Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 09:48 PM GMT

PDB ID	:	4WAF
Title	:	Crystal Structure of a novel tetrahydropyrazolo[1,5-a]pyrazine in an engineered PI3K alpha
Authors	:	Knapp, M.S.; Elling, R.A.
Deposited on	:	2014-08-29
Resolution	:	2.39 Å(reported)

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

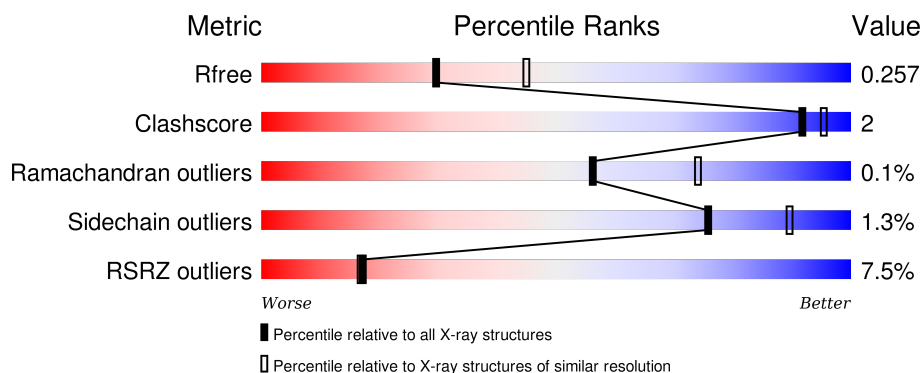|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.39 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		0.257
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		1.3%
RSRZ outliers		7.5%

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$_{free}$	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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	1073	5% / 88% / 7% / 5%
2	B	323	13% / 74% / • / 23%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 10388 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 Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1017	8094	5174	1379	1476	65	0	3	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	232	LYS	MET	engineered mutation	UNP P42336
A	233	LYS	LEU	engineered mutation	UNP P42336
A	800	MET	ILE	engineered mutation	UNP P42336
A	930	VAL	PHE	engineered mutation	UNP P42336
A	1069	HIS	-	expression tag	UNP P42336
A	1070	HIS	-	expression tag	UNP P42336
A	1071	HIS	-	expression tag	UNP P42336
A	1072	HIS	-	expression tag	UNP P42336
A	1073	HIS	-	expression tag	UNP P42336
A	1074	HIS	-	expression tag	UNP P42336

- Molecule 2 is a protein called Phosphatidylinositol 3-kinase regulatory subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	250	1975	1231	353	386	5	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	295	MET	-	initiating methionine	UNP P27986
B	296	GLU	-	expression tag	UNP P27986
B	297	TYR	-	expression tag	UNP P27986
B	298	MET	-	expression tag	UNP P27986
B	299	PRO	-	expression tag	UNP P27986

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	300	MET	-	expression tag	UNP P27986
B	301	GLU	-	expression tag	UNP P27986
B	306	TYR	THR	engineered mutation	UNP P27986

- Molecule 3 is N,N-dimethyl-4-[(6R)-6-methyl-5-(1H-pyrrolo[2,3-b]pyridin-4-yl)-4,5,6,7-tetrahydropyrazolo[1,5-a]pyrazin-3-yl]benzenesulfonamide (three-letter code: 3K6) (formula: $C_{22}H_{24}N_6O_2S$).
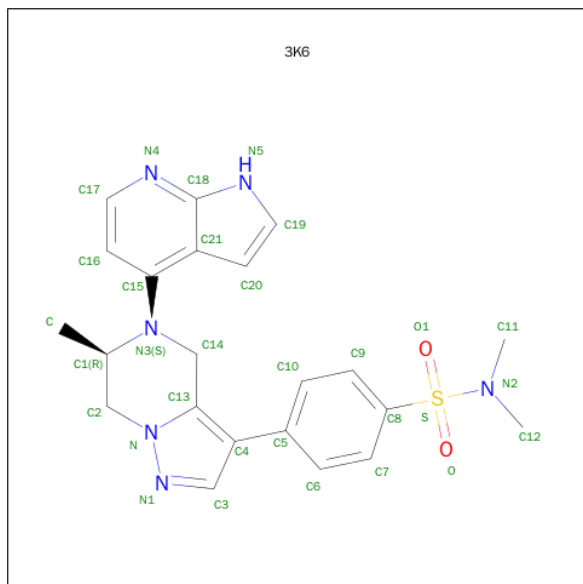
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
3	A	1	31	22	6	2	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	250	Total	O	0	0
			250	250		
4	B	38	Total	O	0	0
			38	38		

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 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: Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform
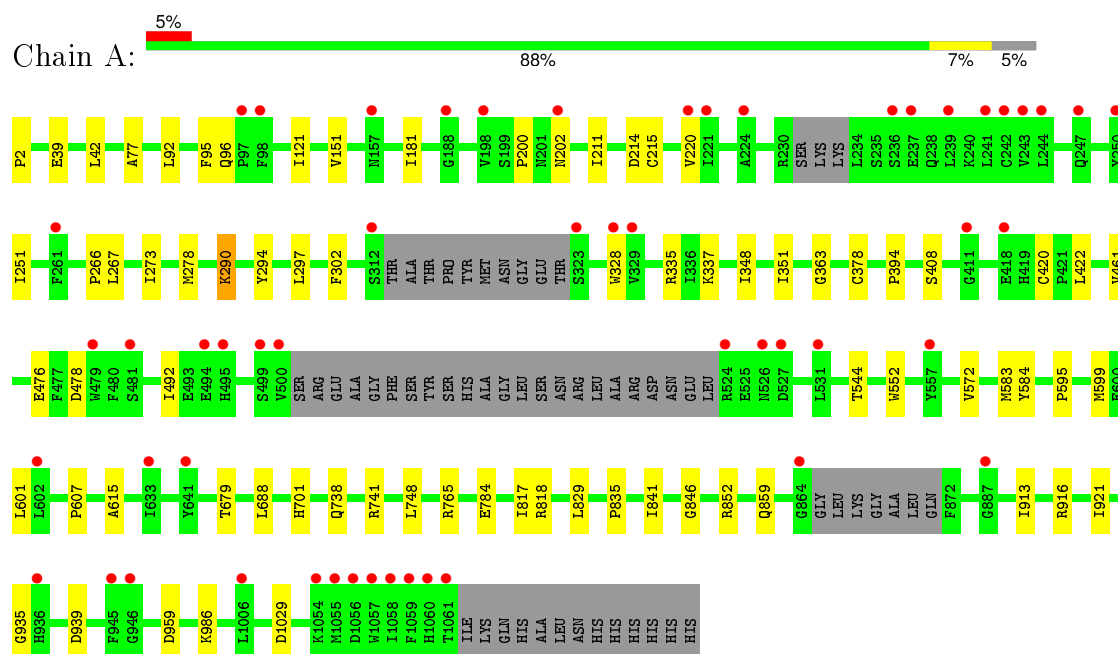
• Molecule 2: Phosphatidylinositol 3-kinase regulatory subunit alpha

Full wwPDB X-ray Structure Validation Report
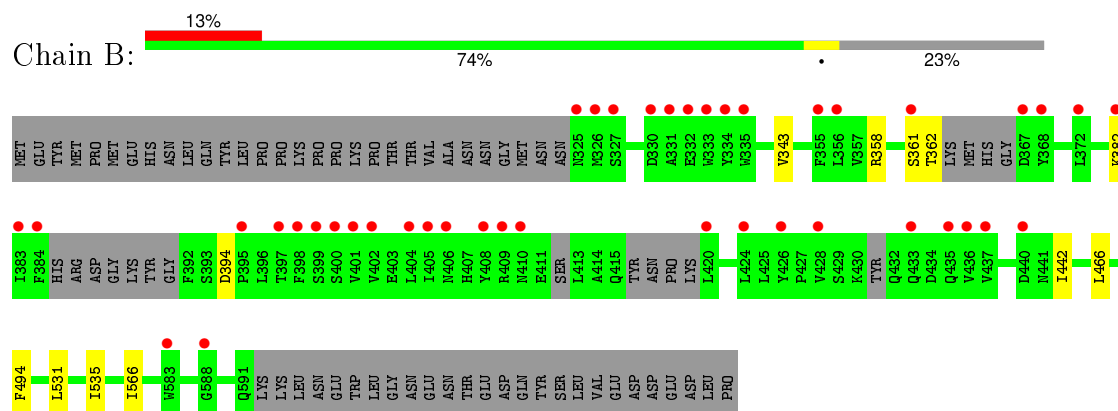
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	104.48Å 106.53Å 133.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.65 – 2.39 48.65 – 2.39	Depositor EDS
% Data completeness (in resolution range)	100.0 (48.65-2.39) 100.0 (48.65-2.39)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.23 (at 2.39Å)	Xtriage
Refinement program	BUSTER 2.11.5	Depositor
R, R_{free}	0.199 , 0.247 0.210 , 0.257	Depositor DCC
R_{free} test set	3010 reflections (5.32%)	DCC
Wilson B-factor (Å²)	53.0	Xtriage
Anisotropy	0.278	Xtriage
Bulk solvent k_{sol}(e/Å³), B_{sol}(Å²)	0.30 , 47.5	EDS
Estimated twinning fraction	0.020 for k,h,-l	Xtriage
L-test for twinning [2]	$< |L| > = 0.48, < L^2 > = 0.32$	Xtriage
Outliers	0 of 59621 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10388	wwPDB-VP
Average B, all atoms (Å²)	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% 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: 3K6

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	0/8279	0.64	0/11232
2	B	0.44	0/2000	0.61	0/2693
All	All	0.45	0/10279	0.63	0/13925

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	8094	0	7872	36	0
2	B	1975	0	1818	6	0
3	A	31	0	24	1	0
4	A	250	0	0	1	0
4	B	38	0	0	1	0
All	All	10388	0	9714	41	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 2.

All (41) 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:121:ILE:HG12	1:A:688:LEU:HB3	1.66	0.75
1:A:181:ILE:HD12	1:A:278:MET:HG2	1.68	0.75
1:A:294:TYR:HA	1:A:297:LEU:HD12	1.76	0.66
1:A:42:LEU:HD21	1:A:92:LEU:HD11	1.80	0.64
2:B:494:PHE:HB3	2:B:535:ILE:HG12	1.79	0.63
1:A:572:VAL:HG21	1:A:583:MET:HG2	1.84	0.60
1:A:351:ILE:HD13	1:A:408:SER:HB2	1.82	0.59
1:A:829:LEU:HD11	1:A:986:LYS:HB3	1.85	0.58
1:A:738:GLN:HB3	1:A:741:ARG:HH21	1.70	0.56
1:A:214:ASP:HA	1:A:266:PRO:HB3	1.87	0.55
1:A:916:ARG:HB3	1:A:921:ILE:HD11	1.91	0.53
1:A:211:ILE:HG12	1:A:215:CYS:SG	2.49	0.52
1:A:267:LEU:HG	1:A:273:ILE:HG13	1.91	0.52
1:A:337:LYS:HB3	1:A:476:GLU:HB3	1.92	0.52
1:A:220:VAL:HG21	1:A:267:LEU:HD13	1.91	0.51
1:A:841:ILE:HG12	1:A:846:GLY:HA2	1.94	0.50
2:B:531:LEU:O	2:B:535:ILE:HG13	2.12	0.49
1:A:151:VAL:HG21	1:A:302:PHE:HB2	1.94	0.49
1:A:420:CYS:HB3	4:B:726:HOH:O	2.12	0.48
1:A:552:TRP:HZ3	1:A:583:MET:HE2	1.77	0.48
1:A:348:ILE:HG22	1:A:378:CYS:HB2	1.98	0.46
1:A:817:ILE:HG22	1:A:835:PRO:HG3	1.97	0.45
2:B:466:LEU:HD12	2:B:566:ILE:HD12	1.97	0.45
1:A:2:PRO:HD2	1:A:77:ALA:HB3	2.00	0.44
3:A:1101:3K6:H1	3:A:1101:3K6:C20	2.48	0.43
1:A:95:PHE:CD2	1:A:96:GLN:HG2	2.53	0.43
1:A:765:ARG:NH1	1:A:784:GLU:HG2	2.33	0.43
1:A:492:ILE:HG21	1:A:584:TYR:HD2	1.84	0.43
1:A:701:HIS:HD2	4:A:1443:HOH:O	2.01	0.43
1:A:461:VAL:HG21	1:A:679:THR:HB	2.01	0.42
1:A:290:LYS:HG2	1:A:294:TYR:CE2	2.55	0.42
1:A:544:THR:HG21	2:B:382:LYS:HB2	2.02	0.42
1:A:601:LEU:HB2	1:A:615:ALA:HB2	2.03	0.41
1:A:363:GLY:N	1:A:607:PRO:HG3	2.35	0.41
1:A:1029:ASP:HB3	2:B:361:SER:HB2	2.02	0.41
1:A:818:ARG:HG2	1:A:835:PRO:HG2	2.03	0.41
1:A:408:SER:HB3	1:A:422:LEU:HD21	2.03	0.41
1:A:200:PRO:C	1:A:202:ASN:H	2.25	0.41
1:A:595:PRO:O	1:A:599:MET:HG3	2.20	0.41
2:B:343:VAL:HG21	2:B:358:ARG:HE	1.86	0.40
1:A:328:TRP:HA	1:A:394:PRO:HB3	2.04	0.40

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	1010/1073 (94%)	982 (97%)	27 (3%)	1 (0%)	56	74
2	B	238/323 (74%)	229 (96%)	9 (4%)	0	100	100
All	All	1248/1396 (89%)	1211 (97%)	36 (3%)	1 (0%)	56	74

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	935	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	875/979 (89%)	864 (99%)	11 (1%)	76	89
2	B	195/301 (65%)	192 (98%)	3 (2%)	72	87
All	All	1070/1280 (84%)	1056 (99%)	14 (1%)	76	89

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	39	GLU
1	A	251	ILE
1	A	290	LYS
1	A	335	ARG
1	A	478	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	748	LEU
1	A	852	ARG
1	A	859	GLN
1	A	913	ILE
1	A	939	ASP
1	A	959	ASP
2	B	362	THR
2	B	394	ASP
2	B	442	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	530	GLN

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 ⓘ

1 ligand is 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	3K6	A	1101	-	29,35,35	0.64	0	35,53,53	0.76	1 (2%)

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	3K6	A	1101	-	-	0/20/32/32	0/4/5/5

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	1101	3K6	C4-C13-N	-3.22	106.83	110.18

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1101	3K6	1	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	1017/1073 (94%)	0.30	53 (5%)	31 31	34, 59, 99, 138	0
2	B	250/323 (77%)	0.71	42 (16%)	2 2	41, 73, 129, 142	0
All	All	1267/1396 (90%)	0.38	95 (7%)	17 17	34, 61, 112, 142	0

All (95) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1059	PHE	7.5
2	B	401	VAL	6.8
1	A	1056	ASP	6.5
1	A	242	CYS	5.4
1	A	243	VAL	5.3
2	B	325	ASN	5.2
2	B	333	TRP	4.8
2	B	400	SER	4.5
2	B	436	VAL	4.3
1	A	524	ARG	4.2
2	B	399	SER	4.2
2	B	326	MET	4.1
1	A	239	LEU	4.1
1	A	1058	ILE	4.1
2	B	331	ALA	4.0
1	A	1057	TRP	4.0
2	B	398	PHE	4.0
2	B	383	ILE	3.9
2	B	428	VAL	3.9
2	B	424	LEU	3.8
2	B	335	TRP	3.7
2	B	368	TYR	3.7
2	B	408	TYR	3.7
1	A	557	TYR	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	236	SER	3.6
1	A	499	SER	3.6
2	B	405	ILE	3.5
2	B	402	VAL	3.5
1	A	1055	MET	3.5
2	B	583	TRP	3.4
1	A	323	SER	3.3
2	B	372	LEU	3.3
1	A	247	GLN	3.3
2	B	327	SER	3.2
1	A	526	ASN	3.2
1	A	1060	HIS	3.2
1	A	241	LEU	3.2
1	A	494	GLU	3.1
2	B	404	LEU	3.1
1	A	479	TRP	3.0
2	B	440	ASP	3.0
1	A	244	LEU	3.0
1	A	481	SER	2.9
2	B	332	GLU	2.9
1	A	188	GLY	2.9
1	A	1061	THR	2.9
1	A	1054	LYS	2.8
1	A	312	SER	2.8
1	A	495	HIS	2.8
2	B	410	ASN	2.8
2	B	355	PHE	2.8
2	B	588	GLY	2.8
1	A	202	ASN	2.8
2	B	437	VAL	2.8
2	B	334	TYR	2.7
2	B	409	ARG	2.7
2	B	395	PRO	2.7
1	A	328	TRP	2.7
2	B	426	TYR	2.6
1	A	887	GLY	2.6
1	A	224	ALA	2.6
2	B	420	LEU	2.6
2	B	382	LYS	2.6
2	B	330	ASP	2.5
1	A	641	TYR	2.5
1	A	527	ASP	2.5

Continued on next page...

Full wwPDB X-ray Structure Validation Report

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	435	GLN	2.5
1	A	946	GLY	2.4
1	A	500	VAL	2.4
1	A	250	TYR	2.4
1	A	98	PHE	2.4
1	A	198	VAL	2.3
1	A	157	ASN	2.3
2	B	367	ASP	2.3
2	B	356	LEU	2.3
1	A	237	GLU	2.3
2	B	361	SER	2.3
1	A	864	GLY	2.2
2	B	433	GLN	2.2
2	B	406	ASN	2.2
1	A	418	GLU	2.2
1	A	633	ILE	2.2
1	A	220	VAL	2.2
2	B	397	THR	2.2
1	A	221	ILE	2.1
1	A	531	LEU	2.1
1	A	97	PRO	2.1
1	A	936	HIS	2.1
1	A	411	GLY	2.1
1	A	329	VAL	2.1
1	A	1006	LEU	2.1
1	A	945	PHE	2.0
2	B	384	PHE	2.0
1	A	602	LEU	2.0
1	A	261	PHE	2.0

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	3K6	A	1101	31/31	0.94	0.16	0.31	41,45,76,78	31

6.5 Other polymers (i)

There are no such residues in this entry.