C	H	N	O	S	0	0	0
			7150	2442	3448	592	658	10			
1	I	490	Total	C	H	N	O	S	0	0	0
			7299	2473	3543	599	675	9			
1	K	489	Total	C	H	N	O	S	0	0	0
			7164	2439	3465	586	663	11			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1538	ARG	-	EXPRESSION TAG	UNP N1P976
A	1539	SER	-	EXPRESSION TAG	UNP N1P976
A	1540	MET	-	EXPRESSION TAG	UNP N1P976
C	1538	ARG	-	EXPRESSION TAG	UNP N1P976
C	1539	SER	-	EXPRESSION TAG	UNP N1P976
C	1540	MET	-	EXPRESSION TAG	UNP N1P976
E	1538	ARG	-	EXPRESSION TAG	UNP N1P976
E	1539	SER	-	EXPRESSION TAG	UNP N1P976
E	1540	MET	-	EXPRESSION TAG	UNP N1P976
G	1538	ARG	-	EXPRESSION TAG	UNP N1P976
G	1539	SER	-	EXPRESSION TAG	UNP N1P976
G	1540	MET	-	EXPRESSION TAG	UNP N1P976
I	1538	ARG	-	EXPRESSION TAG	UNP N1P976
I	1539	SER	-	EXPRESSION TAG	UNP N1P976
I	1540	MET	-	EXPRESSION TAG	UNP N1P976
K	1538	ARG	-	EXPRESSION TAG	UNP N1P976
K	1539	SER	-	EXPRESSION TAG	UNP N1P976

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
K	1540	MET	-	EXPRESSION TAG	UNP N1P976

- Molecule 2 is a protein called GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	50	Total	C	H	N	O	0	0	0
			704	243	328	63	70			
2	D	50	Total	C	H	N	O	0	0	0
			704	242	325	64	73			
2	F	43	Total	C	H	N	O	0	0	0
			619	212	294	56	57			
2	H	45	Total	C	H	N	O	0	0	0
			616	216	286	54	60			
2	J	48	Total	C	H	N	O	0	0	0
			661	229	304	62	66			
2	L	46	Total	C	H	N	O	0	0	0
			607	210	281	58	58			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	413	GLY	-	EXPRESSION TAG	UNP P34909
B	414	PRO	-	EXPRESSION TAG	UNP P34909
B	415	ASP	-	EXPRESSION TAG	UNP P34909
B	416	SER	-	EXPRESSION TAG	UNP P34909
B	417	MET	-	EXPRESSION TAG	UNP P34909
D	413	GLY	-	EXPRESSION TAG	UNP P34909
D	414	PRO	-	EXPRESSION TAG	UNP P34909
D	415	ASP	-	EXPRESSION TAG	UNP P34909
D	416	SER	-	EXPRESSION TAG	UNP P34909
D	417	MET	-	EXPRESSION TAG	UNP P34909
F	413	GLY	-	EXPRESSION TAG	UNP P34909
F	414	PRO	-	EXPRESSION TAG	UNP P34909
F	415	ASP	-	EXPRESSION TAG	UNP P34909
F	416	SER	-	EXPRESSION TAG	UNP P34909
F	417	MET	-	EXPRESSION TAG	UNP P34909
H	413	GLY	-	EXPRESSION TAG	UNP P34909
H	414	PRO	-	EXPRESSION TAG	UNP P34909
H	415	ASP	-	EXPRESSION TAG	UNP P34909
H	416	SER	-	EXPRESSION TAG	UNP P34909
H	417	MET	-	EXPRESSION TAG	UNP P34909

Continued on next page...

Continued from previous page...

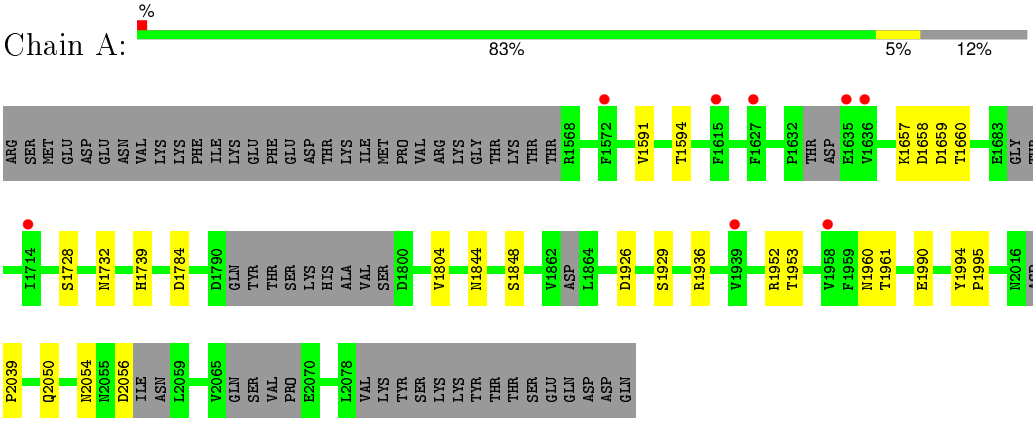
Chain	Residue	Modelled	Actual	Comment	Reference
J	413	GLY	-	EXPRESSION TAG	UNP P34909
J	414	PRO	-	EXPRESSION TAG	UNP P34909
J	415	ASP	-	EXPRESSION TAG	UNP P34909
J	416	SER	-	EXPRESSION TAG	UNP P34909
J	417	MET	-	EXPRESSION TAG	UNP P34909
L	413	GLY	-	EXPRESSION TAG	UNP P34909
L	414	PRO	-	EXPRESSION TAG	UNP P34909
L	415	ASP	-	EXPRESSION TAG	UNP P34909
L	416	SER	-	EXPRESSION TAG	UNP P34909
L	417	MET	-	EXPRESSION TAG	UNP P34909

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.

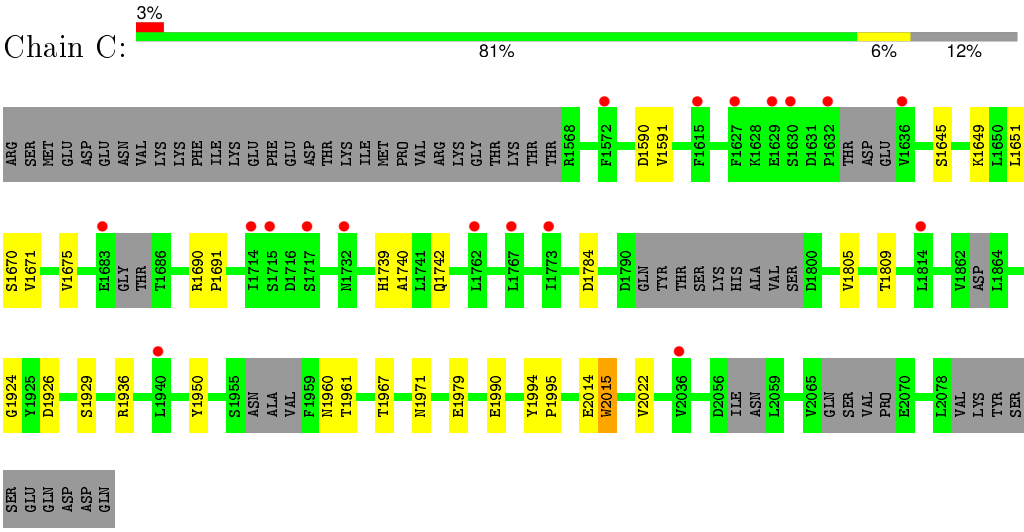
- Molecule 1: CDC39P

Chain A:



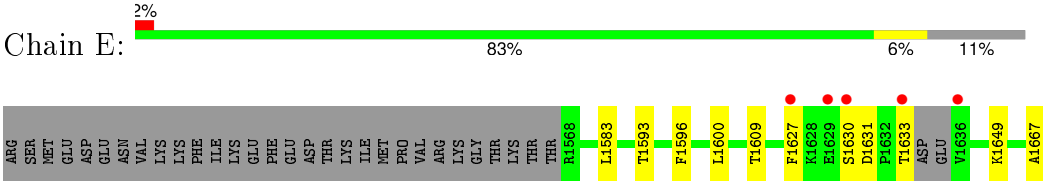
- Molecule 1: CDC39P

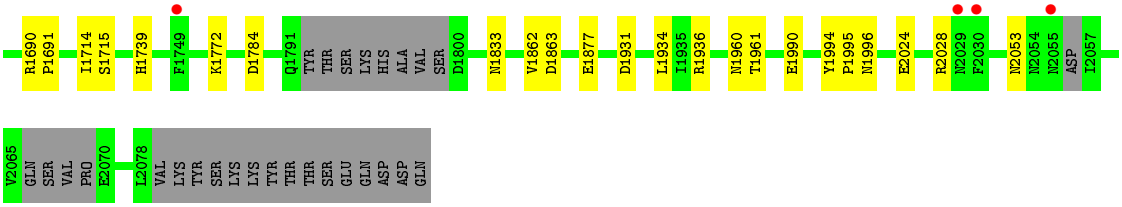
Chain C:



- Molecule 1: CDC39P

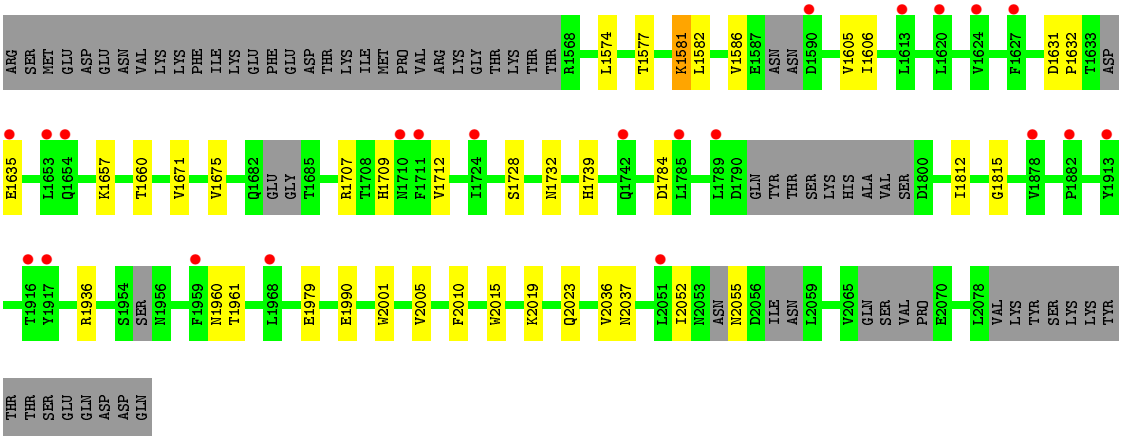
Chain E:





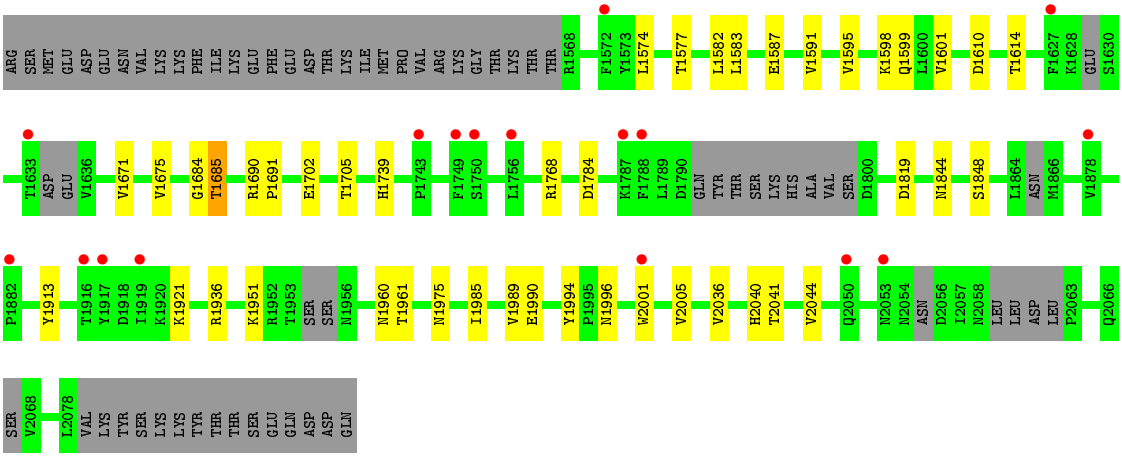
• Molecule 1: CDC39P

Chain G:



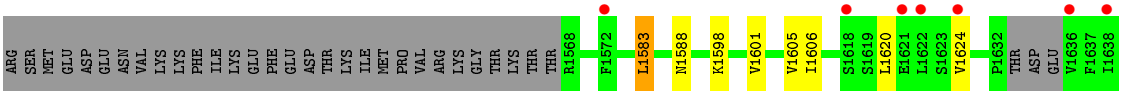
• Molecule 1: CDC39P

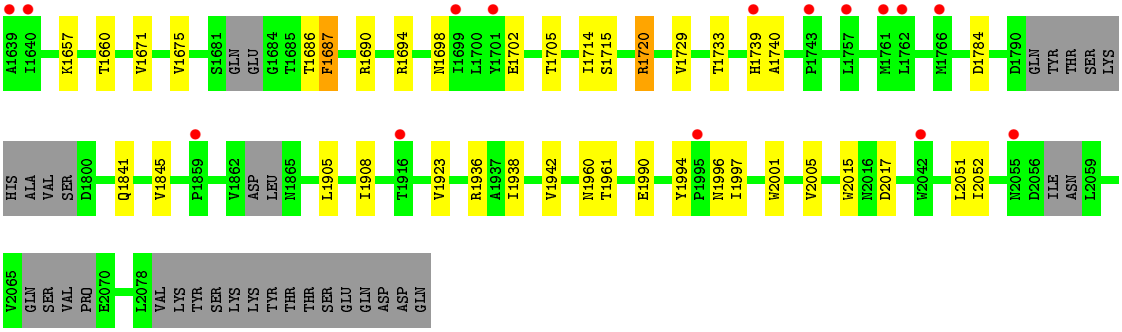
Chain I:



• Molecule 1: CDC39P

Chain K:



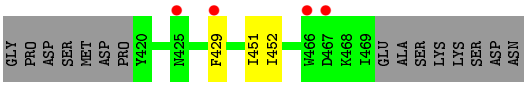


• Molecule 2: GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 4

Chain B:

• Molecule 2: GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 4

Chain D:

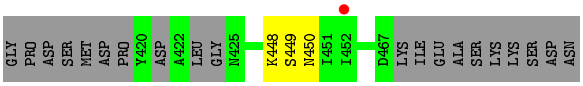


• Molecule 2: GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 4

Chain F:

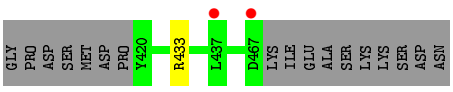
• Molecule 2: GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 4

Chain H:



• Molecule 2: GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 4

Chain J:



• Molecule 2: GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 4

Chain L:


GLY
PRO
ASP
SER
MET
ASP
PRO
TYR
ASP
ALA
LEU
G424
R433
R443
I451
I452
W466
I469
GLU
ALA
SER
LYS
LYS
SER
ASP
ASN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	173.66Å 173.66Å 262.61Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	75.66 – 3.62 86.83 – 3.62	Depositor EDS
% Data completeness (in resolution range)	99.0 (75.66-3.62) 99.6 (86.83-3.62)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.03 (at 3.58Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.266 , 0.319 0.290 , 0.337	Depositor DCC
R_{free} test set	2628 reflections (5.26%)	DCC
Wilson B-factor (Å^2)	133.1	Xtriage
Anisotropy	0.342	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 78.9	EDS
Estimated twinning fraction	0.000 for -h,-k,l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Outliers	8 of 52574 reflections (0.015%)	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	47509	wwPDB-VP
Average B, all atoms (Å^2)	135.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 59.20 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.8292e-05.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	0.21	0/3853	0.36	0/5263
1	C	0.22	0/3824	0.36	0/5223
1	E	0.22	0/3886	0.37	0/5316
1	G	0.22	0/3787	0.36	0/5175
1	I	0.22	0/3843	0.36	0/5246
1	K	0.22	0/3784	0.38	0/5181
2	B	0.24	0/385	0.38	0/527
2	D	0.22	0/388	0.39	0/531
2	F	0.25	0/333	0.39	0/455
2	H	0.26	0/336	0.47	0/458
2	J	0.21	0/366	0.40	0/500
2	L	0.23	0/333	0.39	0/456
All	All	0.22	0/25118	0.37	0/34331

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

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	3765	3584	3581	16	0
1	C	3738	3541	3538	20	0
1	E	3794	3563	3562	19	0
1	G	3702	3448	3444	21	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	I	3756	3543	3540	20	0
1	K	3699	3465	3462	27	0
2	B	376	328	328	4	0
2	D	379	325	325	1	0
2	F	325	294	293	1	0
2	H	330	286	284	1	0
2	J	357	304	304	1	0
2	L	326	281	280	3	0
All	All	24547	22962	22941	126	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 3.

The worst 5 of 126 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:A:1739:HIS:ND1	1:A:1784:ASP:OD2	2.11	0.83
1:C:1926:ASP:O	1:C:1929:SER:OG	2.06	0.71
1:A:1936:ARG:NH1	1:A:1990:GLU:OE1	2.25	0.70
1:E:1593:THR:OG1	1:E:1649:LYS:NZ	2.25	0.70
1:C:1742:GLN:HG2	1:K:1923:VAL:HG11	1.73	0.69

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	473/556 (85%)	455 (96%)	18 (4%)	0	100 100
1	C	471/556 (85%)	446 (95%)	25 (5%)	0	100 100
1	E	486/556 (87%)	455 (94%)	30 (6%)	1 (0%)	52 87

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	G	471/556 (85%)	448 (95%)	23 (5%)	0	100 100
1	I	472/556 (85%)	449 (95%)	23 (5%)	0	100 100
1	K	475/556 (85%)	448 (94%)	27 (6%)	0	100 100
2	B	48/65 (74%)	40 (83%)	8 (17%)	0	100 100
2	D	48/65 (74%)	41 (85%)	7 (15%)	0	100 100
2	F	41/65 (63%)	34 (83%)	7 (17%)	0	100 100
2	H	41/65 (63%)	33 (80%)	7 (17%)	1 (2%)	7 49
2	J	46/65 (71%)	42 (91%)	4 (9%)	0	100 100
2	L	44/65 (68%)	41 (93%)	3 (7%)	0	100 100
All	All	3116/3726 (84%)	2932 (94%)	182 (6%)	2 (0%)	56 89

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	1862	VAL
2	H	449	SER

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	384/525 (73%)	383 (100%)	1 (0%)	94 98
1	C	384/525 (73%)	381 (99%)	3 (1%)	86 95
1	E	383/525 (73%)	381 (100%)	2 (0%)	92 97
1	G	365/525 (70%)	363 (100%)	2 (0%)	92 97
1	I	381/525 (73%)	375 (98%)	6 (2%)	70 89
1	K	369/525 (70%)	364 (99%)	5 (1%)	74 90
2	B	34/59 (58%)	33 (97%)	1 (3%)	50 81
2	D	35/59 (59%)	34 (97%)	1 (3%)	50 81
2	F	30/59 (51%)	30 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	H	28/59 (48%)	28 (100%)	0	100 100
2	J	31/59 (52%)	31 (100%)	0	100 100
2	L	27/59 (46%)	25 (93%)	2 (7%)	17 56
All	All	2451/3504 (70%)	2428 (99%)	23 (1%)	84 93

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	1582	LEU
1	I	1587	GLU
2	L	443	ARG
1	I	1583	LEU
1	I	1685	THR

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

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	489/556 (87%)	0.10	8 (1%) 74 61	74, 114, 155, 182	0
1	C	487/556 (87%)	0.18	19 (3%) 43 30	89, 122, 160, 193	0
1	E	496/556 (89%)	0.07	9 (1%) 71 57	79, 121, 155, 181	0
1	G	489/556 (87%)	0.19	22 (4%) 37 26	84, 126, 163, 193	0
1	I	490/556 (88%)	0.12	17 (3%) 48 34	81, 123, 159, 195	0
1	K	489/556 (87%)	0.21	22 (4%) 37 26	85, 128, 161, 190	0
2	B	50/65 (76%)	-0.11	0 100 100	105, 142, 166, 173	0
2	D	50/65 (76%)	0.15	4 (8%) 15 10	117, 148, 173, 195	0
2	F	43/65 (66%)	-0.23	0 100 100	103, 138, 168, 180	0
2	H	45/65 (69%)	-0.03	1 (2%) 65 50	115, 149, 174, 184	0
2	J	48/65 (73%)	0.02	2 (4%) 40 28	109, 138, 165, 182	0
2	L	46/65 (70%)	0.11	2 (4%) 39 27	109, 139, 167, 172	0
All	All	3222/3726 (86%)	0.13	106 (3%) 50 36	74, 124, 162, 195	0

The worst 5 of 106 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	1630	SER	6.0
2	L	424	GLY	5.9
1	C	1636	VAL	5.7
1	G	1635	GLU	4.4
1	I	1743	PRO	4.2

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 (i)

There are no carbohydrates 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.