



wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 07:54 am BST

PDB ID : 3CWG
Title : Unphosphorylated mouse STAT3 core fragment
Authors : Ren, Z.; Mao, X.; Mertens, C.; Krishnaraj, R.; Qin, J.; Mandal, P.K.; Romanowshi, M.J.; McMurray, J.S.
Deposited on : 2008-04-21
Resolution : 3.05 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

<https://www.wwpdb.org/validation/2017/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
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

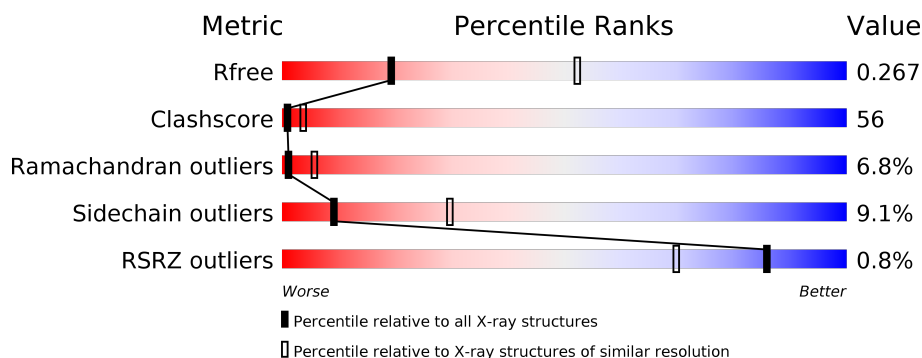
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.05 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}	130704	1754 (3.10-3.02)
Clashscore	141614	1864 (3.10-3.02)
Ramachandran outliers	138981	1794 (3.10-3.02)
Sidechain outliers	138945	1793 (3.10-3.02)
RSRZ outliers	127900	1713 (3.10-3.02)

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 respectively. 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	562	
1	B	562	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 8215 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.

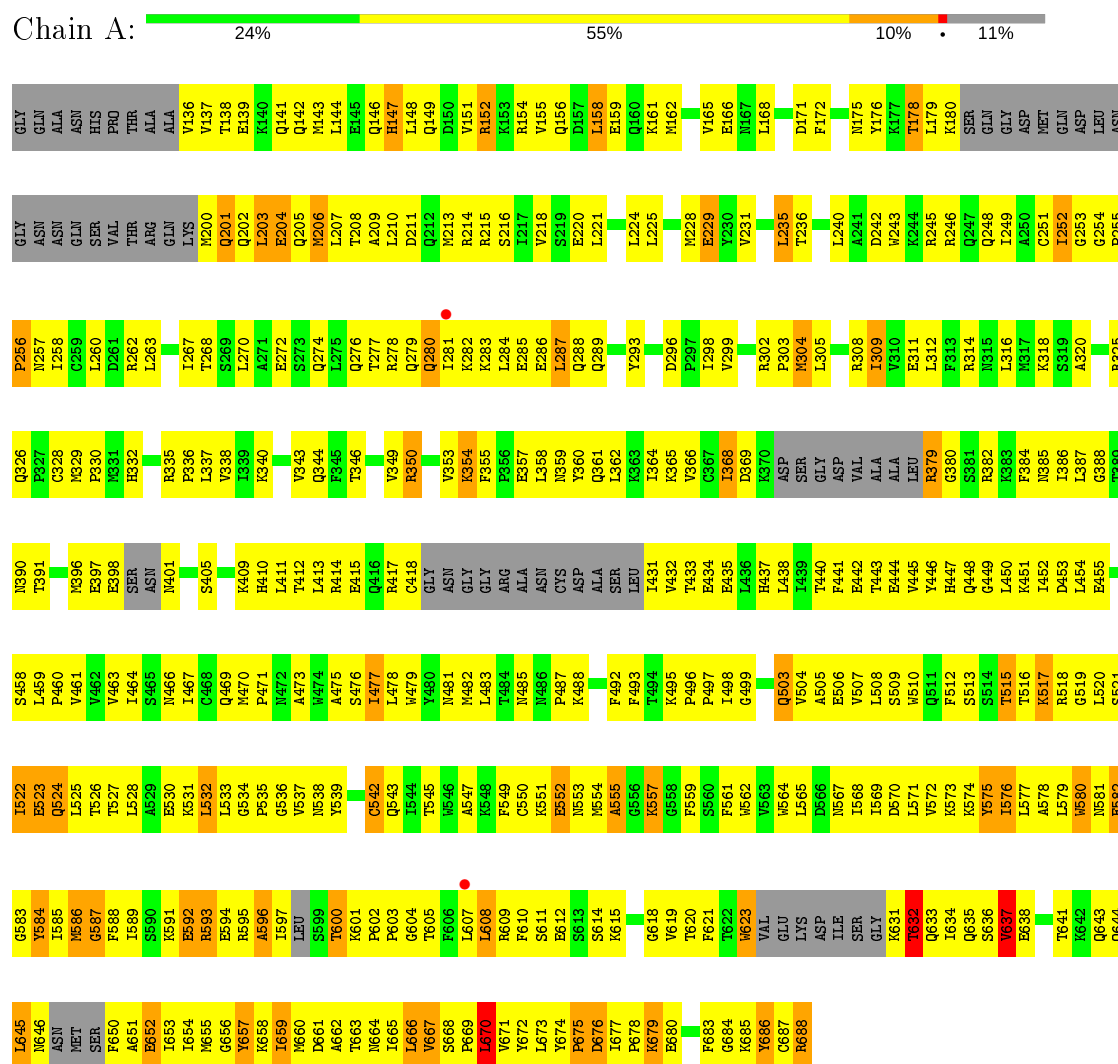
- Molecule 1 is a protein called Signal transducer and activator of transcription 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	501	Total	C	N	O	S	0	0	0
			4082	2615	692	748	27			
1	B	507	Total	C	N	O	S	0	0	0
			4133	2645	704	758	26			

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

- Molecule 1: Signal transducer and activator of transcription 3



- Molecule 1: Signal transducer and activator of transcription 3



G1656	G1657	K1658	I1658	ASP	ALA	T1663	M1664	I1665	L1666	V1667	P1668	P1669	L1670	V1671	L1672	L1673	V1674	P1675	D1676	I1677	P1678	K1679	E1680	E1681	A1682	F1683	G1684	K1685	V1686	C1687	L1688																					
R1595	A1596	I1597	L1598	S1599	T1600	K1601	P1602	P1603	G1604	T1605	F1606	L1607	L1608	F1609	F1610	S1611	E1612	S1613	S1614	K1615	E1616	G1617	G1618	V1619	T1620	F1621	T1622	W1623	V1624	G1625	L1626	L1627																				
G1534	P1535	G1536	V1537	N1538	Y1539	S1540	G1541	C1542	Q1543	I1544	A1547	K1548	K1549	F1549	C1550	K1551	E1552	N1553	M1554	A1555	G1556	K1557	G1558	F1559	S1560	F1561	W1562	V1563	V1564	L1565	D1566	I1567																				
P1470	P1471	M1472	A1473	M1474	S1476	I1477	L1478	W1479	Y1480	M1481	M1482	L1483	T1484	M1485	N1486	P1487	K1488	F1492	F1493	T1494	K1495	P1496	P1497	I1498	G1499	Q1503	V1504	A1505	E1506	V1507	L1508	G1509																				
G1401	L1404	K1409	H1410	L1411	T1412	L1413	R1414	E1415	C1418	GLY	ASN	GLY	GLY	ARG	ALA	ASN	CYS	ASP	ALA	LEU	I1431	V1432	T1433	H1437	L1438	I1439	T1440	F1441	E1442	T1443	E1444	V1445	Y1446																			
H1332	R1335	P1336	L1337	V1338	H1339	K1340	V1343	Q1344	F1345	T1346	R1350	V1353	K1354	F1355	P1356	E1357	L1358	N1359	Y1360	Q1361	L1362	K1365	G1367	T1368	D1369	K1370	E1371	SER	GLY	ASP	VAL	ALA	LEU																			
L1260	D1261	L1262	L1263	I1267	T1268	S1269	A1271	E1272	S1273	Q1274	L1275	T1276	GLY	GLY	GLY	ARG	ALA	ASN	CYS	ASP	ALA	LEU	I1431	V1432	T1433	H1437	L1438	I1439	T1440	F1441	E1442	T1443	E1444	V1445	Y1446																	
SER	V1195	T1196	L1197	H1198	K1199	M1200	Q1201	L1202	Q1203	E1204	Q1205	M1206	L1207	Q1208	A1209	L1210	D1211	Q1212	M1213	L1214	R1215	V1218	S1219	E1220	L1224	L1225	M1228	E1229	Y1230	V1231	T1234	L1235	T1236	D1237	E1238	L1240	A1241	D1242	W1243	R1244	K1245	R1246	Q1247	Q1248	I1249	A1250	C1251	I1252	P1255	P1256	M1257	C1259
GLY	GLN	ALA	ASN	HIS	PRO	THR	ALA	ALA	V1136	V1137	T1138	E1139	K1140	Q1141	Q1142	M1143	L1144	E1145	Q1146	H1147	L1148	V1151	R1152	V1155	Q1156	D1157	L1158	E1159	M1162	V1165	E1166	H1167	L1168	Q1169	F1172	M1175	T1178	L1179	K1180	S1181	Q1182	G1183	D1184	MET	GLN	ASP	LEU	ASN	GLY	ASN	ASN	GLN

4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	254.78Å 254.78Å 123.78Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 3.05 29.70 – 3.02	Depositor EDS
% Data completeness (in resolution range)	91.2 (30.00-3.05) 90.1 (29.70-3.02)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.93 (at 3.00Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.250 , 0.269 0.245 , 0.267	Depositor DCC
R_{free} test set	7413 reflections (8.81%)	wwPDB-VP
Wilson B-factor (Å ²)	71.6	Xtriage
Anisotropy	0.491	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 51.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.41$, $\langle L^2 \rangle = 0.23$	Xtriage
Estimated twinning fraction	0.367 for -h,-k,l	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	8215	wwPDB-VP
Average B, all atoms (Å ²)	97.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.80% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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

5 Model quality

5.1 Standard geometry

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	0.47	0/4159	0.71	0/5608
1	B	0.47	0/4210	0.70	0/5675
All	All	0.47	0/8369	0.70	0/11283

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

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	4082	0	4139	481	0
1	B	4133	0	4195	461	1
All	All	8215	0	8334	929	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 56.

The worst 5 of 929 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:660:MET:HA	1:A:666:LEU:HA	1.37	1.07
1:B:1597:ILE:HG13	1:B:1598:LEU:H	1.22	1.04

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:650:PHE:HA	1:A:653:ILE:HD13	1.39	1.04
1:B:1605:THR:HG22	1:B:1672:TYR:HB2	1.42	1.01
1:A:547:ALA:HA	1:A:551:LYS:HB3	1.43	1.00

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:1575:TYR:OH	1:B:1643:GLN:OE1[5_554]	2.14	0.06

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	485/562 (86%)	362 (75%)	91 (19%)	32 (7%)	1	6
1	B	491/562 (87%)	362 (74%)	95 (19%)	34 (7%)	1	5
All	All	976/1124 (87%)	724 (74%)	186 (19%)	66 (7%)	1	6

5 of 66 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	256	PRO
1	A	368	ILE
1	A	555	ALA
1	A	557	LYS
1	A	667	VAL

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	458/505 (91%)	416 (91%)	42 (9%)	9	29
1	B	464/505 (92%)	422 (91%)	42 (9%)	9	30
All	All	922/1010 (91%)	838 (91%)	84 (9%)	9	30

5 of 84 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	659	ILE
1	B	1203	LEU
1	B	1623	TRP
1	A	676	ASP
1	B	1152	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	503	GLN
1	B	1248	GLN
1	B	1481	ASN
1	A	644	GLN
1	A	279	GLN

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 [i](#)

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	501/562 (89%)	0.10	2 (0%)	92 82	43, 92, 127, 139	0
1	B	507/562 (90%)	0.14	6 (1%)	79 58	36, 101, 131, 143	0
All	All	1008/1124 (89%)	0.12	8 (0%)	86 70	36, 97, 130, 143	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1199	LYS	3.3
1	B	1198	GLN	3.3
1	B	1667	VAL	3.0
1	A	607	LEU	2.4
1	B	1339	ILE	2.2

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.