Full wwPDB NMR Structure Validation Report (i)

Jun 16, 2024 – 08:20 PM EDT

PDB ID	:	5KZO
BMRB ID	:	30147
Title	:	Notch1 transmembrane and associated juxtamembrane segment
Authors	:	Deatherage, C.L.; Lu, Z.; Kroncke, B.
Deposited on	:	2016-07-25

This is a Full wwPDB NMR 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

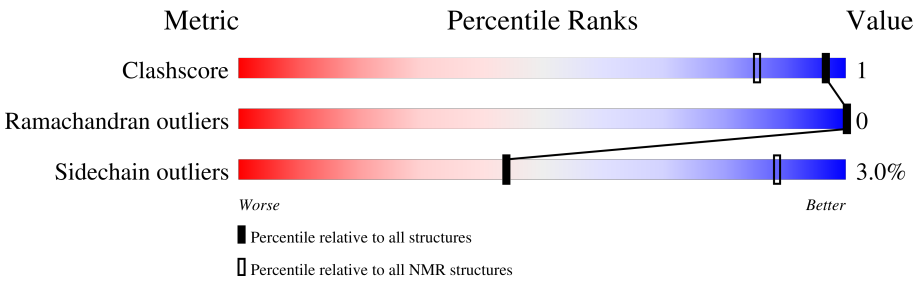
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 34%.

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)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	A	59	37% 8% 41% 14%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 3 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:1732-A:1758 (27)	0.54	3

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 7, 8, 9
2	6, 10

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 825 atoms, of which 412 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Neurogenic locus notch homolog protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	51	825	272	412	73	66	2	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P46531
A	2	GLY	-	expression tag	UNP P46531
A	3	HIS	-	expression tag	UNP P46531
A	4	HIS	-	expression tag	UNP P46531
A	5	HIS	-	expression tag	UNP P46531
A	6	HIS	-	expression tag	UNP P46531
A	7	HIS	-	expression tag	UNP P46531
A	8	HIS	-	expression tag	UNP P46531

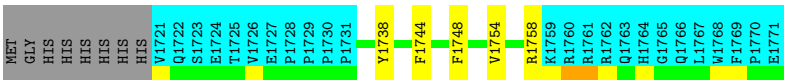
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 37% 8% 41% 14%



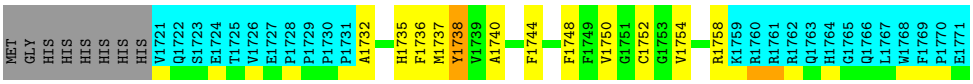
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Neurogenic locus notch homolog protein 1

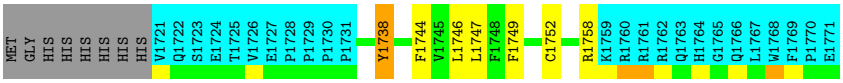
Chain A: 25% 19% • 41% 14%



4.2.2 Score per residue for model 2

• Molecule 1: Neurogenic locus notch homolog protein 1

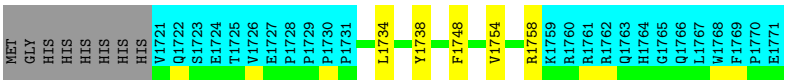
Chain A: 34% 10% • 41% 14%



4.2.3 Score per residue for model 3 (medoid)

• Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 37% 8% 41% 14%



4.2.4 Score per residue for model 4

• Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 32% 14% 41% 14%

4.2.5 Score per residue for model 5

• Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 25% 20% 41% 14%

4.2.6 Score per residue for model 6

• Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 32% 10% • 41% 14%

4.2.7 Score per residue for model 7

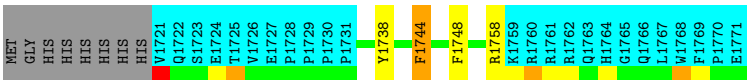
• Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 34% 10% • 41% 14%

4.2.8 Score per residue for model 8

- Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 39% 5% • 41% 14%



4.2.9 Score per residue for model 9

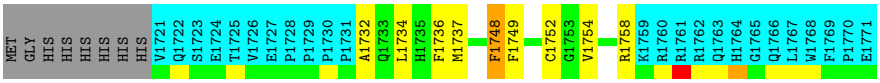
- Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 31% 14% • 41% 14%

4.2.10 Score per residue for model 10

- Molecule 1: Neurogenic locus notch homolog protein 1

Chain A: 31% 14% • 41% 14%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 2000 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	228
Number of shifts mapped to atoms	225
Number of unparsed shifts	0
Number of shifts with mapping errors	3
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	34%

6 Model quality (i)

6.1 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 (average) 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	1.54±0.09	1±1/213 (0.6± 0.6%)	2.10±0.13	9±3/289 (3.0± 1.2%)
All	All	1.55	13/2130 (0.6%)	2.11	86/2890 (3.0%)

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	Planarity
1	A	0.0±0.0	0.9±0.7
All	All	0	9

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	1736	PHE	CE2-CZ	6.71	1.50	1.37	10	2
1	A	1738	TYR	CG-CD2	6.30	1.47	1.39	1	2
1	A	1738	TYR	CD2-CE2	5.91	1.48	1.39	4	1
1	A	1738	TYR	CE1-CZ	5.76	1.46	1.38	3	2
1	A	1753	GLY	CA-C	5.62	1.60	1.51	5	1
1	A	1748	PHE	CG-CD1	5.42	1.46	1.38	4	1
1	A	1736	PHE	CG-CD2	5.41	1.46	1.38	1	1
1	A	1738	TYR	CE2-CZ	5.20	1.45	1.38	6	1
1	A	1736	PHE	CG-CD1	5.15	1.46	1.38	10	1
1	A	1751	GLY	CA-C	5.06	1.59	1.51	4	1

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	1758	ARG	NE-CZ-NH2	14.75	127.67	120.30	5	9
1	A	1738	TYR	CB-CG-CD2	-9.29	115.42	121.00	4	5
1	A	1744	PHE	CB-CG-CD2	9.22	127.26	120.80	6	3
1	A	1758	ARG	NH1-CZ-NH2	-8.70	109.83	119.40	5	2
1	A	1738	TYR	CG-CD1-CE1	-8.00	114.90	121.30	5	1
1	A	1738	TYR	CB-CG-CD1	-7.92	116.25	121.00	5	3
1	A	1744	PHE	CB-CG-CD1	-7.73	115.39	120.80	6	4
1	A	1748	PHE	CB-CG-CD1	7.71	126.20	120.80	6	3
1	A	1749	PHE	CB-CG-CD2	7.55	126.09	120.80	7	1
1	A	1758	ARG	NE-CZ-NH1	-7.54	116.53	120.30	8	5
1	A	1736	PHE	CB-CG-CD2	-7.29	115.69	120.80	1	1
1	A	1750	VAL	CG1-CB-CG2	-7.27	99.28	110.90	5	1
1	A	1736	PHE	CB-CG-CD1	-7.17	115.78	120.80	9	2
1	A	1754	VAL	CA-CB-CG1	7.17	121.65	110.90	9	3
1	A	1754	VAL	CG1-CB-CG2	-7.17	99.44	110.90	9	1
1	A	1748	PHE	CB-CG-CD2	-7.14	115.80	120.80	6	5
1	A	1749	PHE	CB-CG-CD1	-6.79	116.05	120.80	7	2
1	A	1758	ARG	CD-NE-CZ	6.61	132.85	123.60	9	2
1	A	1747	LEU	CB-CG-CD1	-6.39	100.13	111.00	6	2
1	A	1754	VAL	O-C-N	6.31	132.80	122.70	10	1
1	A	1752	CYS	CA-CB-SG	-6.26	102.74	114.00	1	2
1	A	1745	VAL	CA-CB-CG2	6.16	120.14	110.90	4	1
1	A	1754	VAL	CA-CB-CG2	6.12	120.08	110.90	7	1
1	A	1738	TYR	CD1-CG-CD2	6.00	124.50	117.90	5	1
1	A	1738	TYR	CZ-CE2-CD2	-5.89	114.50	119.80	7	1
1	A	1748	PHE	N-CA-CB	5.85	121.13	110.60	10	1
1	A	1735	HIS	CA-CB-CG	5.80	123.47	113.60	7	1
1	A	1733	GLN	CB-CG-CD	5.76	126.57	111.60	5	1
1	A	1732	ALA	CB-CA-C	5.76	118.74	110.10	10	2
1	A	1750	VAL	CA-CB-CG1	5.75	119.53	110.90	1	4
1	A	1738	TYR	CD1-CE1-CZ	5.66	124.89	119.80	2	1
1	A	1747	LEU	CB-CA-C	5.63	120.90	110.20	2	1
1	A	1749	PHE	CB-CA-C	5.62	121.65	110.40	2	1
1	A	1734	LEU	CB-CG-CD1	-5.53	101.60	111.00	3	2
1	A	1737	MET	CA-CB-CG	5.52	122.68	113.30	1	2
1	A	1740	ALA	O-C-N	-5.48	113.93	122.70	1	1
1	A	1740	ALA	N-CA-CB	-5.28	102.71	110.10	6	1
1	A	1748	PHE	CG-CD2-CE2	-5.17	115.11	120.80	3	1
1	A	1750	VAL	CA-C-N	5.12	126.43	116.20	6	1
1	A	1746	LEU	CB-CG-CD1	5.07	119.62	111.00	2	1
1	A	1738	TYR	CG-CD2-CE2	-5.05	117.26	121.30	1	1
1	A	1745	VAL	CA-CB-CG1	5.04	118.46	110.90	5	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	1737	MET	N-CA-CB	5.03	119.66	110.60	10	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	1738	TYR	Sidechain	3
1	A	1744	PHE	Sidechain	3
1	A	1749	PHE	Mainchain,Sidechain	2
1	A	1735	HIS	Sidechain	1

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	207	214	214	0±0
All	All	2070	2140	2140	3

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.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1735:HIS:O	1:A:1739:VAL:HG23	0.57	1.99	4	1
1:A:1748:PHE:O	1:A:1752:CYS:HB3	0.41	2.15	10	1
1:A:1755:LEU:HD12	1:A:1756:LEU:N	0.41	2.31	5	1

6.3 Torsion angles (i)

6.3.1 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 PDB entries followed by that with respect to all NMR

entries. 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	27/59 (46%)	25±1 (93±3%)	2±1 (7±3%)	0±0 (0±0%)	100	100
All	All	270/590 (46%)	252 (93%)	18 (7%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	20/50 (40%)	19±0 (97±2%)	1±0 (3±2%)	44	89
All	All	200/500 (40%)	194 (97%)	6 (3%)	44	89

All 3 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	1755	LEU	3
1	A	1754	VAL	2
1	A	1735	HIS	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 34% for the well-defined parts and 31% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	228
Number of shifts mapped to atoms	225
Number of unparsed shifts	0
Number of shifts with mapping errors	3
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 3 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	8	HIS	C	174.2291	0.25	1
1	A	8	HIS	CA	55.6598	0.25	1
1	A	8	HIS	CB	29.6729	0.25	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	48	-0.38 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	44	0.23 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	45	0.14 ± 0.09	None needed (< 0.5 ppm)
^{15}N	46	0.01 ± 0.27	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 34%, i.e. 129 atoms were assigned a chemical shift out of a possible 380. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	106/137 (77%)	27/56 (48%)	52/54 (96%)	27/27 (100%)
Sidechain	23/187 (12%)	0/128 (0%)	23/55 (42%)	0/4 (0%)
Aromatic	0/56 (0%)	0/28 (0%)	0/27 (0%)	0/1 (0%)
Overall	129/380 (34%)	27/212 (13%)	75/136 (55%)	27/32 (84%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 31%, i.e. 225 atoms were assigned a chemical shift out of a possible 734. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	182/248 (73%)	45/100 (45%)	91/102 (89%)	46/46 (100%)
Sidechain	43/401 (11%)	0/264 (0%)	43/120 (36%)	0/17 (0%)
Aromatic	0/85 (0%)	0/43 (0%)	0/39 (0%)	0/3 (0%)
Overall	225/734 (31%)	45/407 (11%)	134/261 (51%)	46/66 (70%)

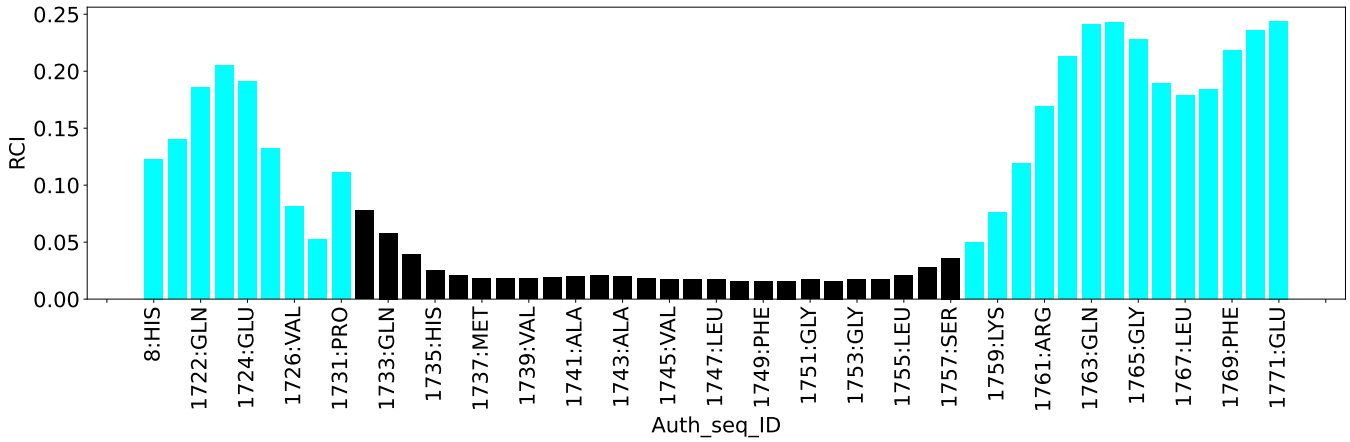
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:


0.25
0.20
0.15
0.10
0.05
0.00
RCI
8:HIS
1722:GLN
1724:GLU
1726:VAL
1731:PRO
1733:GLN
1735:HIS
1737:MET
1739:VAL
1741:ALA
1743:ALA
1745:VAL
1747:LEU
1749:PHE
1751:GLY
1753:GLY
1755:LEU
1757:SER
1759:LYS
1761:ARG
1763:GLN
1765:GLY
1767:LEU
1769:PHE
1771:GLU
Auth_seq_ID