London gp47 – Hypothetical protein or thiamine-binding protein?

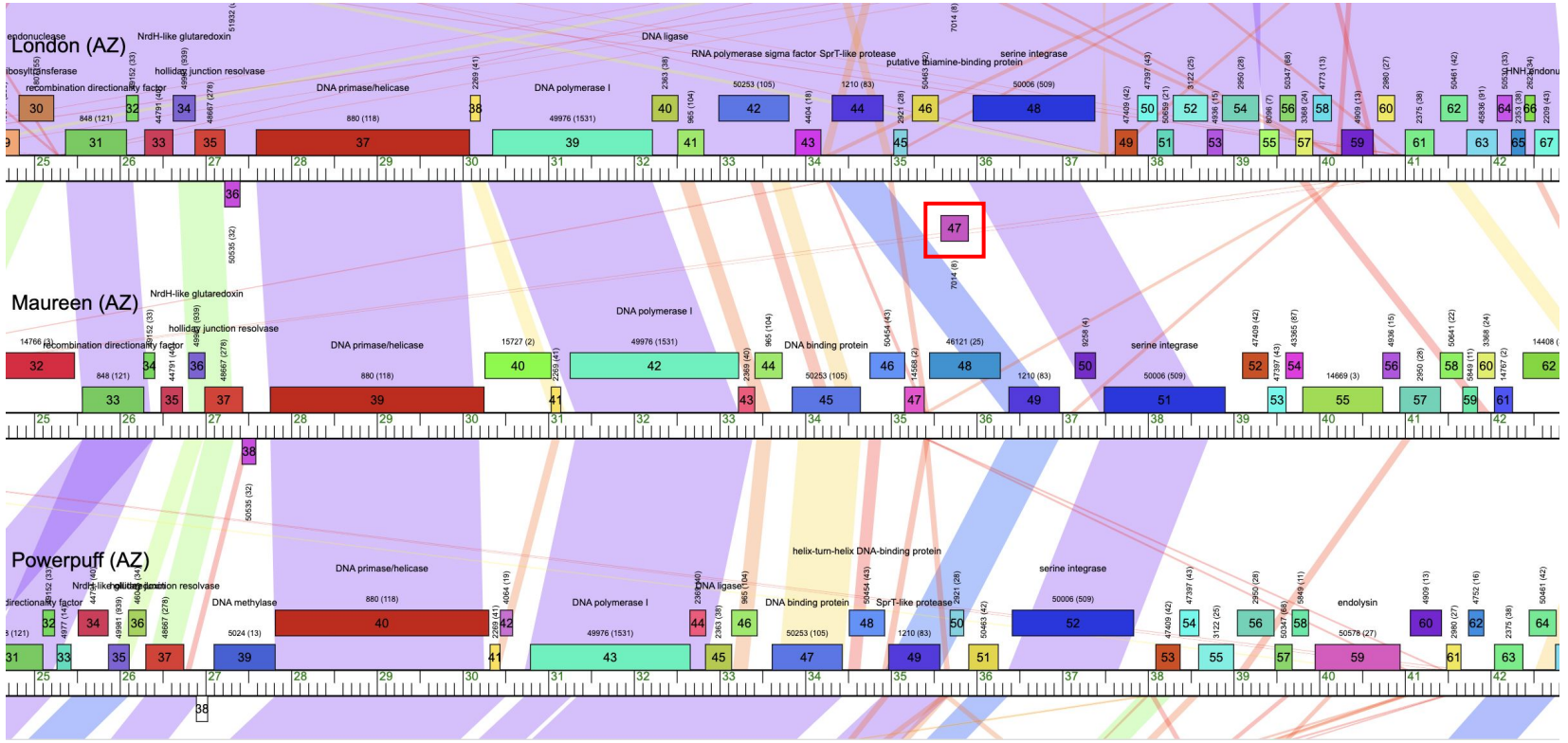
Found in 8 cluster AZ phages,
just upstream of integrase on reverse strand. Pham 7014

Gene	Phage	Accession	Notes	Cluster	Length
Asa16_47	Asa16	MZ681506	putative thiamine-binding protein	AZ	321 bp
Ascela_Draft_52	Ascela			AZ	321 bp
Elezi_47	Elezi	MT639653	putative thiamine-binding protein	AZ	321 bp
Eraser_47	Eraser	MZ747516		AZ	321 bp
Iter_50	Iter	ON208833		AZ	321 bp
London_47	London	MT889366		AZ	321 bp
Niobe_47	Niobe	MZ820087	putative thiamine-binding protein	AZ	321 bp
Nitro_Draft_51	Nitro			AZ	321 bp

Summary:
Good alignment with thiamine-binding proteins.
However,

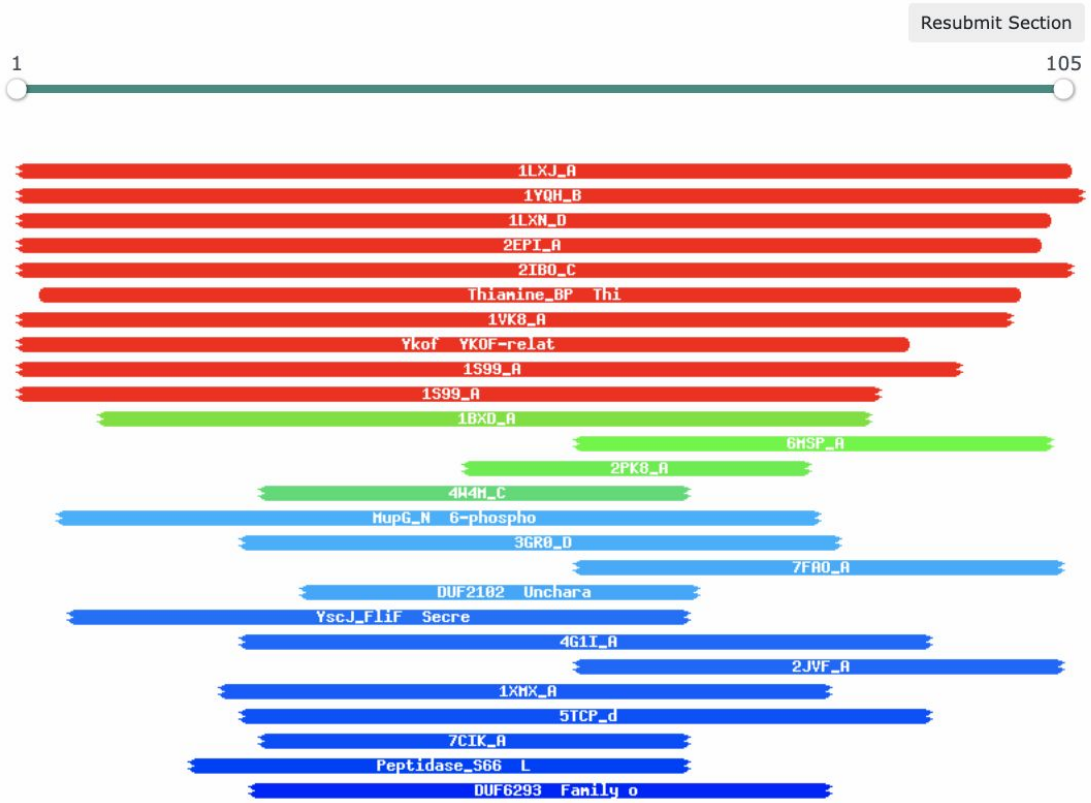
- These are usually part of a multi-gene system. What is the function of this gene by itself?
- What is its function in phage multiplication?
- This domain is in the ACT domain family, which is found in many different proteins with various functions and ligands.

Recommendation: Keep as hypothetical protein for now.



HHPred results
https://toolkit.tuebingen.mpg.de/jobs/5897366_54

Hits to several hypothetical proteins and Bacillus YkoF thiamine-binding protein



1	1LXJ_A	HYPOTHETICAL 11.5KDA PROTEIN IN HTB2-NTH2 INTERGENIC REGION; Hypothetical protein, HTB2-NTH2 intergenic region, Structur	99.92	1.4e-23	138.29	14.1	98	104
2	1YQH_B	IG hypothetical 16092; Bacillus cereus, structural genomics, PSI, Protein Structure Initiative, Midwest Center for Struc	99.92	2.2e-23	138.07	14.4	98	109
3	1LXN_D	HYPOTHETICAL PROTEIN MTH1187; HYPOTHETICAL STRUCTURE, STRUCTURAL GENOMICS, PSI, Protein Structure Initiative, Northeast	99.91	6.7e-23	133.11	13.5	96	99
4	2EPI_A	UPF0045 protein MJ1052; NPPSFA, National Project on Protein Structural and Functional Analyses, RIKEN Structural Genomic	99.89	1e-21	128.08	13	93	100
5	2IBO_C	Hypothetical protein SP2199; alpha-beta protein, Structural Genomics, PSI-2, Protein Structure Initiative, Northeast Str	99.89	1.7e-21	128.43	14	95	104
6	PF01910.20	; Thiamine_BP ; Thiamine-binding protein	99.89	1.4e-21	125.81	12.6	91	92
7	1VK8_A	hypothetical protein TM0486; PROTEIN WITH POSSIBLE ROLE IN CELL WALL BIOGENESIS, STRUCTURAL GENOMICS, JOINT CENTER FOR S	99.87	1.2e-20	125.19	12.7	89	106
8	PF07615.14	; Ykof ; YKOF-related Family	99.83	3.4e-19	112.29	11.9	80	81
9	1S99_A	ykoF; YkoF, thiamin-binding protein, ACT-domain family, LIGAND BINDING PROTEIN; HET: MSE; 1.65A {Bacillus subtilis} SCOP	99.73	8.9e-17	117.08	11	85	200
10	1S99_A	ykoF; YkoF, thiamin-binding protein, ACT-domain family, LIGAND BINDING PROTEIN; HET: MSE; 1.65A {Bacillus subtilis} SCOP	99.68	1.1e-15	111.39	10.7	76	200
11	1BXD_A	PROTEIN (OSMOLARITY SENSOR PROTEIN (ENVZ)); HISTIDINE KINASE, OSMOSENSOR, HIS-ASP PHOSPHORELAY SYSTEM, SIGNAL TRANSDUCTI	75.15	42	22.89	8.7	70	161

Vis Hits Aln | Select All Forward Forward Query A3M Model using selection Download HHR Color Seqs Wrap Seqs

Template alignment | Template 3D Structure | PDBe

9. **1S99_A ykoF; YkoF, thiamin-binding protein, ACT-domain family, LIGAND BINDING PROTEIN; HET: MSE; 1.65A {Bacillus subtilis} SCOP: d.58.48.2**

Probability: 99.73%, E-value: 8.9e-17, Score: 117.08, Aligned cols: 85, Identities: 12%, Similarity: 0.176, Template Neff: 8

```
Q ss_pred            CeEEEEEEecCCCCCCCCCCCCchHHHHHHHHHHHHcCCCeEeCCCCcEEEeCHHHHHHHHHHHHHHHHCCEEEEEE
Q Q_5897366_54   1   MIVAFSVAPSGASASGEAPSDASVHEAVAAAVRVVRESGLPNRTSSMFTELEGDWDAVMDVVKRATEAVAPYGSRVSLVL   80 (106)
Q Consensus      1   ~~~~~sl~P~~~~~~~~~~~~~~~~I~~~i~~l~~sgl~~~~~~~~T~ieG~~~~v~~~i~~~~~~~~~~rv~~~i   80 (106)
                     +.++|+++|++..         ++.++|+++++.|+++|+|++++|+|+++|+|++|++|++.+++.++|+.+++
T Consensus    115 i~a~~sl~Pl~~~~~~~~~~~~~I~~~i~~l~~gl~~~~~~~~T~ieG~~~~v~~~l~~~~~~~~g~rv~~~i   185 (200)
T 1S99_A       115 APCQFALYPMNEP---------DYMGLIMEAVDIAKAQGTFVQGVHYASELDGDAHDVFSTLEAVFRMAEQQTNHITMTV   185 (200)
T ss_dssp            CCEEEEEEETTCT---------THHHHHHHHHHHHHTTCEEEEETTEEEECSBHHHHHHHHHHHHHHHHHCSSEEEEE
T ss_pred            EEEEEEEEECCCc---------chHHHHHHHHHHHHcCCeeeeeCCeEEEcCHHHHHHHHHHHHHHHHhCCeEEEEE

Q ss_pred            EEEeCCCCCCcHHH
Q Q_5897366_54   81  KADIRPGYSGELEG   94 (106)
Q Consensus      81  ~i~~~~~~~~~~~~   94 (106)
                     +++.+.......+.
T Consensus    186 ki~~~~~~~~~~~~  199 (200)
T 1S99_A       186 NLSANSPSRKNRKQ  199 (200)
T ss_dssp            EEEESCTTSCCC--
T ss_pred            EEEecCCChhhccc
```

Template alignment | Template 3D Structure | PDBe

10. **1S99_A ykoF; YkoF, thiamin-binding protein, ACT-domain family, LIGAND BINDING PROTEIN; HET: MSE; 1.65A {Bacillus subtilis} SCOP: d.58.48.2**

Probability: 99.68%, E-value: 1.1e-15, Score: 111.39, Aligned cols: 76, Identities: 20%, Similarity: 0.154, Template Neff: 8

```
Q ss_pred            CeEEEEEEecCCCCCCCCCCCCchHHHHHHHHHHHHcCCCeEeCCCCcEEEeCHHHHHHHHHHHHHHHHCCEEEEEE
Q Q_5897366_54   1   MIVAFSVAPSGASASGEAPSDASVHEAVAAAVRVVRESGLPNRTSSMFTELEGDWDAVMDVVKRATEAVAPYGSRVSLVL   80 (106)
Q Consensus      1   ~~~~~sl~P~~~~~~~~~~~~~~~~I~~~i~~l~~sgl~~~~~~~~T~ieG~~~~v~~~i~~~~~~~~~~rv~~~i   80 (106)
                     +.++|+++|++ .         ++.++|..+++.|+++|+|++++|+|++++|++++|+++++.+++.+.|+.+++
T Consensus     11 i~a~~sl~Pl~~~~~~~~~~~~~I~~~i~~l~~gl~~~~~~~~sT~ieG~~~~v~~~v~~~~~~~~g~rv~~~i   80 (200)
T 1S99_A        11 AGFRFSLYPMT-D---------DFISVIKSALAATDTSKVWTKTDHISTVLRGSIDHVFDAAKAIYLHAANSEQHIVMNG   80 (200)
T ss_dssp            EEEEEEEEECS-T---------THHHHHHHHHHSCCTTSEEEECSSCEEEEECHHHHHHHHHHHHHHHTTCSCEEEEE
T ss_pred            EEEEEEEEECC-C---------chHHHHHHHHHhhhcCCEEEEeCCeEEEEECHHHHHHHHHHHHHHHHhCCCeEEEE

Q ss_pred            EEEeCC
Q Q_5897366_54   81  KADIRP   86 (106)
Q Consensus      81  ~i~~~~   86 (106)
                     +++.+.
T Consensus     81  ~i~~~~   86 (200)
T 1S99_A        81  TFSIGC   86 (200)
T ss_dssp            EEEECC
T ss_pred            EEEecC
```

YkoF has two of these thiamine-binding ACT domains; London_47 has good, full-length matches to both

Regions aligning with London gp47

Thiamine-binding domains

1	1LXJ_A	HYPOTHETICAL 11.5KDA PROTEIN IN HTB2-NTH2 INTERGENIC REGION; Hypothetical protein, HTB2-NTH2 intergenic region, Structur	99.92	1.4e-23	138.29	14.1	98	104
2	1YQH_B	IG hypothetical 16092; Bacillus cereus, structural genomics, PSI, Protein Structure Initiative, Midwest Center for Struc	99.92	2.2e-23	138.07	14.4	98	109
3	1LXN_D	HYPOTHETICAL PROTEIN MTH1187; HYPOTHETICAL STRUCTURE, STRUCTURAL GENOMICS, PSI, Protein Structure Initiative, Northeast	99.91	6.7e-23	133.11	13.5	96	99
4	2EPI_A	UPF0045 protein MJ1052; NPPSFA, National Project on Protein Structural and Functional Analyses, RIKEN Structural Genomic	99.89	1e-21	128.08	13	93	100
5	2IBO_C	Hypothetical protein SP2199; alpha-beta protein, Structural Genomics, PSI-2, Protein Structure Initiative, Northeast Str	99.89	1.7e-21	128.43	14	95	104
6	PF01910.20	; Thiamine_BP ; Thiamine-binding protein	99.89	1.4e-21	125.81	12.6	91	92
7	1VK8_A	hypothetical protein TM0486; PROTEIN WITH POSSIBLE ROLE IN CELL WALL BIOGENESIS, STRUCTURAL GENOMICS, JOINT CENTER FOR S	99.87	1.2e-20	125.19	12.7	89	106
8	PF07615.14	; Ykof ; YKOF-related Family	99.83	3.4e-19	112.29	11.9	80	81
9	1S99_A	ykoF; YkoF, thiamin-binding protein, ACT-domain family, LIGAND BINDING PROTEIN; HET: MSE; 1.65A {Bacillus subtilis} SCOP	99.73	8.9e-17	117.08	11	85	200
10	1S99_A	ykoF; YkoF, thiamin-binding protein, ACT-domain family, LIGAND BINDING PROTEIN; HET: MSE; 1.65A {Bacillus subtilis} SCOP	99.68	1.1e-15	111.39	10.7	76	200
11	1BXD_A	PROTEIN (OSMOLARITY SENSOR PROTEIN (ENVZ)); HISTIDINE KINASE, OSMOSENSOR, HIS-ASP PHOSPHORELAY SYSTEM, SIGNAL TRANSDUCTI	75.15	42	22.89	8.7	70	161

Template alignment | Template 3D Structure | PDBe

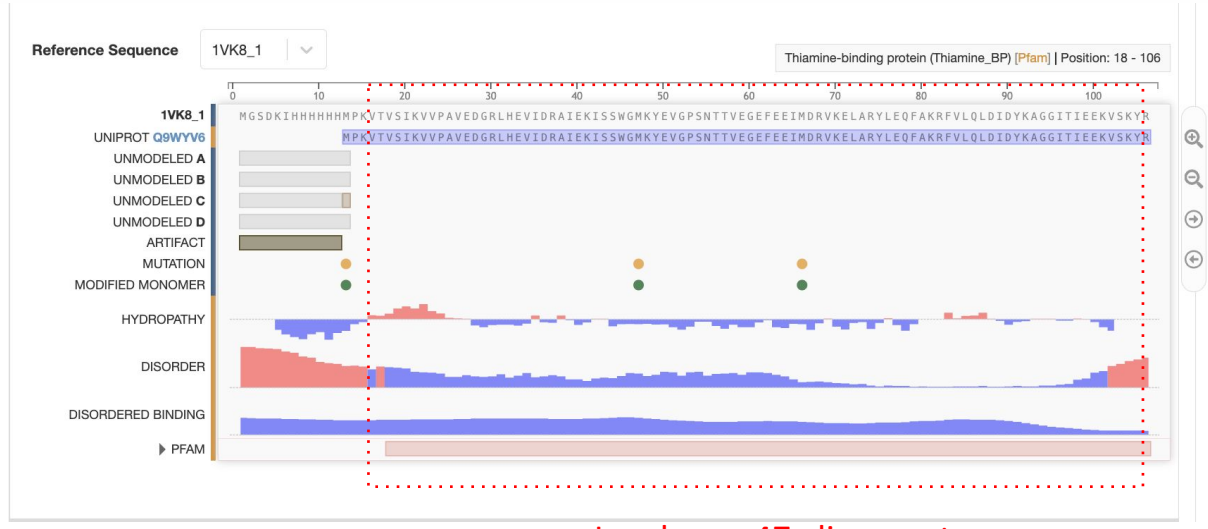
7. 1VK8_A hypothetical protein TM0486; PROTEIN WITH POSSIBLE ROLE IN CELL WALL BIOGENESIS, STRUCTURAL GENOMICS, JOINT CENTER FOR STRUCTURAL GENOMICS, JCSG; HET: UNL, MSE; 1.8A {Thermotoga maritima} SCOP: d.58.48.1

Probability: 99.87%, E-value: 1.2e-20, Score: 125.19, Aligned cols: 89, Identities: 25%, Similarity: 0.366, Template Neff: 7.8

```
Q ss_pred            CeEEEEEEecCCCCCCCCCCCCchHHHHHHHHHHHHHcCCCeEeCCCCcEEEeCHHHHHHHHHHHHHHHHHC-CEEEEE
Q Q_5897366_54   1   MIVAFSVAPSGASASGEAPSDASVHEAVAAAVRVVVRESGLPNRTSSMFTELEGDWDAVMDVVKRATEAVAPYG-SRVSLV   79 (106)
Q Consensus      1   ~~~~~sl~P~~~~~~~~~~~~~~~I~~~i~~l~~sgl~~~~~~~~T~ieG~~~~v~~~i~~~~~~~~~~~~rv~~~   79 (106)
                     ++|+|+++|++|++.+        .++.++|+.+++.|+++||+|+|+++||+|+|+++|+|+++|++|++|++| +.+.+.| .|+.++
T Consensus      16  ~~a~~sl~Pl~~~~~~~~~~~~~~I~~~i~~l~~~gl~~~~~~~~T~ieG~~d~v~~~l~~~~~~~~~~g~~rv~~~   86 (106)
T 1VK8_A         16  VTVSIKVVPAVED--------GRLHEVIDRAIEKISSWGMKYEVGPSNTTVEGEFEEIMDRVKELARYLEQFAKRFVLQ   86 (106)
T ss_dssp            EEEEEEEEESSCG--------GGHHHHHHHHHHHHHTTCSCEEECSSCEEEEECHHHHHHHHHHH-HHHHTTTCSEEEEE
T ss_pred            EEEEEEEEecCCC--------CChHHHHHHHHHHHHCCCeEEEcCCceEEEECHHHHHHHHHHH-HHHHHcCCCEEEEE

Q ss_pred            EEEEeCCCCCCcHHHHHHHH
Q Q_5897366_54   80  LKADIRPGYSGELEGKVERL   99 (106)
Q Consensus      80  i~i~~~~~~~~~~~kv~~v   99 (106)
                     ++++.+.|+ +++++|++++
T Consensus      87  i~i~~~~d~~~s~~~kv~~~  105 (106)
T 1VK8_A         87  LDIDYKAGG-ITIEEKVSKY  105 (106)
T ss_dssp            EEEEETTC-CCHHHHHGGG
T ss_pred            EEEEecCCC-CCHHHHHhhc
```

(See next slide for identification of thiamine binding residues)



London gp47 alignment

London_47 has most but not all of the residues identified as important for thiamine binding

https://www.sciencedirect.com/science/article/abs/pii/S0022283604010174?via%3Dihub

HHPred alignment

Template alignment | Template 3D Structure | PDBe

10. **1S99_A ykoF; YkoF, thiamin-binding protein, ACT-domain family, LIGAND BINDING PROTEIN; HET: MSE; 1.65A {Bacillus subtilis} SCOP: d.58.48.2**

Probability: 99.68%, E-value: 1.1e-15, Score: 111.39, Aligned cols: 76, Identities: 20%, Similarity: 0.154, Template Neff: 8

```
Q ss_pred            CeEEEEEEecCCCCCCCCCCCCchHHHHHHHHHHHHHcCCCeEeCCCCcEEEeCHHHHHHHHHHHHHHHHCCEEEEEE
Q Q_5897366_54   1   MIVAPSVAPSGASASGEAPSDASVHEAVAAAVRVVRESGLPNRTSSMPTELEGDWDAVMDVVKRATEAVAPYGSRVSLVL   80 (106)
Q Consensus      1   ~~~~~sl~P~~~~~~~~~~~~~~~~~I~~~i~~l~~sgl~~~~~~~T~ieG~~~~v~~~i~~~~~~~~~~~rv~~~i   80 (106)
                     +.++|+++|++ .         ++.++|..+++.||++|+|++++|+|+++|++++||++++++.+++.+.||+.+++
T Consensus      11  i~a~~sl~Pl~~~~~~~~~~~~~~~~I~~~i~~l~~gl~~~~~~sT~ieG~~~~v~~~v~~~~~~~~~g~rv~~~i   80 (200)
T 1S99_A         11  AGFRPSLYPMT-D---------DFISVVKSALAAIDTSKVWTKTDHISTVLRGSIDHVFDAAKAIYLHAANSEQHIVMNG   80 (200)
T ss_dssp            EEEEEEEEECS-T---------THHHHHHHHHHSCCTTSEEEECSSCEEEEECHHHHHHHHHHHHHHHHTTCSCEEEEE
T ss_pred            EEEEEEEEECC-C---------chHHHHHHHHHhhhcCCEEEEecCeEEEECHHHHHHHHHHHHHHHHhcCCeEEEEE

Q ss_pred            EEEeCC
Q Q_5897366_54   81  KADIRF   86 (106)
Q Consensus      81  ~i~~~~   86 (106)
                     +++.+.
T Consensus      81  ~i~~~~   86 (200)
T 1S99_A         81  TFSIGG   86 (200)
T ss_dssp            EEEECC
T ss_pred            EEEcCC
```

Alignment with N-terminal domain
It may be missing some of the residues important for dimer formation (labeled with + in Fig. 6)

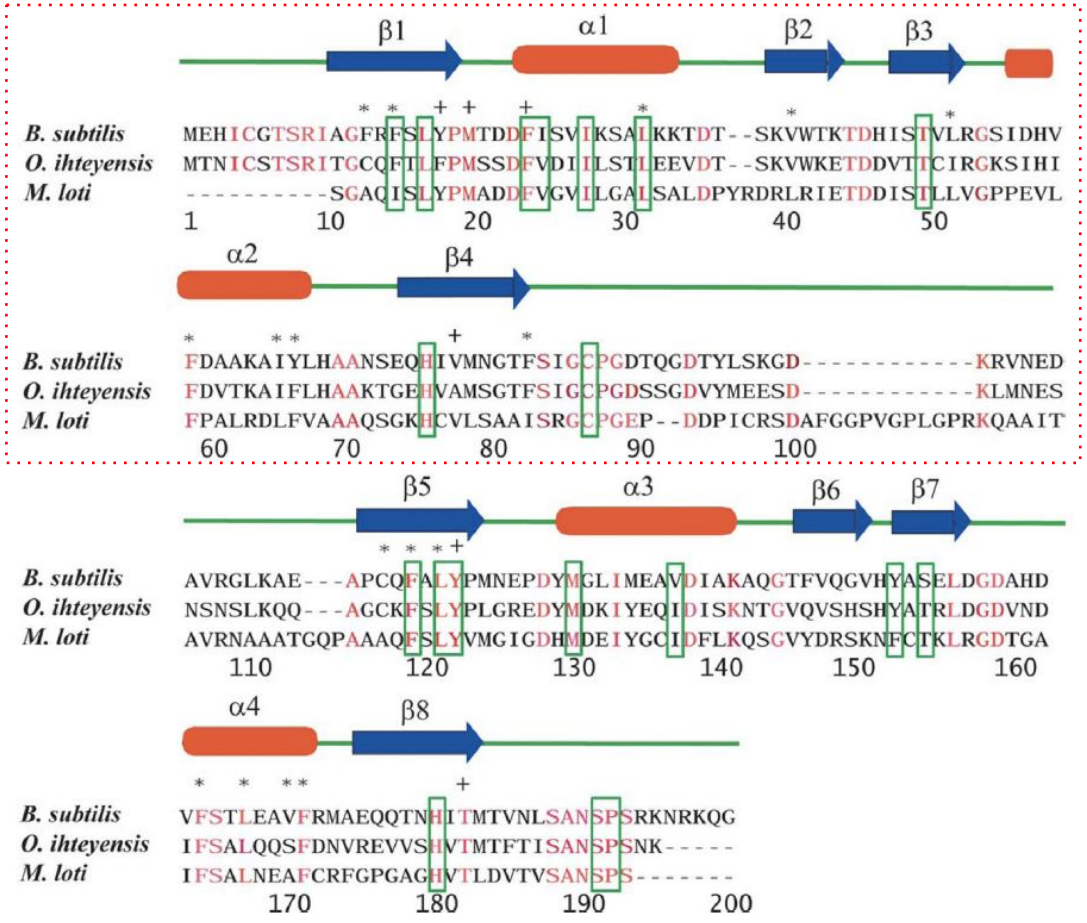
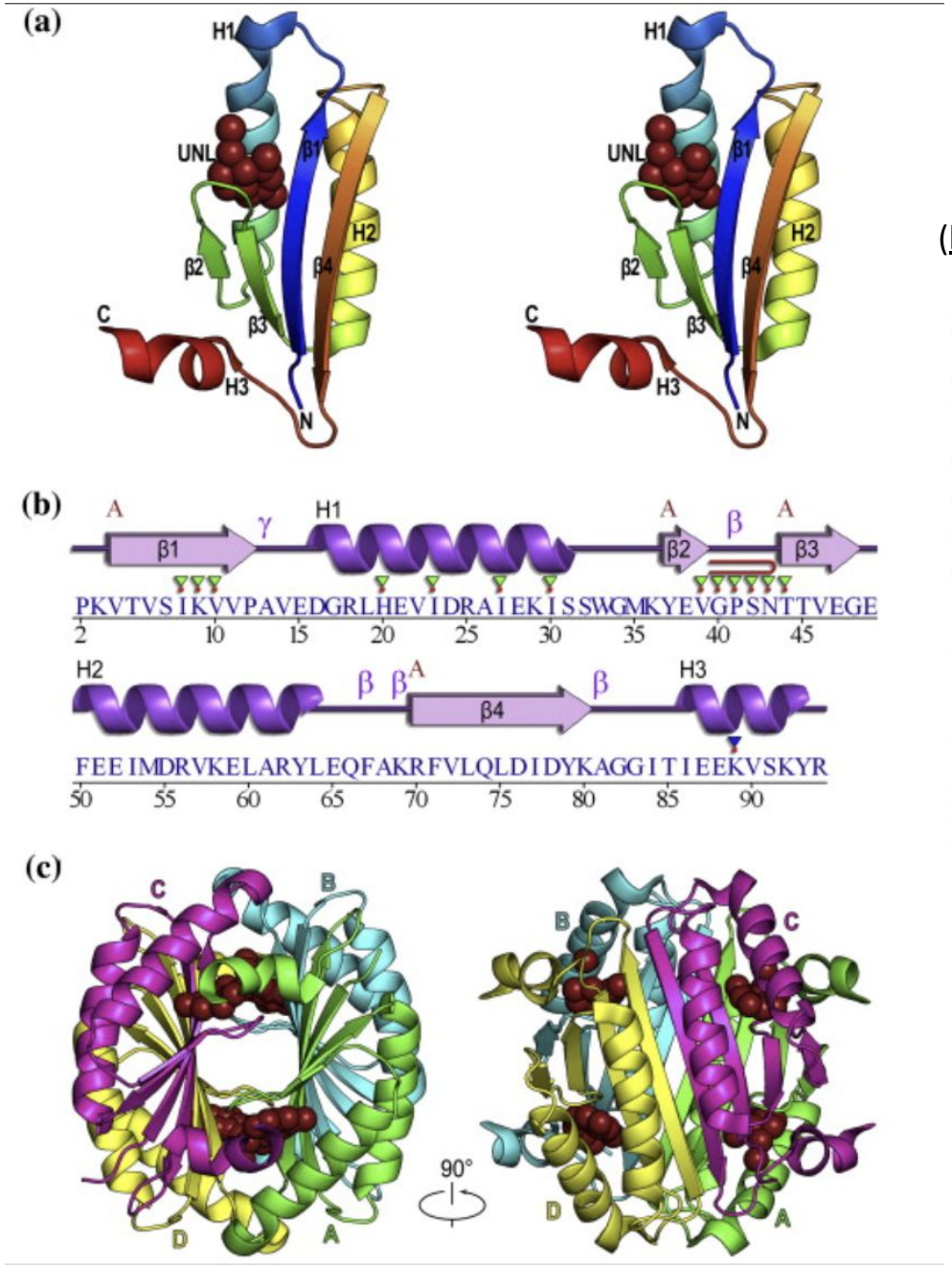


Figure 6.
Sequence alignment of the three YkoF proteins from bacterial species. Residues in the hydrophobic core are labeled with an asterisk, those involved in dimer formation are labeled with crosses and the residues important for thiamin recognition are indicated by green boxes.



(https://www.sciencedirect.com/science/article/pii/S0022283610004936?via%3Dihub

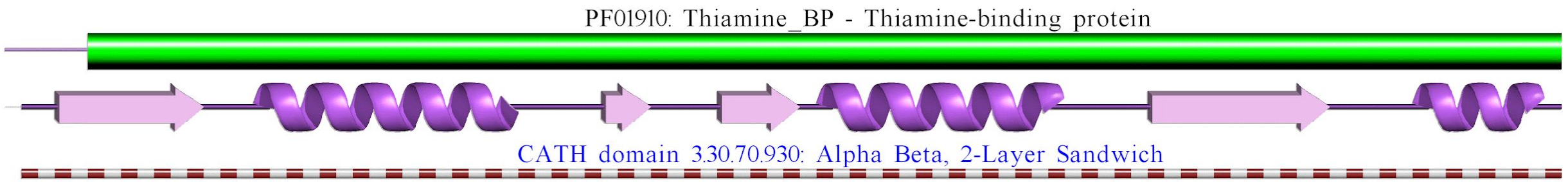
Fig. 1. Crystal structure of TM0486. (a) Stereo ribbon diagram of one TM0486 protomer color-coded from the N-terminus (blue) to the C-terminus (red). Helices H1–H3 and β-strands β1–β4 are indicated. The location of the unknown ligand (UNL) is indicated with red spheres. (b) A diagram showing the secondary-structure elements of TM0486 superimposed on its primary sequence. The labelling of secondary-structure elements is based on PDBsum (http://www.ebi.ac.uk/pdbsum). For TM0486, the α-helices H1–H3, the β-strands in β-sheet A, the β-turns (β), the γ-turns (γ) and the β-hairpins are indicated. Residues interacting with the bound unknown ligand are shown with red dots, green triangles indicate residues interacting with the UNL in the same protomer and the blue triangle indicates the residue interacting with the UNL in the adjacent protomer. (c) A ribbon representation of the TM0486 tetramer showing 222 symmetry. Protomers A, B, C and D are depicted in green, cyan, magenta and yellow, respectively and the unknown ligands are shown as red spheres. The β-sheets from protomers A/B and C/D form a continuous anti-parallel β-sheet with strand order β1, β2, β3, β4, β4′, β3′, β2′, β1′.

Protein ID: Elezi gp47

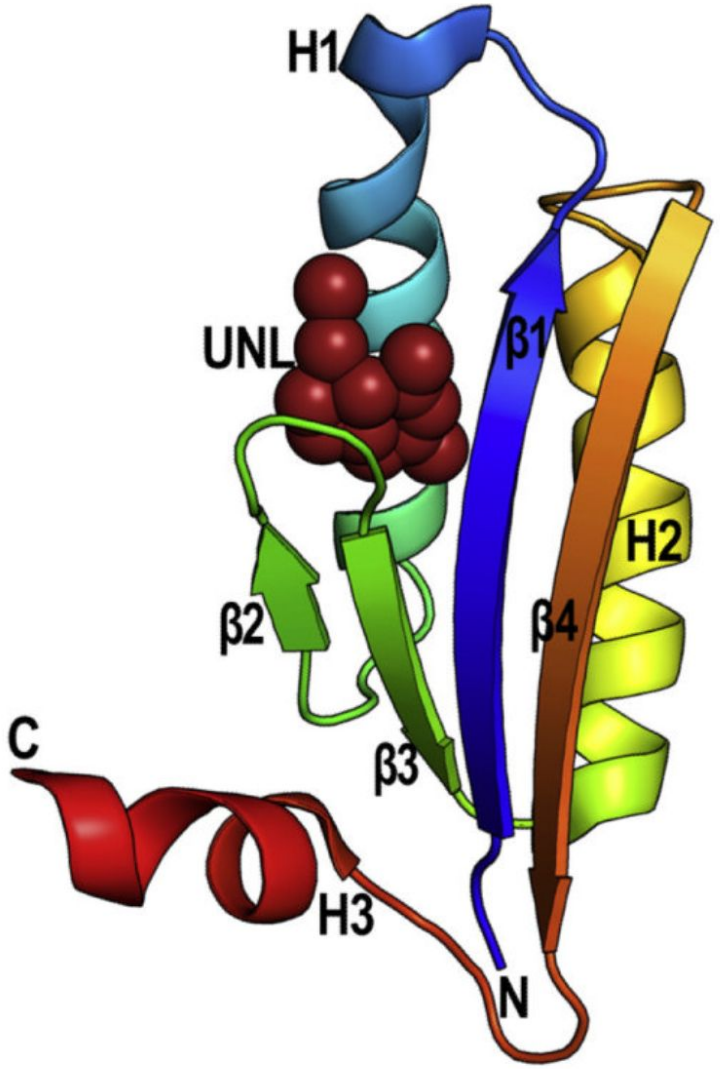
```
AA_QUERY          1    MIVAFSVAPSGAPASGEAPSDASVHEAVAAAVRVVRESGLPNRTSSMFTELEGDWDAVMDVVKRATEAVAPYGSRVSLVLKADIR   85
SS_PSIPRED             EEEEEEEEE              HHHHHHHHHHHHHH     EE     EEE    HHHHHHHHHHHHHHHH     EEEEEEEEEE
SS_SPIDER3           EEEEEEEE                 HHHHHHHHHHHHHH     E        EEEE HHHHHHHHHHHHHHHHH    EEEEEEEEE
SS_PSSPRED4            EEEEEE               HHHHHHHHHHHHHHHH       HHHHHHH    HHHHHHHHHHHHHH     EEEEEEEE
SS_DEEPCNF             EEEEEEE              HHHHHHHHHHHHHH           H        HHHHHHHHHHHHHH     EEEEEEEE
SS_NETSURFP2           EEEEEEEE             HHHHHHHHHHHHHH     EEE    EEEEE HHHHHHHHHHHHHHHHH    EEEEEEEEE
DO_NETSURFPD2
DO_DISOPRED          DDDDDDDDDDDDDDDDD
DO_SPOTD                        DDDDD
DO_IUPRED                       DDDDDDDDD                    D  DD

AA_QUERY          86   PGYSGELEGKVERLERAIEKP   106
SS_PSIPRED              HHHHHHHHHHHH
SS_SPIDER3              HHHHHHHHHHHH
SS_PSSPRED4              HHHHHHHHHHHH
SS_DEEPCNF              HHHHHHHHHHH
SS_NETSURFP2            HHHHHHHHHHH
DO_NETSURFPD2                      D
DO_DISOPRED                        D
DO_SPOTD
DO_IUPRED                  DD   DDD  DD
```

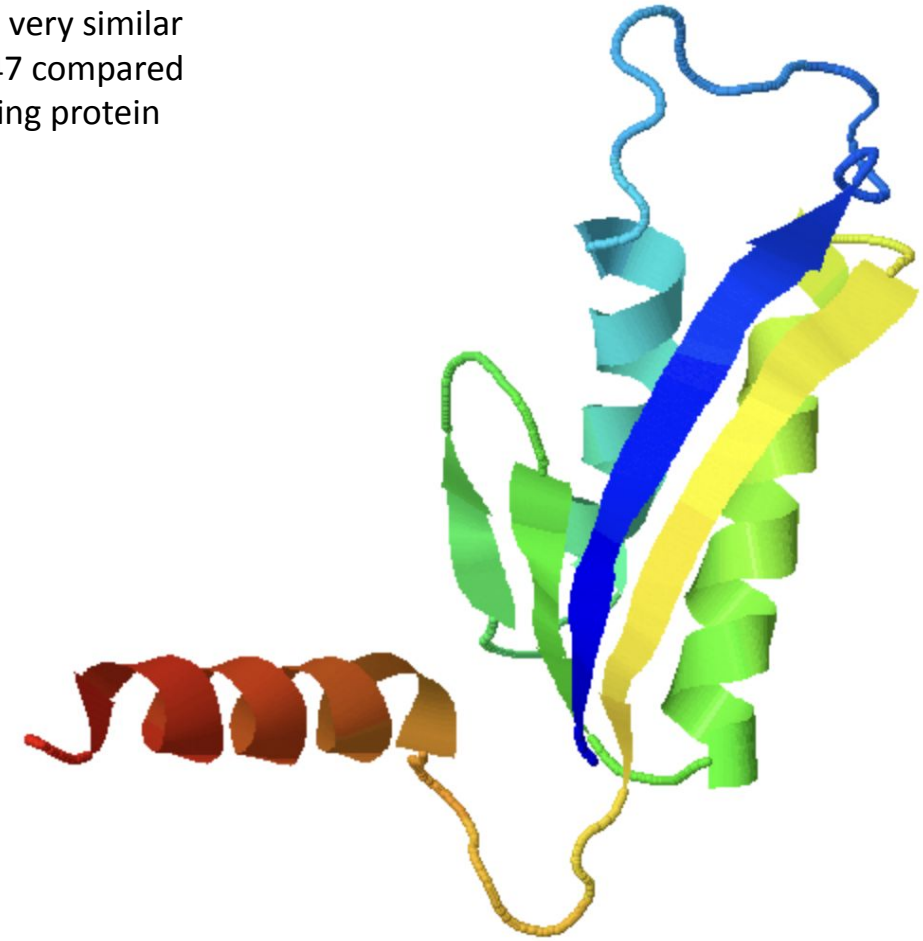
** Same secondary structure predictions…



(a)



AlphaFold predicts very similar structure for Elezi47 compared to a thiamine binding protein



TM0486_1vk8 (TBP)

Elezi_gp47 AlphaFold2 prediction