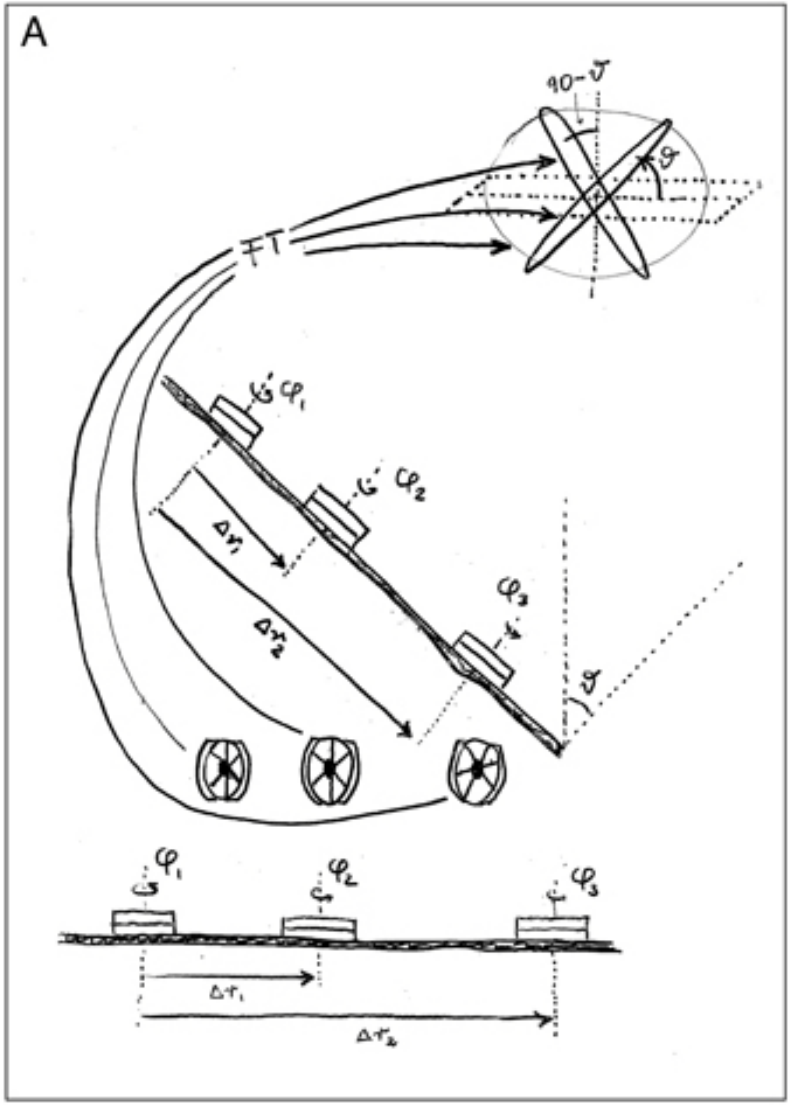
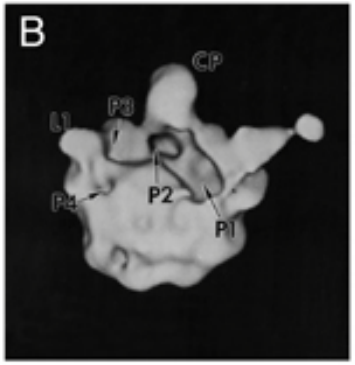
Classification in Real Life (Precise Title to be Announced)

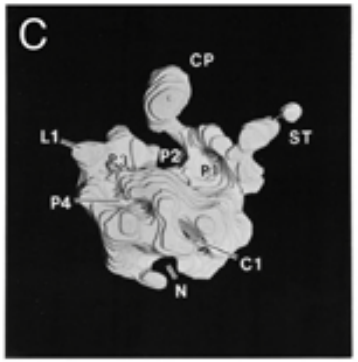
Joachim Frank

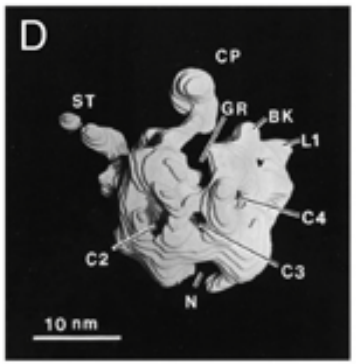
Howard Hughes Medical Institute
Department of Biochemistry and Molecular Biophysics
and Department of Biological Sciences
Columbia University

RANDOM-CONICAL RECONSTRUCTION ~ 30 Years old





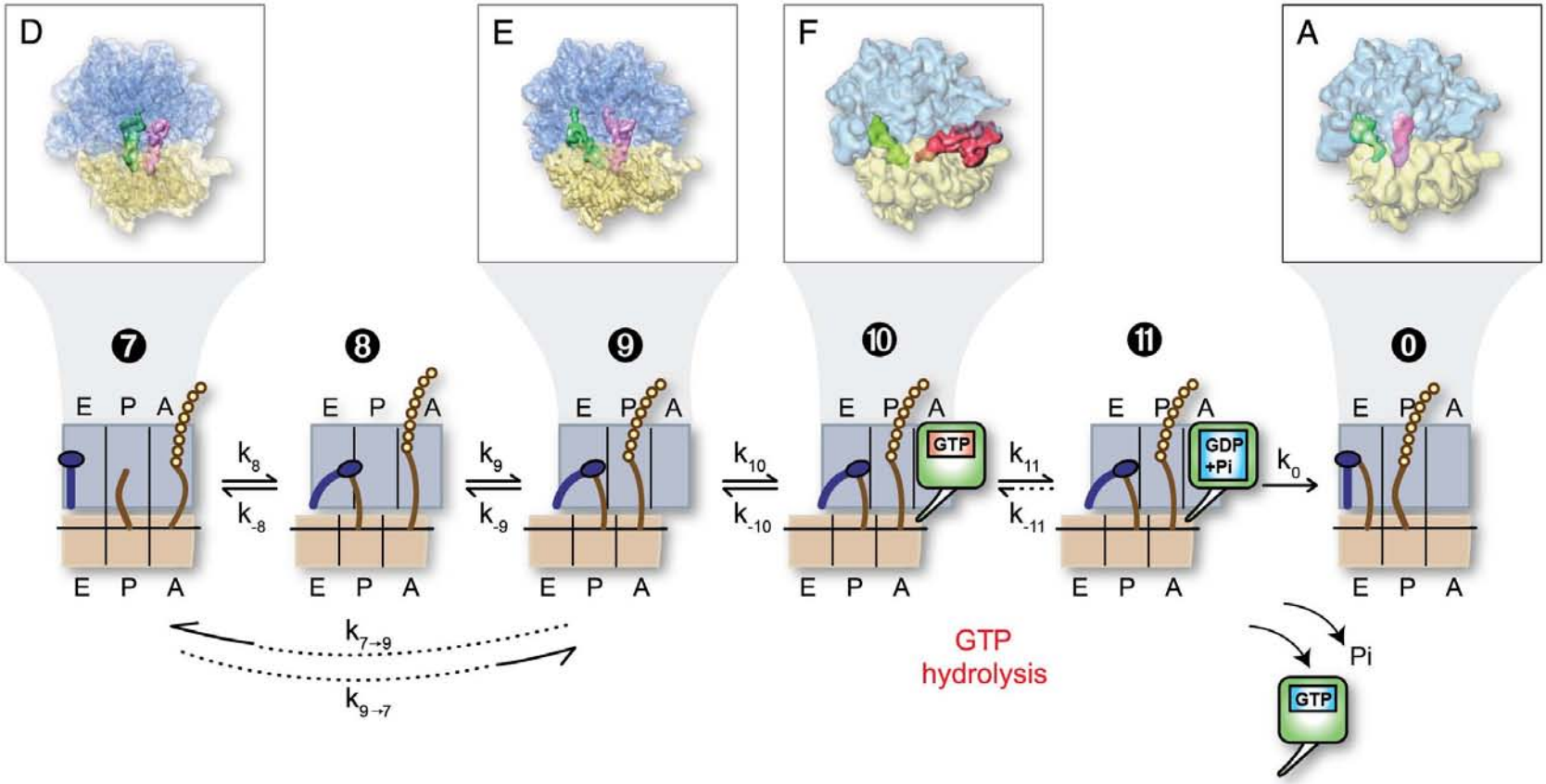




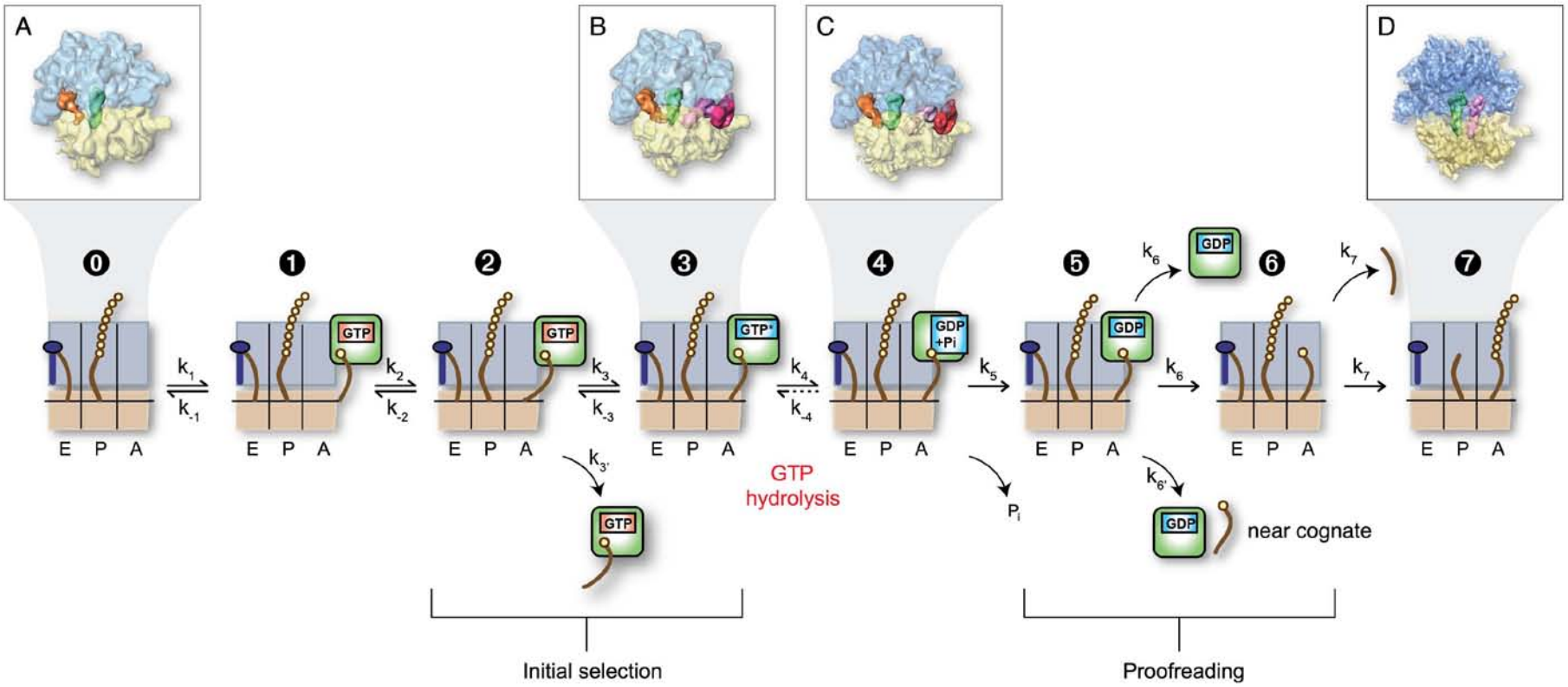
Overhead 1979

Radermacher et al. 1987

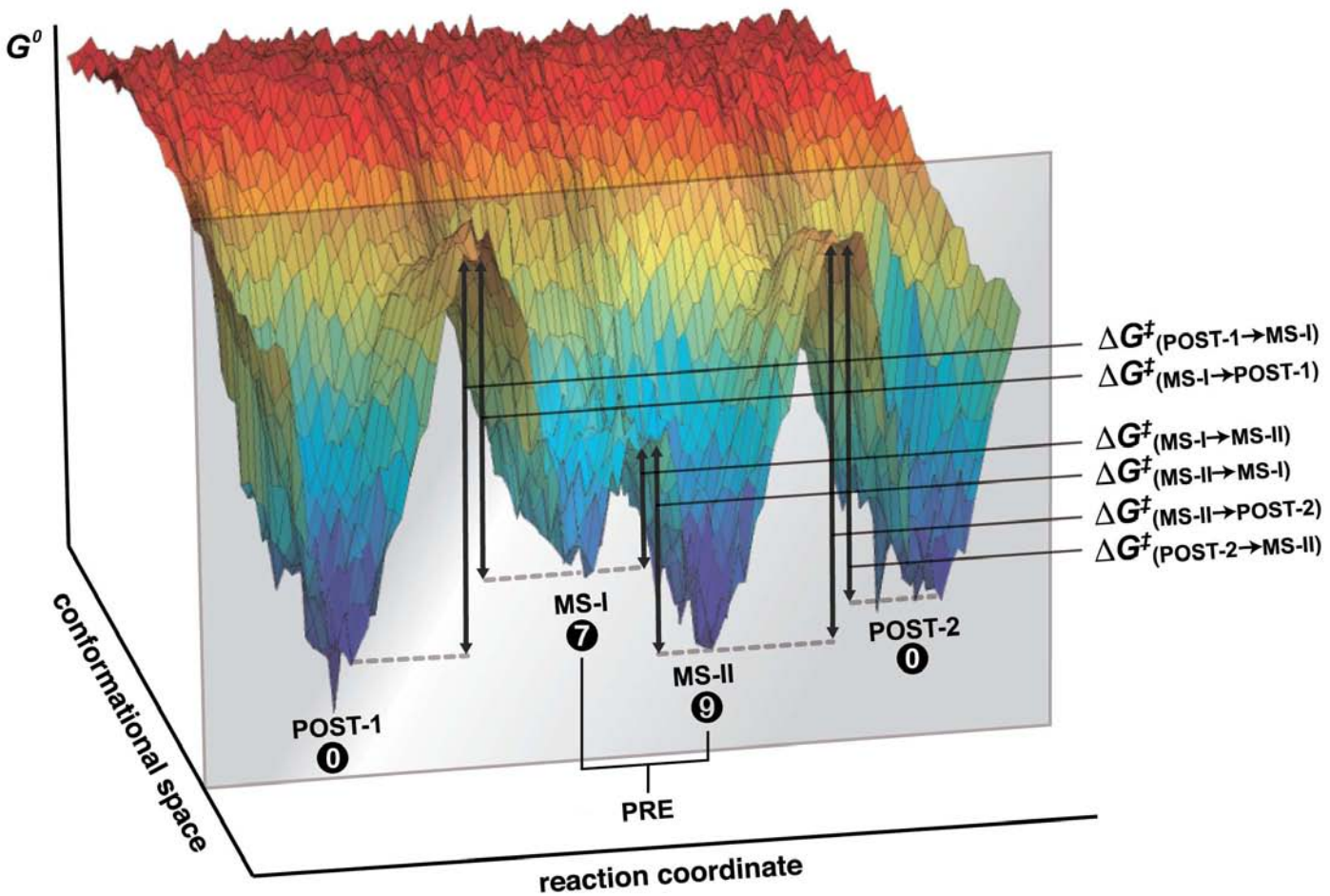
J. Frank, Quart. Rev. Biophys., in press



Translocation


A
B
C
D
0
1
2
3
4
5
6
7
GTP
GTP
GTP
GTP*
GDP +Pi
GDP
GDP
GDP
E P A
E P A
E P A
E P A
E P A
E P A
E P A
E P A
k1
k-1
k2
k-2
k3
k-3
k4
k-4
k5
k6
k7
k3'
k6'
GTP hydrolysis
Pi
near cognate
Initial selection
Proofreading

Decoding


G^0
conformational space
reaction coordinate
POST-1
0
MS-I
7
MS-II
9
POST-2
0
PRE
$\Delta G^{\ddagger}_{(POST\text{-}1 \rightarrow MS\text{-}I)}$
$\Delta G^{\ddagger}_{(MS\text{-}I \rightarrow POST\text{-}1)}$
$\Delta G^{\ddagger}_{(MS\text{-}I \rightarrow MS\text{-}II)}$
$\Delta G^{\ddagger}_{(MS\text{-}II \rightarrow MS\text{-}I)}$
$\Delta G^{\ddagger}_{(MS\text{-}II \rightarrow POST\text{-}2)}$
$\Delta G^{\ddagger}_{(POST\text{-}2 \rightarrow MS\text{-}II)}$

Classification tools

- Supervised (Valle et al., EMBO J. 2002)
- Focused classification (Penczek et al., JSB 2006)
- Hierarchical multi-reference (Schuette et al., EMBO J. 2009)
- Maximum likelihood (Scheres et al., Nat. Methods 2007)
- Bootstrap method (Spahn & Penczek, Cur. Opin. Struct. Biol. 2009; Liao & Frank, in press)

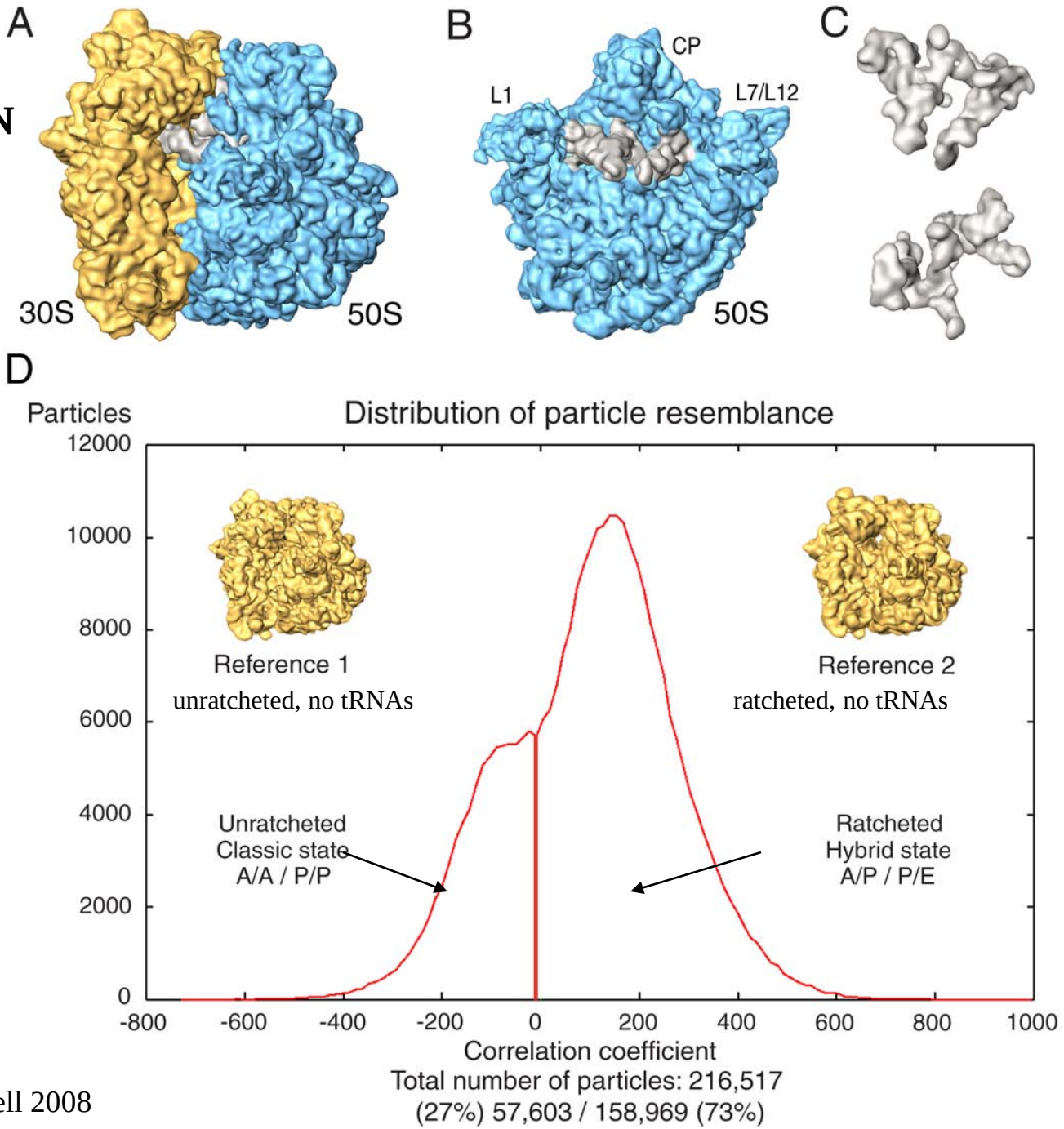
Spontaneous (factor-independent) ratcheting of the ribosome

- Kim et al., Mol. Cell 2007: smFRET studies of pre-translocational ribosome complex show strong Mg^{2+}-dependence of classic → hybrid positions of tRNAs
- 7 mM and above: classical prevails

  3.5 mM: 2/3 are in the hybrid state.

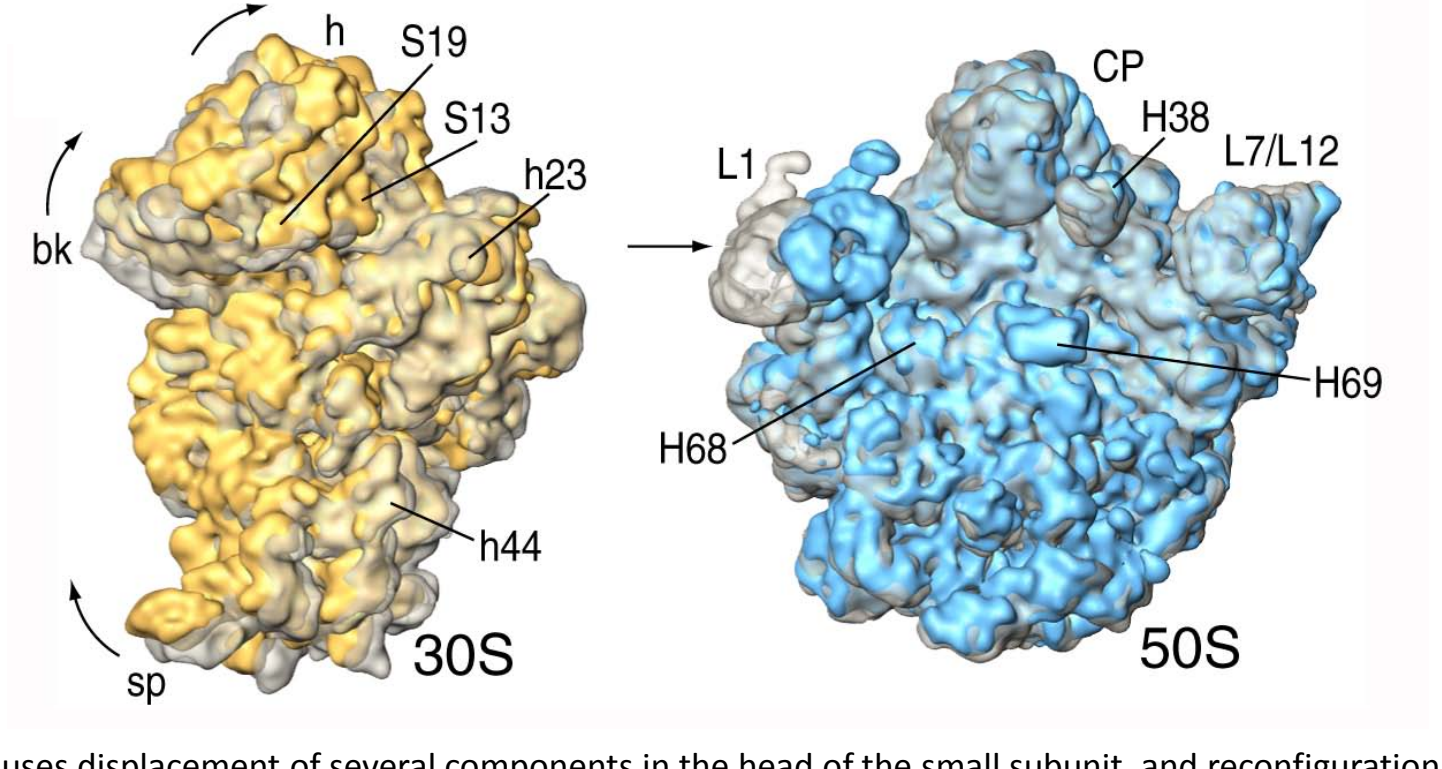
RECONSTRUCTION WITHOUT CLASSIFICATION: tRNAs fused, overlapped

SUPERVISED CLASSIFICATION



Agirrezabala et al., Mol. Cell 2008

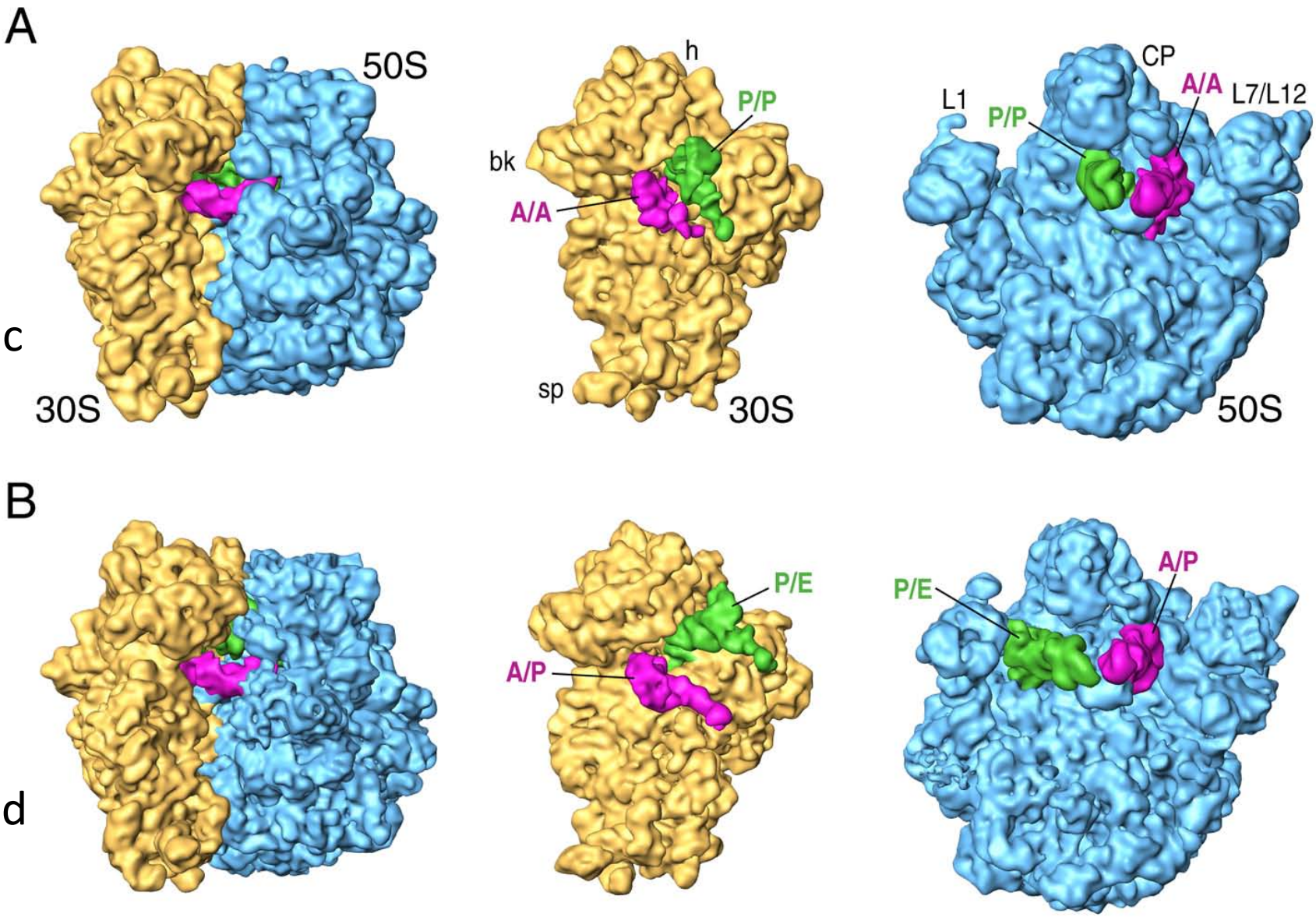
Conformational changes due to spontaneous ratcheting



Rotation causes displacement of several components in the head of the small subunit, and reconfiguration of intersubunit bridges:
Bridge B1b (L5--S13) is remodelled (gliding motion).
Bridge B1a (H38's binding partner S13 is replaced by S19).
Bridge B7a (H68-h23) shifts toward the large subunit.
H38, as well as the central protuberance region where L5 is located, adopt a different conformations.

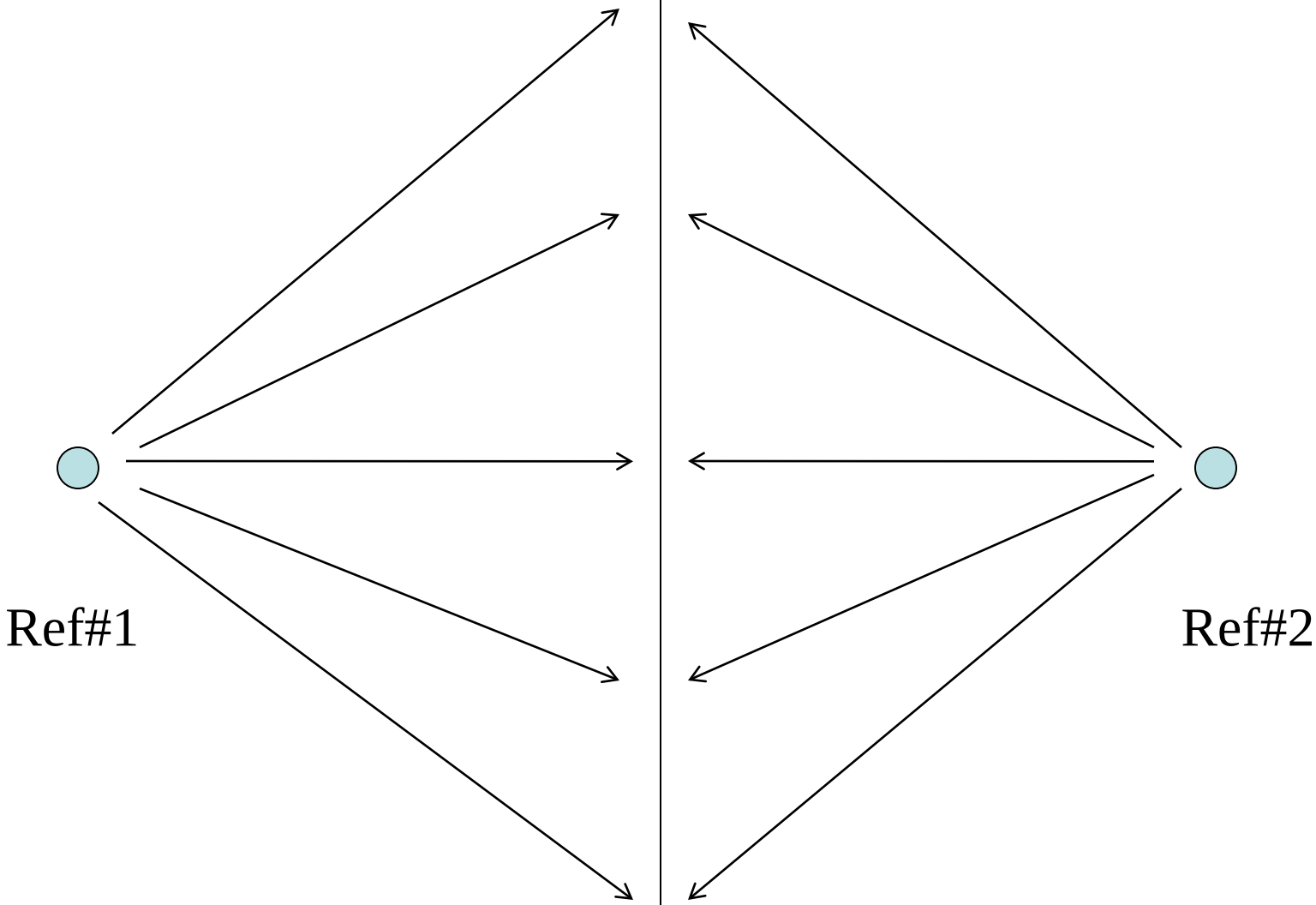
Smaller effects seen in h44, H69.
Large movement of L1 stalk.

Agirrezabala et al., Mol. Cell 2



Agirrezabala et al., Mol. Cell 2008

Neither fish nor fowl


Ref#1
Ref#2

- Richard Henderson:
- Reconstruction is not that much hurt by inclusion of noisy outliers

Xabier Agirrezabala, Jianlin Lei, Rodrigo F. Ortiz-Meoz, Leonardo Trabuco, Klaus Schulten, Rachel Green, and Joachim Frank

Cognate vs. near cognate Trp-tRNA in A/T position, stabilized by kirromycin

Specimen preparation:

Ribosomes programmed with (i) cognate (UGG) or (ii) near-cognate (UGA/stop) codons, loaded with initiation fMet-tRNA^{fMet} in the P site, were incubated with ternary Trp-tRNA^{Trp}•EF-Tu•GTP complexes in the presence of kirromycin.

Cryo-electron microscopy

Data collection with AutoEMation (Lei and Frank, JSB 2005) via 4k x 4k CCD

on FEI 300 kV Polara with effective mag of 100,000 and final pixel size of 1.5Å.

Total # particles: near-cognate -- 359,223 -- heterogeneous

cognate -- 294,671 -- 8.4 Å

initiation-like -- 186,732 -- 8.85 Å

Supervised classification for near-cognate:

Ref 1 – ternary complex removed via soft masking

Ref 2 – ternary complex left in place

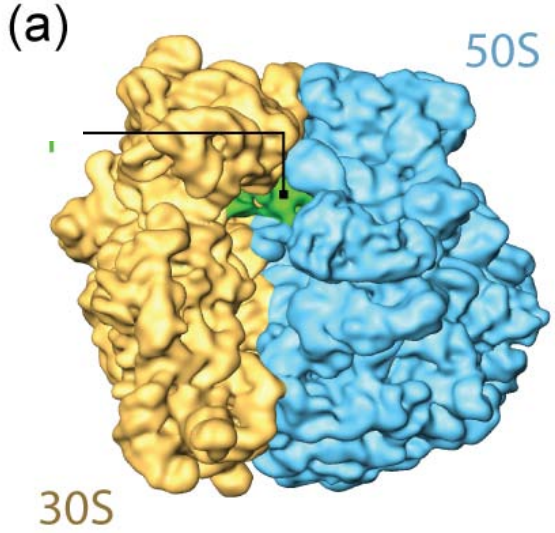
332,410 (=92%) go with Ref 1 8.05 Å

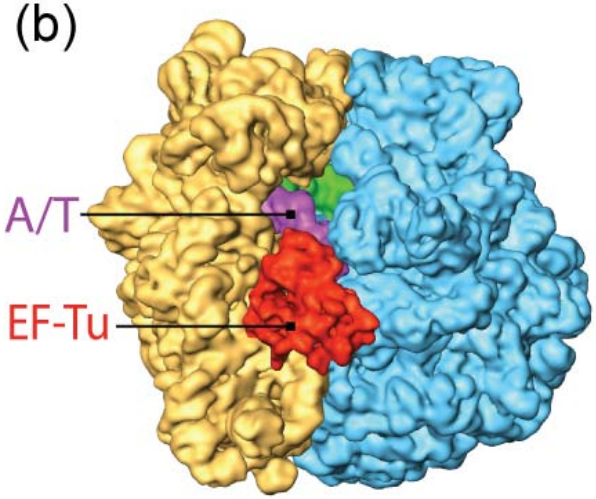
26,873 (=8%) go with Ref 2 13.2 Å

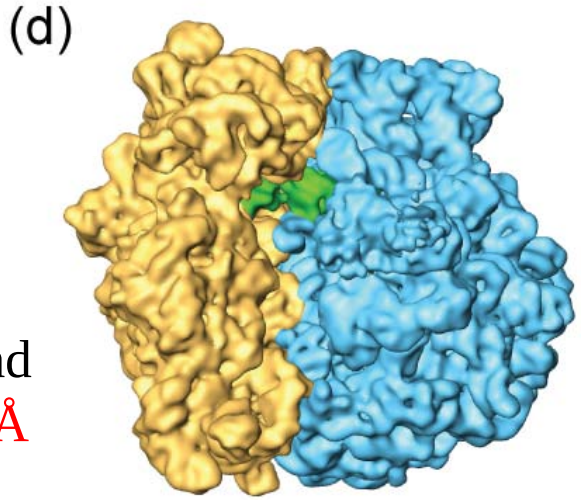
without

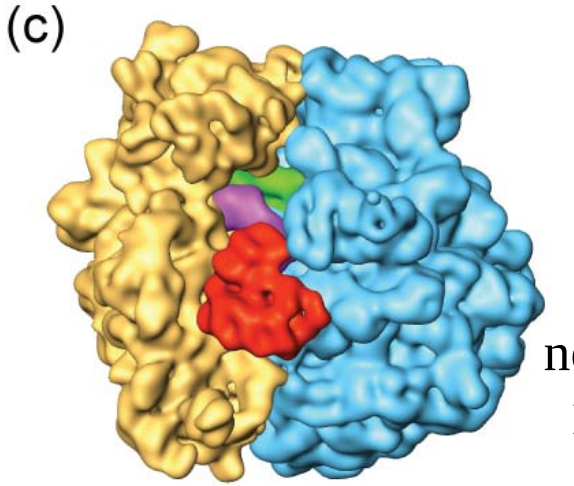
with ternary complex

REFS







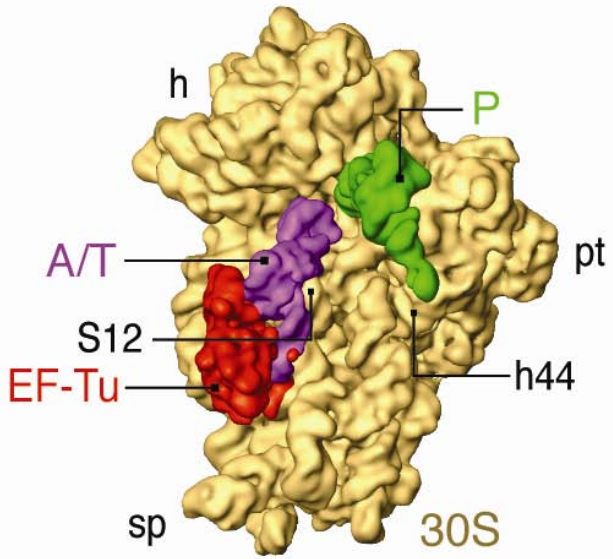


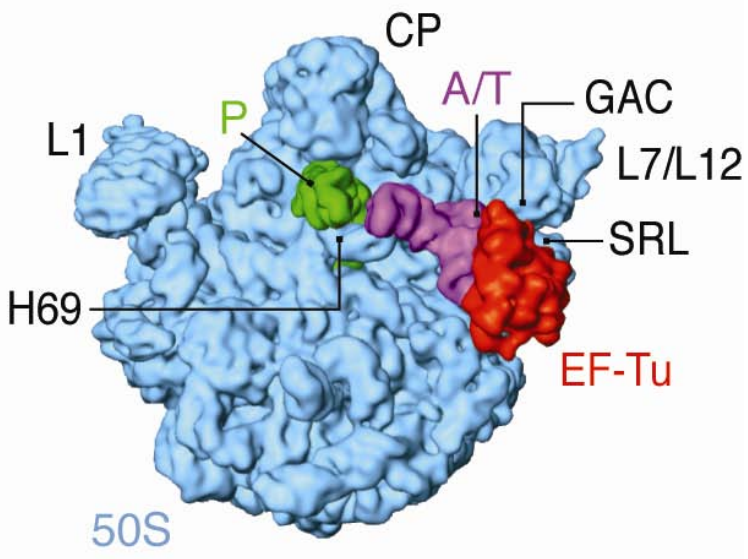
unbound
8.05 Å

near-cognate
13.2 Å

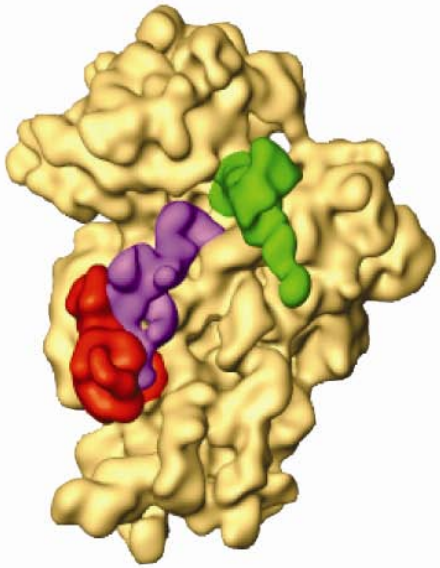
92%

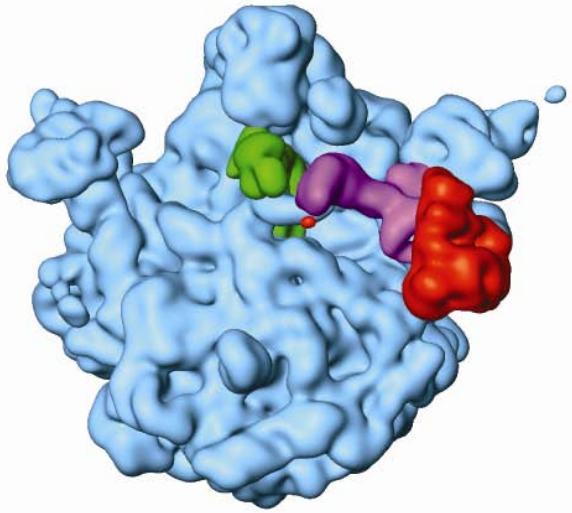
8% of 350,000 images





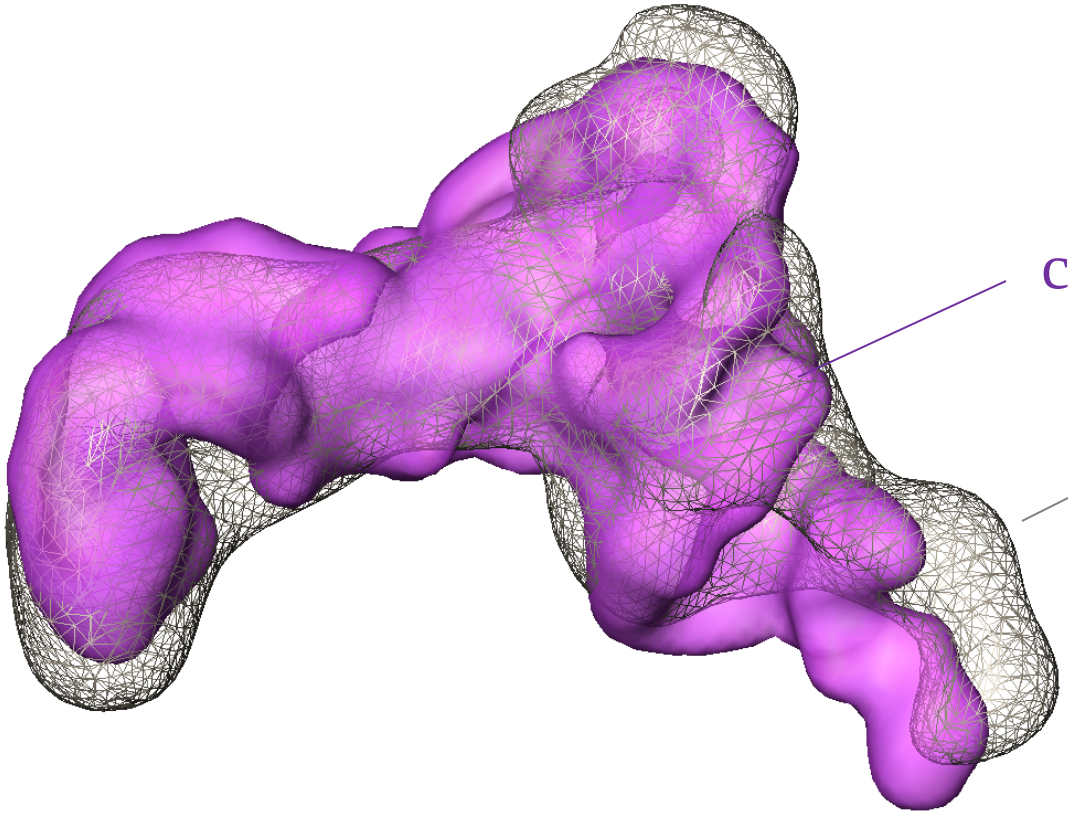
Cognate
8.4 Å





Near-cognate
13.2 Å

Overlay of densities for aa-tRNA



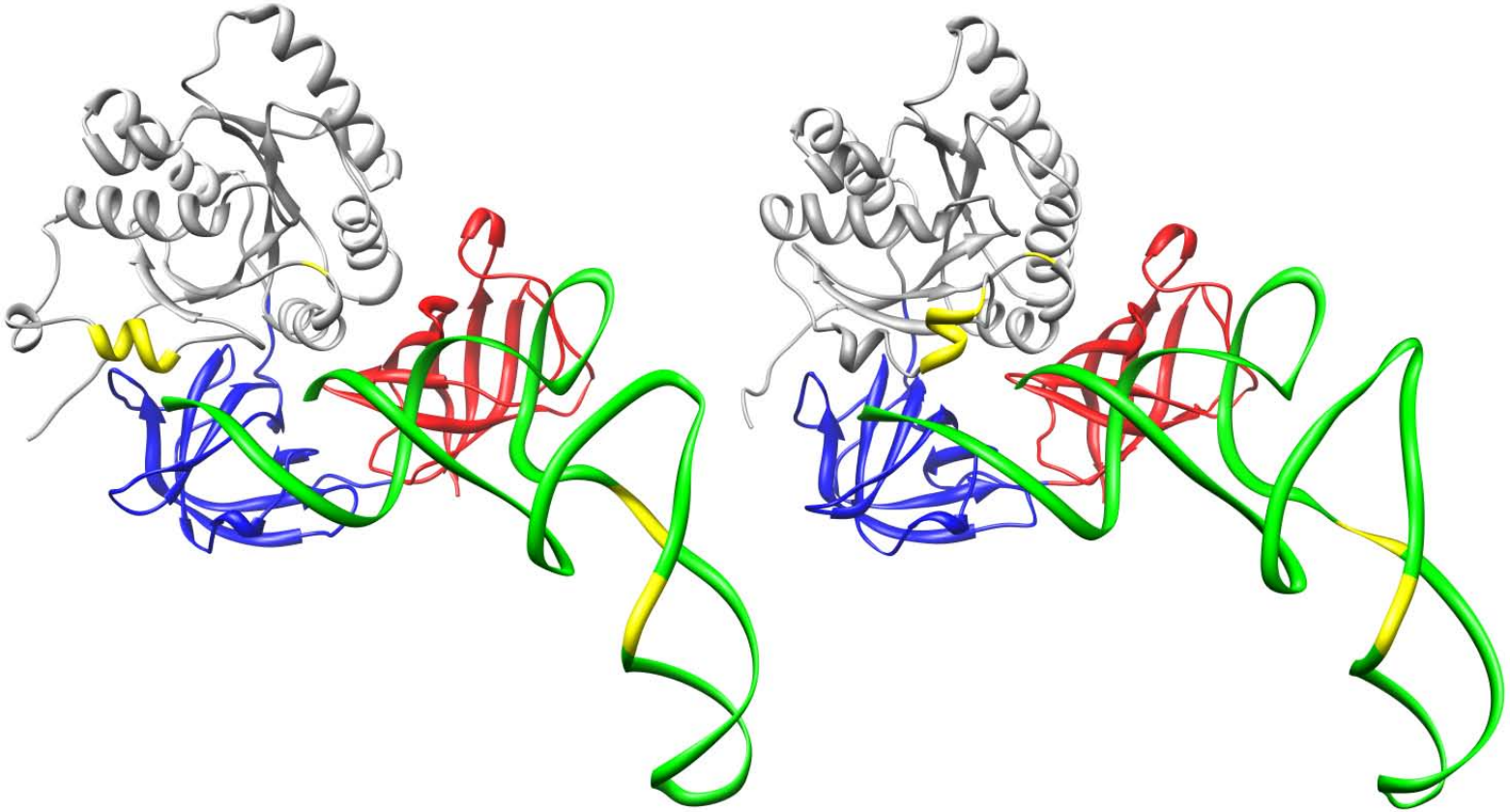
MDFF fitting of observed density for ternary complex (Leonardo Tr

1) Change in anticodon stem loop – kinked, but not as much as in co

2) Change in acceptor arm position on EF-Tu

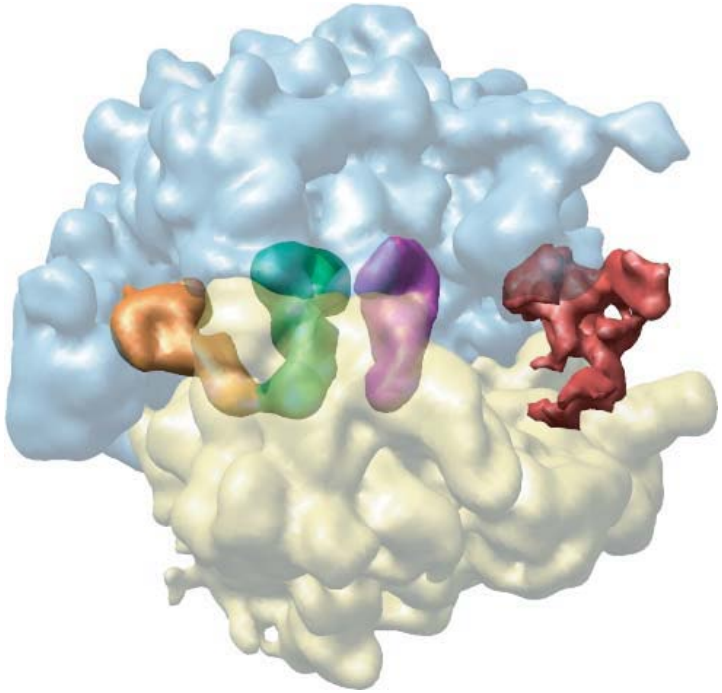
-- OBSERVATIONS (1) and (2) imply difference in conformationa

3) Change in EF-Tu structure (Switch 1)?



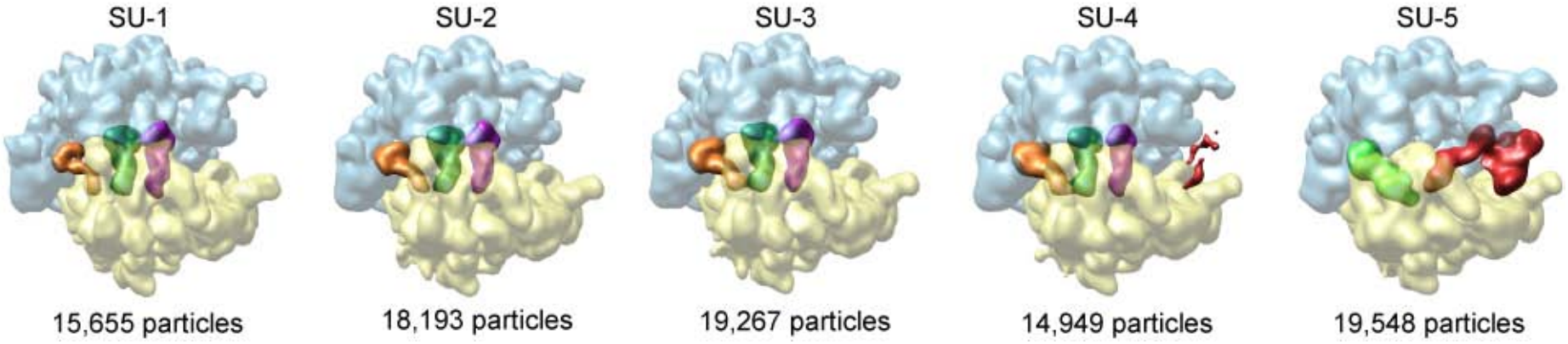
cognate

near-cognate



Reconstruction without classification:
small subunit blurred,
EF-G fragmented

Classes derived by supervised classification (CCF with 2 refs)

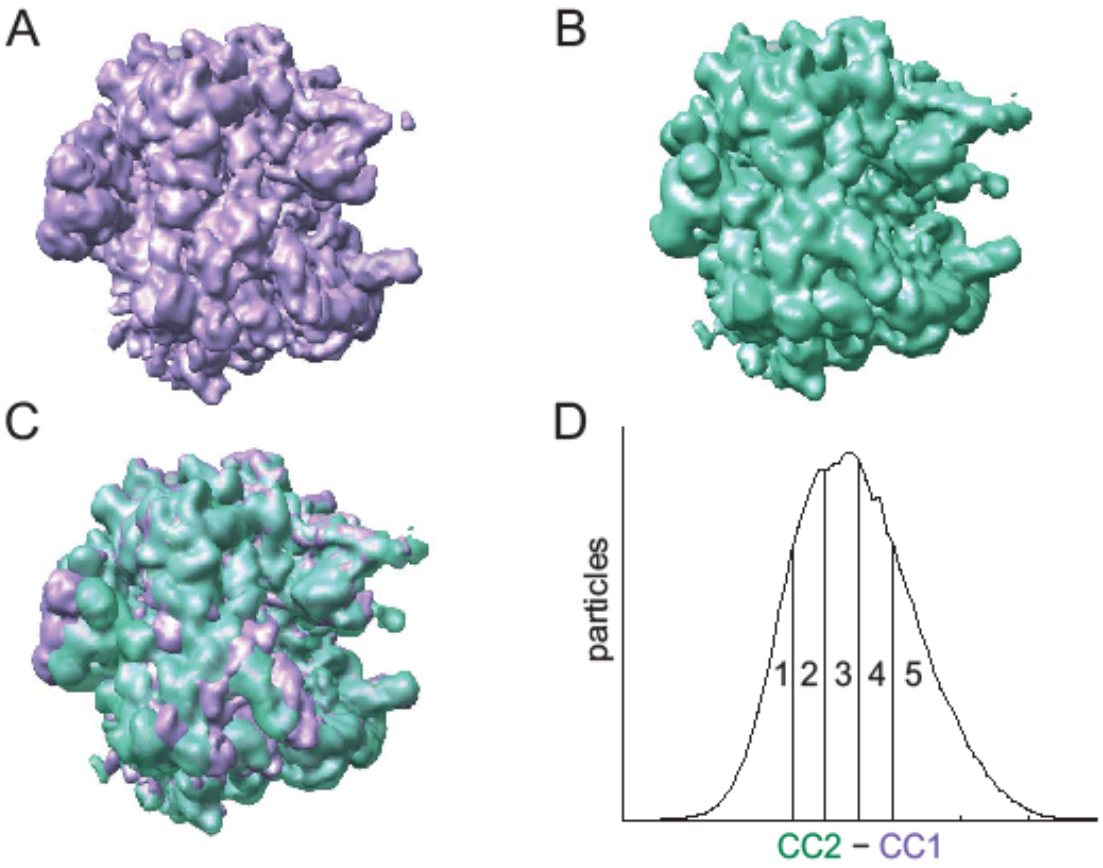


Scheres et al., Nat. Methods 2007

Validation of dual-reference classification:

Equivalent to “R-free”, omit data in reference, and see if they pop up.

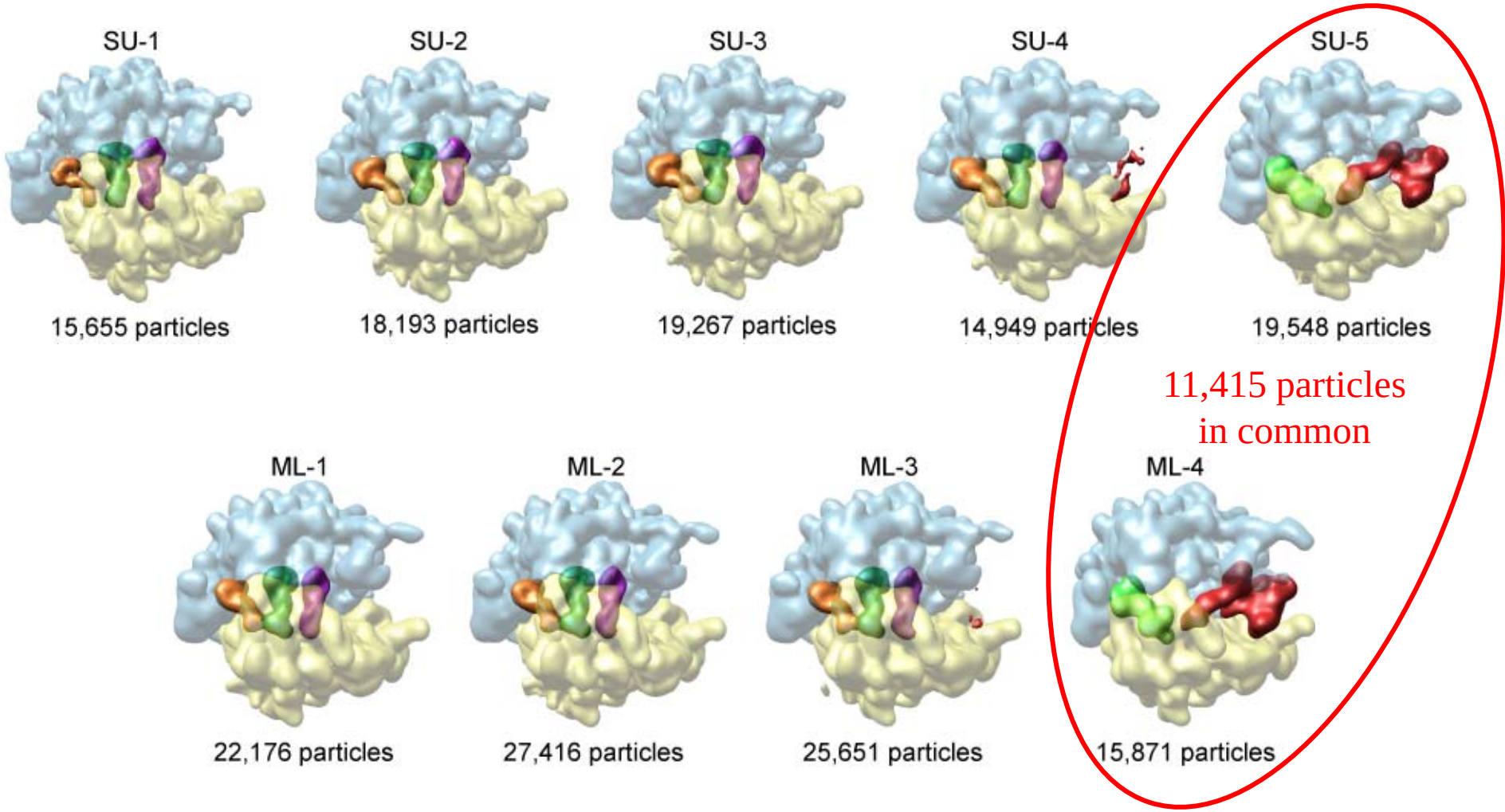
Here: ratcheting and emergence of hybrid positions of tRNA go hand in hand.



Supplementary Figure 3. Supervised classification of the ribosome dataset

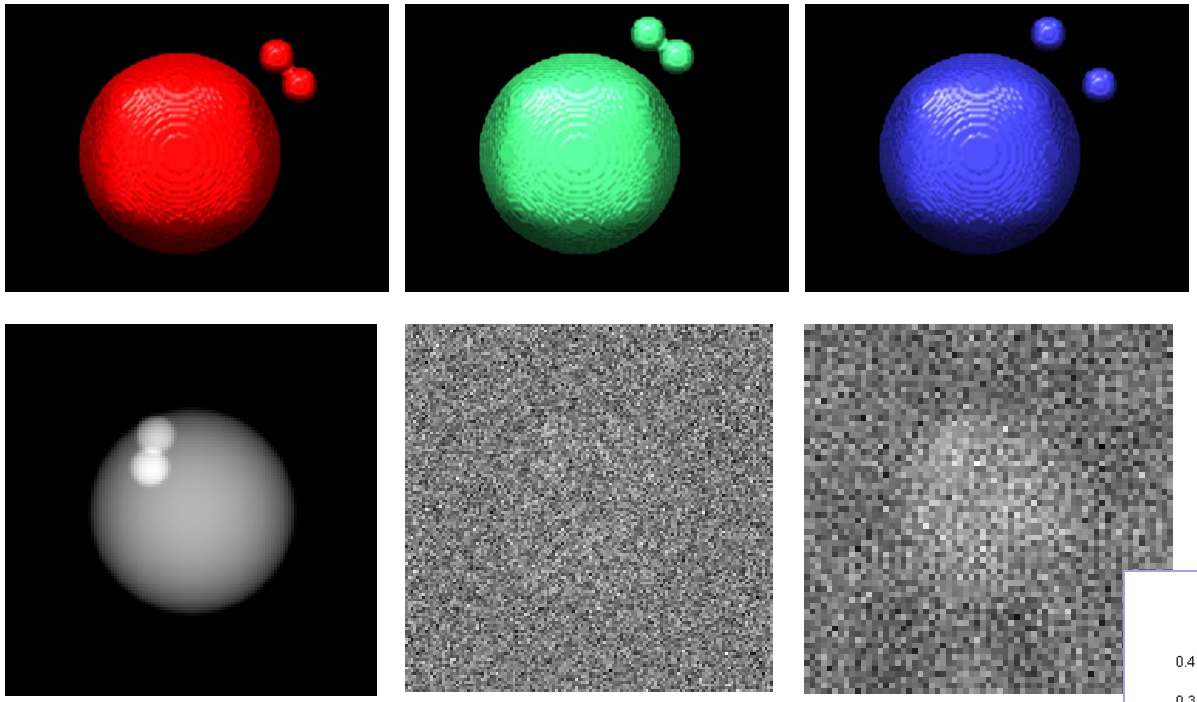
Based on the assumption that the heterogeneity in the data entailed a ratchet motion, we used two reference structures from a previous study (Valle et al. 2003, *Cell* **114**, 123-34) with ribosomes before (**A**) and after (**B**) ratcheting. To avoid any bias introduced by the presence of the ligands, we removed all tRNA and EF-G density from these maps. A superposition of both maps illustrates the ratcheting movement (**C**). The two reference maps were projected according to an even angular distribution with a sampling rate of 15 degrees, and a standard projection matching protocol was used to correlate each of the experimental images with the projection libraries of both references. A histogram of the resulting cross-correlation differences (CC2-CC1: cross-correlation with a reference after

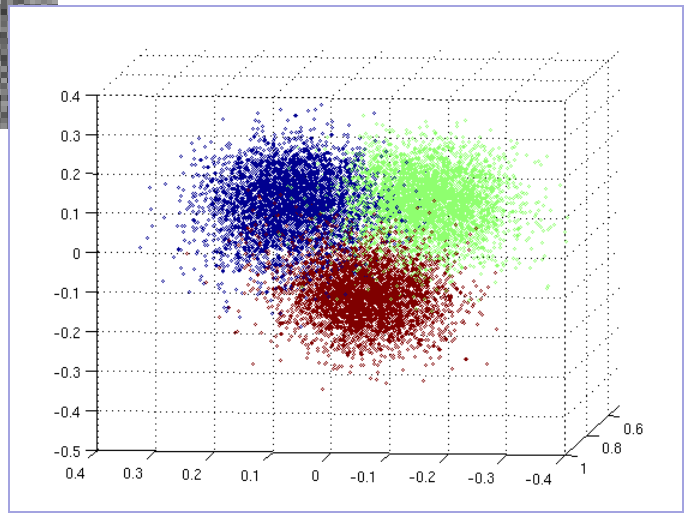
Top: classes derived by Maximum Likelihood-based classification
Bottom: classes derived by supervised classification (CCF with 2 refs)



resolutions: 12-14 Å

Bootstrap Classification





H. Liao and J. Frank, in press

Case Study: Translation Termination in Eukaryotes: 80S Release Complex

Wadsworth Center
Derek Taylor (now Case Western)
Bill Baxter – multi-ref. classification
Jianlin Lei (now Tsinghua) -- AutoEMation
Bob Grassucci -- EM screening
Tapu Shaikh – processing

CNB Madrid
J.M. Carazo
Sjors Scheres

SUNY Downstate Medical Center
Tatyana Pestova -- collaborator
Anett Unbehaun -- sample preparation

Columbia University
Hstau Liao – ML3D
Jie Fu – ML3D

(1) Release of Relief

Translation Termination

- Termination process in bacteria:

(i) RF1 or RF2 bind to ribosome upon encountering stop codon, cleave off polypeptide chain

(ii) RF3 binds to 70S-RFX complex

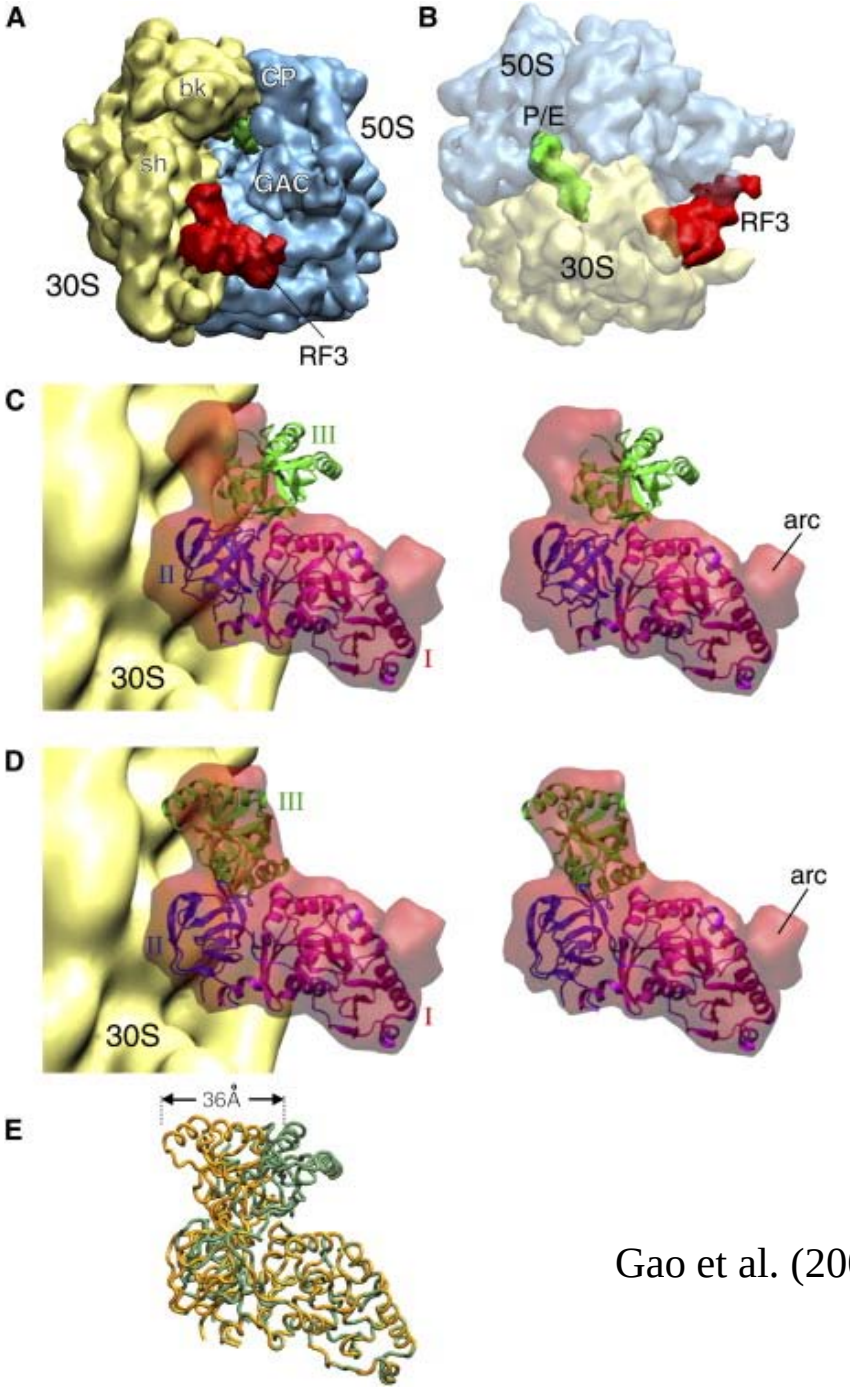
(iii) GTP hydrolysis on RF3; release of RFX and RF3

- Termination process in eukaryotes:

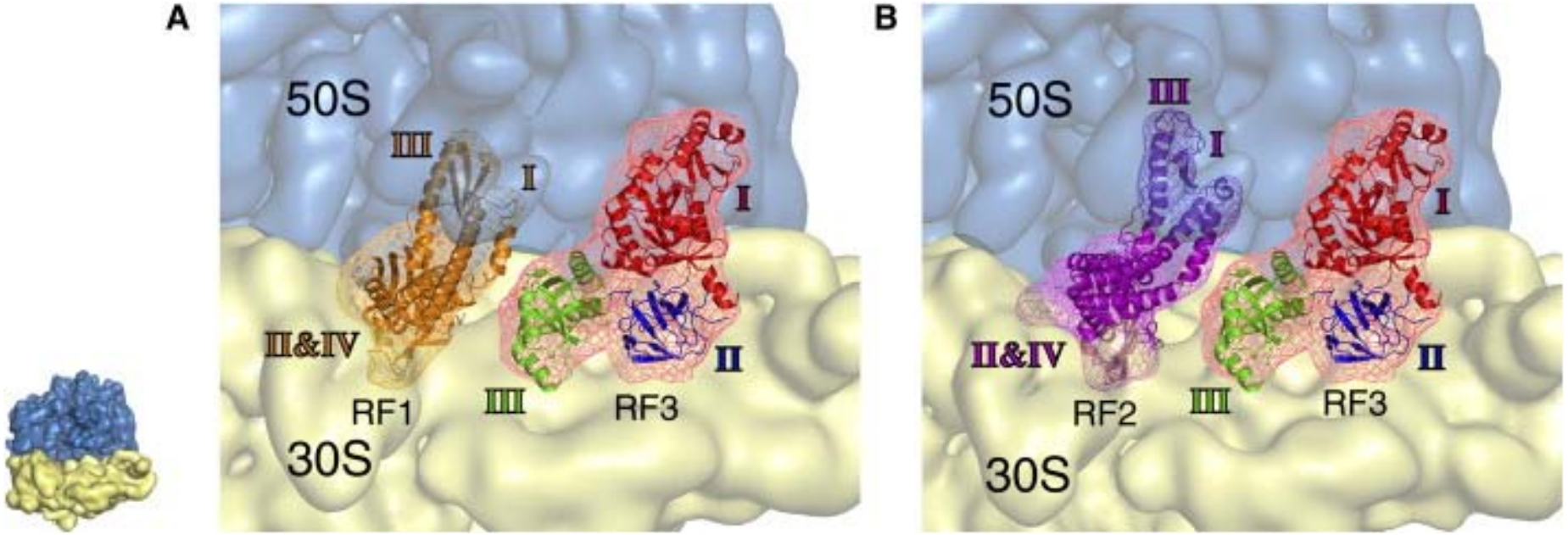
(i) eRF1 binds to stop codon

(ii) eRF3 binds to 80S-eRF3 complex

(iii) GTP hydrolysis on eRF3 → eRF1 cleaves off polypeptide chain

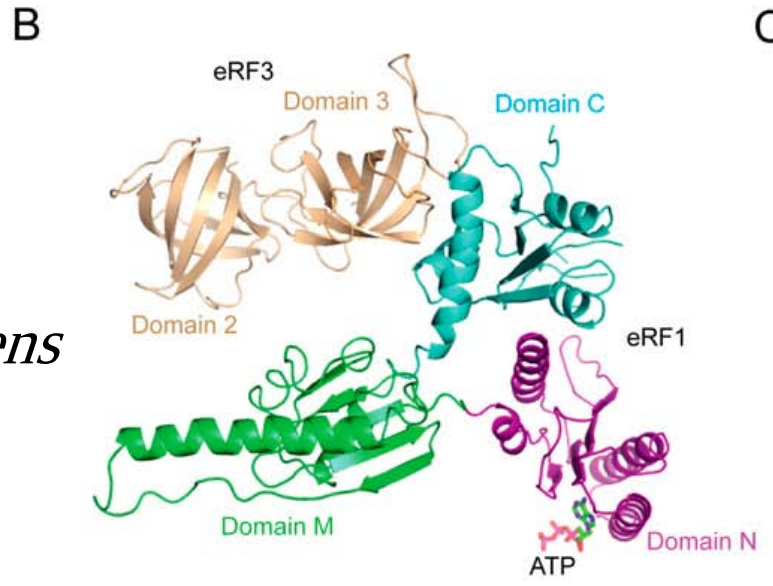


Gao et al. (2007) Cell 129, 929

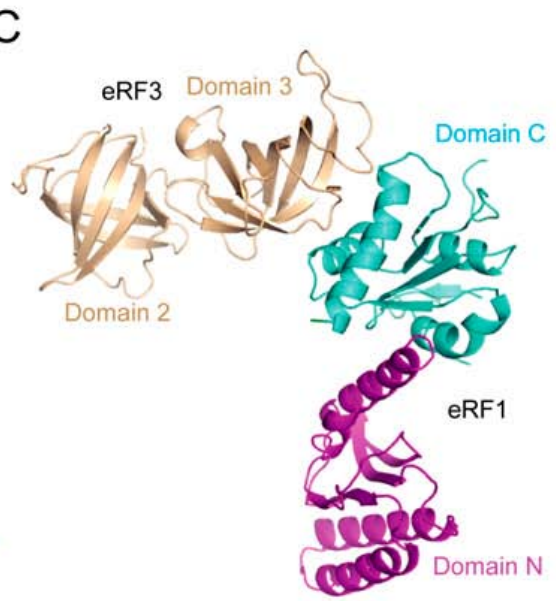


Gao et al. (2007) Cell 129, 929

H. sapiens

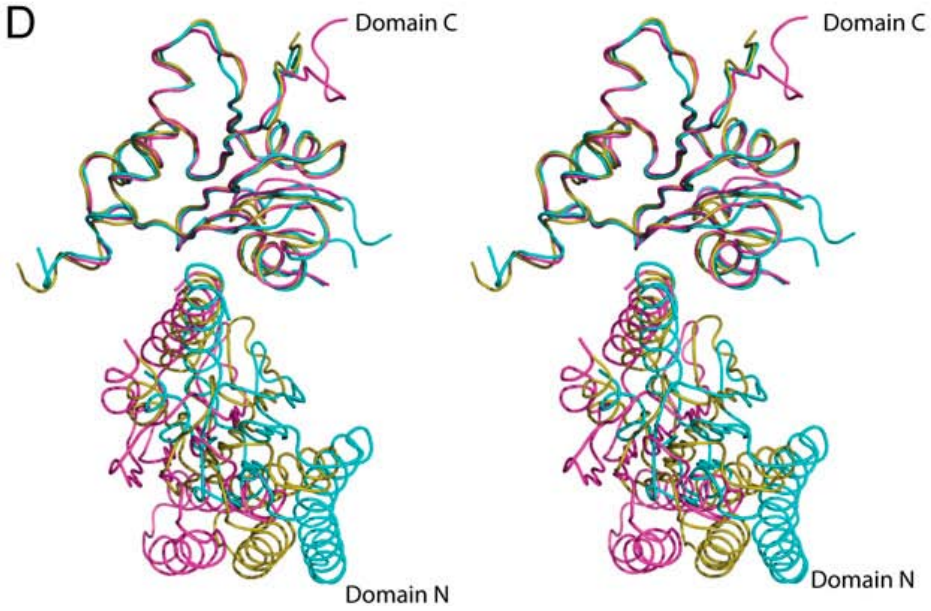


HeRF1/eRF3-23



SpeRF1/eRF3-23

S. pombe



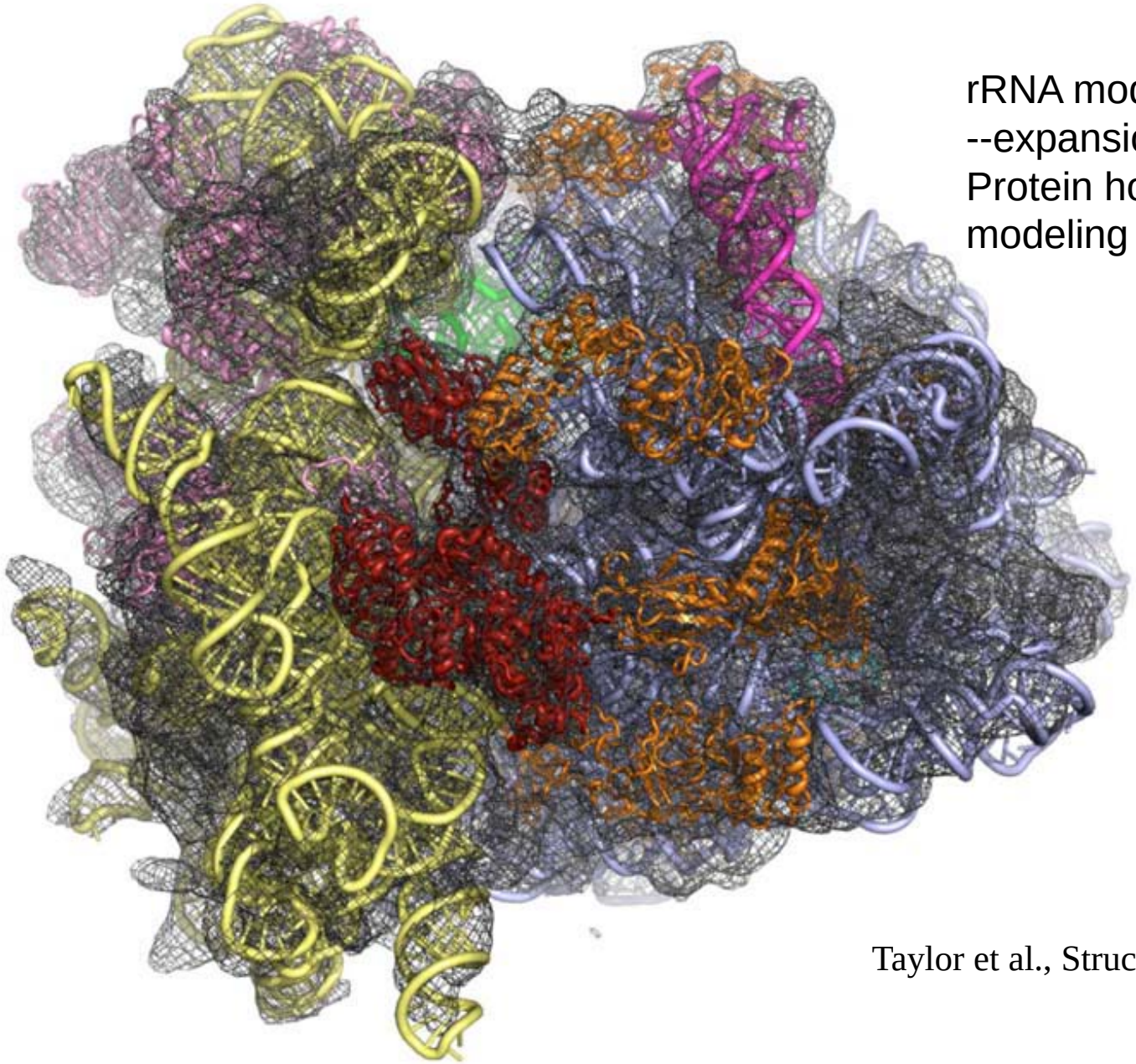
Cheng et al. Gen. & Development 2009

Structural insights into eRF3 and stop codon recognition by eRF1

Zhihong Cheng, Kazuki Saito, Andrey V. Pisarev, Miki Wada, Vera P. Pisareva, Tatyana V. Pestova, Michal Gajda, Adam Round, Chunguang Kong, Mengkiat Lim, Yoshikazu Nakamura, Dmitri I. Svergun, Koichi Ito, and Haiwei Song.

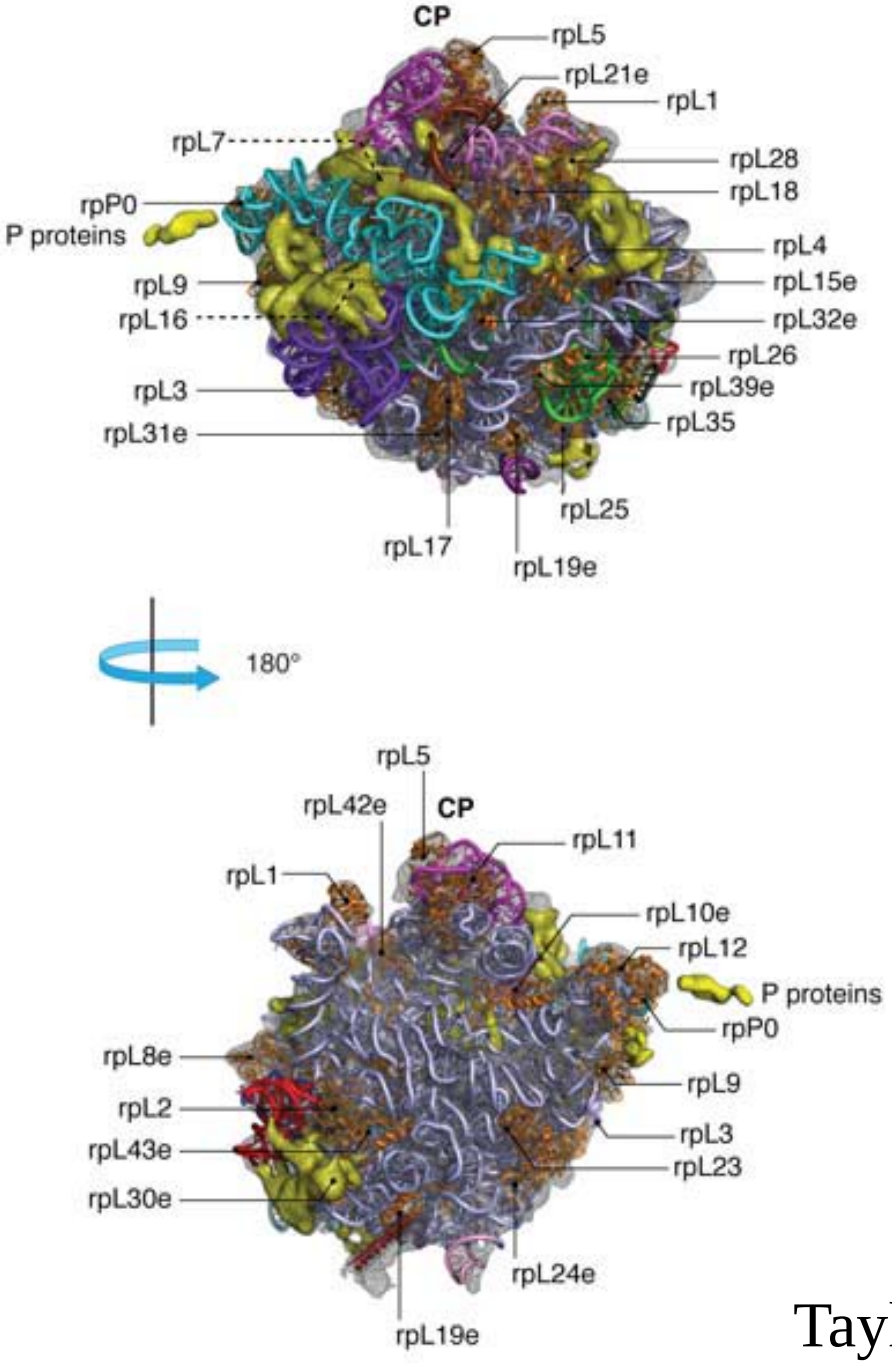
GENES & DEVELOPMENT 23:1106–1118 (2009)

Comprehensive (95% complete) model of the 80S ribosome

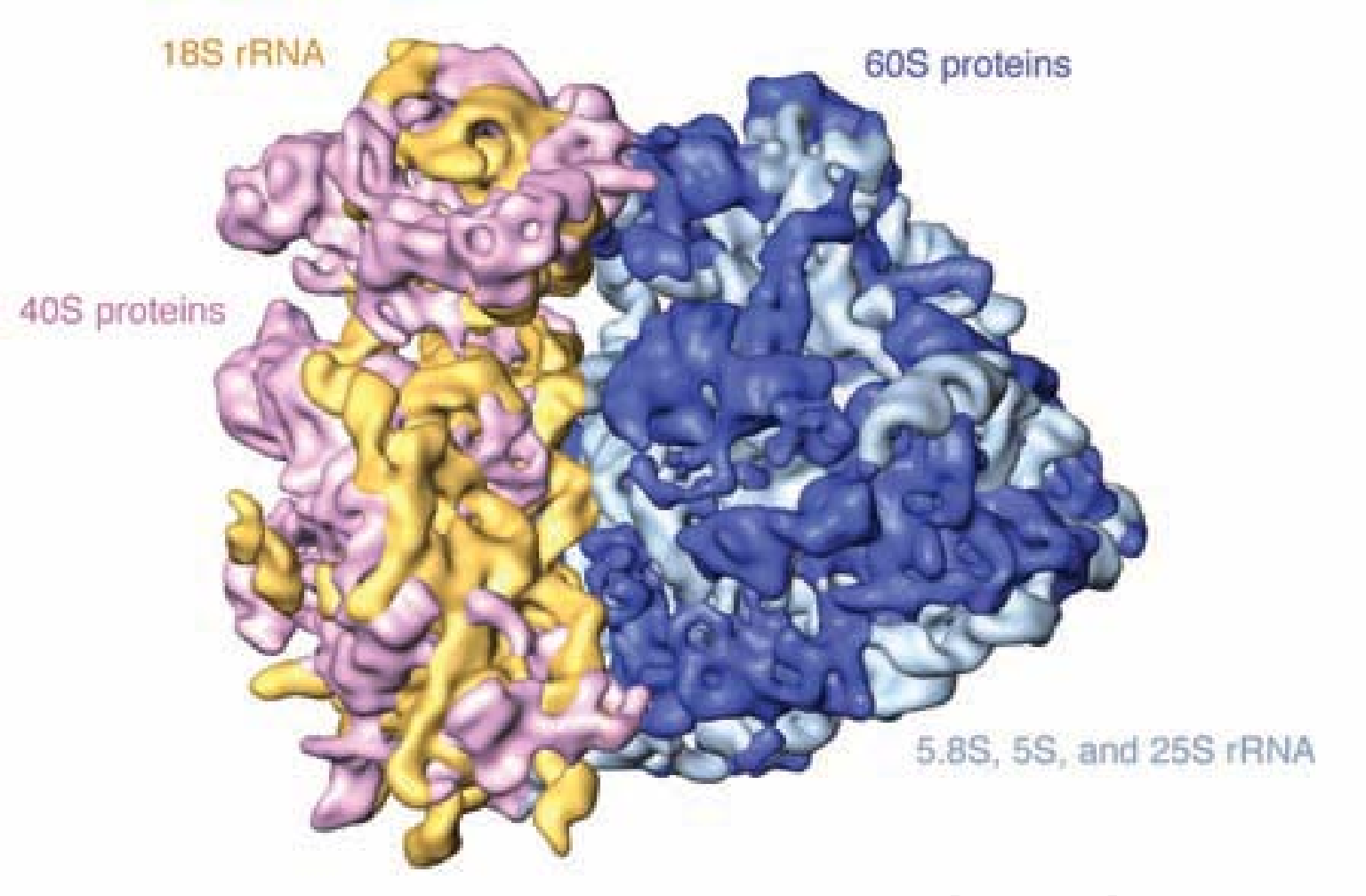


rRNA modeling
--expansion segments
Protein homology modeling

Taylor et al., Structure, in press

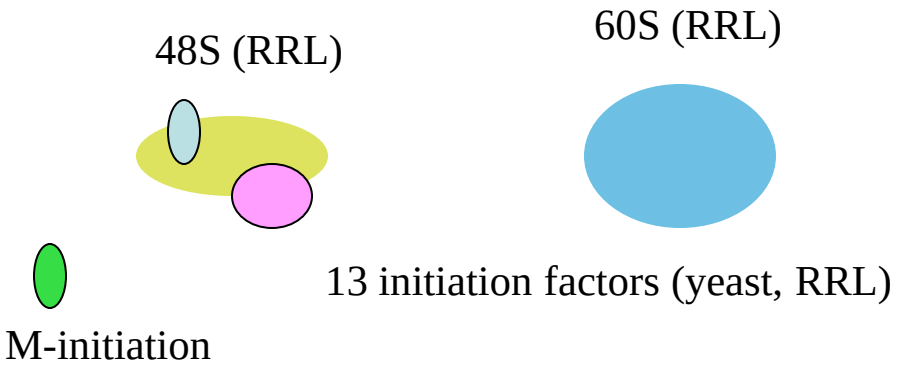


Taylor et al., Structure, in p



Taylor et al., Structure, in pres

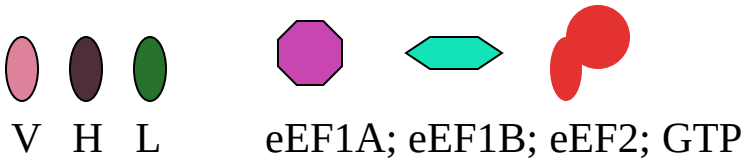
1. Mix



2. Purify

---

3. Add elongation factors

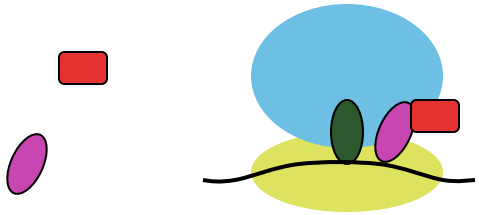


---

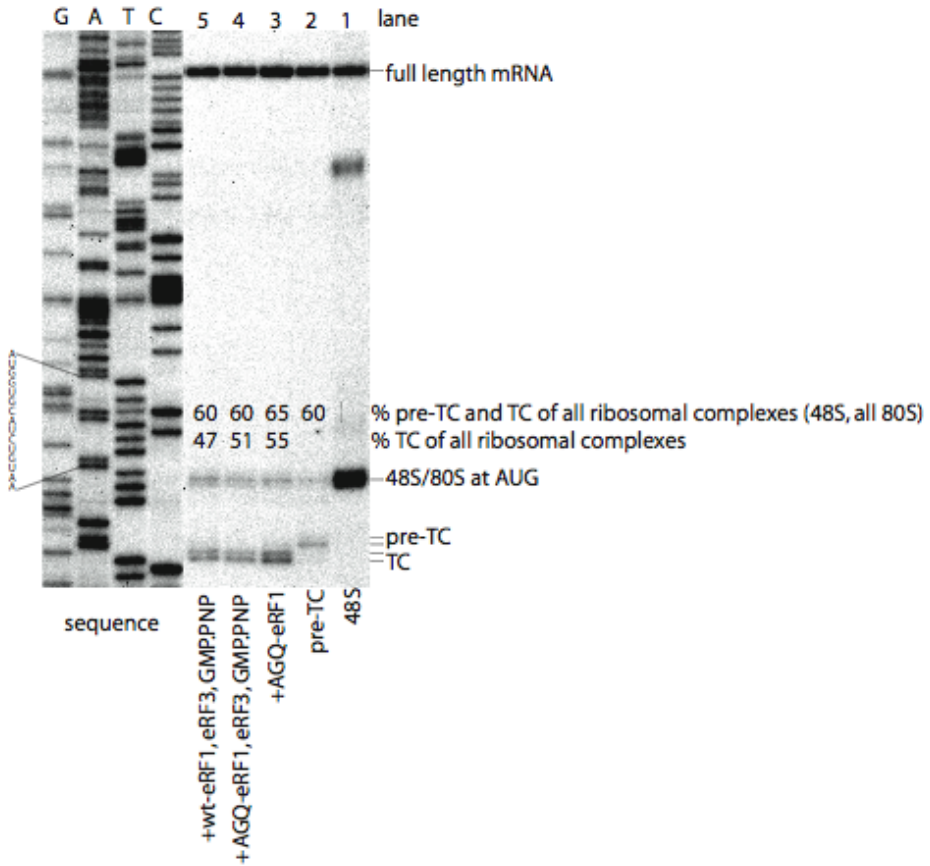
4. Purify

---

5. Add Release Factors; GDPNP



Assembly of ribosomal complexes on mRNA: 5′-(G(CAA)$_{4-}$-(β-globin 5′-UTR)-AUG-GUG-CAU-CUG-UAA-3′-UTR
Met Val His Leu



pre-TC... pre-termination complex (stop codon UAA in the ribosomal A site, the P site contains tRNA- MVHL tetrapeptide)

TC... termination complex (incubation of pre-TC with termination factors and 3 mM GMP.PNP, as indicated)

Challenges:

Limited References, Multiple factors

- 70S much smaller than mammalian 80S
- release of peptide is different in two systems
- eRF1, eRF3, eRF1-eRF3
    - binding of different factors induces conformational changes in the ribosome.

Start with pre-termination complex (no factors)

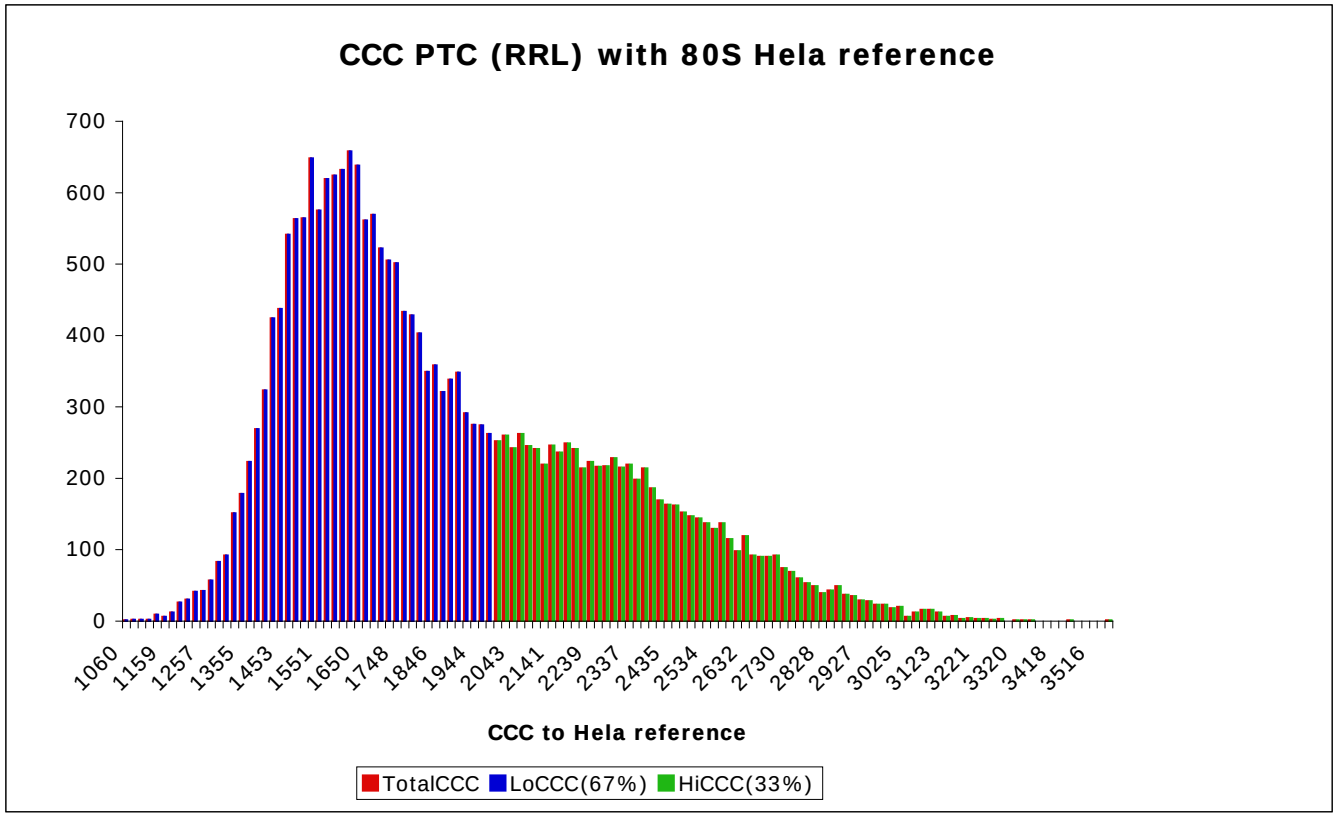
Only 35% are actually programmed.

80S - Rabbit Reticulocyte Lysate - using HeLa 80S reference

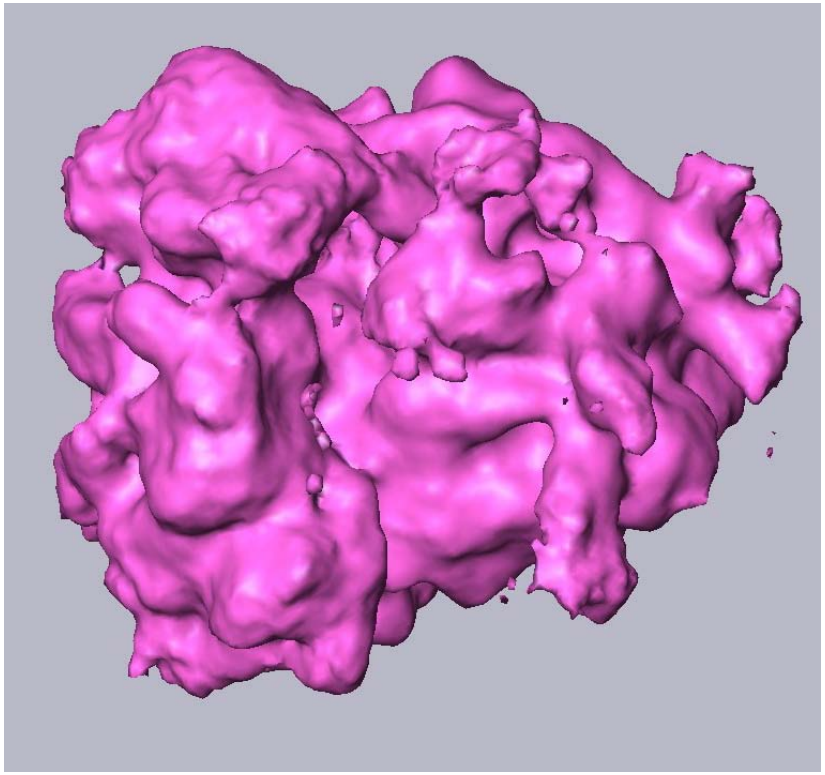
22,816 particles

7541 particles to P-site
Model ~22Å (33%)

15,275 particles to E-site
Model ~26Å (67%)

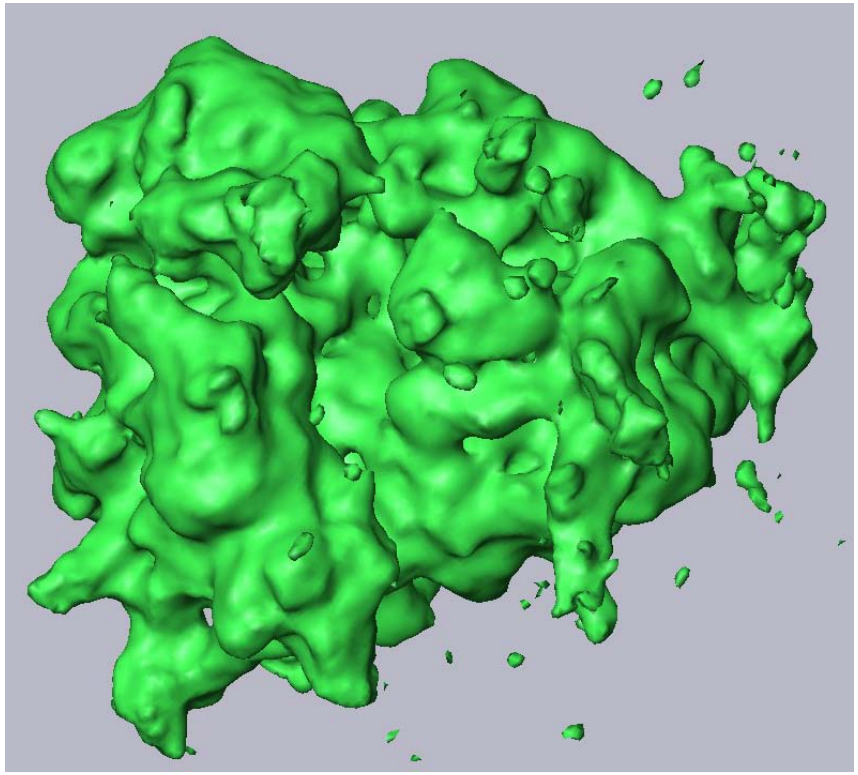


26Å
15,275 particles



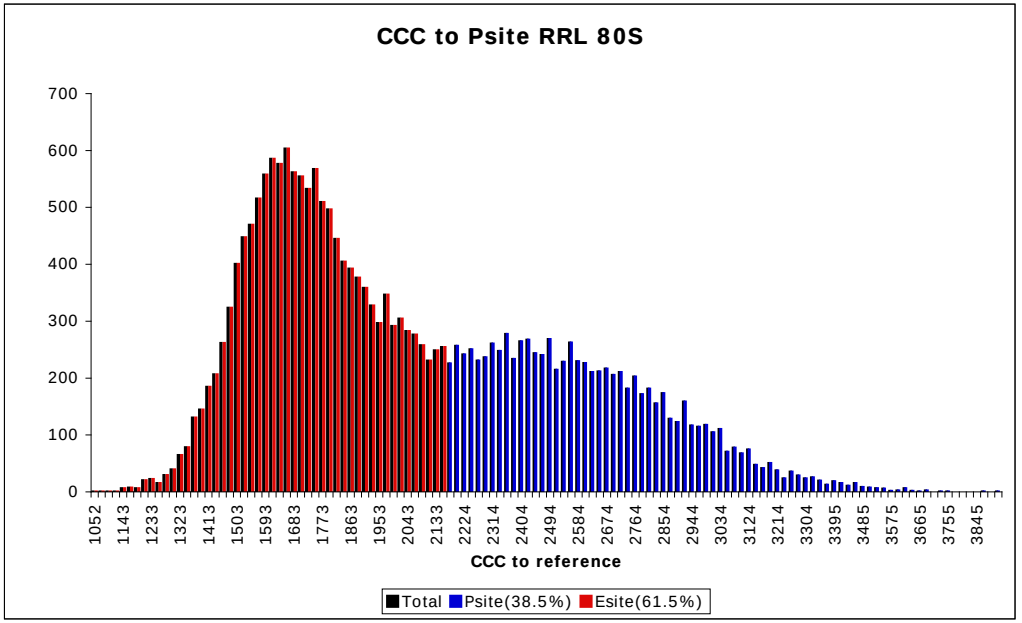
E-site tRNA
Non-specific

22Å
7,541 particles

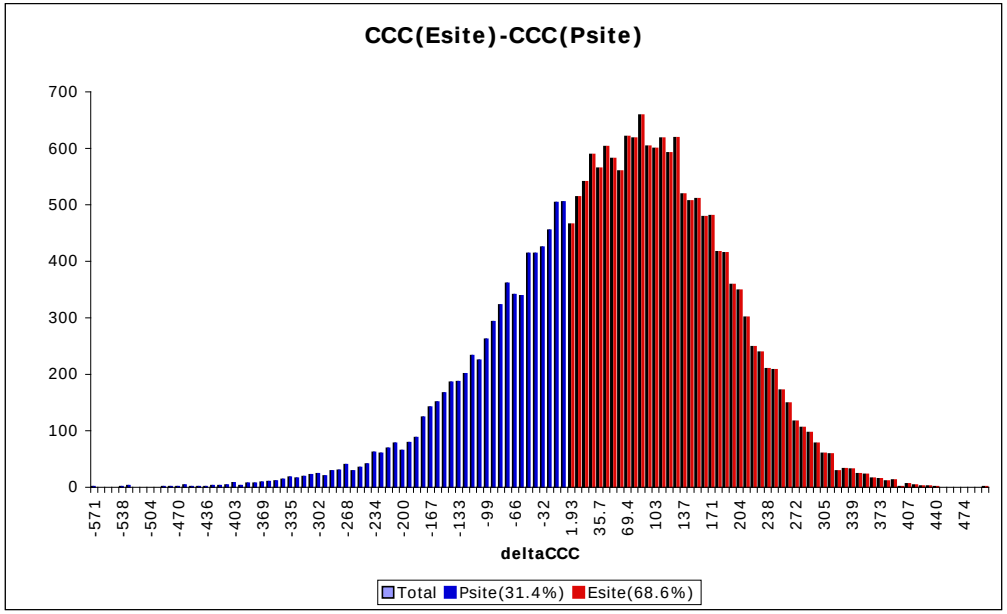


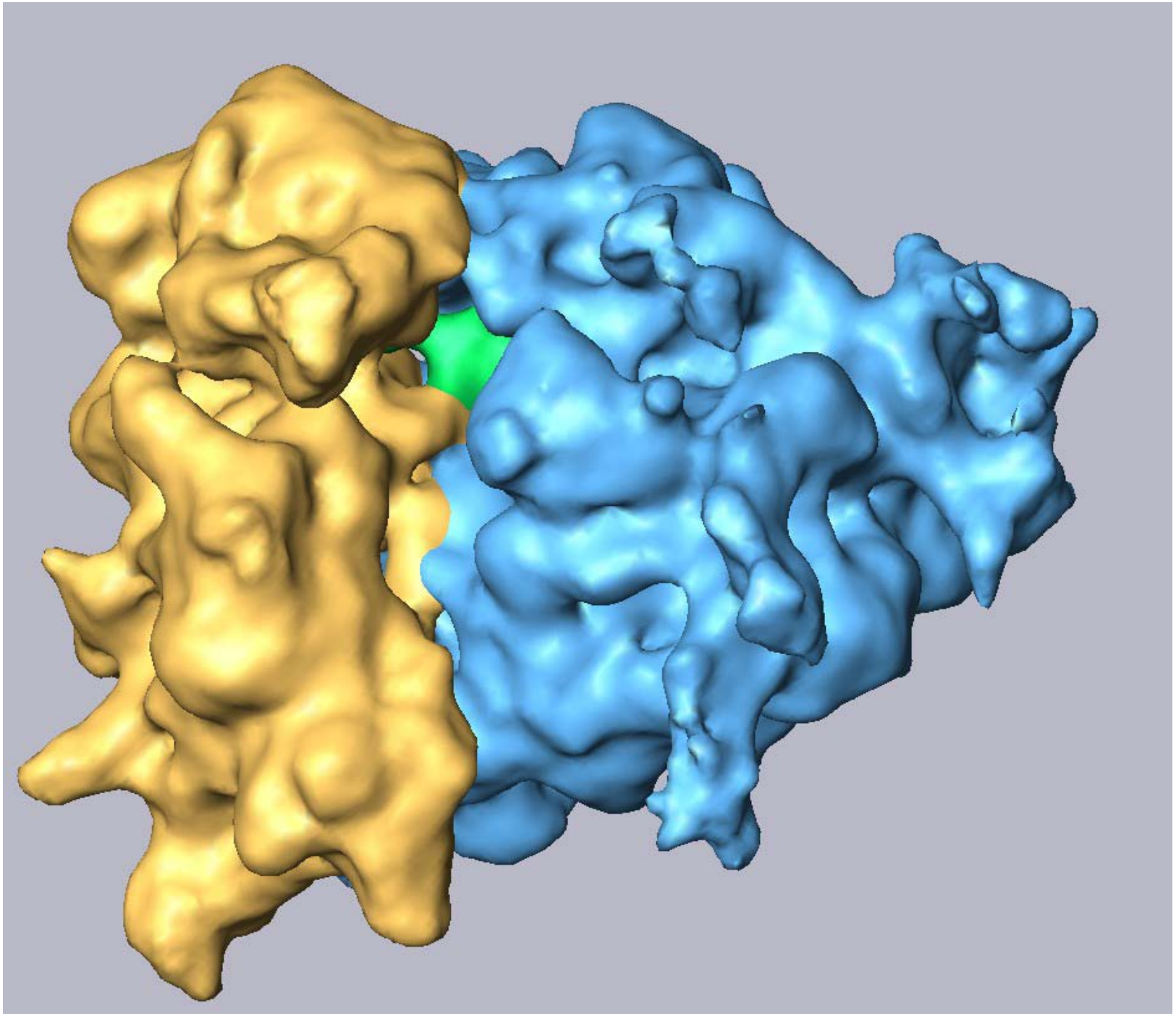
P-site tRNA
Programmed ribosome

Alignment to
P-site model
38% P; 62% E

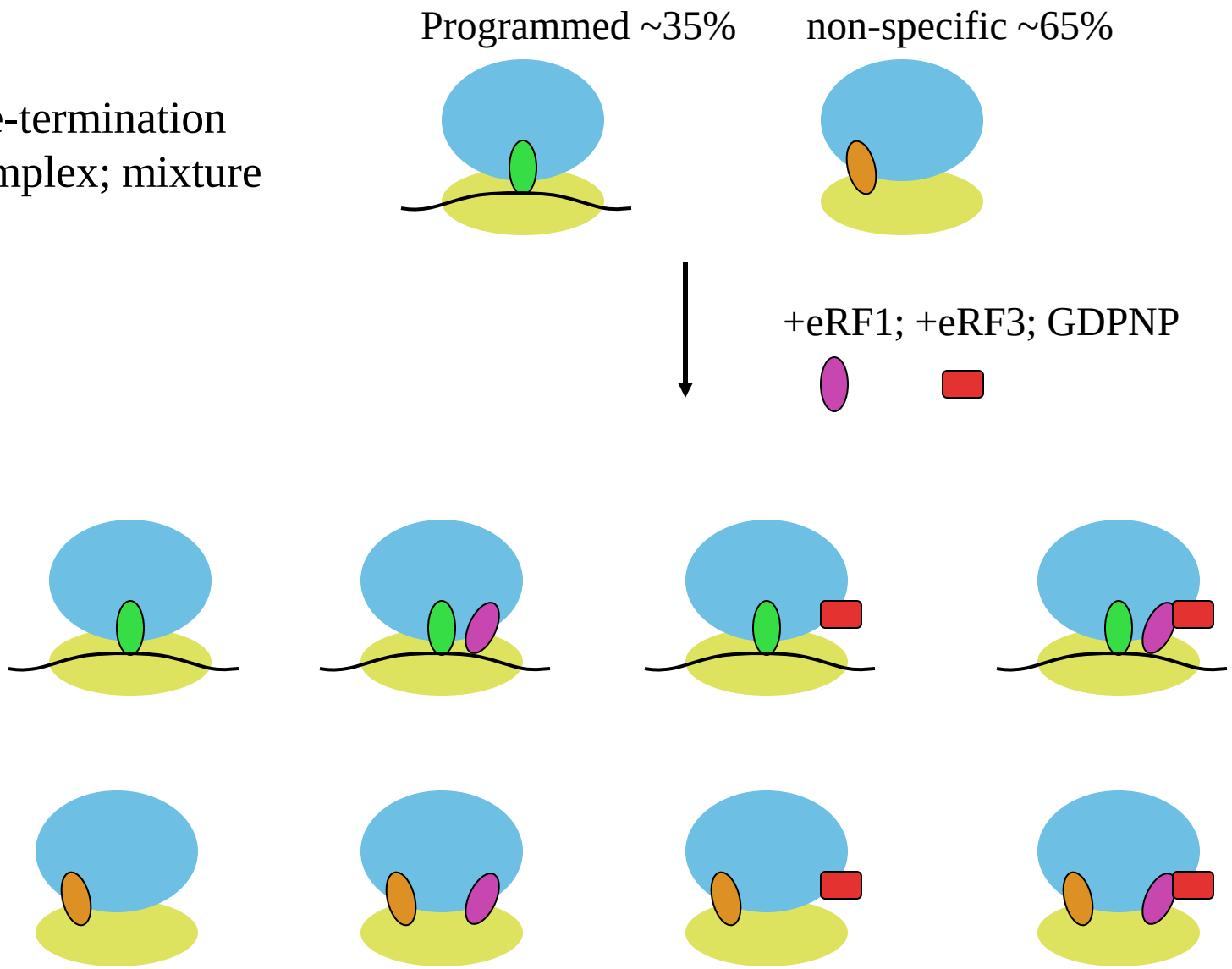


Supervised Class
P-site vs E-site
32% P; 68% E





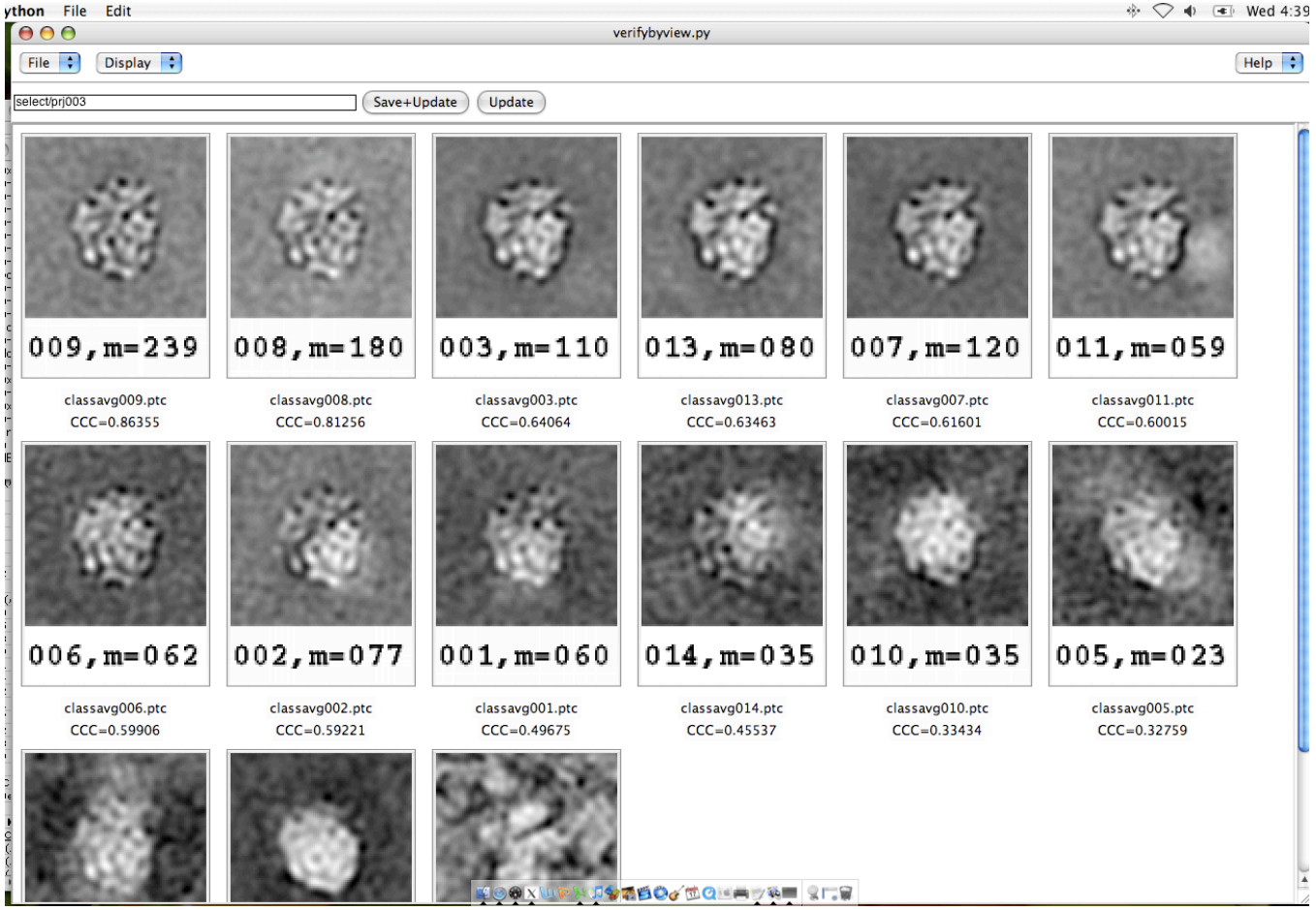
Pre-termination complex; mixture



Particle Verification using Multivariate Data Analysis and Classification

Auto-Emation/Polara → 10 days, 10,000 micrographs CCD

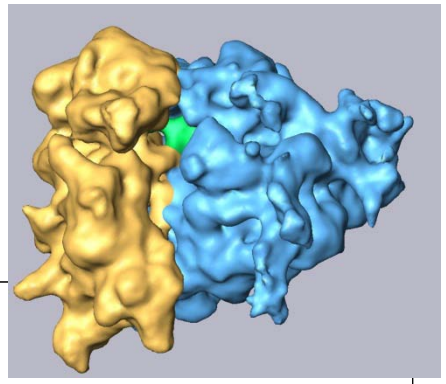
~1M particles selected, 430K verified

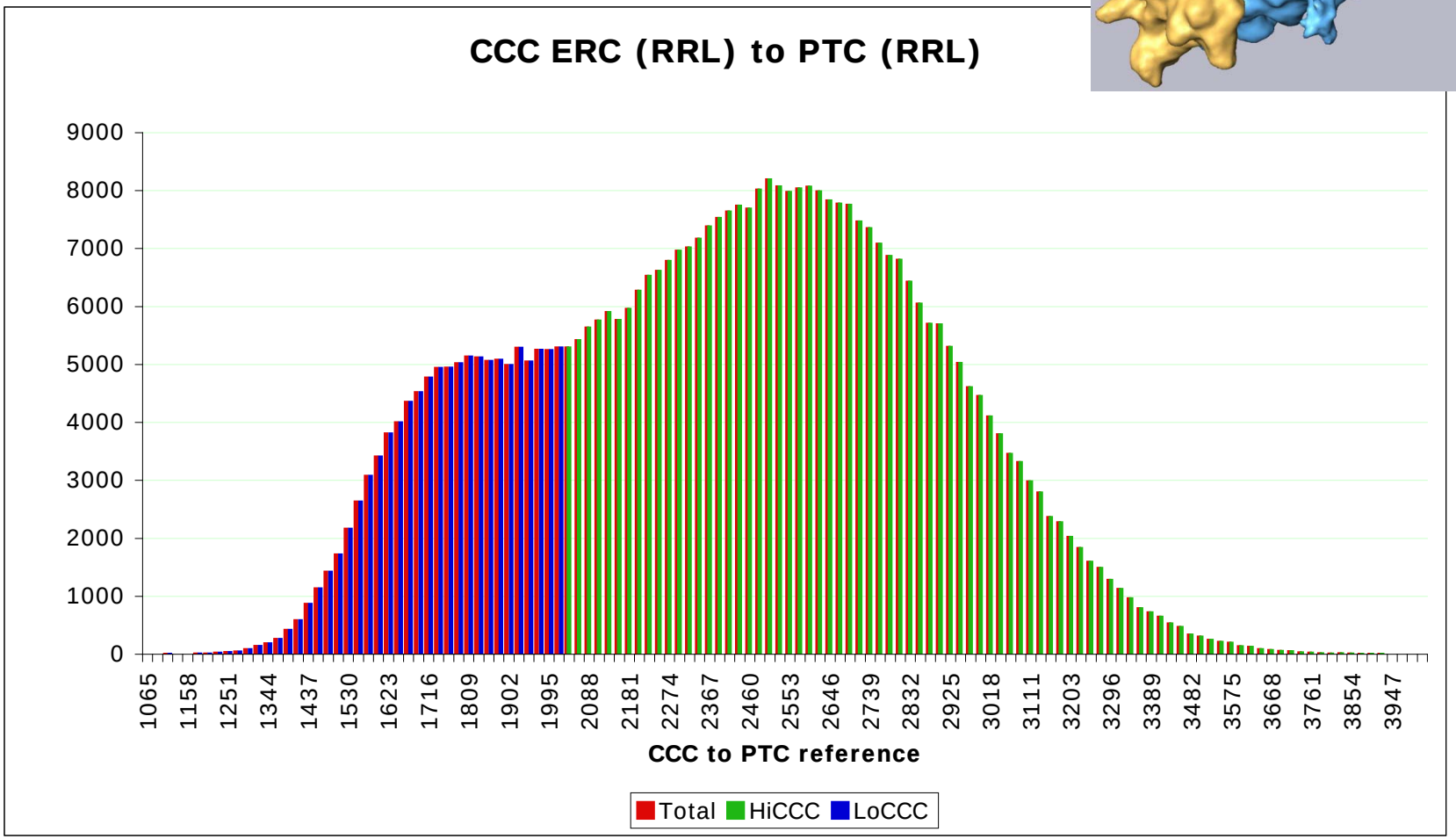


Shaikh *et al.* (2008) *JSB*

Eukaryotic Release Complex

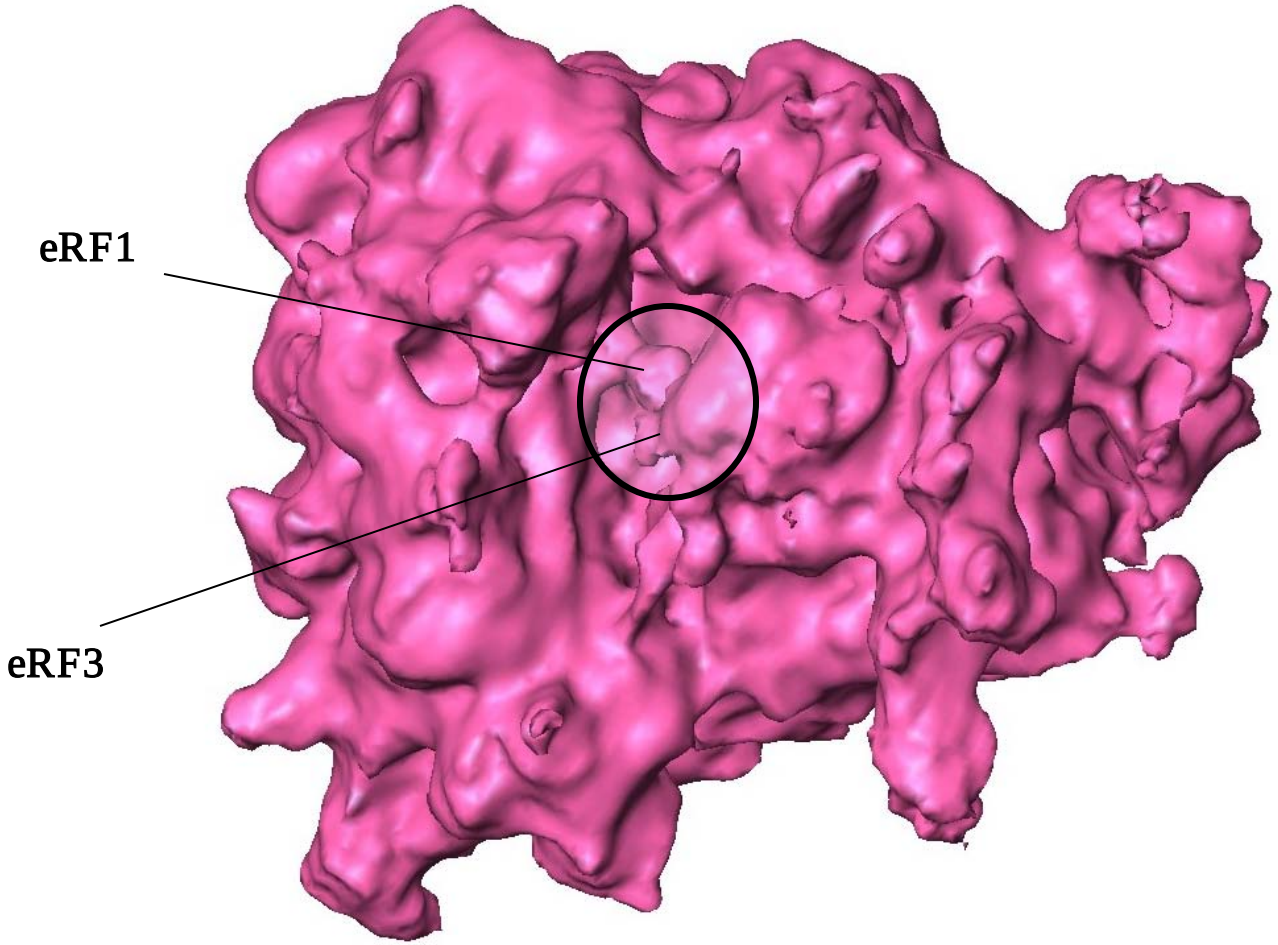
430,167 Total particles verified
106,111 particles in LO CCC class
324,056 particles in HI CCC class





Supervised classification for Factor Density:

~195K have either eRF1, eRF3, or both

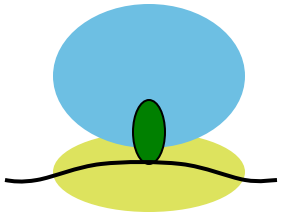


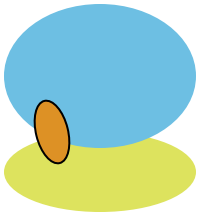
~192K have no factor binding

Programmed ~35%

non-specific ~65%

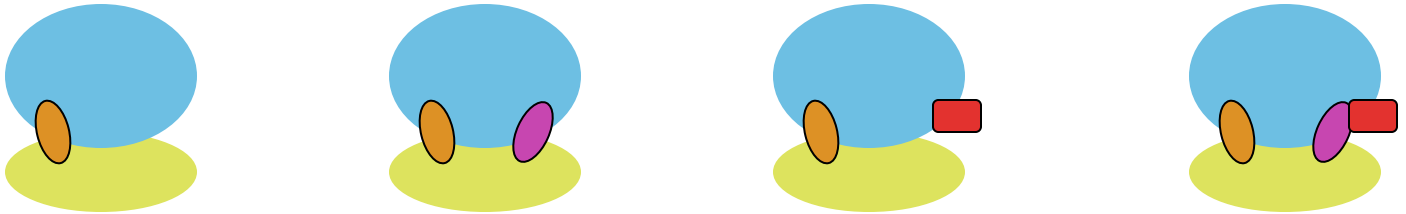
Pre-termination complex; mixture





+eRF1; +eRF3; GDPNP

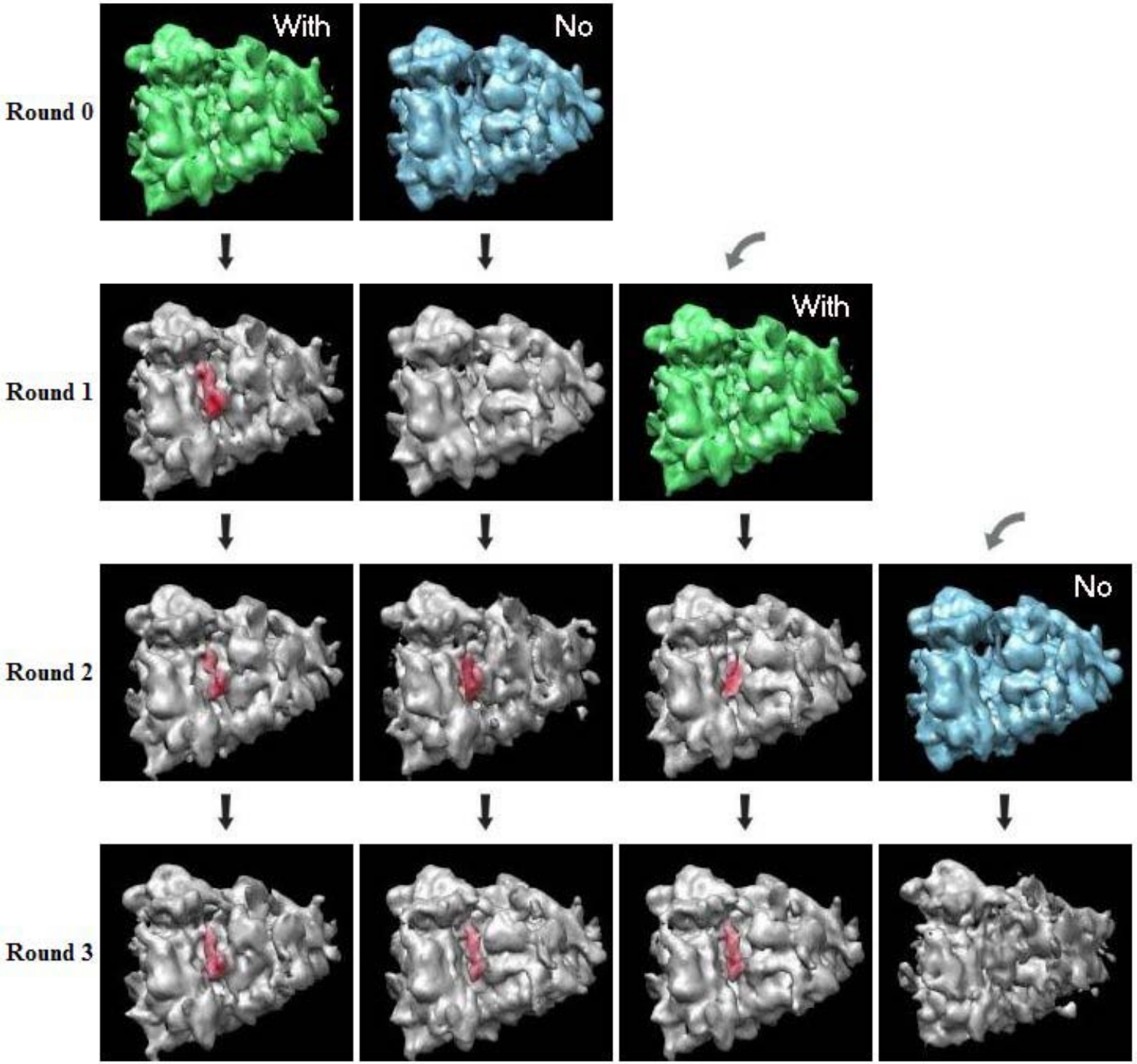
~195K



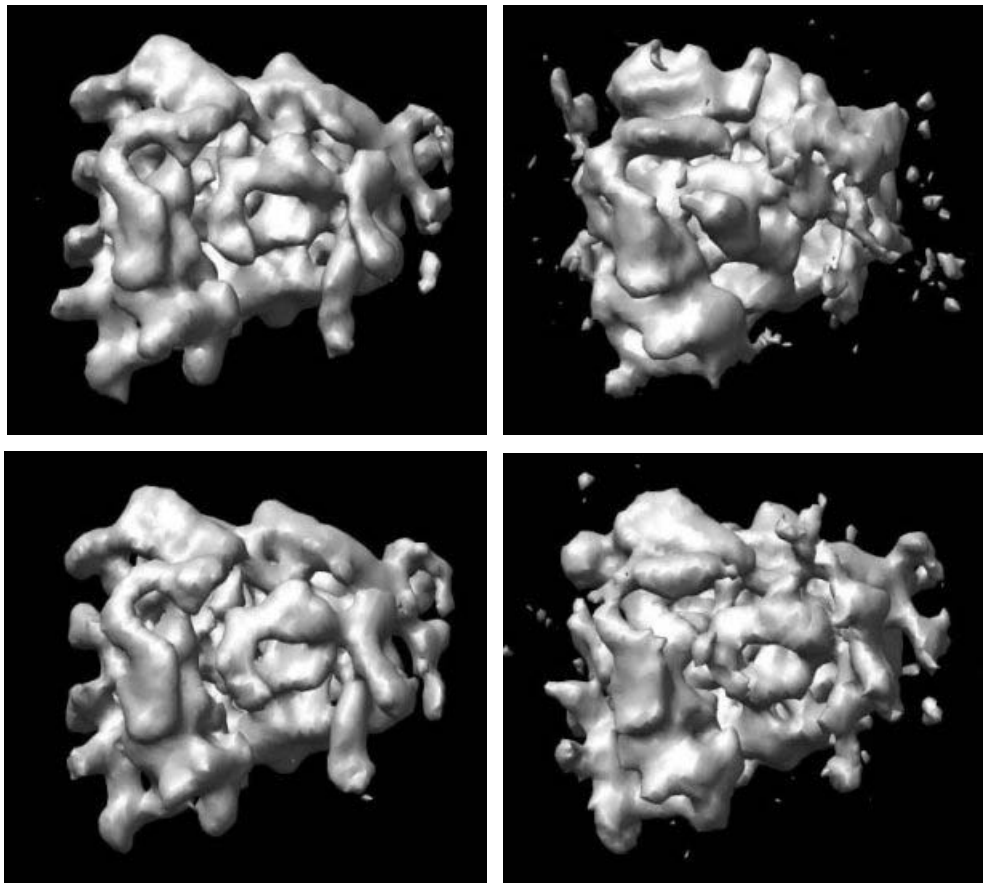
~192K

- Multi-reference (Bill Baxter)
- ML3D (Hstau Liao)

Round 0	**80_eRF1 volume**	**No factor volume**		
Align, reconstruct	↓	↓	↩	
Round 1	**81,058** **12.7 Å** *both*	**45,941** **14.9 Å** *E-site tRNA*	**80_eRF1 volume**	
Align, reconstruct	↓	↓	↓	↩
Round 2	**26,799** **15.7 Å** *both*	**47,344** **13.9 Å** *lower*	**57,904** **13.9 Å** *upper*	**No factor volume**
Align, reconstruct	↓	↓	↓	↓
Round 3	**37,223** **15.2 Å** *lower +*	**31,164** **14.8 Å** *Lower +*	**48,248** **13.8 Å** *low + upper*	**14,470** **18.0 Å** *E-site tRNA*


With
No
Round 0
Round 1
With
Round 2
No
Round 3

ML3D: Hstau applied the maximum-likelihood algorithm (ML3D) to the 200k P-site particle set (downsampled to 76 pixels). This yielded 4 volumes, two of which were distorted and noisy, while the others had densities near the GAC.



volume ML1, 46839 particles,	volume ML2, 63152 particles
volume ML3, 40983 particles	volume ML4, 44458 particles

Figure 2. Volumes from maximum likelihood classification.ML1 : good structure, has large lower factor, no E-site tRNAML2 : noisy structure, malformed, and considerably rotated re volume ML1. Difficult to tell if there is E-site tRNA with so many bridges and distorted small subunit.ML3 : medium quality structure, noisier than ML1, has large lower factor, no E-site tRNAML4 : very similar to ML1, slightly rotated.

Given that there were 2 promising MA volumes (MA1 and MA3) and 2 promising ML volumes (ML1 and ML4), I expected that there would be significant overlap between these pairs. However when particles from intersecting sets were counted, it was found that particles for each ML volume were scattered across all MA volumes:

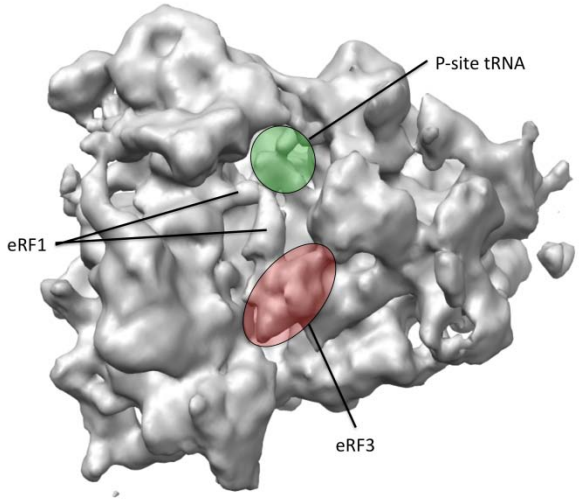
Volumes	**ML1**	**ML2**	**ML3**	**ML4**	**totals**
MA1	***14990***	29959	16280	***13603***	74832
MA2	11518	13027	7956	10368	42869
MA3	***16085***	17152	12440	***15242***	60919
MA4	4246	3014	4307	5245	16812
Totals	46839	63152	40983	44458	195432

Table 2 numbers of particles in overlapping ML and MA sets.
Reconstructions were made of selected intersection sets, specifically, those that corresponded to maximum-likelihood volumes ML1 and ML4, and multireference volumes MA1 and MA3 (bold italics in table). After discarding some particles for the above-mentioned reason, these four volumes were obtained:
MA1_ML1 : 12611 particles, 17.3 A resolution
MA1_ML4 : 11440 particles, 17.7 A
MA3_ML1 : 14140 particles, 16.7 A
MA3_ML4 : 13875 particles, 17.6 A

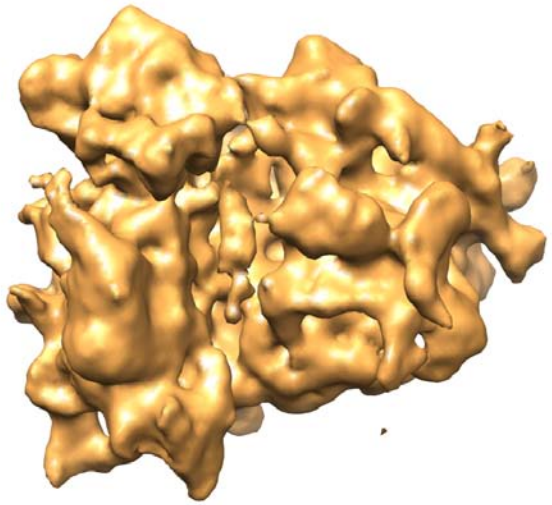
Difference in Termination Mechanisms Bacteria vs. Eukaryotes

Bacteria: RF1 or RF2 binds to stop codon at decoding center and interacts with PTC to cleave peptide bond & release the chain. After that, GTPase RF3 binds to cause release of RF1 or RF2.

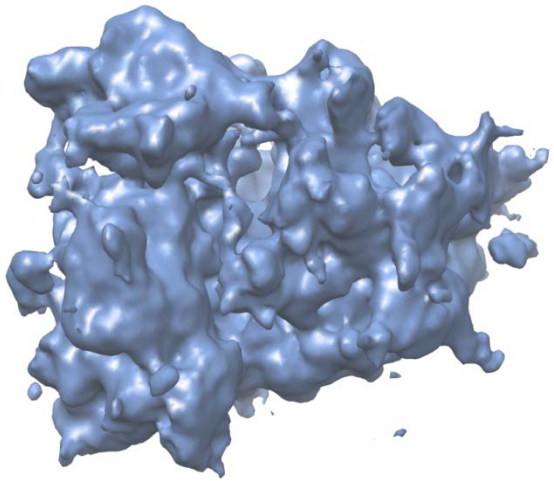
Eukaryotes: eRF1 binds to stop codon at decoding center, but it requires the binding & GTP hydrolysis of eRF3 before it will cleave the peptide bond.



eRF1 + eRF3



eRF1 only



eRF3 only

- How was variability detected?

  (i) local blurring, (ii) appearance of physically impossible density regions (fragmented or overlap[pewd density of ligands)
- How were various populations sorted and averaged?

  see above
- What were the thought processes and decisions made along the way?

  panic
- How were the various problems that were encountered solved?

  tenacity
- What is the pipeline in terms of new approaches?

  data collection needs to be streamlined – screen at the very first opportunity (data coming from EM)
- What does not work?

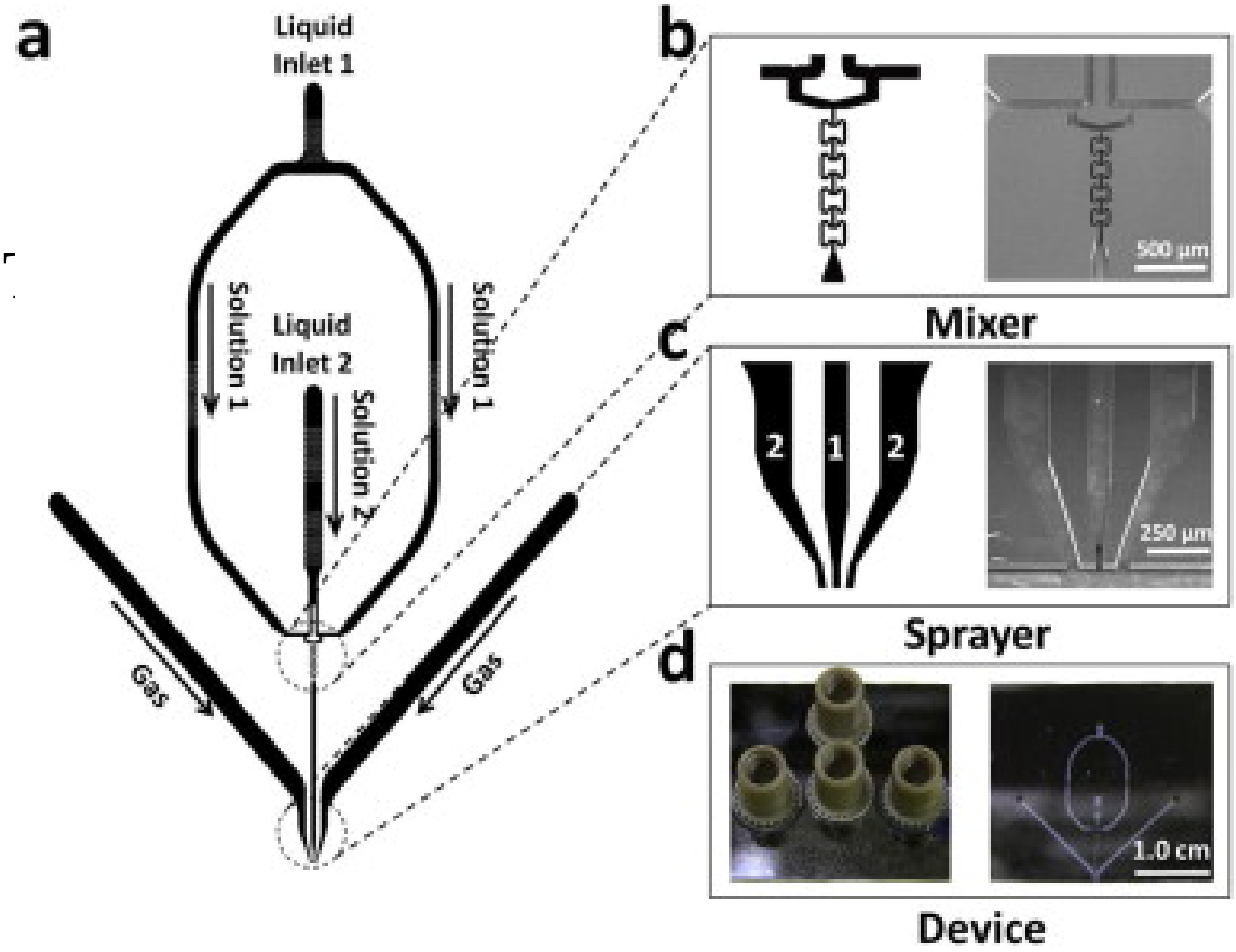
Time-resolved cryo-EM

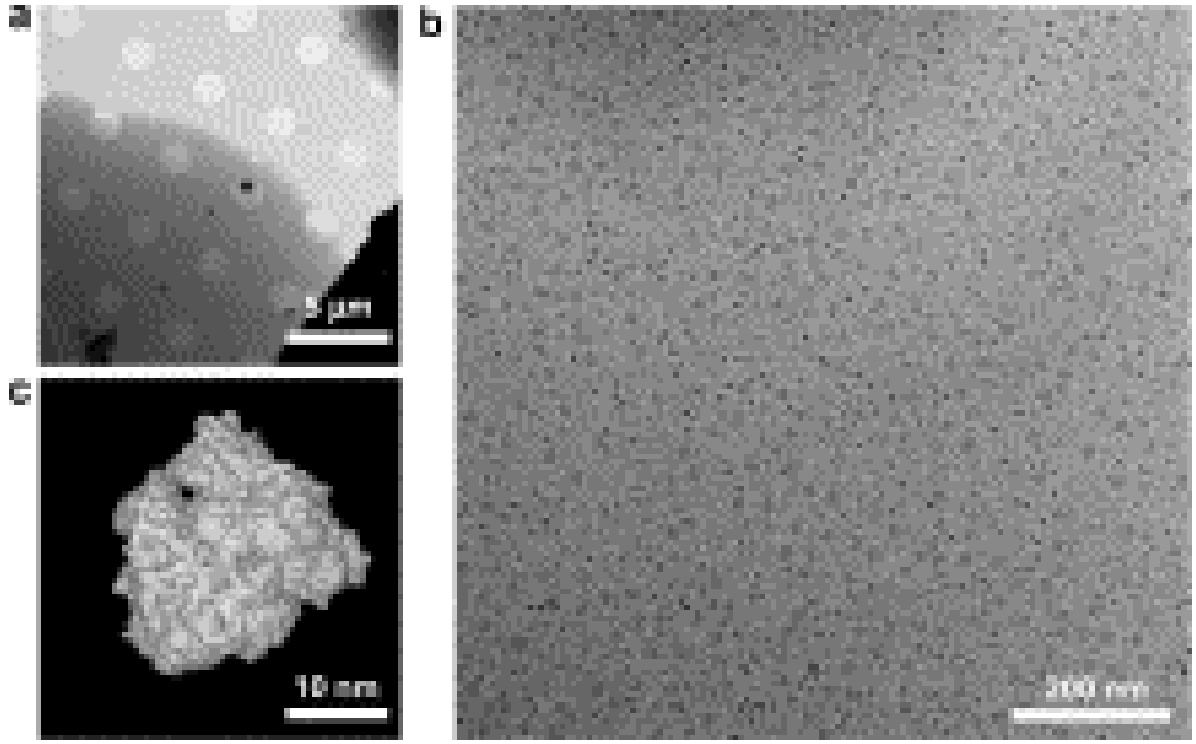
Monolithic microfluidic mixing–spraying devices for time-resolved cryo-electron microscopy

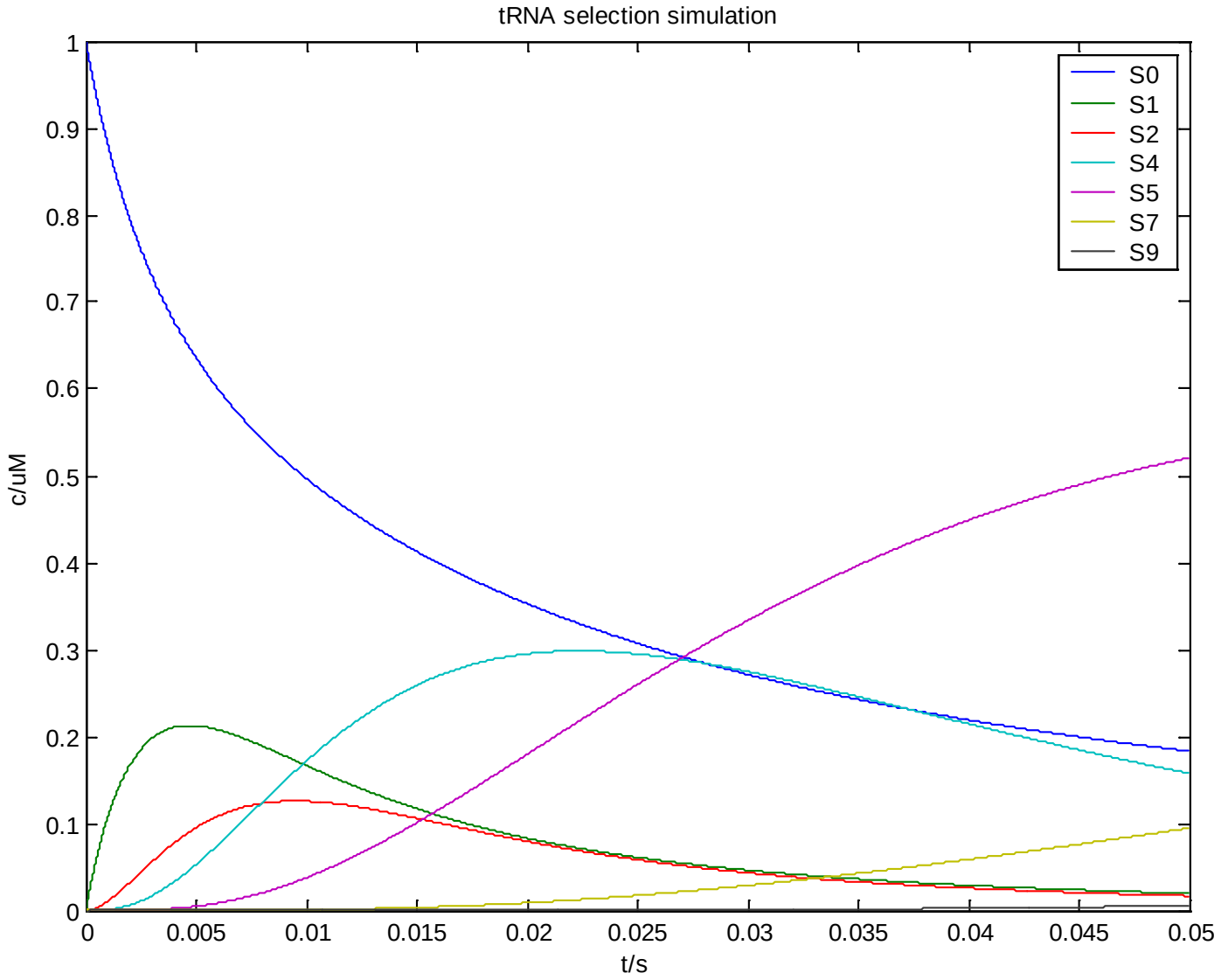
Zonghuan Lu, Tanvir R. Shaikh, David Barnard, Xing Meng, Hisham Mohamed, Aymen Yassin, Carmen A. Mannella, Rajendra K. Agrawal, Toh-Ming Lu and Terence Wagenknecht

J. Struct. Biol. 2009

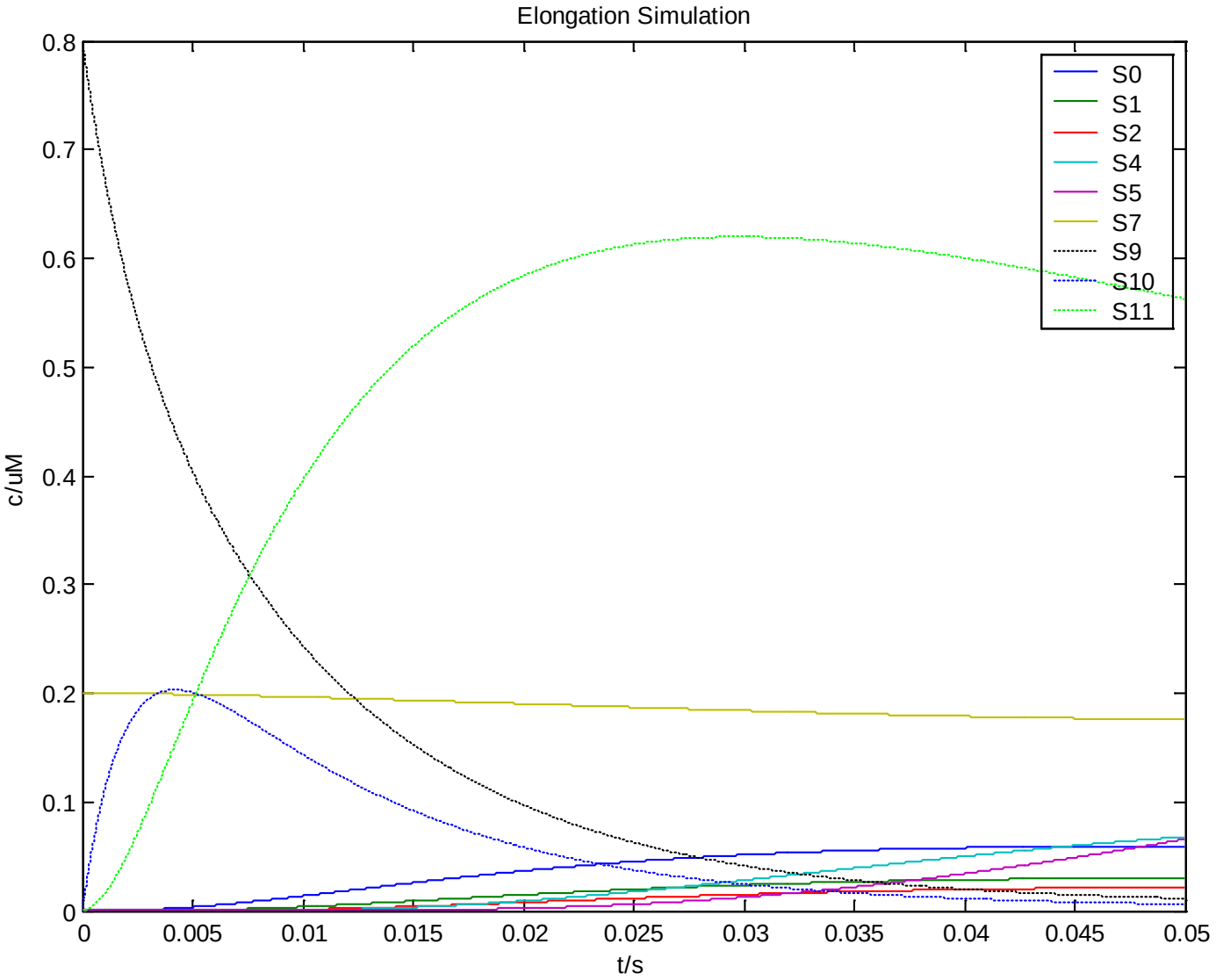
Resource for the Visualization of Biological Complexity, Wadsworth Center, Albany


a
Liquid
Inlet 1
Solution 1
Liquid
Inlet 2
Solution 2
Solution 1
Gas
Gas
b
500 μm
Mixer
c
2
1
2
250 μm
Sprayer
d
1.0 cm
Device


a
5 μm
b
200 nm
c
10 nm


tRNA selection simulation
S0
S1
S2
S4
S5
S7
S9
c/uM
t/s
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
0.045
0.05

Decoding



Decoding + Translocation

Using NVIDIA GPU hardware and the CUDA programming architecture:

Acceleration of supervised classification inherent in projection matching.