W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 17, 2022 – 09:11 PM EDT

PDB ID	:	3WTS
Title	:	Crystal structure of the complex comprised of ETS1, RUNX1, CBFBETA, and the tcralpha gene enhancer DNA
Authors	:	Shiina, M.; Hamada, K.; Ogata, K.
Deposited on	:	2014-04-21
Resolution	:	2.35 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 (i)) were used in the production of this report:

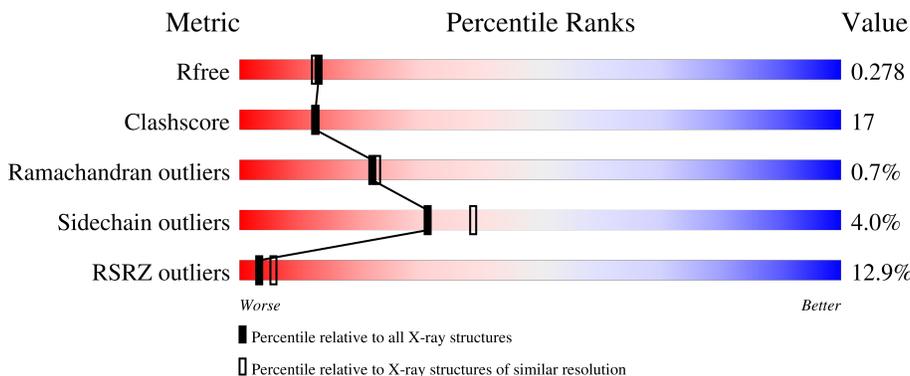|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.29
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 2.35 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	Percentile Ranks	Value
Rfree		0.278
Clashscore		17
Ramachandran outliers		0.7%
Sidechain outliers		4.0%
RSRZ outliers		12.9%

Worse   *Better*

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	205	% 43% 13% • 42%
1	F	205	44% 12% • 42%
2	B	142	18% 53% 37% • 8%
2	G	142	10% 65% 25% • 9%
3	C	166	2% 56% 14% • 29%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	166	32% / 33% / 27% / · / 39%
4	D	15	40% / 47% / 13%
4	I	15	27% / 60% / 13%
5	E	15	87% / 13%
5	J	15	80% / 20%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 7094 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 Runt-related transcription factor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	119	Total 922	C 579	N 170	O 168	S 5	0	0	0
1	F	118	Total 914	C 574	N 169	O 167	S 4	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	59	MET	-	expression tag	UNP Q03347
A	94	LYS	LEU	engineered mutation	UNP Q03347
F	59	MET	-	expression tag	UNP Q03347
F	94	LYS	LEU	engineered mutation	UNP Q03347

- Molecule 2 is a protein called Core-binding factor subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	130	Total 1071	C 671	N 195	O 199	S 6	0	0	0
2	G	129	Total 1062	C 666	N 193	O 197	S 6	0	0	0

- Molecule 3 is a protein called Protein C-ets-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	118	Total 967	C 627	N 165	O 171	S 4	0	0	0
3	H	101	Total 853	C 553	N 147	O 149	S 4	0	0	0

- Molecule 4 is a DNA chain called 5'-D(*GP*AP*AP*GP*CP*CP*AP*CP*AP*TP*CP*CP*TP*CP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	15	Total	C	N	O	P	0	0	0
			299	144	54	87	14			
4	I	15	Total	C	N	O	P	0	0	0
			299	144	54	87	14			

- Molecule 5 is a DNA chain called 5'-D(*AP*GP*AP*GP*GP*AP*TP*GP*TP*GP*GP*CP*TP*TP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	15	Total	C	N	O	P	0	0	0
			310	148	59	89	14			
5	J	15	Total	C	N	O	P	0	0	0
			310	148	59	89	14			

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	18	Total	O	0	0
			18	18		
6	B	3	Total	O	0	0
			3	3		
6	C	16	Total	O	0	0
			16	16		
6	F	13	Total	O	0	0
			13	13		
6	G	3	Total	O	0	0
			3	3		
6	D	7	Total	O	0	0
			7	7		
6	E	14	Total	O	0	0
			14	14		
6	I	2	Total	O	0	0
			2	2		
6	J	11	Total	O	0	0
			11	11		

3 Residue-property plots ⓘ

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: Runt-related transcription factor 1

Chain A:

- Molecule 1: Runt-related transcription factor 1

Chain F:

- Molecule 2: Core-binding factor subunit beta

Chain B:

- Molecule 2: Core-binding factor subunit beta

Chain G:

MET P2 D7 Q8 K11 E15 R19 K20 L21 S22 R23 I27 F32 R35 Q45 E54 I55 A56 F57 T60 G61 T62 N63 L66 Q67 F68 F69 P70 A71 SER TRP GLN GLY GLU GLN ARG GLN THR P81 S82 R83 E84 Y85 V86 D87 L88 G93 K94

V95 Y96 L97 K98 L103 C107 V108 I109 L116 L119 D120 F127 E130 R131 A132 Q133 D136 A139 GLN GLN ALA

- Molecule 3: Protein C-ets-1

Chain C:

SER LEU GLN ARG VAL PRO SER TYR ASP SER PHE ASP SER GLU ASP TYR PRO ALA ALA LEU PRO ASN HIS LYS PRO LYS GLY THR PHE LYS ASP TYR VAL ARG ASP ARG ALA ASP LEU ASN LYS ASP LYS P319 A323 A324 A325 L326 A327 T330 G331 S332 L345 T346 I354

K364 D369 N380 K381 P382 K383 M384 N385 Y386 E387 K388 L389 Y397 K404 V411 D417 L421 L422 T425 E428 K436 PRO ASP ALA ASP GLU

- Molecule 3: Protein C-ets-1

Chain H:

SER LEU GLN ARG VAL PRO SER TYR ASP SER PHE ASP SER GLU ASP TYR PRO ALA ALA LEU PRO ASN HIS LYS PRO LYS GLY THR PHE LYS ASP TYR VAL ARG ASP ARG ALA ASP LEU ASN LYS ASP LYS PRO VAL ILE PRO ALA ALA LEU ALA GLY TYR THR GLY SER G333 P334 I335

Q336 L337 W338 Q339 F340 L341 L342 E343 L344 L345 T346 D347 K348 S349 C350 Q351 S352 F353 I354 S355 W356 T357 G358 D359 G360 W361 E362 F363 K364 L365 S366 D367 P368 D369 E370 V371 A372 R373 R374 W375 G376 K377 N380 K381 P382 K383 M384 N385 Y386 E387 K388 L389 S390 L393 Y397 H403

K404 T405 A406 G407 K408 R409 Y410 V411 V412 R413 F414 V415 C16 D417 L418 Q419 S420 L421 L422 G423 Y424 T425 P426 E427 E428 E429 H430 A431 M432 L433 ASP VAL LYS PRO ASP ALA ASP GLU

- Molecule 4: 5'-D(*GP*AP*AP*GP*CP*CP*AP*CP*AP*TP*CP*CP*TP*CP*T)-3'

Chain D:

G1 G4 C8 A9 T10 C11 C12 T13 C14 T15

- Molecule 4: 5'-D(*GP*AP*AP*GP*CP*CP*AP*CP*AP*TP*CP*CP*TP*CP*T)-3'

Chain I:

G1 A2 A3 G4 A7 C8 A9 T10 C11 C12 T13 C14 T15

- Molecule 5: 5'-D(*AP*GP*AP*GP*GP*AP*TP*GP*TP*GP*GP*CP*TP*TP*C)-3'

Chain E:

A101 T113 T114 C115

• Molecule 5: 5'-D(*AP*GP*AP*GP*GP*AP*TP*GP*TP*GP*GP*CP*TP*TP*C)-3'

Chain J:

80%   20%

A101 G102 G111 C115

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	78.72Å 102.06Å 194.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.99　–　2.35 44.99　–　2.35	Depositor EDS
% Data completeness (in resolution range)	98.8 (44.99-2.35) 98.9 (44.99-2.35)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.01 (at 2.34Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.249　,　0.279 0.249　,　0.278	Depositor DCC
R_{free} test set	6608 reflections (10.13%)	wwPDB-VP
Wilson B-factor (Å^2)	64.1	Xtriage
Anisotropy	0.140	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 46.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	7094	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.21% 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 ⓘ

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	0.40	0/941	0.62	0/1278
1	F	0.48	0/933	0.68	0/1268
2	B	0.34	0/1093	0.49	0/1466
2	G	0.36	0/1084	0.50	0/1454
3	C	0.44	0/994	0.58	0/1342
3	H	0.32	0/877	0.43	0/1181
4	D	0.61	0/334	0.94	1/512 (0.2%)
4	I	0.58	0/334	0.88	1/512 (0.2%)
5	E	0.52	0/348	0.82	0/537
5	J	0.47	0/348	0.74	0/537
All	All	0.43	0/7286	0.63	2/10087 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
4	D	0	2
4	I	0	2
5	E	0	1
All	All	0	5

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	4	DG	N9-C1'-C2'	-5.84	101.50	112.60
4	I	4	DG	N9-C1'-C2'	-5.10	102.91	112.60

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	D	12	DC	Sidechain
4	D	4	DG	Sidechain
5	E	113	DT	Sidechain
4	I	4	DG	Sidechain
4	I	9	DA	Sidechain

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	922	0	931	33	0
1	F	914	0	922	23	0
2	B	1071	0	1034	50	0
2	G	1062	0	1026	31	0
3	C	967	0	966	20	0
3	H	853	0	847	47	0
4	D	299	0	170	13	0
4	I	299	0	170	16	0
5	E	310	0	171	2	0
5	J	310	0	171	5	0
6	A	18	0	0	2	0
6	B	3	0	0	0	0
6	C	16	0	0	1	0
6	D	7	0	0	2	0
6	E	14	0	0	0	0
6	F	13	0	0	0	0
6	G	3	0	0	2	0
6	I	2	0	0	0	0
6	J	11	0	0	0	0
All	All	7094	0	6408	224	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 17.

The worst 5 of 224 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:82:ASN:HD21	1:A:118:ARG:HH11	1.14	0.94
1:A:69:ASN:HD21	1:A:95:GLY:H	1.17	0.90
3:H:345:LEU:HB3	3:H:356:TRP:HE1	1.38	0.86
4:I:10:DT:H2"	4:I:11:DC:H5'	1.63	0.79
1:A:69:ASN:ND2	1:A:95:GLY:H	1.80	0.79

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	117/205 (57%)	111 (95%)	6 (5%)	0	100	100
1	F	116/205 (57%)	111 (96%)	5 (4%)	0	100	100
2	B	126/142 (89%)	109 (86%)	16 (13%)	1 (1%)	19	20
2	G	125/142 (88%)	112 (90%)	13 (10%)	0	100	100
3	C	116/166 (70%)	114 (98%)	2 (2%)	0	100	100
3	H	99/166 (60%)	82 (83%)	13 (13%)	4 (4%)	3	1
All	All	699/1026 (68%)	639 (91%)	55 (8%)	5 (1%)	22	23

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	H	382	PRO
3	H	383	LYS
2	B	116	LEU
3	H	368	PRO
3	H	334	PRO

5.3.2 Protein sidechains ⓘ

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	101/180 (56%)	94 (93%)	7 (7%)	15	15
1	F	100/180 (56%)	96 (96%)	4 (4%)	31	39
2	B	113/123 (92%)	110 (97%)	3 (3%)	44	55
2	G	112/123 (91%)	107 (96%)	5 (4%)	27	33
3	C	102/145 (70%)	98 (96%)	4 (4%)	32	40
3	H	91/145 (63%)	89 (98%)	2 (2%)	52	63
All	All	619/896 (69%)	594 (96%)	25 (4%)	31	39

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

Mol	Chain	Res	Type
1	F	96	ASP
1	F	162	TYR
3	H	384	MET
1	F	139	ARG
2	G	7	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	163	HIS
2	G	134	GLN
2	G	67	GLN
3	H	336	GLN
1	A	132	ASN

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 monosaccharides 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 ⓘ

6.1   Protein, DNA and RNA chains ⓘ

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	119/205 (58%)	0.63	2 (1%)	70	78	42, 57, 81, 113	0
1	F	118/205 (57%)	0.51	1 (0%)	86	91	38, 51, 71, 94	0
2	B	130/142 (91%)	1.22	26 (20%)	1	2	62, 87, 113, 127	0
2	G	129/142 (90%)	1.01	14 (10%)	5	9	56, 78, 106, 111	0
3	C	118/166 (71%)	0.58	4 (3%)	45	57	38, 58, 87, 95	0
3	H	101/166 (60%)	2.59	53 (52%)	0	0	61, 116, 154, 158	0
4	D	15/15 (100%)	-0.06	0	100	100	46, 51, 60, 61	0
4	I	15/15 (100%)	0.04	0	100	100	50, 61, 75, 78	0
5	E	15/15 (100%)	-0.15	0	100	100	41, 48, 61, 61	0
5	J	15/15 (100%)	-0.06	0	100	100	45, 56, 68, 71	0
All	All	775/1086 (71%)	0.97	100 (12%)	3	5	38, 67, 125, 158	0

The worst 5 of 100 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	H	433	LEU	9.0
3	H	353	PHE	8.8
3	H	429	LEU	8.2
3	H	421	LEU	8.1
1	A	59	MET	6.8

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.