Gene cloning:
Expression of protein and RNA

Controlled gene expression

I. proteins
- Various tags and their functions
- A protocol for purification of 6xHis tagged proteins expressed in *E. coli*

II. RNA
- For in vitro uses
- For RNA interference (RNAi) in vivo

Guide to readings:

1) *30 MC4 Protein expression*. Intro to protein expression in bacterial and eukaryotic systems, vectors, optimization.

2) *32 MC4 Epitope tags.* A list and description of tags/handles that can be added to recombinant proteins.

3) *Protein expression tags Malhotra 2009.* More good info on tags.

4) *31 MC4 His tag protein purification*. A detailed protocol for production of 6 x His tagged proteins.

5) RNAi delivery methods 2015. Trends in RNAi –based therapies

6) *SARS 1 vaccine 2009.* Using spike protein for immune response

7) *Hotez and Bottazzi 2020*. Production of affordable SARS CoV2 protein antigen vaccine.

Controlled protein or RNA production is important

1) Production of proteins & RNA:
- Biochemical studies (enzyme function)
- Structural studies (crystallography, NMR
- Commercially useful proteins or RNAs
- Protein pharmaceuticals
- Induction of RNA interference (RNAi)

2) Living systems
- Cell biology -- protein localization
- Transgenic plants and animals
- Human gene therapy

Strategy for protein “overexpression”

Clone

Express

Purify

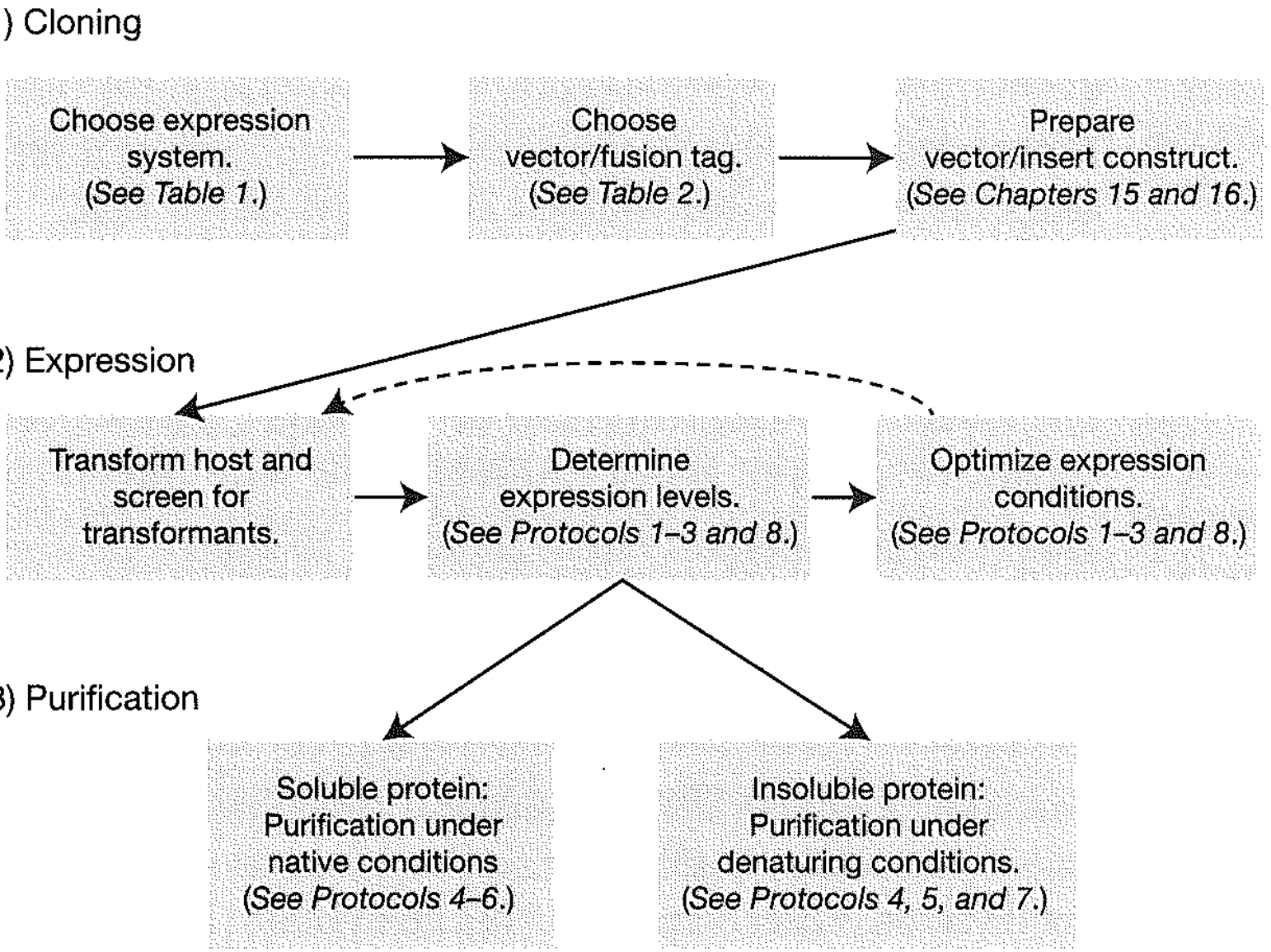


FIGURE 1. **Overall strategy for recombinant protein expression and purification.**

Expression vectors for protein/RNA purification

- Make lots of protein or RNA
- Control timing of synthesis
- Control amount of synthesis

- Optimized for transcription, translation
- Facilitate post-translational handling by adding “ tags”

The most common expression platform is bacterial, specifically *E. coli*

- Lots of vectors available
- Lots of host *E. coli* strains available
- Large variety of tags for proteins

- Straightforward, easily adaptable protocols
  - Add gene to expression plasmid
  - Transform host strain with expression plasmid
  - Grow cells to mid-log (exponential phase)
  - Induce protein expression
  - Purify the protein

Alternative expression hosts sometimes needed

Eukaryotic protein expression in bacteria not always successful:

- Some proteins do not fold properly
- May lack post-translational modifications required for activity, especially glycosylation
- Protein might be toxic to bacteria

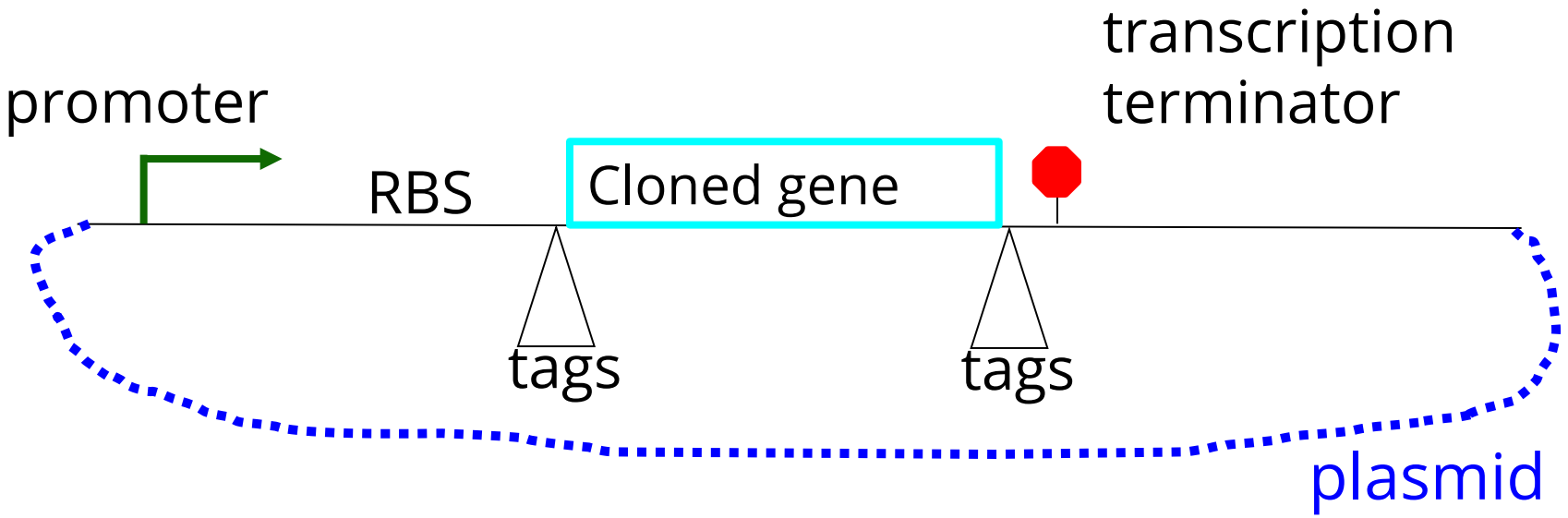
Some alternatives to bacterial expression:

1) Baculovirus-infected insect cell culture
2) *Pichia pastoris*, an easily cultured yeast
3) Mammalian cell culture

4) Cell-free translation system

See Table 1 in *30MC4 Protein Expression* for comparison of these methods

A bacterial protein expression vector



1) Promoter for transcription
2) Transcription terminator
3) Ribosome binding site (RBS) for translation
4) Epitope tags for purificationor detection of the protein

To maximize synthesis of gene product:

1) Bring cells to mid-log phase (all cells in the process of growth and division) without expressing the gene: cell physiology is optimized

2) When conditions are right, " induce" the gene, by turning on transcription/mRNA synthesis
   - lots of ribosomes available
   - protein synthetic machinery is at a maximum to cope with new mRNA for induced gene
   - Ideally, massive amount of new protein will be made

Inducible promoters for bacterial gene expression

Promoter	Transcription regulation	Induction
Lac	Lac repressor	IPTG (lactose analogue)
Trp/lac	Lac repressor	IPTG
T7	Lac repressor, T7 RNA polymerase	IPTG
pBAD	AraC repressor	arabinose
Lambda P_L	Temperature sensitive lambda repressor	Shift from 30°C to 37°C

Fusion protein tags: protein purification/detection

- Tags can be added to the N- or C-terminus, or can be internal to the protein sequence (on a surface loop)

- Tags can be engineered to be removable

- More than one tag can be added to a protein, for "Tandem Affinity Purification"

- Peptide tags. Example: FLAG tag, detected by monoclonal antibodies

- Whole proteins. Biotin carboxylase: covalently attaches to biotin, biotin binds to streptavidin which can be immobilized on columns or beads

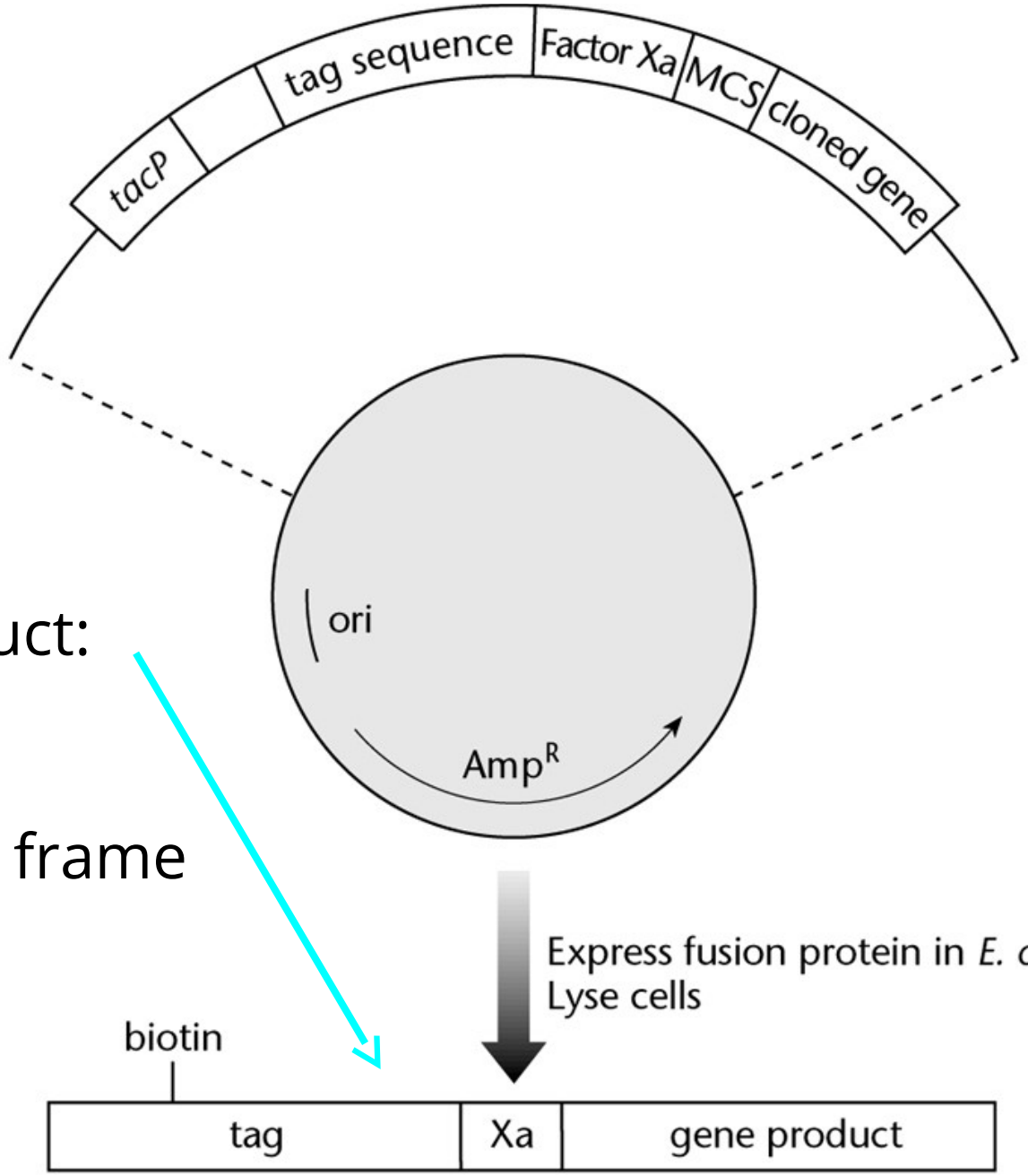
Epitope tags as purification handles

Protein tag	**Size**	**What the tag sticks to**	**What causes release**
Polyhistidine (6-10xHis)	6-10 a.a.	Immobilized Ni, Co, Cu, Zn ions	Imidazole >100 mM
Glutathione S-transferase (GST)	211 a.a.	Immobilized glutathione	Reduced glutathione
FLAG-tag: DYKDDDDK	8 a.a.	Anti-FLAG antibody	FLAG peptide or low pH
Protein A	280 a.a.	Any immunoglobulin G	Protease
Calmodulin binding protein	26 a.a.	Immobilized calmodulin	EGTA 2mM
Acceptor peptide (**biotin** added by biotin ligase)	15 a.a.	Avidin/streptavidin	Biotin

Tags for protein folding/solubility

Protein tag	Size	What the tag does	Additional facts
SUMO (small ubiquitin-like modifier)	98 a.a.	Improve protein folding/solubility	Can be removed by SUMO protease
MBP: maltose binding protein	396 a.a.	Helps folding/ solubility AND sticks to cross-linked amylose for purification	Released from amylose by adding maltose
Trx: thioredoxin	109 a.a.	Improve protein folding/solubility	Extremely soluble, even at 40% of total cell protein, prevents inclusion bodies

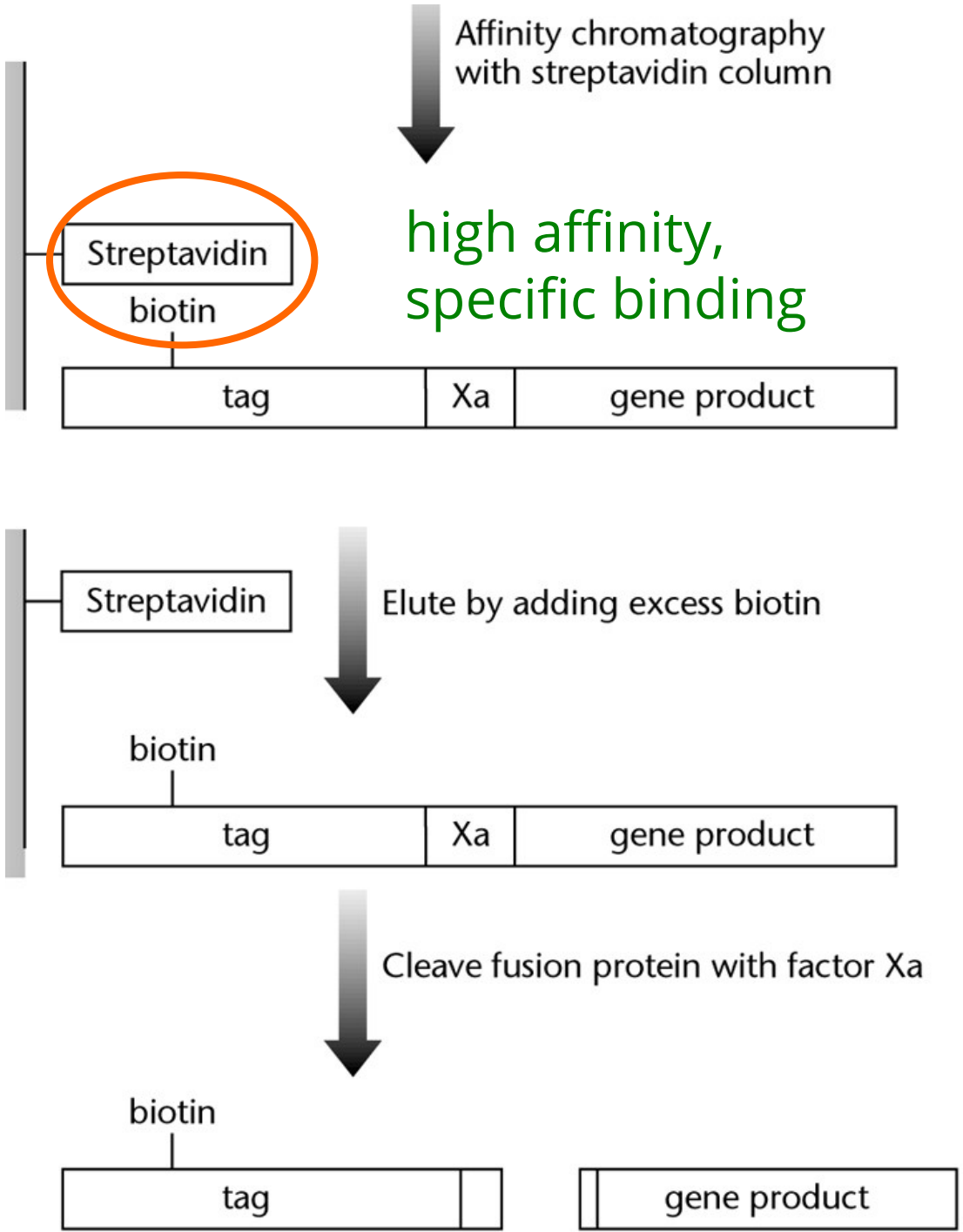
Biotin tag for protein purification



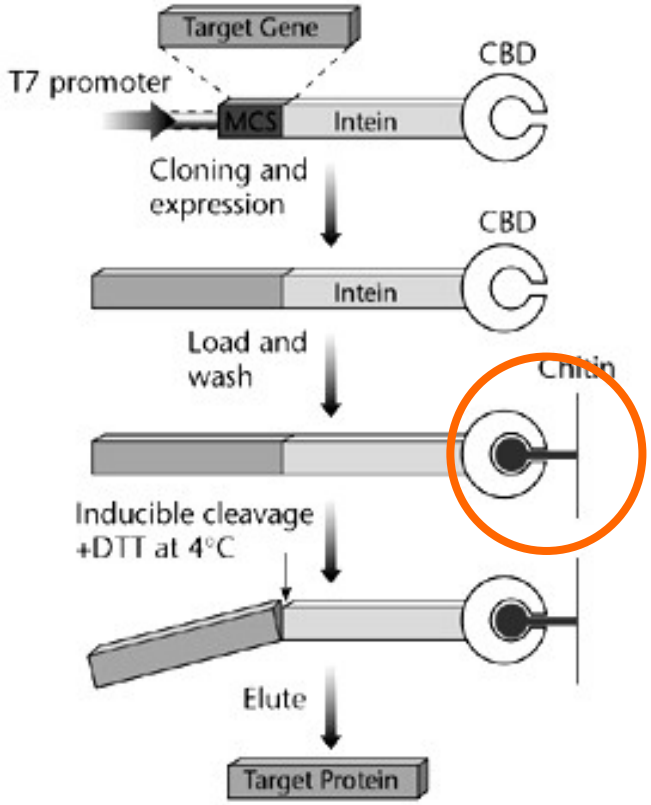
Translation product:

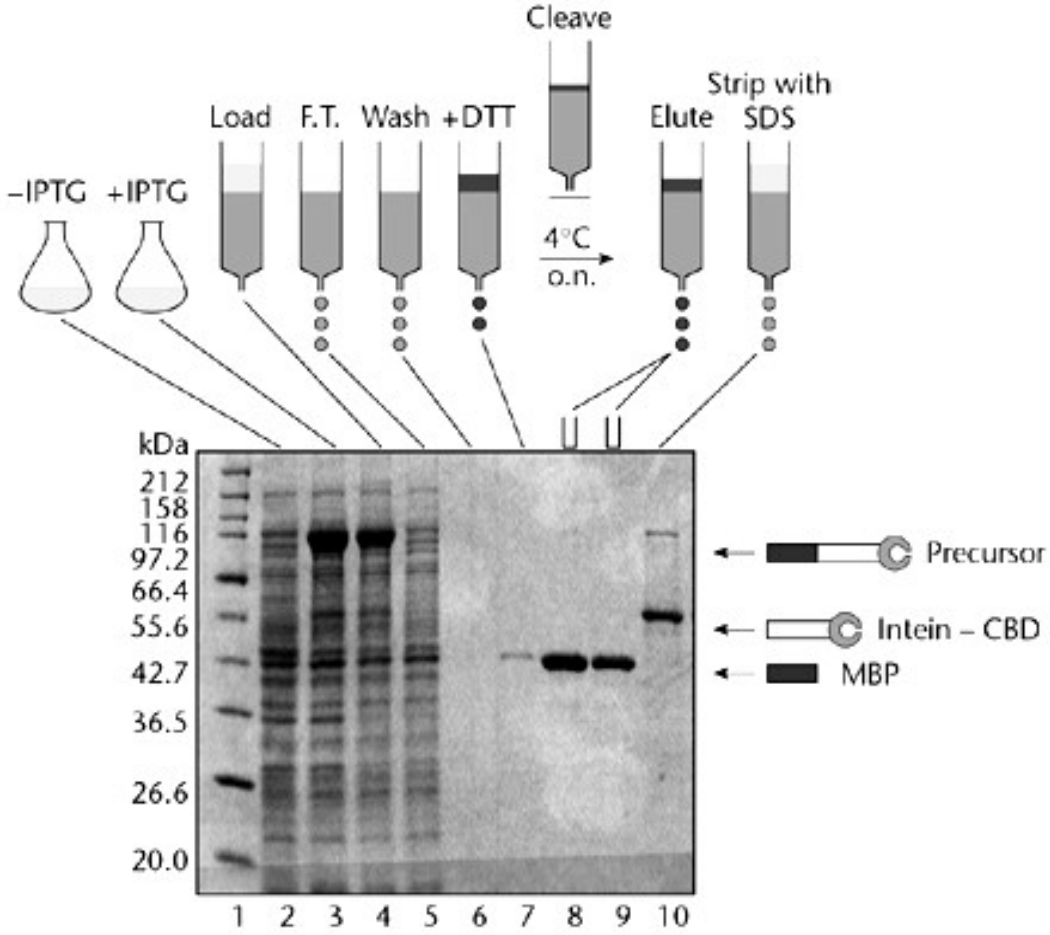
Junctions have to maintain reading frame for translation

Biotin tag for protein purification



Another protein purification scheme--removable tag





Lane 1: Protein Marker.
Lane 2: Crude extract from uninduced cells.
Lane 3: Crude extract from cells, induced at 15°C for 16 hours.
Lane 4: Clarified crude extract from induced cells.
Lane 5: Chitin column flow through (F.T.).
Lane 6: Chitin column wash.
Lane 7: Quick DTT wash to distribute DTT evenly throughout the chitin column.
Lanes 8-9: Fraction of eluted MBP after stopping column flow and inducing a self-cleavage reaction at 4°C overnight.
Lane 10: SDS stripping of remaining proteins bound to chitin column (mostly the cleaved intein-CBD fusion).

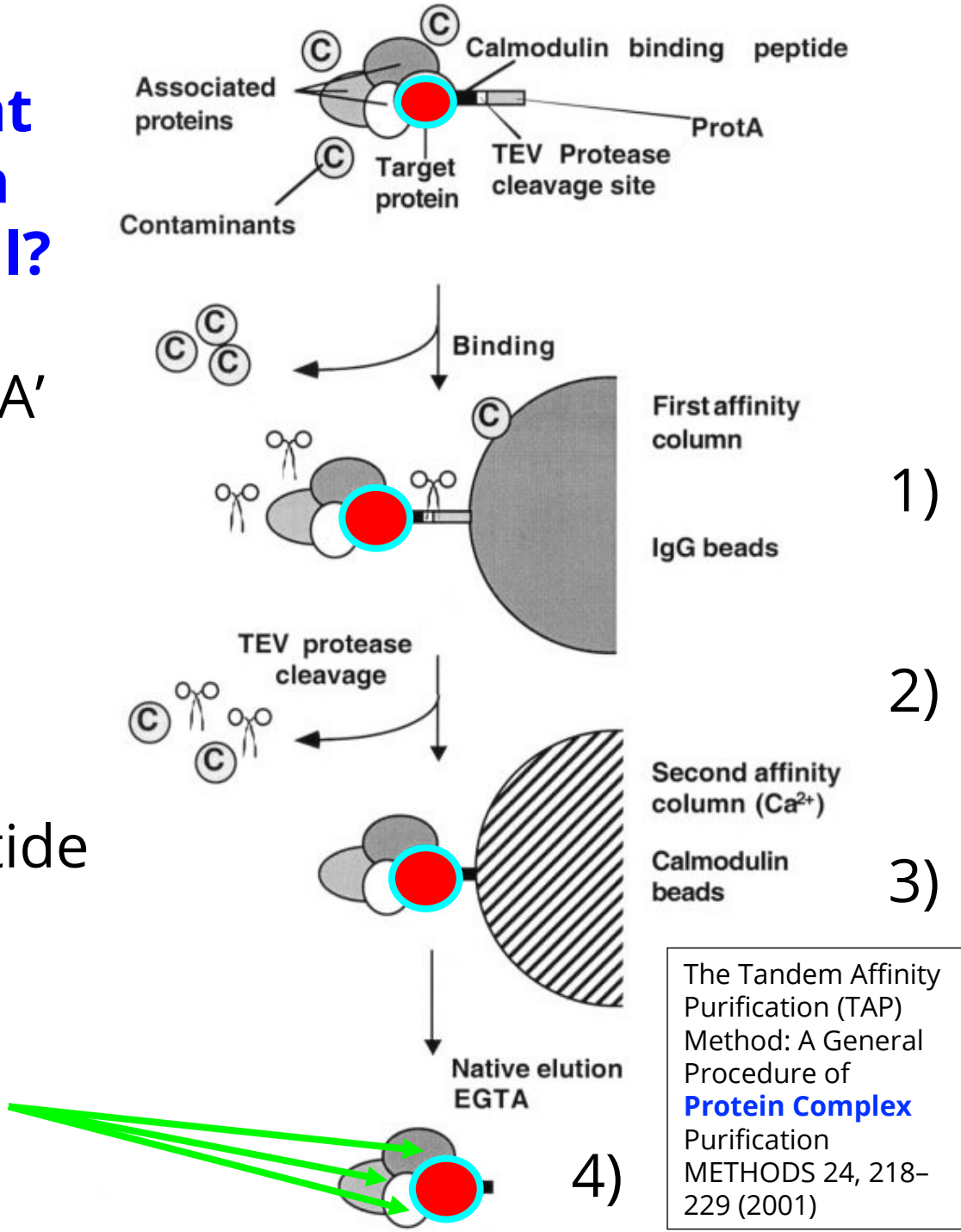
Tandem affinity purification (TAP): what does my target protein interact with in the cell?

1)IgG beads for ‘protein A’ tag

2)TEV protease cleavage

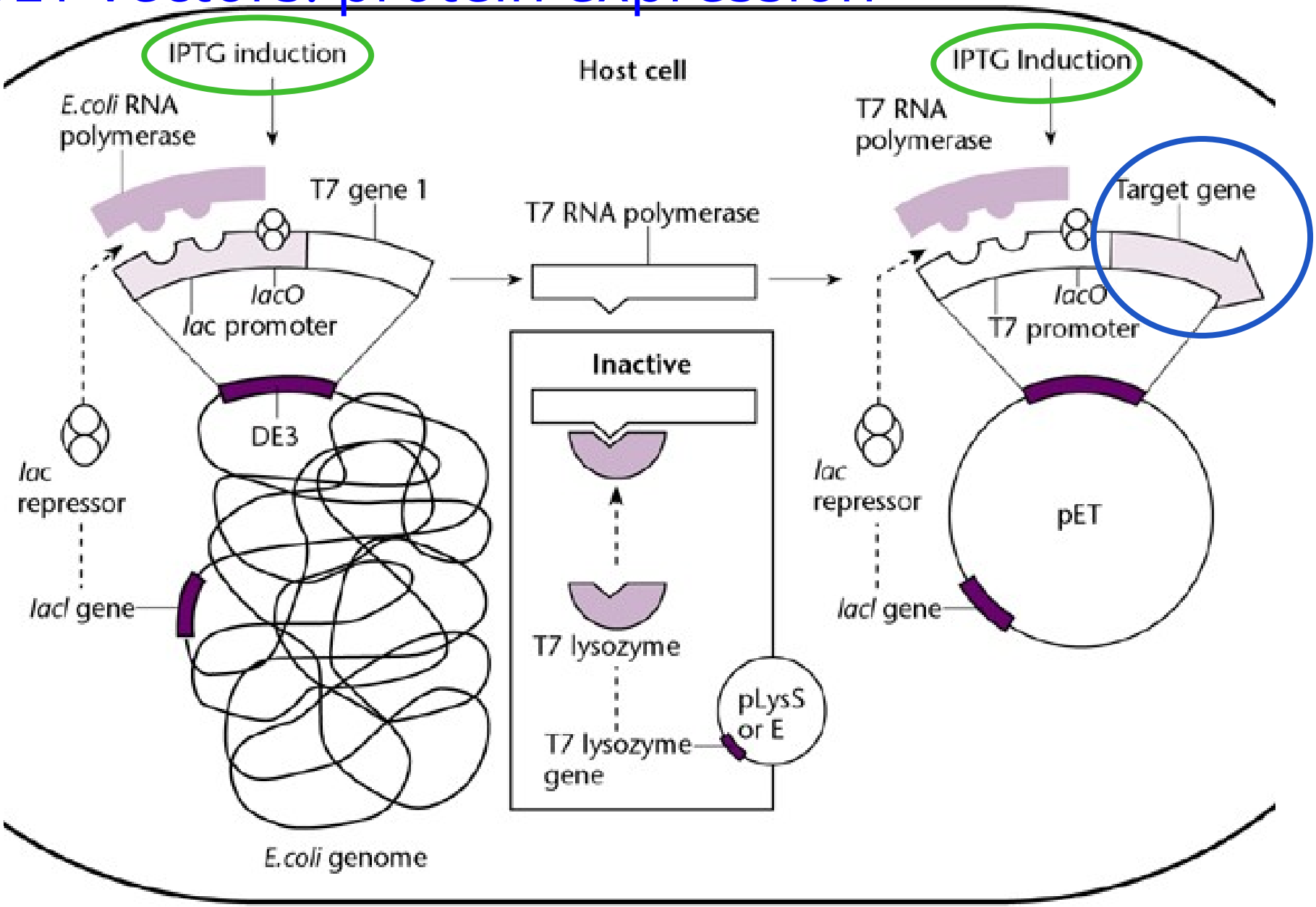
3)Calmodulin beads for calmodulin binding peptide

4)Pure protein and associated complex



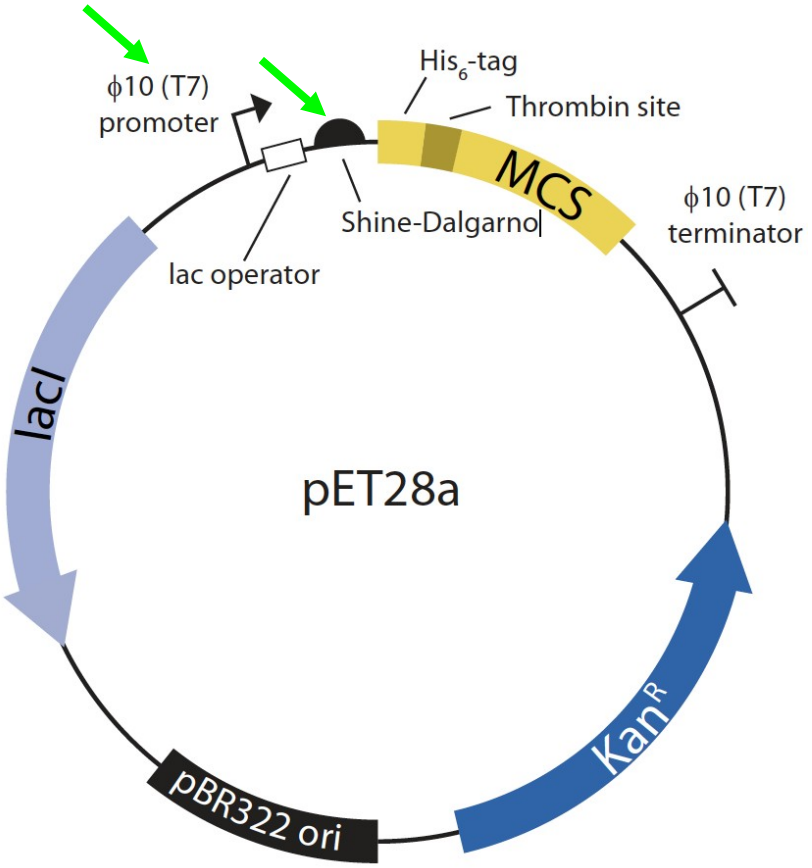
The Tandem Affinity Purification (TAP) Method: A General Procedure of **Protein Complex** Purification METHODS 24, 218–229 (2001)

pET vectors: protein expression

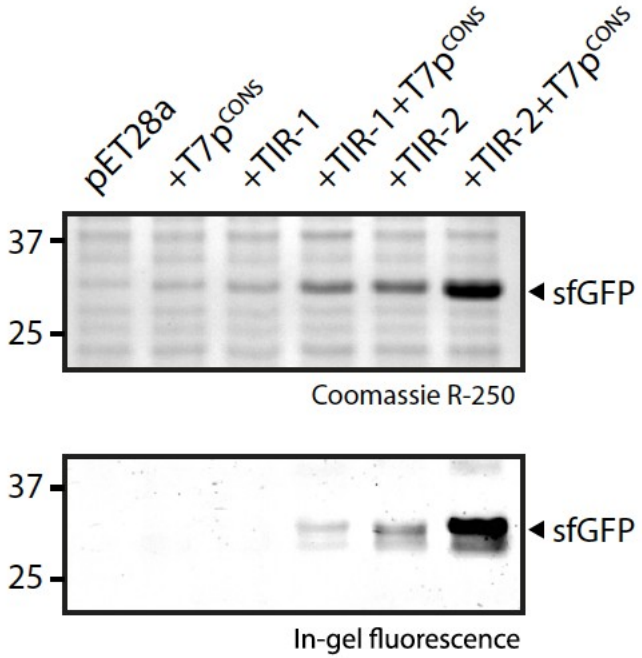


+IPTG = high T7 RNAP, high target gene expression

Updating pET vectors: new developments in 2020
How to get even more protein expression



1) Change T7 promoter to the ϕ10 consensus sequence
2) Alter translation initiation region (TIR) by searching through library of TIR mutants



Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*

Patrick J. Shilling[1✉], Kiavash Mirzadeh[1,2], Alister J. Cumming[1], Magnus Widesheim[1], Zoe Köck[1,3] & Daniel O. Daley[1✉]

COMMUNICATIONS BIOLOGY | (2020)3:214 | https://doi.org/10.1038/s42003-020-0939-8 | www.nature.com/commsbio

Purification of a 6xHis tagged protein, start to finish

1) Clone gene into pET vectors
   - Design primers, PCR
   - Choose vector, ligate gene in frame to vector
   - Sequence to verify the clone
   - Transform the plasmid into BL21 strain
     - T7 RNA polymerase gene
     - Lacks proteases

2) Overexpress the protein
   - Grow the cells to mid-log phase
   - Add IPTG to induce overexpression of the gene
     - Induction temperature may vary (15 – 42°C)
     - Concentration of IPTG varies (0.01 – 2.0 mM)
     - Time of induction varies (1 hour to overnight)

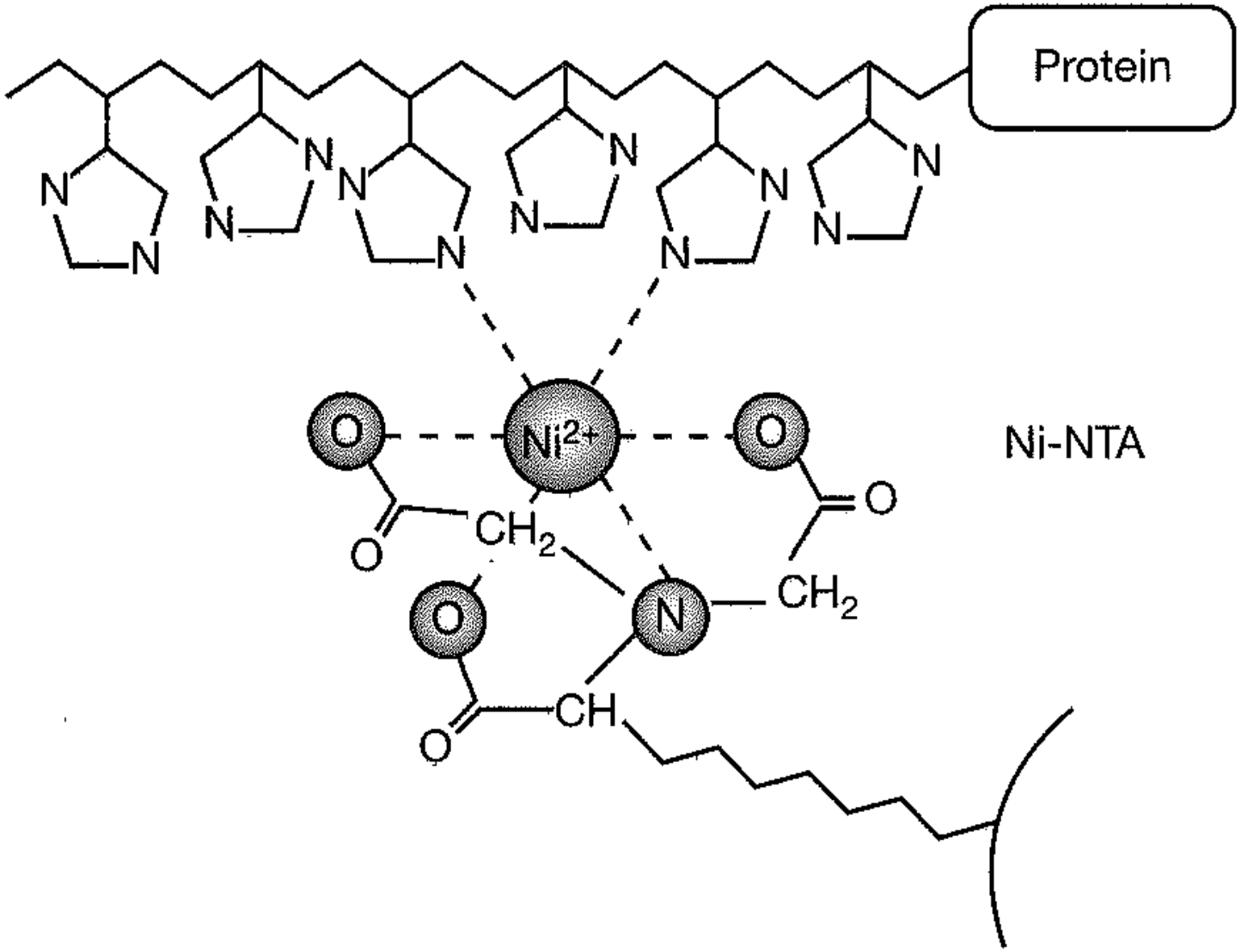
Purification of a 6xHis tagged protein, start to finish

3) Break open the cells
- lysozyme, sonication
- add protease inhibitors
- spin to get a clarified “lysate”

4) Affinity purification of protein
- apply lysate to a Nickel column or to Nickel beads
- wash with low concentration imidazole (20 mM)
- elute with high concentration imidazole (200 mM)

5) Check protein purity and concentration by SDS-PAGE

6xHis interaction with Ni^{++}

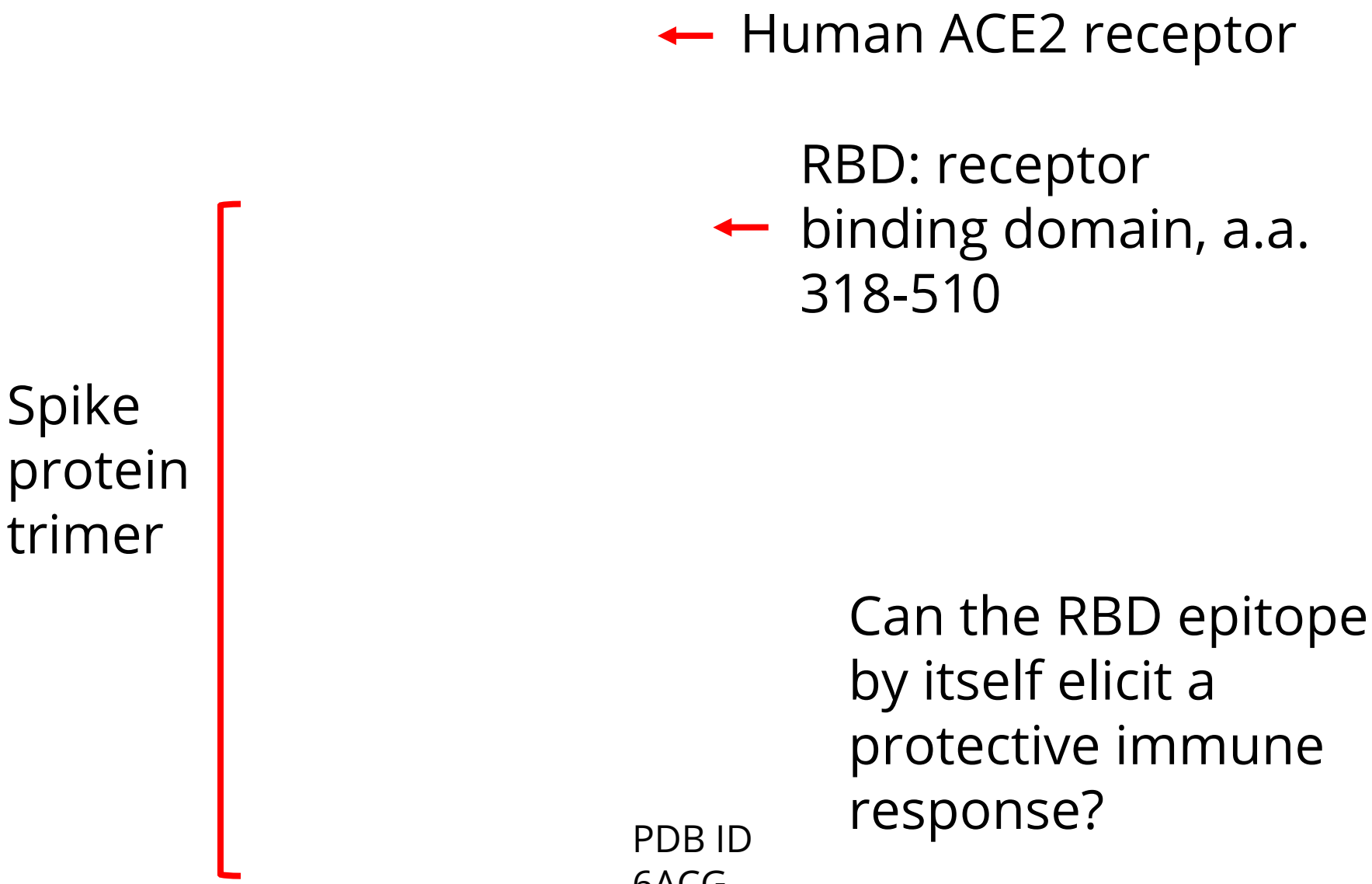


Protein purification and vaccines: SARS CoV-1 in 2009

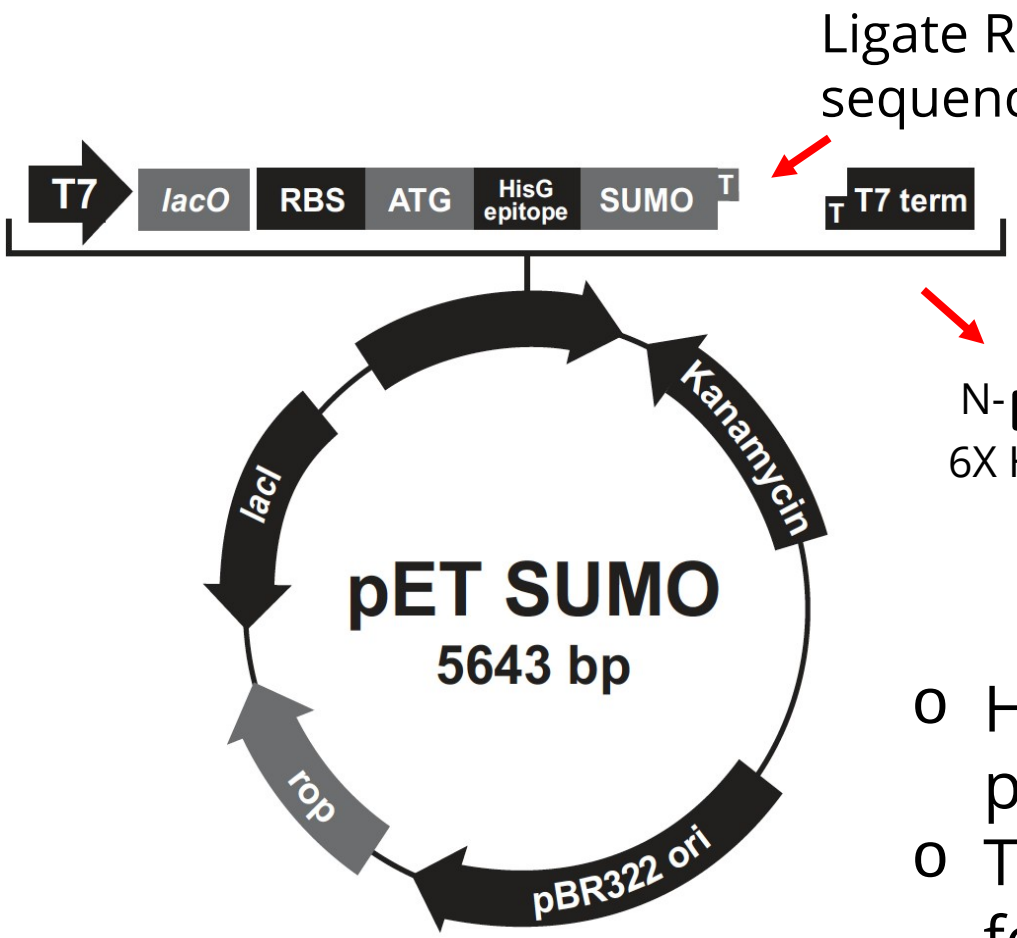
- Many different vaccines for SARS Cov-2 are in development now, for example:
  - Inactivated SARS CoV2 virus (Sinopharm, CoronaVac)
  - Vector-driven protein expression (Sputnik V)
  - Nucleic acid (RNA or DNA) vaccine (Moderna, Pfizer/BioNTech)
  - Antigenic proteins (especially the spike protein)

- Production of antigenic proteins?
  - Example: SARS CoV-1, outbreak in 2003, thousands of infections, hundreds of deaths
  - No new cases since 2004
  - Can spike protein epitope (including only the receptor binding domain) confer protective immunity?
  - Du et al. 2009, doi:10.1016/j.virol.2009.07.018

Structure of SARS CoV-1 spike with ACE2 receptor



Making the RBD: protein expression in *E. coli*



- Has a His tag for purification
- The SUMO motif assists in folding and solubility
- SUMO motif gets cut away by SUMO protease

Comparison of *E. coli* produced RBD to baculovirus and human cell line expressed versions

- RBD was also expressed human cell line, as well as in baculovirus infected insect cell line

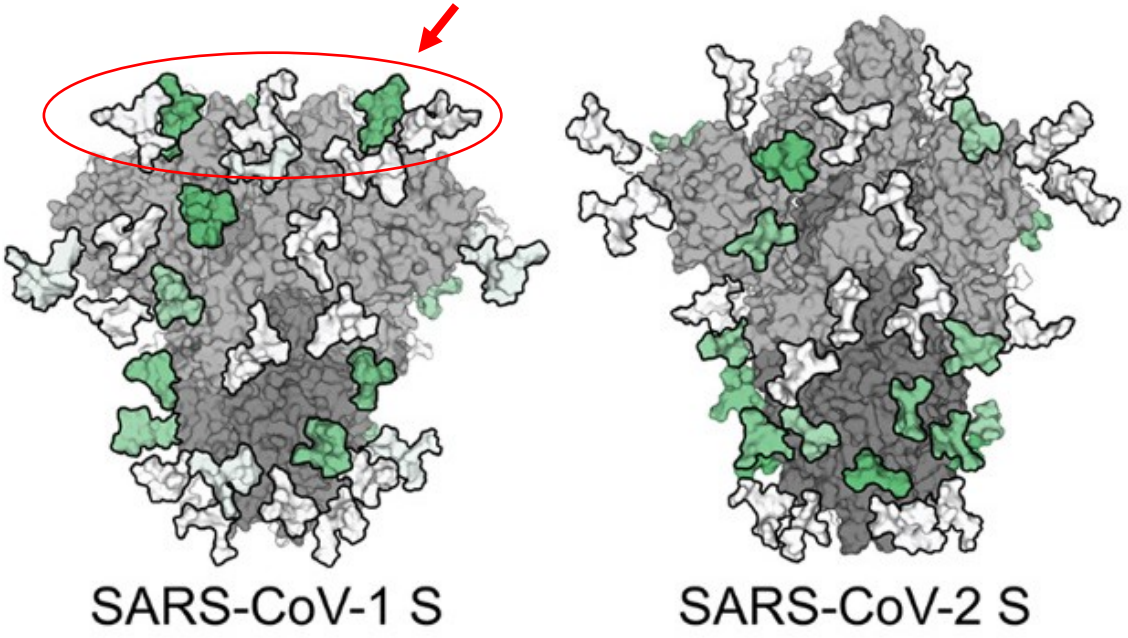
- All three versions were used to vaccinate mice
  - 20 micrograms purified RBG, followed by two boosts of 10 micrograms each in 3 week intervals
  - The mice were tested for:
    - humoral immune response (antibodies)
    - protective immunity

- All three yielded humoral (antibody-mediated) response, although less so with *E. coli*-made RBD
- All three vaccines conferred similar protective immunity (in mice) to SARS CoV-1 infection

Advantage and disadvantage to *E. coli* expression?

- Using *E. coli,* lots of doses of the vaccine could be made cheaply & easily
- However, the protein would not be glycosylated in *E. coli*, this may explain the lower humoral response in mice

Glycosylation map shows sites (in green and white) that overlap the RBD



Y. Watanabe et al., Science 10.1126/science.abb9983 (2020).

2020-21 update:

- Spike (RBD) protein currently being tested as low cost (~$1/dose) vaccine alternative, see Hotez & Bottazzi 2020

- Vaccine testing in progress in India

- Both SARS CoV and SARS CoV2 RBD proteins being tested

- Proteins produced using the yeast *Pichia pastoris*
  - can do glycosylation
  - Proteins secreted into growth medium, so easier to purify

Protein overexpression: pitfalls and limitations

1) Low or no gene expression

a) mRNA synthesis and stability:
   i. Premature transcription terminator
   ii. Unstable mRNA

b) Protein synthesis and stability
   i. Gene has lots of rare codons: if a low abundance codon for *E. coli* is used a lot in gene, translation rate will be slow (limited [tRNA])
      i. Check codon usage in *E. coli* at: http://people.mbi.ucla.edu/sumchan/caltor.html
   ii. Protein product is toxic to the cell: gene deletions occur
   iii. Protein is degraded by cellular protease

E. coli rpoB

ATGGTTTACTCCTATACCGAGAAAAAACGTATTCGTAAGGATTTTGGTAA
ACGTCCACAAGTTCTGGATGTACCTTATCTCCTTTCTATCCAGCTTGACT
CGTTTCAGAAATTTATCGAGCAAGATCCTGAAGGGCAGTATGGTCTGGAA
GCTGCTTTCCGTTCCGTATTCCCGATTCAGAGCTACAGCGGTAATTCCGA
GCTGCAATACGTCAGCTACCGCCTTGGCGAACCGGTGTTTGACGTCCAGG
AATGTCAAATCCGTGGCGTGACCTATTCCGCACCGCTGCGCGTTAAACTG
CGTCTGGTGATCTATGAGCGCGAAGCGCCGGAAGGCACCGTAAAAGACAT
TAAAGAACAAGAAGTCTACATGGGCGAAATTCCGCTCATGACAGACAACG
GTACCTTTGTTATCAACGGTACTGAGCGTGTTATCGTTTCCCAGCTGCAC
CGTAGTCCGGGCGTCTTCTTTGACTCCGACAAAGGTAAAACCCACTCTTC
GGGTAAAGTGCTGTATAACGCGCGTATCATCCCTTACCGTGGTTCCTGGC
TGGACTTCGAATTCGATCCGAAGGACAACCTGTTCGTACGTATCGACCGT
CGCCGTAAACTGCCTGCGACCATCATTCTGCGCGCCCTGAACTACACCAC
AGAGCAGATCCTCGACCTGTTCTTTGAAAAAGTTATCTTTGAAATCCGTG
ATAACAAGCTGCAGATGGAACTGGTGCCGGAACGCCTGCGTGGTGAAACC
GCATCTTTTGACATCGAAGCTAACGGTAAAGTGTACGTAGAAAAAGGCCG
CCGTATCACTGCGCGCCACATTCGCCAGCTGGAAAAAGACGACGTCAAAC
TGATCGAAGTCCCGGTTGAGTACATCGCAGGTAAAGTGGTTGCTAAAGAC
TATATTGATGAGTCTACCGGCGAGCTGATCTGCGCAGCGAACATGGAGCT
GAGCCTGGATCTGCTGGCTAAGCTGAGCCAGTCTGGTCACAAGCGTATCG
AAACGCTGTTCACCAACGATCTGGATCACGGCCCATATATCTCTGAAACC
TTACGTGTCGACCCAACTAACGACCGTCTGAGCGCACTGGTAGAAATCTA
CCGCATGATGCGCCCTGGCGAGCCGCCGACTCGTGAAGCAGCTGAAAGCC
TGTTCGAGAACCTGTTCTTCTCCGAAGACCGTTATGACTTGTCTGCGGTT
GGTCGTATGAAGTTCAACCGTTCTCTGCTGCGCGAAGAAATCGAAGGTTC
CGGTATCCTGAGCAAAGACGACATCATTGATGTTATGAAAAAGCTCATCG
ATATCCGTAACGGTAAAGGCGAAGTCGATGATATCGACCACCTCGGCAAC
CGTCGTATCCGTTCCGTTGGCGAAATGGCGGAAAACCAGTTCCGCGTTGG
CCTGGTACGTGTAGAGCGTGCGGTGAAAGAGCGTCTGTCTCTGGGCGATC
TGGATACCCTGATGCCACAGGATATGATCAACGCCAAGCCGATTTCCGCA
GCAGTGAAAGAGTTCTTCGGTTCCAGCCAGCTGTCTCAGTTTATGGACCA
GAACAACCCGCTGTCTGAGATTACGCACAAACGTCGTATCTCCGCACTCG
GCCCAGGCGGTCTGACCCGTGAACGTGCAGGCTTCGAAGTTCGAGACGTA
CACCCGACTCACTACGGTCGCGTATGTCCAATCGAAACCCCTGAAGGTCC
GAACATCGGTCTGATCAACTCTCTGTCCGTGTACGCACAGACTAACGAAT
ACGGCTTCCTTGAGACTCCGTATCGTAAAGTGACCGACGGTGTTGTAACT
GACGAAATTCACTACCTGTCTGCTATCGAAGAAGGCAACTACGTTATCGC
CCAGGCGAACTCCAACTTGGATGAAGAAGGCCACTTCGTAGAAGACCTGG
TAACTTGCCGTAGCAAAGGCGAATCCAGCTTGTTCAGCCGCGACCAGGTT
GACTACATGGACGTATCCACCCAGCAGGTGGTATCCGTCGGTGCGTCCCT
GATCCCGTTCCTGGAACACGATGACGCCAACCGTGCATTGATGGGTGCGA
ACATGCAACGTCAGGCCGTTCCGACTCTGCGCGCTGATAAGCCGCTGGTT
GGTACTGGTATGGAACGTGCTGTTGCCGTTGACTCCGGTGTAACTGCGGT
AGCTAAACGTGGTGGTGTCGTTCAGTACGTGGATGCTTCCCGTATCGTTA
TCAAAGTTAACGAAGACGAGATGTATCCGGGTGAAGCAGGTATCGACATC
TACAACCTGACCAAATACACCCGTTCTAACCAGAACACCTGTATCAACCA
GATGCCGTGTGTGTCTCTGGGTGAACCGGTTGAACGTGGCGACGTGCTGG
CAGACGGTCCGTCCACCGACCTCGGTGAACTGGCGCTTGGTCAGAACATG
CGCGTAGCGTTCATGCCGTGGAATGGTTACAACTTCGAAGACTCCATCCT
CGTATCCGAGCGTGTTGTTCAGGAAGACCGTTTCACCACCATCCACATTC
AGGAACTGGCGTGTGTCCCGTGACACCAAGCTGGGTCCGGAAGAGATC
ACCGCTGACATCCCGAACGTGGGTGAAGCTGCGCTCTCCAAACTGGATGA
ATCCGGTATCGTTTACATTGGTGCGGAAGTGACCGGTGGCGACATTCTGG
TTGGTAAGGTAACGCCGAAAGGTGAAACTCAGCTGACCCCAGAAGAAAAA
CTGCTGCGTGCGATCTTCGGTGAGAAAGCCTCTGACGTTAAAGACTCTTC
TCTGCGCGTACCAAACGGTGTATCCGGTACGGTTATCGACGTTCAGGTCT
TTACTCGCGATGGCGTAGAAAAAGACAAACGTGCGCTGGAAATCGAAGAA
ATGCAGCTCAAACAGGCGAAGAAAGACCTGTCTGAAGAACTGCAGATCCT
CGAAGCGGGTCTGTTCAGCCGTATCCGTGCTGTGCTGGTAGCCGGTGGCG
TTGAAGCTGAGAAGCTCGACAAACTGCCGCGCGATCGCTGGCTGGAGCTG
GGCCTGACAGACGAAGAGAAACAAAATCAGCTGGAACAGCTGGCTGAGCA
GTATGACGAACTGAAACACGAGTTCGAGAAGAAACTCGAAGCGAAACGCC
GCAAAATCACCCAGGGCGACGATCTGGCACCGGGCGTGCTGAAGATTGTT
AAGGTATATCTGGCGGTTAAACGCCGTATCCAGCCTGGTGACAAGATGGC
AGGTCGTCACGGTAACAAGGGTGTAATTTCTAAGATCAACCCGATCGAAG
ATATGCCTTACGATGAAAACGGTACGCCGGTAGACATCGTACTGAACCCG
CTGGGCGTACCGTCTCGTATGAACATCGGTCAGATCCTCGAAACCCACCT
GGGTATGGCTGCGAAAGGTATCGGCGACAAGATCAACGCCATGCTGAAAC
AGCAGCAAGAAGTCGCGAAACTGCGCGAATTCATCCAGCGTGCGTACGAT
CTGGGCGCTGACGTTCGTCAGAAAGTTGACCTGAGTACCTTCAGCGATGA
AGAAGTTATGCGTCTGGCTGAAAACCTGCGCAAAGGTATGCCAATCGCAA
CGCCGGTGTTCGACGGTGCGAAAGAAGCAGAAATTAAAGAGCTGCTGAAA
CTTGGCGACCTGCCGACTTCCGGTCAGATCCGCCTGTACGATGGTCGCAC
TGGTGAACAGTTCGAGCGTCCGGTAACCGTTGGTTACATGTACATGCTGA
AACTGAACCACCTGGTCGACGACAAGATGCACGCGCGTTCCACCGGTTCT
TACAGCCTGGTTACTCAGCAGCCGCTGGGTGGTAAGGCACAGTTCGGTGG
TCAGCGTTTCGGGGAGATGGAAGTGTGGGCGCTGGAAGCATACGGCGCAG
CATACACCCTGCAGGAAATGCTCACCGTTAAGTCTGATGACGTGAACGGT
CGTACCAAGATGTATAAAAACATCGTGGACGGCAACCATCAGATGGAGCC
GGGCATGCCAGAATCCTTCAACGTATTGTTGAAAGAGATTCGTTCGCTGG
GTATCAACATCGAACTGGAAGACGAGTAA

Archaeal RNAP subunit B

ATGAGAGGTCCGACTGTTGTAGATGTTACTCCCGACGATCTTTGGATTGT
AATGGAGAGCTATTGGCAAGAGAAAGGGCTTGTGAGACAGCATCTTGACT
CATACAATGCCTTTATTGAGAGAGGACTTCAAGAGGTGGTAAATGAGTTT
GGTGGAATAAAGCCGGATATTCCAGACTTTGAAGTAAAGTTCGGTAAGAT
TAGAGTAGGAGAACCCGAGTTTCAAGAACCTCATGGTCAAAGAAAGCCCC
TTTATCCAATGGATGCAAGAATAAGAAACCTAACCTACGCCGCACCACTC
TACCTAGAAATGATTCCAGTTATAAAAGGAGTTGAGCAAGATCCAGTAGA
GGTTAGAATTGGAGAACTTCCAATAATGCTAAAATCTAAAGTTTGTAGGC
TGTACGGACTTAGCGATGAAGAATTAATAAAGCTTGGAGAGGATCCCAAG
GATCCAGGTGGTTACTTCATAATCAACGGTTCTGAAAGAGTTATAGTCTC
TATTGAAGACATAGCTCCCAATAGGACTTTGGTAGAGAAGGACGAAAGGC
AAGAAAAGTACATAGCAAAGGTCTTTTCTTATAGACATGGGTATAGGGCT
TTAGTTGCTGTTGAGAGAAGAAAGGATGGAATTCTTTACGTTGATATACC
TAACGTACCTAAGCCTGTTAAATTCGTTTATGTGATGAGAGCTCTTGGGC
TGGAGAGAGATAAGGATATCGTTGAGGCAGTAGGTAATGATCCAGAAGTT
CAACAAATTATGTTTGATAACCTTGAGGATGCCAGCGATATAACTACCCA
ACAAGAGGCTCTTGAATTCATAGGTAAGCTTGTAGCTCCTGGGCAGGCCA
GAGAATACAGACTTAAGAGGGCAGAATATGTAATTGATAACAACTTGTTG
CCTCACATGGGTGTTTCTCCAGAGGATAGAATAAAGAAGGCTTACTACCT
GGGAATGATGGCTCTTAAGGTCATCGAGCTCTCCCTCGGAAGAAGACCTG
AAGATGACAAAGATCATTATGCAAATAAGAGGCTAAGGCTGGCTGGAGAT
CTTCTAAAGGACTTGTTCCGTGTAGCATTTTCTCAACTTGTTAAAGACAT
ACAATATCAAATGACCAAGACATACCAAAGAAAAGGGGACAAATATACTT
TTGGAAACATCCATAGATTTATTAGGAACTCCATAAGGCCAGACGTTCTC
ACCGAGAGAATTGAACATGCTCTTGCAACTGGAGCTTGGCCTGGTGGGAG
AACGGGTGTCAGTCAGTTGCTAGATAGAACTAACTACATGTCTACATTAT
CCCACCTTAGAAGAGTCACGTCTCCACTAGACAGAGAACAGCCCCACTTC
GAGGCTAGAGACCTTCACGGAACTCACTGGGGAAGAATTTGTCCAACGGA
GACTCCTGAAGGTCCAAACTGTGGTCTTGTGAAGAATTTAGCATTGATGG
CTCAAATAACAACAGCCGTTCCTGAAGAGGAGATTAGGGAGTACTTAATG
TCTTTGGGAATAGTTCCAATTGAAGTTAGAAGACCAGATCCAACATTGTG
GAGAGTTTACCTTAATGGTGTCCTCATCGGAACAATAGAGGATGGAAAAG
CTTTAGTTCAGAGAATTAGGCAAGACAGAAGAAGTGGAAAGATCAGCGAT
GTTATCAATGTGGCATACTATGAGGACGTTAAAGAAGTCTACGTTAACAG
CGACGATGGTAGGGTTAGAAGGCCTCTAATCATTGTTGAAAATGGCGTTC
CTAAACTAACTAAAGAGCATGTTCAAGCAGTTAAAGAAGGAAGGCTCAAG
TGGAGTGACTTAATTAAAATGGGAATTATCGAATACCTTGACGCAGAAGA
GGAGGAAAACGCATATGTTGCAATGTGGCCCTGGGAAGTTACTGAAGAGC
ACACTCACTTAGAATTAATGCCTGCAGCCATTCTTGGCCTACCGGCTTCC
CTTGTTCCCTATCCTGAACACAACGCAGCTCCAAGAAACACGTATGGAGC
AGGTATGGCCAAGCAGAGCCTTGGACTTGGATGGGCAAACTTTAGAATTA
GGGTAGATACTAGAGGTCACTTACTTCATTATCCACAGATTCCACTGGTG
AACTCAAGAATAATGAAAGCAGTAGGATATGAGGATAGGCCTGCTGGTCA
GAACTTCGTGGTTGCAGTCTTGGCTTACCAAGGTTATAACATGGAGGATG
CTATCATAATAAACAAGGCCTCCATAGAGAGAGGACTCGCCAGATCAACT
TTCTTCAGAACGTATGAAGCTGAGGAGAAGAAATACCTTGGAGGTCAGAC
TGACAAGTTTGAAATTCCGGATCCTACTGTAAGAGGTTACAGAGGAGAAA
AGTACTATAGAAACCTTGATGAAGATGGTCTCATATTCCCAGAATCAAAA
GTCGAGGGGAAGGACGTTTTAGTAGGAAGAACTTCTCCACCAAGGTTCCT
TGAAGAGAGAGGTCTTGGAACTGTGTCTCTTCAGGAGAGAAGAGAGACTA
GTGTGGCTGTAAGGCCGAGTGAGAAGGGAGTTGTTGATAAAGTAATAATA
ACCGAGACGGGAGATGGAACTAAGTTAGTTAAGGTTACTGTGAGAGACTT
GAGAATCCCAGAACTCGGAGACAAATTCGCTTCAAGACATGGGCAGAAAG
GTGTCATTGGACTTATTGTGCCCCAGGAGGACATGCCTTGGACCGAGAGT
GGAATTGTACCCGATCTCATAGTTAATCCACACGGTATACCATCAAGAAT
GACCGTGGGTCAGCTAATTGAAGCTATAGGTGGTAAGGTTGCTTCACTTA
AGGGAAGGAGAGTTGATGGAACAGCATTTATCGGAGAACCAGAAGAGAAG
CTAAGAAAAGAGCTAGAAGAGCTTGGATTCAAGCATACGGGTAGAGAAAT
AATGTATGATGGTATAACAGGGAAGAGATTAGAGGCTGATATATTCATTG
GTGTGATCTACTATCAGAGGCTCCACCACATGGTTGCAGATAAGATACAT
GCACGTTCAAGAGGTCCTGTCCAAGTTCTAACTAAGCAACCGACGGAAGG
TAGAGCAAGAGAGGGTGGTCTAAGATTTGGAGAAATGGAGAGAGACGTTC
TTGTTGGCCATGGAGCTGCCATGTTACTAATAGAGCGTTTACTTGAGGAG
AGCGATAAGACTGAAGTATGGGTATGTGAGAACTGTGGACACATAGCACT
AGAGGACAAGAGAAGAAGAAGAGTCTACTGTCCTGTTTGTGGAGAAGAAG
AAAGAATTAGCAAAGTAGAGATGAGCTACGCGTTCAAATTGTTGCTTGAT
GAGTTGAAGGCCATGGTTATTAGACCTAAGTTAAACCTCTCAGAGAGGGT
GTGA

Protein overexpression: pitfalls and limitations

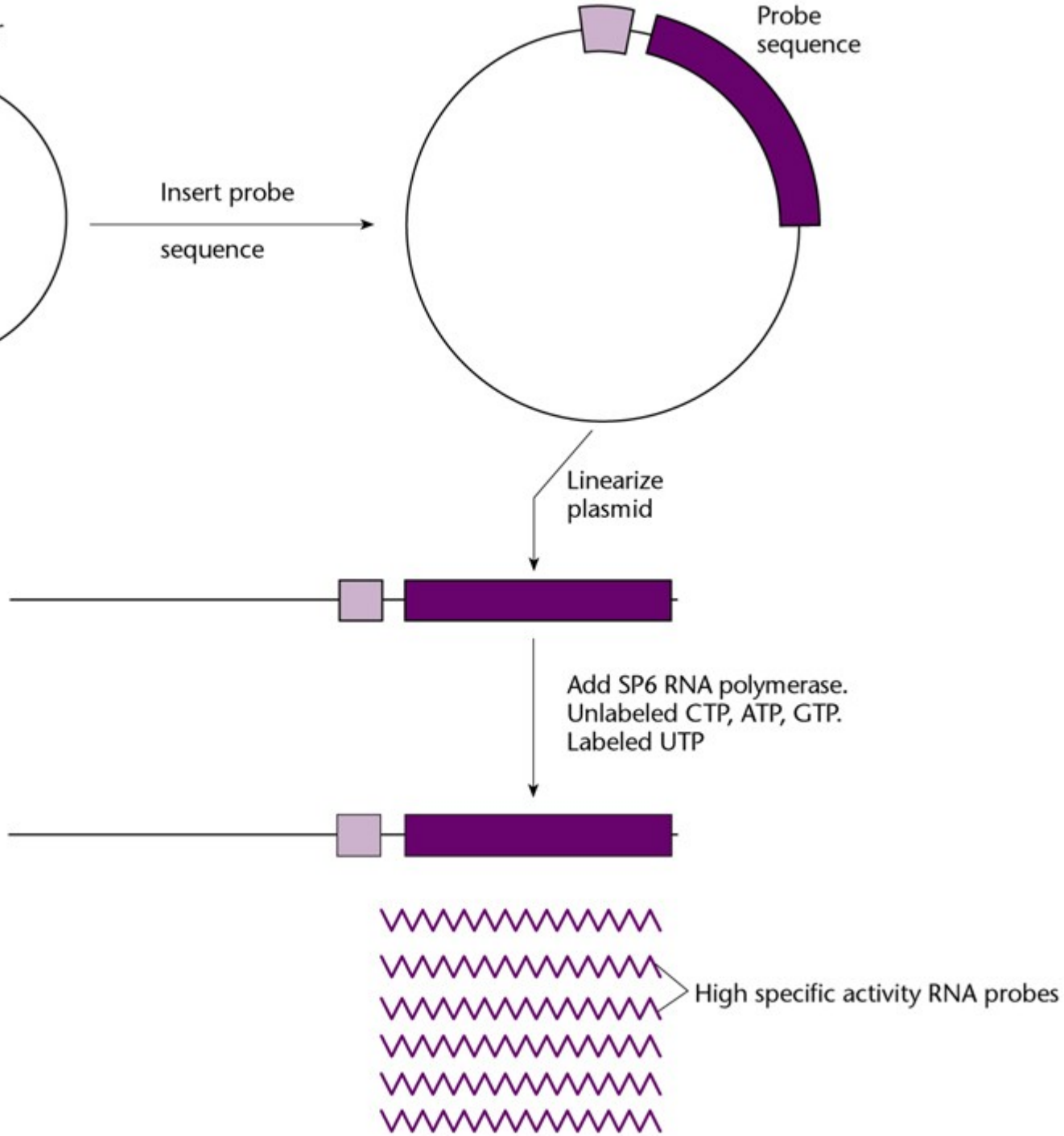
2) Protein product is not functional or hard to purify

a) *Inactive protein*
   i. Lacking necessary cofactor
   ii. Does not fold properly
   iii. Needs posttranslational modification
      - Glycosylation
      - Acetylation
      - Phosphorylation
   iv. Functions as part of a multi-protein complex

b) *Protein purification or detection difficulties*: hard to separate or distinguish from host proteins
   - Add epitope tag

RNA expression vectors:

in vitro RNA expression (from purified plasmids)



Therapeutic RNAs: RNA interference (RNAi)

- Micro RNAs and small interfering RNAs can be used to knock down expression of disease-causing genes.

- The current best RNAi therapeutic target is the liver
  - because of the liver's physiological role in removing particles from circulation, the RNAi therapeutic agent is very likely to reach it

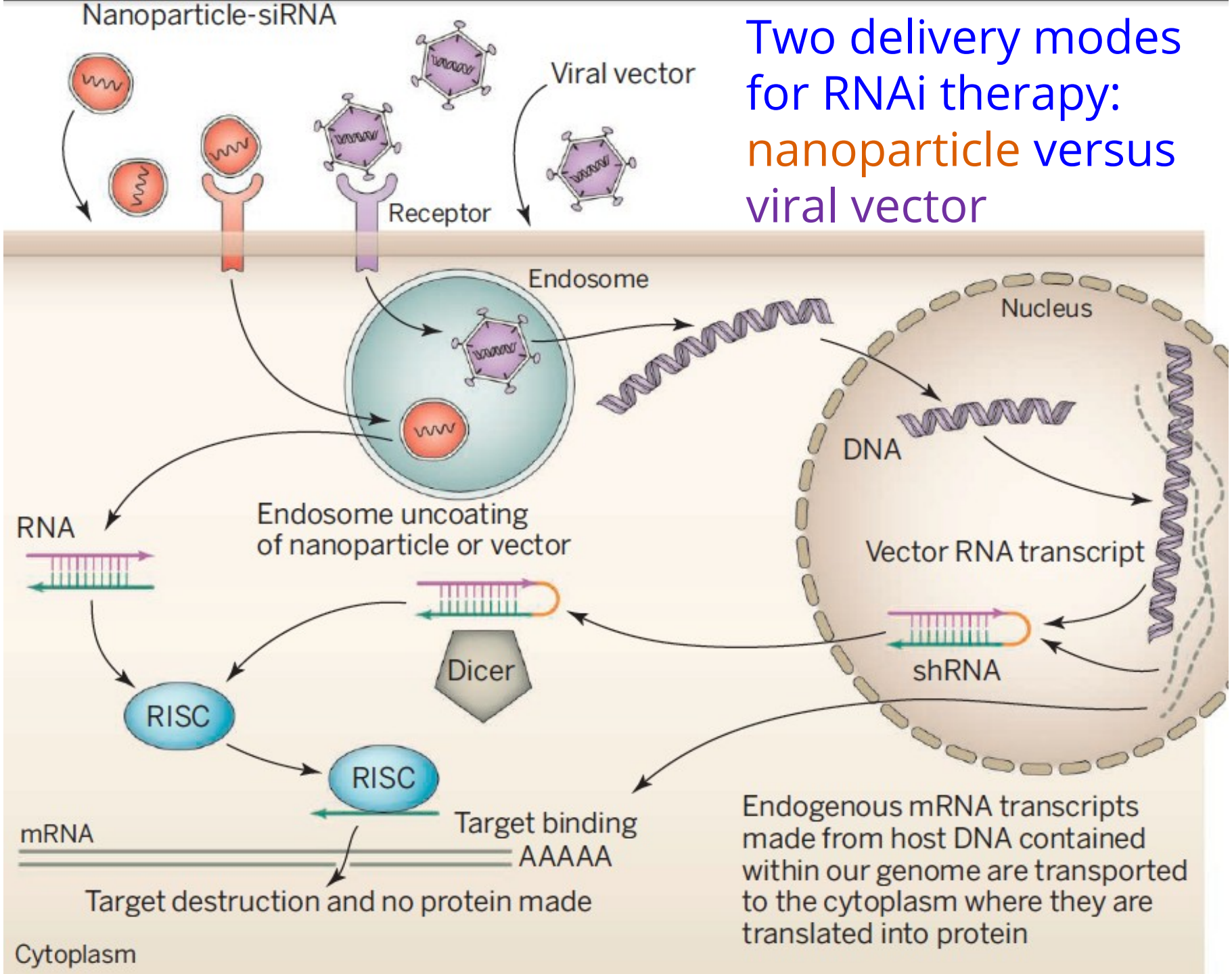
- Some diseases that can be targeted at the liver:
  - Transthyretin, causing amyloid neuropathy
  - Hepatitis B
  - Liver cancer
  - Hepatic porphyria (heme biosynthesis disorder)

Two modes of delivery of therapeutic RNAi

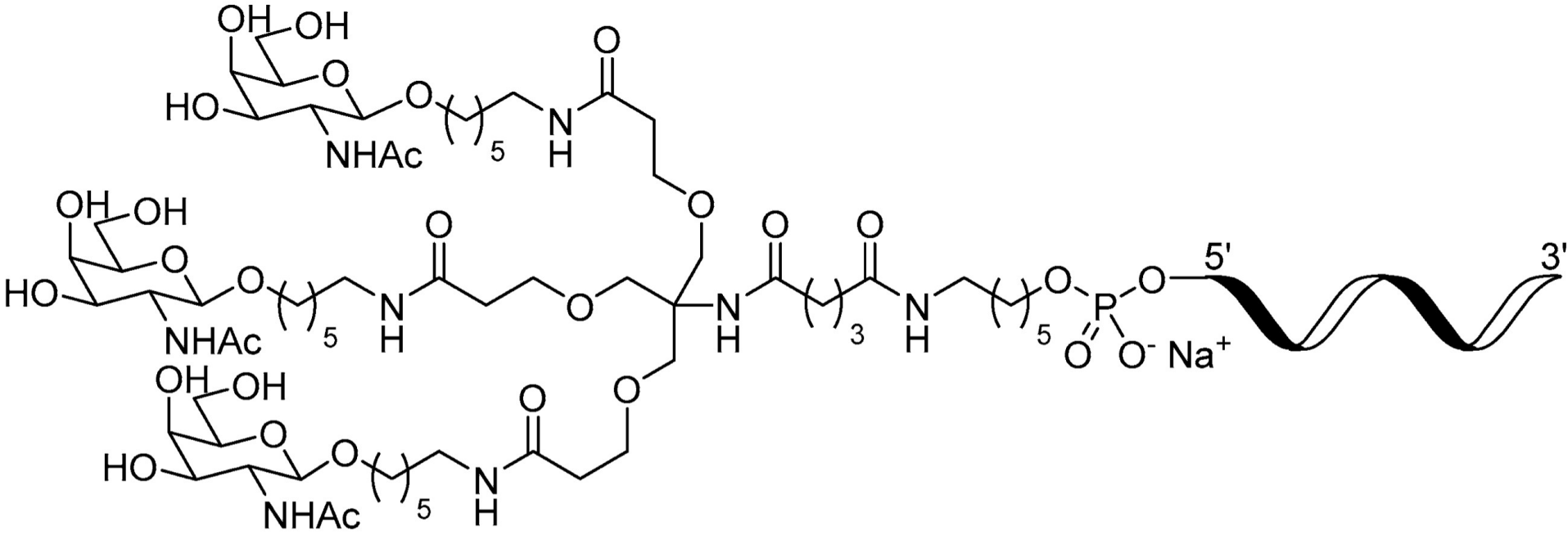
1) Premade RNAi precursors produced by *in vitro* transcription or chemical synthesis
   - Lipid nanoparticles (LNPs)
   - GalNac -RNA conjugates, which induce receptor mediated endocytosis

2) RNA expression vectors (recombinant virus) that make RNAi precursors upon entry into cells
   - Theoretically better for life-long treatments or hard to reach tissues
   - Precise control of expression is hard to achieve

Two delivery modes for RNAi therapy: nanoparticle versus viral vector

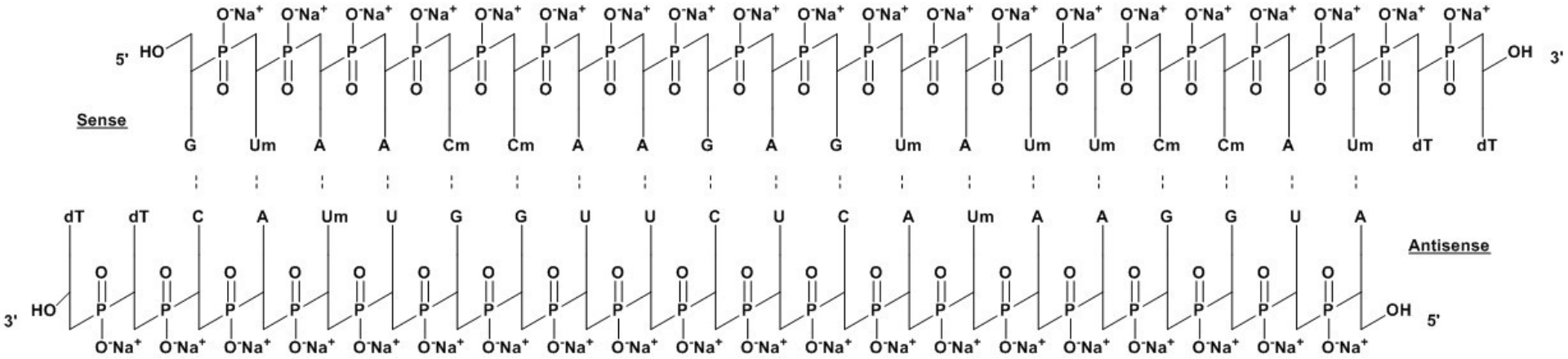


Nanoparticle (NP) example: N-acetylgalactosamine (GalNac)/RNA conjugate



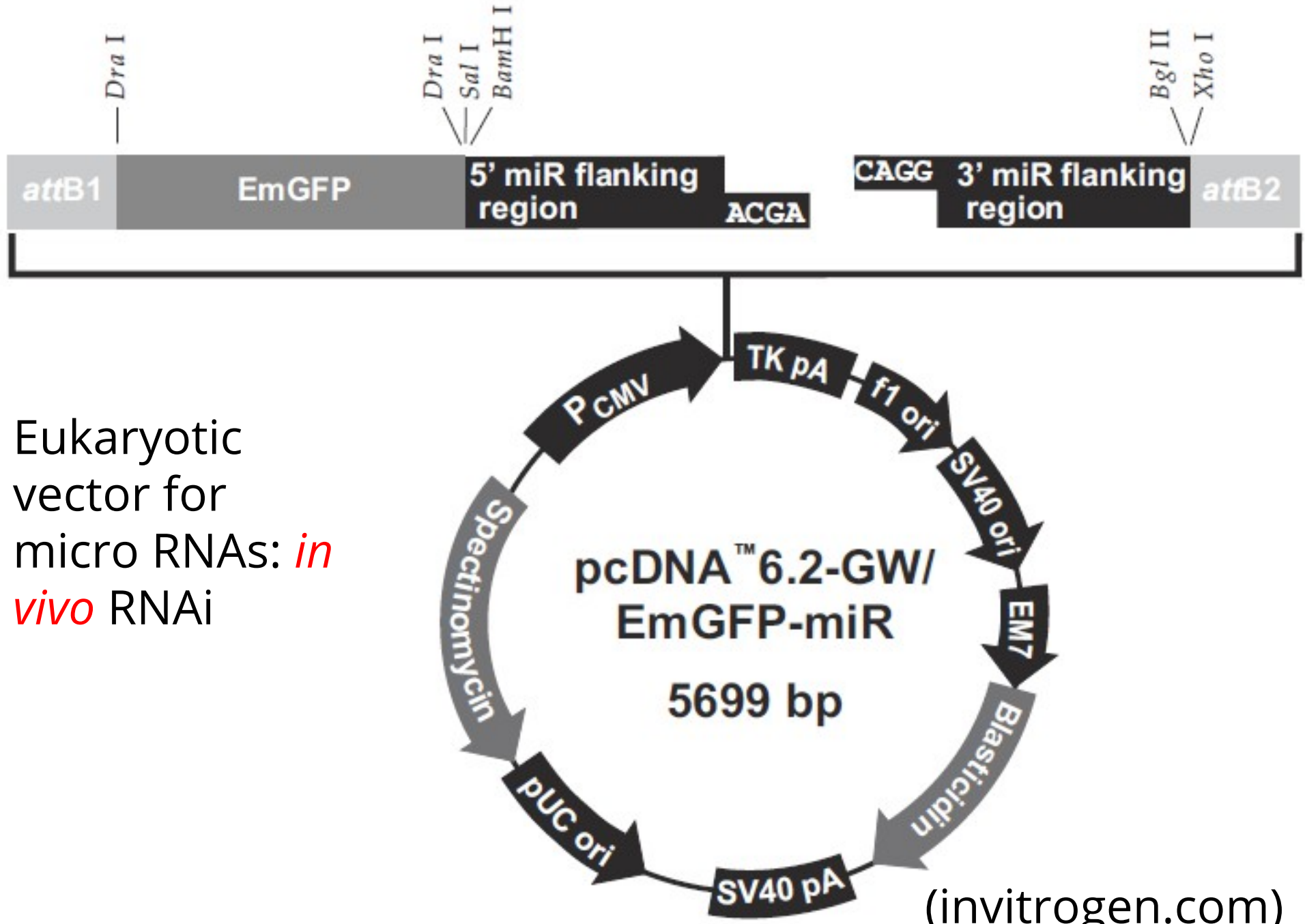
- Binds to the asialoglycoprotein receptor that is highly expressed on hepatocytes
- Results in rapid endocytosis

Patisiran: first RNAi drug approved by FDA (2018)



A, adenosine; C, cytidine; G, guanosine; U, uridine; Cm, 2’-O-methylcytidine; Um, 2’-O-methyluridine; dT, thymidine (modified nucleotides help tame dendritic cell immune response)

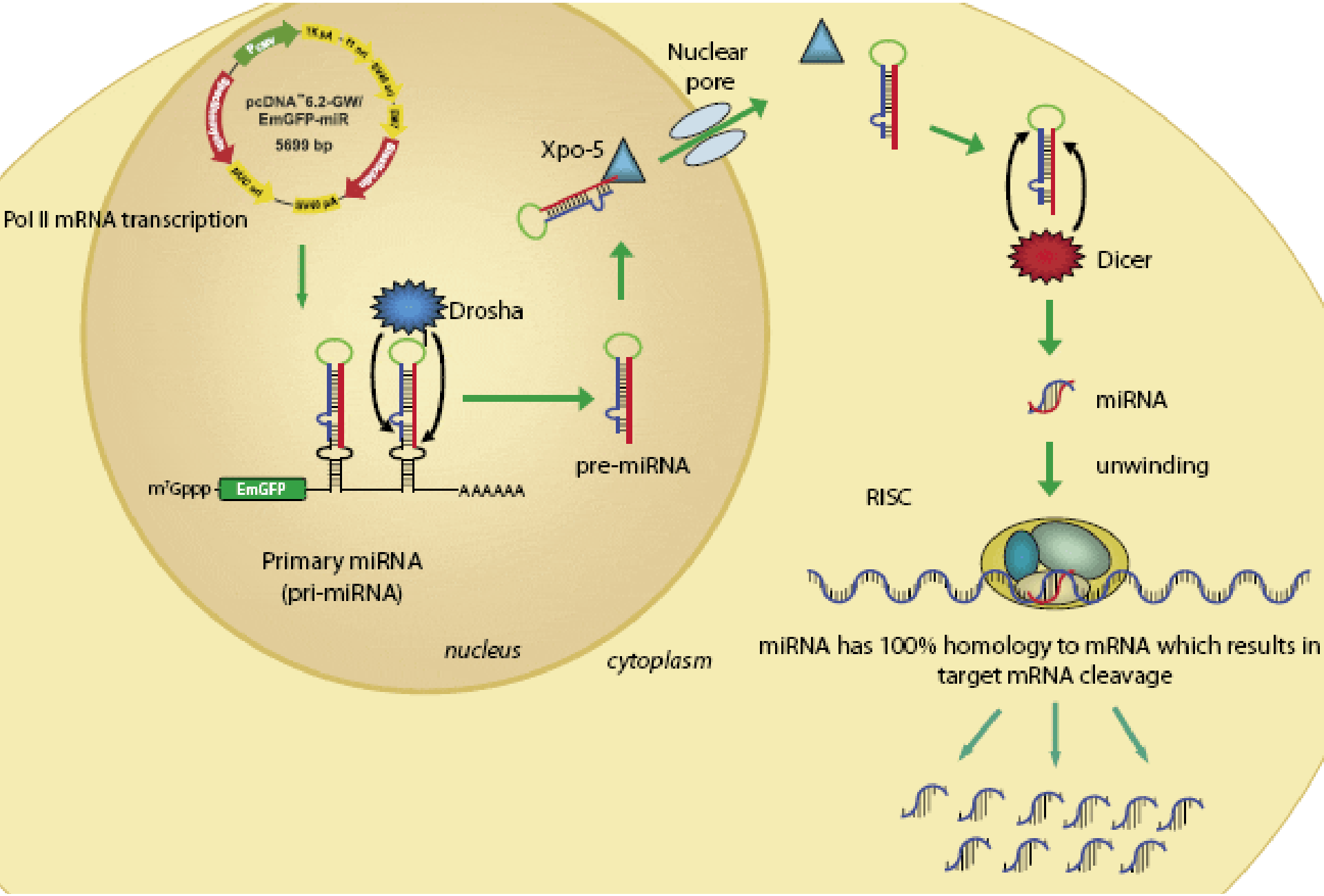
- Formulated as lipid NPs for delivery to hepatocytes
- RNAi specifically binds to a genetically conserved sequence in the 3’-untranslated region (3’-UTR) of mutant and wild-type transthyretin (TTR) messenger RNA (mRNA)
- Treats transthyretin amyloidosis (peripheral neuropathy)



Eukaryotic vector for micro RNAs: *in vivo* RNAi

(invitrogen.com)

Directed mRNA degradation by engineered miRNA



Gene cloning:
Expression of protein and RNA

Controlled gene expression

I. proteins
- Various tags and their functions
- A protocol for purification of 6xHis tagged proteins expressed in *E. coli*

II. RNA