wwPDB X-ray Structure Validation Summary Report (i)

May 17, 2020 – 05:40 pm BST

PDB ID	:	4GO6
Title	:	Crystal structure of HCF-1 self-association sequence 1
Authors	:	Park, J.; Lammers, F.; Herr, W.; Song, J.
Deposited on	:	2012-08-18
Resolution	:	2.70 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:

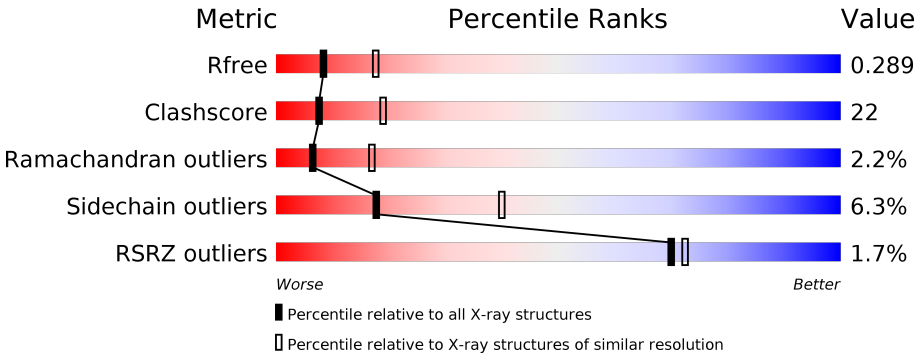
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	45	2% 38% 44% 9% 9%
1	C	45	2% 47% 38% • 11%
2	B	232	45% 24% • 28%
2	D	232	2% 50% 21% • 25%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3295 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 HCF N-terminal chain 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	41	Total	C	N	O	0	0	0
			320	203	55	62			
1	C	40	Total	C	N	O	0	0	0
			311	198	54	59			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	358	GLY	-	EXPRESSION TAG	UNP P51610
A	359	SER	-	EXPRESSION TAG	UNP P51610
C	358	GLY	-	EXPRESSION TAG	UNP P51610
C	359	SER	-	EXPRESSION TAG	UNP P51610

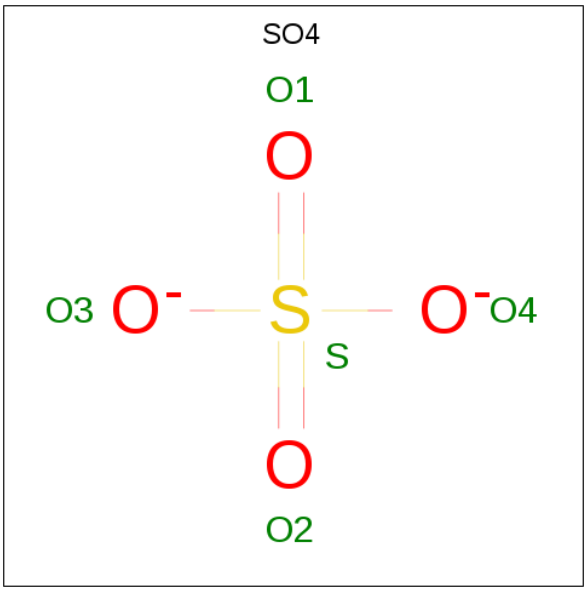
- Molecule 2 is a protein called HCF C-terminal chain 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	168	Total	C	N	O	S	Se	0	0	0
			1295	833	222	232	5	3			
2	D	173	Total	C	N	O	S	Se	0	0	0
			1316	845	223	241	5	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1804	GLY	-	EXPRESSION TAG	UNP P51610
B	1805	SER	-	EXPRESSION TAG	UNP P51610
D	1804	GLY	-	EXPRESSION TAG	UNP P51610
D	1805	SER	-	EXPRESSION TAG	UNP P51610

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	C	1	Total O S 5 4 1	0	0
3	D	1	Total O S 5 4 1	0	0

- Molecule 4 is water.

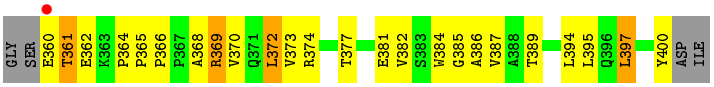
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	3	Total O 3 3	0	0
4	B	12	Total O 12 12	0	0
4	C	2	Total O 2 2	0	0
4	D	16	Total O 16 16	0	0

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: HCF N-terminal chain 1

Chain A:

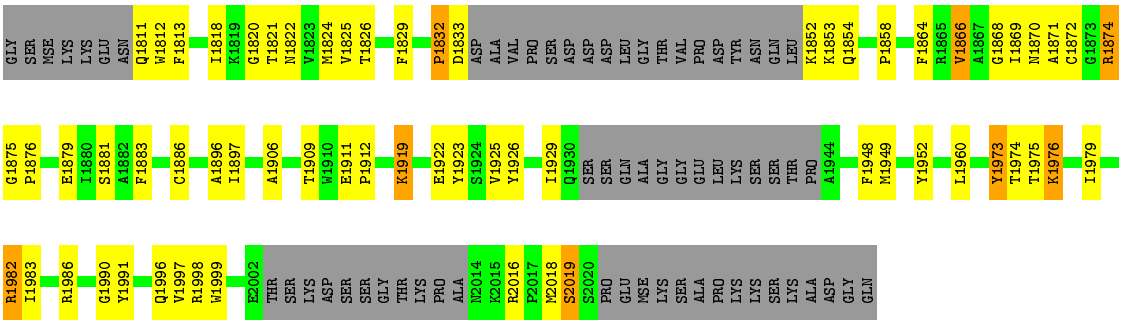


- Molecule 1: HCF N-terminal chain 1

Chain C:

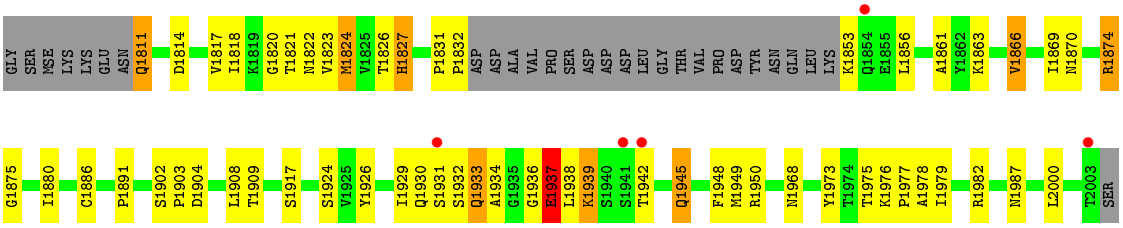
- Molecule 2: HCF C-terminal chain 1

Chain B:



- Molecule 2: HCF C-terminal chain 1

Chain D:



LYS
ASP
SER
SER
GLY
THR
LYS
PRO
ALA
ASN
LYS
ARG
PRO
MSE
SER
SER
PRO
GLU
MSE
LYS
SER
ALA
PRO
LYS
LYS
SER
LYS
ALA
ASP
GLY
GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	94.90Å 183.51Å 86.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.94 – 2.70 33.21 – 2.70	Depositor EDS
% Data completeness (in resolution range)	93.6 (19.94-2.70) 97.0 (33.21-2.70)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.67 (at 2.68Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.231 , 0.282 0.240 , 0.289	Depositor DCC
R_{free} test set	2035 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	45.2	Xtriage
Anisotropy	0.791	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 62.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	3295	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 4.99% 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

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.35	0/327	0.64	0/448
1	C	0.36	0/318	0.64	0/436
2	B	0.38	0/1327	0.65	0/1794
2	D	0.40	0/1351	0.70	1/1834 (0.1%)
All	All	0.39	0/3323	0.67	1/4512 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	1933	GLN	N-CA-C	5.48	125.79	111.00

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	320	0	320	29	0
1	C	311	0	314	20	0
2	B	1295	0	1292	70	0
2	D	1316	0	1309	59	0
3	B	10	0	0	0	0
3	C	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	5	0	0	0	0
4	A	3	0	0	0	0
4	B	12	0	0	0	0
4	C	2	0	0	0	0
4	D	16	0	0	6	0
All	All	3295	0	3235	143	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 22.

The worst 5 of 143 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)
2:B:1919:LYS:HD3	2:B:1919:LYS:H	1.10	1.14
2:B:1811:GLN:HG3	2:B:1812:TRP:H	1.26	0.98
2:B:1919:LYS:HD3	2:B:1919:LYS:N	1.85	0.91
1:C:381:GLU:HB2	2:D:1824:MSE:HE2	1.53	0.91
1:A:368:ALA:HB3	1:A:385:GLY:HA3	1.53	0.90

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
1	A	39/45 (87%)	33 (85%)	4 (10%)	2 (5%)	2 3
1	C	38/45 (84%)	34 (90%)	3 (8%)	1 (3%)	5 13
2	B	160/232 (69%)	145 (91%)	12 (8%)	3 (2%)	8 20
2	D	169/232 (73%)	150 (89%)	16 (10%)	3 (2%)	8 21
All	All	406/554 (73%)	362 (89%)	35 (9%)	9 (2%)	6 17

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	361	THR
1	A	369	ARG
2	D	1937	GLU
1	C	369	ARG
2	B	1871	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	
1	A	35/38 (92%)	31 (89%)	4 (11%)	5	13
1	C	34/38 (90%)	32 (94%)	2 (6%)	19	43
2	B	140/188 (74%)	133 (95%)	7 (5%)	24	51
2	D	142/188 (76%)	133 (94%)	9 (6%)	18	40
All	All	351/452 (78%)	329 (94%)	22 (6%)	18	40

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	1976	LYS
1	C	394	LEU
2	D	1937	GLU
2	B	1982	ARG
1	C	382	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	1996	GLN
2	D	1968	ASN
2	D	1930	GLN
2	B	1854	GLN
2	D	1945	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

4 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	SO4	D	2101	-	4,4,4	0.26	0	6,6,6	0.06	0
3	SO4	B	2101	-	4,4,4	0.19	0	6,6,6	0.12	0
3	SO4	C	501	-	4,4,4	0.31	0	6,6,6	0.10	0
3	SO4	B	2102	-	4,4,4	0.25	0	6,6,6	0.10	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

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	41/45 (91%)	0.29	1 (2%)	59	60	42, 65, 89, 130	0
1	C	40/45 (88%)	0.13	1 (2%)	57	59	39, 63, 84, 113	0
2	B	165/232 (71%)	0.04	0	100	100	32, 55, 92, 124	0
2	D	171/232 (73%)	0.03	5 (2%)	51	52	19, 49, 86, 142	0
All	All	417/554 (75%)	0.07	7 (1%)	70	72	19, 55, 89, 142	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	1941	SER	3.5
1	C	388	ALA	2.9
2	D	1942	THR	2.8
1	A	360	GLU	2.7
2	D	2003	THR	2.6

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	SO4	D	2101	5/5	0.93	0.09	87,93,104,113	0
3	SO4	C	501	5/5	0.94	0.11	68,75,78,82	0
3	SO4	B	2101	5/5	0.96	0.12	69,71,74,77	0
3	SO4	B	2102	5/5	0.97	0.15	62,66,70,80	0

6.5 Other polymers (i)

There are no such residues in this entry.