Full wwPDB X-ray Structure Validation Report (i)

Jan 31, 2016 – 09:52 PM GMT

PDB ID	:	1R14
Title	:	Carbohydrate recognition and neck domains of surfactant protein A (Sp-A) containing samarium
Authors	:	Head, J.F.; Mealy, T.R.; McCormack, F.X.; Seaton, B.A.
Deposited on	:	2003-09-23
Resolution	:	2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i)) were used in the production of this report:

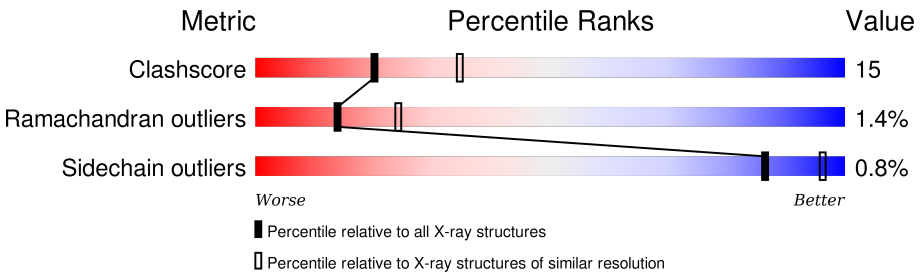
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.50 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))
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	148	77% 19% • •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1163 atoms, of which 0 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 Pulmonary surfactant-associated protein A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	145	1143	714	189	231	9	0	0	0

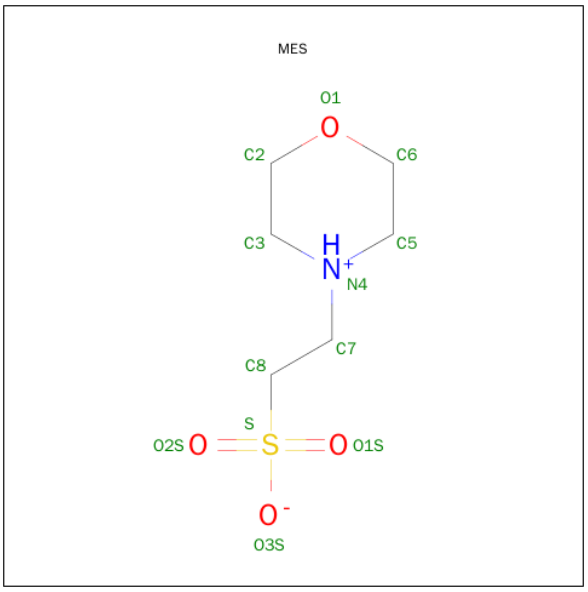
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	187	SER	ASN	ENGINEERED	UNP P08427

- Molecule 2 is SAMARIUM (III) ION (three-letter code: SM) (formula: Sm).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Sm		
2	A	2	2	2	0	0

- Molecule 3 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
3	A	1	12	6	1	4	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	6	6	6	0	0

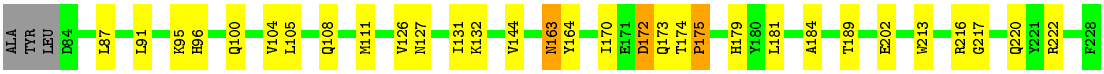
3 Residue-property plots (i)

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.

Note EDS was not executed.

- Molecule 1: Pulmonary surfactant-associated protein A

Chain A:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, α, β, γ	97.68Å 97.68Å 44.40Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 2.50	Depositor
% Data completeness (in resolution range)	96.2 (50.00-2.50)	Depositor
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.222 , 0.247	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1163	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

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

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	0.36	0/1166	0.57	0/1576

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 (i)

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	1143	0	1079	31	0
2	A	2	0	0	0	0
3	A	12	0	13	2	0
4	A	6	0	0	5	0
All	All	1163	0	1092	33	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 15.

All (33) 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)
3:A:700:MES:H32	4:A:406:HOH:O	1.92	0.68
1:A:163:ASN:HD22	1:A:164:TYR:H	1.46	0.64
1:A:216:ARG:HG2	1:A:217:GLY:H	1.68	0.58
1:A:202:GLU:HG2	1:A:216:ARG:HA	1.86	0.58
1:A:189:THR:HA	1:A:213:TRP:CH2	2.40	0.56
1:A:173:GLN:HG2	1:A:174:THR:HG23	1.89	0.55
1:A:104:VAL:O	1:A:108:GLN:HG2	2.07	0.53
1:A:189:THR:HA	1:A:213:TRP:CZ3	2.44	0.52
1:A:216:ARG:HG2	1:A:217:GLY:N	2.24	0.52
1:A:216:ARG:CG	1:A:217:GLY:H	2.22	0.51
1:A:126:VAL:HG23	1:A:131:ILE:HG13	1.95	0.49
1:A:173:GLN:O	1:A:175:PRO:HD3	2.14	0.48
1:A:172:ASP:HB3	1:A:173:GLN:OE1	2.14	0.48
1:A:163:ASN:HD22	1:A:164:TYR:N	2.11	0.47
1:A:105:LEU:HB3	1:A:111:MET:HG2	1.95	0.47
1:A:164:TYR:CG	4:A:406:HOH:O	2.67	0.47
1:A:216:ARG:CG	1:A:217:GLY:N	2.78	0.46
1:A:163:ASN:ND2	1:A:164:TYR:H	2.10	0.46
1:A:127:ASN:HA	1:A:220:GLN:O	2.16	0.46
1:A:91:LEU:O	1:A:95:LYS:HG3	2.16	0.45
3:A:700:MES:H52	4:A:406:HOH:O	2.17	0.45
1:A:170:ILE:HG13	1:A:179:HIS:CE1	2.52	0.45
1:A:96:HIS:O	1:A:100:GLN:HG3	2.16	0.44
1:A:173:GLN:HG2	1:A:174:THR:N	2.33	0.44
1:A:164:TYR:HB2	1:A:222:ARG:HD2	1.99	0.44
1:A:132:LYS:HD3	1:A:181:LEU:HD21	2.01	0.43
1:A:164:TYR:CE2	4:A:406:HOH:O	2.70	0.43
1:A:202:GLU:CG	1:A:216:ARG:HA	2.49	0.42
1:A:164:TYR:CD2	4:A:406:HOH:O	2.57	0.42
1:A:144:VAL:O	1:A:144:VAL:HG13	2.20	0.41
1:A:163:ASN:ND2	1:A:164:TYR:N	2.69	0.41
1:A:179:HIS:HB2	1:A:184:ALA:O	2.21	0.41
1:A:87:LEU:C	1:A:87:LEU:HD13	2.41	0.41

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	143/148 (97%)	133 (93%)	8 (6%)	2 (1%)	14 24

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	172	ASP
1	A	175	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	124/126 (98%)	123 (99%)	1 (1%)	86 96

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	163	ASN

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	162	ASN
1	A	163	ASN

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 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	MES	A	700	-	11,12,12	1.75	3 (27%)	14,16,16	1.13	1 (7%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	MES	A	700	-	-	0/6/14/14	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	700	MES	C5-N4	2.74	1.54	1.46
3	A	700	MES	C3-N4	2.98	1.55	1.46
3	A	700	MES	C7-N4	3.11	1.54	1.47

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	700	MES	O1S-S-C8	2.36	108.92	106.91

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	700	MES	2	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.