wwPDB X-ray Structure Validation Summary Report (i)

Aug 23, 2023 – 05:10 AM EDT

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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

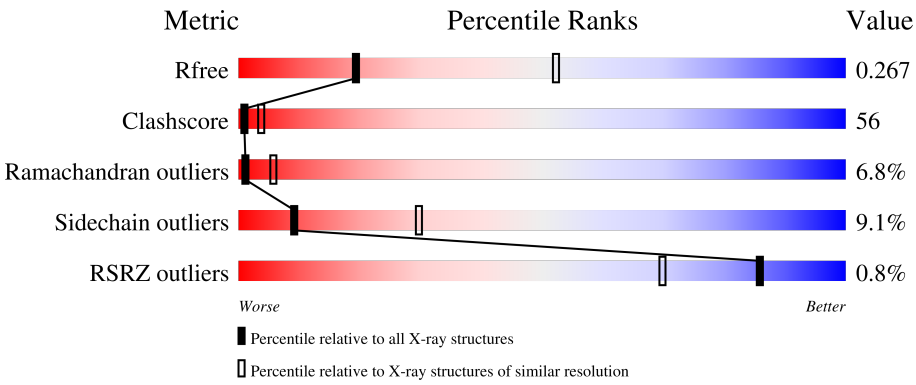
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 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 of 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 <=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 (i)

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.

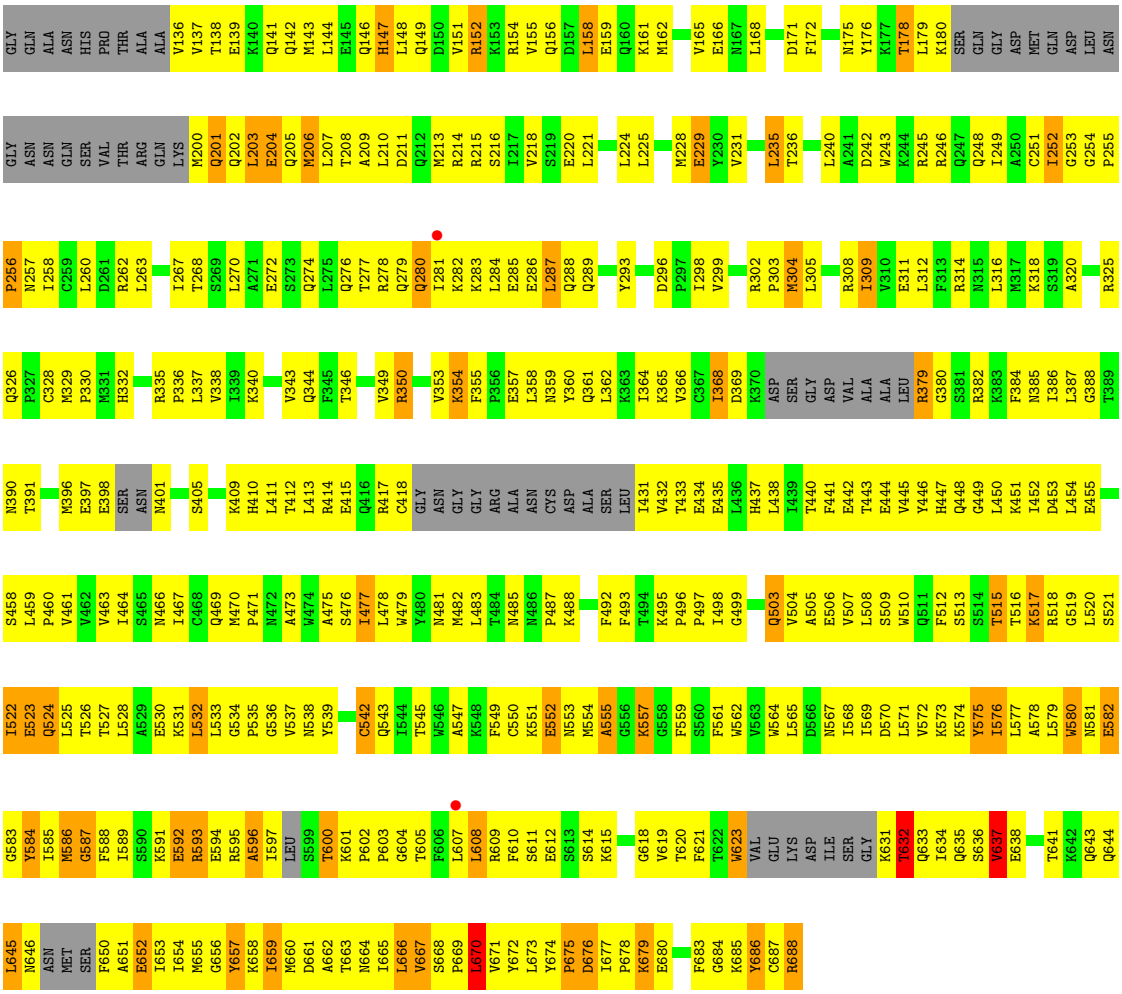
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	501	4082	2615	692	748	27	0	0	0
			Total	C	N	O	S			
1	B	507	4133	2645	704	758	26	0	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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

Chain A: 24% 55% 10% • 11%



• Molecule 1: Signal transducer and activator of transcription 3

Chain B: % 27% 53% 9% • 10%

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 N1167 L1168 Q1169 F1172 N1175 T1178 L1179 K1180 S1181 Q1182 G1183 D1184 MET GLN ASP LEU ASN GLY ASN ASN GLN
SER V1195 T1196 R1197 Q1198 K1199 M1200 Q1201 Q1202 L1203 E1204 Q1205 M1206 L1207 T1208 A1209 L1210 D1211 Q1212 M1213 R1214 R1215 V1218 S1219 E1220 L1224 L1225 M1228 E1229 Y1230 V1231 T1234 L1235 T1236 D1237 E1238 E1239 L1240 A1241 D1242 W1243 K1244 R1245 R1246 Q1247 Q1248 I1249 A1250 C1251 I1252 P1255 P1256 N1257 I1258 C1259
L1260 D1261 R1262 L1263 I1267 T1268 S1269 L1270 A1271 E1272 S1273 Q1274 L1275 Q1276 T1277 R1278 Q1279 Q1280 I1281 K1282 K1283 L1284 E1285 E1286 L1287 Q1288 Q1289 Y1293 D1296 P1297 I1298 V1299 R1302 P1303 M1304 R1308 E1311 L1312 F1313 R1314 N1315 L1316 M1317 K1318 S1319 A1320 R1325 Q1326 P1327 C1328 M1329 P1330 M1331
H1332 R1335 P1336 L1337 V1338 I1339 K1340 V1343 Q1344 F1345 T1346 R1350 V1353 K1354 F1355 P1356 E1357 L1358 N1359 Y1360 Q1361 L1362 K1365 V1366 C1367 I1368 D1369 K1370 ASP SER GLY ASP VAL ALA ALA LEU R1379 G1380 S1381 R1382 K1383 F1384 N1385 L1386 L1387 G1388 T1389 N1390 T1391 M1396 E1397 E1398 SER ASN
N1401 L1404 K1409 H1410 L1411 T1412 L1413 R1414 E1415 C1418 GLY ASN GLY GLY ARG ALA ASN CYS ASP ALA SER LEU I1431 V1432 T1433 H1437 L1438 I1439 T1440 F1441 E1442 T1443 E1444 V1445 Y1446 H1447 Q1448 G1449 L1450 K1451 I1452 D1453 L1454 E1455 S1458 L1459 P1460 V1461 V1462 V1463 N1466 I1467 C1468 Q1469
M1470 P1471 N1472 A1473 W1474 A1475 S1476 I1477 L1478 W1479 Y1480 N1481 M1482 L1483 T1484 N1485 N1486 P1487 K1488 F1492 F1493 T1494 K1495 P1496 P1497 I1498 G1499 Q1503 V1504 A1505 E1506 V1507 L1508 S1509 W1510 Q1511 F1512 S1513 S1514 T1515 T1516 K1517 R1518 G1519 L1520 S1521 I1522 E1523 Q1524 L1525 T1526 T1527 L1528 A1529 E1530 K1531 L1532 L1533
G1534 P1535 G1536 V1537 N1538 Y1539 S1540 G1541 C1542 Q1543 I1544 A1547 K1548 F1549 C1550 K1551 E1552 N1553 M1554 A1555 G1556 K1557 G1558 F1559 S1560 F1561 W1562 V1563 W1564 L1565 D1566 N1567 I1568 I1569 D1570 L1571 V1572 K1573 K1574 Y1575 I1576 L1577 A1578 L1579 W1580 N1581 E1582 G1583 Y1584 I1585 M1586 G1587 F1588 I1589 S1590 K1591 E1592 R1593 E1594
R1595 A1596 I1597 L1598 S1599 T1600 K1601 P1602 P1603 G1604 T1605 F1606 L1607 L1608 R1609 F1610 S1611 E1612 S1613 S1614 K1615 E1616 G1617 G1618 V1619 T1620 F1621 T1622 W1623 VAL GLU LYS ASP ILE SER GLY K1631 T1632 Q1633 I1634 Q1635 S1636 V1637 E1638 T1641 K1642 Q1643 Q1644 L1645 N1646 ASN MET SER F1650 A1651 E1652 I1653 I1654 M1655
G1656 Y1657 K1658 ILE MET ASP ALA T1663 N1664 I1665 L1666 V1667 S1668 P1669 L1670 V1671 Y1672 L1673 Y1674 P1675 D1676 I1677 P1678 K1679 E1680 E1681 A1682 F1683 G1684 K1685 Y1686 C1687 R1688

4 Data and refinement statistics (i)

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
$< I/\sigma(I) >$ [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]	$< \|L\| > = 0.41$, $< L^2 > = 0.23$	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.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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	0.47	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 (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	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	B	1379	ARG
1	B	1580	TRP
1	B	1431	ILE
1	B	1523	GLU
1	B	1597	ILE

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

Mol	Chain	Res	Type
1	B	1167	ASN
1	B	1280	GLN
1	B	1481	ASN
1	B	1279	GLN
1	B	1332	HIS

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 monosaccharides 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 monosaccharides 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.