



Full wwPDB X-ray Structure Validation Report

Mar 8, 2018 – 11:39 pm GMT

PDB ID : 2ITT
Title : Crystal structure of EGFR kinase domain L858R mutation in complex with AEE788
Authors : Yun, C.-H.; Boggon, T.J.; Li, Y.; Woo, S.; Greulich, H.; Meyerson, M.; Eck, M.J.
Deposited on : 2006-05-25
Resolution : 2.73 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the  symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

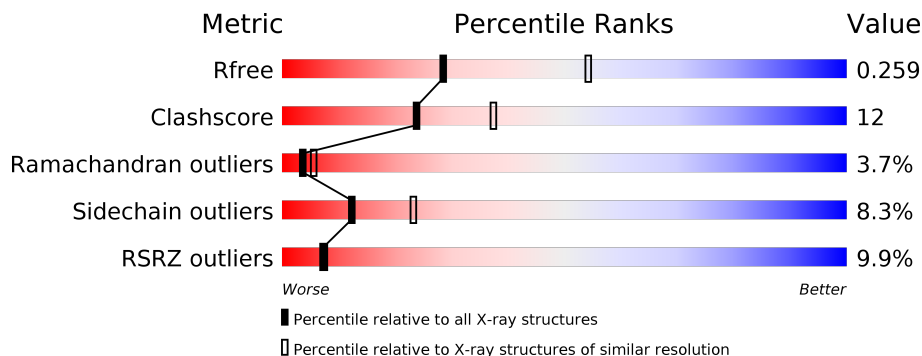
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.73 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}	111664	1033 (2.76-2.72)
Clashscore	122126	1084 (2.76-2.72)
Ramachandran outliers	120053	1064 (2.76-2.72)
Sidechain outliers	120020	1065 (2.76-2.72)
RSRZ outliers	108989	1008 (2.76-2.72)

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 $\leq 5\%$. 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	327	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2526 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.

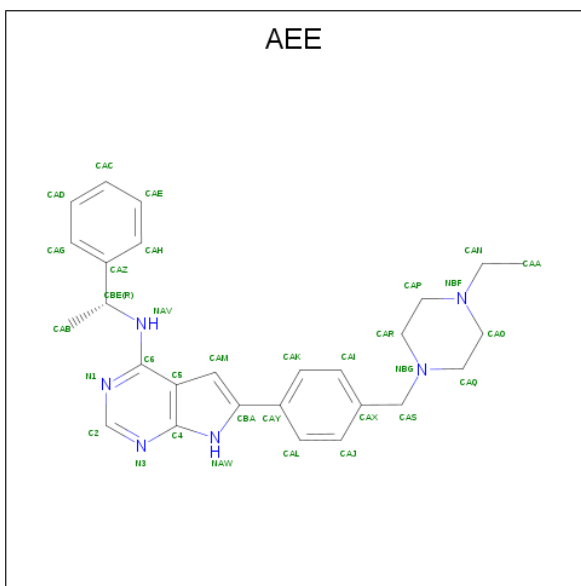
- Molecule 1 is a protein called EPIDERMAL GROWTH FACTOR RECEPTOR.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	302	2419	1552	407	441	19	0	1	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	858	ARG	LEU	engineered mutation	UNP P00533

- Molecule 2 is 6-{4-[(4-ETHYLPIPERAZIN-1-YL)METHYL]PHENYL}-N-[(1R)-1-PHENYLETHYL]-7H-PYRROLO[2,3-D]PYRIMIDIN-4-AMINE (three-letter code: AEE) (formula: C₂₇H₃₂N₆).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	N		
2	A	1	33	27	6	0	0

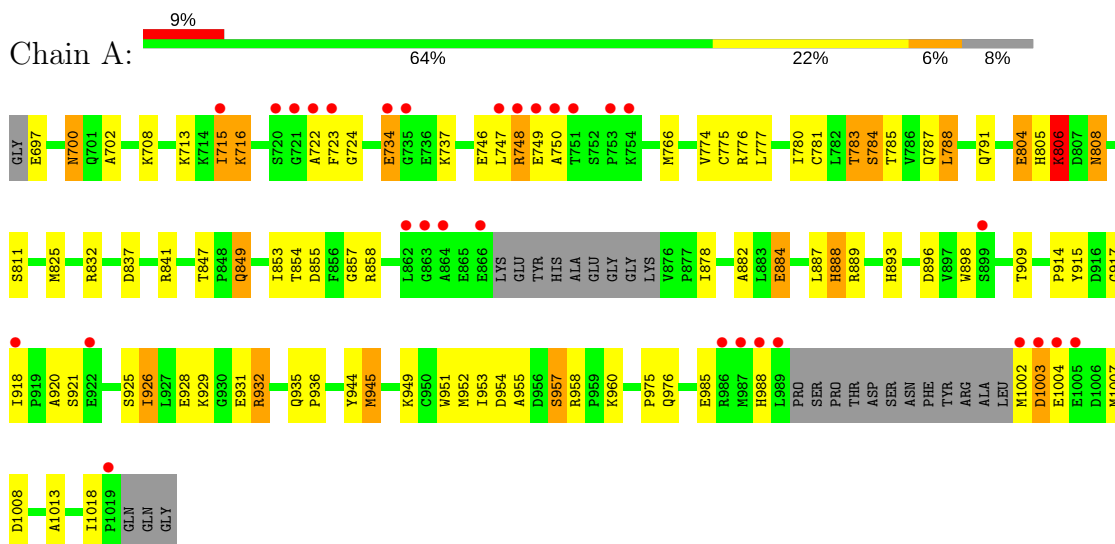
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	74	Total	O	0	0
			74	74		

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 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: EPIDERMAL GROWTH FACTOR RECEPTOR



4 Data and refinement statistics i

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, α , β , γ	145.72Å 145.72Å 145.72Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.99 – 2.73 24.99 – 2.73	Depositor EDS
% Data completeness (in resolution range)	99.6 (24.99-2.73) 99.6 (24.99-2.73)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.99 (at 2.72Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.208 , 0.262 0.205 , 0.259	Depositor DCC
R_{free} test set	987 reflections (7.17%)	wwPDB-VP
Wilson B-factor (Å ²)	61.4	Xtrriage
Anisotropy	0.000	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 62.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	0.024 for -l,-k,-h	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	2526	wwPDB-VP
Average B, all atoms (Å ²)	57.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.85% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: AEE

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	0.87	1/2468 (0.0%)	0.87	3/3339 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	748	ARG	CZ-NH1	5.51	1.40	1.33

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	748	ARG	NE-CZ-NH2	-5.96	117.32	120.30
1	A	932	ARG	NE-CZ-NH1	-5.81	117.39	120.30
1	A	784	SER	N-CA-C	-5.17	97.04	111.00

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	2419	0	2453	59	0
2	A	33	0	32	3	0
3	A	74	0	0	4	0
All	All	2526	0	2485	59	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 12.

All (59) 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:775[B]:CYS:SG	3:A:3041:HOH:O	2.11	1.07
1:A:887:LEU:HB3	1:A:888:HIS:ND1	2.10	0.67
1:A:737:LYS:HG3	1:A:737:LYS:O	1.96	0.64
1:A:700:ASN:C	1:A:700:ASN:HD22	2.00	0.63
1:A:776:ARG:HB3	3:A:3018:HOH:O	1.98	0.63
1:A:918:ILE:HG21	1:A:926:ILE:HD12	1.82	0.62
1:A:882:ALA:HB1	1:A:884:GLU:OE2	2.00	0.61
1:A:1002:MET:HG3	3:A:3067:HOH:O	2.01	0.61
1:A:766:MET:HB3	1:A:777:LEU:HB2	1.83	0.61
1:A:766:MET:HE3	2:A:2020:AEE:HAE	1.84	0.59
1:A:708:LYS:NZ	1:A:734:GLU:OE2	2.34	0.56
1:A:766:MET:CE	2:A:2020:AEE:HAE	2.36	0.56
1:A:945:MET:O	1:A:949:LYS:HG3	2.04	0.56
1:A:932:ARG:NH1	1:A:951:TRP:O	2.38	0.55
1:A:783:THR:OG1	1:A:785:THR:O	2.25	0.55
1:A:893:HIS:O	1:A:896:ASP:HB2	2.06	0.55
1:A:917:GLY:O	1:A:918:ILE:C	2.46	0.54
1:A:909:THR:HG22	1:A:936:PRO:HB3	1.88	0.54
1:A:847:THR:OG1	1:A:849:GLN:HG2	2.08	0.54
1:A:804:GLU:HG3	1:A:804:GLU:O	2.09	0.53
1:A:781:CYS:HB3	1:A:787:GLN:HE21	1.74	0.53
1:A:791:GLN:HG3	1:A:1013:ALA:HB3	1.89	0.53
1:A:746:GLU:HB2	1:A:787:GLN:HG2	1.92	0.52
1:A:837:ASP:HB2	1:A:858:ARG:HD3	1.91	0.52
1:A:887:LEU:HB3	1:A:888:HIS:CE1	2.44	0.52
1:A:715:ILE:O	1:A:716:LYS:HB2	2.10	0.52
1:A:788:LEU:HB3	2:A:2020:AEE:HAC	1.91	0.52
1:A:724:GLY:CA	1:A:748:ARG:H	2.22	0.52
1:A:780:ILE:HG22	1:A:788:LEU:HD12	1.92	0.52
1:A:1003:ASP:H	1:A:1007:MET:HE1	1.75	0.51
1:A:985:GLU:CD	1:A:985:GLU:H	2.14	0.50
1:A:700:ASN:C	1:A:700:ASN:ND2	2.64	0.50
1:A:697:GLU:N	3:A:3005:HOH:O	2.44	0.49
1:A:954:ASP:O	1:A:957:SER:HB2	2.12	0.49
1:A:952:MET:O	1:A:958:ARG:HD2	2.11	0.49
1:A:929:LYS:HB2	1:A:929:LYS:HE3	1.62	0.48
1:A:724:GLY:HA3	1:A:748:ARG:H	1.78	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:926:ILE:HG23	1:A:931:GLU:HB3	1.96	0.48
1:A:882:ALA:HA	1:A:898:TRP:CD2	2.49	0.47
1:A:697:GLU:OE2	1:A:832:ARG:NH1	2.47	0.47
1:A:808:ASN:ND2	1:A:988:HIS:O	2.49	0.46
1:A:887:LEU:C	1:A:888:HIS:ND1	2.69	0.46
1:A:935:GLN:HA	1:A:944:TYR:CD2	2.51	0.46
1:A:949:LYS:O	1:A:958:ARG:HG3	2.16	0.45
1:A:955:ALA:HA	1:A:958:ARG:HD3	1.98	0.45
1:A:854:THR:HG23	1:A:855:ASP:N	2.33	0.44
1:A:700:ASN:ND2	1:A:702:ALA:H	2.16	0.43
1:A:811:SER:OG	1:A:975:PRO:HB2	2.18	0.43
1:A:724:GLY:HA3	1:A:747:LEU:HA	2.01	0.42
1:A:805:HIS:O	1:A:806:LYS:C	2.57	0.42
1:A:887:LEU:CB	1:A:888:HIS:ND1	2.79	0.42
1:A:749:GLU:O	1:A:750:ALA:C	2.57	0.42
1:A:781:CYS:HB3	1:A:787:GLN:HB2	2.02	0.42
1:A:781:CYS:SG	1:A:783:THR:HG23	2.59	0.42
1:A:878:ILE:HG13	1:A:920:ALA:HB1	2.00	0.42
1:A:825:MET:SD	1:A:853:ILE:HD13	2.60	0.42
1:A:749:GLU:HA	1:A:749:GLU:OE2	2.20	0.41
1:A:914:PRO:HB2	1:A:915:TYR:CD2	2.56	0.40
1:A:1018:ILE:HD13	1:A:1018:ILE:HA	1.83	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	297/327 (91%)	255 (86%)	31 (10%)	11 (4%)	4 5

All (11) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	723	PHE
1	A	734	GLU
1	A	716	LYS
1	A	722	ALA
1	A	784	SER
1	A	857	GLY
1	A	806	LYS
1	A	889	ARG
1	A	953	ILE
1	A	715	ILE
1	A	926	ILE

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	266/287 (93%)	244 (92%)	22 (8%)	12 23

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	700	ASN
1	A	713	LYS
1	A	774	VAL
1	A	783	THR
1	A	788	LEU
1	A	804	GLU
1	A	806	LYS
1	A	808	ASN
1	A	841	ARG
1	A	849	GLN
1	A	884	GLU
1	A	888	HIS
1	A	921	SER
1	A	925	SER
1	A	928	GLU
1	A	945	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	957	SER
1	A	960	LYS
1	A	976	GLN
1	A	1003	ASP
1	A	1004	GLU
1	A	1008	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	700	ASN
1	A	756	ASN
1	A	791	GLN
1	A	808	ASN

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](#)

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
2	AEE	A	2020	-	36,37,37	1.59	5 (13%)	43,51,51	2.17	7 (16%)

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	AEE	A	2020	-	-	0/18/28/28	0/5/5/5

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2020	AEE	CAY-CBA	-4.23	1.42	1.48
2	A	2020	AEE	C4-N3	-2.61	1.33	1.37
2	A	2020	AEE	CAM-CBA	-2.58	1.35	1.39
2	A	2020	AEE	CAS-CAX	2.56	1.56	1.51
2	A	2020	AEE	C6-N1	3.35	1.38	1.34

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	2020	AEE	N3-C2-N1	-8.16	121.88	128.86
2	A	2020	AEE	C5-C6-N1	-5.63	117.50	121.35
2	A	2020	AEE	CAX-CAS-NBG	-2.72	107.96	113.14
2	A	2020	AEE	CAB-CBE-NAV	2.04	113.01	108.99
2	A	2020	AEE	CAR-NBG-CAQ	2.14	113.59	108.87
2	A	2020	AEE	C2-N3-C4	3.19	120.83	113.33
2	A	2020	AEE	C2-N1-C6	6.77	120.95	116.52

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	2020	AEE	3	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, 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	302/327 (92%)	0.32	30 (9%) 7 7	33, 52, 95, 113	17 (5%)

All (30) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	721	GLY	5.2
1	A	988	HIS	4.9
1	A	864	ALA	4.7
1	A	723	PHE	4.7
1	A	748	ARG	4.6
1	A	866	GLU	4.6
1	A	863	GLY	4.5
1	A	986	ARG	4.3
1	A	720	SER	4.2
1	A	1003	ASP	3.7
1	A	862	LEU	3.6
1	A	750	ALA	3.6
1	A	754	LYS	3.6
1	A	747	LEU	3.4
1	A	1005	GLU	3.3
1	A	1019	PRO	3.2
1	A	722	ALA	3.2
1	A	749	GLU	3.1
1	A	1004	GLU	2.8
1	A	751	THR	2.8
1	A	987	MET	2.7
1	A	1002	MET	2.6
1	A	753	PRO	2.6
1	A	735	GLY	2.5
1	A	715	ILE	2.4
1	A	899	SER	2.3
1	A	922	GLU	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	989	LEU	2.2
1	A	734	GLU	2.1
1	A	918	ILE	2.1

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. 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	B-factors(\AA^2)	Q<0.9
2	AEE	A	2020	33/33	0.87	0.24	57,68,89,90	0

6.5 Other polymers [i](#)

There are no such residues in this entry.