wwPDB X-ray Structure Validation Summary Report (i)

May 16, 2020 – 01:42 pm BST

PDB ID	:	1S9K
Title	:	Crystal Structure of Human NFAT1 and Fos-Jun on the IL-2 ARRE1 Site
Authors	:	Wang, D.; Stroud, J.C.; Chen, L.
Deposited on	:	2004-02-04
Resolution	:	3.10 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 following versions of software and data (see references (1)) were used in the production of this report:

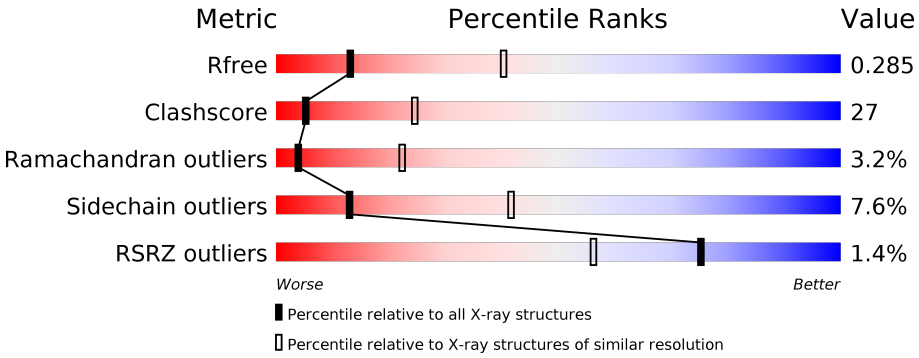
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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% 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	20	15% 85%
2	B	20	5% 80% 15%
3	C	280	2% 49% 44% 6%
4	D	53	57% 40% .
5	E	52	62% 38%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3887 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 Human IL-2 ARRE1 Promoter Element, Plus Strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	20	Total	C	N	O	P	0	0	0
			413	200	76	118	19			

- Molecule 2 is a DNA chain called Human IL-2 ARRE1 Promoter Element, Minus Strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	20	Total	C	N	O	P	0	0	0
			401	196	68	118	19			

- Molecule 3 is a protein called Nuclear factor of activated T-cells, cytoplasmic 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	280	Total	C	N	O	S	0	0	0
			2204	1383	404	408	9			

- Molecule 4 is a protein called Proto-oncogene protein c-fos.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	53	Total	C	N	O	S	0	0	0
			443	262	92	87	2			

- Molecule 5 is a protein called Transcription factor AP-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	52	Total	C	N	O	S	0	0	0
			426	257	92	74	3			

3 Residue-property plots (i)

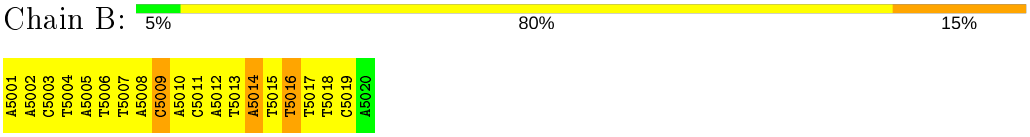
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Human IL-2 ARRE1 Promoter Element, Plus Strand

Chain A:

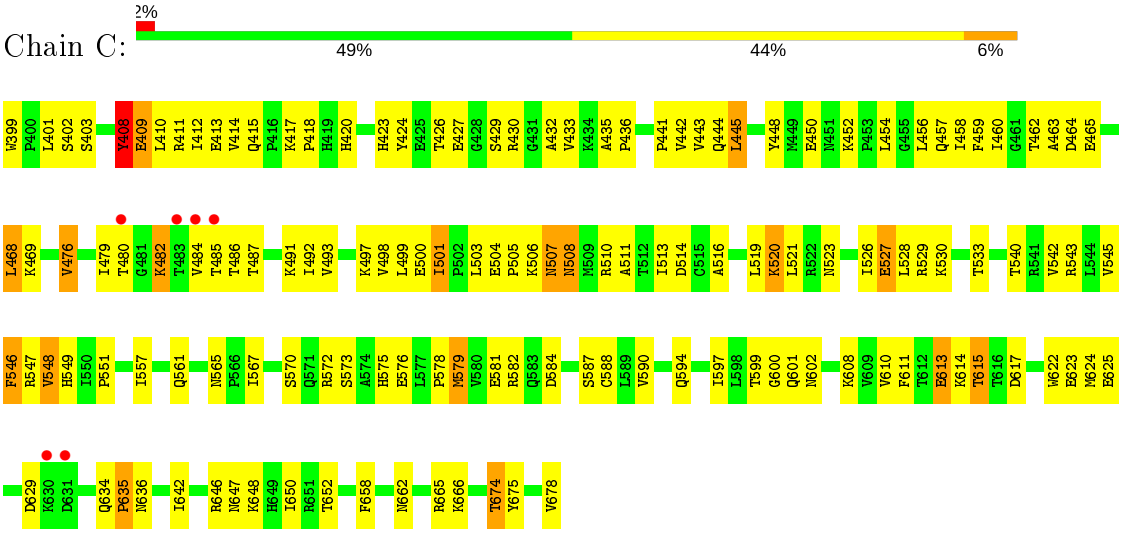
- Molecule 2: Human IL-2 ARRE1 Promoter Element, Minus Strand

Chain B:



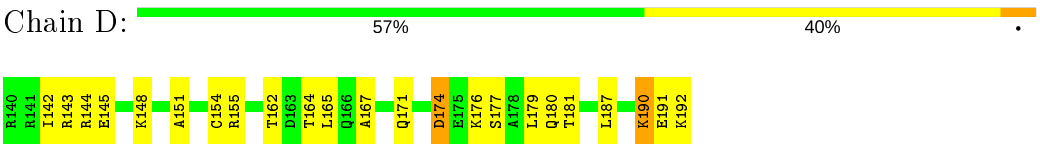
- Molecule 3: Nuclear factor of activated T-cells, cytoplasmic 2

Chain C:



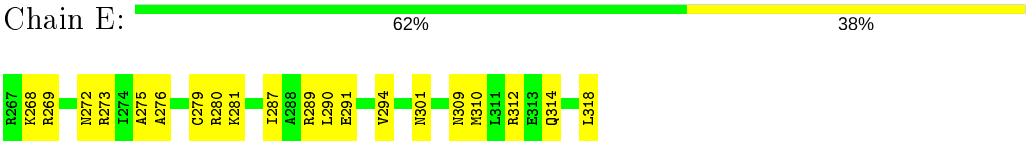
- Molecule 4: Proto-oncogene protein c-fos

Chain D:



• Molecule 5: Transcription factor AP-1

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	64.77Å 86.26Å 84.04Å 90.00° 111.24° 90.00°	Depositor
Resolution (Å)	30.00 – 3.10 29.39 – 3.05	Depositor EDS
% Data completeness (in resolution range)	(Not available) (30.00-3.10) 95.3 (29.39-3.05)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.22 (at 3.05Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.242 , 0.275 0.250 , 0.285	Depositor DCC
R_{free} test set	1578 reflections (9.77%)	wwPDB-VP
Wilson B-factor (Å^2)	48.5	Xtriage
Anisotropy	0.934	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 50.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.035 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	3887	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.60% 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	0.59	0/464	0.81	0/716
2	B	0.63	0/448	0.83	0/688
3	C	0.46	0/2253	0.76	1/3050 (0.0%)
4	D	1.16	1/442 (0.2%)	0.68	0/583
5	E	1.24	1/425 (0.2%)	0.68	0/558
All	All	0.72	2/4032 (0.0%)	0.76	1/5595 (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
2	B	0	3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	E	279	CYS	CB-SG	-24.04	1.41	1.82
4	D	154	CYS	CB-SG	-23.07	1.43	1.82

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	408	TYR	N-CA-C	-5.50	96.15	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	5009	DC	Sidechain
2	B	5014	DA	Sidechain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
2	B	5016	DT	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	413	0	230	26	0
2	B	401	0	230	23	0
3	C	2204	0	2168	136	0
4	D	443	0	461	16	0
5	E	426	0	472	24	0
All	All	3887	0	3561	198	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 198 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)
3:C:613:GLU:HB2	3:C:622:TRP:HB3	1.36	1.04
2:B:5005:DA:H1'	2:B:5006:DT:H5'	1.51	0.92
3:C:441:PRO:HG2	3:C:513:ILE:HB	1.60	0.82
3:C:521:LEU:N	3:C:521:LEU:HD12	1.95	0.81
4:D:162:THR:HG22	5:E:287:ILE:HG12	1.65	0.78

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
3	C	278/280 (99%)	237 (85%)	32 (12%)	9 (3%)	4 22
4	D	51/53 (96%)	45 (88%)	4 (8%)	2 (4%)	3 18
5	E	50/52 (96%)	39 (78%)	10 (20%)	1 (2%)	7 31
All	All	379/385 (98%)	321 (85%)	46 (12%)	12 (3%)	4 22

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	480	THR
3	C	590	VAL
3	C	527	GLU
3	C	629	ASP
3	C	635	PRO

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
3	C	238/250 (95%)	215 (90%)	23 (10%)	8 30
4	D	47/47 (100%)	45 (96%)	2 (4%)	29 62
5	E	45/45 (100%)	45 (100%)	0	100 100
All	All	330/342 (96%)	305 (92%)	25 (8%)	13 41

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

Mol	Chain	Res	Type
3	C	507	ASN
3	C	520	LYS
4	D	165	LEU
3	C	508	ASN
3	C	546	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such

sidechains are listed below:

Mol	Chain	Res	Type
3	C	539	ASN
3	C	565	ASN
3	C	676	HIS
3	C	523	ASN
3	C	669	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

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, 95^{th} 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	20/20 (100%)	-0.46	0 100 100	29, 39, 68, 81	0
2	B	20/20 (100%)	-0.50	0 100 100	28, 46, 60, 70	0
3	C	280/280 (100%)	-0.21	6 (2%) 63 43	20, 55, 112, 151	0
4	D	53/53 (100%)	-0.20	0 100 100	25, 56, 116, 138	0
5	E	52/52 (100%)	-0.09	0 100 100	24, 63, 116, 126	0
All	All	425/425 (100%)	-0.22	6 (1%) 75 56	20, 53, 112, 151	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	483	THR	14.8
3	C	485	THR	6.0
3	C	484	VAL	4.5
3	C	631	ASP	4.1
3	C	480	THR	3.1

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 carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.