



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:54 AM GMT

PDB ID : 3K4A
Title : Crystal structure of selenomethionine substituted E. coli beta-glucuronidase
Authors : Wallace, B.D.; Orans, J.; Redinbo, M.R.
Deposited on : 2009-10-05
Resolution : 2.90 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 ⓘ 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 (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

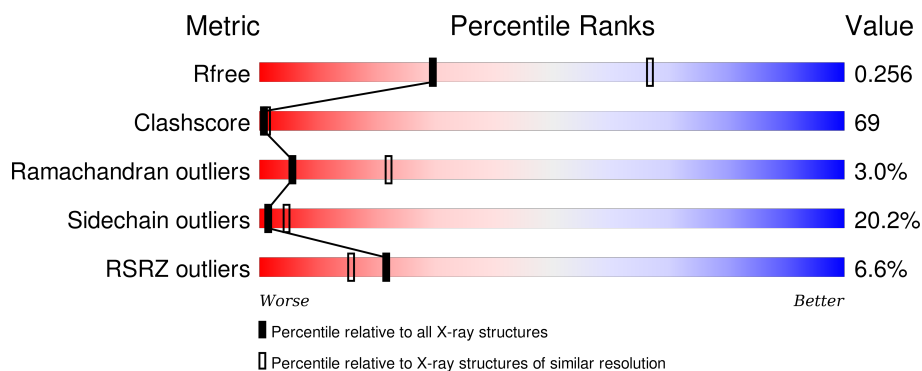
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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	1451 (2.90-2.90)
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

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 $\leq 5\%$. 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	605	<div> <div>5%</div> <div>33%</div> <div>51%</div> <div>14%</div> <div>..</div> </div>
1	B	605	<div> <div>8%</div> <div>25%</div> <div>55%</div> <div>16%</div> <div>..</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 9761 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 Beta-glucuronidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	597	Total	C	N	O	S	Se	0	0	0
			4779	3034	827	896	9	13			
1	B	597	Total	C	N	O	S	Se	0	0	0
			4768	3026	826	894	9	13			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	SER	-	EXPRESSION TAG	UNP P05804
A	0	HIS	-	EXPRESSION TAG	UNP P05804
B	-1	SER	-	EXPRESSION TAG	UNP P05804
B	0	HIS	-	EXPRESSION TAG	UNP P05804

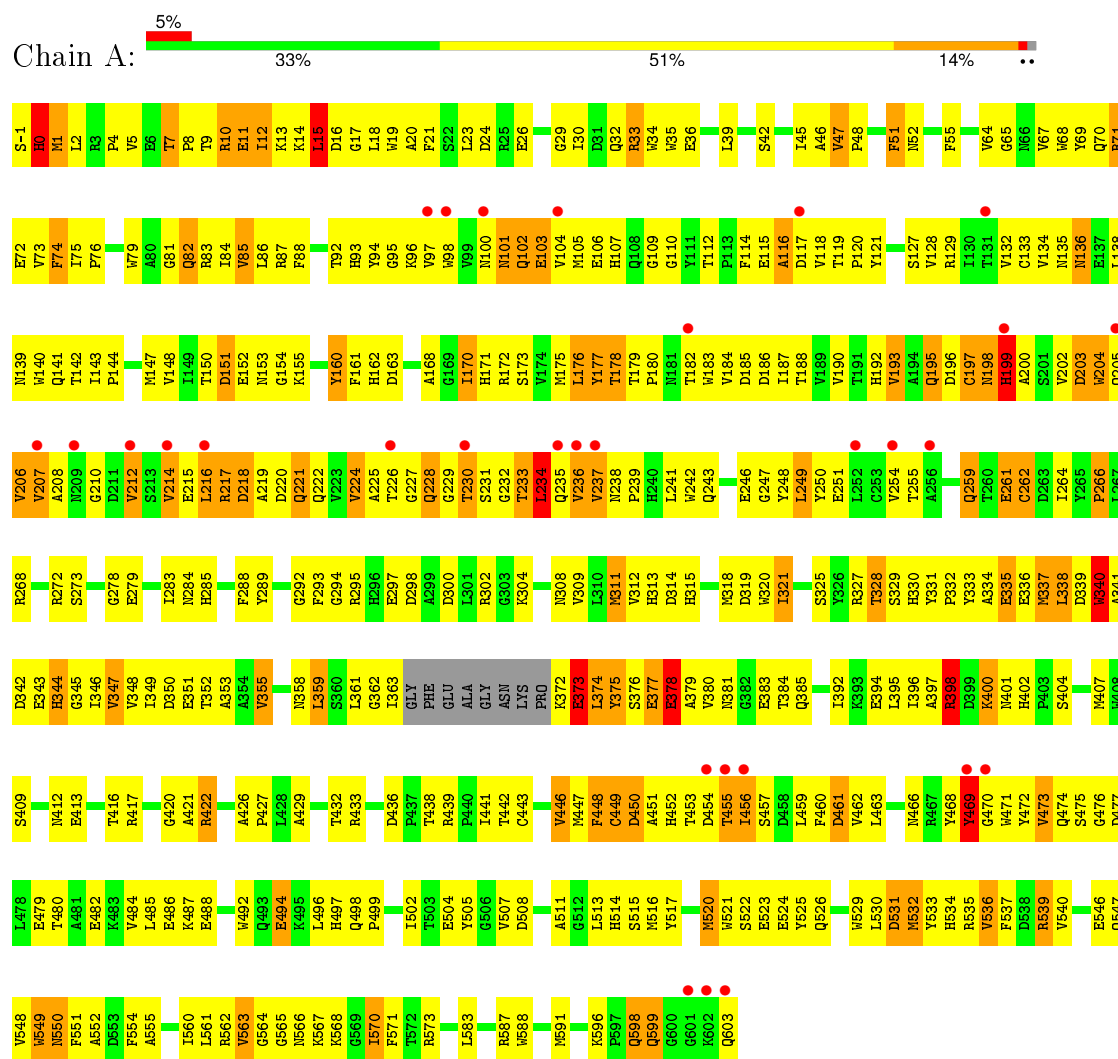
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	123	Total	O	0	0
			123	123		
2	B	91	Total	O	0	0
			91	91		

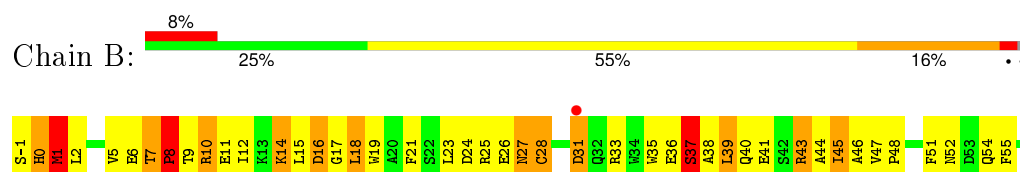
3 Residue-property plots

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 ($\text{RSRZ} > 2$). 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: Beta-glucuronidase



• Molecule 1: Beta-glucuronidase



L583	K586	K587	M588	T589	G590	M591	N592	F593	G594	E595	K596	P597	Q598	Q599	G600	Q603																																																	
A511	Y517	D518	D519	M520	M521	S522	S523	E524	Y525	Q526	W529	L530	D531	M532	Y533	H534	R535	D537	D538	R539	Y540	S541	A542	G545	E546	Q547	V548	N550	F551	A552	D553	F554	A555	T556	S557	Q558	G559	I560	L561	R562	Y563	G564	G565	M566	K567	K568	G569	I570	R573	K578															
T384	Q385	Q386	A387	H388	D389	Q390	A391	I392	K393	E394	S395	A396	E397	R398	D399	L399	V400	H401	H402	P403	S404	V405	M407	M408	S409	I410	A411	T346	V347	V348	I349	D350	E351	T352	A353	P354	V355	G356	F357	R417	P418	Q419	G420	A421	R422	E423	F424	F425	A429	E430	A431	T432	R433	K434	L435	D436	P437	T438	R439	P440	I441	T442	C443	V444	N445
V446	M447	F448	C449	D450	A451	H452	T453	D454	T455	L456	A457	D458	L459	F460	D461	V462	L463	C464	H465	V466	R467	V468	Y469	G470	W471	W472	W473	Q474	S475	G476	A481	E482	K483	V484	L485	E486	K487	E488	L489	L490	A491	W492	Q493	E494	T495	K496	L497	I500	I501	I502	I503	E504	D508	T509	L540	W541	W542								
M135	M136	E137	L138	M139	W140	Q141	T142	T143	P144	P145	M146	G147	V148	I149	T150	D151	E152	M153	G154	K155	K156	K157	Q158	H162	D163	F164	Y167	A168	G169	I170	H171	R172	S173	M174	M175	L176	Y177	T178	T179	P180	M181	T182	M183	V184	D185	D186	I187	T188	V189	V190	T191	C192	V193	A194	Q195	D196	C197								
H198	H199	A200	S201	D202	V203	K204	Q205	V206	V207	A208	N209	G210	M211	V212	S213	V214	E215	L216	R217	D218	A219	D220	Q221	Q222	V223	V224	A225	T226	G227	Q228	G229	T230	S231	L234	Q235	V236	V237	R238	P239	E240	L241	K242	Q243	P244	G245	E246	G247	Y248	L249	Y250	E251	L252	C253	V254	T255	A256	K257	S258							
Q259	T260	E261	C262	D263	L264	V265	P266	L267	R268	V269	G270	L271	R272	S273	V274	A275	V276	Q280	F281	L282	L283	N284	R285	K286	F287	F288	Y289	F290	T291	G292	F293	G294	R295	E296	E297	D298	A299	R300	L301	R302	G303	K304	G305	F306	D307	LYS	PRO	K308	V309	L310	K311	D314	H315	K318	D319	R320	I321	G322							
A323	N324	S325	Y326	R327	T328	D329	H330	Y331	F332	E333	A334	E335	E336	R337	L338	D339	V340	A341	D342	E343	I346	V347	V348	I349	D350	E351	T352	A353	P354	V355	G356	F357	R358	L359	S360	L361	G362	I363	GLY	PHE	GLU	ALA	GLY	ASN	LYS	PRO	K372	E373	L374	Y375	S376	E377	E378	A379	V380	N381	K382	E383							
T384	Q385	Q386	A387	H388	D389	Q390	A391	I392	K393	E394	S395	A396	E397	R398	D399	L399	V400	H401	H402	P403	S404	V405	M407	M408	S409	I410	A411	T346	V347	V348	I349	D350	E351	T352	A353	P354	V355	G356	F357	R417	P418	Q419	G420	A421	R422	E423	F424	F425	A429	E430	A431	T432	R433	K434	L435	D436	P437	T438	R439	P440	I441	T442	C443	V444	N445
V446	M447	F448	C449	D450	A451	H452	T453	D454	T455	L456	A457	D458	L459	F460	D461	V462	L463	C464	H465	V466	R467	V468	Y469	G470	W471	W472	W473	Q474	S475	G476	A481	E482	K483	V484	L485	E486	K487	E488	L489	L490	A491	W492	Q493	E494	T495	K496	L497	I500	I501	I502	I503	E504	D508	T509	L540	W541	W542								
A511	Y517	D518	D519	M520	M521	S522	S523	E524	Y525	Q526	W529	L530	D531	M532	Y533	H534	R535	D537	D538	R539	Y540	S541	A542	G545	E546	Q547	V548	N550	F551	A552	D553	F554	A555	T556	S557	Q558	G559	I560	L561	R562	Y563	G564	G565	M566	K567	K568	G569	I570	R573	K578															
L583	K586	K587	M588	T589	G590	M591	N592	F593	G594	E595	K596	P597	Q598	Q599	G600	Q603																																																	

4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	169.00Å 77.26Å 126.58Å 90.00° 125.02° 90.00°	Depositor
Resolution (Å)	32.11 – 2.90 48.11 – 2.90	Depositor EDS
% Data completeness (in resolution range)	100.0 (32.11-2.90) 98.9 (48.11-2.90)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.59 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, R_{free}	0.242 , 0.282 0.256 , 0.256	Depositor DCC
R_{free} test set	1495 reflections (5.34%)	DCC
Wilson B-factor (Å ²)	54.2	Xtriage
Anisotropy	0.687	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 61.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 29542 reflections	Xtriage
F_o, F_c correlation	0.88	EDS
Total number of atoms	9761	wwPDB-VP
Average B, all atoms (Å ²)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.55% 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, 0.375 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	0.98	8/4893 (0.2%)	1.05	19/6636 (0.3%)
1	B	0.82	2/4882 (0.0%)	0.94	8/6623 (0.1%)
All	All	0.91	10/9775 (0.1%)	1.00	27/13259 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	6
1	B	0	5
All	All	0	11

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	340	TRP	CB-CG	-9.12	1.33	1.50
1	A	1	MSE	CG-SE	-6.18	1.74	1.95
1	B	1	MSE	CG-SE	-5.88	1.75	1.95
1	A	532	MSE	CG-SE	-5.87	1.75	1.95
1	A	311	MSE	CG-SE	-5.72	1.76	1.95

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	374	LEU	N-CA-C	8.35	133.55	111.00
1	A	199	HIS	N-CA-C	7.98	132.56	111.00
1	A	573	ARG	NE-CZ-NH1	-7.83	116.38	120.30
1	A	198	ASN	N-CA-C	7.50	131.25	111.00
1	A	294	GLY	N-CA-C	-7.17	95.18	113.10

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	116	ALA	Peptide
1	A	197	CYS	Peptide
1	A	373	GLU	Peptide
1	A	378	GLU	Peptide
1	A	455	THR	Peptide

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	4779	0	4554	600	1
1	B	4768	0	4532	690	4
2	A	123	0	0	73	3
2	B	91	0	0	59	0
All	All	9761	0	9086	1279	4

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 69.

The worst 5 of 1279 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:B:502:ILE:HD12	1:B:537:PHE:CE2	1.49	1.48
1:A:455:THR:CG2	1:A:456:ILE:HG22	1.50	1.41
1:B:198:ASN:CB	1:B:236:VAL:HG22	1.59	1.32
1:B:444:VAL:CG1	1:B:466:ASN:HD21	1.47	1.26
1:A:456:ILE:CG1	1:A:459:LEU:HB2	1.64	1.26

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:152:GLU:OE2	2:A:685:HOH:O[2_554]	1.03	1.17
1:B:152:GLU:CD	2:A:685:HOH:O[2_554]	1.50	0.70

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:152:GLU:OE1	2:A:685:HOH:O[2_554]	1.83	0.37
1:A:599:GLN:O	1:B:603:GLN:C[2_564]	2.17	0.03

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	593/605 (98%)	522 (88%)	58 (10%)	13 (2%)	8	31
1	B	593/605 (98%)	494 (83%)	76 (13%)	23 (4%)	4	15
All	All	1186/1210 (98%)	1016 (86%)	134 (11%)	36 (3%)	5	22

5 of 36 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	228	GLN
1	B	7	THR
1	B	119	THR
1	B	123	ILE
1	B	144	PRO

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	505/501 (101%)	417 (83%)	88 (17%)	2	7
1	B	502/501 (100%)	387 (77%)	115 (23%)	1	3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1007/1002 (100%)	804 (80%)	203 (20%)	1 4

5 of 203 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	10	ARG
1	B	122	VAL
1	B	463	LEU
1	B	18	LEU
1	B	61	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 43 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	32	GLN
1	B	102	GLN
1	B	514	HIS
1	B	52	ASN
1	B	54	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 [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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	584/605 (96%)	0.11	30 (5%) 32 25	18, 45, 78, 103	0
1	B	584/605 (96%)	0.49	47 (8%) 15 10	39, 70, 89, 99	0
All	All	1168/1210 (96%)	0.30	77 (6%) 22 16	18, 61, 86, 103	0

The worst 5 of 77 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	227	GLY	6.5
1	B	204	TRP	6.4
1	B	206	VAL	5.9
1	A	98	TRP	5.2
1	B	469	TYR	4.9

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.