



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 15, 2026 – 11:02 AM UTC

PDB ID : 4C5E / pdb_00004c5e
Title : Crystal structure of the minimal Pho-Sfmbt complex (P21 spacegroup)
Authors : Alfieri, C.; Glatt, S.; Mueller, C.W.
Deposited on : 2013-09-11
Resolution : 1.95 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

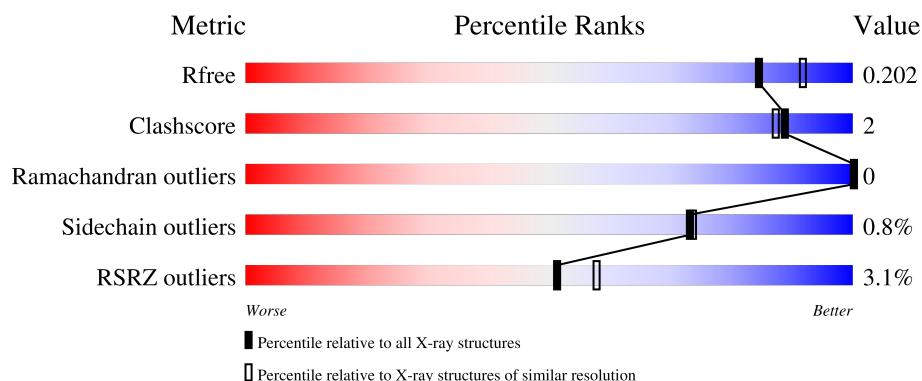
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.95 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}	180053	3494 (1.96-1.96)
Clashscore	190562	3612 (1.96-1.96)
Ramachandran outliers	187476	3587 (1.96-1.96)
Sidechain outliers	187428	3587 (1.96-1.96)
RSRZ outliers	180081	3495 (1.96-1.96)

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 $\leq 5\%$. 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	451	 2% 93% 5%
1	B	451	 2% 93%
1	C	451	 2% 91% 6%
1	D	451	 4% 92%
2	E	32	 3% 88% 6% 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	32	16% 88% 6% 6%
2	G	32	12% 94% 6%
2	H	32	12% 88% 6% 6%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 16286 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 POLYCOMB PROTEIN SFMBT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	440	Total	C	N	O	S	1	2	0
			3479	2229	578	646	26			
1	B	439	Total	C	N	O	S	0	5	0
			3480	2229	578	647	26			
1	C	438	Total	C	N	O	S	0	2	0
			3456	2216	573	641	26			
1	D	437	Total	C	N	O	S	0	2	0
			3409	2186	568	630	25			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	530	MET	-	expression tag	UNP Q9VK33
B	530	MET	-	expression tag	UNP Q9VK33
C	530	MET	-	expression tag	UNP Q9VK33
D	530	MET	-	expression tag	UNP Q9VK33

- Molecule 2 is a protein called POLYCOMB PROTEIN PHO.

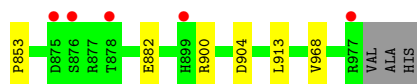
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	30	Total	C	N	O	S	0	0	0
			238	150	43	42	3			
2	F	30	Total	C	N	O	S	0	0	0
			232	147	40	42	3			
2	G	30	Total	C	N	O	S	1	0	0
			231	147	40	41	3			
2	H	30	Total	C	N	O	S	0	0	0
			233	148	40	42	3			

There are 16 discrepancies between the modelled and reference sequences:

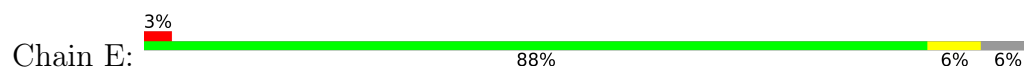
Chain	Residue	Modelled	Actual	Comment	Reference
E	141	GLY	-	expression tag	UNP Q8ST83
E	142	ALA	-	expression tag	UNP Q8ST83
E	143	MET	-	expression tag	UNP Q8ST83
E	144	ALA	-	expression tag	UNP Q8ST83
F	141	GLY	-	expression tag	UNP Q8ST83
F	142	ALA	-	expression tag	UNP Q8ST83
F	143	MET	-	expression tag	UNP Q8ST83
F	144	ALA	-	expression tag	UNP Q8ST83
G	141	GLY	-	expression tag	UNP Q8ST83
G	142	ALA	-	expression tag	UNP Q8ST83
G	143	MET	-	expression tag	UNP Q8ST83
G	144	ALA	-	expression tag	UNP Q8ST83
H	141	GLY	-	expression tag	UNP Q8ST83
H	142	ALA	-	expression tag	UNP Q8ST83
H	143	MET	-	expression tag	UNP Q8ST83
H	144	ALA	-	expression tag	UNP Q8ST83

- Molecule 3 is water.

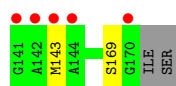
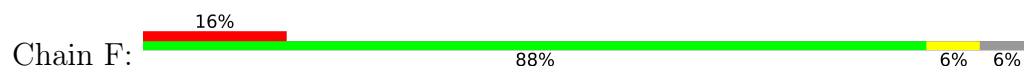
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	427	Total O 427 427	0	0
3	B	409	Total O 409 409	0	0
3	C	384	Total O 384 384	0	0
3	D	256	Total O 256 256	0	0
3	E	21	Total O 21 21	0	0
3	F	10	Total O 10 10	0	0
3	G	7	Total O 7 7	0	0
3	H	14	Total O 14 14	0	0



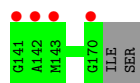
• Molecule 2: POLYCOMB PROTEIN PHO



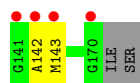
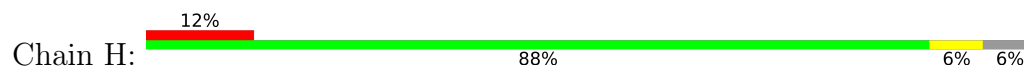
• Molecule 2: POLYCOMB PROTEIN PHO



• Molecule 2: POLYCOMB PROTEIN PHO



• Molecule 2: POLYCOMB PROTEIN PHO



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	87.51Å 147.88Å 97.88Å 90.00° 112.26° 90.00°	Depositor
Resolution (Å)	48.60 – 1.95 48.60 – 1.95	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.60-1.95) 99.7 (48.60-1.95)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.66 (at 1.95Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.179 , 0.198 0.184 , 0.202	Depositor DCC
R_{free} test set	8327 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	28.8	Xtriage
Anisotropy	0.256	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 47.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	16286	wwPDB-VP
Average B, all atoms (Å ²)	33.0	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 30.08 % 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 1.3892e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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](#)

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	0.29	0/3582	0.67	0/4880
1	B	0.28	0/3583	0.65	0/4883
1	C	0.27	0/3556	0.66	0/4843
1	D	0.28	0/3509	0.67	0/4783
2	E	0.25	0/243	0.62	0/324
2	F	0.28	0/237	0.61	0/317
2	G	0.25	0/236	0.56	0/316
2	H	0.24	0/238	0.58	0/318
All	All	0.28	0/15184	0.66	0/20664

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	3479	0	3272	14	0
1	B	3480	0	3267	13	0
1	C	3456	0	3245	15	0
1	D	3409	0	3182	11	0
2	E	238	0	233	1	0
2	F	232	0	222	1	0
2	G	231	0	219	0	0
2	H	233	0	224	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	427	0	0	2	0
3	B	409	0	0	0	0
3	C	384	0	0	2	0
3	D	256	0	0	2	0
3	E	21	0	0	0	0
3	F	10	0	0	1	0
3	G	7	0	0	0	0
3	H	14	0	0	0	0
All	All	16286	0	13864	54	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 (54) 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:C:722:PHE:CE2	1:C:722:PHE:CG	2.41	0.99
1:B:814:SER:OG	1:B:816[B]:CYS:SG	2.51	0.69
1:B:720:ARG:NH1	1:B:725:ASP:O	2.27	0.67
1:A:832:ARG:NH1	1:A:844:ASP:OD2	2.26	0.67
1:D:853:PRO:O	1:D:900:ARG:NH2	2.27	0.67
1:B:814:SER:HG	1:B:816[B]:CYS:HG	1.46	0.63
2:F:143:MET:O	3:F:2001:HOH:O	2.16	0.62
1:A:720:ARG:NH1	1:A:725:ASP:O	2.32	0.62
1:A:875:ASP:HB3	1:A:878:THR:H	1.65	0.62
1:D:613:GLU:OE2	3:D:2055:HOH:O	2.16	0.61
1:A:814:SER:OG	1:A:816[B]:CYS:SG	2.58	0.61
1:C:613:GLU:OE1	3:C:2059:HOH:O	2.17	0.59
1:C:580:CYS:HA	1:C:648:ARG:HG3	1.87	0.55
1:B:904[A]:ASP:OD1	1:B:904[A]:ASP:N	2.40	0.54
1:A:789:TYR:OH	1:A:851:LYS:NZ	2.41	0.54
1:C:814:SER:OG	1:C:816[B]:CYS:SG	2.67	0.53
1:A:627:CYS:HB2	1:A:967:LEU:HD13	1.91	0.53
1:A:759:ARG:NH2	3:A:2222:HOH:O	2.44	0.51
1:A:787:ASP:OD1	1:C:639:THR:OG1	2.26	0.49
1:B:816[A]:CYS:SG	1:B:835:SER:HB3	2.53	0.48
1:B:574:GLU:HG3	1:B:634:VAL:HA	1.97	0.47
1:C:627:CYS:HB2	1:C:967:LEU:HD13	1.97	0.46
1:A:613:GLU:OE1	3:A:2064:HOH:O	2.20	0.46
1:D:548:ASN:ND2	3:D:2015:HOH:O	2.45	0.46
2:E:152:LEU:HD11	2:E:163:SER:HB2	1.98	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:577:ASN:ND2	1:D:645:ILE:O	2.49	0.45
1:D:760:MET:HA	1:D:760:MET:HE2	1.99	0.44
1:B:752:ALA:HB3	1:B:757:LEU:HD13	1.99	0.44
1:C:724:SER:OG	1:C:725:ASP:N	2.49	0.44
1:B:919:MET:HA	1:B:919:MET:HE2	1.99	0.43
1:A:816[A]:CYS:SG	1:A:835:SER:HB3	2.58	0.43
1:C:919:MET:HA	1:C:919:MET:HE2	2.00	0.43
1:D:756:TYR:O	1:D:760:MET:HG2	2.18	0.43
1:B:627:CYS:HB2	1:B:967:LEU:HD13	2.00	0.43
1:C:555:PHE:CD1	1:C:961:PRO:HG3	2.54	0.43
1:C:784:PHE:HB3	1:C:851:LYS:HD2	2.01	0.43
1:A:555:PHE:CD1	1:A:961:PRO:HG3	2.54	0.42
1:C:596:TRP:CG	1:C:597:VAL:H	2.36	0.42
1:D:904:ASP:OD1	1:D:904:ASP:N	2.44	0.42
1:C:541:ARG:NH2	3:C:2015:HOH:O	2.46	0.42
1:D:580:CYS:HA	1:D:648:ARG:HG3	2.01	0.42
1:B:797:SER:HB3	1:B:897:LEU:HD22	2.02	0.42
1:C:722:PHE:CE2	1:C:723:ASP:HB3	2.55	0.42
1:A:584:GLU:OE2	1:A:642:LYS:NZ	2.44	0.41
1:D:693:LEU:HD11	1:D:719:LEU:HD22	2.01	0.41
1:D:821:MET:SD	1:D:832:ARG:HG2	2.60	0.41
1:A:920:ASP:OD2	1:A:922:ARG:NH2	2.53	0.41
1:C:574:GLU:HG3	1:C:634:VAL:HA	2.02	0.41
1:B:555:PHE:CD1	1:B:961:PRO:HG3	2.56	0.41
1:D:816:CYS:SG	1:D:835:SER:HB3	2.60	0.41
1:C:586:ILE:HG21	1:C:592:PRO:HA	2.04	0.41
1:B:904[A]:ASP:OD2	2:H:142:ALA:N	2.49	0.40
1:B:757:LEU:HD12	1:B:757:LEU:HA	1.91	0.40
1:A:844:ASP:OD1	1:A:844:ASP:N	2.54	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	438/451 (97%)	426 (97%)	12 (3%)	0	100	100
1	B	440/451 (98%)	431 (98%)	9 (2%)	0	100	100
1	C	436/451 (97%)	426 (98%)	10 (2%)	0	100	100
1	D	435/451 (96%)	424 (98%)	11 (2%)	0	100	100
2	E	28/32 (88%)	27 (96%)	1 (4%)	0	100	100
2	F	28/32 (88%)	26 (93%)	2 (7%)	0	100	100
2	G	28/32 (88%)	27 (96%)	1 (4%)	0	100	100
2	H	28/32 (88%)	27 (96%)	1 (4%)	0	100	100
All	All	1861/1932 (96%)	1814 (98%)	47 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	372/392 (95%)	372 (100%)	0	100	100
1	B	372/392 (95%)	367 (99%)	5 (1%)	61	58
1	C	368/392 (94%)	366 (100%)	2 (0%)	81	82
1	D	358/392 (91%)	355 (99%)	3 (1%)	73	74
2	E	24/26 (92%)	24 (100%)	0	100	100
2	F	23/26 (88%)	22 (96%)	1 (4%)	26	15
2	G	22/26 (85%)	22 (100%)	0	100	100
2	H	23/26 (88%)	22 (96%)	1 (4%)	26	15
All	All	1562/1672 (93%)	1550 (99%)	12 (1%)	73	74

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	648	ARG
1	B	706	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	757	LEU
1	B	772	GLU
1	B	913	LEU
1	C	913	LEU
1	C	923	LEU
1	D	882	GLU
1	D	913	LEU
1	D	968	VAL
2	F	169	SER
2	H	143	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	566	ASN
1	A	899	HIS
1	A	980	HIS
1	B	566	ASN
1	B	899	HIS
1	C	566	ASN
1	C	872	ASN
1	C	899	HIS
1	C	980	HIS
1	D	566	ASN
1	D	692	ASN
2	H	150	GLN

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 oligosaccharides 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, 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	440/451 (97%)	-0.12	10 (2%) 61 68	11, 25, 50, 67	3 (0%)
1	B	439/451 (97%)	-0.13	9 (2%) 63 70	11, 27, 56, 73	5 (1%)
1	C	438/451 (97%)	-0.05	7 (1%) 70 77	13, 30, 52, 75	2 (0%)
1	D	437/451 (96%)	0.24	19 (4%) 40 46	12, 39, 65, 77	2 (0%)
2	E	30/32 (93%)	-0.03	1 (3%) 49 55	14, 25, 46, 53	0
2	F	30/32 (93%)	0.38	5 (16%) 4 4	18, 33, 61, 77	0
2	G	30/32 (93%)	0.46	4 (13%) 7 8	24, 36, 53, 62	1 (3%)
2	H	30/32 (93%)	0.25	4 (13%) 7 8	17, 29, 52, 58	0
All	All	1874/1932 (96%)	0.00	59 (3%) 51 58	11, 30, 58, 77	13 (0%)

All (59) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	770	VAL	5.1
1	C	588	PRO	4.8
2	H	170	GLY	4.4
1	A	724	SER	3.7
1	C	761	LEU	3.6
1	D	875	ASP	3.5
2	H	143	MET	3.5
1	D	770	VAL	3.4
1	B	724	SER	3.4
1	C	534	HIS	3.3
1	A	761	LEU	3.2
1	B	877	ARG	3.2
2	F	142	ALA	3.1
1	B	876	SER	3.1
1	D	878	THR	3.1
1	B	532	PRO	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	837	GLN	3.0
2	G	170	GLY	3.0
1	A	878	THR	3.0
1	B	763	GLY	3.0
1	D	589	GLY	2.9
2	H	141	GLY	2.9
2	H	142	ALA	2.8
1	B	588	PRO	2.8
2	G	143	MET	2.8
1	D	761	LEU	2.8
1	B	878	THR	2.8
1	D	534	HIS	2.8
2	G	142	ALA	2.8
1	D	839	ASP	2.7
1	D	977	ARG	2.7
1	A	762	ALA	2.7
1	D	588	PRO	2.7
1	D	590	GLN	2.6
2	F	141	GLY	2.6
1	B	762	ALA	2.6
1	D	722	PHE	2.6
1	A	583	ILE	2.5
1	A	875	ASP	2.5
1	C	581	ASP	2.5
2	F	144	ALA	2.4
1	A	534	HIS	2.4
1	D	587	GLN	2.4
2	F	143	MET	2.4
1	C	723	ASP	2.3
1	D	585	VAL	2.3
1	C	977	ARG	2.3
1	D	899	HIS	2.3
2	E	170	GLY	2.3
1	D	755	ASP	2.2
1	C	722	PHE	2.2
1	D	725	ASP	2.2
1	A	582	SER	2.1
2	G	141	GLY	2.1
1	B	761	LEU	2.1
1	D	876	SER	2.0
1	A	786	PHE	2.0
2	F	170	GLY	2.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	838	PRO	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 oligosaccharides 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.