Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 05:50 PM GMT

PDB ID	:	4JNY
Title	:	Crystal structure of PutA86-630 mutant D370A complexed with L-Tetrahydro-2-furoic acid
Authors	:	Tanner, J.J.
Deposited on	:	2013-03-16
Resolution	:	1.90 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

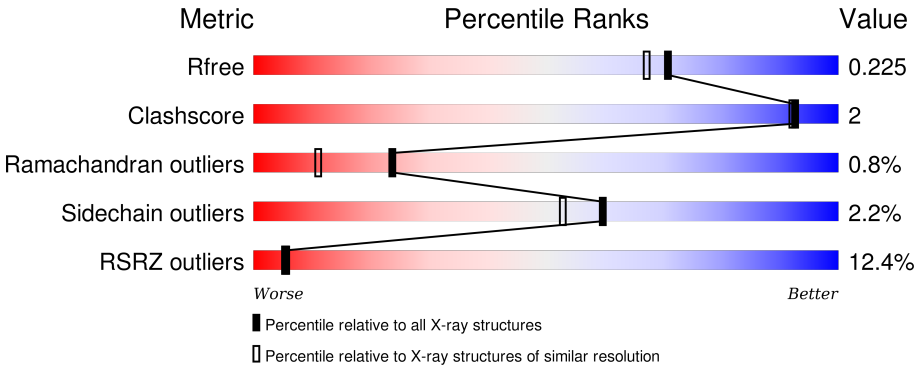
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 1.90 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	4755 (1.90-1.90)
Clashscore	102246	5398 (1.90-1.90)
Ramachandran outliers	100387	5338 (1.90-1.90)
Sidechain outliers	100360	5339 (1.90-1.90)
RSRZ outliers	91569	4766 (1.90-1.90)

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	602	10% 76% 5% 18%

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	1PE	A	2004	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4033 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.

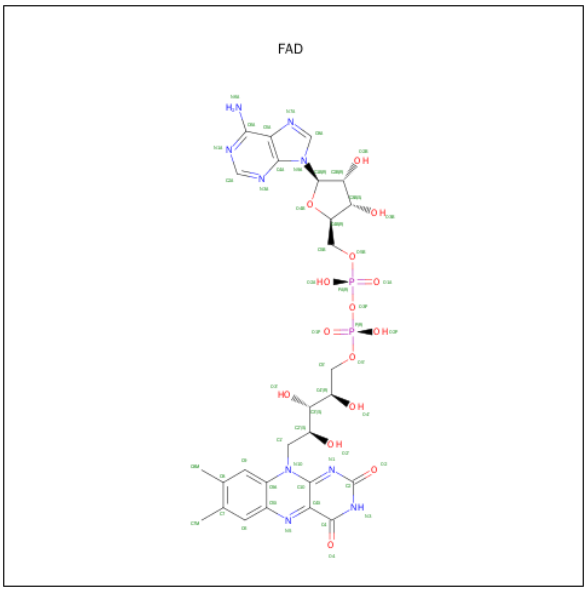
- Molecule 1 is a protein called Bifunctional protein PutA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	491	Total	C	N	O	S	0	3	0
			3732	2369	655	690	18			

There are 19 discrepancies between the modelled and reference sequences:

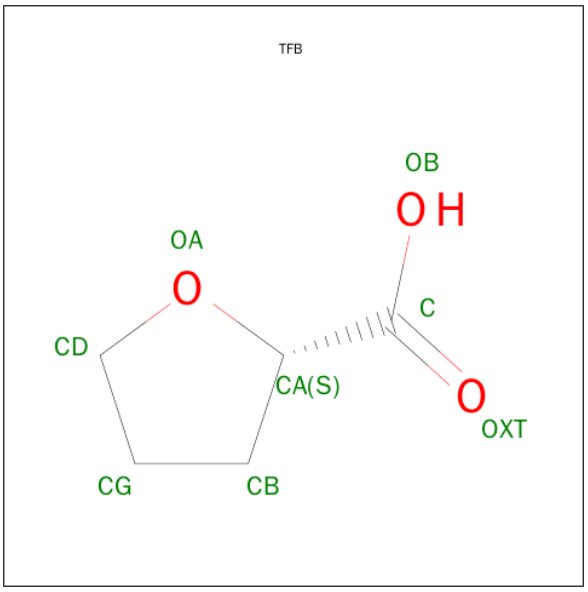
Chain	Residue	Modelled	Actual	Comment	Reference
A	370	ALA	ASP	ENGINEERED MUTATION	UNP P09546
A	670	SER	-	EXPRESSION TAG	UNP P09546
A	671	SER	-	EXPRESSION TAG	UNP P09546
A	672	SER	-	EXPRESSION TAG	UNP P09546
A	673	VAL	-	EXPRESSION TAG	UNP P09546
A	674	ASP	-	EXPRESSION TAG	UNP P09546
A	675	LYS	-	EXPRESSION TAG	UNP P09546
A	676	LEU	-	EXPRESSION TAG	UNP P09546
A	677	ALA	-	EXPRESSION TAG	UNP P09546
A	678	ALA	-	EXPRESSION TAG	UNP P09546
A	679	ALA	-	EXPRESSION TAG	UNP P09546
A	680	LEU	-	EXPRESSION TAG	UNP P09546
A	681	GLU	-	EXPRESSION TAG	UNP P09546
A	682	HIS	-	EXPRESSION TAG	UNP P09546
A	683	HIS	-	EXPRESSION TAG	UNP P09546
A	684	HIS	-	EXPRESSION TAG	UNP P09546
A	685	HIS	-	EXPRESSION TAG	UNP P09546
A	686	HIS	-	EXPRESSION TAG	UNP P09546
A	687	HIS	-	EXPRESSION TAG	UNP P09546

- Molecule 2 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula: $C_{27}H_{33}N_9O_{15}P_2$).



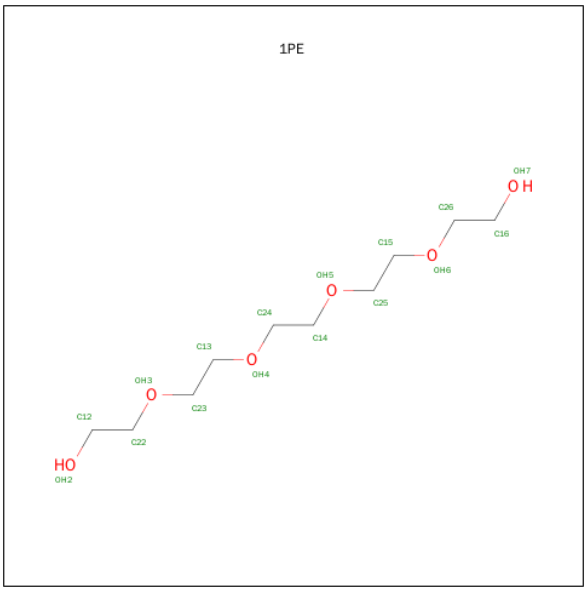
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			53	27	9	15	2		

- Molecule 3 is TETRAHYDROFURAN-2-CARBOXYLIC ACID (three-letter code: TFB) (formula: $C_5H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			8	5	3		

- Molecule 4 is PENTAETHYLENE GLYCOL (three-letter code: 1PE) (formula: $C_{10}H_{22}O_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 16 10 6	0	0
4	A	1	Total C O 10 6 4	0	0

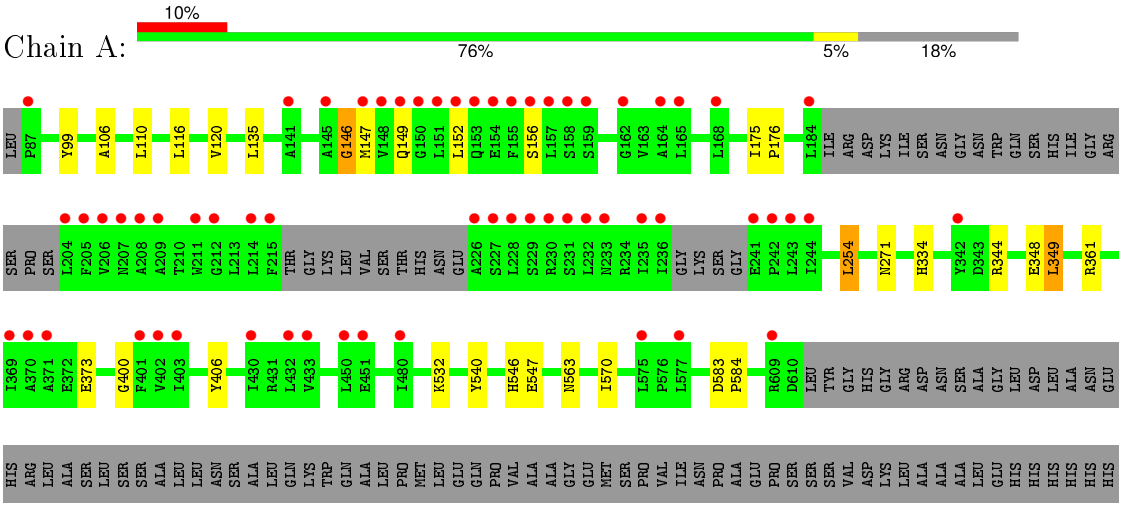
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	214	Total O 214 214	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: Bifunctional protein PutA



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	73.33Å 142.32Å 146.59Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.83 – 1.90 39.83 – 1.89	Depositor EDS
% Data completeness (in resolution range)	99.2 (39.83-1.90) 99.3 (39.83-1.89)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.43 (at 1.89Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.190 , 0.219 0.196 , 0.225	Depositor DCC
R_{free} test set	3061 reflections (5.32%)	DCC
Wilson B-factor (Å^2)	30.8	Xtriage
Anisotropy	0.623	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 52.2	EDS
Estimated twinning fraction	0.003 for -h,-l,-k	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Outliers	0 of 60902 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4033	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.36% 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

Bond lengths and bond angles in the following residue types are not validated in this section: 1PE, FAD, TFB

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	0.34	0/3809	0.50	0/5172

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	3732	0	3651	14	0
2	A	53	0	31	2	0
3	A	8	0	7	0	0
4	A	26	0	34	0	0
5	A	214	0	0	0	0
All	All	4033	0	3723	15	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 (15) 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:344:ARG:NH2	1:A:348:GLU:OE1	2.40	0.54
1:A:334:HIS:HB2	1:A:349:LEU:HG	1.90	0.53
1:A:540:TYR:HB2	2:A:2001:FAD:HM72	1.91	0.52
1:A:116:LEU:HD23	1:A:120:VAL:HG12	1.95	0.48
1:A:546:HIS:CE1	1:A:547:GLU:HG3	2.50	0.47
1:A:146:GLY:O	1:A:149:GLN:N	2.48	0.45
1:A:106:ALA:O	1:A:110:LEU:HG	2.17	0.45
2:A:2001:FAD:H4'	2:A:2001:FAD:H1'1	1.83	0.44
1:A:175:ILE:HA	1:A:176:PRO:HD3	1.86	0.44
1:A:135:LEU:HD23	1:A:570:ILE:HD13	1.99	0.43
1:A:135:LEU:HA	1:A:135:LEU:HD12	1.83	0.43
1:A:583:ASP:HA	1:A:584:PRO:HD3	1.91	0.41
1:A:373:GLU:H	1:A:373:GLU:CD	2.23	0.41
1:A:254:LEU:HA	1:A:254:LEU:HD12	1.94	0.41

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	486/602 (81%)	470 (97%)	12 (2%)	4 (1%)	24 11

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	147	MET
1	A	152	LEU
1	A	156	SER
1	A	146	GLY

5.3.2 Protein sidechains

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	362/489 (74%)	354 (98%)	8 (2%)	60 53

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	99	TYR
1	A	254	LEU
1	A	271	ASN
1	A	349	LEU
1	A	361	ARG
1	A	406	TYR
1	A	532	LYS
1	A	563	ASN

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	298	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 (i)

4 ligands are 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	$\#\|Z\| > 2$
2	FAD	A	2001	-	48,58,58	1.93	13 (27%)	54,89,89	2.42	12 (22%)
3	TFB	A	2002	-	5,8,8	2.27	1 (20%)	2,10,10	0.46	0
4	1PE	A	2003	-	15,15,15	0.54	0	14,14,14	1.47	0
4	1PE	A	2004	-	9,9,15	0.51	0	8,8,14	1.45	0

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
2	FAD	A	2001	-	-	0/30/50/50	0/6/6/6
3	TFB	A	2002	-	-	0/0/11/11	0/1/1/1
4	1PE	A	2003	-	-	0/13/13/13	0/0/0/0
4	1PE	A	2004	-	-	0/7/7/13	0/0/0/0

All (14) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	2002	TFB	CB-CA	-4.59	1.46	1.53
2	A	2001	FAD	O2'-C2'	-2.88	1.36	1.43
2	A	2001	FAD	O4B-C4B	-2.42	1.39	1.45
2	A	2001	FAD	PA-O5B	-2.21	1.48	1.59
2	A	2001	FAD	PA-O2A	-2.20	1.45	1.54
2	A	2001	FAD	C4X-C10	2.02	1.44	1.41
2	A	2001	FAD	C2A-N3A	2.50	1.36	1.32
2	A	2001	FAD	C4-N3	2.92	1.38	1.33
2	A	2001	FAD	C5X-N5	3.36	1.40	1.35
2	A	2001	FAD	C9A-N10	3.47	1.43	1.38
2	A	2001	FAD	C6A-N6A	3.56	1.45	1.34
2	A	2001	FAD	C4X-N5	4.25	1.40	1.33

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2001	FAD	C4-C4X	4.33	1.49	1.41
2	A	2001	FAD	O4-C4	5.02	1.36	1.24

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	2001	FAD	N3A-C2A-N1A	-12.61	119.24	128.89
2	A	2001	FAD	C4-C4X-C10	-5.76	116.26	119.94
2	A	2001	FAD	C4A-C5A-N7A	-2.25	107.41	109.48
2	A	2001	FAD	O2P-P-O5'	-2.00	98.35	108.46
2	A	2001	FAD	C4X-C4-N3	-2.00	120.85	123.59
2	A	2001	FAD	C2A-N1A-C6A	2.17	122.64	118.77
2	A	2001	FAD	C2B-C1B-N9A	2.26	117.74	114.29
2	A	2001	FAD	O5'-P-O1P	2.27	118.44	109.62
2	A	2001	FAD	C4X-N5-C5X	2.38	119.50	116.76
2	A	2001	FAD	C5X-C9A-N10	3.08	119.96	117.62
2	A	2001	FAD	C4-C4X-N5	3.85	123.40	118.72
2	A	2001	FAD	C4-N3-C2	5.77	120.24	115.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	2001	FAD	2	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

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	491/602 (81%)	0.57	61 (12%) 5 5	22, 39, 88, 106	0

All (61) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	155	PHE	8.8
1	A	205	PHE	8.3
1	A	228	LEU	6.9
1	A	232	LEU	6.4
1	A	244	ILE	6.3
1	A	204	LEU	6.2
1	A	235	ILE	6.1
1	A	151	LEU	5.9
1	A	229	SER	5.7
1	A	152	LEU	5.6
1	A	230	ARG	5.3
1	A	209	ALA	5.1
1	A	342	TYR	4.4
1	A	402	VAL	4.4
1	A	206	VAL	4.3
1	A	162	GLY	4.2
1	A	403	ILE	4.1
1	A	233	ASN	4.1
1	A	226	ALA	4.1
1	A	159	SER	4.0
1	A	242	PRO	4.0
1	A	609	ARG	4.0
1	A	215	PHE	3.9
1	A	243	LEU	3.9
1	A	148	VAL	3.8
1	A	208	ALA	3.7
1	A	145	ALA	3.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	450	LEU	3.5
1	A	157	LEU	3.4
1	A	87	PRO	3.3
1	A	153	GLN	3.2
1	A	147	MET	3.2
1	A	231	SER	3.1
1	A	164	ALA	3.0
1	A	158	SER	3.0
1	A	236	ILE	2.9
1	A	165	LEU	2.9
1	A	154	GLU	2.8
1	A	150	GLY	2.8
1	A	211	TRP	2.8
1	A	451	GLU	2.8
1	A	214	LEU	2.7
1	A	432	LEU	2.6
1	A	369	ILE	2.6
1	A	370	ALA	2.6
1	A	207	ASN	2.5
1	A	433	VAL	2.5
1	A	241	GLU	2.5
1	A	401	PHE	2.5
1	A	575	LEU	2.5
1	A	227	SER	2.4
1	A	168	LEU	2.3
1	A	184	LEU	2.2
1	A	480	ILE	2.2
1	A	577	LEU	2.2
1	A	141	ALA	2.2
1	A	156	SER	2.2
1	A	430	ILE	2.1
1	A	212	GLY	2.1
1	A	149	GLN	2.1
1	A	371	ALA	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, 95^{th} 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
4	1PE	A	2004	10/16	0.85	0.16	3.14	61,64,67,67	0
3	TFB	A	2002	8/8	0.98	0.23	1.81	26,29,30,31	0
2	FAD	A	2001	53/53	0.96	0.15	0.06	22,26,28,30	0
4	1PE	A	2003	16/16	0.85	0.14	0.04	56,67,71,75	0

6.5 Other polymers (i)

There are no such residues in this entry.