



Full wwPDB NMR Structure Validation Report ⓘ

Mar 6, 2022 – 07:35 AM EST

PDB ID : 2KC0
Title : Solution structure of the factor H binding protein
Authors : Cantini, F.; Veggi, D.; Dragonetti, S.; Savino, S.; Scarselli, M.; Romagnoli, G.;
Pizza, M.; Banci, L.; Rappuoli, R.
Deposited on : 2008-12-13

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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : 2.27
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.27

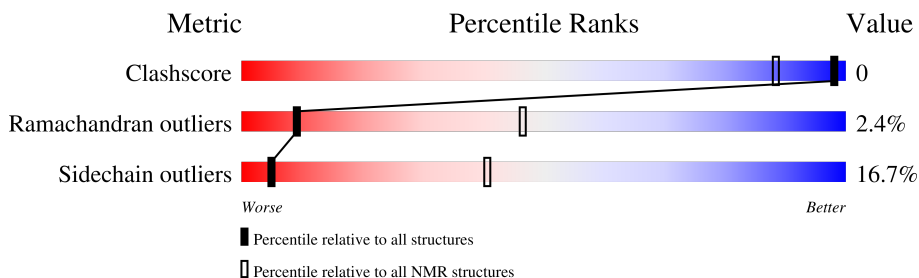
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment was not calculated.

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	255	 83% 10% • 5%

2 Ensemble composition and analysis i

This entry contains 25 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:14-A:24, (236)	A:31-A:255	0.83	4

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 4 clusters and 2 single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 5, 9, 12, 13, 14, 15, 19, 21, 24
2	6, 8, 16, 17, 20, 22, 23
3	18, 25
4	1, 7
Single-model clusters	10; 11

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 3648 atoms, of which 1820 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called lipoprotein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	242	3648	1135	1820	329	363	1	0

There are 7 discrepancies between the modelled and reference sequences:

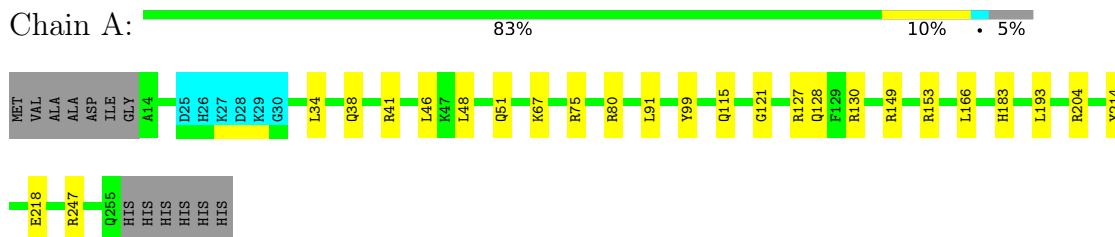
Chain	Residue	Modelled	Actual	Comment	Reference
A	7	MET	-	expression tag	UNP Q6VRZ6
A	256	HIS	-	expression tag	UNP Q6VRZ6
A	257	HIS	-	expression tag	UNP Q6VRZ6
A	258	HIS	-	expression tag	UNP Q6VRZ6
A	259	HIS	-	expression tag	UNP Q6VRZ6
A	260	HIS	-	expression tag	UNP Q6VRZ6
A	261	HIS	-	expression tag	UNP Q6VRZ6

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: lipoprotein

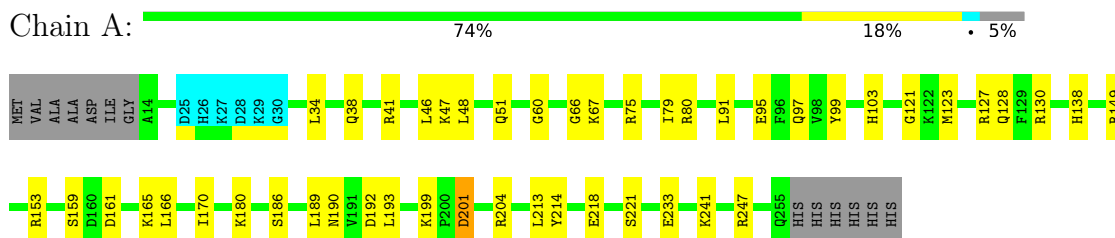


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

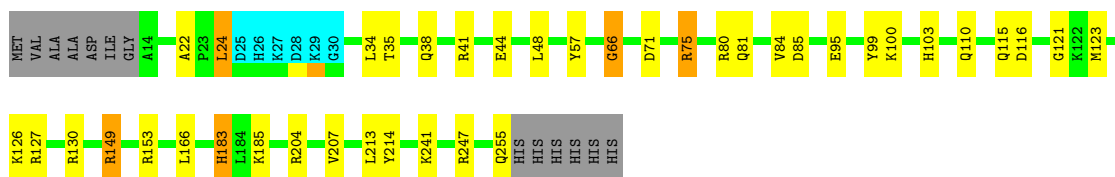
- Molecule 1: lipoprotein



4.2.2 Score per residue for model 2

- Molecule 1: lipoprotein

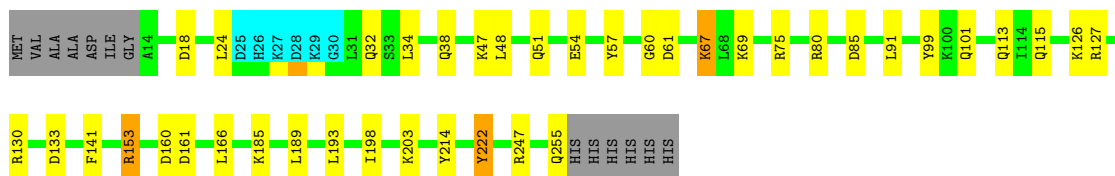




4.2.3 Score per residue for model 3

- Molecule 1: lipoprotein

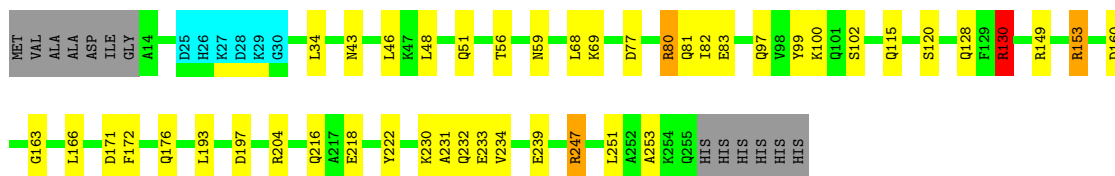
Chain A: 77% 15% 5%



4.2.4 Score per residue for model 4 (medoid)

- Molecule 1: lipoprotein

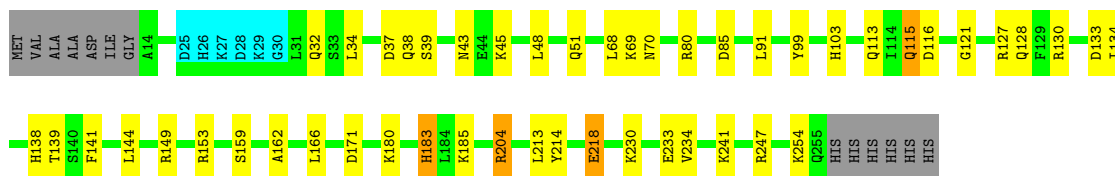
Chain A: 75% 16% 5%



4.2.5 Score per residue for model 5


- Molecule 1: lipoprotein

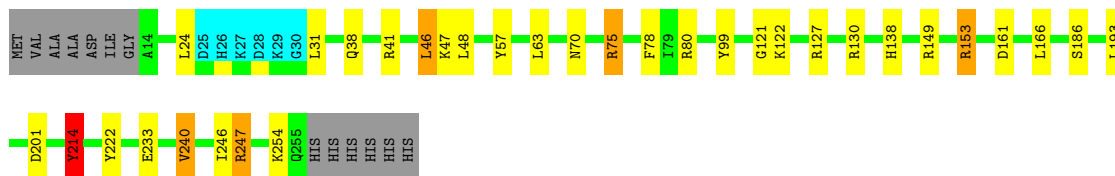
Chain A: 73% 18% 5%



4.2.6 Score per residue for model 6


- Molecule 1: lipoprotein

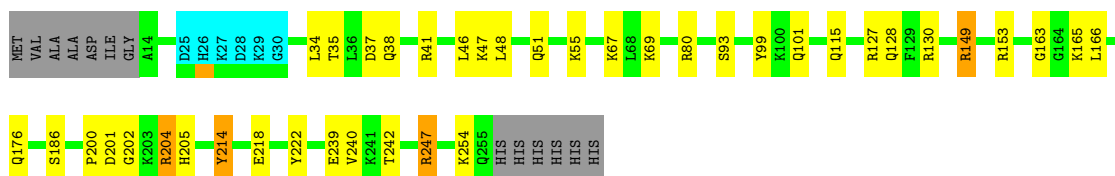
Chain A:  80% 11% 5%



4.2.7 Score per residue for model 7


- Molecule 1: lipoprotein

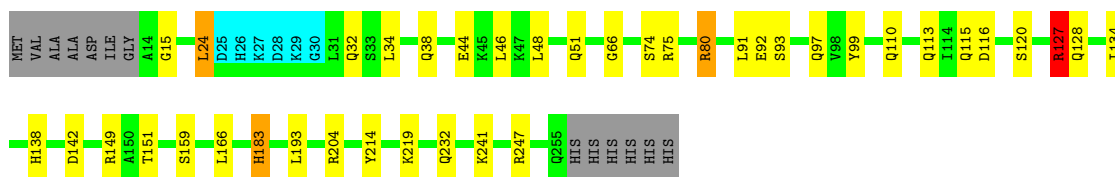
Chain A:  77% 14% 5%



4.2.8 Score per residue for model 8


- Molecule 1: lipoprotein

Chain A:  77% 14% 5%



4.2.9 Score per residue for model 9

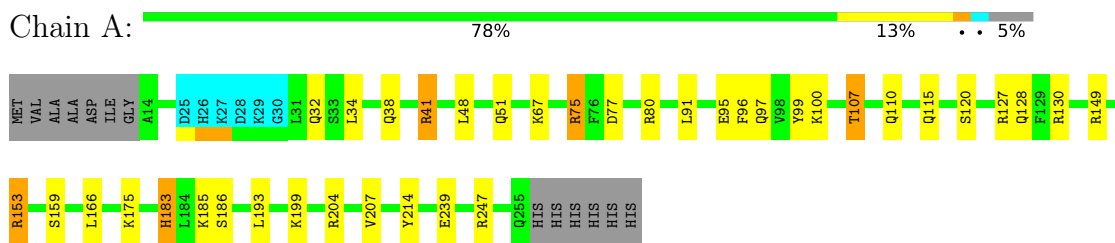
- Molecule 1: lipoprotein

Chain A:  79% 12% 5%



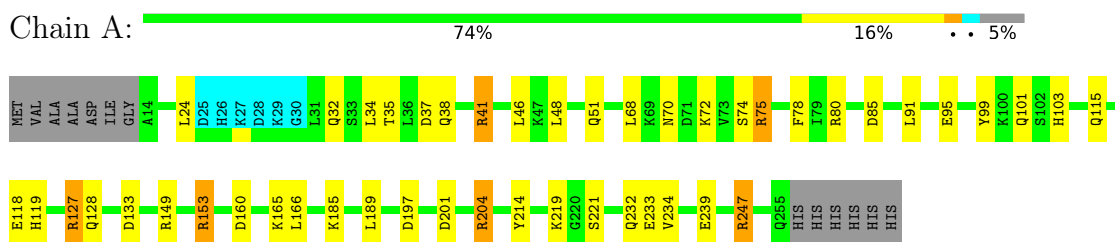
4.2.10 Score per residue for model 10

- Molecule 1: lipoprotein



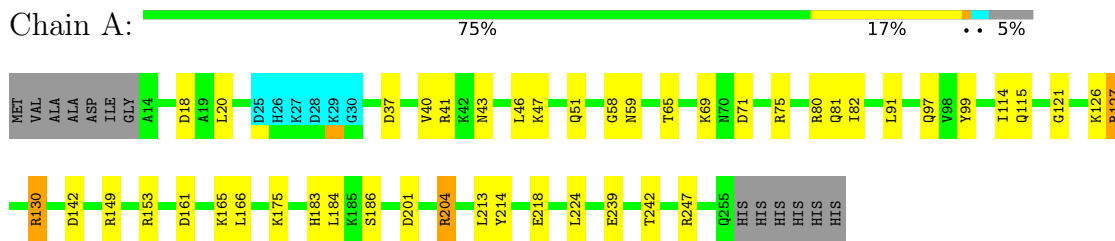
4.2.11 Score per residue for model 11

- Molecule 1: lipoprotein



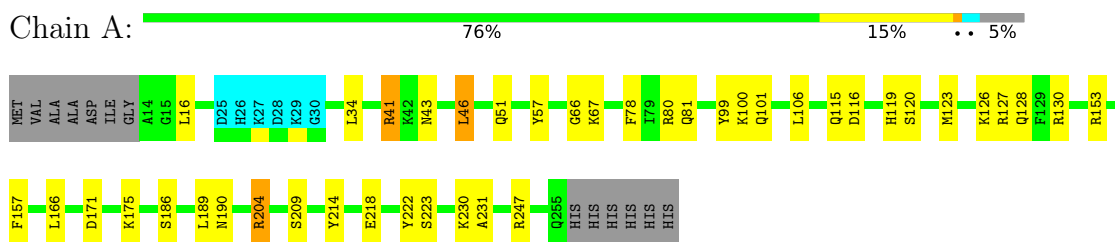
4.2.12 Score per residue for model 12

- Molecule 1: lipoprotein



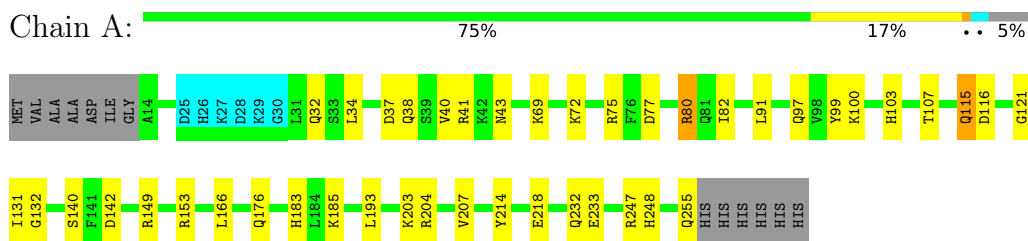
4.2.13 Score per residue for model 13

- Molecule 1: lipoprotein



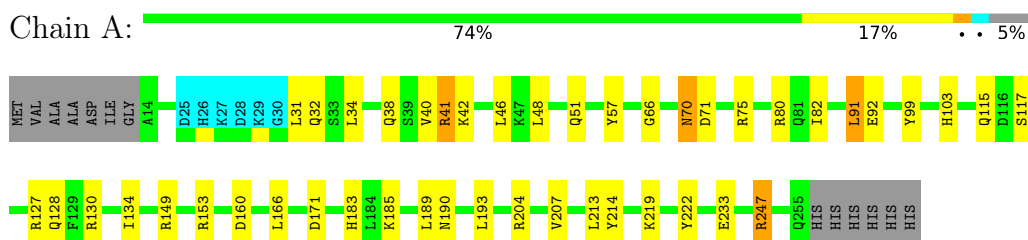
4.2.14 Score per residue for model 14

- Molecule 1: lipoprotein



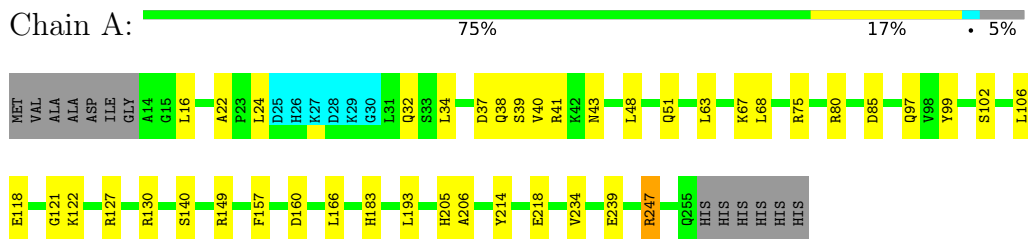
4.2.15 Score per residue for model 15

- Molecule 1: lipoprotein



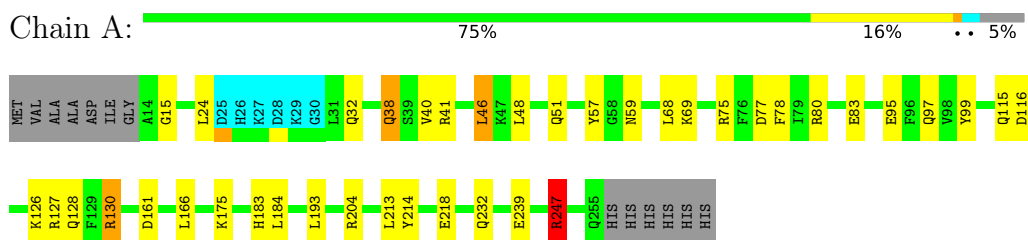
4.2.16 Score per residue for model 16

- Molecule 1: lipoprotein



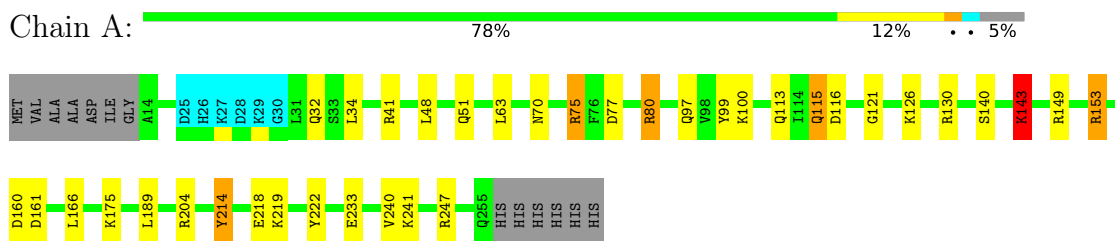
4.2.17 Score per residue for model 17

- Molecule 1: lipoprotein



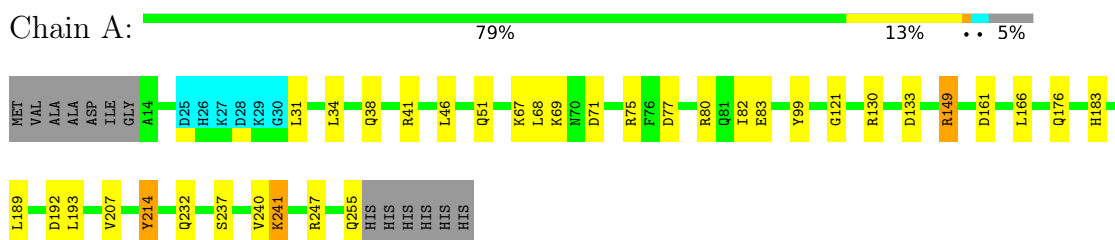
4.2.18 Score per residue for model 18

- Molecule 1: lipoprotein



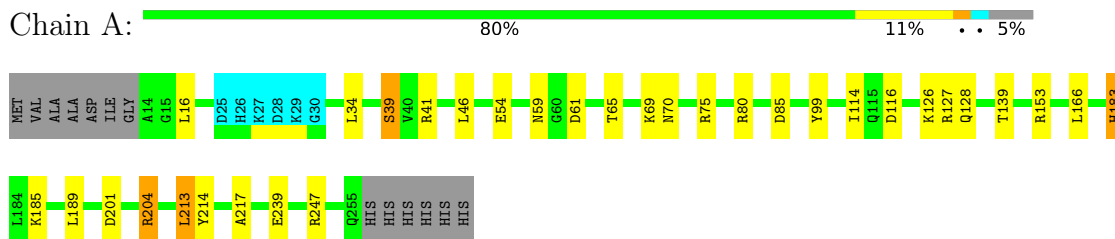
4.2.19 Score per residue for model 19

- Molecule 1: lipoprotein



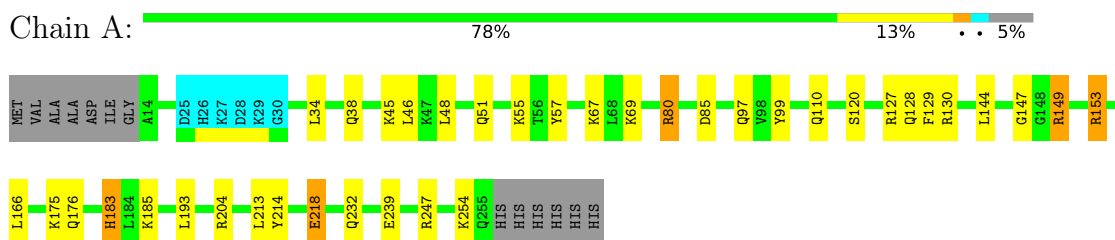
4.2.20 Score per residue for model 20

- Molecule 1: lipoprotein



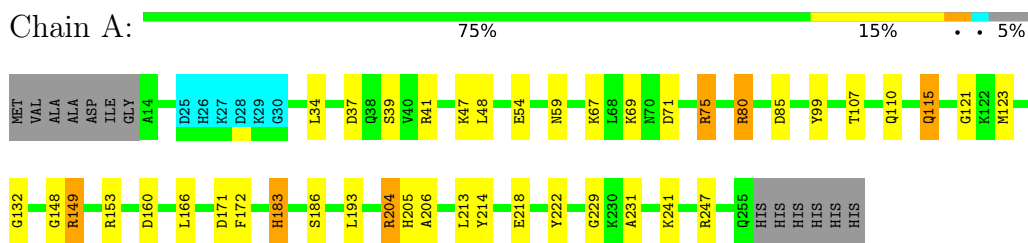
4.2.21 Score per residue for model 21

- Molecule 1: lipoprotein



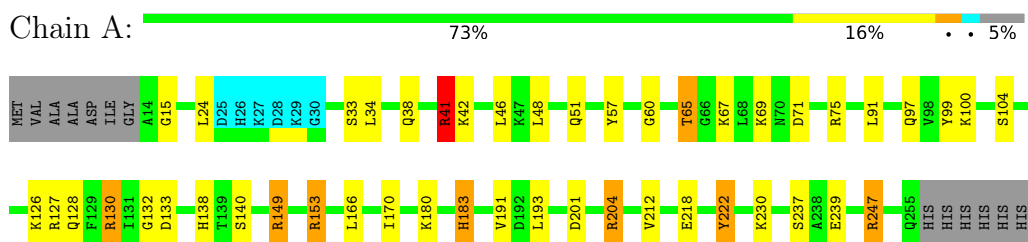
4.2.22 Score per residue for model 22

- Molecule 1: lipoprotein



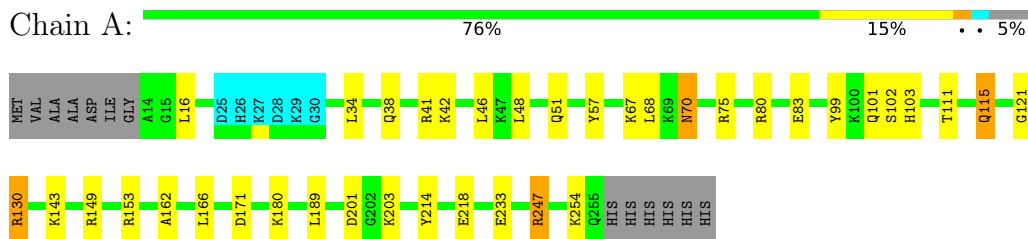
4.2.23 Score per residue for model 23

- Molecule 1: lipoprotein



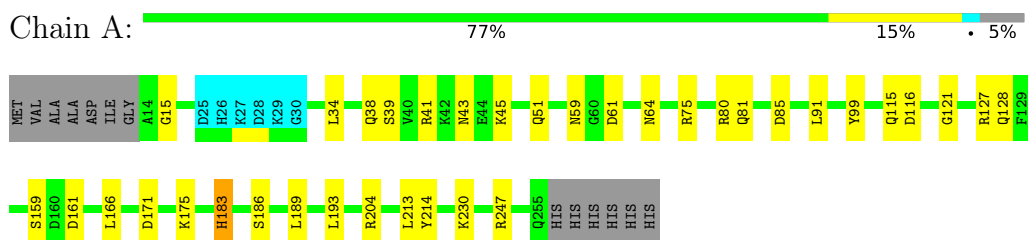
4.2.24 Score per residue for model 24

- Molecule 1: lipoprotein



4.2.25 Score per residue for model 25

- Molecule 1: lipoprotein



5 Refinement protocol and experimental data overview

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

Of the 900 calculated structures, 25 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CYANA	structure solution	
CYANA	refinement	

No chemical shift data was provided.

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	0.69±0.01	0±0/1804 (0.0± 0.0%)	1.13±0.02	8±2/2422 (0.3± 0.1%)
All	All	0.69	0/45100 (0.0%)	1.13	200/60550 (0.3%)

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	1.1±0.9
All	All	0	28

There are no bond-length outliers.

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	247	ARG	NE-CZ-NH1	10.64	125.62	120.30	23	22
1	A	80	ARG	NE-CZ-NH1	10.15	125.37	120.30	11	22
1	A	41	ARG	NE-CZ-NH1	9.75	125.17	120.30	2	17
1	A	127	ARG	NE-CZ-NH1	9.67	125.14	120.30	8	20
1	A	153	ARG	NE-CZ-NH1	9.36	124.98	120.30	13	17
1	A	75	ARG	NE-CZ-NH1	9.30	124.95	120.30	11	14
1	A	149	ARG	NE-CZ-NH1	9.22	124.91	120.30	8	16
1	A	130	ARG	NE-CZ-NH1	8.96	124.78	120.30	2	20
1	A	204	ARG	NE-CZ-NH1	8.33	124.46	120.30	10	17
1	A	127	ARG	NE-CZ-NH2	-7.95	116.32	120.30	8	4
1	A	247	ARG	CD-NE-CZ	7.01	133.41	123.60	23	1
1	A	149	ARG	NE-CZ-NH2	-6.79	116.91	120.30	22	3
1	A	80	ARG	NE-CZ-NH2	-6.72	116.94	120.30	16	3
1	A	204	ARG	NE-CZ-NH2	-6.15	117.23	120.30	20	3
1	A	115	GLN	CB-CA-C	6.15	122.69	110.40	24	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	57	TYR	CB-CG-CD2	-6.03	117.38	121.00	3	1
1	A	41	ARG	NE-CZ-NH2	-5.89	117.36	120.30	17	1
1	A	127	ARG	CD-NE-CZ	5.52	131.33	123.60	15	1
1	A	247	ARG	NE-CZ-NH2	-5.51	117.55	120.30	13	1
1	A	172	PHE	CB-CG-CD2	-5.46	116.97	120.80	4	1
1	A	75	ARG	NE-CZ-NH2	-5.42	117.59	120.30	15	2
1	A	214	TYR	CB-CG-CD1	-5.34	117.80	121.00	6	1
1	A	222	TYR	CB-CG-CD2	-5.32	117.81	121.00	23	1
1	A	115	GLN	N-CA-CB	-5.31	101.04	110.60	14	3
1	A	149	ARG	CD-NE-CZ	5.20	130.88	123.60	22	1
1	A	48	LEU	CB-CG-CD2	5.14	119.73	111.00	7	1
1	A	153	ARG	NE-CZ-NH2	-5.08	117.76	120.30	22	1
1	A	57	TYR	CB-CG-CD1	-5.03	117.98	121.00	24	1
1	A	41	ARG	CD-NE-CZ	5.02	130.62	123.60	2	1
1	A	143	LYS	CA-CB-CG	5.00	124.41	113.40	18	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	57	TYR	Sidechain	8
1	A	127	ARG	Sidechain	3
1	A	222	TYR	Sidechain	3
1	A	41	ARG	Sidechain	3
1	A	80	ARG	Sidechain	3
1	A	149	ARG	Sidechain	2
1	A	247	ARG	Sidechain	2
1	A	214	TYR	Sidechain	1
1	A	130	ARG	Sidechain	1
1	A	115	GLN	Sidechain	1
1	A	75	ARG	Sidechain	1

6.2 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 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	1780	1776	1775	1±1
All	All	44500	44400	44375	26

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:46:LEU:HD13	1:A:78:PHE:CZ	0.72	2.18	17	2
1:A:46:LEU:HD13	1:A:78:PHE:CE2	0.69	2.22	13	2
1:A:96:PHE:CE1	1:A:107:THR:HG21	0.67	2.24	10	1
1:A:96:PHE:CZ	1:A:107:THR:HG21	0.60	2.31	10	1
1:A:214:TYR:CD1	1:A:240:VAL:HG21	0.54	2.37	19	2
1:A:115:GLN:HE21	1:A:116:ASP:N	0.51	2.04	18	1
1:A:24:LEU:H	1:A:24:LEU:HD23	0.48	1.68	9	1
1:A:103:HIS:CD2	1:A:232:GLN:NE2	0.47	2.82	14	2
1:A:91:LEU:HD13	1:A:91:LEU:H	0.46	1.71	15	1
1:A:214:TYR:CE1	1:A:240:VAL:HG11	0.46	2.45	7	1
1:A:115:GLN:HE21	1:A:115:GLN:C	0.45	2.15	18	1
1:A:240:VAL:HG22	1:A:241:LYS:O	0.45	2.11	19	1
1:A:193:LEU:HD21	1:A:222:TYR:CE1	0.44	2.47	3	1
1:A:106:LEU:HD11	1:A:157:PHE:CZ	0.44	2.47	16	2
1:A:46:LEU:HD11	1:A:78:PHE:CE2	0.43	2.48	6	1
1:A:213:LEU:CD1	1:A:217:ALA:N	0.43	2.82	20	1
1:A:24:LEU:HD23	1:A:66:GLY:H	0.42	1.74	2	1
1:A:69:LYS:CG	1:A:70:ASN:H	0.42	2.27	20	1
1:A:38:GLN:H	1:A:38:GLN:NE2	0.40	2.14	17	1
1:A:115:GLN:C	1:A:115:GLN:HE21	0.40	2.20	22	1
1:A:214:TYR:CE1	1:A:240:VAL:HG21	0.40	2.50	6	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	234/255 (92%)	201±4 (86±2%)	28±4 (12±2%)	6±2 (2±1%)	9	46

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	5850/6375 (92%)	5019 (86%)	690 (12%)	141 (2%)	9 46

All 49 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	121	GLY	14
1	A	183	HIS	13
1	A	120	SER	6
1	A	70	ASN	6
1	A	66	GLY	5
1	A	59	ASN	5
1	A	68	LEU	5
1	A	161	ASP	4
1	A	159	SER	4
1	A	15	GLY	4
1	A	127	ARG	4
1	A	60	GLY	3
1	A	103	HIS	3
1	A	204	ARG	3
1	A	24	LEU	3
1	A	67	LYS	3
1	A	163	GLY	3
1	A	231	ALA	3
1	A	162	ALA	3
1	A	102	SER	3
1	A	65	THR	3
1	A	132	GLY	3
1	A	42	LYS	3
1	A	39	SER	3
1	A	201	ASP	2
1	A	22	ALA	2
1	A	160	ASP	2
1	A	230	LYS	2
1	A	218	GLU	2
1	A	206	ALA	2
1	A	69	LYS	2
1	A	203	LYS	1
1	A	253	ALA	1
1	A	144	LEU	1
1	A	200	PRO	1
1	A	202	GLY	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	91	LEU	1
1	A	58	GLY	1
1	A	142	ASP	1
1	A	117	SER	1
1	A	118	GLU	1
1	A	122	LYS	1
1	A	143	LYS	1
1	A	147	GLY	1
1	A	71	ASP	1
1	A	148	GLY	1
1	A	229	GLY	1
1	A	133	ASP	1
1	A	180	LYS	1

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	183/198 (92%)	153±5 (83±3%)	30±5 (17±3%)	5	41
All	All	4575/4950 (92%)	3813 (83%)	762 (17%)	5	41

All 141 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	99	TYR	25
1	A	166	LEU	25
1	A	34	LEU	22
1	A	214	TYR	22
1	A	38	GLN	19
1	A	51	GLN	19
1	A	115	GLN	19
1	A	48	LEU	17
1	A	128	GLN	17
1	A	46	LEU	15
1	A	218	GLU	15
1	A	193	LEU	14

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	153	ARG	13
1	A	91	LEU	11
1	A	97	GLN	11
1	A	213	LEU	11
1	A	183	HIS	11
1	A	67	LYS	10
1	A	189	LEU	10
1	A	126	LYS	10
1	A	247	ARG	10
1	A	130	ARG	10
1	A	239	GLU	10
1	A	233	GLU	9
1	A	75	ARG	9
1	A	85	ASP	9
1	A	185	LYS	9
1	A	32	GLN	9
1	A	69	LYS	9
1	A	186	SER	8
1	A	241	LYS	8
1	A	100	LYS	8
1	A	116	ASP	8
1	A	149	ARG	8
1	A	171	ASP	8
1	A	204	ARG	8
1	A	175	LYS	8
1	A	201	ASP	7
1	A	43	ASN	7
1	A	37	ASP	7
1	A	47	LYS	6
1	A	123	MET	6
1	A	24	LEU	6
1	A	77	ASP	6
1	A	222	TYR	6
1	A	95	GLU	5
1	A	138	HIS	5
1	A	71	ASP	5
1	A	81	GLN	5
1	A	110	GLN	5
1	A	207	VAL	5
1	A	255	GLN	5
1	A	101	GLN	5
1	A	113	GLN	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	160	ASP	5
1	A	82	ILE	5
1	A	176	GLN	5
1	A	232	GLN	5
1	A	254	LYS	5
1	A	16	LEU	5
1	A	40	VAL	5
1	A	161	ASP	4
1	A	165	LYS	4
1	A	54	GLU	4
1	A	133	ASP	4
1	A	80	ARG	4
1	A	83	GLU	4
1	A	234	VAL	4
1	A	219	LYS	4
1	A	41	ARG	4
1	A	107	THR	4
1	A	140	SER	4
1	A	180	LYS	3
1	A	190	ASN	3
1	A	35	THR	3
1	A	44	GLU	3
1	A	61	ASP	3
1	A	39	SER	3
1	A	45	LYS	3
1	A	68	LEU	3
1	A	134	ILE	3
1	A	230	LYS	3
1	A	31	LEU	3
1	A	63	LEU	3
1	A	205	HIS	3
1	A	203	LYS	3
1	A	114	ILE	3
1	A	170	ILE	2
1	A	192	ASP	2
1	A	199	LYS	2
1	A	221	SER	2
1	A	103	HIS	2
1	A	18	ASP	2
1	A	141	PHE	2
1	A	197	ASP	2
1	A	139	THR	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	122	LYS	2
1	A	55	LYS	2
1	A	93	SER	2
1	A	242	THR	2
1	A	74	SER	2
1	A	92	GLU	2
1	A	142	ASP	2
1	A	224	LEU	2
1	A	72	LYS	2
1	A	118	GLU	2
1	A	119	HIS	2
1	A	184	LEU	2
1	A	70	ASN	2
1	A	143	LYS	2
1	A	237	SER	2
1	A	79	ILE	1
1	A	159	SER	1
1	A	84	VAL	1
1	A	198	ILE	1
1	A	56	THR	1
1	A	102	SER	1
1	A	216	GLN	1
1	A	251	LEU	1
1	A	240	VAL	1
1	A	246	ILE	1
1	A	151	THR	1
1	A	78	PHE	1
1	A	20	LEU	1
1	A	59	ASN	1
1	A	209	SER	1
1	A	223	SER	1
1	A	131	ILE	1
1	A	248	HIS	1
1	A	120	SER	1
1	A	129	PHE	1
1	A	144	LEU	1
1	A	127	ARG	1
1	A	172	PHE	1
1	A	33	SER	1
1	A	65	THR	1
1	A	104	SER	1
1	A	191	VAL	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	212	VAL	1
1	A	111	THR	1
1	A	64	ASN	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

No chemical shift data were provided