



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:59 AM GMT

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
<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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

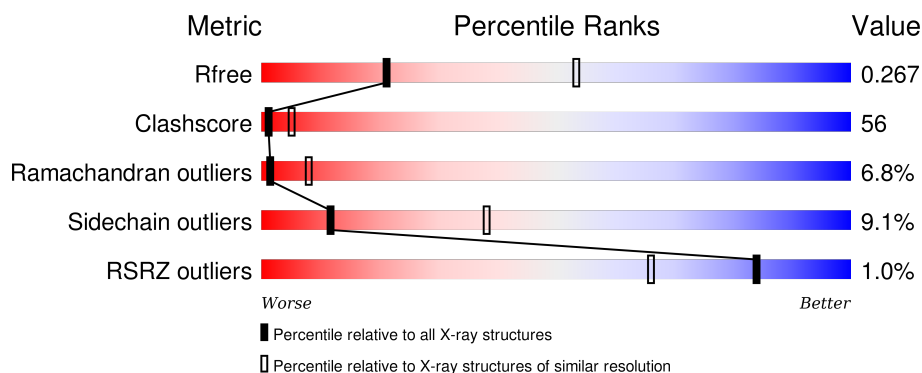
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

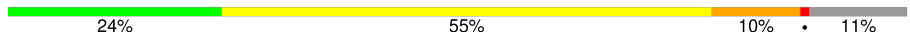
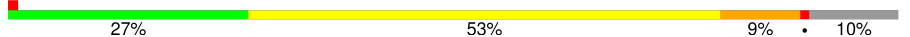
The reported resolution of this entry is 3.05 Å.

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}	91344	1191 (3.10-3.02)
Clashscore	102246	1303 (3.10-3.02)
Ramachandran outliers	100387	1254 (3.10-3.02)
Sidechain outliers	100360	1254 (3.10-3.02)
RSRZ outliers	91569	1197 (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. 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	 24% 55% 10% • 11%
1	B	562	 27% 53% 9% • 10%

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.

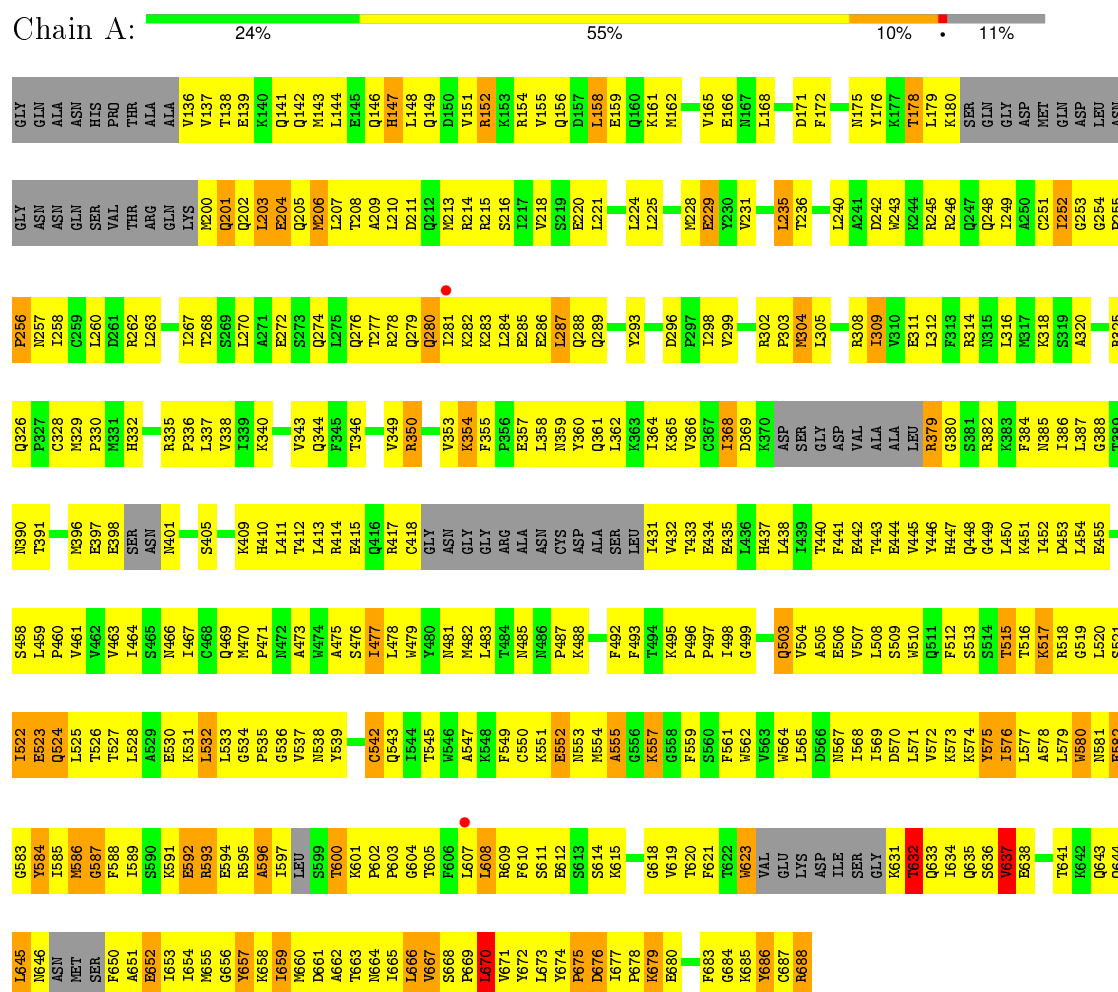
- Molecule 1 is a protein called Signal transducer and activator of transcription 3.

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			

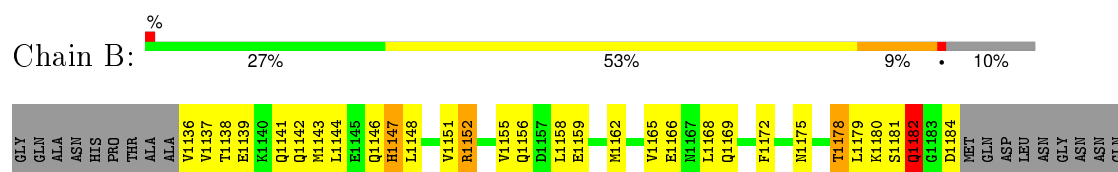
3 Residue-property plots

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

• Molecule 1: Signal transducer and activator of transcription 3



• Molecule 1: Signal transducer and activator of transcription 3



G1656	R1595	G1534	M1470	H1401	H1332	L1260	SER
Y1657	A1596	P1535	P1471	L1404	R1335	D1261	V1195
K1658	I1597	G1536	M1472	L1404	P1336	L1262	T1196
ILE	L1598	V1537	A1473	K1409	L1337	L1263	R1197
MET	S1599	M1538	W1474	H1410	L1338	I1267	Q1198
ASP	T1600	Y1539	A1475	L1411	I1339	T1268	K1199
ALA	K1601	S1540	S1476	T1412	K1340	M1200	M1200
T1603	P1602	G1541	I1477	L1413	L1340	Q1201	Q1201
M1604	P1603	C1542	L1478	R1414	V1343	Q1202	L1202
I1605	G1604	Q1543	W1479	R1415	Q1344	L1203	L1203
L1606	T1605	I1544	Y1480	E1415	E1272	E1204	E1204
V1607	F1606	M1481	M1481	C1418	F1345	Q1205	Q1205
L1607	L1607	M1482	M1482	GLY	T1346	M1206	M1206
L1608	L1608	L1483	L1483	ASN	R1350	L1207	L1207
F1609	F1549	T1484	T1484	GLY	V1353	T1208	T1208
F1610	C1550	M1485	M1485	GLY	GLY	A1209	A1209
S1611	K1551	M1486	M1486	ARG	V1353	L1210	L1210
E1612	E1552	P1487	P1487	ALA	F1355	D1211	D1211
S1613	M1553	K1488	K1488	ALA	P1356	Q1212	Q1212
S1614	M1554			ASN	F1356	M1213	M1213
K1615	A1555	F1492	F1492	CYS	E1357	M1214	M1214
K1616	G1556	F1493	F1493	ASP	L1358	R1215	R1215
G1617	K1557	T1494	T1494	ALA	M1359		
G1618	G1558	K1495	K1495	SER	Y1360		
V1619	F1559	P1496	P1496	LEU	Q1361	V1218	V1218
T1620	S1560	P1497	P1497	LEU	E1286	S1219	S1219
F1621	F1561	I1498	I1498	I1431	L1287	E1220	E1220
T1622	W1562	G1499	G1499	T1433	Q1288		
W1623	W1563			T1433	Q1289	L1224	L1224
VAL	V1564	Q1503	Q1503	H1437	V1366	L1225	L1225
GLU	L1565	V1504	V1504	L1438	C1367		
LYS	D1566	A1505	A1505	L1439	I1368	M1228	M1228
ASP	M1567	E1506	E1506	T1440	D1369	E1229	E1229
ILE	I1568	V1507	V1507	F1441	K1370	Y1230	Y1230
SER	I1569	L1508	L1508	E1442	ASP	I1298	I1298
GLY	D1570	S1509	S1509	T1443	SER	V1299	V1299
K1631	L1571	W1510	W1510	T1444	GLY	T1234	T1234
T1632	V1572	Q1511	Q1511	E1444	ASP	L1235	L1235
Q1633	K1573	F1512	F1512	V1445	VAL	T1236	T1236
I1634	K1574	S1513	S1513	T1446	ALA	D1237	D1237
Q1635	Y1575	S1514	S1514	H1447	ALA	E1238	E1238
S1636	I1576	T1515	T1515	Q1448	LEU	L1240	L1240
V1637	L1577	T1516	T1516	G1449	R1379	E1239	E1239
E1638	A1578	K1517	K1517	L1450	G1380	L1241	L1241
	L1579	A1518	A1518	K1451	S1381	D1242	D1242
T1641	W1580	G1519	G1519	I1452	A1382	W1243	W1243
K1642	M1581	L1520	L1520	D1453	K1383	R1244	R1244
Q1643	E1582	S1521	S1521	L1454	F1384	R1245	R1245
Q1644	G1583	I1522	I1522	E1455	M1385	R1246	R1246
L1645	Y1584	E1523	E1523		T1386	Q1247	Q1247
N1646	I1585	Q1524	Q1524	S1458	L1387	Q1248	Q1248
ASN	M1586	L1525	L1525	L1459	G1388	I1249	I1249
MET	G1587	T1526	T1526	P1460	T1389	A1250	A1250
SER	F1588	T1527	T1527	V1461	M1390	C1251	C1251
F1650	I1589	L1528	L1528	V1462	T1391	I1252	I1252
A1651	S1590	A1529	A1529	V1463	M1396		
E1652	K1591	E1530	E1530	M1466	P1397	P1255	P1255
I1653	E1592	K1531	K1531	I1467	E1397	P1256	P1256
I1654	R1593	L1532	L1532	C1468	E1398	M1257	M1257
M1655	E1594	L1533	L1533	Q1469	SER	P1330	P1330

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 (10.20%)	DCC
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
Estimated twinning fraction	0.367 for -h,-k,l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.41$, $\langle L^2 \rangle = 0.23$	Xtriage
Outliers	0 of 84177 reflections	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, 0.375 respectively for untwinned datasets, and 0.333, 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	8
1	B	491/562 (87%)	362 (74%)	95 (19%)	34 (7%)	1	8
All	All	976/1124 (87%)	724 (74%)	186 (19%)	66 (7%)	1	8

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%)	11	37
1	B	464/505 (92%)	422 (91%)	42 (9%)	12	38
All	All	922/1010 (91%)	838 (91%)	84 (9%)	12	38

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

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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%)	0.12	2 (0%) 93 84	43, 92, 127, 139	0
1	B	507/562 (90%)	0.16	8 (1%) 74 52	36, 101, 131, 143	0
All	All	1008/1124 (89%)	0.14	10 (0%) 84 66	36, 97, 130, 143	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1199	LYS	3.4
1	B	1198	GLN	3.3
1	B	1667	VAL	3.1
1	A	607	LEU	2.4
1	B	1339	ILE	2.3

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.