Full wwPDB NMR Structure Validation Report (i)

May 29, 2020 – 04:10 am BST

PDB ID	:	3ZOB
Title	:	Solution structure of chicken Engrailed 2 homeodomain
Authors	:	Carlier, L.; Balayssac, S.; Cantrelle, F.X.; Khemtemourian, L.; Chassaing, G.; Joliot, A.; Lequin, O.
Deposited on	:	2013-02-21

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

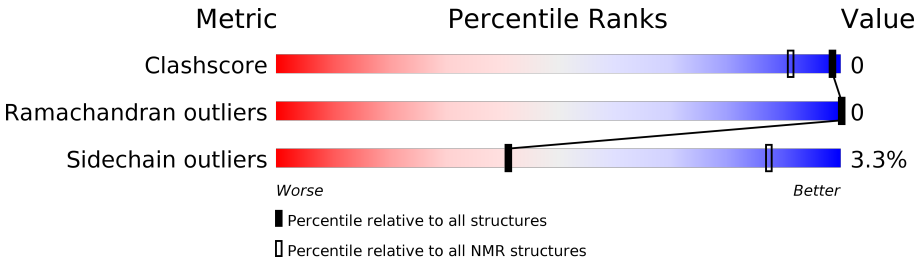
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance

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

The overall completeness of chemical shifts assignment is 95%.

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	67	67% 33%

2 Ensemble composition and analysis

This entry contains 20 models. Model 15 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:13-A:57 (45)	0.37	15

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 and 3 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1125 atoms, of which 572 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called HOMEOBOX PROTEIN ENGRAILED-2.

Mol	Chain	Residues	Atoms						Trace
1	A	67	Total	C	H	N	O	S	0
			1125	343	572	108	101	1	

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP Q05917
A	2	PRO	-	expression tag	UNP Q05917
A	3	MET	-	expression tag	UNP Q05917
A	4	ALA	-	expression tag	UNP Q05917
A	5	SER	-	expression tag	UNP Q05917
A	66	GLN	-	expression tag	UNP Q05917
A	67	ALA	-	expression tag	UNP Q05917

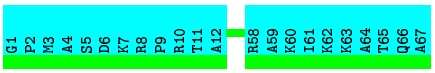
4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: HOMEOBOX PROTEIN ENGRAILED-2

Chain A:



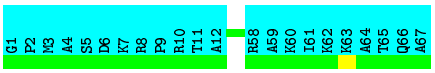
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

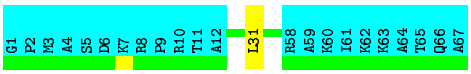
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

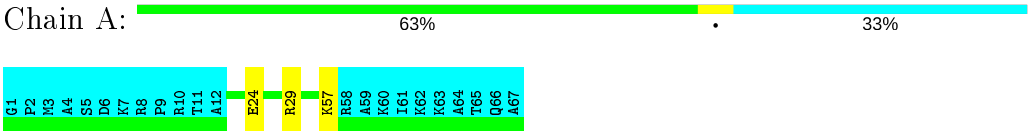
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

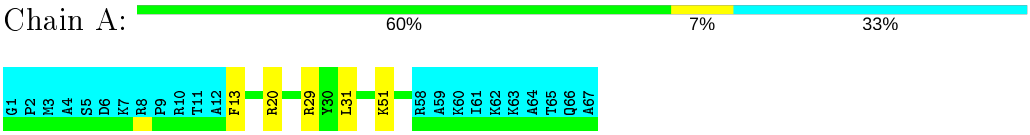
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

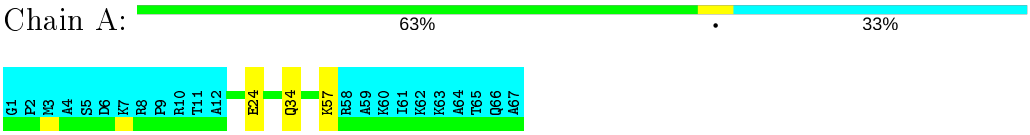
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

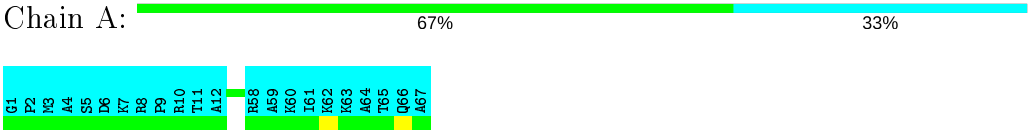
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

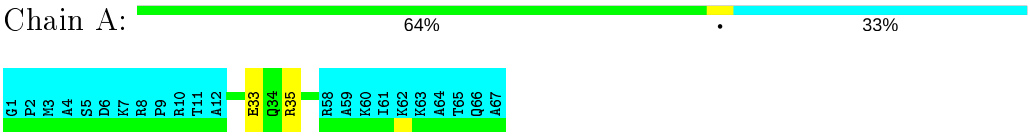
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

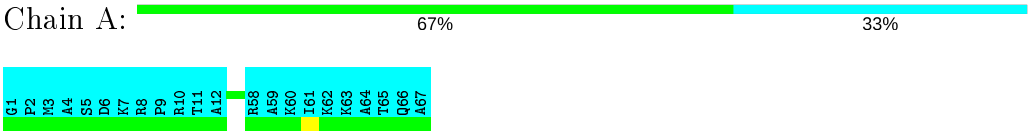
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

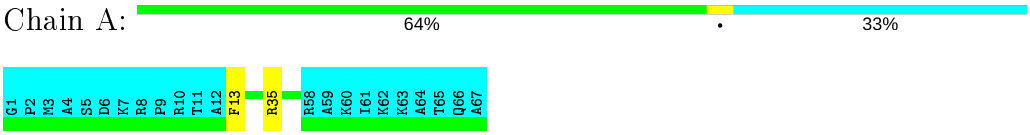
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

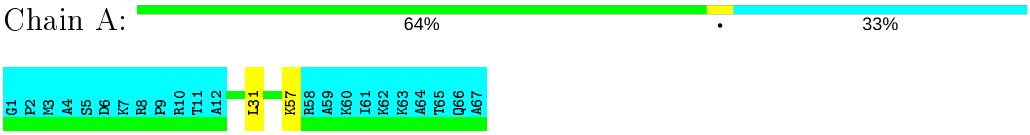
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

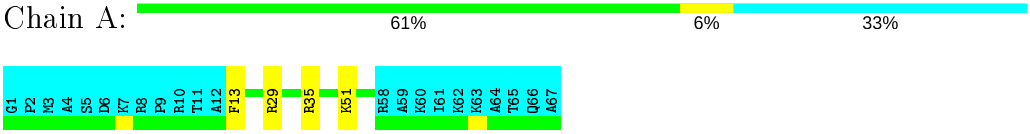
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

Chain A:



4.2.12 Score per residue for model 12

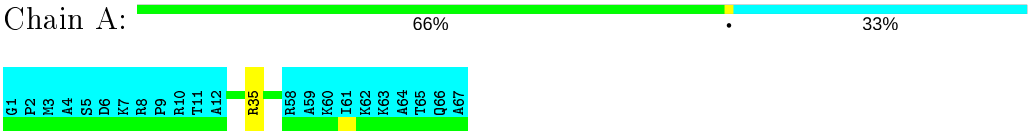
• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

Chain A:

4.2.13 Score per residue for model 13

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

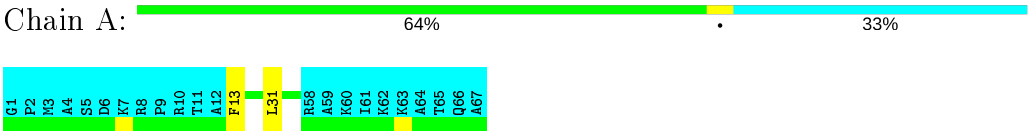
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

Chain A:



4.2.15 Score per residue for model 15 (medoid)

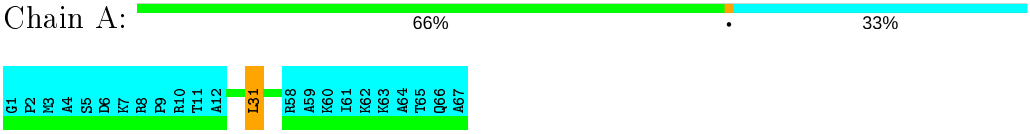
• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

Chain A:

4.2.16 Score per residue for model 16

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

Chain A:



4.2.17 Score per residue for model 17

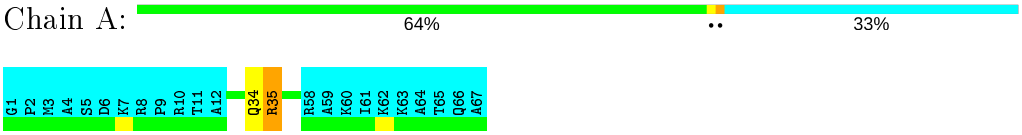
• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

Chain A:

4.2.18 Score per residue for model 18

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

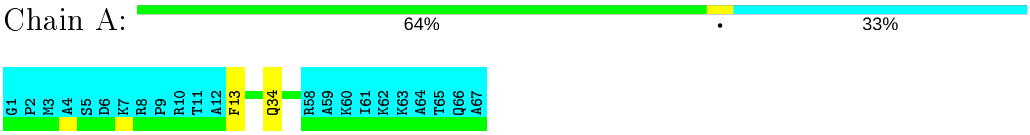
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

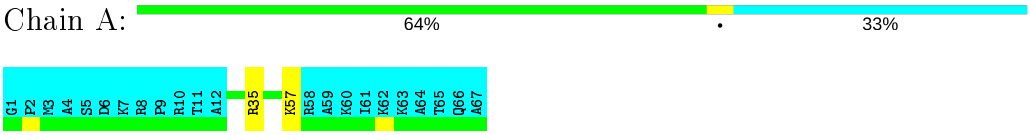
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: HOMEOBOX PROTEIN ENGRAILED-2

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *LOWEST ENERGY*.

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

Software name	Classification	Version
XPLOR-NIH	refinement	
DYANA	structure solution	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	924
Number of shifts mapped to atoms	924
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	95%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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 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	387	386	386	0±0
All	All	7740	7720	7720	1

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:31:LEU:H	1:A:31:LEU:HD22	0.41	1.76	16	1

5.2 Torsion angles (i)

5.2.1 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	45/67 (67%)	42±1 (93±2%)	3±1 (7±2%)	0±0 (0±0%)	100	100
All	All	900/1340 (67%)	837 (93%)	63 (7%)	0 (0%)	100	100

There are no Ramachandran outliers.

5.2.2 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	41/57 (72%)	40±1 (97±3%)	1±1 (3±3%)	41	87
All	All	820/1140 (72%)	793 (97%)	27 (3%)	41	87

All 9 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	31	LEU	5
1	A	29	ARG	4
1	A	57	LYS	4
1	A	20	ARG	3
1	A	34	GLN	3
1	A	35	ARG	3
1	A	24	GLU	2
1	A	51	LYS	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	33	GLU	1

5.2.3 RNA

There are no RNA molecules in this entry.

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

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

5.4 Carbohydrates

There are no carbohydrates in this entry.

5.5 Ligand geometry

There are no ligands in this entry.

5.6 Other polymers

There are no such molecules in this entry.

5.7 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 95% for the well-defined parts and 93% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *En2HD_1.bmrb.csh*

6.1.1 Bookkeeping

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	924
Number of shifts mapped to atoms	924
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	66	-0.26 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	64	0.15 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	66	-0.44 ± 0.14	None needed (< 0.5 ppm)
^{15}N	64	-0.16 ± 0.18	None needed (< 0.5 ppm)

6.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 95%, i.e. 598 atoms were assigned a chemical shift out of a possible 630. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	225/225 (100%)	90/90 (100%)	90/90 (100%)	45/45 (100%)
Sidechain	326/358 (91%)	192/211 (91%)	118/121 (98%)	16/26 (62%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	47/47 (100%)	25/25 (100%)	21/21 (100%)	1/1 (100%)
Overall	598/630 (95%)	307/326 (94%)	229/232 (99%)	62/72 (86%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 93%, i.e. 846 atoms were assigned a chemical shift out of a possible 912. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	326/331 (98%)	130/132 (98%)	132/134 (99%)	64/65 (98%)
Sidechain	473/534 (89%)	282/317 (89%)	171/177 (97%)	20/40 (50%)
Aromatic	47/47 (100%)	25/25 (100%)	21/21 (100%)	1/1 (100%)
Overall	846/912 (93%)	437/474 (92%)	324/332 (98%)	85/106 (80%)

6.1.4 Statistically unusual chemical shifts (i)

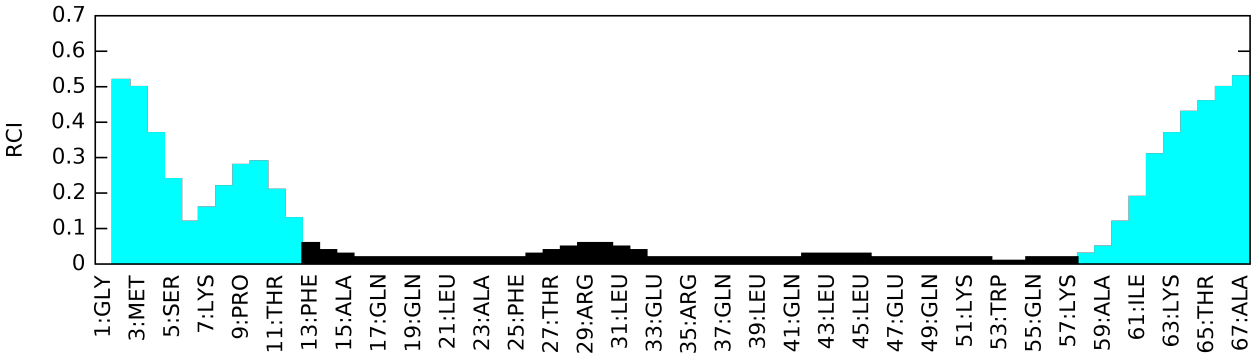
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	57	LYS	HG2	-0.24	2.67 – 0.07	-6.2
1	A	57	LYS	HG3	-0.37	2.76 – -0.04	-6.2
1	A	21	LEU	HB3	-0.37	3.34 – -0.26	-5.3
1	A	57	LYS	HD3	0.38	2.75 – 0.45	-5.3
1	A	57	LYS	HB3	0.37	3.10 – 0.40	-5.1

6.1.5 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.

Random coil index (RCI) for chain A:


RCI
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1:GLY
3:MET
5:SER
7:LYS
9:PRO
11:THR
13:PHE
15:ALA
17:GLN
19:GLN
21:LEU
23:ALA
25:PHE
27:THR
29:ARG
31:LEU
33:GLU
35:ARG
37:GLN
39:LEU
41:GLN
43:LEU
45:LEU
47:GLU
49:GLN
51:LYS
53:TRP
55:GLN
57:LYS
59:ALA
61:ILE
63:LYS
65:THR
67:ALA