



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 07:18 PM GMT

PDB ID : 4RTD
Title : Escherichia coli alpha-2-macroglobulin activated by porcine elastase
Authors : Fyfe, C.D.; Grinter, R.; Roszak, A.W.; Josts, I.; Cogdell, R.J.; Walker, D.
Deposited on : 2014-11-14
Resolution : 3.65 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : unknown
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026982
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20026982

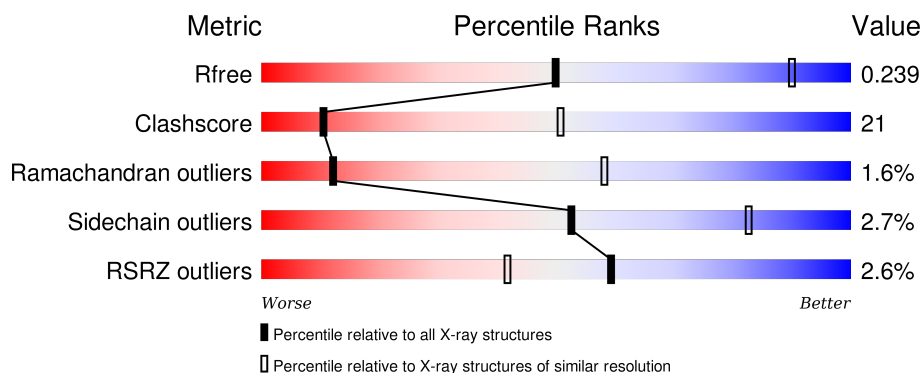
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

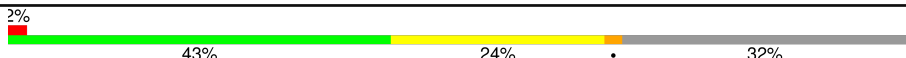
The reported resolution of this entry is 3.65 Å.

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))
R_{free}	91344	1010 (3.82-3.50)
Clashscore	102246	1125 (3.82-3.50)
Ramachandran outliers	100387	1079 (3.82-3.50)
Sidechain outliers	100360	1078 (3.82-3.50)
RSRZ outliers	91569	1017 (3.82-3.50)

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 $\leq 5\%$. 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.

Mol	Chain	Length	Quality of chain
1	A	1639	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 8699 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 Uncharacterized lipoprotein YfhM.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1122	Total	C	N	O	S	Se	0	0	0
			8699	5497	1505	1677	1	19			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	219	PRO	SER	ENGINEERED MUTATION	UNP P76578
A	606	ARG	GLN	ENGINEERED MUTATION	UNP P76578
A	1587	ASN	SER	ENGINEERED MUTATION	UNP P76578
A	1654	LEU	-	EXPRESSION TAG	UNP P76578
A	1655	GLU	-	EXPRESSION TAG	UNP P76578
A	1656	HIS	-	EXPRESSION TAG	UNP P76578
A	1657	HIS	-	EXPRESSION TAG	UNP P76578
A	1658	HIS	-	EXPRESSION TAG	UNP P76578
A	1659	HIS	-	EXPRESSION TAG	UNP P76578
A	1660	HIS	-	EXPRESSION TAG	UNP P76578
A	1661	HIS	-	EXPRESSION TAG	UNP P76578

L1615	M1399	I1276	G1167	V1066	V962	N885
A1616	K1400	I1283	K1168	A1071	N963	G886
R1617	L1401	L1284	P1169	A1076	N966	E887
V1619	L1410	R1285	P1170	P1077	Q971	K888
T1620	L1413	L1287	N1171	G1078	A972	P889
P1621	L1432	M1292	I1173	V1079	V892	N892
P1627	Q1416	M1293	A1174	R1080	N893	N896
M1630	A1417	M1296	R1175	P1086	V896	V899
V1631	E1420	P1297	Y1176	P1087	Q980	L904
E1632	E1427	A1298	I1177	R1088	G981	N905
S1633	S1428	L1301	K1178	A1089	V985	I906
M1634	N1429	K1302	E1179	L1090	T986	T907
V1635	N1445	A1303	Y1185	L1180	L987	D908
V1636	F1432	A1304	L1188	E1097	D991	Y909
P1637	Q1443	K1305	L1195	I1098	T1004	W915
Q1638	K1444	Y1311	S1198	I1102	M1011	Q921
W1639	W1445	A1326	L1199	G1103	E1012	K922
W1640	Q1446	L1327	Y1200	L1105	ARG	ARG
A1641	T1457	R1328	T1201	A1106	TTR	GLY
L1649	K1460	E1329	N1202	L1107	V1015	ALA
I1650	A1461	I1330	Q1205	P1108	I1016	ASP
V1651	Q1462	W1331	L1206	G1109	V1017	ASP
R1652	S1464	D1336	L1209	E1110	A1018	I1E
P1653	ALA	A1337	I1209	T1111	V1021	TTR
LEU	ASN	A1338	G1210	I1112	I1021	ASP
GLU	Q1463	G1340	I1211	A1113	I1022	I1E
HIS	S1464	L1341	K1212	D1114	M1027	TTR
HIS	L1471	L1342	K1218	K1120	M1031	GLN
HIS	L1474	L1343	S1222	I1121	A1082	VAL
HIS	S1477	Q1344	S1223	R1124	S1033	I1E
HIS	N1478	L1345	V1223	P1125	T1036	GLU
HIS	S1479	L1350	G1226	T1131	L1039	GLY
HIS	P1483	K1351	I1227	V1132	L1043	GLN
HIS	L1484	D1355	S1228	N1133	T1044	ARG
HIS	W1485	R1358	R1229	A1137	A943	LEU
HIS	L1486	R1371	L1230	L1138	A944	ARG
HIS	R1487	I1377	L1231	Q1139	L945	ARG
HIS	M1488	W1378	Q1232	P1140	R946	ARG
HIS	D1489	L1365	M1233	G1141	F947	ARG
HIS	A1490	R1371	A1241	E1142	G948	ARG
HIS	S1491	I1377	M1258	I1146	G949	ARG
HIS	N1503	Q1506	V1262	P1147	D950	ARG
HIS	V1504	R1509	R1263	S1155	Q1051	ARG
HIS	L1505	G1383	T1273	T1055	V1055	ARG
HIS	Q1506	R1387	D1274	L1063	G957	ARG
HIS	E1606	G1513	A1275		G958	ARG
HIS	Y1607	G1516			K959	ARG
HIS	Q1608				P960	ARG
HIS	T1611				P961	ARG

4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α , β , γ	176.06 Å 176.06 Å 161.13 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.87 – 3.65 46.87 – 3.65	Depositor EDS
% Data completeness (in resolution range)	99.9 (46.87-3.65) 100.0 (46.87-3.65)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.83 (at 3.66 Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.177 , 0.238 0.183 , 0.239	Depositor DCC
R_{free} test set	1033 reflections (5.25%)	DCC
Wilson B-factor (Å ²)	126.2	Xtriage
Anisotropy	0.162	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 113.6	EDS
Estimated twinning fraction	0.037 for h,-h-k,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 20712 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	8699	wwPDB-VP
Average B, all atoms (Å ²)	144.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.14% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 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 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	0.50	0/8863	0.75	3/12024 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	460	MSE	CG-SE-CE	7.75	115.94	98.90
1	A	1105	LEU	CA-CB-CG	5.31	127.52	115.30
1	A	1488	MSE	CA-CB-CG	-5.20	104.46	113.30

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	8699	0	8589	370	0
All	All	8699	0	8589	370	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 21.

The worst 5 of 370 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)
1:A:639:ILE:HG21	1:A:738:TRP:O	1.73	0.89
1:A:396:LEU:HD22	1:A:906:ILE:HD11	1.56	0.87
1:A:1045:ASN:OD1	1:A:1047:THR:HG22	1.77	0.84
1:A:1504:VAL:HG21	1:A:1639:TRP:CD1	2.15	0.81
1:A:1106:ALA:HA	1:A:1113:ALA:HA	1.62	0.81

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	1112/1639 (68%)	989 (89%)	105 (9%)	18 (2%)	12	58

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	642	GLU
1	A	730	GLU
1	A	1169	PRO
1	A	1178	LYS
1	A	947	PHE

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
-----	-------	----------	-----------	----------	-------------

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	927/1340 (69%)	902 (97%)	25 (3%)	52 82

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1155	SER
1	A	1228	SER
1	A	1607	TYR
1	A	1198	SER
1	A	1274	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1139	GLN
1	A	1172	ASN
1	A	1470	GLN
1	A	1590	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	1103/1639 (67%)	0.05	29 (2%) 59 42	85, 139, 205, 267	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1108	PRO	4.9
1	A	943	ALA	4.0
1	A	616	GLY	3.3
1	A	615	PRO	3.2
1	A	651	VAL	3.1

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

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

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.