W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 10:31 pm BST

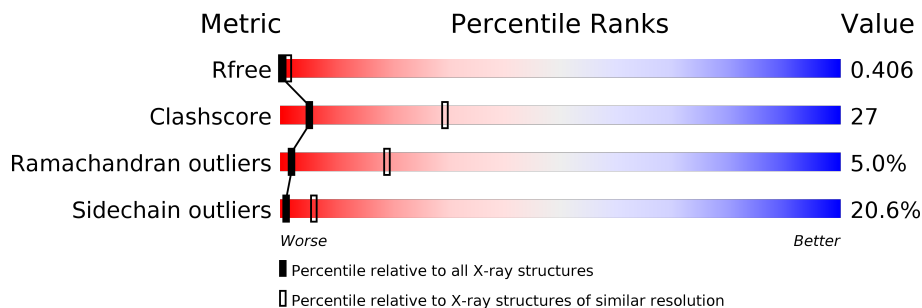|		
PDB ID	:	4XRS
Title	:	Heterodimeric complex of transcription factors MEIS1 and DLX3 on specific DNA
Authors	:	Jorma, A.; Yin, Y.; Nitta, K.R.; Dave, K.; Enge, M.; Kivioja, T.; Popov, A.; Morgunova, E.; Taipale, J.
Deposited on	:	2015-01-21
Resolution	:	3.50 Å(reported)

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 3.50 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		0.406
Clashscore		27
Ramachandran outliers		5.0%
Sidechain outliers		20.6%

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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)

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 <=5%

Mol	Chain	Length	Quality of chain
1	M	15	47% 53%
2	D	17	47% 47% 6%
3	E	18	44% 56%
4	L	17	35% 53% 6% 6%
5	A	58	47% 36% 14% •
5	B	58	41% 48% 9% •
6	G	56	30% 50% 16% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	I	56	32% 50% 18%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 3247 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 DNA chain called DNA (5'-D(P*CP*AP*AP*TP*TP*AP*TP*CP*CP*TP *GP*TP*CP*AP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	M	15	Total	C	N	O	P	0	0	0
			303	146	52	90	15			

- Molecule 2 is a DNA chain called DNA (5'-D(P*AP*CP*AP*AP*TP*TP*AP*TP*CP*CP *TP*GP*TP*CP*AP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	17	Total	C	N	O	P	0	0	0
			343	165	60	101	17			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	18	Total	C	N	O	P	0	0	0
			374	179	67	110	18			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	L	16	Total	C	N	O	P	0	0	0
			332	159	60	97	16			

- Molecule 5 is a protein called Homeobox protein Meis1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	A	56	Total	C	N	O	S	0	0	0
			468	301	87	78	2			
5	B	58	Total	C	N	O	S	0	0	0
			487	314	89	82	2			

• Molecule 6 is a protein called Homeobox protein DLX-3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
6	G	56	472	304	90	78	0	0	0
6	I	56	Total	C	N	O	0	0	0
			468	301	89	78			

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. 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. 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.

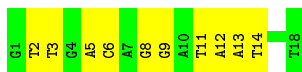- Molecule 1: DNA (5'-D(P*CP*AP*AP*TP*TP*AP*TP*CP*CP*TP*GP*TP*CP*AP*A)-3')

Chain M:

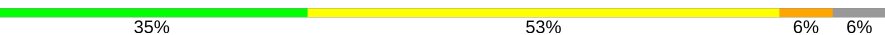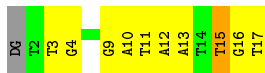- Molecule 2: DNA (5'-D(P*AP*CP*AP*AP*TP*TP*AP*TP*CP*CP*TP*GP*TP*CP*AP*AP*C)-3')

Chain D:

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

Chain E:

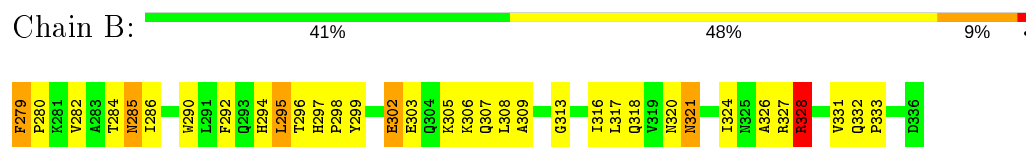- Molecule 4: DNA (5'-D(P*TP*TP*GP*AP*CP*AP*GP*GP*AP*TP*AP*AP*TP*TP*GP*T)-3')

Chain L:

- Molecule 5: Homeobox protein Meis1

Chain A:

• Molecule 5: Homeobox protein Meis1

Chain B:

41%   48%   9%

F279 P280 K281 V282 A283 T284 N285 I286 W290 L291 F292 Q293 H294 I295 T296 H297 P298 Y299 E302 E303 Q304 K305 K306 Q307 L308 A309 G313 I316 L317 Q318 V319 N320 N321 I324 N325 A326 R327 R328 V331 Q332 P333 D336

• Molecule 6: Homeobox protein DLX-3

Chain G:

30%   50%   16%

K131 P132 R133 T134 I135 Y136 S137 S138 Y139 Q140 L141 A142 A143 L144 Q145 R146 R147 K150 A151 L154 A155 E158 R159 A160 E161 L162 A163 A164 Q165 L166 G167 L168 T169 Q172 V173 Q178 R181 S182 K183 K186

• Molecule 6: Homeobox protein DLX-3

Chain I:

32%   50%   18%

K131 P132 R133 T134 I135 Y136 S137 S138 Y139 Q140 L141 L144 Q145 R146 R147 F148 Q149 K150 L154 A155 L156 P157 E158 R159 A160 E161 L162 Q165 L166 G167 L168 T169 Q170 K174 I175 W176 F177 Q178 M179 R180 R181 F184 K185 K186

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	69.64Å　　69.84Å　　116.89Å 90.00°　　90.00°　　90.00°	Depositor
Resolution (Å)	49.31　–　3.50 49.31　–　2.97	Depositor EDS
% Data completeness (in resolution range)	88.1 (49.31-3.50) 89.0 (49.31-2.97)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.06 (at 2.96Å)	Xtriage
Refinement program	PHENIX phenix.refine: 1.9_1692	Depositor
R, R_{free}	0.331　,　0.359 0.389　,　0.406	Depositor DCC
R_{free} test set	529 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å^2)	101.7	Xtriage
Anisotropy	0.765	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.15 , 76.5	EDS
L-test for twinning [2]	$<|L|> = 0.47, <L^2> = 0.30$	Xtriage
Estimated twinning fraction	0.001 for k,h,-l	Xtriage
Reported twinning fraction	0.080 for k,h,-l	Depositor
Outliers	2 of 10985 reflections (0.018%)	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	3247	wwPDB-VP
Average B, all atoms (Å^2)	136.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 22.09 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.1274e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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	M	0.72	0/338	1.07	0/518
2	D	0.76	0/383	1.12	1/587 (0.2%)
3	E	0.70	0/419	1.11	0/646
4	L	0.64	0/372	1.10	1/573 (0.2%)
5	A	0.33	0/480	0.56	0/651
5	B	0.32	0/500	0.64	0/679
6	G	0.35	0/481	0.74	0/644
6	I	0.34	0/477	0.67	0/640
All	All	0.53	0/3450	0.89	2/4938 (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
5	B	0	1
6	G	0	1
All	All	0	2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	33	DC	O4'-C1'-N1	6.28	112.40	108.00
4	L	15	DT	O4'-C1'-N1	5.40	111.78	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	B	328	ARG	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
6	G	181	ARG	Peptide

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	M	303	0	171	13	0
2	D	343	0	193	15	0
3	E	374	0	206	18	0
4	L	332	0	183	15	0
5	A	468	0	479	20	0
5	B	487	0	491	33	0
6	G	472	0	497	35	0
6	I	468	0	486	40	0
All	All	3247	0	2706	156	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 27.

The worst 5 of 156 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)
6:I:156:LEU:O	6:I:160:ALA:N	2.03	0.91
5:A:309:ALA:O	5:A:313:GLY:N	2.01	0.91
4:L:13:DA:O3'	6:G:131:LYS:N	2.05	0.89
5:B:279:PHE:N	5:B:284:THR:HG1	1.75	0.84
3:E:12:DA:N7	6:I:179:ASN:ND2	2.29	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	
5	A	54/58 (93%)	45 (83%)	7 (13%)	2 (4%)	3	26
5	B	56/58 (97%)	51 (91%)	4 (7%)	1 (2%)	8	41
6	G	54/56 (96%)	42 (78%)	8 (15%)	4 (7%)	1	11
6	I	54/56 (96%)	38 (70%)	12 (22%)	4 (7%)	1	11
All	All	218/228 (96%)	176 (81%)	31 (14%)	11 (5%)	2	19

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
6	G	138	SER
6	G	140	GLN
5	A	309	ALA
6	G	134	THR
6	G	160	ALA

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	
5	A	51/53 (96%)	40 (78%)	11 (22%)	1	5
5	B	53/53 (100%)	47 (89%)	6 (11%)	6	27
6	G	48/48 (100%)	35 (73%)	13 (27%)	0	3
6	I	47/48 (98%)	36 (77%)	11 (23%)	1	4
All	All	199/202 (98%)	158 (79%)	41 (21%)	1	6

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	G	135	ILE
6	G	158	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
6	I	169	THR
6	G	137	SER
6	G	138	SER

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

Mol	Chain	Res	Type
5	A	320	ASN
5	A	321	ASN
6	G	172	GLN
6	I	140	GLN

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 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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3   Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4   Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5   Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.