Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 10:00 PM GMT

PDB ID	:	4WIJ
Title	:	HUMAN SPLICING FACTOR, CONSTRUCT 1
Authors	:	lee, M.; bond, c.s.
Deposited on	:	2014-09-26
Resolution	:	3.49 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:

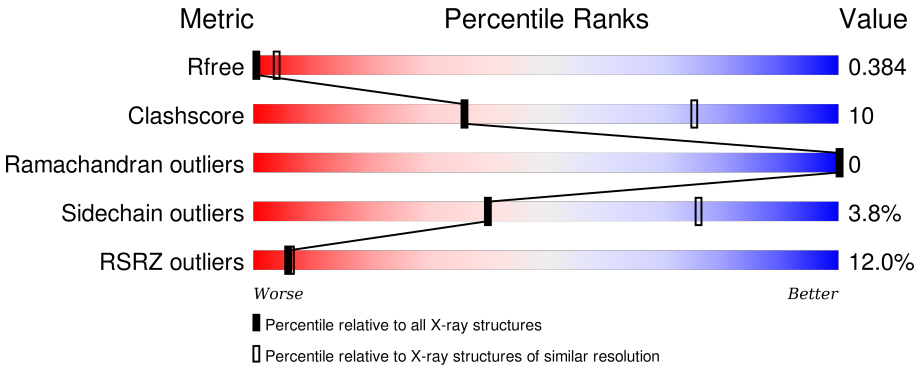
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 3.49 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	1051 (3.60-3.40)
Clashscore	102246	1157 (3.60-3.40)
Ramachandran outliers	100387	1120 (3.60-3.40)
Sidechain outliers	100360	1121 (3.60-3.40)
RSRZ outliers	91569	1058 (3.60-3.40)

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	326	5% 76% 14% • 9%
1	B	326	17% 74% 17% • 8%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 4978 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 Splicing factor, proline- and glutamine-rich.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	298	Total	C	N	O	S	0	1	0
			2473	1532	453	473	15			
1	B	301	Total	C	N	O	S	0	1	0
			2505	1548	463	479	15			

There are 6 discrepancies between the modelled and reference sequences:

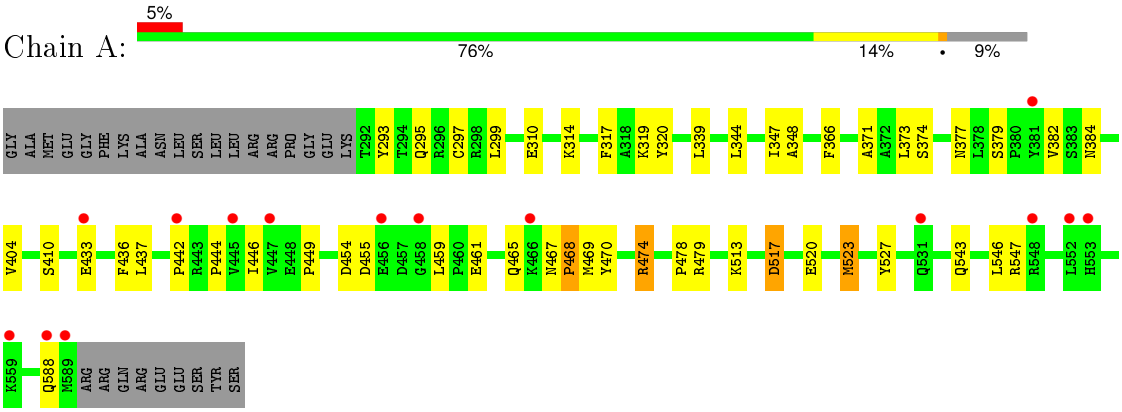
Chain	Residue	Modelled	Actual	Comment	Reference
A	273	GLY	-	expression tag	UNP P23246
A	274	ALA	-	expression tag	UNP P23246
A	275	MET	-	expression tag	UNP P23246
B	273	GLY	-	expression tag	UNP P23246
B	274	ALA	-	expression tag	UNP P23246
B	275	MET	-	expression tag	UNP P23246

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.

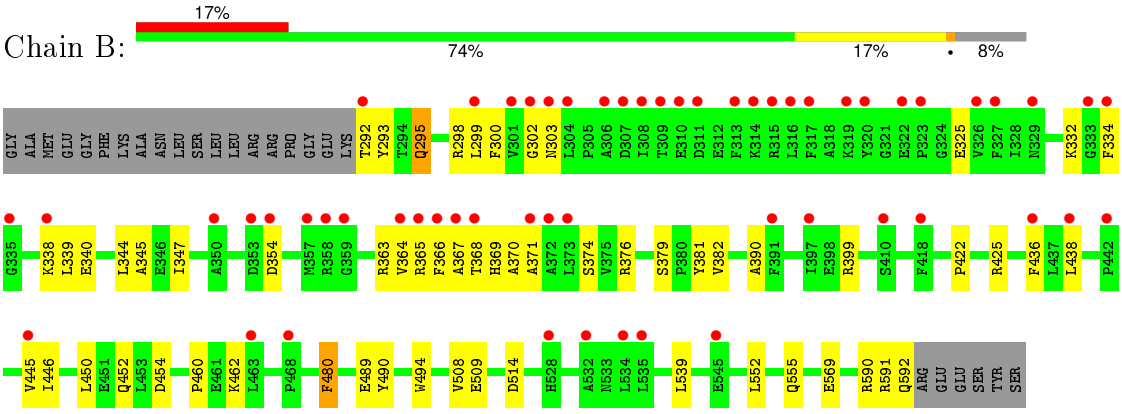
- Molecule 1: Splicing factor, proline- and glutamine-rich

Chain A:



- Molecule 1: Splicing factor, proline- and glutamine-rich

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	66.58Å 66.58Å 398.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.08 – 3.49 47.08 – 3.49	Depositor EDS
% Data completeness (in resolution range)	99.4 (47.08-3.49) 99.7 (47.08-3.49)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.87 (at 3.48Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.272 , 0.335 0.320 , 0.384	Depositor DCC
R_{free} test set	605 reflections (4.93%)	DCC
Wilson B-factor (Å^2)	103.8	Xtriage
Anisotropy	0.667	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 67.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.39$, $< L^2 > = 0.21$	Xtriage
Outliers	0 of 12261 reflections	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	4978	wwPDB-VP
Average B, all atoms (Å^2)	127.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.09% 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

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	0.37	1/2516 (0.0%)	0.62	2/3366 (0.1%)
1	B	0.35	0/2548	0.56	1/3407 (0.0%)
All	All	0.36	1/5064 (0.0%)	0.59	3/6773 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	468	PRO	N-CD	5.93	1.56	1.47

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	474	ARG	CB-CA-C	5.70	121.81	110.40
1	B	364	VAL	N-CA-C	-5.55	96.00	111.00
1	A	454	ASP	CB-CA-C	-5.53	99.34	110.40

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	2473	0	2440	53	0
1	B	2505	0	2469	55	0
All	All	4978	0	4909	95	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 10.

All (95) 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:470:TYR:CE1	1:A:474:ARG:HD3	1.60	1.33
1:B:300:PHE:N	1:B:367:ALA:HB2	1.73	1.03
1:A:320:TYR:O	1:A:344:LEU:HD22	1.55	1.02
1:B:332:LYS:HB2	1:B:334:PHE:HE1	1.30	0.94
1:A:470:TYR:CE1	1:A:474:ARG:CD	2.50	0.94
1:B:552:LEU:HD23	1:B:555:GLN:OE1	1.66	0.94
1:A:470:TYR:HE1	1:A:474:ARG:CD	1.81	0.93
1:B:299:LEU:HD12	1:B:339:LEU:HD11	1.55	0.88
1:A:470:TYR:HE1	1:A:474:ARG:HD3	1.10	0.88
1:A:470:TYR:CD1	1:A:474:ARG:HD3	2.08	0.88
1:B:303:ASN:HD21	1:B:363:ARG:NH1	1.73	0.85
1:A:467:ASN:OD1	1:A:468:PRO:HD3	1.77	0.84
1:B:365:ARG:HD2	1:B:454:ASP:OD1	1.78	0.83
1:B:300:PHE:H	1:B:367:ALA:HB2	1.42	0.81
1:B:332:LYS:HB2	1:B:334:PHE:CE1	2.15	0.81
1:B:293:TYR:O	1:B:338:LYS:HD3	1.81	0.79
1:B:591:ARG:O	1:B:592:GLN:HB2	1.82	0.78
1:A:478:PRO:HB3	1:B:438:LEU:HD23	1.68	0.74
1:B:298:ARG:NH2	1:B:325:GLU:OE2	2.24	0.70
1:A:467:ASN:OD1	1:A:468:PRO:CD	2.43	0.67
1:B:376:ARG:HG3	1:B:446:ILE:HB	1.75	0.67
1:A:377:ASN:ND2	1:B:494:TRP:CZ3	2.64	0.66
1:B:365:ARG:HD2	1:B:454:ASP:CG	2.15	0.65
1:B:339:LEU:HD13	1:B:345:ALA:HA	1.78	0.65
1:A:467:ASN:CG	1:A:468:PRO:CD	2.64	0.65
1:A:317:PHE:CE1	1:A:348:ALA:HB1	2.34	0.62
1:A:317:PHE:HE1	1:A:348:ALA:HB1	1.63	0.62
1:A:455:ASP:HB2	1:B:399:ARG:HH22	1.63	0.61
1:B:368:THR:HB	1:B:452:GLN:HB3	1.82	0.61
1:A:461:GLU:O	1:A:465:GLN:HG3	2.01	0.61
1:A:523:MET:O	1:A:527:TYR:HB3	2.01	0.60
1:A:468:PRO:HD2	1:A:469:MET:H	1.67	0.60
1:A:467:ASN:HD22	1:B:381:TYR:HE1	1.48	0.60
1:B:436:PHE:CE2	1:B:438:LEU:HD21	2.37	0.59
1:A:470:TYR:HE1	1:A:474:ARG:NH1	2.01	0.59
1:A:377:ASN:HB2	1:A:446:ILE:HD12	1.83	0.59
1:A:461:GLU:HG2	1:A:465:GLN:OE1	2.01	0.59

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:303:ASN:HD22	1:B:363:ARG:HG3	1.68	0.59
1:B:332:LYS:CB	1:B:334:PHE:CE1	2.86	0.59
1:B:303:ASN:ND2	1:B:363:ARG:NH1	2.52	0.56
1:A:470:TYR:CE1	1:A:474:ARG:NH1	2.75	0.54
1:A:465:GLN:NE2	1:A:470:TYR:OH	2.41	0.54
1:B:299:LEU:CD1	1:B:339:LEU:HD11	2.32	0.54
1:A:465:GLN:HB3	1:A:470:TYR:CE2	2.43	0.53
1:A:436:PHE:HD1	1:B:480:PHE:CD2	2.26	0.53
1:A:455:ASP:O	1:B:399:ARG:NH1	2.43	0.52
1:A:319:LYS:O	1:A:319:LYS:CG	2.57	0.52
1:A:344:LEU:HA	1:A:347:ILE:HD12	1.92	0.52
1:A:478:PRO:HB3	1:B:438:LEU:CD2	2.40	0.51
1:B:438:LEU:CD1	1:B:445:VAL:HG21	2.41	0.51
1:A:470:TYR:HE1	1:A:474:ARG:CZ	2.24	0.51
1:B:295:GLN:O	1:B:298:ARG:HB2	2.12	0.49
1:A:442:PRO:O	1:B:490:TYR:HE2	1.95	0.49
1:A:467:ASN:ND2	1:B:381:TYR:CE1	2.77	0.49
1:A:295:GLN:HE22	1:A:371:ALA:CB	2.25	0.49
1:A:293:TYR:HA	1:A:339:LEU:O	2.13	0.49
1:B:303:ASN:ND2	1:B:363:ARG:HG3	2.27	0.49
1:A:373:LEU:HD23	1:A:449:PRO:HA	1.94	0.48
1:A:467:ASN:CG	1:A:468:PRO:HD2	2.33	0.48
1:B:295:GLN:HG2	1:B:369:HIS:ND1	2.29	0.48
1:B:292:THR:HG22	1:B:340:GLU:O	2.13	0.48
1:A:436:PHE:HD1	1:B:480:PHE:HD2	1.62	0.48
1:B:422:PRO:HA	1:B:425:ARG:HD2	1.96	0.48
1:A:442:PRO:O	1:B:490:TYR:CE2	2.67	0.47
1:B:339:LEU:CD1	1:B:345:ALA:HA	2.43	0.47
1:B:299:LEU:C	1:B:367:ALA:HB2	2.32	0.47
1:A:446:ILE:HD11	1:B:494:TRP:HZ3	1.80	0.47
1:B:370:ALA:O	1:B:371:ALA:HB3	2.15	0.47
1:A:470:TYR:HE1	1:A:474:ARG:HH11	1.62	0.47
1:A:470:TYR:HE1	1:A:474:ARG:NE	2.12	0.47
1:B:325:GLU:CD	1:B:338:LYS:HE3	2.35	0.47
1:B:438:LEU:HD12	1:B:445:VAL:HG21	1.97	0.47
1:A:299:LEU:HD23	1:A:366:PHE:HA	1.98	0.46
1:A:517:ASP:HA	1:A:520:GLU:HB2	1.96	0.46
1:B:344:LEU:HA	1:B:347:ILE:HD12	1.97	0.46
1:B:591:ARG:O	1:B:592:GLN:CB	2.59	0.45
1:B:365:ARG:O	1:B:366:PHE:C	2.55	0.45
1:A:468:PRO:CD	1:A:469:MET:H	2.30	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:319:LYS:O	1:A:319:LYS:HG2	2.16	0.44
1:B:374:SER:HB2	1:B:450:LEU:HD12	1.98	0.44
1:A:543:GLN:HE21	1:A:547:ARG:HG3	1.82	0.44
1:B:390:ALA:HB1	1:B:438:LEU:CD2	2.48	0.44
1:A:295:GLN:HE22	1:A:371:ALA:HB3	1.83	0.44
1:A:379:SER:HB2	1:A:382:VAL:HG23	2.00	0.43
1:A:310:GLU:HG2	1:A:314:LYS:HE3	2.00	0.43
1:A:404:VAL:HG12	1:A:410:SER:HA	2.00	0.43
1:B:302:GLY:O	1:B:303:ASN:HB2	2.18	0.43
1:B:365:ARG:CD	1:B:454:ASP:CG	2.87	0.42
1:B:460:PRO:HB2	1:B:462:LYS:HG2	2.00	0.42
1:A:467:ASN:ND2	1:B:381:TYR:OH	2.53	0.42
1:B:590:ARG:O	1:B:590:ARG:HG2	2.18	0.42
1:A:377:ASN:ND2	1:A:444:PRO:O	2.53	0.41
1:B:379:SER:HB3	1:B:382:VAL:HG23	2.02	0.41
1:B:390:ALA:HB1	1:B:438:LEU:HD21	2.03	0.41
1:A:436:PHE:O	1:A:437:LEU:HD23	2.21	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	297/326 (91%)	288 (97%)	9 (3%)	0	100	100
1	B	300/326 (92%)	290 (97%)	10 (3%)	0	100	100
All	All	597/652 (92%)	578 (97%)	19 (3%)	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	264/286 (92%)	253 (96%)	11 (4%)	36 74
1	B	267/286 (93%)	257 (96%)	10 (4%)	41 76
All	All	531/572 (93%)	510 (96%)	21 (4%)	40 75

All (21) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	297	CYS
1	A	374	SER
1	A	384	ASN
1	A	433	GLU
1	A	459	LEU
1	A	479	ARG
1	A	513	LYS
1	A	517	ASP
1	A	523	MET
1	A	546	LEU
1	A	588	GLN
1	B	295	GLN
1	B	354	ASP
1	B	480	PHE
1	B	489	GLU
1	B	508	VAL
1	B	509	GLU
1	B	514[A]	ASP
1	B	514[B]	ASP
1	B	539	LEU
1	B	569	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	295	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	329	ASN
1	A	377	ASN
1	A	465	GLN
1	A	467	ASN
1	A	471	GLN
1	A	543	GLN
1	B	303	ASN
1	B	492	GLN
1	B	503	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

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, 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	298/326 (91%)	0.38	15 (5%) 32 25	49, 98, 168, 181	0
1	B	301/326 (92%)	0.94	57 (18%) 2 2	75, 129, 239, 251	0
All	All	599/652 (91%)	0.66	72 (12%) 6 6	49, 117, 235, 251	0

All (72) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	329	ASN	7.7
1	B	335	GLY	7.5
1	B	320	TYR	7.3
1	B	334	PHE	7.0
1	A	442	PRO	6.0
1	B	372	ALA	5.5
1	B	304	LEU	5.4
1	B	468	PRO	5.4
1	B	319	LYS	5.3
1	B	364	VAL	5.0
1	B	326	VAL	4.7
1	B	358	ARG	4.7
1	B	366	PHE	4.3
1	B	365	ARG	4.2
1	B	528	HIS	4.2
1	B	391	PHE	4.0
1	B	368	THR	4.0
1	B	311	ASP	4.0
1	B	303	ASN	4.0
1	B	397	ILE	3.9
1	B	354	ASP	3.9
1	B	317	PHE	3.8
1	B	313	PHE	3.8
1	B	301	VAL	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	316	LEU	3.7
1	B	302	GLY	3.6
1	B	299	LEU	3.5
1	B	438	LEU	3.4
1	B	307	ASP	3.3
1	B	322	GLU	3.2
1	B	292	THR	3.2
1	B	445	VAL	3.0
1	B	338	LYS	3.0
1	B	535	LEU	3.0
1	B	463	LEU	3.0
1	B	442	PRO	2.9
1	B	314	LYS	2.8
1	A	559	LYS	2.8
1	A	466	LYS	2.7
1	B	323	PRO	2.7
1	B	310	GLU	2.7
1	B	315	ARG	2.7
1	A	447	VAL	2.6
1	B	371	ALA	2.6
1	A	531	GLN	2.6
1	B	367	ALA	2.6
1	A	589	MET	2.6
1	B	359	GLY	2.5
1	B	308	ILE	2.5
1	B	534	LEU	2.5
1	B	306	ALA	2.5
1	A	553	HIS	2.5
1	A	456	GLU	2.5
1	B	350	ALA	2.5
1	B	532	ALA	2.4
1	B	436	PHE	2.4
1	A	381	TYR	2.3
1	A	552	LEU	2.3
1	A	433	GLU	2.3
1	B	327	PHE	2.3
1	A	548	ARG	2.2
1	B	545	GLU	2.2
1	B	309	THR	2.2
1	A	458	GLY	2.2
1	A	588	GLN	2.1
1	B	353	ASP	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	445	VAL	2.1
1	B	410	SER	2.1
1	B	357	MET	2.1
1	B	418	PHE	2.1
1	B	373	LEU	2.1
1	B	333	GLY	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.