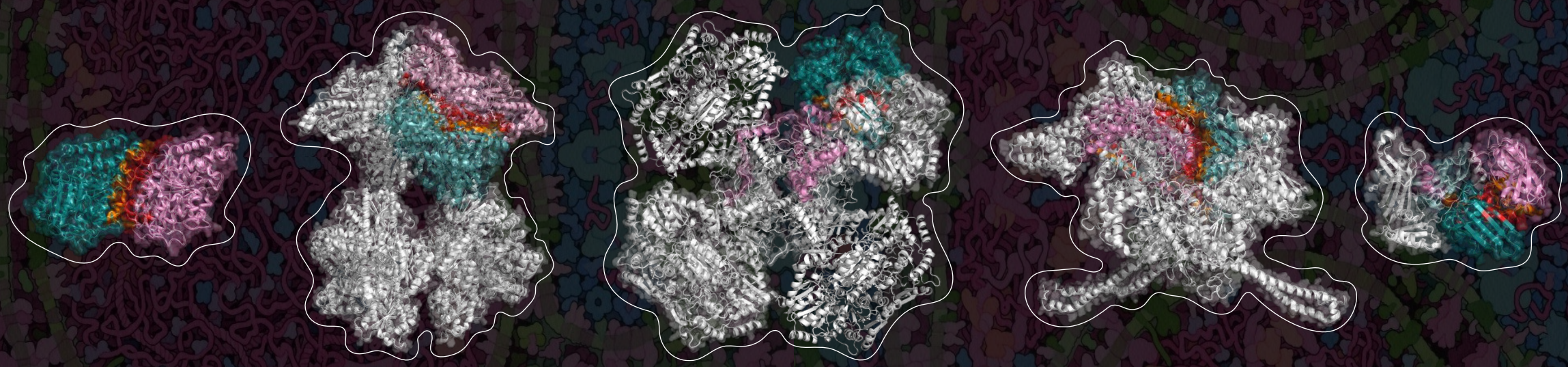
Co-translational protein complex assembly:

integrating PDB structural data and quantitative mass spectrometry data to identify candidate complexes



Faculty of Life Science, The University of Manchester, United Kingdom

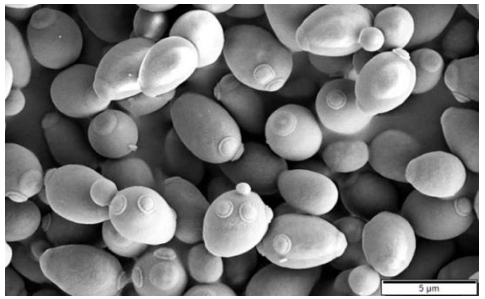
Linqing Hu, Simon Hubbard's Lab

Contact: liniqnghu120120@outlook.com

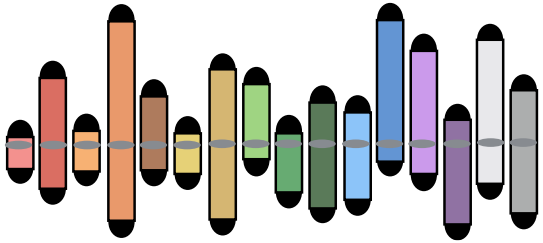
Background 1: Complex Co-translational assembly is worth studying

About *Saccharomyces cerevisiae* (Yeast) …

- The whole genome: **12.1 Mb** (16 Chromosomes)
- Gene number: **6470** --- **6014** are protein-coding



5 µm



n = 16

- In one expression, produce **~ 3500** individual proteins
  --- takes **~75%** of the cell's energy budget
- Form **>600** protein complexes

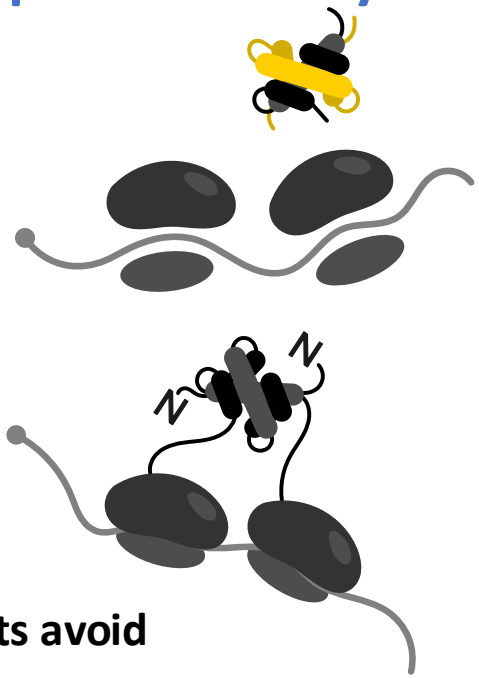
Complex assembly

--- highly regulated process

--- save material, energy and time

Two types of protein complex assembly :

- **Post-translational assembly (Post-TA)**
- **Co-translational assembly (Co-TA)**



Post-TA does not explain:
how unassembled protein subunits avoid
- non-specific interactions
- aggregation
- quality control sequestration to proteases and chaperones
- navigate crowded and occluded cellular environments

Complex Co-TA is worth studying!

Background 2: Complex co-translational assembly has 6 modes

3 key factors

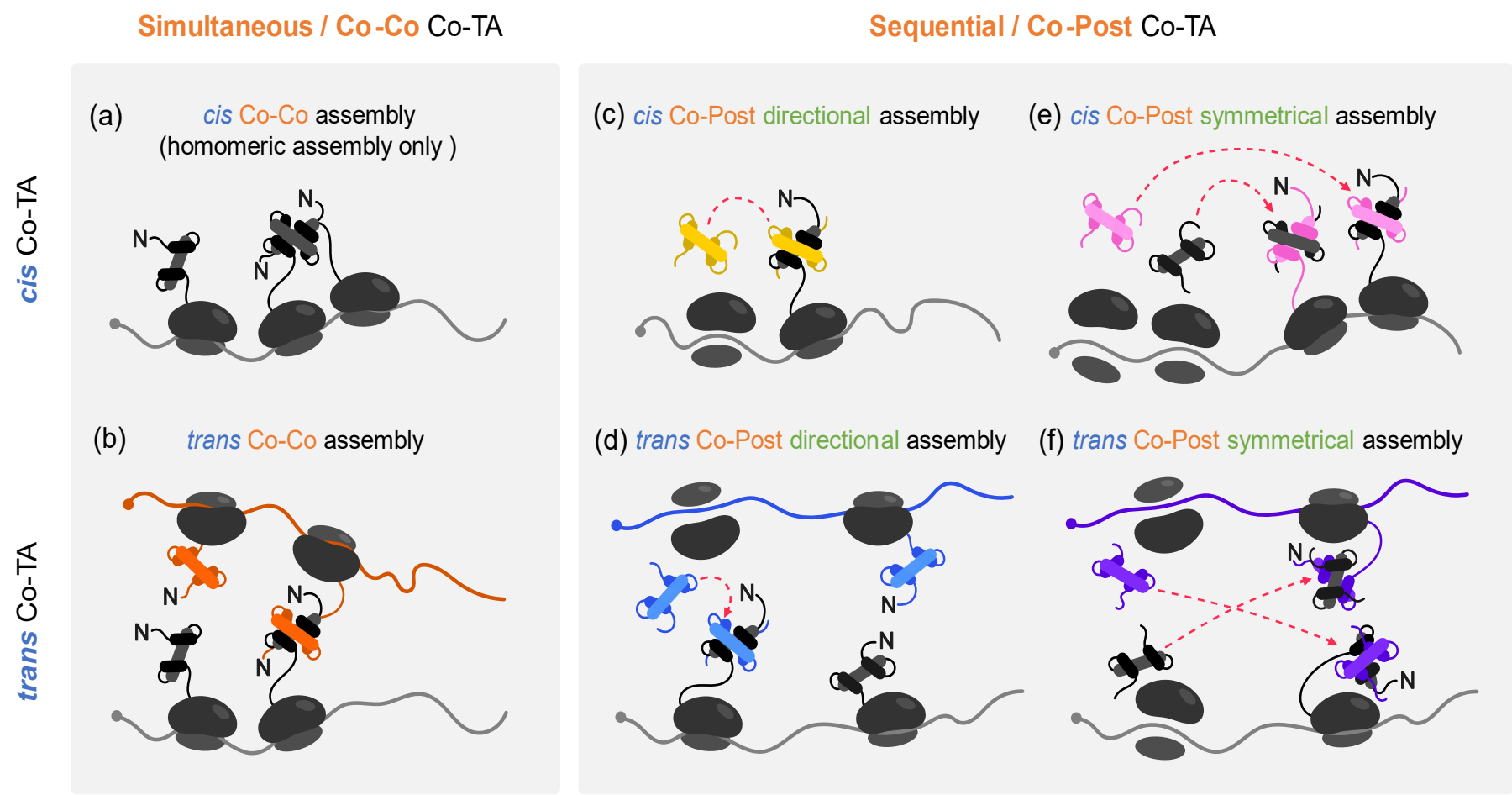
(1) processing orders:
simultaneous (Co-co)
vs
sequential (Co-post)

(2) ribosome translation locations:
on the same mRNA (*cis*)
vs
on different mRNAs (*trans*)

(3) the subunits assembly order
directional
vs
symmetrical

Complex type: • Homomeric • Heteromeric

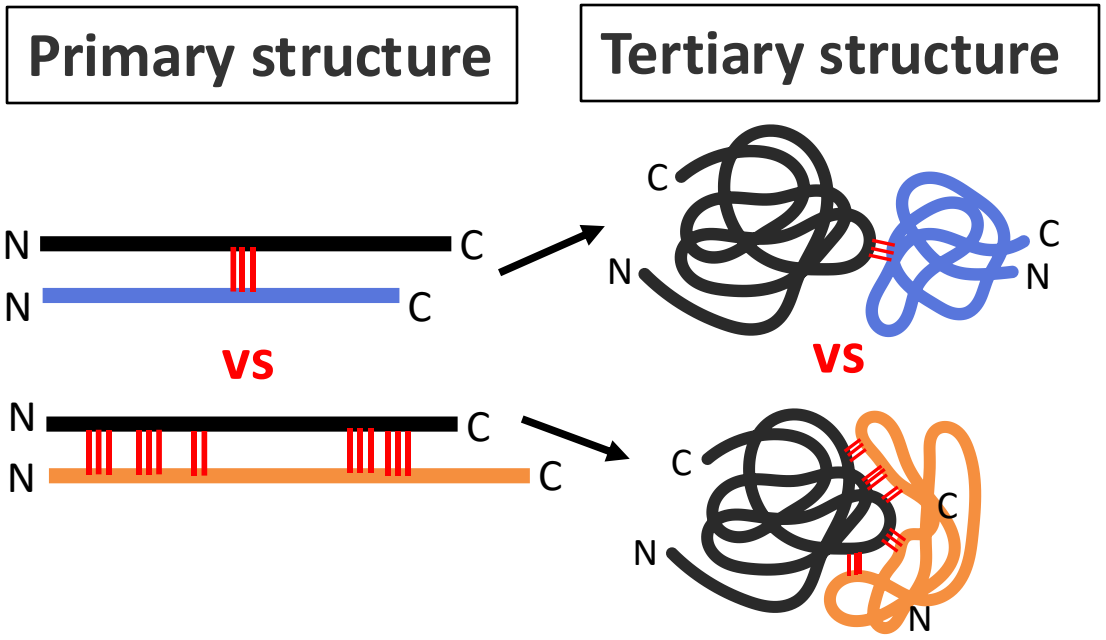
Focus of this study



Previous study:

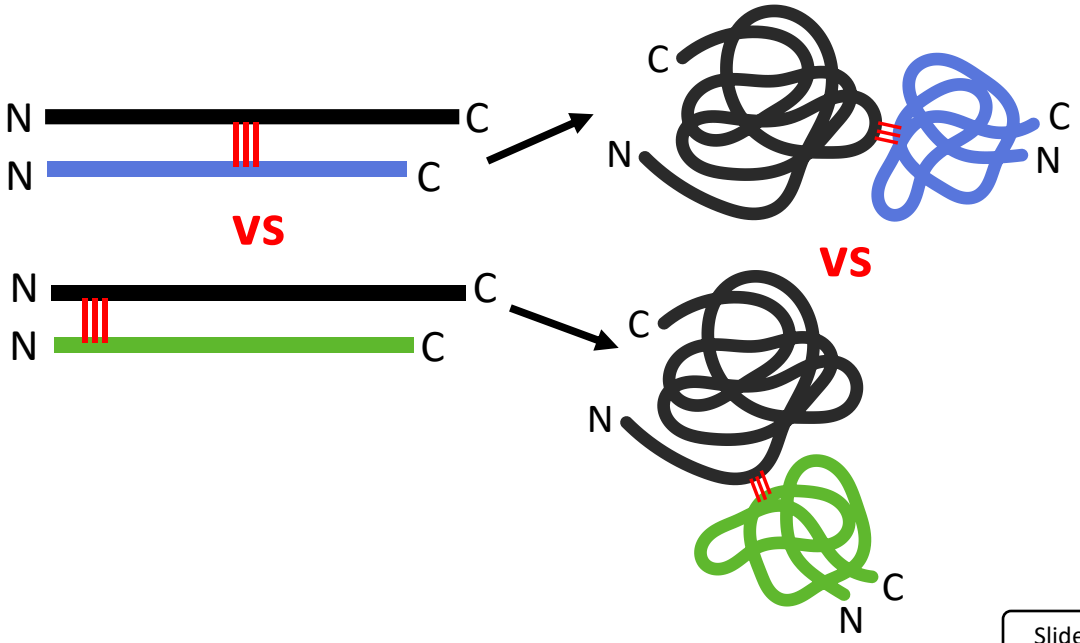
The mechanisms of complexes Co-TA may be distinguishable by **the areas of the interfaces** involved in the protein complex from one previous study

(Badonyi and Marsh, 2022 --- PMID: 35899946)



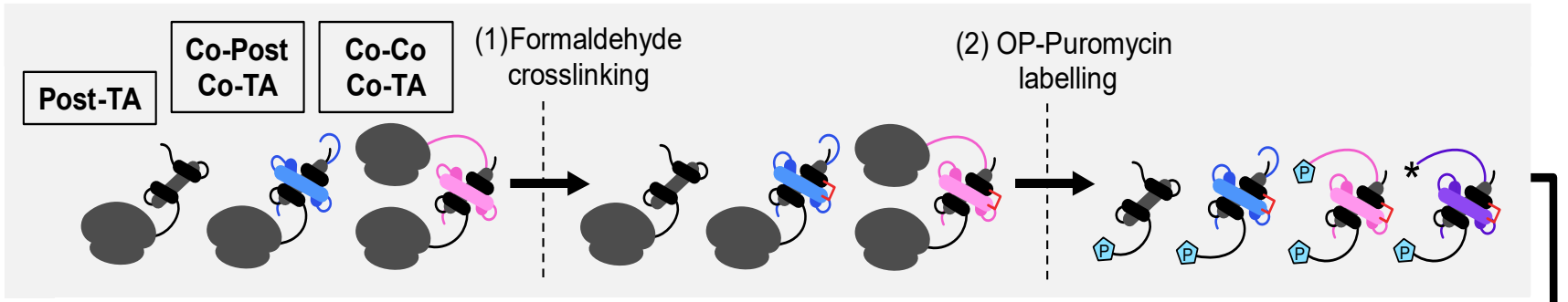
My Hypothesis:

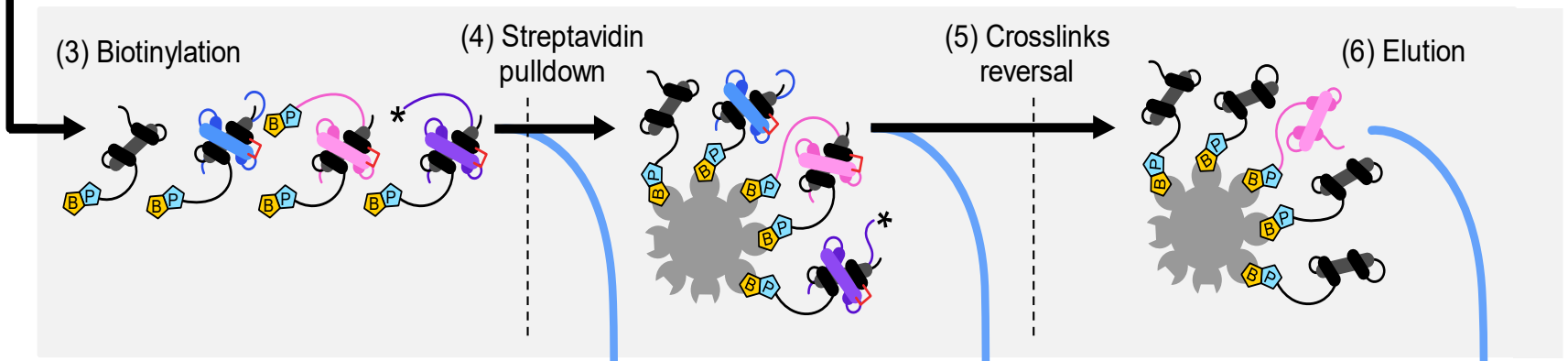
The mechanisms of complexes Co-TA may be distinguishable by **the relative location of the buried surfaces** of the component subunits in their protein sequence



Experiment 0: Sample preparation

Yeast S2883 extract





3 lists of samples

(1) The candidates

(2) The positive control

(3) The negative control

The experiments were performed and the three lists were given by Robert Crawford from Hubbard' lab

Experiment --- the flowchart shows the procedure

1 Download PDB files

1.1
Find available PDB model

↓

1.2
Confirm PDB model applicability through EBI-MUSCLE alignment

↓

1.3
Download the PDB file

2 Calculate ASA values

2.1
Extract the **spatial location data** of:
(1) The individual subunits × 2
(2) The two-subunit complex

↓

2.2
Run external program 'naccess', get **the absolute solvent accessibility (ASA)** of:
(1) The individual subunits × 2
(2) The two-subunit complex

↓

2.3
Subtract the ASA data of the two-subunit complex by the data of the two individual subunits respectively, to determine **the buried residues** and **their ASA data (ASA_{buried})** in the two subunits

3 Mapping data

3.1 --- Data Preparation
- Extract the following date from PDB file:
(1) Residue indices
(2) Residue three-letter codes
(3) The ASA_{buried} of residues
- Generate single-letter code PDB sequence

↓

3.2 --- Sequence alignment
- Load the UniProt sequence
- Align the PDB and the UniProt sequences
- Find the best alignment

↓

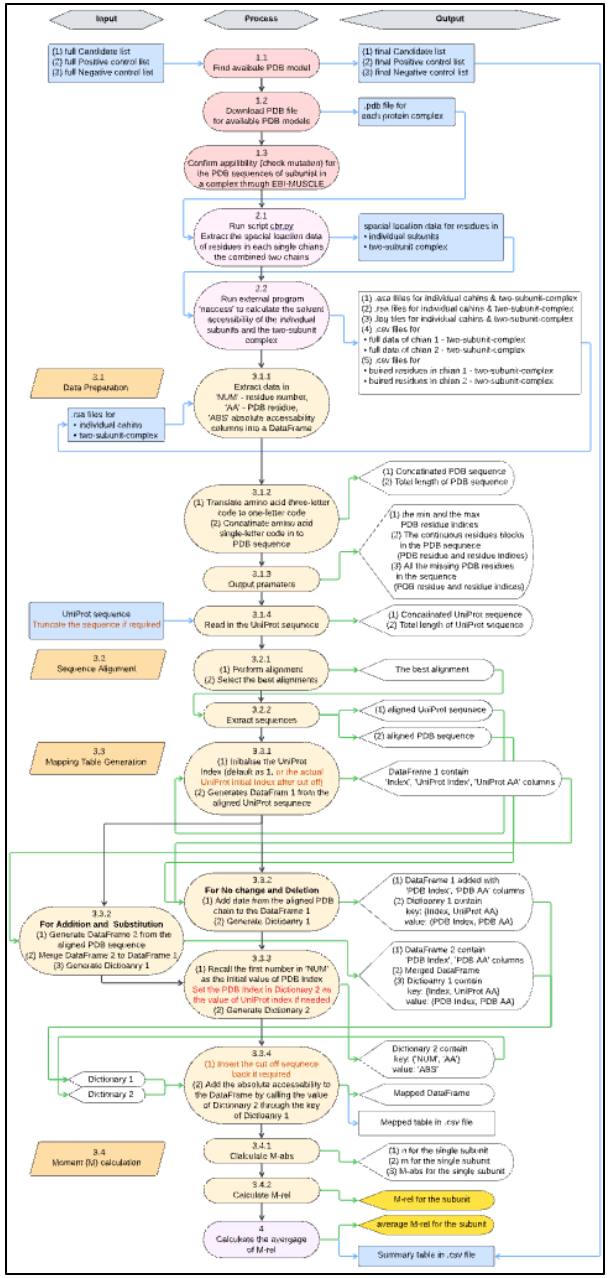
3.3 --- Generate mapped table
- Extract residue indices and residue code from the aligned PDB sequence
- Extract residue indices and residue code from the aligned UniProt sequence
- Integrate:
(1) the PDB sequence data
(2) the UniProt sequence data
(3) the ABS_{buried} data

↓

3.4 --- Moment (M) calculation
Calculate **the relative Moment (M-rel)** of each individual subunit

4 Calculate average M-rel

4
Calculate **the average M-rel** of each subunit if there are replicates



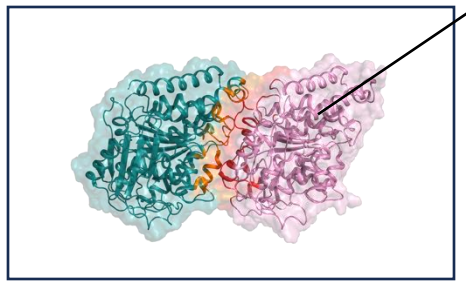
Experiment 1: Identify available and applicable complex PDB model

1 Download PDB files

1.1
Find available PDB model

1.2
Confirm PDB model applicability through EBI-MUSCLE alignment

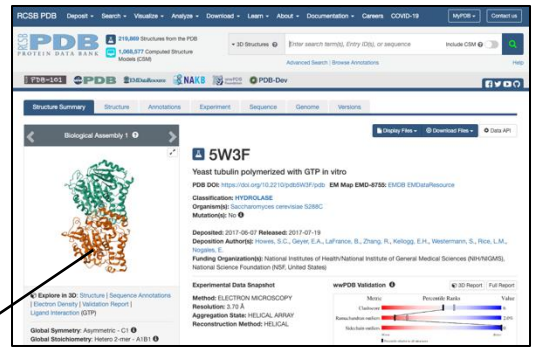
1.3
Download the PDB file



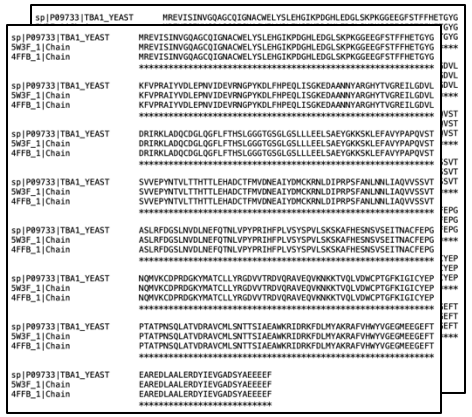
E.g. **The candidate list**

Number	Sample	Uniprot	Gene Name	ORF	Complex Name	Notes	Mer
1	release_XL	P09733	TUB1	YML085C	Tubulin alpha-beta heterodimeric complex, TUB1 variant	candidate	2
2	release_XL	P28834	IDH1	YNL037C	Mitochondrial isocitrate dehydrogenase complex (NAD+)	candidate	2
3	release_XL	P16861	PFK1	YGR240C	6-phosphofructokinase complex	known	2
4	release_XL	P15624	FRS1	YLR060W	Phenylalanyl-tRNA synthetase complex	candidate	2
5	release_XL	P07149	FAS1	YKL182W	Fatty-acyl-CoA synthase	known	2
...							

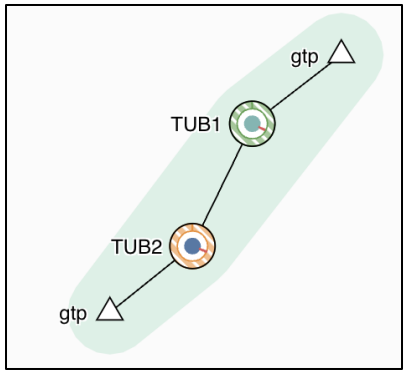
(1) Search in PDB database



(2) EBI-MUSCLE alignment



(3) Confirm interaction by EBI Complex Portal

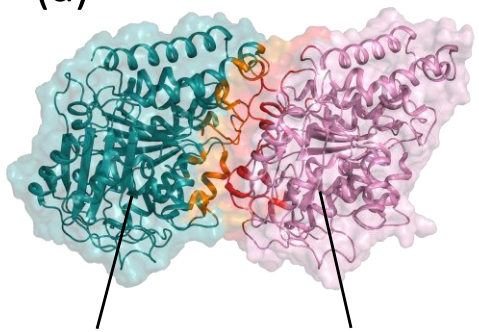


Result: **The final candidate list**

Number	Complex Name	PDB model	Uniprot 1	Gene 1	Chian 1	Uniprot 2	Gene 2	Chian 2
1	Tubulin alpha-beta heterodimeric complex, TUB1 variant	5W3F	P09733	TUB1	A	P02557	TUB2	B
2	Mitochondrial isocitrate dehydrogenase complex (NAD+)	3BLV	P28834	IDH1	A/C/E/G	P28241	IDH2	B/D/F/H
3	6-phosphofructokinase complex	3O8O	P16861	PFK1	A/C/E/G	P16862	PFK2	B/D/F/H
	Phenylalanyl-tRNA synthetase complex	NO	P15624	FRS1	-	-	-	-
4	Fatty-acyl-CoA synthase	2UV8	P07149	FAS1	A/B/C	P07149	FAS1	G/H/I
...								

Experiment 2: Buried residues visualisation

(a)

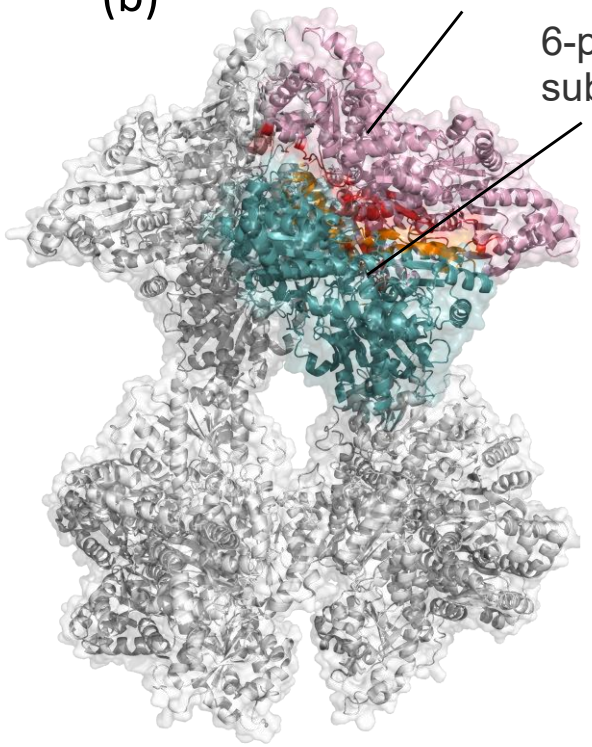


Tubulin-α subunit (Chain A)

Tubulin-β subunit (Chain B)

Tubulin alpha-beta heterodimeric complex, TUB1 variant
PDB: 5W3F

(b)

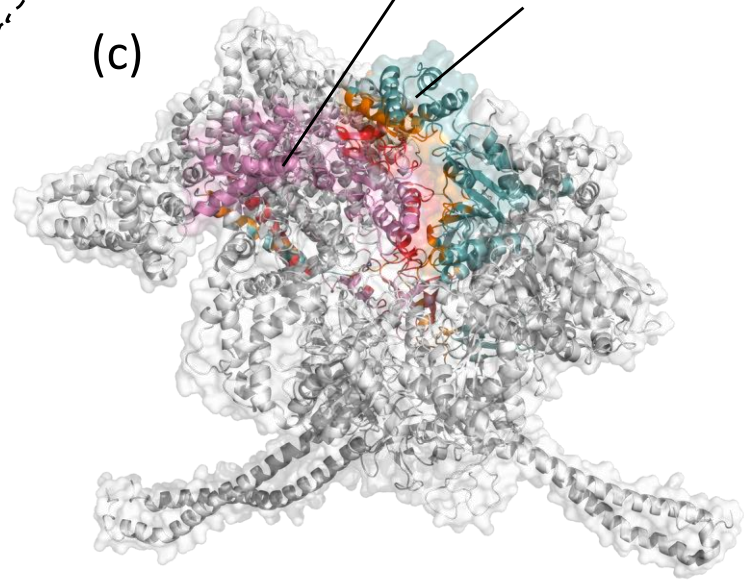


6-phosphofructokinase subunit-α (Chain A)

6-phosphofructokinase subunit-β (Chain B)

6-phosphofructokinase complex (PFK)
PDB: 3O8O

(c)



RPT1 isoform 2 (Chain I)

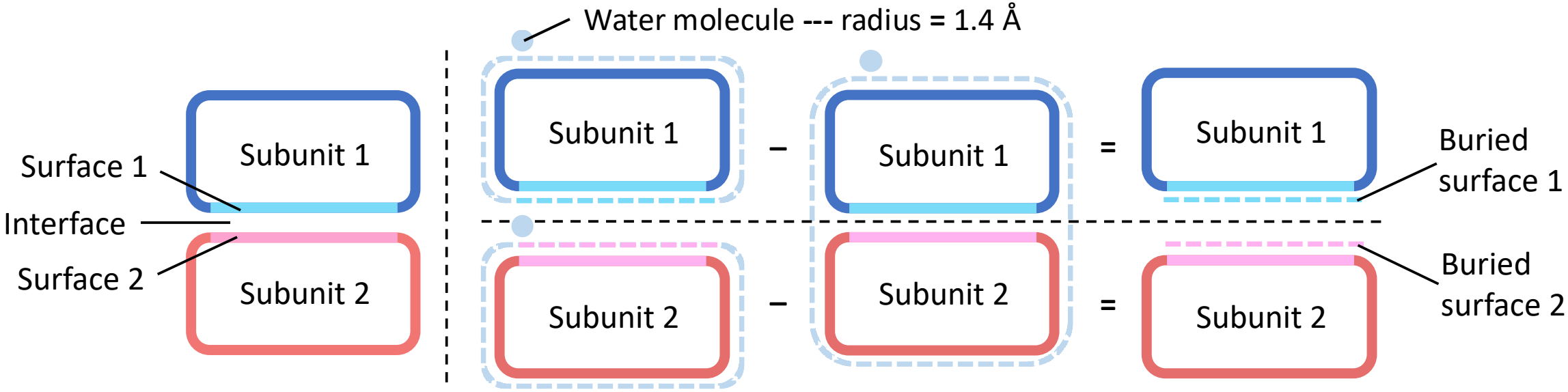
RPT1 isoform 1 (Chain H)

26S proteasome complex
PDB: 7QO4

Simple complex

Complicate complex

Experiment 2: Calculate **Absolute Solvent Accessability (ASA)**



Experiment 2: Calculate Absolute Solvent Accessability (ASA)

2 Calculate ASA values

2.1
Extract the **spatial location data** of:
(1) The individual subunits × 2
(2) The two-subunit complex

↓

2.2
Run external program 'naccess', get **the absolute solvent accessibility (ASA)** of:
(1) The individual subunits × 2
(2) The two-subunit complex

↓

2.3
Subtract the ASA data of the two-subunit complex by the data of the two individual subunits respectively, to determine **the buried residues** and **their ASA data (ASA_{buried})** in the two subunits

⇩

(1) Download PDB file  **5w3f.pdb**

↓

(2) Extract spatial location data (into .pdb files)

Complex contain chain A & B

```
ATOM      1  N   MET A   1      297.004 429.686 338.449  1.00 30.00           N
ATOM      2  CA  MET A   1      296.696 429.564 337.028  1.00 30.00           C
ATOM      3  C   MET A   1      297.349 428.322 336.429  1.00 30.00           C
ATOM      4  O   MET A   1      298.315 427.758 336.943  1.00 30.00           O
ATOM      5  CB  MET A   1      297.150 430.813 336.272  1.00 20.00           C
```

Chain B

```
ATOM   6768  N   MET B   1      298.832 430.781 293.952  1.00 30.00           N
```

Chain A

```
ATOM      1  N   MET A   1      297.004 429.686 338.449  1.00 30.00           N
ATOM      2  CA  MET A   1      296.696 429.564 337.028  1.00 30.00           C
ATOM      3  C   MET A   1      297.349 428.322 336.429  1.00 30.00           C
ATOM      4  O   MET A   1      298.315 427.758 336.943  1.00 30.00           O
ATOM      5  CB  MET A   1      297.150 430.813 336.272  1.00 20.00           C
ATOM      6  CG  MET A   1      296.390 432.075 336.646  1.00 20.00           C
ATOM      7  SD  MET A   1      294.621 431.943 336.325  1.00 20.00           S
ATOM      8  CE  MET A   1      294.605 431.856 334.537  1.00 20.00           C
ATOM      9  N   ARG A   2      296.816 427.867 335.300  1.00 87.74           N
ATOM     10  CA  ARG A   2      297.340 426.696 334.630  1.00 84.28           C
...
```

↓

(3) Run 'naccess' to calculate the ASA data (into .rsa files)

Complex contain chain A & B

```
REM RES _ NUM         All-atoms
REM                     ABS   REL
RES MET A   1         136.04  70.1
RES ARG A   2          92.00  38.5
RES GLU A   3           4.41   2.6
```

Chain B

```
REM RES _ NUM         All-atoms
REM                     ABS   REL
RES MET B   1          81.69  42.1
```

Chain A

```
REM RES _ NUM         All-atoms
REM                     ABS   REL
RES MET A   1         136.04  70.1
RES ARG A   2          92.00  38.5
RES GLU A   3           4.41   2.6
...
```

(4) Subtract data by equation 1, save in .csv files

Equation 1

$$(1)\ ASA_{buried\text{-}sub1} = ASA_{sub1} - ASA_{complex\text{-}sub1}$$
$$(2)\ ASA_{buried\text{-}sub2} = ASA_{sub2} - ASA_{complex\text{-}sub2}$$

↓

(1) subtracted data

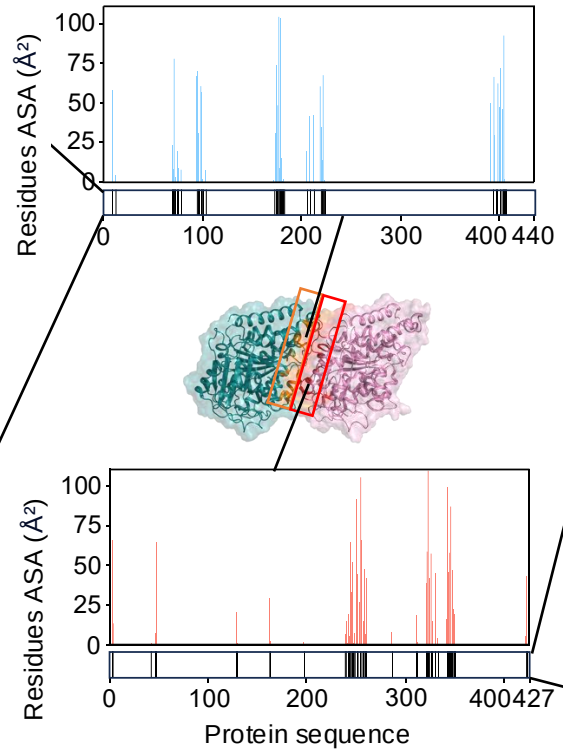
Result:
- **5w3f_A-AB.csv**
- **5w3f_B-AB.csv**

(2) Buried residues

Chain A: 440 aa, 43 buried residues
Chain B: 427 aa, 51 buried residues

Result: Calculated **Absolute Solvent Accessability (ASA)**

REM	RES_NUM			All-atoms
REM	**AA**	**CHAIN**	**NUM**	**ASA**
RES	GLN	A	11	59.35
RES	GLN	A	15	3.92
RES	GLU	A	72	23.66
RES	PRO	A	73	8.19
RES	ASN	A	74	80.05
…				
RES	HIS	A	407	46.89
RES	TRP	A	408	95.12
RES	TYR	A	409	1.66



REM	RES_NUM			All-atoms
REM	**AA**	**CHAIN**	**NUM**	**ASA**
RES	MET	B	1	65.54
RES	ARG	B	2	13.07
RES	ASP	B	41	0.33
RES	GLU	B	45	6.64
RES	ARG	B	46	64.26
…				
RES	THR	B	351	18.81
RES	GLN	B	424	4.76
RES	TYR	B	425	42.5

Experiment 3 & 4: Calculate **average relative Moment (M-rel)**

3 Mapping data

3.1 --- Data Preparation
- Extract the following date from PDB file:

(1) Residue indices
(2) Residue three-letter codes
(3) The ASA_{buried} of residues
- Generate single-letter code PDB sequence

3.2 --- Sequence alignment
- Load the UniProt sequence
- Align the PDB and the UniProt sequences
- Find the best alignment

3.3 --- Generate mapped table
- Extract residue indices and residue code from the aligned PDB sequence
- Extract residue indices and residue code from the aligned UniProt sequence
- Integrate:
  (1) the PDB sequence data
  (2) the UniProt sequence data
  (3) the ABS_{buried} data

3.4 --- Moment (M) calculation
Calculate **the relative Moment (M-rel)** of each individual subunit

4 Calculate average M-rel

4
Calculate **the average M-rel** of each subunit if there are replicates

(1) Generate alignment by Biopython package

```
Alignment:

target      0 MREIIHISTGQ   EYQQYQEATVEDDEEVDENGDFGAPQNQDEPITENFE 457
            0 ||||||||||| ... |||||||------------------------------ 457
query       0 MREIIHISTGQ   EYQQYQE------------------------------ 427
```

Unique Alignment relationship

(2) Form the mapped tables

Uniprot data — link — PDB data + ASA values

Chain A

Exported result in file 5w3f_A_P09733_cal.csv

	Index	UniProt_Index	UniProt_AA	PDB_Index	PDB_AA	ABS
0	1	1	M	1	M	0.0
1	2	2	R	2	R	0.0
2	3	3	E	3	E	0.0
3	4	4	V	4	V	0.0
4	5	5	I	5	I	0.0
..	...	...	...	...	...	...
442	443	443	E		-	
443	444	444	E		-	
444	445	445	E		-	
445	446	446	E		-	
446	447	447	F		-	

[447 rows x 6 columns]

Chain B

455	456	456	F		-
456	457	457	E		-

[457 rows x 6 columns]

(3) Calculate M-rel by Equation 2

Equation 2 --- M-rel

$$M = \sum_{i-1}^{n} \Delta a \times \Delta d$$

$$with\ \Delta a = \frac{a}{\sum_{i=1}^{n} a}\ and\ \Delta d = (i - m) \div n$$

$\boldsymbol{n}$ – the total number of residues in the protein sequence
$\boldsymbol{m}$ – the absolute middle value, which is n divided by 2 then plus 0.5, disregarding to its odevity
$\boldsymbol{i}$ – the location index of each residue
$\boldsymbol{d}$ - the distance of each residue to the absolute middle point
$\boldsymbol{a}$ - the absolute ASA of the corresponding residue

Experiment 3 & 4: Calculate **average relative Moment (M-rel)**

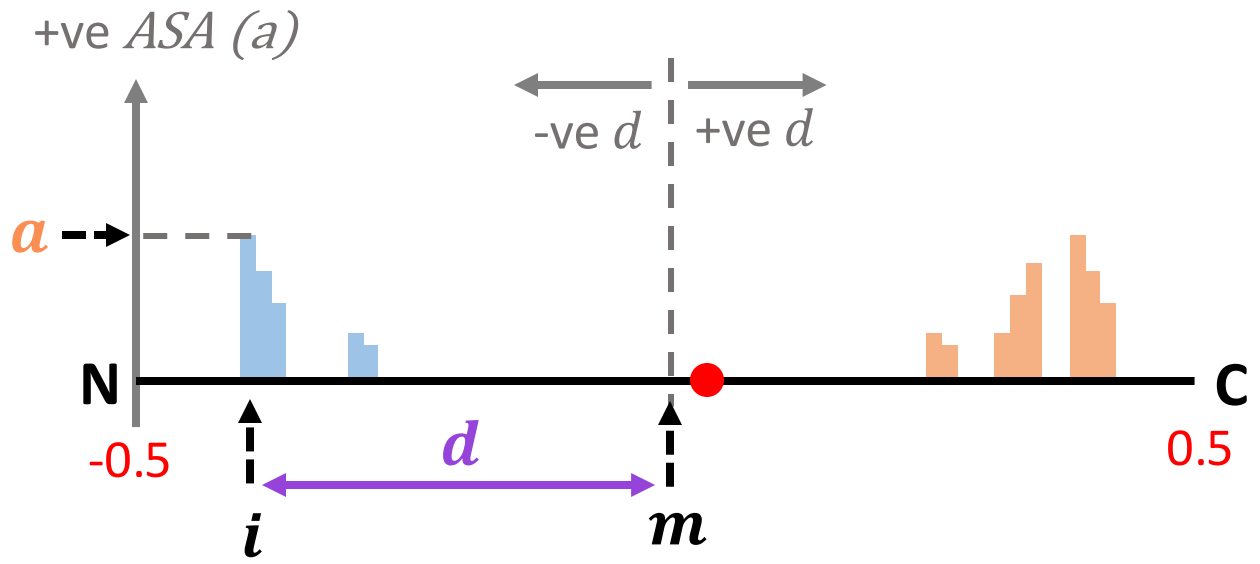
(3) Calculate M-rel by Equation 2

Equation 2 --- M-rel

$$M = \sum_{i-1}^{n} \Delta a \times \Delta d$$

$$with\ \Delta a = \frac{a}{\sum_{i=1}^{n} a}\ and\ \Delta d = (i - m) \div n$$

- n – the total number of residues in the protein sequence
- m – the absolute middle value, which is n divided by 2 then plus 0.5, disregarding to its odevity
- i – the location index of each residue
- d - the distance of each residue to the absolute middle point
- a - the absolute ASA of the corresponding residue



$$\sum \text{r:} \left(\text{▮} \div \text{▮▮▮▮} = \Delta a \right) \times \left(\frac{d}{} \div \text{N} \text{——} \text{C} = \Delta d \right)$$

e.g. 0.03 *e.g.* -0.41

Result: The final list of PDB complex models used as samples

Group	Sample number	PDB ID	Complex name
Candidate	1	1JK9	SOD1-CCS1 superoxide dismutase heterodimer
	2	7X3K	Ribosome-associated complex
	3	4CRN	Translation release factor ERF1-ERF3 complex
	4	5W3F	Tubulin alpha-beta heterodimeric complex, TUB1 variant
	5	3BLV	Mitochondrial isocitrate dehydrogenase complex (NAD+)
	6	1F60	Elongation factor eEF1 complex, variant CAM1
	7	8DAR	CDC48-RAD23-UFD2 complex
	8	1ID3	Nucleosome, variant HTA1-HTB2
Positive control	1	3O8O	6-phosphofructokinase complex (PFK)
	2	2UV8	Fatty-acyl-CoA synthase (FAS)
	3	6O07	NatA N-alpha-acetyltransferase complex (NatA)
	4	2HRK	Methionyl glutamyl tRNA synthetase complex (MetGluRS)
	5	6ZL0	COPII vesicle coat complex (COPII)
	6	7QO4	26S proteasome complex
	7	7WOO	Nuclear pore complex (NPC)
Negative control	1	6U9D	Acetolactate synthase complex
	2	2VDU	tRNA (guanine-N(7)-)-methyltransferase
	4	6EM5	PeBoW complex
	6	6I52	Replication protein A complex
	7	6YLE	RIX1 complex
	8	3T4N	Snf1 protein kinase complex variant GAL83
	9	2WWA	SSH1 translocon complex

Number of complexes

- Candidate: 8 out of 39
- +ve control: 7 out of 14
- -ve control: 8 out of 29

23 82

Directional
sequential *trans* Co-TA

Symmetrical
sequential *trans* Co-TA

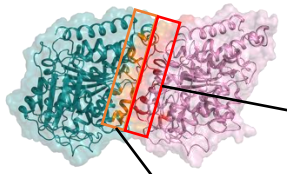
Methods summary

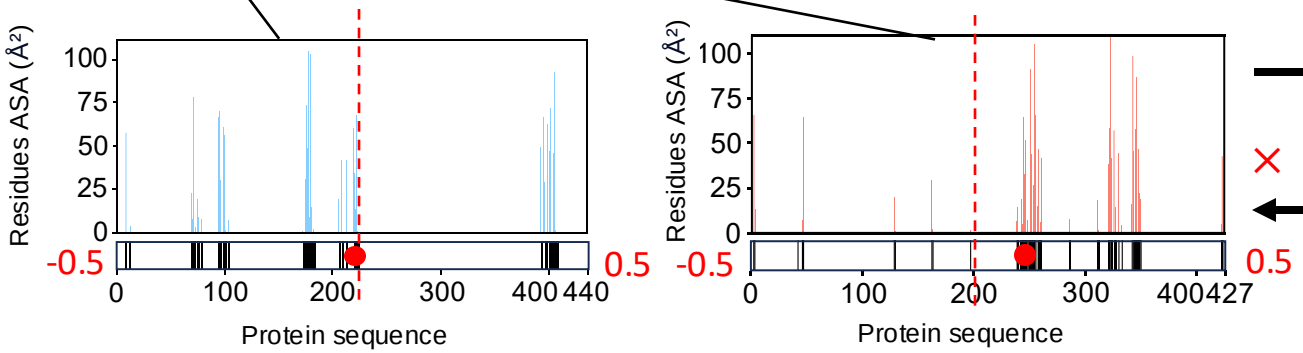
1 Download PDB files for all applicable complex models

1abc.pdb × 23

						X	Y	Z			
ATOM	1	N	MET	A	1	297.004	429.686	338.449	1.00	30.00	N
ATOM	2	CA	MET	A	1	296.696	429.564	337.028	1.00	30.00	C
ATOM	3	C	MET	A	1	297.349	428.322	336.429	1.00	30.00	C
ATOM	4	O	MET	A	1	298.315	427.758	336.943	1.00	30.00	O
ATOM	5	CB	MET	A	1	297.150	430.813	336.272	1.00	20.00	C
ATOM	6	CG	MET	A	1	296.390	432.075	336.646	1.00	20.00	C
ATOM	7	SD	MET	A	1	294.621	431.943	336.325	1.00	20.00	S
ATOM	8	CE	MET	A	1	294.605	431.856	334.537	1.00	20.00	C
ATOM	9	N	ARG	A	2	296.816	427.867	335.300	1.00	87.74	N
ATOM	10	CA	ARG	A	2	297.340	426.696	334.630	1.00	84.28	C
...											

2 Calculate ASA values, determined the buried residues





× 47

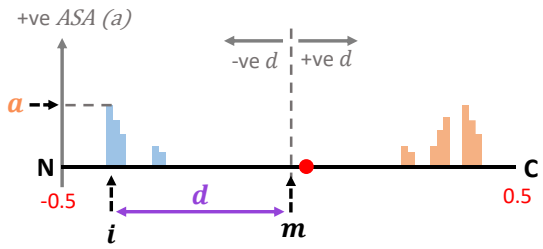
3 Mapping data & 4 Calculated the average M-rel

Quantitative Mass Spectrometry Data — Uniprot data — link — PDB data + ASA values — Structural data

Exported result in file 5w3f_A_P09733_cal.csv

	Index	UniProt_Index	UniProt_AA	PDB_Index	PDB_AA	ABS
0	1	1	M	1	M	0.0
1	2	2	R	2	R	0.0
2	3	3	E	3	E	0.0
3	4	4	V	4	V	0.0
4	5	5	I	5	I	0.0
..	...	...	...	...	...	...
442	443	443	E		–	
443	444	444	E		–	
444	445	445	E		–	
445	446	446	E		–	
446	447	447	F		–	

[447 rows x 6 columns]



$\sum r : \left(\text{|} \div \text{|}\!\text{|}\!\text{|} = \Delta a \right) \times \left(\frac{d}{\text{N}\text{——}\text{C}} = \Delta d \right)$

e.g. 0.03 *e.g.* -0.41

Result:

The full data

Group	Sample number	PDB ID	Combination	Chain 1								Chain 2							
				PDB chain	UniProt ID	Complex	uried residue	n	m	M-rel	Average	PDB chain	UniProt ID	Complex	uried residue	n	m	M-rel	Average
Candidate	1	1JK9		A	P00445	A-AB	35	154.0	77.5	0.08	0.08	B	P40202	B-AB	36	249.0	125.0	0.24	0.24
	2	7X3K		A	P32527	A-AB	55	433.0	217.0	-0.41	-0.41	B	P38788	B-AB	89	538.0	269.5	0.25	0.25
	3	4CRN		P	P05453	P-PX	48	685.0	342.0	0.21	0.21	X	P12385	X-PX	43	437.0	219.0	0.10	0.10
	4	5W3F		A	P09733	A-AB	43	447.0	224.0	-0.01	-0.01	B	P02557	B-AB	51	457.0	229.0	0.09	0.09
	5	3BLV		ACEG	P28834	A-AB C-CD E-EF G-GH	59 59 59 58	360.0	180.5	0.01 0.01 0.01 0.01	0.01	BDFH	P28241	B-AB D-CD F-EF H-GH	57 58 54 55	369.0	185.0	0.05 0.05 0.05 0.05	0.05
	6	1F60		A	P02994	A-AB	54	458.0	229.5	-0.15	-0.15	B	P32471	B-AB	47	206.0	103.5	0.35	0.35
	7	8DAR		ABCDEF	P25694	A-AG B-BG C-CG D-DG E-EG F-FG	1 28 22 38 8 6	835.0	418.0	-0.156 -0.177 -0.151 -0.171 -0.110 -0.153	-0.15	G	P33755	G-AG G-BG G-CG G-DG E-EG F-FG	4 27 23 24 8 4	580.0	290.5	0.27 -0.05 -0.17 -0.24 -0.14 0.27	-0.01
	8	1ID3		AE	P61830	A-AB E-EF	56 60	136.0	68.5	0.14 0.13	0.14	BF	P02309	B-AB F-EF	52 53	103.0 103.0	52.0 52.0	0.04 0.04	0.04
Positive control	1	3O8O		A	P16861	A-AB	79	987.0	494.0	0.09	0.09	B	P16862	B-AB	84	959.0	480.0	0.10	0.10
	2	2UV8		ABC	P19097	A-AG B-BH C-CI	109 109 108	1887.0	944.0	-0.30 -0.30 -0.30	-0.30	GHI	P07149	G-AG H-BH I-CI	157 157 156	2051.0	1026.0	0.29 0.29 0.29	0.29
	3	6O07		A	P07347	A-AB	103	854.0	427.5	-0.04	-0.04	B	P12945	B-AB	81	238.0	119.5	-0.10	-0.10
	4	2HRK		A	P46655	A-AB	21	708.0	354.5	-0.32	-0.32	B	P46672	B-AB	23	376.0	188.5	-0.28	-0.28
	5	6ZL0		AC	P38968	A-AB C-CD	51 51	1273.0	637.0	-0.13 -0.13	-0.13	BD	Q04491	B-AB D-CD	56 56	297.0	149.0	-0.01 -0.01	-0.01
	6	7QO4		H	P33299	H-HI	117			-0.02	-0.02	I	P40327	I-HI	109	437.0	219.0	-0.05	-0.05
	7	7WOO		HK	P48837	H-HI K-KL	106 113	541.0	271.0	0.26 0.27	0.27	IL	P14907	I-HI L-KL	98 90	823.0	412.0	0.37 0.38	0.38
Negative control	1	6U9D		BEFIJMNQRU	P07342	A-AC B-BD E-EG F-FH I-IK J-JL M-MO N-NP Q-QS R-RT U-UW V-VX	24 24 23 23 23 27 25 22 24 24 25 26	687.0	344.0	-0.07 -0.07 -0.06 -0.06 -0.07 -0.07 -0.06 -0.06 -0.07 -0.07 -0.07 -0.06	-0.07	DGHKLOPSTW	P25605	C-AC D-BD G-EG H-FH K-IK L-JL O-MO P-NP S-QS T-RT W-UW X-VX	20 20 21 20 20 20 21 20 20 20 21 20	309.0	155.0	-0.10 -0.10 -0.10 -0.09 -0.10 -0.10 -0.10 -0.10 -0.10 -0.10 -0.10 -0.10	-0.10
	2	2VDU		BD	Q03774	B-BE D-DF	28 27	444.0	222.5	0.00 0.00	0.00	EF	Q12009	E-BE F-DF	26 26	286.0	143.5	0.00 0.00	0.00
	3	6EM5	1	m	Q04660	m-mn	42	807.0	404.0	-0.02	-0.02	n	P53261	n-mn	50	605.0	303.0	-0.28	-0.28
	4		2	m	Q04660	m-mp	4			0.24	0.24	p	Q12024	p-mp	7	460.0	230.5	-0.15	-0.15
	5	6I52		B	P26754	B-BC	30	273.0	137.0	-0.06	-0.06	C	P22336	C-BC	28	621.0	311.0	0.41	0.41
	6	6YLE		AB	P53877	A-AC B-BD	80 75	555.0	278.0	-0.26 -0.28	-0.27	CD	P38883	C-AC D-BD	91 93	763.0	382.0	-0.07 -0.04	-0.05
	7	3T4N		A	P06782	A-AC	44	633.0	317.0	0.31	0.31	C	P12904	C-AC	58	322.0	161.5	0.12	0.12
	8	2WWA		A	P38353	A-AC	24	490.0	245.5	-0.36	-0.36	C	P52871	C-AC	17	88.0	44.5	0.26	0.26

Result:

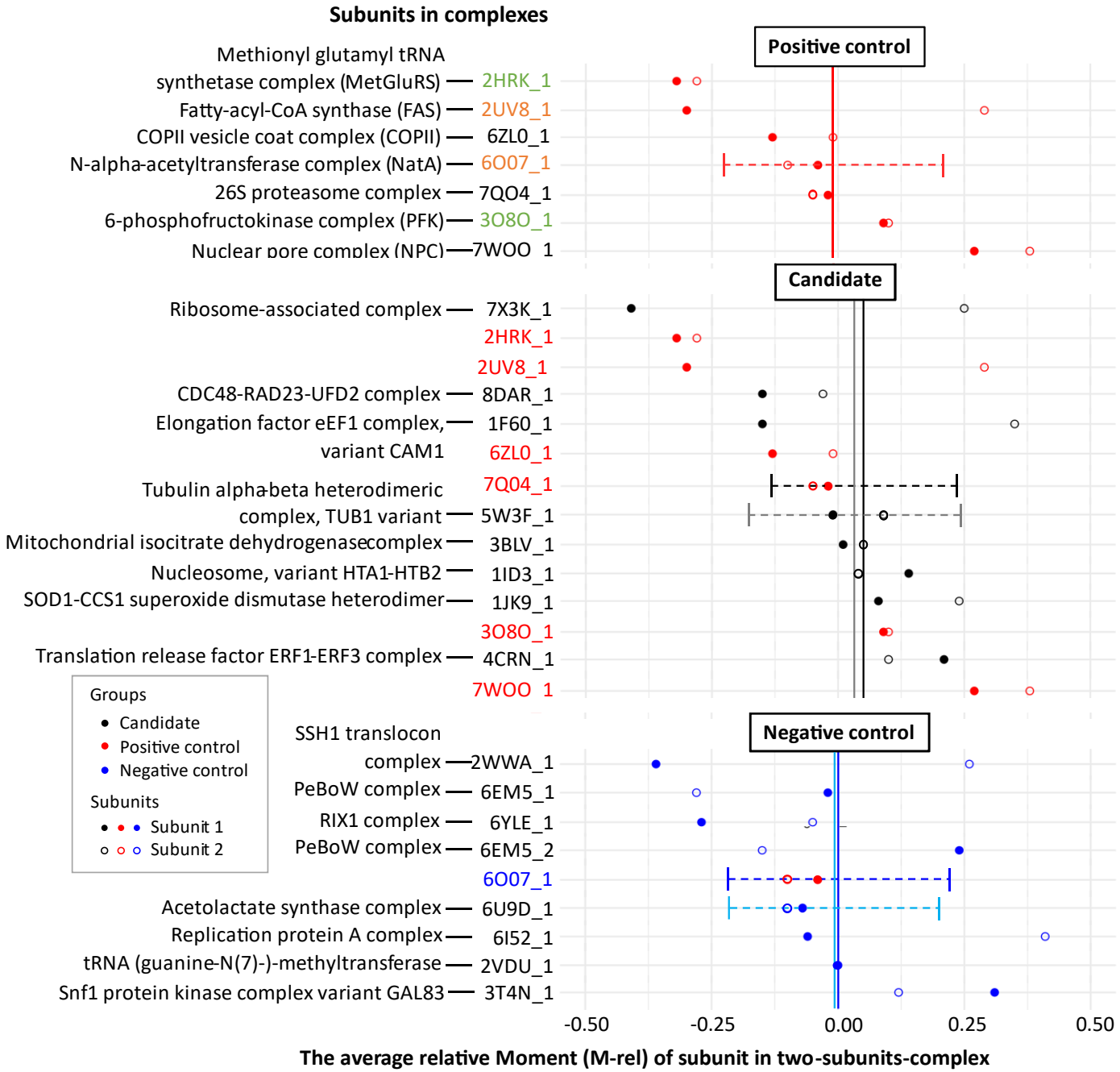
Average M-rel visualisation

- **Red label**: the positive control complexes that were also found in the candidate.
- **Blue label**: the positive control complex that was also found in the negative control.
- **Orange label**: complexes known to undergo directional sequential *trans* Co-TA
- **Green label**: complexes known to undergo symmetrical sequential *trans* Co-TA

- **Blue and white cells on the same line**: the candidate subunit found in the elute sample and in the release sample

- **Yellow cells**: the subunits that have negative average M-rel values

- **Green text**: the complex has two subunits only, no other non-interested subunits



Conclusion 1

There were No Significant Difference between the three lists

Average M-rel group mean & standard deviation:

- **Positive control**: 0.0093 ±0.2174 (red lines)

'Essential':

- **Candidates**: -0.0501 ±0.1844 (black lines)
- **Negative control**: -0.0004 ±0.2190 (dark blue lines)

'Include':

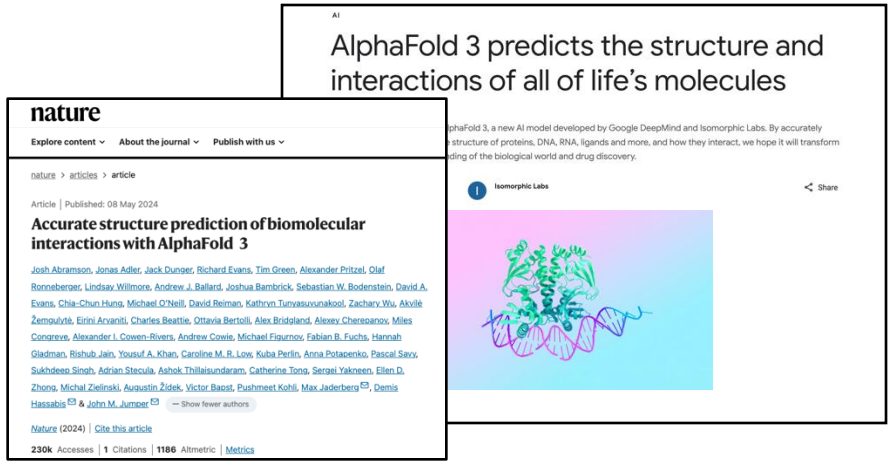
- **Candidates**: 0.0331 ±0.2106 (grey lines)
- **Negative control**: -0.0081 ± 0.2072 (light blue lines)

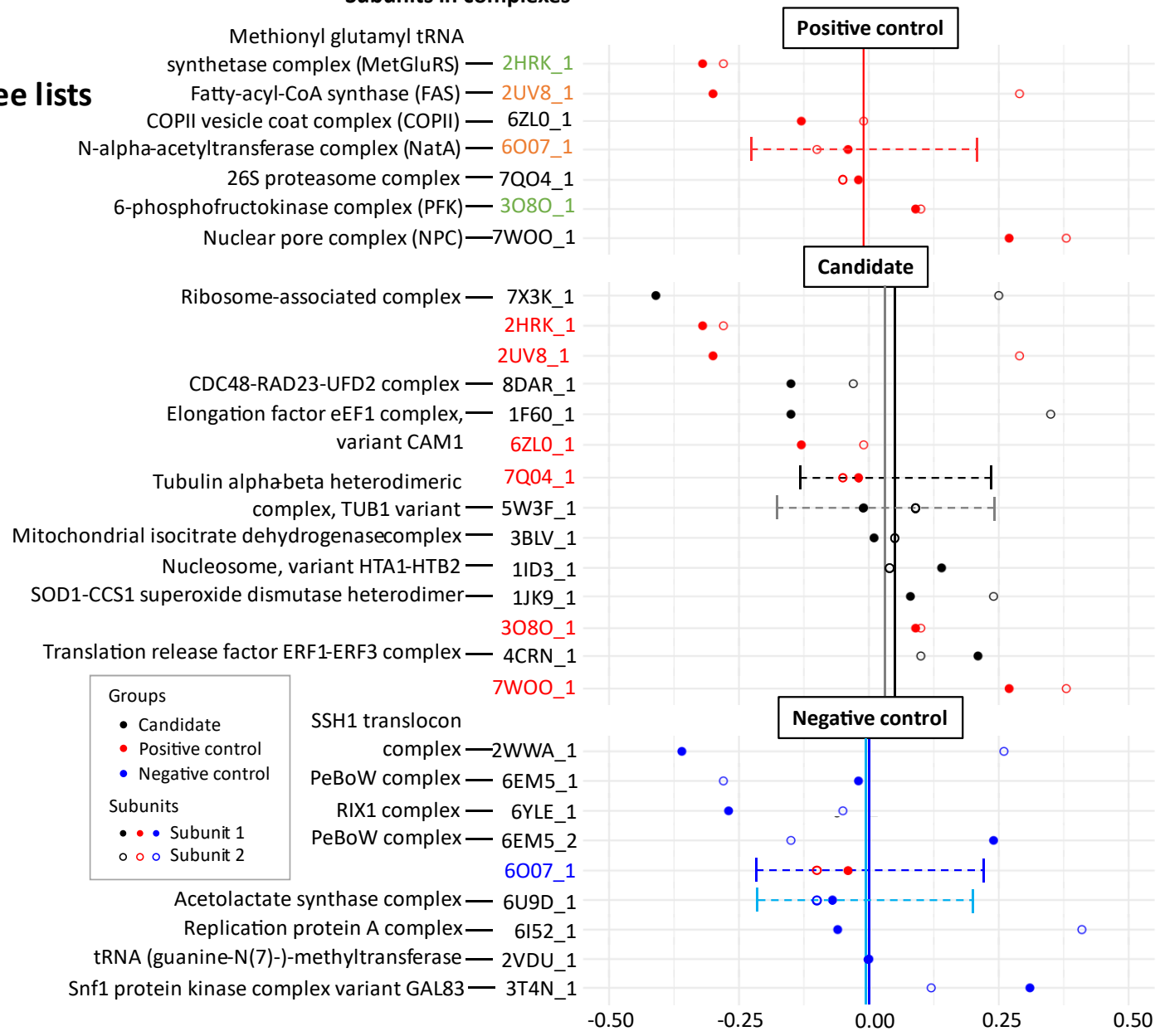
Likely due to

the small number of samples

--- constraint the persuasion of M-rel calculation

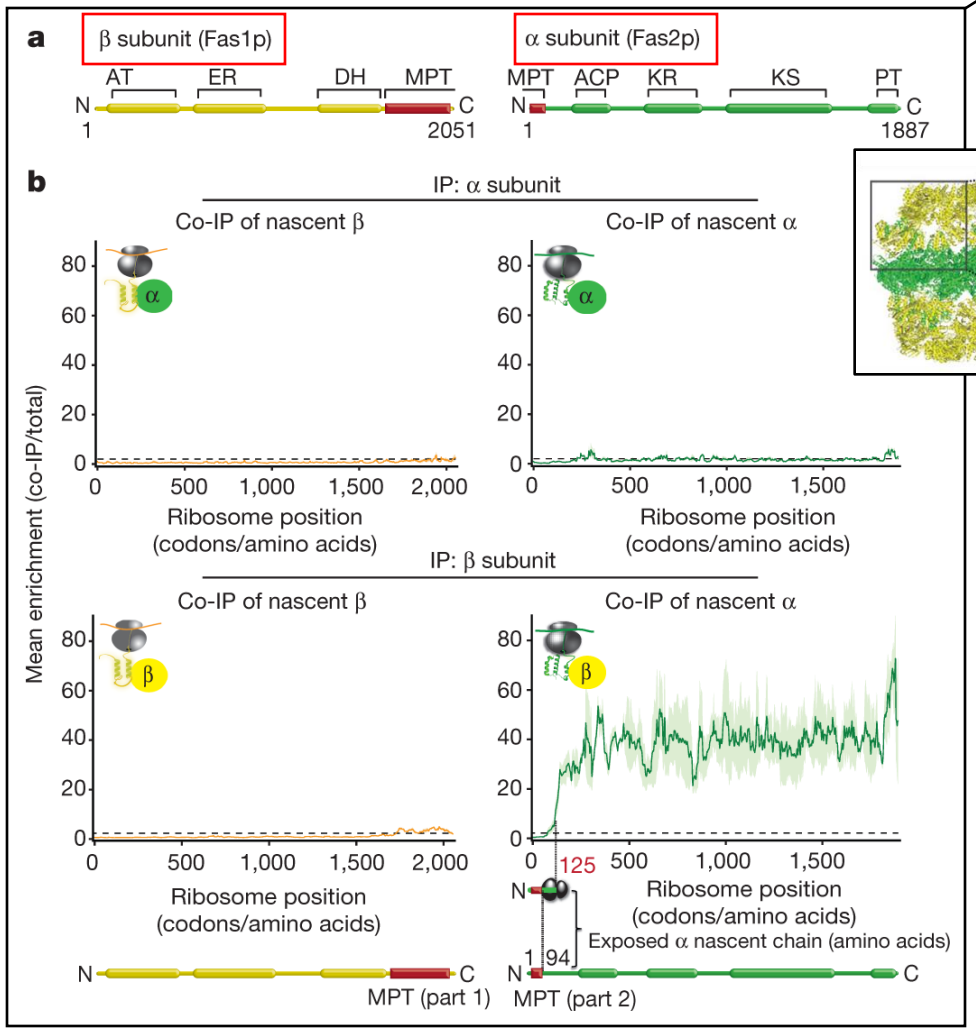
Improvement: Alpha Fold 3 & Deep learning

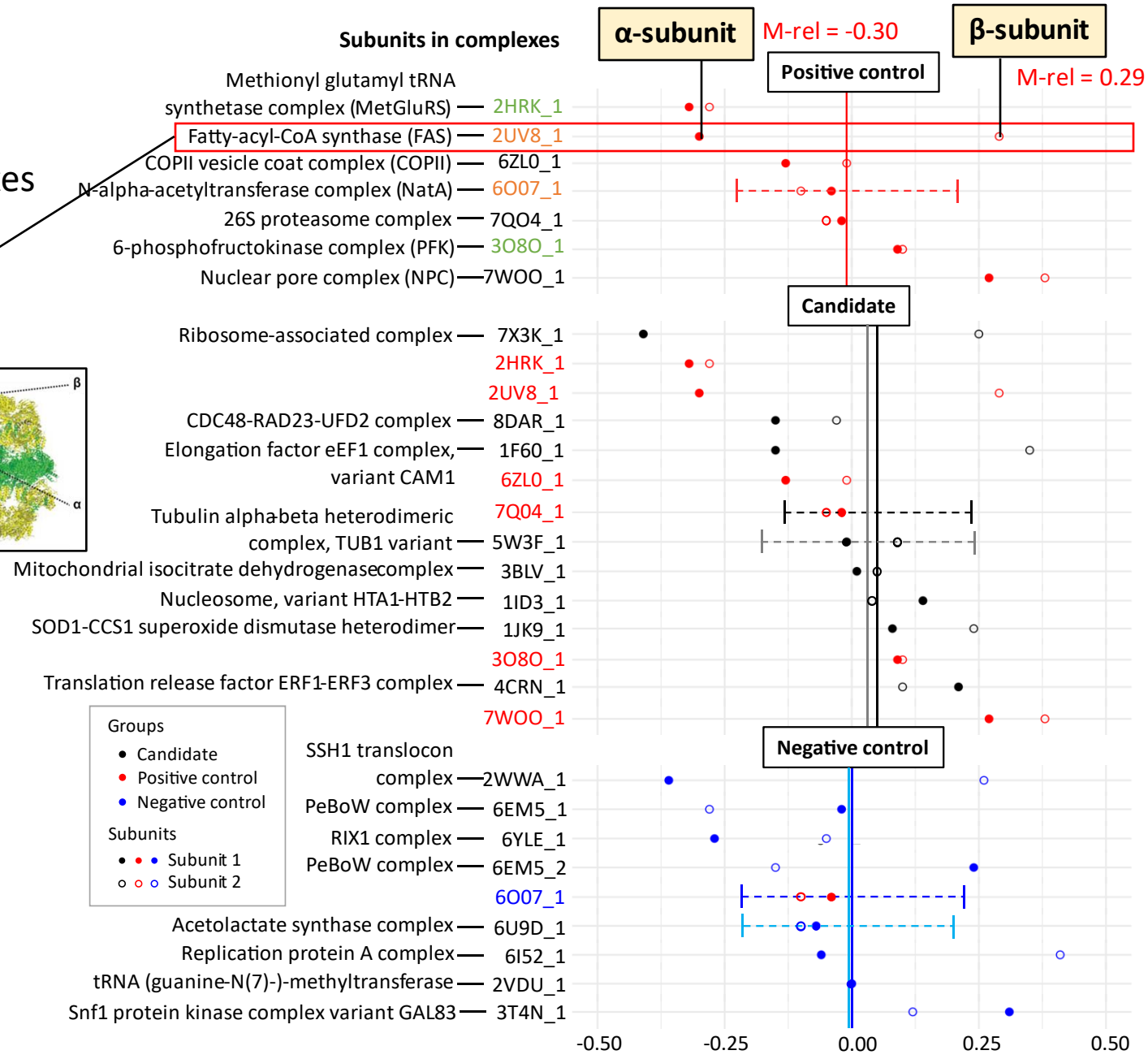




Conclusion 2

Subunit **average M-rel** can help to suggest
the potential *trans* Co-TA mechanism for complexes
--- true for (FAS, PDB: 2UV8)

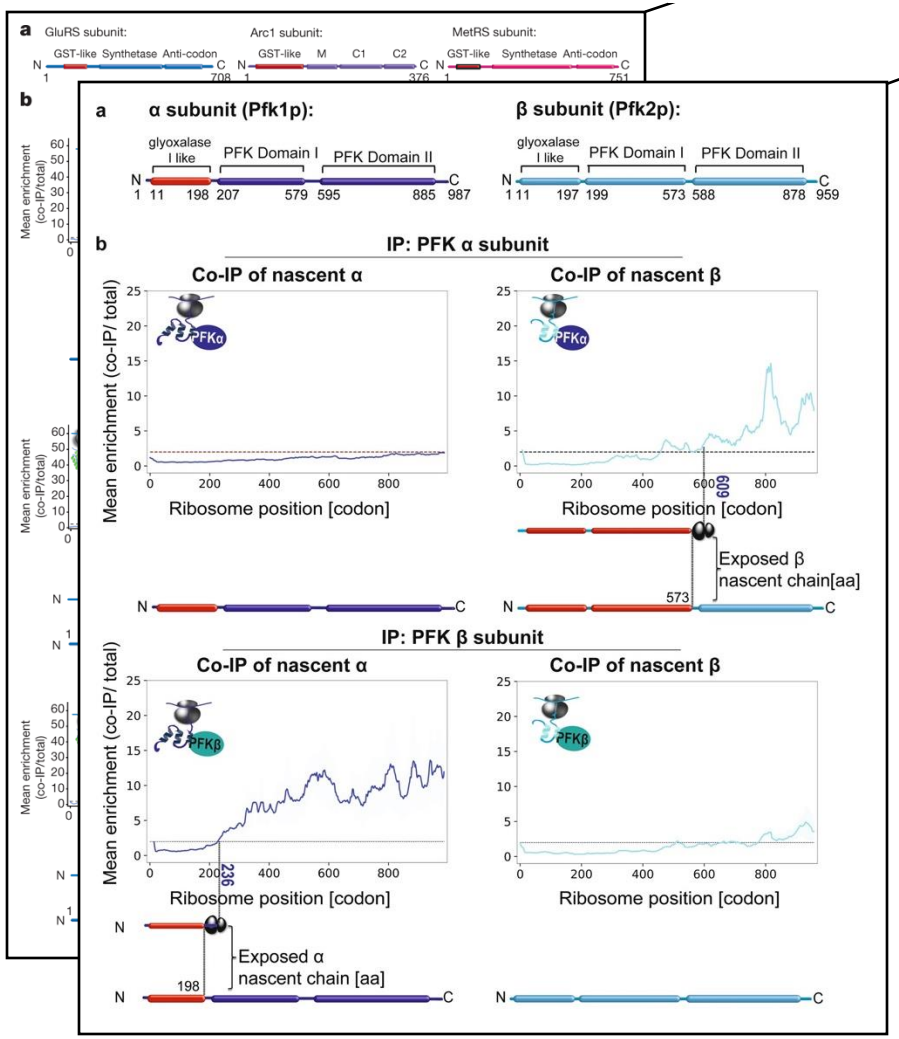




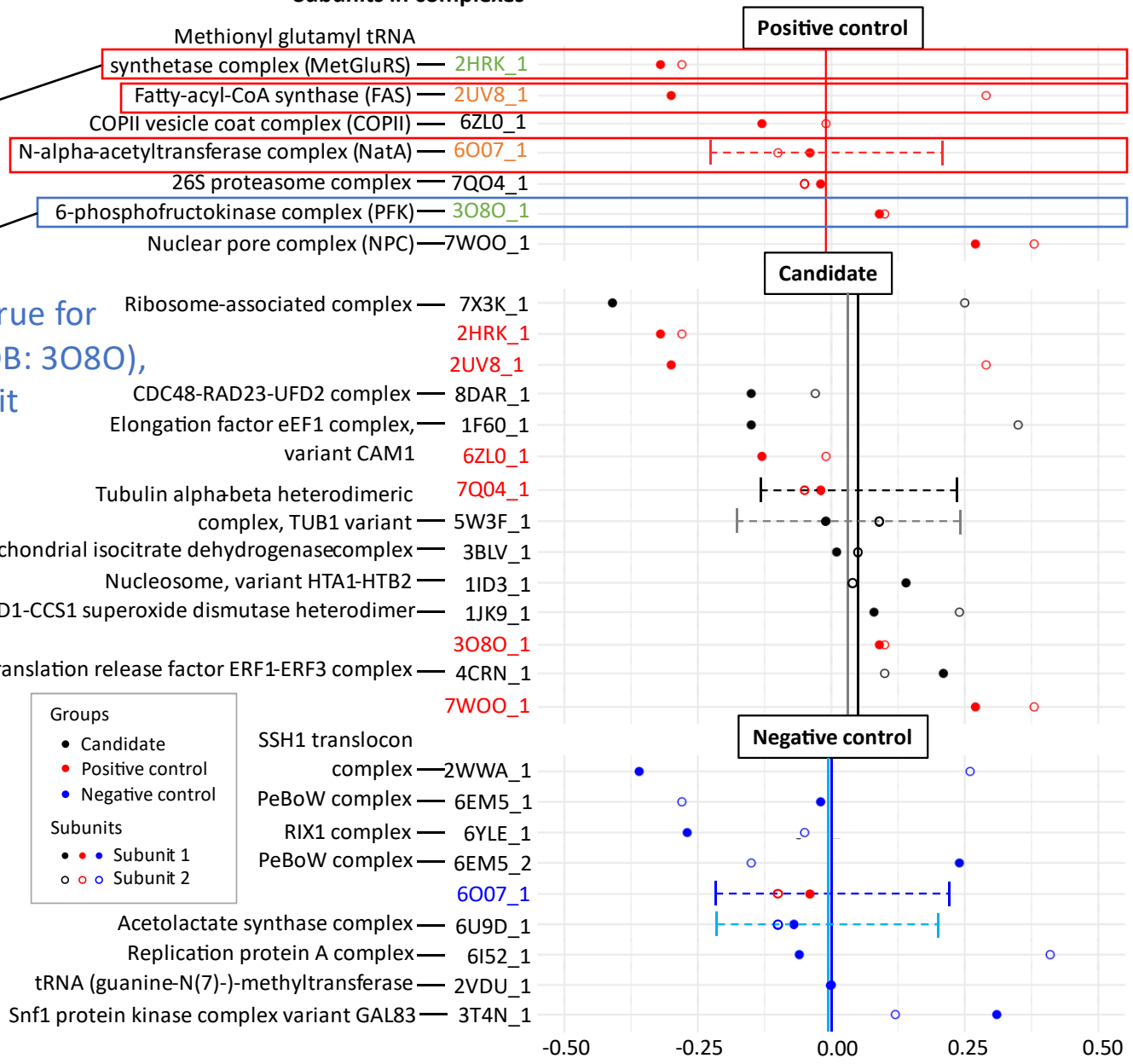
Shiber et al. (2018) --- PMID: 30158700

Conclusion 2

Subunit **average M-rel** can help to suggest
the potential *trans* Co-TA mechanism for complexes
--- true for (MetGluRS, PDB: 2HRK), (NatA, PDB: 6O07)



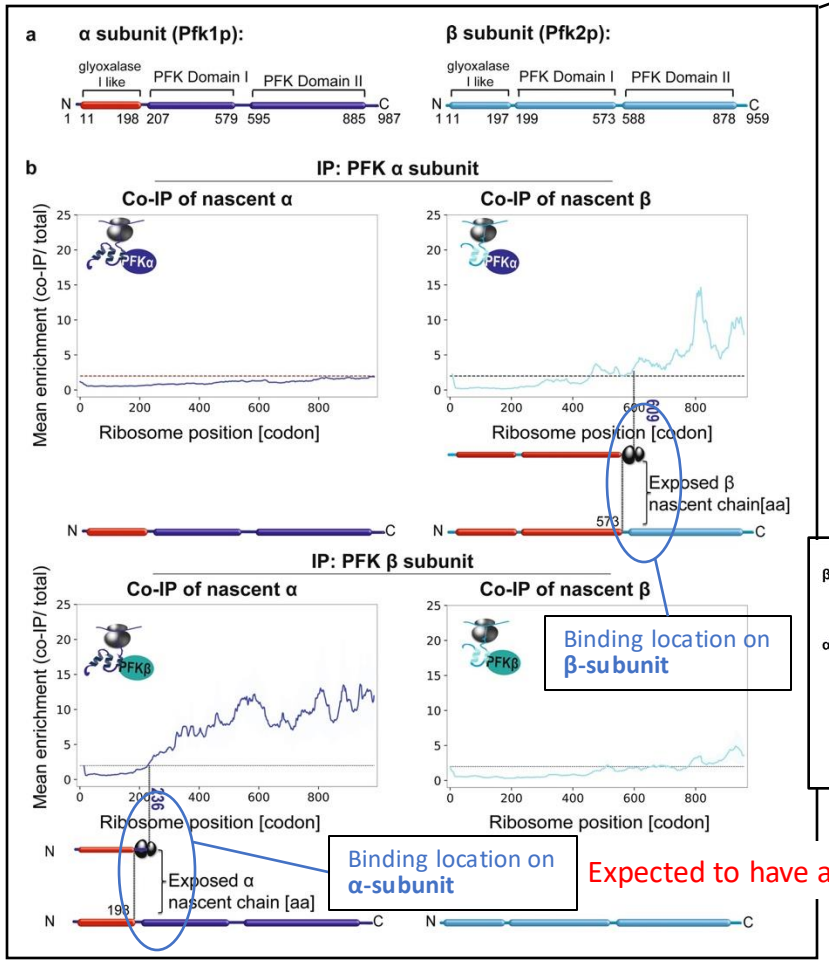
--- less true for (PFK, PDB: 3O8O), α subunit



Shiber et al. (2018) --- PMID: 30158700

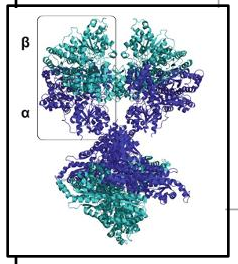
Conclusion 2

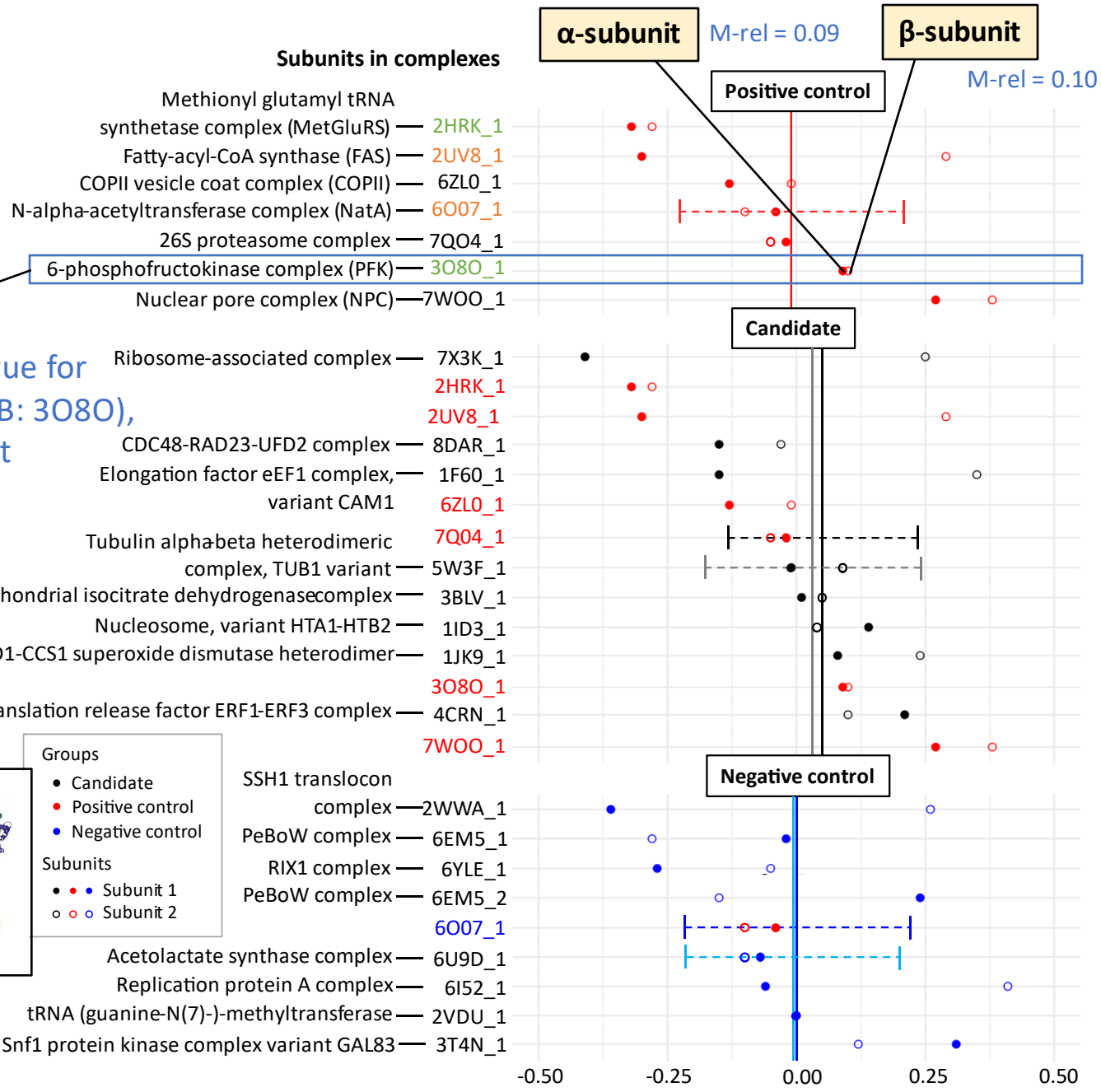
Subunit **average M-rel** can help to suggest
the potential *trans* Co-TA mechanism for complexes
--- true for (MetGluRS, PDB: 2HRK), (NatA, PDB: 6O07)



Expected to have a -ve M-rel

--- less true for (PFK, PDB: 3O8O), α subunit





Shiber et al. (2018) --- PMID: 30158700

Extension

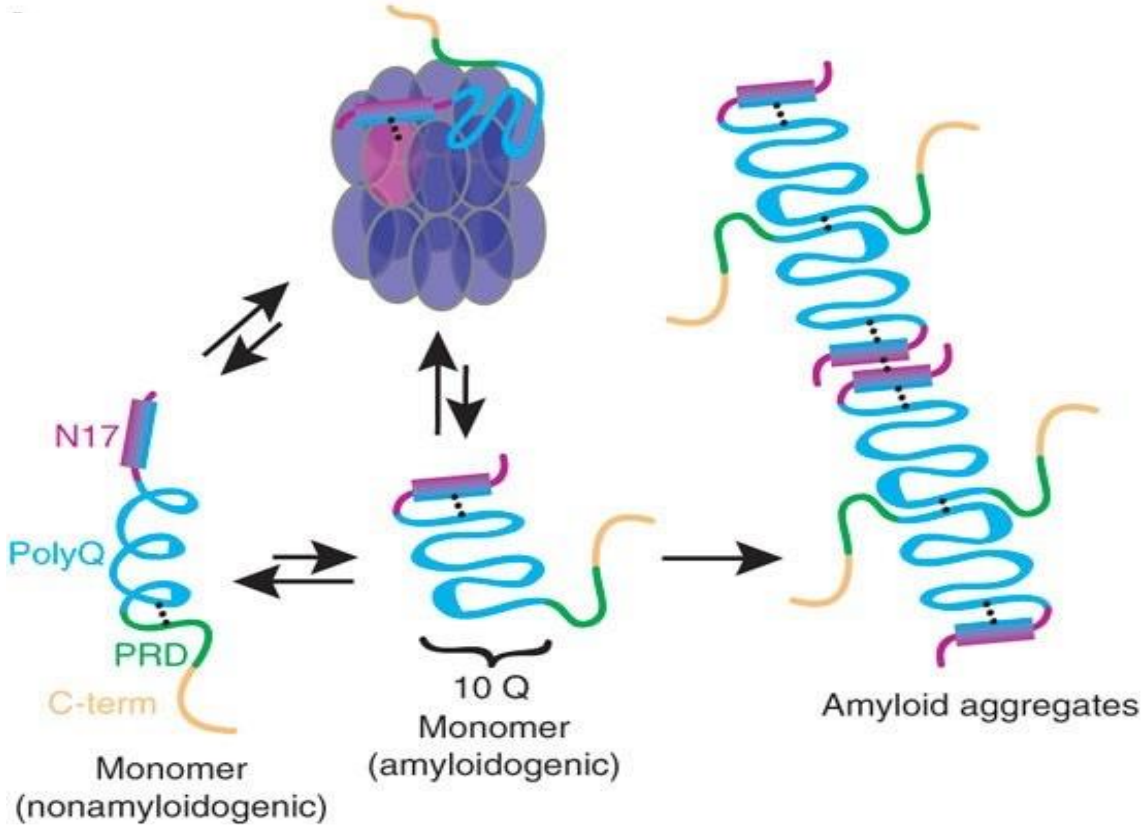
Understanding the mechanism of Co-TA could help to tackle diseases

Case 1:

Huntington diseases

The short N-terminal domain of the Huntington protein

- Protein oligomerization
- Subsequent significant acceleration of amyloid formation



Tam et al. (2009) --- PMID: 19915590

Extension

Understanding the mechanism of Co-TA could help to tackle diseases

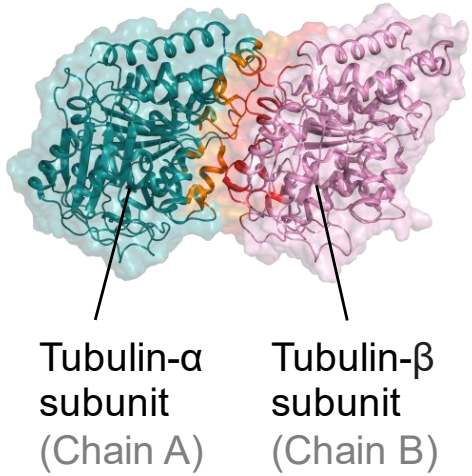
Case 2:

Pathogenic missense mutations

α-tubulin gene:
- *TUBA1A*

β-tubulin genes:
- *TUBB4A*
- *TUBB2A*
- *TUBB3*

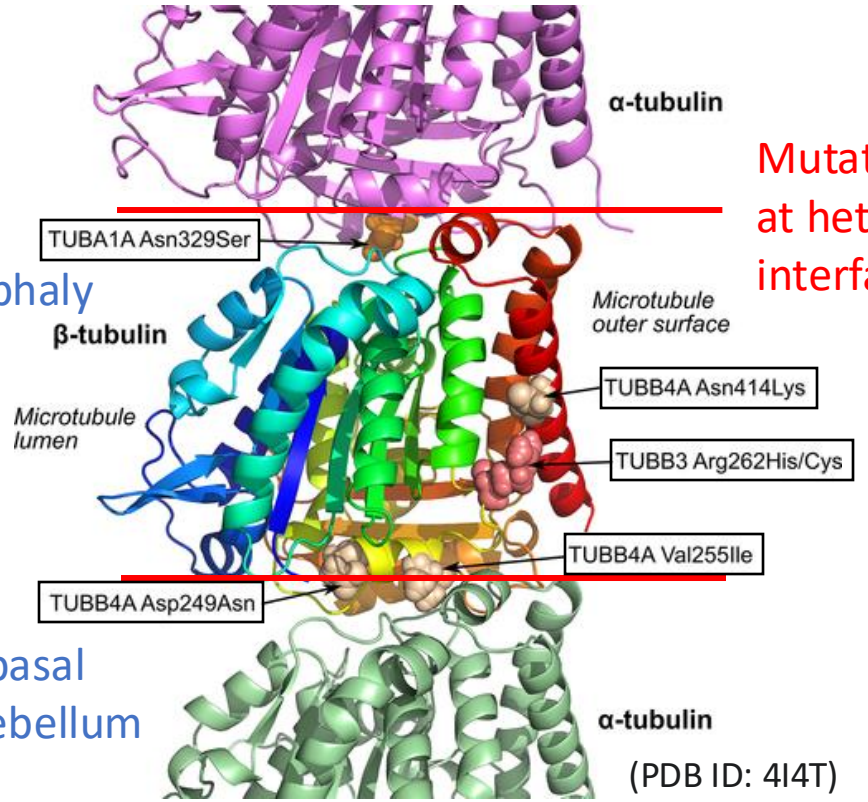
→ Brain malformations



Tubulin-α subunit (Chain A)

Tubulin-β subunit (Chain B)

Severe lissencephaly with cerebellar hypoplasia



Mutations enrich at heteromeric interfaces

Atrophy of the basal ganglia and cerebellum (H-ABC)

(PDB ID: 4I4T)

Bergendahl et al. (2019) --- PMID: 31219644

Thank you for listening!

Acknowledgement

I want to thank my supervisor, Professor Simon Hubbard for his kind guidance and advice throughout the project.

I would also like to thank Dr Robert Crawford for his generosity in providing source data and his insight and support throughout this project.

Faculty of Life Science, The University of Manchester, United Kingdom

Linqing Hu, Simon Hubbard's Lab

Contact: liniqnghu120120@outlook.com