Full wwPDB X-ray Structure Validation Report ⓘ

Oct 15, 2023 – 09:37 AM EDT

PDB ID	:	7UA8
Title	:	Pfs230 D1 domain in complex with 230AL-20
Authors	:	Tang, W.K.; Tolia, N.H.
Deposited on	:	2022-03-11
Resolution	:	2.80 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

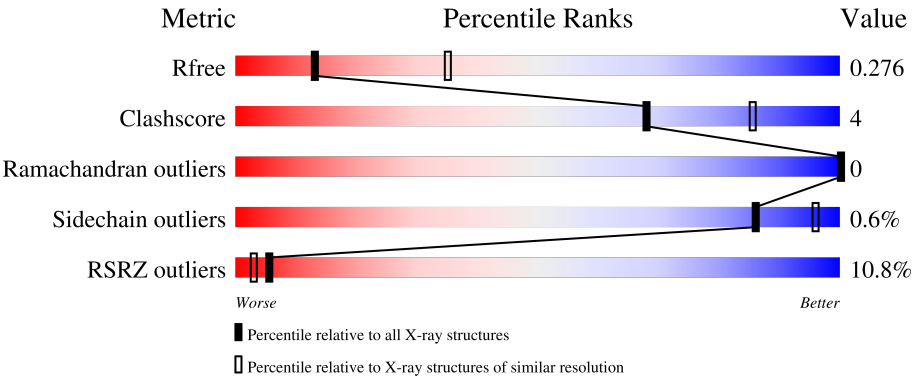
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.80 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}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	191	23% 87% 5% 8%
1	B	191	17% 84% 8% 8%
2	H	256	3% 79% 10% 12%
2	I	256	2% 76% 13% 11%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 12570 atoms, of which 6231 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 Gametocyte surface protein P230.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	176	Total	C	H	N	O	S	0	0	0
			2825	901	1418	219	283	4			
1	B	176	Total	C	H	N	O	S	0	0	0
			2825	901	1418	219	283	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	585	GLN	ASN	conflict	UNP P68874
B	585	GLN	ASN	conflict	UNP P68874

- Molecule 2 is a protein called 230Al-20.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	H	226	Total	C	H	N	O	S	0	0	0
			3453	1109	1699	296	341	8			
2	I	228	Total	C	H	N	O	S	0	0	0
			3460	1114	1696	298	344	8			

- Molecule 3 is water.

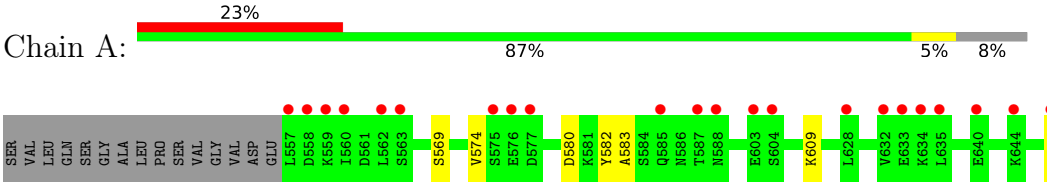
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	H	4	Total	O	0	0
			4	4		
3	I	3	Total	O	0	0
			3	3		

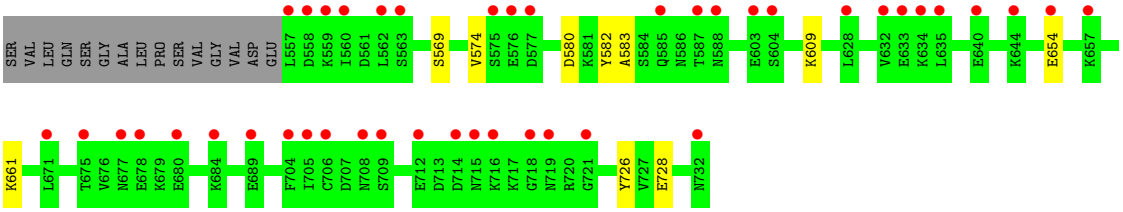
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Gametocyte surface protein P230

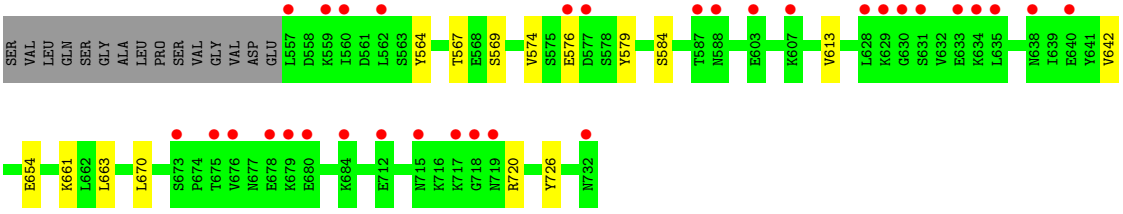
Chain A:





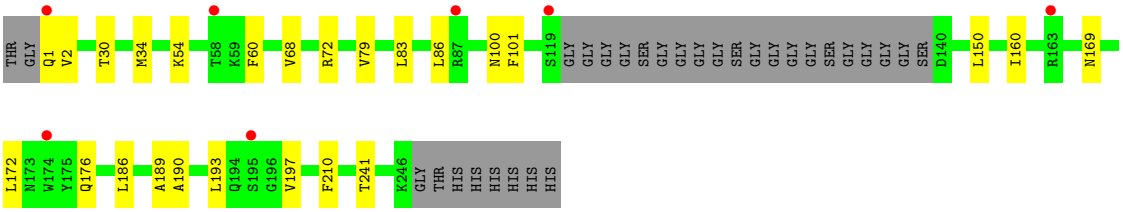
• Molecule 1: Gametocyte surface protein P230

Chain B:



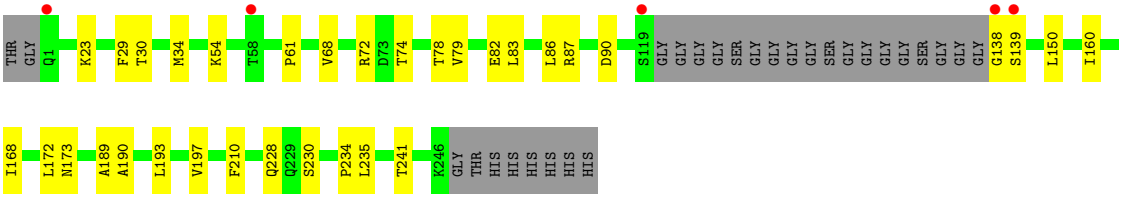
• Molecule 2: 230Al-20

Chain H:



• Molecule 2: 230Al-20

Chain I:


THR GLY Q1 K23 F29 T30 M34 K54 T58 P61 V68 R72 D73 T74 T78 V79 E82 L83 L86 R87 D90 S119 GLY GLY GLY GLY SER GLY GLY GLY GLY SER GLY GLY GLY GLY SER GLY GLY GLY G138 S139 L150 I160
I168 L172 N173 A189 A190 L193 V197 P210 Q228 Q229 S230 P234 L235 T241 K245 GLY THR HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	153.47Å 165.12Å 44.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.82 – 2.80 19.82 – 2.80	Depositor EDS
% Data completeness (in resolution range)	96.6 (19.82-2.80) 97.3 (19.82-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.99 (at 2.79Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158, PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.249 , 0.274 0.250 , 0.276	Depositor DCC
R_{free} test set	1393 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	59.6	Xtriage
Anisotropy	1.018	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 44.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	12570	wwPDB-VP
Average B, all atoms (Å^2)	73.0	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 44.41 % 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 1.5329e-04. 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, 0.333 respectively for untwinned datasets, and 0.375, 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	0.28	0/1433	0.52	0/1938
1	B	0.28	0/1433	0.54	0/1938
2	H	0.32	0/1796	0.60	0/2436
2	I	0.31	0/1806	0.60	0/2449
All	All	0.30	0/6468	0.57	0/8761

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	1407	1418	1418	8	0
1	B	1407	1418	1418	7	0
2	H	1754	1699	1701	18	0
2	I	1764	1696	1709	20	0
3	H	4	0	0	0	0
3	I	3	0	0	0	0
All	All	6339	6231	6246	48	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 4.

All (48) 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)
2:H:172:LEU:HD22	2:H:210:PHE:CG	2.23	0.73
2:I:172:LEU:HD22	2:I:210:PHE:CG	2.28	0.68
2:H:160:ILE:HG12	2:H:241:THR:HG21	1.81	0.62
2:I:160:ILE:HG12	2:I:241:THR:HG21	1.86	0.58
2:I:173:ASN:ND2	2:I:228:GLN:OE1	2.36	0.57
2:I:83:LEU:HD23	2:I:86:LEU:HD21	1.88	0.56
1:B:654:GLU:OE2	1:B:661:LYS:NZ	2.28	0.55
2:H:83:LEU:HD23	2:H:86:LEU:HD21	1.89	0.54
2:I:138:GLY:O	2:I:139:SER:HB3	2.07	0.54
2:H:1:GLN:HG3	2:H:2:VAL:H	1.73	0.53
1:B:584:SER:HB3	1:B:613:VAL:HG22	1.91	0.53
1:A:582:TYR:HA	2:H:169:ASN:ND2	2.25	0.51
2:I:168:ILE:HD11	2:I:210:PHE:CE2	2.45	0.51
1:A:583:ALA:H	2:H:169:ASN:ND2	2.07	0.51
2:H:150:LEU:HD12	2:H:150:LEU:C	2.32	0.50
2:I:61:PRO:HB3	2:I:234:PRO:HB3	1.94	0.50
1:A:728:GLU:OE2	2:H:101:PHE:N	2.41	0.50
2:H:83:LEU:HD23	2:H:86:LEU:CD2	2.43	0.49
1:A:654:GLU:OE2	1:A:661:LYS:NZ	2.33	0.49
2:I:172:LEU:HD22	2:I:210:PHE:CD2	2.48	0.48
1:A:569:SER:OG	1:A:574:VAL:O	2.22	0.48
2:I:83:LEU:HD23	2:I:86:LEU:CD2	2.44	0.48
2:H:193:LEU:HD11	2:H:197:VAL:CG1	2.44	0.47
1:B:569:SER:OG	1:B:574:VAL:O	2.30	0.47
1:B:564:TYR:OH	1:B:642:VAL:HG11	2.14	0.47
2:H:1:GLN:HG3	2:H:2:VAL:N	2.30	0.46
2:H:189:ALA:O	2:H:190:ALA:HB3	2.15	0.46
2:I:23:LYS:HG2	2:I:78:THR:OG1	2.15	0.46
2:I:189:ALA:O	2:I:190:ALA:HB3	2.16	0.46
2:I:193:LEU:HD11	2:I:197:VAL:CG1	2.46	0.45
1:B:567:THR:O	1:B:720:ARG:NH2	2.49	0.45
1:B:576:GLU:HG2	1:B:579:TYR:CE2	2.54	0.43
1:A:728:GLU:OE2	2:H:100:ASN:HA	2.19	0.43
2:I:68:VAL:HA	2:I:82:GLU:O	2.19	0.43
2:H:176:GLN:HB2	2:H:186:LEU:HD11	2.01	0.42
2:I:230:SER:HA	2:I:235:LEU:HD22	2.00	0.42
1:B:663:LEU:HG	1:B:670:LEU:HD22	2.02	0.42
2:I:87:ARG:O	2:I:90:ASP:HB2	2.19	0.42
1:A:583:ALA:H	2:H:169:ASN:HD21	1.68	0.41
2:I:68:VAL:HG23	2:I:82:GLU:O	2.20	0.41
1:A:580:ASP:HB3	1:A:609:LYS:HD3	2.02	0.41
2:H:30:THR:O	2:H:54:LYS:HG3	2.20	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:I:34:MET:HG3	2:I:79:VAL:HG21	2.03	0.41
2:H:34:MET:HG3	2:H:79:VAL:HG21	2.02	0.41
2:I:150:LEU:HD12	2:I:150:LEU:C	2.41	0.41
2:I:30:THR:O	2:I:54:LYS:HE3	2.20	0.41
2:H:60:PHE:CE1	2:H:68:VAL:HG13	2.56	0.40
2:I:29:PHE:HE2	2:I:74:THR:HG22	1.87	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	174/191 (91%)	169 (97%)	5 (3%)	0	100	100
1	B	174/191 (91%)	169 (97%)	5 (3%)	0	100	100
2	H	222/256 (87%)	211 (95%)	11 (5%)	0	100	100
2	I	224/256 (88%)	213 (95%)	11 (5%)	0	100	100
All	All	794/894 (89%)	762 (96%)	32 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	165/177 (93%)	164 (99%)	1 (1%)	86 96
1	B	165/177 (93%)	164 (99%)	1 (1%)	86 96
2	H	192/204 (94%)	191 (100%)	1 (0%)	88 96
2	I	193/204 (95%)	192 (100%)	1 (0%)	88 96
All	All	715/762 (94%)	711 (99%)	4 (1%)	86 96

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	726	TYR
2	H	72	ARG
1	B	726	TYR
2	I	72	ARG

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

Mol	Chain	Res	Type
2	H	169	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	176/191 (92%)	1.24	43 (24%)	0	0	61, 98, 140, 159	0
1	B	176/191 (92%)	1.06	32 (18%)	1	1	63, 90, 125, 145	0
2	H	226/256 (88%)	0.28	7 (3%)	49	39	46, 67, 84, 92	1 (0%)
2	I	228/256 (89%)	0.19	5 (2%)	62	52	46, 66, 87, 97	1 (0%)
All	All	806/894 (90%)	0.64	87 (10%)	5	3	46, 76, 123, 159	2 (0%)

All (87) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	715	ASN	6.3
1	A	635	LEU	5.9
1	A	603	GLU	5.6
1	B	715	ASN	5.5
1	A	634	LYS	5.3
1	A	560	ILE	5.1
1	A	557	LEU	4.9
1	A	721	GLY	4.8
1	B	560	ILE	4.7
1	A	718	GLY	4.6
1	A	714	ASP	4.6
1	B	634	LYS	4.6
1	A	604	SER	4.5
1	A	587	THR	4.5
1	B	718	GLY	4.4
2	I	119	SER	4.3
1	B	587	THR	4.3
1	B	732	ASN	4.2
2	H	119	SER	4.0
1	A	732	ASN	4.0
1	A	708	ASN	3.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	631	SER	3.8
1	A	719	ASN	3.7
1	B	628	LEU	3.6
1	B	603	GLU	3.5
1	B	629	LYS	3.5
1	B	684	LYS	3.5
1	A	559	LYS	3.4
1	B	557	LEU	3.3
1	B	633	GLU	3.2
1	B	640	GLU	3.2
1	A	633	GLU	3.2
1	B	630	GLY	3.1
1	A	562	LEU	3.1
1	A	716	LYS	3.0
2	I	138	GLY	3.0
1	A	684	LYS	3.0
1	A	689	GLU	3.0
1	A	671	LEU	3.0
1	A	678	GLU	2.9
2	I	58	THR	2.9
1	A	706	CYS	2.8
1	A	628	LEU	2.8
1	A	640	GLU	2.8
2	H	1	GLN	2.7
1	B	638	ASN	2.7
1	A	704	PHE	2.7
1	B	607	LYS	2.7
1	B	576	GLU	2.7
1	B	717	LYS	2.7
2	H	58	THR	2.6
1	B	673	SER	2.6
1	B	577	ASP	2.6
1	B	588	ASN	2.6
1	B	678	GLU	2.6
1	A	677	ASN	2.5
1	B	559	LYS	2.5
1	B	675	THR	2.5
2	H	163	ARG	2.5
2	H	195	SER	2.5
1	B	712	GLU	2.5
1	A	654	GLU	2.5
1	A	709	SER	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	558	ASP	2.4
1	A	588	ASN	2.4
1	A	632	VAL	2.4
1	A	712	GLU	2.4
1	B	635	LEU	2.4
2	H	87	ARG	2.3
1	A	705	ILE	2.3
1	B	562	LEU	2.3
2	H	174	TRP	2.3
1	A	577	ASP	2.3
1	A	675	THR	2.3
1	B	680	GLU	2.3
1	A	680	GLU	2.3
1	A	563	SER	2.2
2	I	139	SER	2.2
1	A	657	LYS	2.2
2	I	1	GLN	2.2
1	A	585	GLN	2.2
1	A	644	LYS	2.1
1	A	575	SER	2.1
1	B	676	VAL	2.1
1	A	576	GLU	2.1
1	B	679	LYS	2.1
1	B	719	ASN	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 monosaccharides 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.