WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 01:21 pm GMT

PDB ID	:	4AYM
Title	:	Structure of a complex between CCPs 6 and 7 of Human Complement Factor H and Neisseria meningitidis FHbp Variant 3 P106A mutant
Authors	:	Johnson, S.; Tan, L.; van der Veen, S.; Caesar, J.; Goicoechea De Jorge, E.; Everett, R.J.; Bai, X.; Exley, R.M.; Ward, P.N.; Ruivo, N.; Trivedi, K.; Cumber, E.; Jones, R.; Newham, L.; Staunton, D.; Borrow, R.; Pickering, M.; Lea, S.M.; Tang, C.M.
Deposited on	:	2012-06-21
Resolution	:	3.00 Å(reported)

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

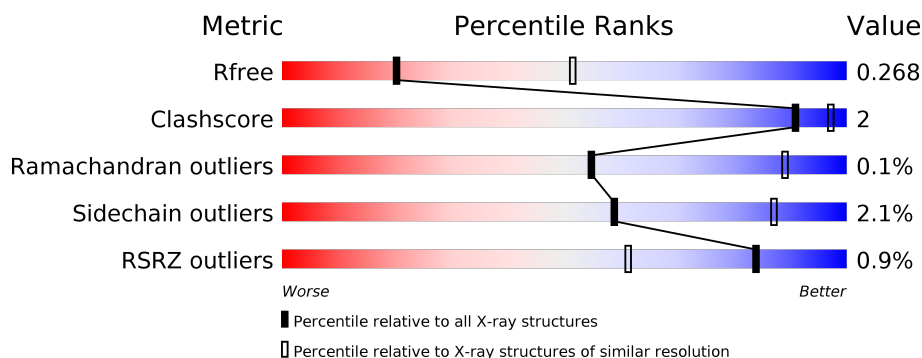|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

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 3.00 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		0.268
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		2.1%
RSRZ outliers		0.9%

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$_{free}$	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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. 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	125	92% · 5%
1	B	125	86% 8% 6%
1	E	125	87% 7% · 5%
1	F	125	90% 5% 5%
2	C	270	80% 6% · 13%
2	D	270	83% 6% 11%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 7506 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 COMPLEMENT FACTOR H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	119	Total	C	N	O	S	0	0	0
			968	618	166	174	10			
1	B	118	Total	C	N	O	S	0	0	0
			960	612	165	173	10			
1	E	119	Total	C	N	O	S	0	0	0
			968	618	166	174	10			
1	F	119	Total	C	N	O	S	0	0	0
			968	618	166	174	10			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	319	MET	-	EXPRESSION TAG	UNP P08603
A	320	GLY	-	EXPRESSION TAG	UNP P08603
A	402	HIS	TYR	VARIANT	UNP P08603
B	319	MET	-	EXPRESSION TAG	UNP P08603
B	320	GLY	-	EXPRESSION TAG	UNP P08603
B	402	HIS	TYR	VARIANT	UNP P08603
E	319	MET	-	EXPRESSION TAG	UNP P08603
E	320	GLY	-	EXPRESSION TAG	UNP P08603
E	402	HIS	TYR	VARIANT	UNP P08603
F	319	MET	-	EXPRESSION TAG	UNP P08603
F	320	GLY	-	EXPRESSION TAG	UNP P08603
F	402	HIS	TYR	VARIANT	UNP P08603

- Molecule 2 is a protein called FACTOR H BINDING PROTEIN.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	234	Total	C	N	O	0	0	0
			1794	1122	316	356			
2	D	240	Total	C	N	O	0	0	0
			1842	1150	324	368			

WORLDWIDE PDB
PROTEIN DATA BANK

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	61	MET	-	EXPRESSION TAG	UNP Q19KF7
C	62	GLY	-	EXPRESSION TAG	UNP Q19KF7
C	63	PRO	-	EXPRESSION TAG	UNP Q19KF7
C	64	ASP	-	EXPRESSION TAG	UNP Q19KF7
C	65	SER	-	EXPRESSION TAG	UNP Q19KF7
C	66	ASP	-	EXPRESSION TAG	UNP Q19KF7
C	67	ARG	-	EXPRESSION TAG	UNP Q19KF7
C	68	LEU	-	EXPRESSION TAG	UNP Q19KF7
C	69	GLN	-	EXPRESSION TAG	UNP Q19KF7
C	70	GLN	-	EXPRESSION TAG	UNP Q19KF7
C	71	ARG	-	EXPRESSION TAG	UNP Q19KF7
C	72	ARG	-	EXPRESSION TAG	UNP Q19KF7
C	321	LEU	-	EXPRESSION TAG	UNP Q19KF7
C	322	GLU	-	EXPRESSION TAG	UNP Q19KF7
C	323	HIS	-	EXPRESSION TAG	UNP Q19KF7
C	324	HIS	-	EXPRESSION TAG	UNP Q19KF7
C	325	HIS	-	EXPRESSION TAG	UNP Q19KF7
C	326	HIS	-	EXPRESSION TAG	UNP Q19KF7
C	327	HIS	-	EXPRESSION TAG	UNP Q19KF7
C	328	HIS	-	EXPRESSION TAG	UNP Q19KF7
C	106	ALA	PRO	ENGINEERED MUTATION	UNP Q19KF7
D	61	MET	-	EXPRESSION TAG	UNP Q19KF7
D	62	GLY	-	EXPRESSION TAG	UNP Q19KF7
D	63	PRO	-	EXPRESSION TAG	UNP Q19KF7
D	64	ASP	-	EXPRESSION TAG	UNP Q19KF7
D	65	SER	-	EXPRESSION TAG	UNP Q19KF7
D	66	ASP	-	EXPRESSION TAG	UNP Q19KF7
D	67	ARG	-	EXPRESSION TAG	UNP Q19KF7
D	68	LEU	-	EXPRESSION TAG	UNP Q19KF7
D	69	GLN	-	EXPRESSION TAG	UNP Q19KF7
D	70	GLN	-	EXPRESSION TAG	UNP Q19KF7
D	71	ARG	-	EXPRESSION TAG	UNP Q19KF7
D	72	ARG	-	EXPRESSION TAG	UNP Q19KF7
D	321	LEU	-	EXPRESSION TAG	UNP Q19KF7
D	322	GLU	-	EXPRESSION TAG	UNP Q19KF7
D	323	HIS	-	EXPRESSION TAG	UNP Q19KF7
D	324	HIS	-	EXPRESSION TAG	UNP Q19KF7
D	325	HIS	-	EXPRESSION TAG	UNP Q19KF7
D	326	HIS	-	EXPRESSION TAG	UNP Q19KF7
D	327	HIS	-	EXPRESSION TAG	UNP Q19KF7
D	328	HIS	-	EXPRESSION TAG	UNP Q19KF7
D	106	ALA	PRO	ENGINEERED MUTATION	UNP Q19KF7

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	O 1	0	0
3	B	1	Total 1	O 1	0	0
3	C	2	Total 2	O 2	0	0
3	D	2	Total 2	O 2	0	0

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 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.
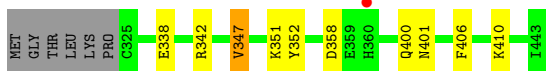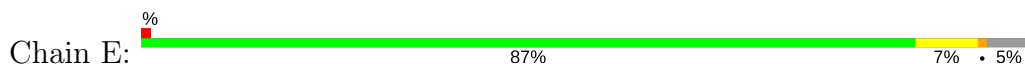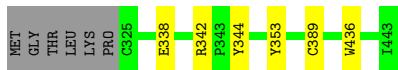
- Molecule 1: COMPLEMENT FACTOR H
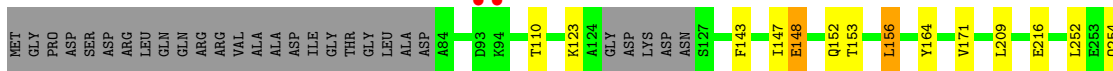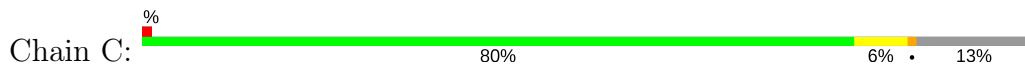

- Molecule 1: COMPLEMENT FACTOR H
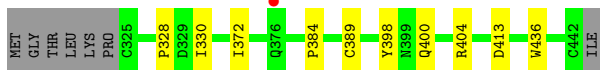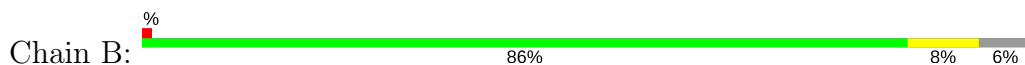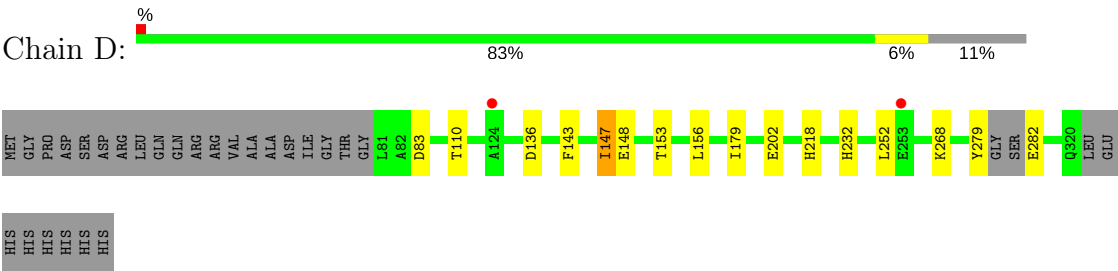

- Molecule 1: COMPLEMENT FACTOR H

- Molecule 1: COMPLEMENT FACTOR H

- Molecule 2: FACTOR H BINDING PROTEIN

• Molecule 2: FACTOR H BINDING PROTEIN

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	78.09Å 83.49Å 361.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 3.00 180.72 – 2.85	Depositor EDS
% Data completeness (in resolution range)	99.9 (15.00-3.00) 99.9 (180.72-2.85)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.60 (at 2.86Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, R_{free}	0.254 , 0.257 0.268 , 0.268	Depositor DCC
R_{free} test set	1223 reflections (5.32%)	DCC
Wilson B-factor (Å^2)	57.0	Xtriage
Anisotropy	0.400	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 37.0	EDS
L-test for twinning[2]	$<|L|> = 0.47, <L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	7506	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.84% 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	0.36	0/1007	0.59	0/1373
1	B	0.37	0/999	0.61	0/1362
1	E	0.36	0/1007	0.61	0/1373
1	F	0.36	0/1007	0.63	0/1373
2	C	0.38	0/1819	0.72	0/2442
2	D	0.40	0/1868	0.71	1/2510 (0.0%)
All	All	0.38	0/7707	0.66	1/10433 (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	282	GLU	C-N-CA	5.50	135.45	121.70

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	968	0	877	2	0
1	B	960	0	866	6	0
1	E	968	0	877	5	0
1	F	968	0	877	4	0
2	C	1794	0	1785	8	0
2	D	1842	0	1828	6	0

Continued on next page...

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

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	2	0	0	0	0
3	D	2	0	0	0	0
All	All	7506	0	7110	29	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 2.

The worst 5 of 29 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:E:347:VAL:HG22	1:E:351:LYS:HG3	1.74	0.68
2:C:147:ILE:HG12	2:C:156:LEU:HD13	1.80	0.64
2:C:254:GLN:HG2	1:E:352:TYR:OH	1.98	0.63
1:B:330:ILE:HD11	1:B:372:ILE:HD13	1.87	0.57
2:D:148:GLU:HG2	2:D:153:THR:HG22	1.90	0.54

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	117/125 (94%)	114 (97%)	3 (3%)	0	100	100
1	B	116/125 (93%)	112 (97%)	3 (3%)	1 (1%)	20	62
1	E	117/125 (94%)	114 (97%)	3 (3%)	0	100	100
1	F	117/125 (94%)	113 (97%)	4 (3%)	0	100	100
2	C	228/270 (84%)	226 (99%)	2 (1%)	0	100	100
2	D	236/270 (87%)	232 (98%)	4 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	931/1040 (90%)	911 (98%)	19 (2%)	1 (0%)	55	89

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	404	ARG

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	104/109 (95%)	104 (100%)	0	100	100
1	B	103/109 (94%)	101 (98%)	2 (2%)	62	88
1	E	104/109 (95%)	102 (98%)	2 (2%)	62	88
1	F	104/109 (95%)	104 (100%)	0	100	100
2	C	189/218 (87%)	182 (96%)	7 (4%)	39	76
2	D	194/218 (89%)	188 (97%)	6 (3%)	45	80
All	All	798/872 (92%)	781 (98%)	17 (2%)	59	87

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	268	LYS
2	C	307	ILE
2	D	147	ILE
2	C	216	GLU
2	D	202	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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, 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(Å²)	Q<0.9
1	A	119/125 (95%)	-0.01	1 (0%)	86	64	33, 57, 83, 93	0
1	B	118/125 (94%)	-0.05	1 (0%)	86	64	31, 58, 83, 96	0
1	E	119/125 (95%)	0.09	1 (0%)	86	64	40, 60, 82, 96	0
1	F	119/125 (95%)	-0.28	0	100	100	23, 41, 61, 76	0
2	C	234/270 (86%)	0.04	4 (1%)	70	42	33, 62, 102, 118	0
2	D	240/270 (88%)	-0.21	2 (0%)	86	64	28, 49, 76, 93	0
All	All	949/1040 (91%)	-0.07	9 (0%)	84	61	23, 55, 83, 118	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	360	HIS	3.3
1	B	376	GLN	3.2
2	C	269	SER	2.9
2	D	253	GLU	2.7
2	C	295	ARG	2.5

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.5 Other polymers ⓘ

There are no such residues in this entry.