Full wwPDB NMR Structure Validation Report (i)

Apr 16, 2023 – 06:59 AM EDT

PDB ID	:	2L92
BMRB ID	:	17434
Title	:	Solution structure of the C-terminal domain of H-NS like protein Bv3F
Authors	:	Li, Y.; Xia, B.
Deposited on	:	2011-01-29

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:

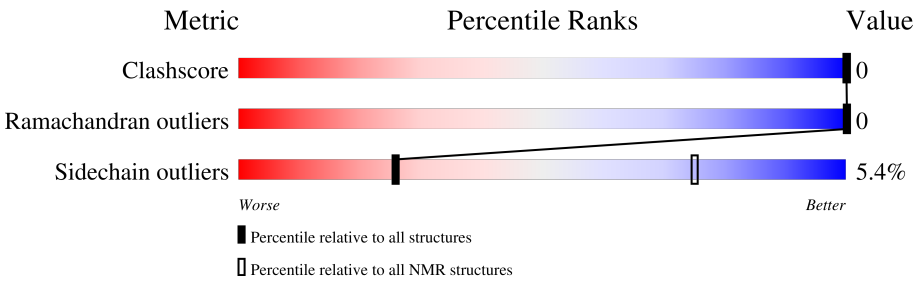
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.2

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 84%.

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	50	26% • 60% 12%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *minimized 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:76-A:87, A:103-A:104 (14)	0.08	19

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 5 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 4, 8, 11, 16, 18, 19
2	2, 5, 9, 12, 15, 17
3	10, 14, 20
4	3, 13
5	6, 7

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 679 atoms, of which 339 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone family protein nucleoid-structuring protein H-NS.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	44	679	218	339	60	61	1	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	113	LEU	-	expression tag	UNP A4JS72
A	114	GLU	-	expression tag	UNP A4JS72
A	115	HIS	-	expression tag	UNP A4JS72
A	116	HIS	-	expression tag	UNP A4JS72
A	117	HIS	-	expression tag	UNP A4JS72
A	118	HIS	-	expression tag	UNP A4JS72
A	119	HIS	-	expression tag	UNP A4JS72
A	120	HIS	-	expression tag	UNP A4JS72

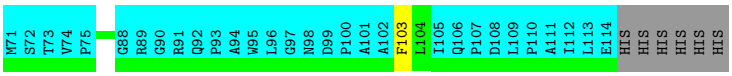
4 Residue-property plots (i)

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: Histone family protein nucleoid-structuring protein H-NS

Chain A:



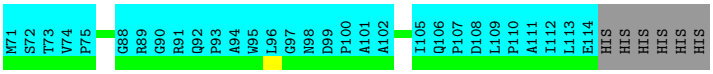
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

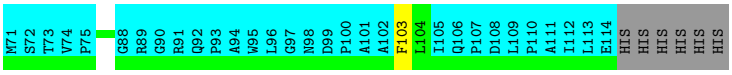
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 24% • 60% 12%

M71 S72 T73 V74 P75 R78 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.4 Score per residue for model 4

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 28% 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.5 Score per residue for model 5

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.6 Score per residue for model 6

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.7 Score per residue for model 7

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 28% 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.8 Score per residue for model 8

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.9 Score per residue for model 9

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.10 Score per residue for model 10

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.11 Score per residue for model 11

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.12 Score per residue for model 12

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.13 Score per residue for model 13

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.14 Score per residue for model 14

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.15 Score per residue for model 15

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.16 Score per residue for model 16

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.17 Score per residue for model 17

• Molecule 1: Histone family protein nucleoid-structuring protein H-NS

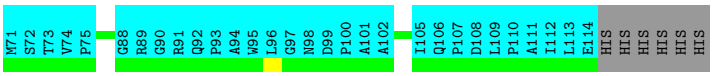
Chain A: 26% • 60% 12%

M71 S72 T73 V74 P75 G88 R89 G90 R91 Q92 P93 A94 W95 L96 G97 N98 D99 P100 A101 A102 F103 L104 I105 Q106 P107 D108 L109 P110 A111 I112 L113 E114 HIS HIS HIS HIS HIS HIS

4.2.18 Score per residue for model 18

- Molecule 1: Histone family protein nucleoid-structuring protein H-NS

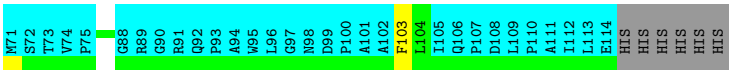
Chain A: 28% 60% 12%



4.2.19 Score per residue for model 19 (medoid)

- Molecule 1: Histone family protein nucleoid-structuring protein H-NS

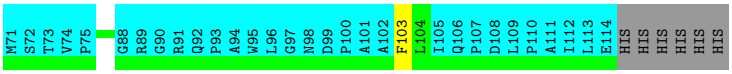
Chain A: 26% • 60% 12%



4.2.20 Score per residue for model 20

- Molecule 1: Histone family protein nucleoid-structuring protein H-NS

Chain A: 26% • 60% 12%



5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 20 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
Amber	geometry optimization	
Amber	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	471
Number of shifts mapped to atoms	471
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

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	0.82±0.01	0±0/123 (0.0± 0.0%)	1.06±0.02	0±0/167 (0.0± 0.1%)
All	All	0.82	0/2460 (0.0%)	1.06	1/3340 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	78	ARG	NE-CZ-NH1	5.47	123.03	120.30	3	1

There are no chirality outliers.

There are no planarity outliers.

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	118	116	116	0±0
All	All	2360	2320	2320	-

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 -.

There are no clashes.

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	14/50 (28%)	14±1 (96±5%)	0±1 (4±5%)	0±0 (0±0%)	100	100
All	All	280/1000 (28%)	270 (96%)	10 (4%)	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	12/41 (29%)	11±0 (95±4%)	1±0 (5±4%)	26	75
All	All	240/820 (29%)	227 (95%)	13 (5%)	26	75

All 1 unique residues with a non-rotameric sidechain are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	103	PHE	13

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 84% for the well-defined parts and 80% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *Bv3F_fin_2.str*

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	471
Number of shifts mapped to atoms	471
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	9

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	44	-0.03 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	40	0.08 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	35	-0.18 ± 0.16	None needed (< 0.5 ppm)
^{15}N	35	-0.04 ± 0.76	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 84%, i.e. 166 atoms were assigned a chemical shift out of a possible 198. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	64/69 (93%)	27/28 (96%)	25/28 (89%)	12/13 (92%)
Sidechain	72/98 (73%)	44/63 (70%)	28/30 (93%)	0/5 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	30/31 (97%)	15/15 (100%)	14/15 (93%)	1/1 (100%)
Overall	166/198 (84%)	86/106 (81%)	67/73 (92%)	13/19 (68%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 80%, i.e. 471 atoms were assigned a chemical shift out of a possible 589. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	197/212 (93%)	83/86 (97%)	79/88 (90%)	35/38 (92%)
Sidechain	233/334 (70%)	139/218 (64%)	91/102 (89%)	3/14 (21%)
Aromatic	41/43 (95%)	21/21 (100%)	18/20 (90%)	2/2 (100%)
Overall	471/589 (80%)	243/325 (75%)	188/210 (90%)	40/54 (74%)

7.1.4 Statistically unusual chemical shifts (i)

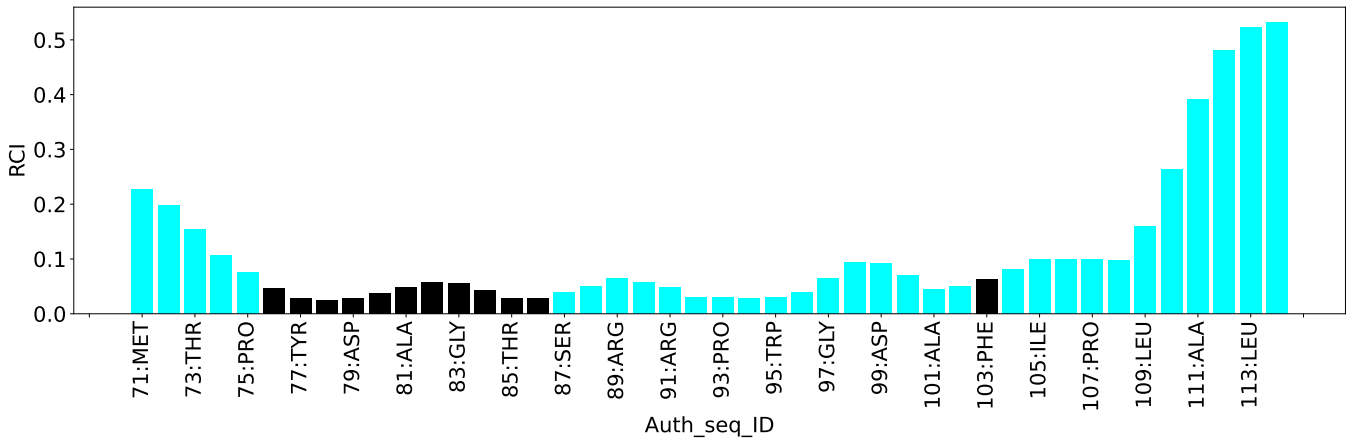
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	112	ILE	CD1	64.87	5.18 – 21.60	31.4
1	A	103	PHE	CE2	118.52	124.80 – 136.72	-10.3
1	A	103	PHE	CE1	118.52	124.17 – 137.29	-9.3
1	A	86	TRP	CZ2	125.32	107.20 – 121.33	7.8
1	A	86	TRP	CE3	133.99	111.58 – 129.41	7.6
1	A	88	GLY	HA3	1.30	2.08 – 5.71	-7.2
1	A	95	TRP	CD1	114.77	117.34 – 135.80	-6.4
1	A	93	PRO	HD3	1.39	1.76 – 5.48	-6.0
1	A	103	PHE	CZ	121.57	121.82 – 136.66	-5.2

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:


RCI
0.0
0.1
0.2
0.3
0.4
0.5
71:MET
73:THR
75:PRO
77:TYR
79:ASP
81:ALA
83:GLY
85:THR
87:SER
89:ARG
91:ARG
93:PRO
95:TRP
97:GLY
99:ASP
101:ALA
103:PHE
105:ILE
107:PRO
109:LEU
111:ALA
113:LEU
Auth_seq_ID