



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 3, 2025 – 01:39 PM EDT

PDB ID : 4G9I / pdb\_00004g9i  
Title : Crystal structure of T.kodakarensis HypF  
Authors : Tominaga, T.; Watanabe, S.; Matsumi, R.; Atomi, H.; Imanaka, T.; Miki, K.  
Deposited on : 2012-07-24  
Resolution : 4.50 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](mailto: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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0rc1  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 2.0rc1  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.006 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.43.1

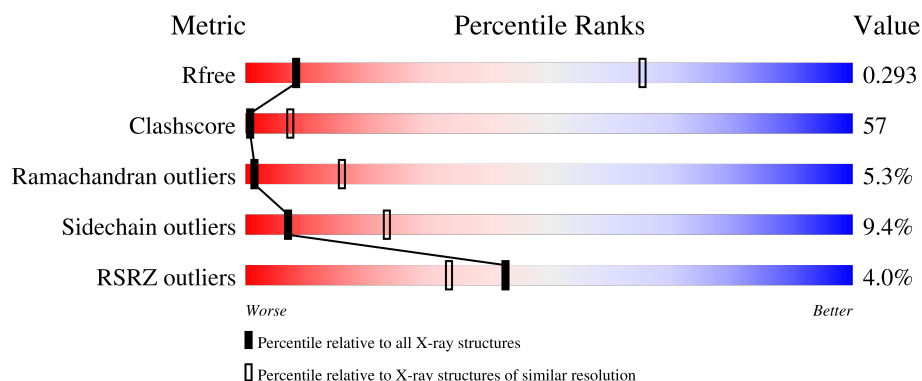
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.50 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	164625	1050 (5.10-3.90)
Clashscore	180529	1106 (5.10-3.90)
Ramachandran outliers	177936	1006 (5.10-3.90)
Sidechain outliers	177891	1008 (5.12-3.88)
RSRZ outliers	164620	1046 (5.10-3.90)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	772	
1	B	772	
1	C	772	
1	D	772	
1	E	772	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	772	<div><div></div><div>5%</div><div>26%</div><div>64%</div><div>9%</div><div></div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 36065 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 Hydrogenase maturation protein HypF.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	766	Total	C	N	O	S	Se	0	0	0
			6032	3862	1028	1113	11	18			
1	B	762	Total	C	N	O	S	Se	0	0	0
			6008	3847	1024	1108	11	18			
1	C	756	Total	C	N	O	S	Se	0	0	0
			5964	3819	1017	1100	11	17			
1	D	765	Total	C	N	O	S	Se	0	0	0
			6027	3859	1027	1112	11	18			
1	E	762	Total	C	N	O	S	Se	0	0	0
			6008	3847	1024	1108	11	18			
1	F	762	Total	C	N	O	S	Se	0	0	0
			6008	3847	1024	1108	11	18			

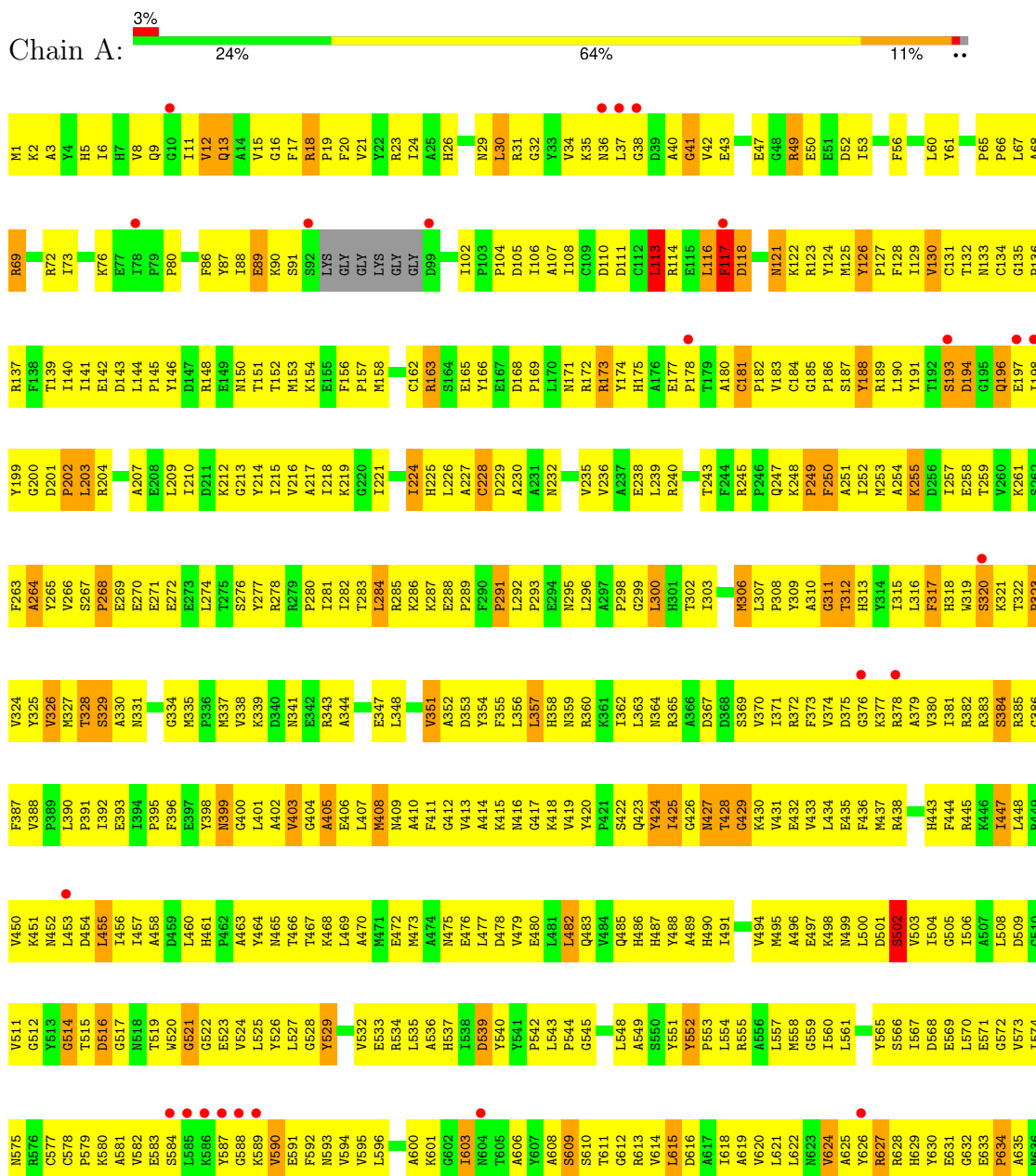
- Molecule 2 is ZINC ION (CCD ID: ZN) (formula: Zn).

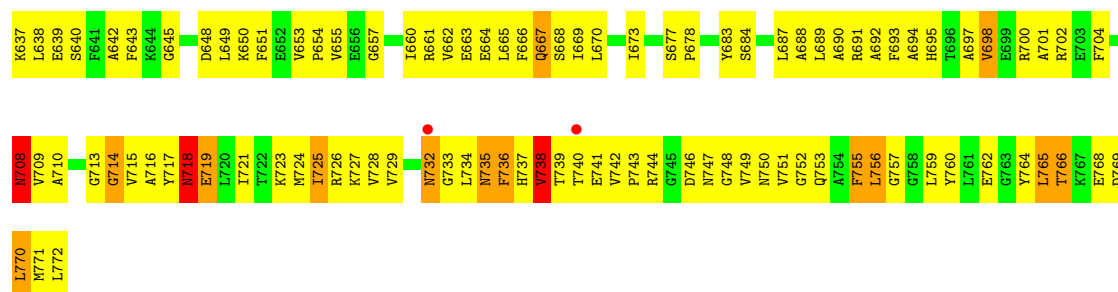
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Zn	0	0
			3	3		
2	B	3	Total	Zn	0	0
			3	3		
2	C	3	Total	Zn	0	0
			3	3		
2	D	3	Total	Zn	0	0
			3	3		
2	E	3	Total	Zn	0	0
			3	3		
2	F	3	Total	Zn	0	0
			3	3		

### 3 Residue-property plots

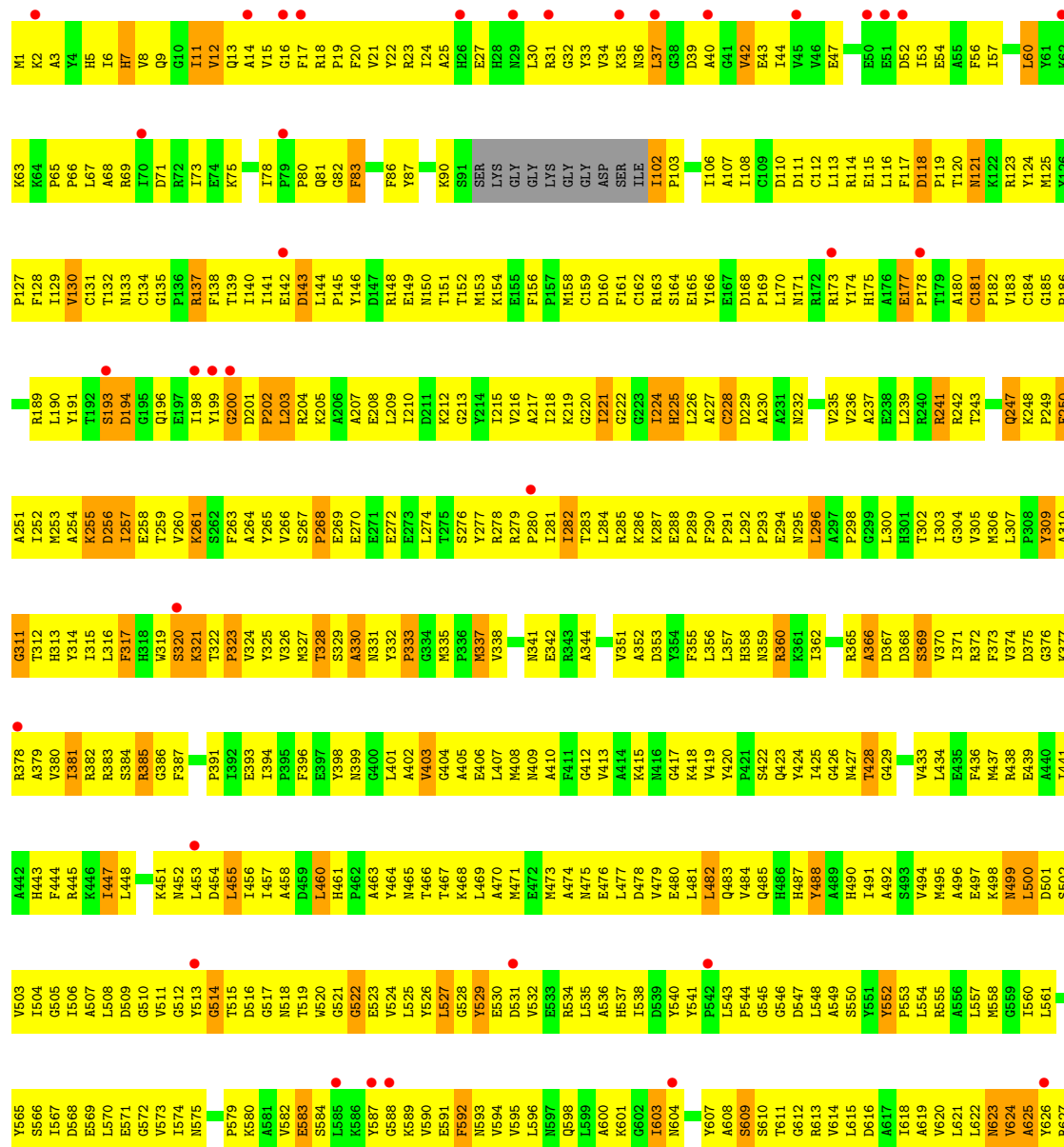
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.

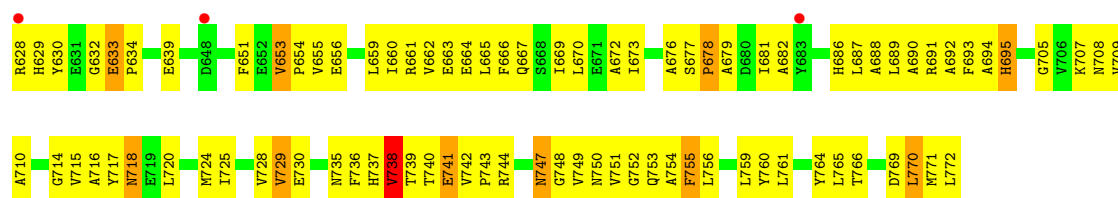
• Molecule 1: Hydrogenase maturation protein HypF



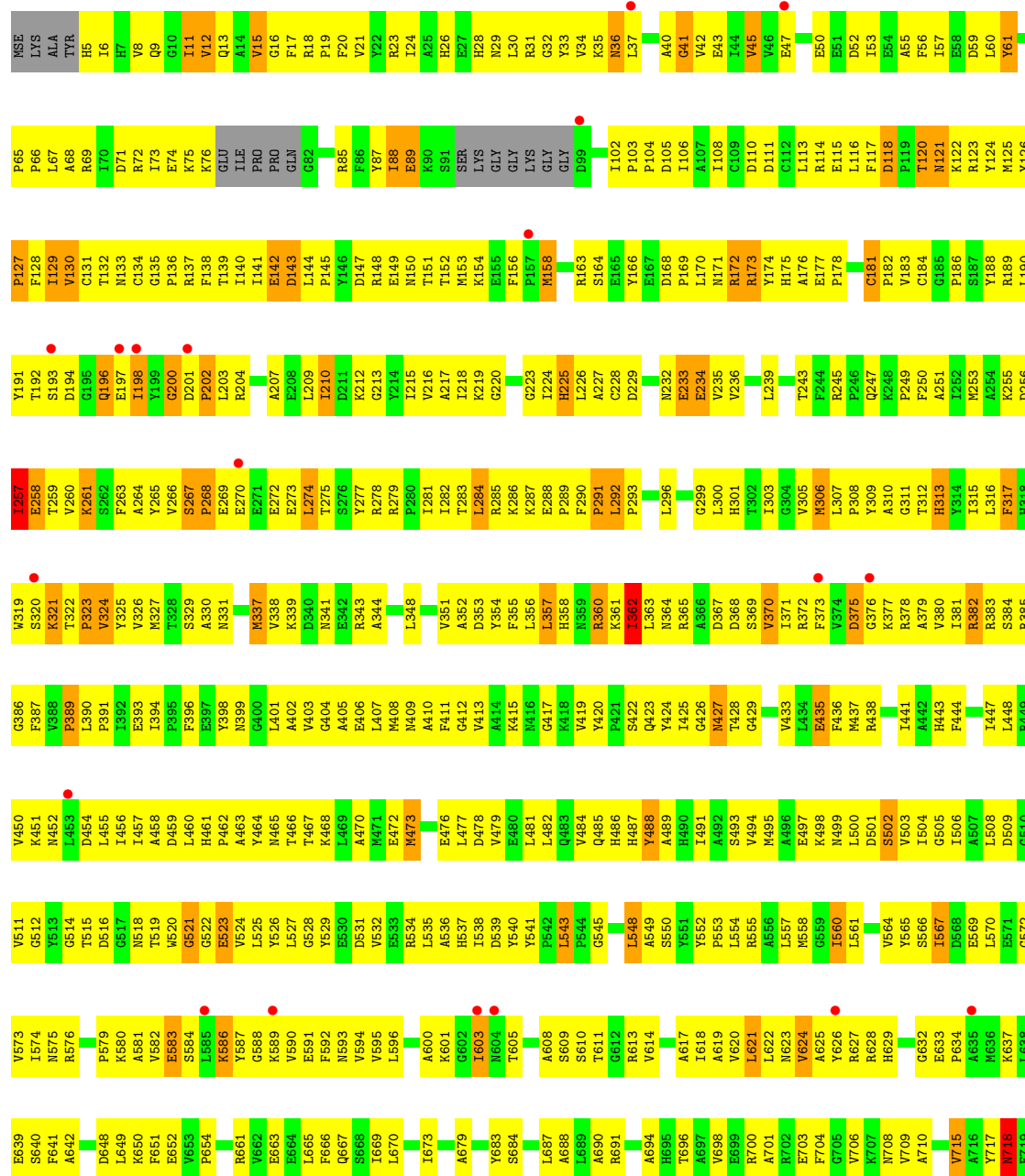


● Molecule 1: Hydrogenase maturation protein HypF



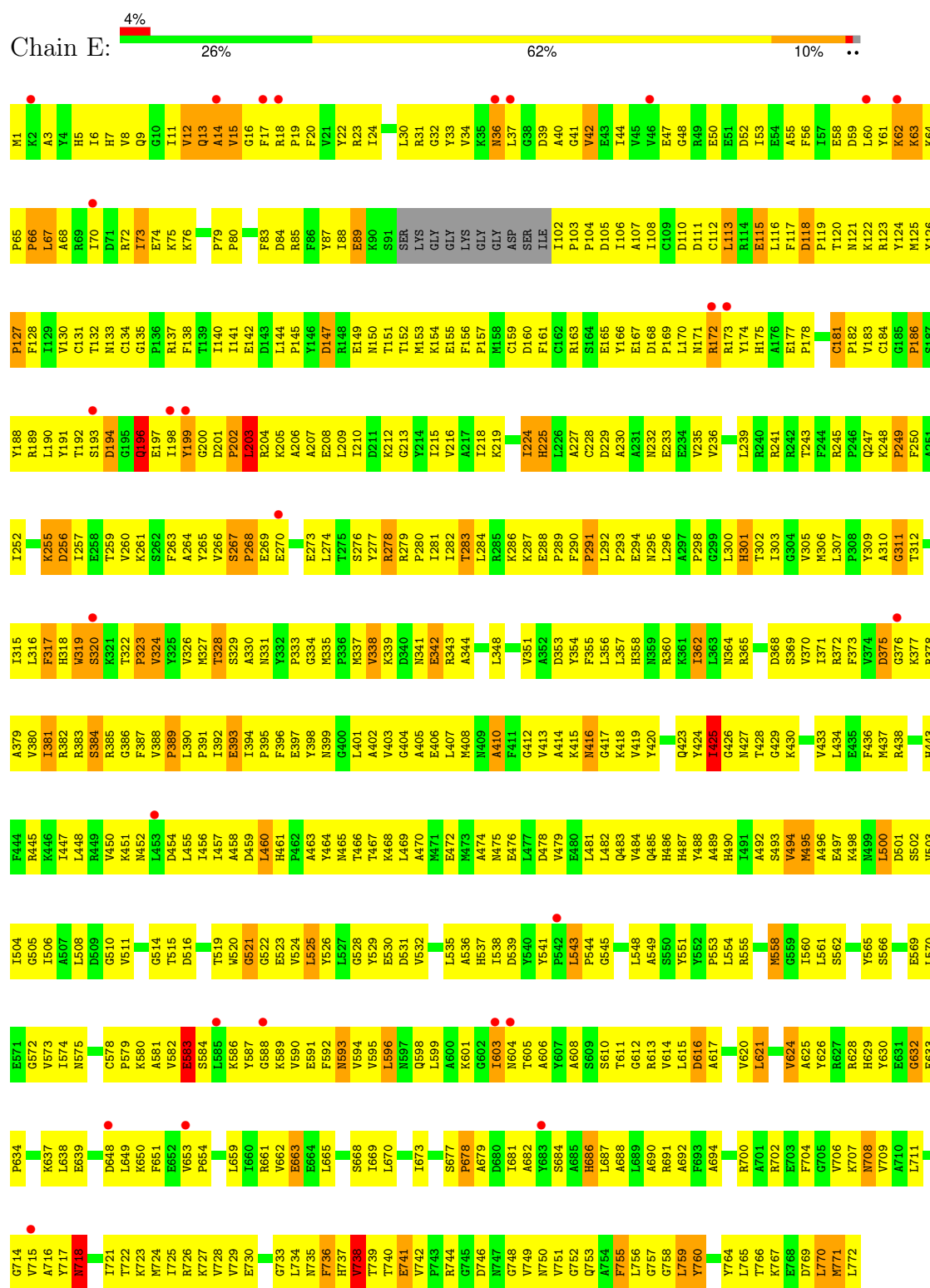


● Molecule 1: Hydrogenase maturation protein HypF









• Molecule 1: Hydrogenase maturation protein HypF



T722	G645	S584	T519	K459	V388	K321	T192	V129	K64	M1
K723	D648	L585	W520	D459	P389	T322	S193	V130	P65	K2
M724	L649	K586	G521	L460	L390	P323	D194	C131	P66	A3
I725	L650	Y587	G522	H461	P391	V324	G195	T132	L67	Y4
K726	K650	G588	E523	P462	E393	V325	Q196	M133	H5	H6
K727	F651	K589	E524	A463	E393	V326	P263	C134	I6	I7
V728	E582	V590	L525	Y464	I394	M327	A264	E197	H7	H8
V729	V653	E591	V526	M465	P395	T328	Y265	R137	V8	V9
E730	E658	F592	L527	T466	F396	S329	G200	F138	E73	Q9
A731	E659	N593	G528	T467	E397	A330	P267	T139	K75	G10
N732	L659	V594	Y529	K468	Y398	N331	P202	T140	K76	I11
G733	V595	V595	V532	L469	N399	Y332	L203	I141	V12	V13
L734	L596	L596	A470	A470	G400	P333	K204	E142	Q13	Q14
N735	H597	H597	M471	L401	G394	G334	K205	D143	A14	A15
F736	Q598	Q598	E472	A402	M335	L274	A206	P145	F83	V15
F737	L665	L665	M473	V403	P336	T275	A207	P146	D84	G16
V738	F666	A600	A474	A404	M337	S276	E208	Y146	H85	F17
V739	Q667	K601	N475	A405	V338	Y277	L209	E147	R86	R18
T740	S668	G602	E476	E406	K339	R278	I210	R148	P19	P20
E741	I669	I603	L477	L407	D340	R279	D211	E149	F87	F21
V742	P678	M604	D478	M408	N341	P280	K212	N150	I88	F22
F743	P542	P542	V479	N409	E342	I281	T151	E151	E89	V21
R744	P544	P544	E480	G409	R343	I282	T152	T152	K90	R23
G748	S683	A608	L481	L481	A344	T283	I215	M153	SER	I24
V749	S684	S609	L482	V413	L348	L284	V216	K154	LYS	A25
N750	H686	S610	L548	A414	K348	R285	A217	E155	GLY	H26
V751	L687	G612	A549	K415	D353	K286	I218	F156	GLY	E27
G752	A688	R613	E550	H486	Y354	K287	K219	P157	LYS	L30
Q753	L689	V614	Y551	H487	F355	E288	G220	GLY	R31	R31
A754	A690	L615	Y552	Y488	F355	P289	GLY	G220	L30	L31
F755	R691	D616	P553	H489	L356	F290	P291	C159	GLY	G32
L756	A692	A617	L554	H490	L357	L292	I224	R163	ASP	Y33
F757	F693	I618	R555	A492	H368	L292	H225	S164	ILE	F34
G758	A694	A619	A556	S493	N359	P293	A227	E165	K35	N36
L759	H695	V620	L557	V494	R360	E294	C228	P104	P103	L37
Y760	T696	L621	M558	I425	N295	N295	D229	D105	L37	G38
A697	A697	L622	G553	L496	P298	P298	A230	P169	D39	D39
Y764	A701	R623	L560	E497	G299	G299	A231	L170	A40	A40
L765	R702	V624	L561	T428	L300	L300	E231	N171	G41	G41
K767	E703	A625	S562	N498	H301	H301	E233	R172	V42	V42
E768	F704	Y626	X563	G429	T302	T302	E234	R173	E43	E43
D769	G705	R627	D501	V433	I303	I303	V235	Y174	I44	I44
L770	V706	R628	S502	L434	G304	G304	V236	H175	V45	V45
M771	V706	H629	Y565	L434	K377	K377	A237	A176	G48	G48
L772	N708	K635	V503	L434	R372	R372	E237	E177	R49	R49
	V709	E631	G505	H443	F373	F373	E238	E177	E50	E50
	A710	G632	G506	F444	V374	V374	L239	P178	F51	F51
	L711	E633	A507	R445	D375	D375		L179	D52	D52
	S712	P634	L508	K446	G376	G376	R242	F117	I53	I53
	S712	A635	D509	L447	K377	K377	T243	D118	E54	E54
	G713	M636	G510	L448	R378	R378	Q247	P119	F56	F56
	G714	K637	V511	L448	A379	A379	K248	N121	I57	I57
	V715	C577	G512	K451	V380	V380	P249	G185		
	A716	C578	V513	M452	I381	I381	G248	P186		
	V717	P579	G514	L453	R382	R382	F250	R123		
	N718	F641	T515	D454	R383	R383	A251	Y183		
	E719	A642	T516	L455	S384	S384	I252	M125		
	L720	V582	G517	L456	R385	R385	M253	L190		
	I721	E583	N518	I457	G386	G386	P127	K63		
		K644			F387	F387	K255			

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	265.81Å 265.81Å 693.71Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	15.00 – 4.50 15.00 – 4.50	Depositor EDS
% Data completeness (in resolution range)	90.5 (15.00-4.50) 92.0 (15.00-4.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.17	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.46 (at 4.45Å)	Xtriage
Refinement program	CNS 1.3	Depositor
R, $R_{free}$	0.279 , 0.300 0.298 , 0.293	Depositor DCC
$R_{free}$ test set	4204 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	143.9	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 121.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	36065	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	186.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.63% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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](#)

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

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	0.30	0/6151	0.79	11/8308 (0.1%)
1	B	0.30	0/6127	0.82	15/8275 (0.2%)
1	C	0.29	0/6080	0.80	11/8208 (0.1%)
1	D	0.29	0/6146	0.81	12/8301 (0.1%)
1	E	0.30	0/6127	0.81	11/8275 (0.1%)
1	F	0.30	0/6127	0.82	14/8275 (0.2%)
All	All	0.30	0/36758	0.81	74/49642 (0.1%)

There are no bond length outliers.

The worst 5 of 74 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	624	VAL	N-CA-C	11.54	121.38	110.53
1	F	624	VAL	N-CA-C	9.74	120.55	110.62
1	B	200	GLY	N-CA-C	9.05	125.88	114.66
1	B	120	THR	N-CA-C	8.43	120.55	111.36
1	C	120	THR	N-CA-C	8.36	120.39	111.28

There are no chirality outliers.

There are no planarity outliers.

### 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	A	6032	0	5969	722	1
1	B	6008	0	5954	707	0
1	C	5964	0	5909	693	0
1	D	6027	0	5970	693	0
1	E	6008	0	5954	646	0
1	F	6008	0	5955	687	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
2	C	3	0	0	0	0
2	D	3	0	0	0	0
2	E	3	0	0	0	0
2	F	3	0	0	0	0
All	All	36065	0	35711	4092	1

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 57.

The worst 5 of 4092 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:11:ILE:HG21	1:A:69:ARG:H	1.24	1.02
1:D:9:GLN:HE21	1:D:41:GLY:HA2	1.17	1.02
1:D:485:GLN:HE21	1:D:487:HIS:H	1.04	1.01
1:D:451:LYS:HA	1:D:477:LEU:HD21	1.41	0.99
1:F:394:ILE:HD11	1:F:419:VAL:HB	1.41	0.99

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:204:ARG:NH1	1:A:204:ARG:NH1[12_554]	1.95	0.25

## 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	762/772 (99%)	530 (70%)	179 (24%)	53 (7%)	1	12
1	B	758/772 (98%)	565 (74%)	145 (19%)	48 (6%)	1	14
1	C	750/772 (97%)	540 (72%)	170 (23%)	40 (5%)	1	16
1	D	761/772 (99%)	580 (76%)	148 (19%)	33 (4%)	2	18
1	E	758/772 (98%)	566 (75%)	149 (20%)	43 (6%)	1	15
1	F	758/772 (98%)	562 (74%)	171 (23%)	25 (3%)	3	21
All	All	4547/4632 (98%)	3343 (74%)	962 (21%)	242 (5%)	1	16

5 of 242 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	194	ASP
1	A	203	LEU
1	A	264	ALA
1	A	311	GLY
1	B	42	VAL

### 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	632/627 (101%)	571 (90%)	61 (10%)	6	22
1	B	630/627 (100%)	580 (92%)	50 (8%)	10	29
1	C	626/627 (100%)	564 (90%)	62 (10%)	6	22
1	D	632/627 (101%)	571 (90%)	61 (10%)	6	22
1	E	630/627 (100%)	573 (91%)	57 (9%)	8	25
1	F	630/627 (100%)	567 (90%)	63 (10%)	6	21
All	All	3780/3762 (100%)	3426 (91%)	354 (9%)	7	23

5 of 354 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	760	TYR
1	E	738	VAL
1	E	75	LYS
1	E	317	PHE
1	F	181	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 118 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	718	ASN
1	F	452	ASN
1	D	483	GLN
1	F	331	ASN
1	E	708	ASN

### 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 oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 18 ligands modelled in this entry, 18 are monoatomic - leaving 0 for Mogul analysis.

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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q < 0.9
1	A	748/772 (96%)	0.30	26 (3%)	47	36	97, 155, 228, 284	0
1	B	744/772 (96%)	0.49	39 (5%)	34	27	115, 179, 293, 361	0
1	C	739/772 (95%)	0.28	22 (2%)	52	39	118, 171, 229, 257	0
1	D	747/772 (96%)	0.34	27 (3%)	46	35	125, 189, 267, 339	0
1	E	744/772 (96%)	0.36	28 (3%)	44	34	110, 183, 285, 327	0
1	F	744/772 (96%)	0.47	35 (4%)	37	30	138, 191, 298, 348	0
All	All	4466/4632 (96%)	0.38	177 (3%)	43	33	97, 179, 272, 361	0

The worst 5 of 177 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	198	ILE	6.1
1	D	198	ILE	6.0
1	E	198	ILE	6.0
1	A	37	LEU	5.6
1	F	604	ASN	5.4

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

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

### 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 6.4 Ligands [i](#)

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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
2	ZN	B	1003	1/1	0.92	0.07	180,180,180,180	0
2	ZN	F	1002	1/1	0.95	0.08	213,213,213,213	0
2	ZN	D	1003	1/1	0.96	0.06	202,202,202,202	0
2	ZN	D	1001	1/1	0.97	0.10	157,157,157,157	0
2	ZN	A	1003	1/1	0.97	0.07	178,178,178,178	0
2	ZN	C	1003	1/1	0.97	0.05	156,156,156,156	0
2	ZN	B	1002	1/1	0.98	0.03	187,187,187,187	0
2	ZN	A	1001	1/1	0.98	0.14	144,144,144,144	0
2	ZN	C	1001	1/1	0.98	0.09	143,143,143,143	0
2	ZN	E	1002	1/1	0.98	0.04	175,175,175,175	0
2	ZN	F	1001	1/1	0.98	0.07	187,187,187,187	0
2	ZN	C	1002	1/1	0.98	0.10	147,147,147,147	0
2	ZN	F	1003	1/1	0.98	0.04	179,179,179,179	0
2	ZN	E	1003	1/1	0.99	0.04	183,183,183,183	0
2	ZN	B	1001	1/1	0.99	0.07	170,170,170,170	0
2	ZN	E	1001	1/1	0.99	0.08	152,152,152,152	0
2	ZN	D	1002	1/1	0.99	0.06	149,149,149,149	0
2	ZN	A	1002	1/1	1.00	0.05	113,113,113,113	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.