Co-translational protein complex assembly: integrating structural data and quantitative mass spectrometry data to identify candidates

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

Linqing Hu, Simon Hubbard's Lab linqinghu120120@outlook.com

The University of Manchester

Background

- Multimeric protein complexes are important for cellular functions across all domains. Their assembly are mediated by groups of physically associated proteins with precisely regulated stoichiometry [1].
- Since the conventional thought of the protein complexes formation through diffusion and random collision of subunits within the cytoplasm can not explain how unassembled subunits avoid (i) non-specific interactions (ii) aggregation (iii) quality control sequestration to proteases and chaperones and (iv) navigate crowded and occluded cellular environments [2]. The process of **co-translational assembly (Co-TA**), which has the assembly start while at least one of the component subunits is still undergoing ribosome translation, was suggested, and found to happen prevalently for both the homomeric and heteromeric complexes [3].
- Previous studies have raised **three factors that are important to Co-TA**:
  (1) the Co-TA processing orders: **simultaneous (Co-co)** vs **sequential (Co-post)**
  (2) the ribosome translation locations: **on the same mRNA (*cis*)** vs **on different mRNAs (*trans*)**
  (3) the subunits assembly orders: **directional** vs **symmetrical**
  Overall, **6** possible modes of the Co-TA mechanism were given (Figure 1).
  They were found favoured by the homomeric and the heteromeric complexes differently.

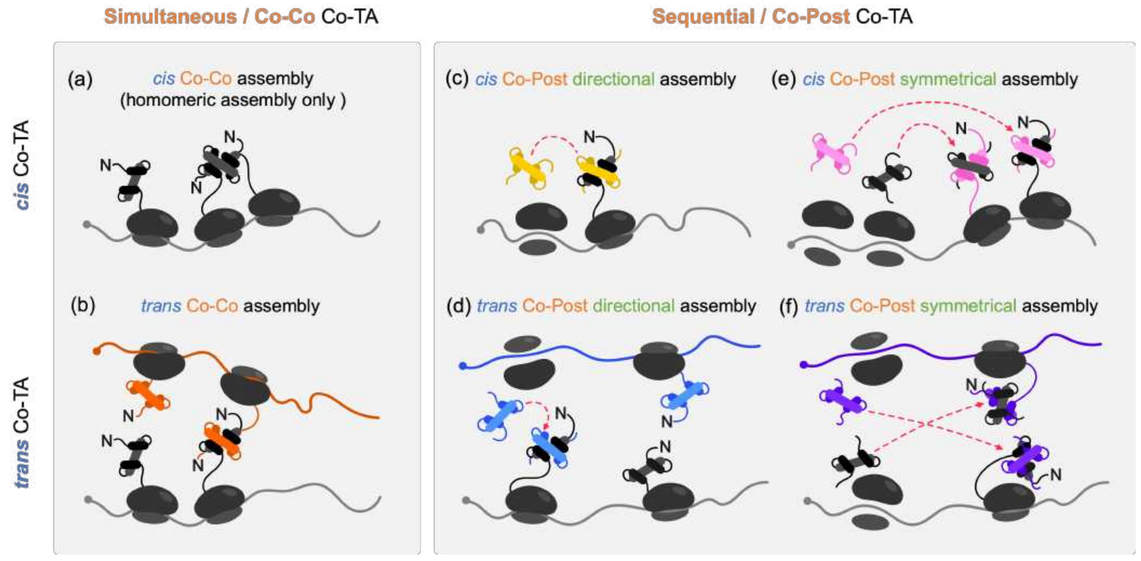


Figure 1. All possible co-translational assembly (Co-TA) mechanisms for both the homomeric and the heteromeric complexes.

- In addition to the factors mentioned above, **the localisation of interface buried residues --- towards the N(amino)-terminus or the C(carboxy)-terminus** of a subunit seems also important for both types of complexes to employ, although different, Co-TA mechanisms. However, biological evidence, especially for heteromeric complexes, lacks evidence to support this idea.

Hypothesis and Aims

- Similar to the hypothesised that the type of co-translational interactions may be distinguishable by **the areas of the interfaces** involved in the protein complex from one previous study [4]. This study focuses on **heteromeric complexes** and hypothesises that 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.
- This study raised the concept of subunit **relative Moment (M-rel)** to verify the hypothesis, aimed to develop a computational method first to determine the buried residues in each subunit of a two-subunit complex, then calculate the M-rel of each of these subunits to verify the hypothesis.

Methods

For comparison, three lists of yeast complexes were given by Dr Robert Crawford from Simon Hubbard's lab:

(1) The candidates: containing subunits with evidence from a novel experiment method (Figure 2) that indicated the involvement in complex Co-TA. If a subunit (nascent chain) in the elution sample was found to have a partner subunit (nascent chain or complete protein) of a complex in either the elution sample or the release sample. The respective complex that the two subunits belong to would be treated as a candidate.

The origin of the partner subunit gives evidence for the status of the partner subunit and the mode of Co-TA.
- If the partner subunit is a complete protein --- it will not appear in the elution sample, but in the release sample, to be involved in **sequential / Co-Post Co-TA**.
- If the partner subunit is a nascent chain, it should only appear in the elution sample, although it might also appear in the release sample in reality, due to the limited (about 25%) labelling efficiency of OPP to the ribosome synthesising ends of the nascent proteins. But no matter what, the partner subunit undergoes **simultaneous / Co-Co Co-TA.**

(2) The positive control: contains protein complexes that were known to be involved in the co-translational assembly from previous literature. This list was expected to be used as a standard to confirm whether a candidate undergoes Co-TA.

(3) The negative control: containing protein complexes with their subunits only found in the input but not in the elution and release samples (which did not have evidence for Co-TA) were given. The composition of the initial candidate list and the initial negative control list were mutually exclusive. This list was used to confirm that subunits of the complex can be detected by label-free MS.

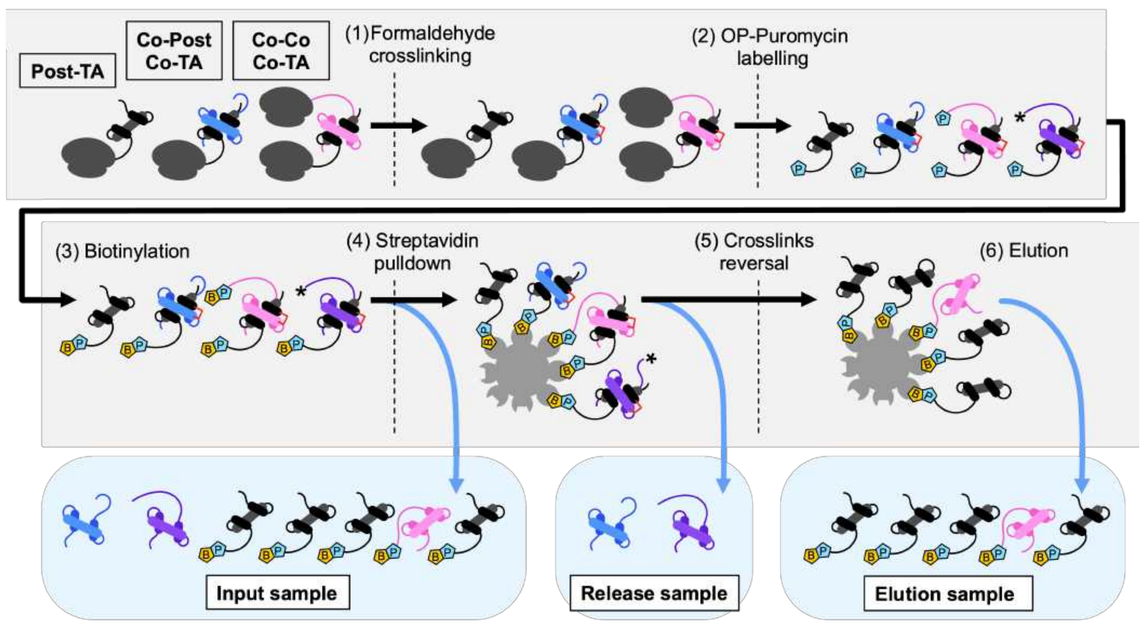
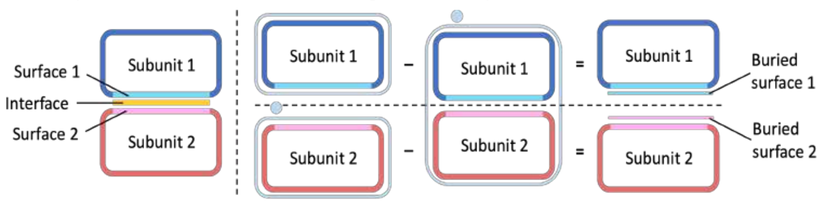


Figure 2. The workflow to catch subunits of protein complexes involved in co-translational assembly.
Three types of protein: (1) the one that undergoes **post-translational assembly (Post-TA)**, (2) the one that undergoes **directional Co-translational assembly (Co-Post Co-TA)**, and (3) the one that undergoes simultaneous **Co-translational assembly (Co-Co Co-TA)** in **Yeast S2883 extract** were input and run through the 6 steps.

- **The absolute solvent accessibility (ASA)** for the individual subunits in solvent status **× 2** and the two-subunit complex are calculated as shown in Figure 3 and Equation 1a, 1b.



Equation 1

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

Figure 3. A schematic diagram shows the calculation of ASA for buried residues in the two individual subunits in a two-subunit complex
The interface (yellow) and the buried surfaces (light blue and light pink) of the two subunits (dark blue and dark pink) are labelled in a two-subunit protein complex. The buried surfaces are determined by subtracting the 'rolling ball' crossed surfaces (pale blue) of chains in the complex from the surfaces of chains in individual proteins.

- **The relative Moment (M-rel)** reflects the overall contribution of **the relative location** and **the ASA in Å²** of the buried residues to the buried surface of a subunit in the two-subunit complex. It was calculated by equation 3 that was evolved from equation 2.

Equation 2

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

Equation 3

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

- $\boldsymbol{n}$ − 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 the odevity of n
- $\boldsymbol{i}$ − 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

Overall, the working flow is summarised in Figure 4 below.

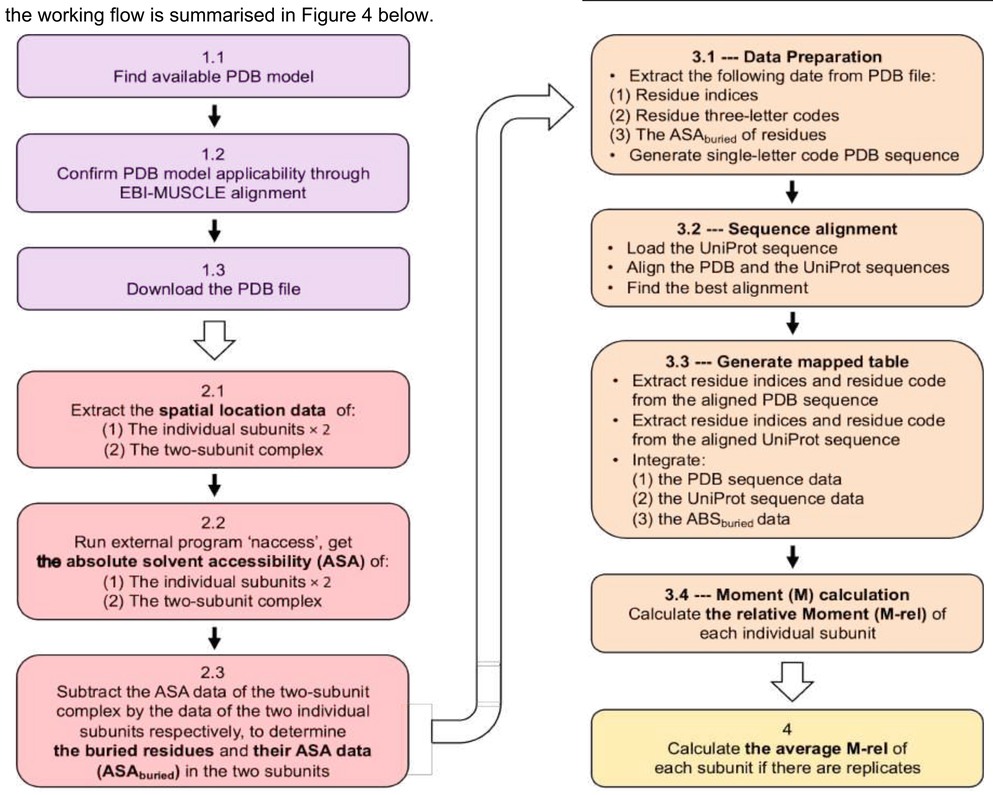
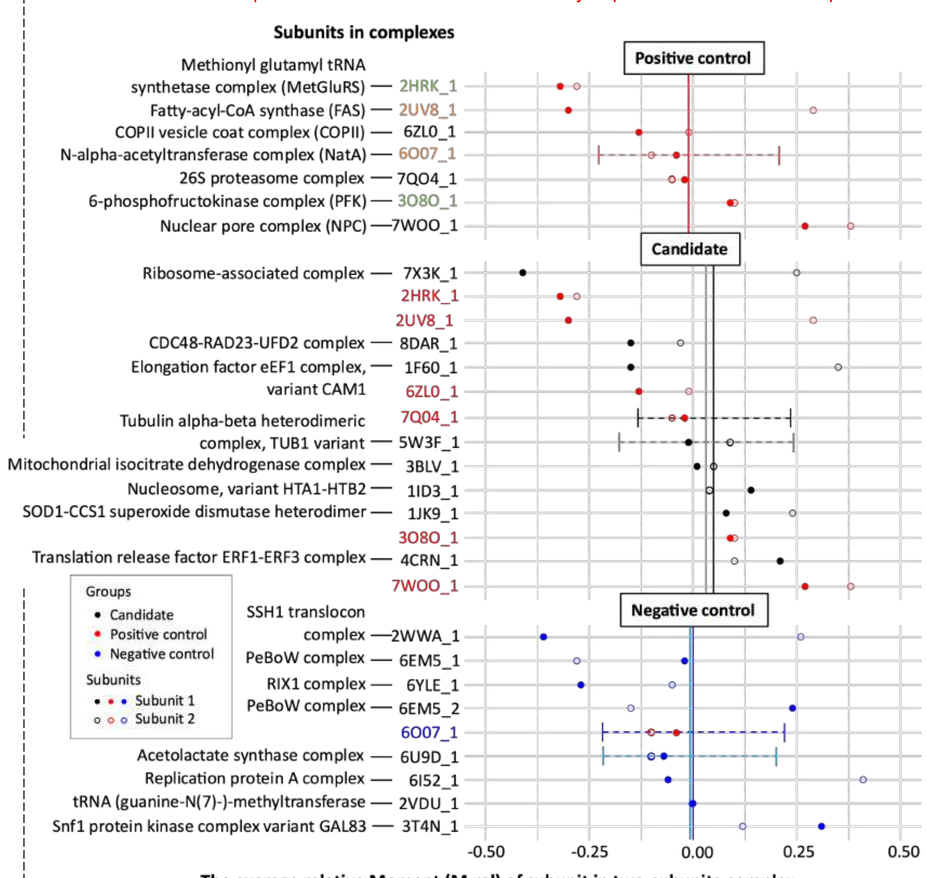


Figure 4. A flowchart shows the experimental process of identifying useable PDB models (purple boxes), calculating the solvent accessibility of subunits (red boxes), and two-subunit complexes, and determining the (average) relative Moment of subunits (orange and yellow boxes)

Results

- Overall, 71 complexes were identified as the candidates to look for PDB models, with 62 of them newly found. All the complexes in the positive control, the complexes consisting of less and equal to the 4 different subunits in the candidate, and the complexes consisting of less and equal to the 4 different subunits in the negative control that have available and applicable PDB complex models are shown in Figure 5 (8 out of 39, 7 out of 14, and 8 out of 29 respectively).
- Comparing the calculated M-rel for subunits in the two-subunit complex in the candidate, the positive control and the negative control three lists, there was no significant difference between them, no matter whether include the sample in the candidate and the negative control which was also found in the positive control (known as 'include') or not (known as 'essential'). This result could be due to the small number of PDB samples available. **AlphaFold** and **Deep Learning** can be adopted to increase sample numbers for calculation in the next step.
- However, the location of the subunit M-rel still has the potential to give hints for suggesting the potential *trans* Co-TA mechanism for complexes, since the locations of M-rel consist of the binding sites previously identified for all the subunits in Fatty-acyl-CoA synthase (2UV8), N-alpha-acetyltransferase complex (6O07), Methionyl glutamyl tRNA synthetase complex (2HRK), and the β subunit of 6-phosphofructokinase complex (3O8O) [5].
- Even though, the observed subunits' average M-rel do not satisfy all proposals for using them to predict the Co-TA mechanisms, e.g. if suggesting that a complex that has one subunit with an extremely big M-rel undergoes sequential *trans* Co-TA, as an inconsistent trend is found between Fatty-acyl-CoA synthase (2UV8) and N-alpha-acetyltransferase complex (6O07), as N-alpha-acetyltransferase complex does not have a subunit that has the assembly domain very towards the C-terminus. This could mean that sequential *trans* Co-TA may tend to happen in complexes that have a subunit with the assembly domain very towards the C-terminus, but having a subunit with the assembly domain very towards the C-terminus is not a necessary condition for sequential *trans* Co-TA. So, whether the average M-rel can be used to predict the Co-TA mechanisms entirely depends on how the assumptions are made.



Subunit 1	ave M-rel	Subunit 2	ave M-rel
GUS1	-0.32	ARC1	-0.28
FAS1	-0.30	FAS2	0.29
SEC31	-0.13	SEC13	-0.01
ARD1	-0.04	NAT1	-0.10
RPT1	-0.02	RPT2	-0.05
PFK1	0.09	PFK2	0.10
NUP57	0.27	NSP1	0.38
ZUO1	-0.41	SSZ1	0.25
CDC48	-0.15	NPL4	-0.01
EF1A	-0.15	EF1B	0.35
TUB1	-0.01	TUB2	0.09
IDH1	0.01	IDH2	0.05
HHT1	0.14	HHF1	0.04
SOD1	0.08	CCS1	0.24
SUP35	0.21	SUP45	0.10
SSH1	-0.36	SBH2	0.26
ERB1	-0.02	NOP7	-0.28
IPI3	-0.27	RIX1	-0.05
ERB1	0.24	YTM1	-0.15
ILV2	-0.07	ILV6	-0.10
RFA2	-0.06	RFA1	0.41
TRM82	0.00	TRM8	0.00
SNF1	0.31	SNF4	0.12

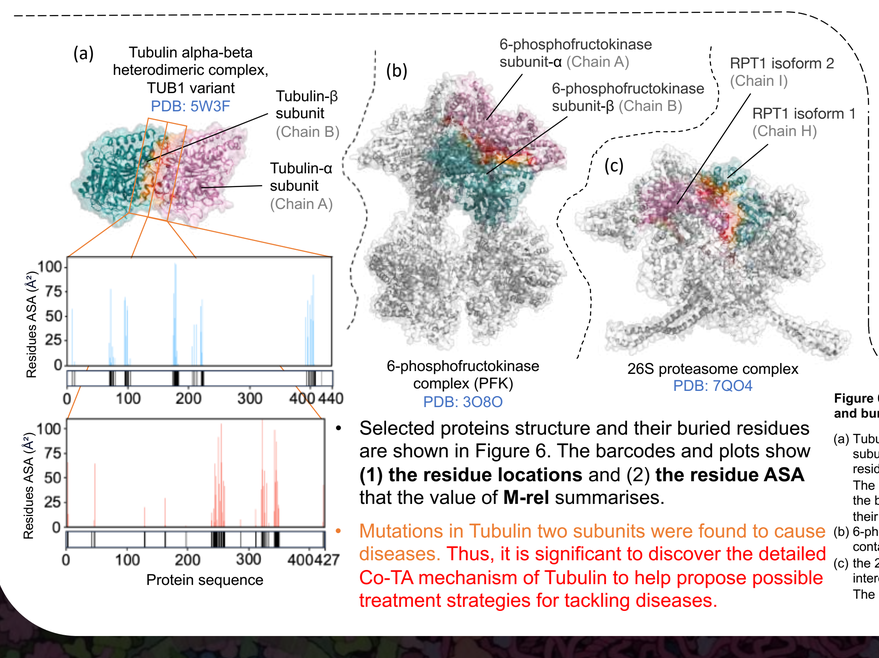
Figure 5. The distribution of the average relative Moment (M-rel) of subunits in two-subunits-complexes

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)

- **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



- Selected proteins structure and their buried residues are shown in Figure 6. The barcodes and plots show **(1) the residue locations** and (2) **the residue ASA** that the value of **M-rel** summarises.
- Mutations in Tubulin two subunits were found to cause diseases. Thus, it is significant to discover the detailed Co-TA mechanism of Tubulin to help propose possible treatment strategies for tackling diseases.

Figure 6. A visualisation of protein complex structures and buried residues

(a) Tubulin heteromeric complex that contains only two subunits (subunit 1: blue, subunit 2: pink). The buried residues are highlighted in orange and red respectively. The binding sites are boxed in orange. The barcodes of the buried residues on the scale of protein sequence and their ASA are shown.
(b) 6-phosphofructokinase heteromeric complex that contains 4 copies of the same α-β two-subunit complex.
(c) the 26S proteosome heteromeric complex contains the interested two subunits and the non-interested subunits. The non-interested subunits are coloured in white.

References

[1] Kamenova et al. (2019) --- PMID: 30988355
[2] Shieh et al. (2015) --- PMID: 26405228
[3] Bernardini and Tora (2023) --- PMID: 38061625
[4] Badonyi and Marsh (2022) --- PMID: 35899946
[5] Shiber et al. (2018) --- PMID: 30158700