wwPDB X-ray Structure Validation Summary Report (i)

Jan 19, 2017 – 06:20 PM EST

PDB ID	:	5KO2
Title	:	Mouse pgp 34 linker deleted mutant Hg derivative
Authors	:	Xia, D.; Esser, L.; Zhou, F.
Deposited on	:	2016-06-29
Resolution	:	3.30 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

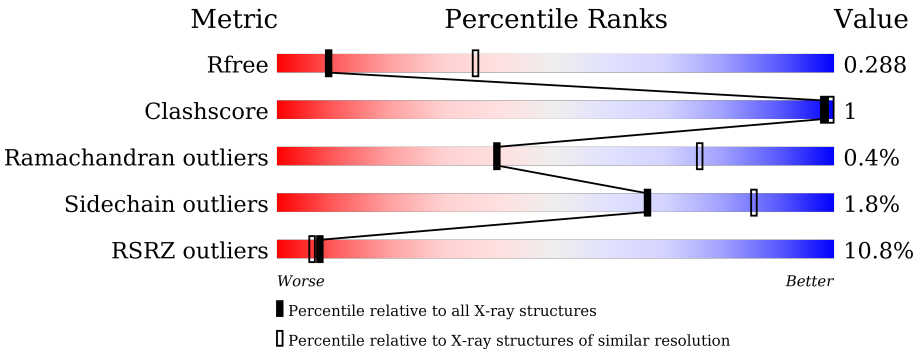
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

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 3.30 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	2060 (3.40-3.20)
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)

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	1248	10% 91% · 5%
1	B	1248	11% 90% · 5%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 37085 atoms, of which 18697 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 Multidrug resistance protein 1A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1185	Total	C	H	N	O	S	0	0	0
			18539	5907	9350	1559	1685	38			
1	B	1184	Total	C	H	N	O	S	0	0	0
			18532	5905	9347	1558	1684	38			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	552	GLN	GLU	engineered mutation	UNP P21447
A	?	-	MET	deletion	UNP P21447
A	?	-	SER	deletion	UNP P21447
A	?	-	SER	deletion	UNP P21447
A	?	-	LYS	deletion	UNP P21447
A	?	-	ASP	deletion	UNP P21447
A	?	-	SER	deletion	UNP P21447
A	?	-	GLY	deletion	UNP P21447
A	?	-	SER	deletion	UNP P21447
A	?	-	SER	deletion	UNP P21447
A	?	-	LEU	deletion	UNP P21447
A	?	-	ILE	deletion	UNP P21447
A	?	-	ARG	deletion	UNP P21447
A	?	-	ARG	deletion	UNP P21447
A	?	-	ARG	deletion	UNP P21447
A	?	-	SER	deletion	UNP P21447
A	?	-	THR	deletion	UNP P21447
A	?	-	ARG	deletion	UNP P21447
A	?	-	LYS	deletion	UNP P21447
A	?	-	SER	deletion	UNP P21447
A	?	-	ILE	deletion	UNP P21447
A	?	-	CYS	deletion	UNP P21447
A	?	-	GLY	deletion	UNP P21447
A	?	-	PRO	deletion	UNP P21447
A	?	-	HIS	deletion	UNP P21447

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	ASP	deletion	UNP P21447
A	?	-	GLN	deletion	UNP P21447
A	?	-	ASP	deletion	UNP P21447
A	?	-	ARG	deletion	UNP P21447
A	?	-	LYS	deletion	UNP P21447
A	?	-	LEU	deletion	UNP P21447
A	?	-	SER	deletion	UNP P21447
A	?	-	THR	deletion	UNP P21447
A	?	-	LYS	deletion	UNP P21447
A	?	-	GLU	deletion	UNP P21447
A	1197	GLN	GLU	engineered mutation	UNP P21447
A	1277	HIS	-	expression tag	UNP P21447
A	1278	HIS	-	expression tag	UNP P21447
A	1279	HIS	-	expression tag	UNP P21447
A	1280	HIS	-	expression tag	UNP P21447
A	1281	HIS	-	expression tag	UNP P21447
A	1282	HIS	-	expression tag	UNP P21447
B	552	GLN	GLU	engineered mutation	UNP P21447
B	?	-	MET	deletion	UNP P21447
B	?	-	SER	deletion	UNP P21447
B	?	-	SER	deletion	UNP P21447
B	?	-	LYS	deletion	UNP P21447
B	?	-	ASP	deletion	UNP P21447
B	?	-	SER	deletion	UNP P21447
B	?	-	GLY	deletion	UNP P21447
B	?	-	SER	deletion	UNP P21447
B	?	-	SER	deletion	UNP P21447
B	?	-	LEU	deletion	UNP P21447
B	?	-	ILE	deletion	UNP P21447
B	?	-	ARG	deletion	UNP P21447
B	?	-	ARG	deletion	UNP P21447
B	?	-	ARG	deletion	UNP P21447
B	?	-	SER	deletion	UNP P21447
B	?	-	THR	deletion	UNP P21447
B	?	-	ARG	deletion	UNP P21447
B	?	-	LYS	deletion	UNP P21447
B	?	-	SER	deletion	UNP P21447
B	?	-	ILE	deletion	UNP P21447
B	?	-	CYS	deletion	UNP P21447
B	?	-	GLY	deletion	UNP P21447
B	?	-	PRO	deletion	UNP P21447
B	?	-	HIS	deletion	UNP P21447

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	ASP	deletion	UNP P21447
B	?	-	GLN	deletion	UNP P21447
B	?	-	ASP	deletion	UNP P21447
B	?	-	ARG	deletion	UNP P21447
B	?	-	LYS	deletion	UNP P21447
B	?	-	LEU	deletion	UNP P21447
B	?	-	SER	deletion	UNP P21447
B	?	-	THR	deletion	UNP P21447
B	?	-	LYS	deletion	UNP P21447
B	?	-	GLU	deletion	UNP P21447
B	1197	GLN	GLU	engineered mutation	UNP P21447
B	1277	HIS	-	expression tag	UNP P21447
B	1278	HIS	-	expression tag	UNP P21447
B	1279	HIS	-	expression tag	UNP P21447
B	1280	HIS	-	expression tag	UNP P21447
B	1281	HIS	-	expression tag	UNP P21447
B	1282	HIS	-	expression tag	UNP P21447

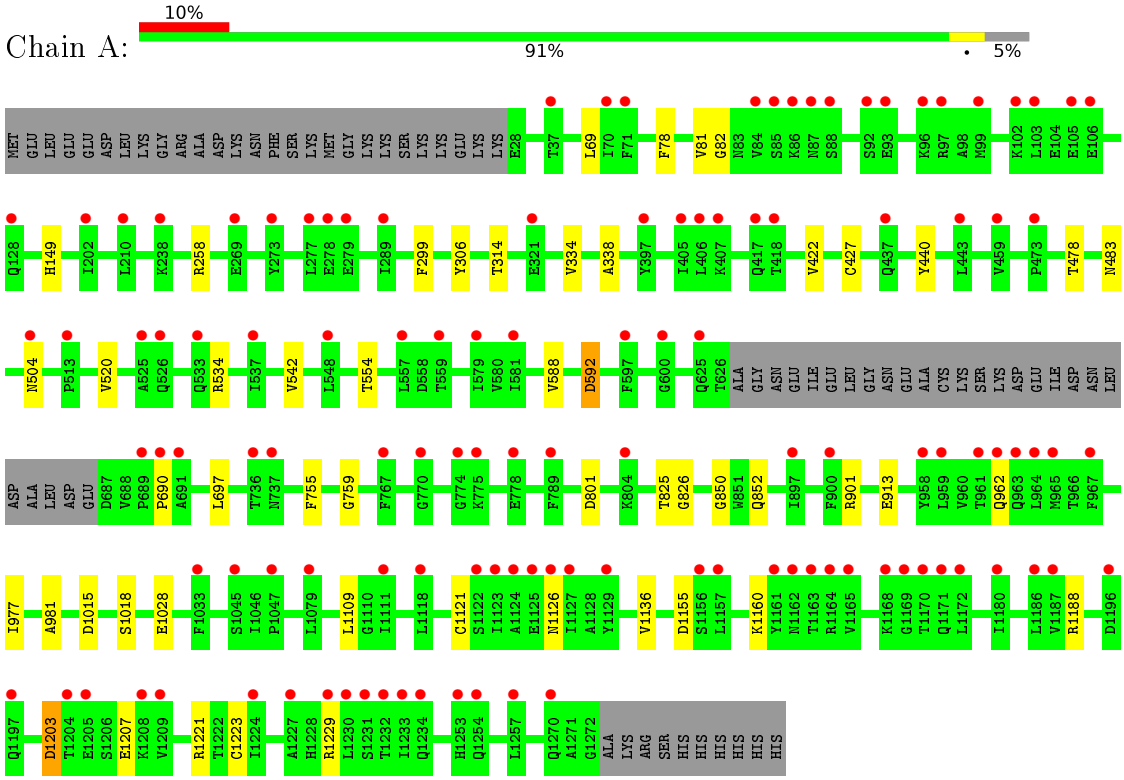
- Molecule 2 is MERCURY (II) ION (three-letter code: HG) (formula: Hg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	7	Total Hg 7 7	0	0
2	A	7	Total Hg 7 7	0	0

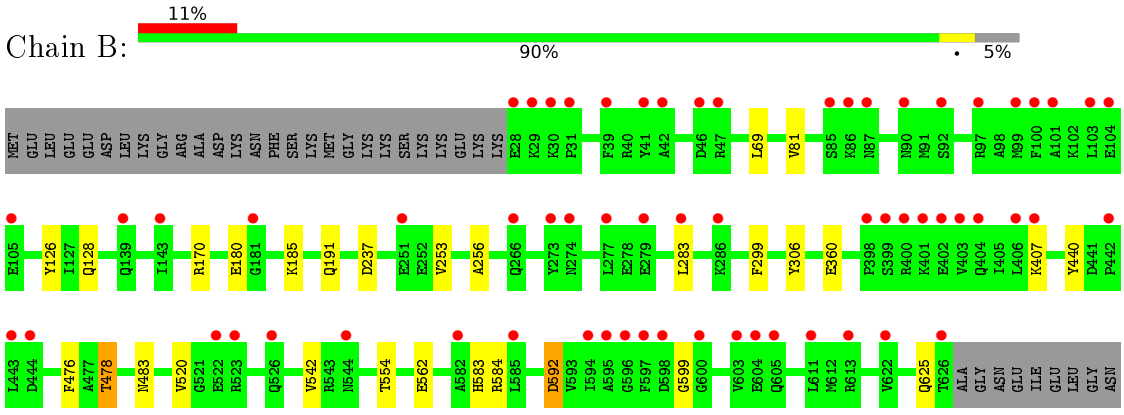
3 Residue-property plots (i)

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: Multidrug resistance protein 1A



- Molecule 1: Multidrug resistance protein 1A




GLU ALA CYS LYS SER LYS ASP GLU ILE ASP ASN LEU ASP ALA LEU ASP GLU D687 F693 L697 N700 F707 T736 P741 Q744 N747 F751 R785 Y786 L814 A815 N816 D817 G850 W851 Q852 L857 L875 Q878 R901 N928 F934
S939 Q942 Y949 L964 E968 L972 I977 A981 A991 D1015 S1018 L1022 E1028 Q1050 L1062 L1079 V1090 E1096 I1097 L1109 Q1114 I1123 N1132 S1133 R1134 V1135 V1136 S1137 E1140 A1148 N1149 I1150 H1151 Q1152 F1153 I1154
D1155 S1156 L1157 P1158 D1159 K1160 Y1161 N1162 T1163 R1164 V1165 K1168 G1169 T1170 Q1171 R1188 Q1189 P1190 H1191 I1192 L1193 L1194 L1195 D1196 T1199 D1203 T1204 E1207 K1208 V1209 V1210 Q1211 R1221 T1222 C1223 I1224 V1225 I1226 A1227 H1228 R1229 L1230 S1231 T1232 I1233 K1248 Q1254 Y1263 V1269 Q1270
A1271 GLY ALA LYS ARG SER HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	98.19Å 114.74Å 375.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.54 – 3.30 49.11 – 3.29	Depositor EDS
% Data completeness (in resolution range)	92.2 (33.54-3.30) 90.5 (49.11-3.29)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.37 (at 3.33Å)	Xtriage
Refinement program	PHENIX (dev_2443: ???)	Depositor
R, R_{free}	0.243 , 0.285 0.256 , 0.288	Depositor DCC
R_{free} test set	1977 reflections (3.29%)	DCC
Wilson B-factor (Å^2)	105.4	Xtriage
Anisotropy	0.186	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 41.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	37085	wwPDB-VP
Average B, all atoms (Å^2)	145.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.48% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HG

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	0/9358	0.50	0/12650
1	B	0.29	0/9354	0.50	0/12645
All	All	0.29	0/18712	0.50	0/25295

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	9189	9350	9367	16	0
1	B	9185	9347	9364	21	0
2	A	7	0	0	0	0
2	B	7	0	0	0	0
All	All	18388	18697	18731	37	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 1.

The worst 5 of 37 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:1134:ARG:NH1	1:B:1140:GLU:OE2	2.26	0.68
1:B:1189:GLN:OE1	1:B:1221:ARG:NH2	2.29	0.65
1:B:440:TYR:OH	1:B:901:ARG:NH2	2.30	0.65
1:A:440:TYR:OH	1:A:901:ARG:NH2	2.31	0.63
1:A:1207:GLU:OE2	1:A:1229:ARG:NH1	2.33	0.61

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	1181/1248 (95%)	1139 (96%)	37 (3%)	5 (0%)	39	76
1	B	1180/1248 (95%)	1140 (97%)	35 (3%)	5 (0%)	39	76
All	All	2361/2496 (95%)	2279 (96%)	72 (3%)	10 (0%)	39	76

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1136	VAL
1	B	1160	LYS
1	A	1160	LYS
1	B	1136	VAL
1	A	690	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	977/1031 (95%)	957 (98%)	20 (2%)	63	84
1	B	977/1031 (95%)	961 (98%)	16 (2%)	70	87
All	All	1954/2062 (95%)	1918 (98%)	36 (2%)	66	85

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1121	CYS
1	B	69	LEU
1	B	1090	VAL
1	A	1155	ASP
1	B	81	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	87	ASN
1	B	910	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 14 ligands modelled in this entry, 14 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	1185/1248 (94%)	0.70	120 (10%)	9	8	65, 119, 214, 325	0
1	B	1184/1248 (94%)	0.73	136 (11%)	6	5	62, 128, 218, 349	0
All	All	2369/2496 (94%)	0.71	256 (10%)	8	6	62, 123, 218, 349	0

The worst 5 of 256 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1170	THR	12.0
1	A	85	SER	10.2
1	A	690	PRO	9.8
1	A	87	ASN	9.0
1	A	1123	ILE	8.9

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} 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	LLDF	B-factors(Å^2)	Q<0.9
2	HG	A	1307	1/1	0.81	0.24	0.63	415,415,415,415	0
2	HG	B	1302	1/1	0.93	0.21	0.03	321,321,321,321	0
2	HG	B	1307	1/1	0.69	0.28	-0.06	536,536,536,536	0
2	HG	B	1305	1/1	0.61	0.18	-0.57	347,347,347,347	0
2	HG	A	1305	1/1	0.78	0.10	-1.09	303,303,303,303	0
2	HG	B	1304	1/1	0.93	0.15	-1.60	385,385,385,385	0
2	HG	A	1303	1/1	0.99	0.07	-1.64	127,127,127,127	0
2	HG	A	1301	1/1	0.96	0.08	-1.70	274,274,274,274	0
2	HG	B	1303	1/1	0.98	0.11	-1.89	142,142,142,142	0
2	HG	A	1304	1/1	0.93	0.13	-2.64	228,228,228,228	0
2	HG	A	1302	1/1	0.97	0.03	-3.47	286,286,286,286	0
2	HG	A	1306	1/1	0.90	0.06	-	368,368,368,368	0
2	HG	B	1306	1/1	0.79	0.07	-	414,414,414,414	0
2	HG	B	1301	1/1	0.96	0.06	-	182,182,182,182	0

6.5 Other polymers

There are no such residues in this entry.