TU Graz

MOL.911

Cell Engineering

Cell Engineering

General strategies:

- Knock out of specific genes
  - Gene disruption
  - Mutagenesis
- Down-regulation of specific genes
  - Antisense expression
  - Manipulation of regulatory elements
- Overexpression of homologous genes
- Integration and expression of heterologous genes
  - Co-expression of helper proteins
  - Introduction of metabolic pathway steps

Engineering of E.coli for Expression of Proteins by adding tRNAs for rare codons

Rosetta™ and Rosetta 2 host strains are BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*. (13–17).

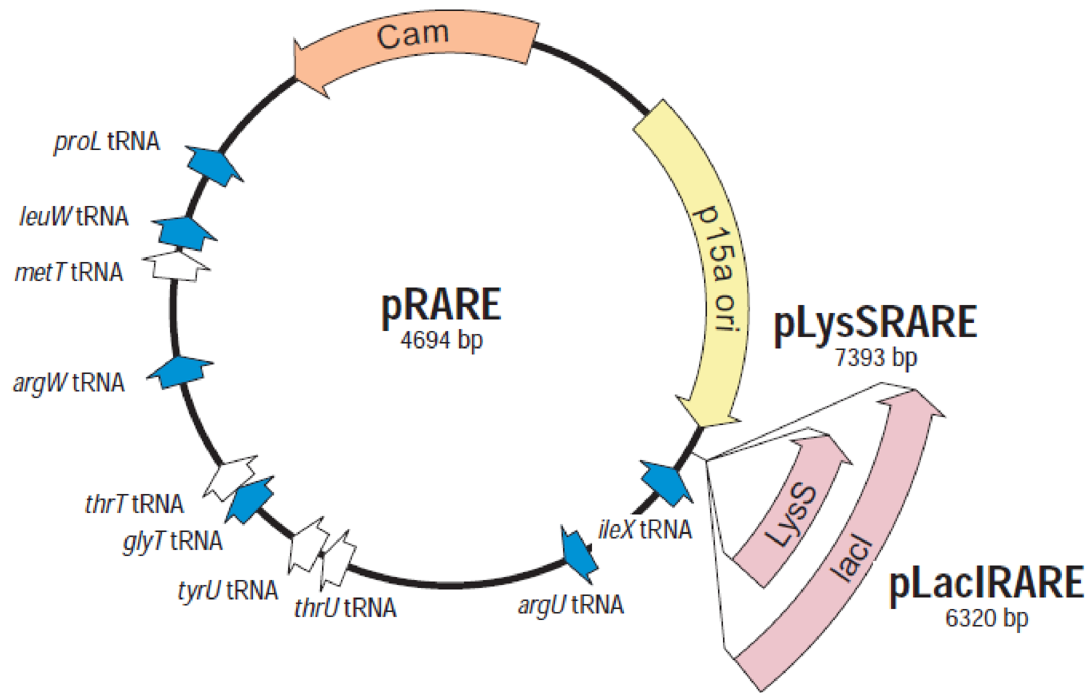
The original Rosetta strains supply tRNAs for the codons AUA, AGG, AGA, CUA, CCC, and GGA on a compatible chloramphenicol-resistant plasmid, pRARE (18).

The Rosetta 2 strains supply a seventh rare codon (CGG) in addition to the six found in the original Rosetta strains (19).

By supplying rare codons, the Rosetta strains provide for "universal" translation, where translation would otherwise be limited by the codon usage of *E. coli*. (15, 16, 20, 21).

The tRNA genes are driven by their native promoters (18).

Rosetta Strains



lysS

T7 lysozyme, a natural inhibitor of T7 RNA polymerase

Map of pRARE plasmid family
The basic structure of pRARE is indicated. pLysSRARE and pLacIRARE contain the genes encoding T7 lysozyme (LysS) and lac repressor
(lacI), respectively. Also indicated are chloramphenicol resistance gene (Cam), origin of replication (p15a ori) and tRNA genes. tRNA
genes corresponding to rare codons in E. coli are indicated in blue. pRARE is derived from pRIG (11).

*Novagen, in**Nova**tions 12*

lysS

T7 lysozyme, a natural inhibitor of T7 RNA polymerase

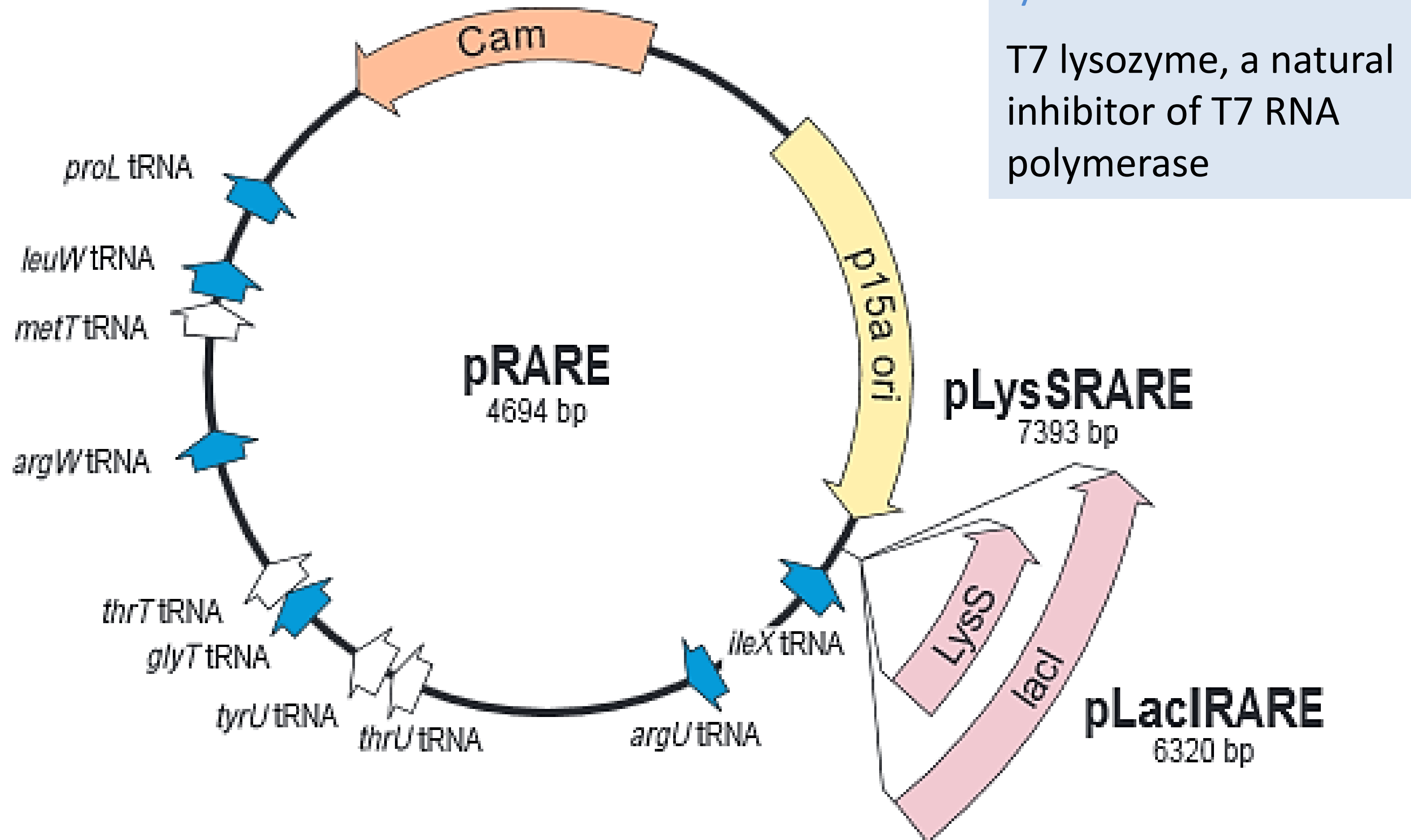


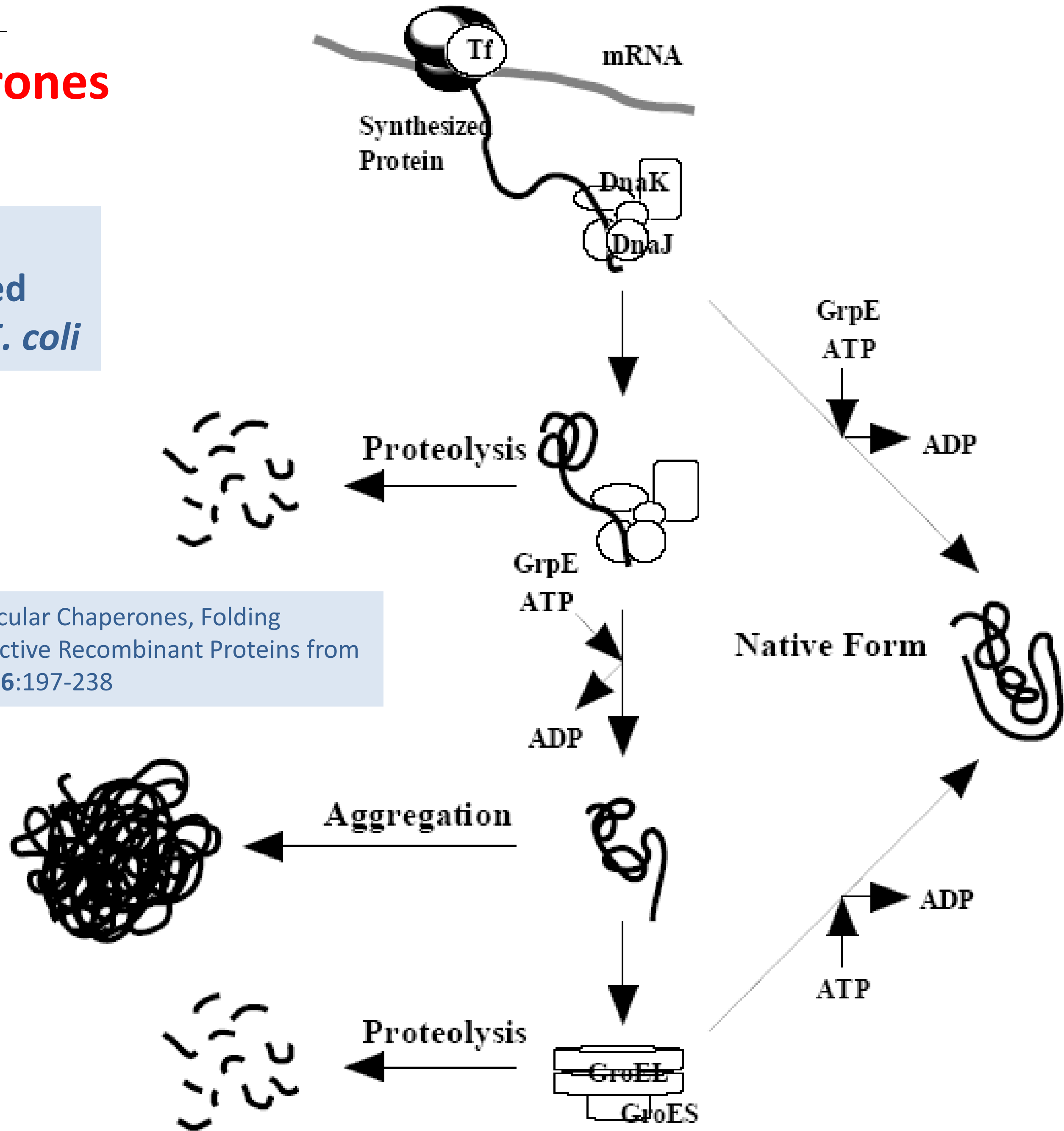
Figure 1. Map of pRARE plasmid family

The basic structure of pRARE is indicated. pLysSRARE and pLacIRARE contain the genes encoding T7 lysozyme (LysS) and *lac* repressor (lacI), respectively. Also indicated are chloramphenicol resistance gene (Cam), origin of replication (p15a ori) and tRNA genes. tRNA genes corresponding to rare codons in *E. coli* are indicated in blue. pRARE is derived from pRIG (11).

*Novagen, in**Nova**tions 12*

E.coli chaperones

Possible model for chaperone - assisted protein folding in *E. coli*



Thomas, J. G., et al.(1997) Molecular Chaperones, Folding Catalysts, and the Recovery of Active Recombinant Proteins from *E.coli*. Appl. Biochem. Biotech, **66**:197-238

http://www.takara-bio.co.jp

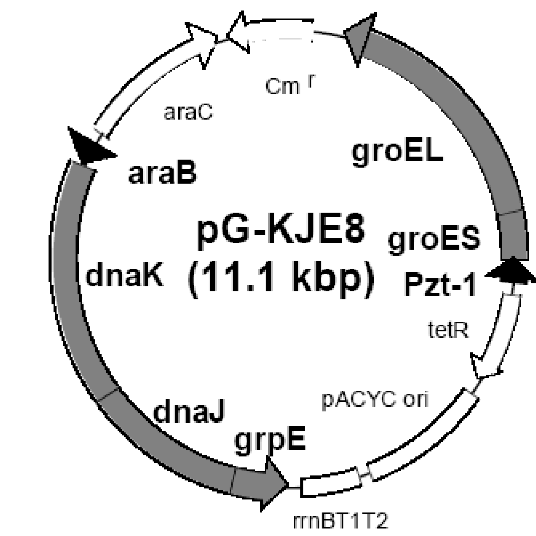
E.coli chaperones

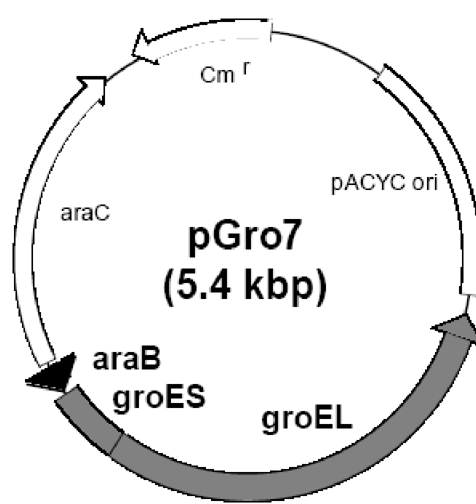
Chaperone Plasmid Set

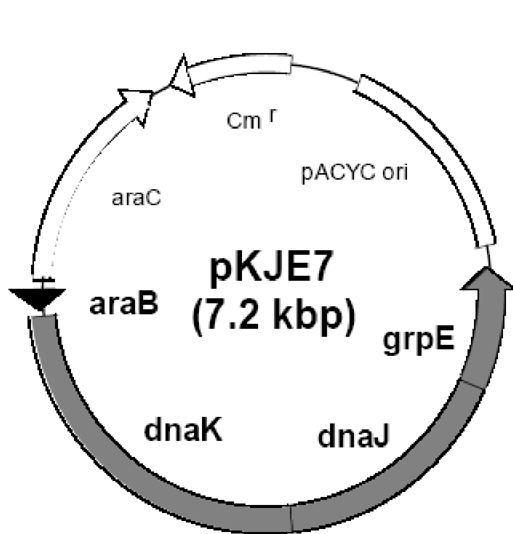
Cat.# 3340
v.0401

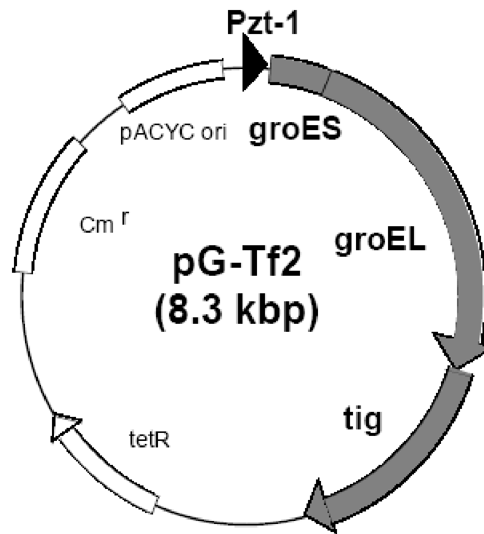
No.	Plasmid	Chaperone	Promoter	Inducer	Resistant Marker	References
1	pG-KJE8	dnaK-dnaJ-grpE groES-groEL	*araB* *Pzt1*	L-Arabinose Tetracyclin	Cm	2 , 3
2	pGro7	groES-groEL	*araB*	L-Arabinose	Cm	2
3	pKJE7	dnaK-dnaJ-grpE	*araB*	L-Arabinose	Cm	2
4	pG-Tf2	groES-groEL-tig	*Pzt1*	Tetracyclin	Cm	3
5	pTf16	tig	*araB*	L-Arabinose	Cm	3

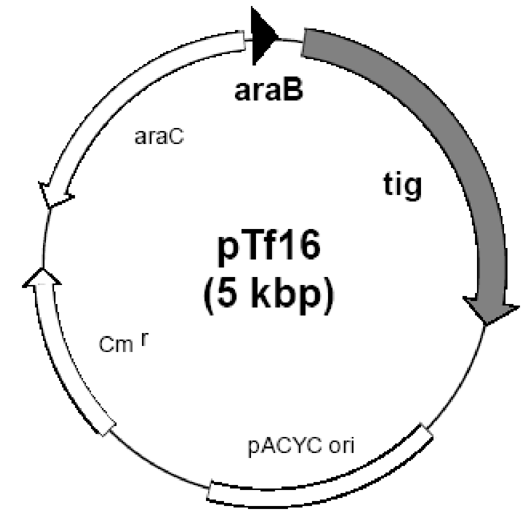
E.coli Chaperones





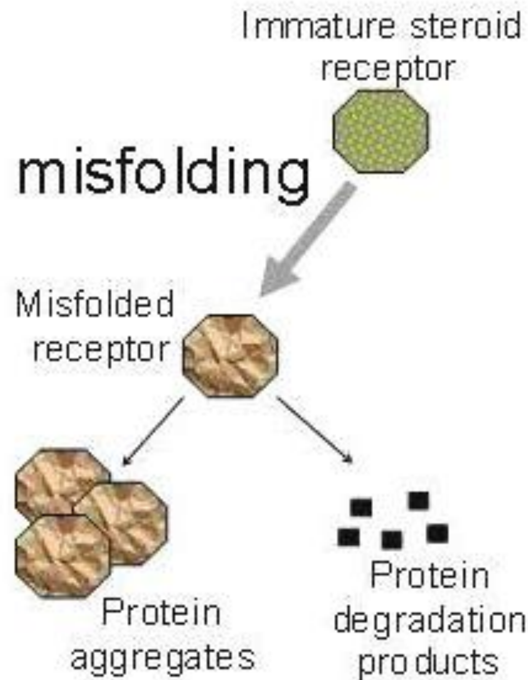


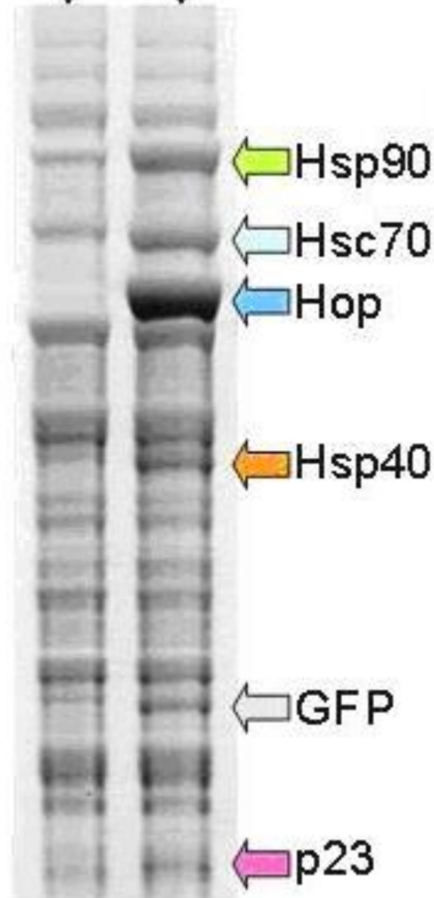


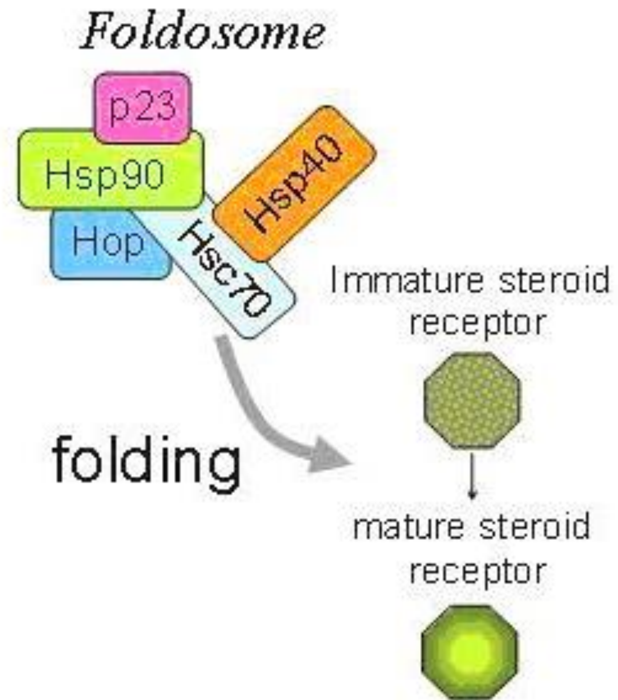


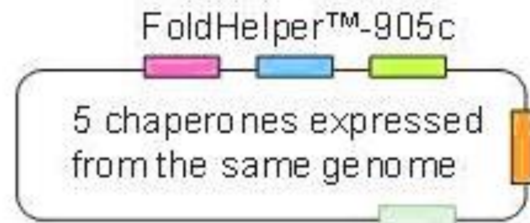
http://www.takara-bio.co.jp

Mammalian Chaperones





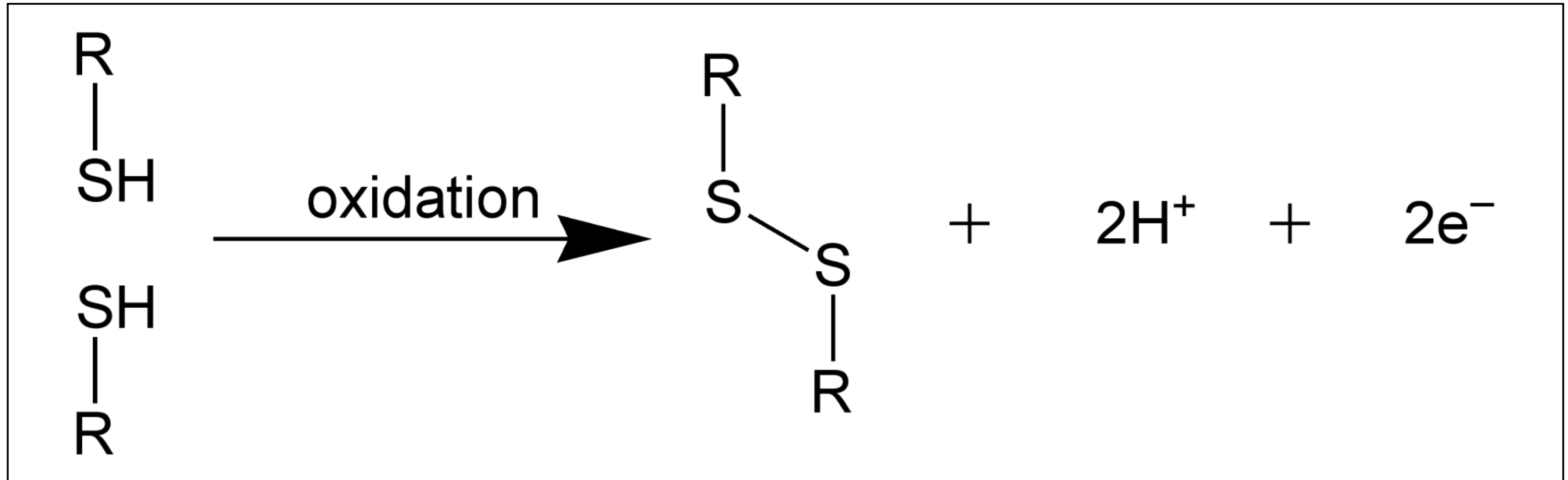




http://www.abvector.com/MultiChaperoneSystems.htm

1-12-16

Engineering for Expression of Proteins with Disulfide bonds



Thioredoxin (TRX) superfamily

- Consists of proteins containing one or more „TRX-like“ domains
- Redox-active members have a „CXXC“ catalytic motif
- TRX-like redox-active proteins can be:
  - reductants of disulfide bonds (eg. TRX itself)
  - oxidants of SH-groups (eg. PDI, DsbA)
  - disulfide isomerases (eg. PDI, DsbC)

Engineering for Expression of Proteins with Disulfide bonds

Origami™ host strains are K-12 derivatives that have mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which greatly enhance disulfide bond formation in the cytoplasm.

Studies have shown that expression in Origami (DE3) yielded 10-fold more active protein than in another host even though overall expression levels were similar.

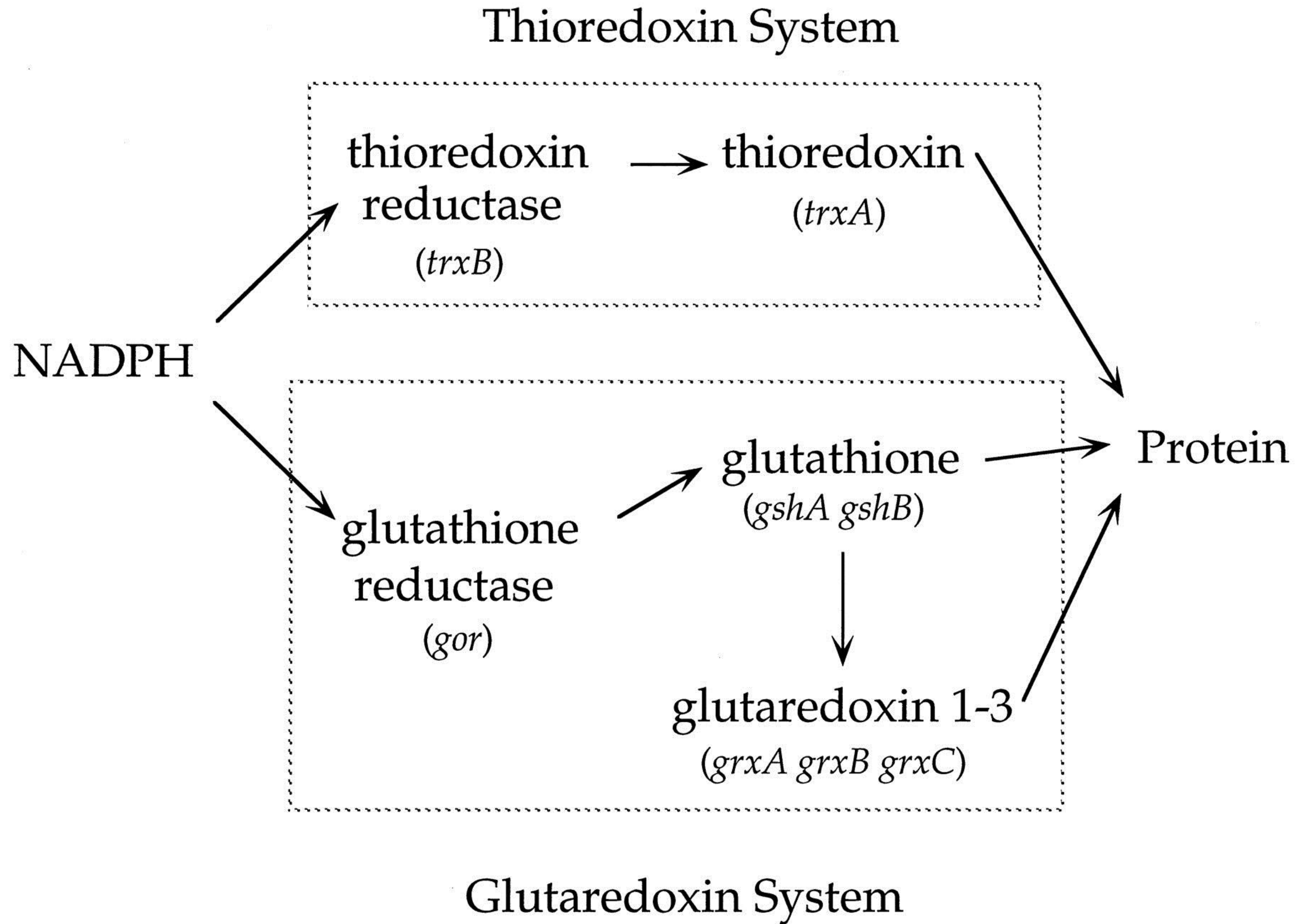
Origami hosts are compatible with ampicillin resistant plasmids and are ideal for use with pET-32 vectors, since the thioredoxin fusion tag further enhances the formation of disulfide bonds in the cytoplasm.

The *trxB* and *gor* mutations are selectable on kanamycin and tetracycline, respectively; therefore these strains cannot be used with plasmids carrying kanamycin- or tetracycline-resistance genes.

To reduce the possibility of disulfide bond formation between molecules, hosts containing the *trxB/gor* mutations are only recommended for the expression of proteins that require disulfide bond formation for proper folding.

J Biol Chem. 1997 Jun 20;272(25):15661-7.
The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm

Known components of the thioredoxin system (top) and glutaredoxin system (bottom).The genes encoding the components of these systems are shown inparentheses.

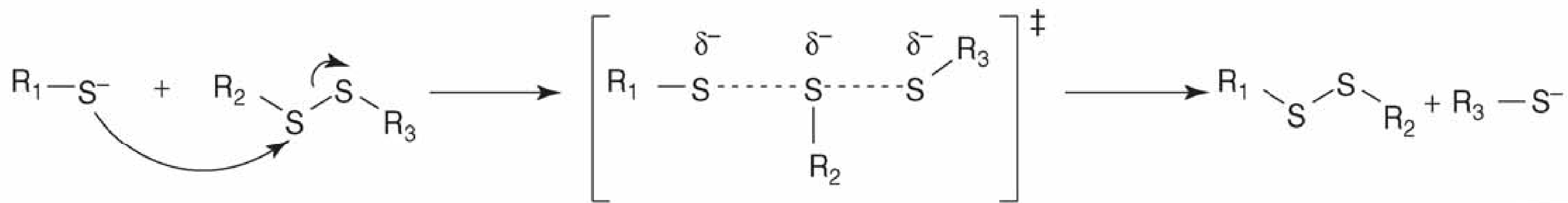


Prinz W A et al. J. Biol. Chem. 1997;272:15661-15667

©1997 by American Society for Biochemistry and Molecular Biology

Disulfide isomerisation

1) nucleophilic attack of a disulfide bond by an thiolate anion
2) transition state
3) formation of a mixed disulfide between PDI and the substrate protein – a substrate thiol is now free to attack another protein disulfide bond
4) isomerization reaction is driven by energy minimization:
   the native disulfide bond is favored and forms more quickly than the potential re-oxidation of the same bond
5) PDI is released unchanged – in it‘s reduced state



Engineering of E.coli for Expression of Proteins with Disulfide bonds

Protein Disulfide Isomerases

E.coli

DsbC
DsbG

Saccharomyces cerevisiae,
Pichia pastoris

PDI

E.coli

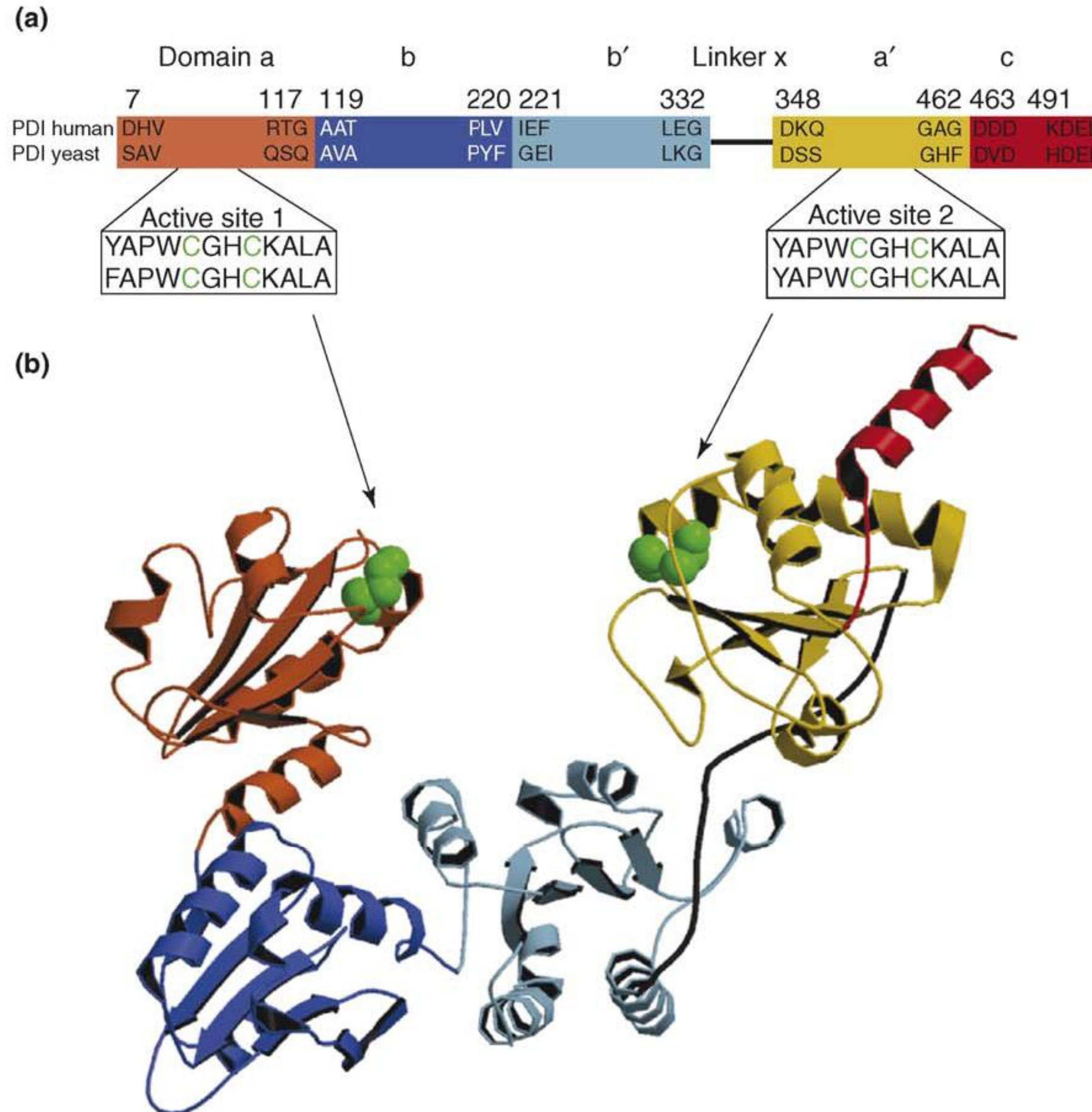
DsbA: Disulfide oxidase

PDI - Protein disulfide isomerase

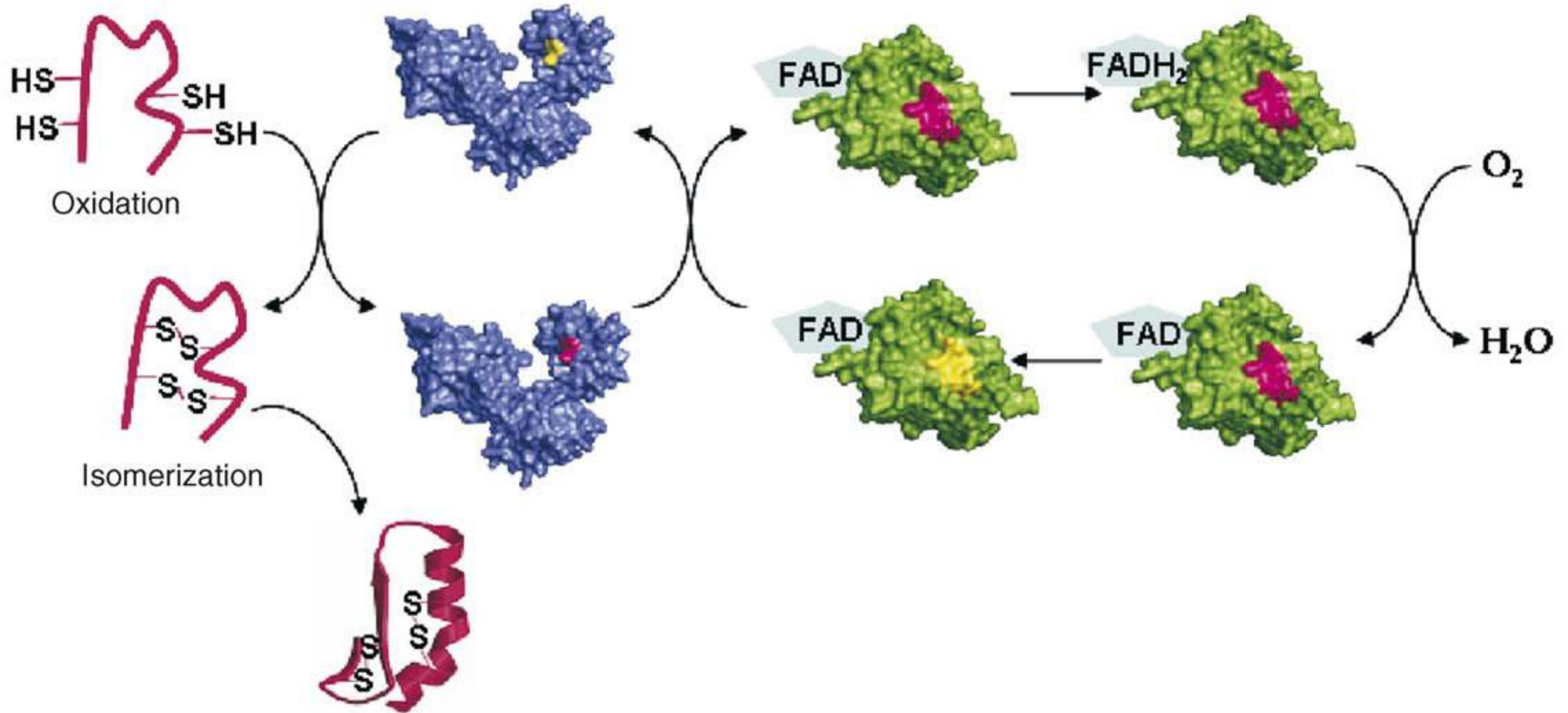
- PDI family comprises of several members
  more than a dozen members in humans alone
- PDI (aka PDI-1) is the most abundant member and constitutes ~0.8% of total cellular proteins in mammalian cells and yeast

- PDI is a monomer containing 4 TRX-like domains: a, b, b´, a´; linker (x) and a c-terminal extension domain (c)
  organized in the order: **abb´xa´c**
- only a and a´ have the catalytic CXXC motif

PDI - Protein disulfide isomerase



Oxidative folding in eukaryotes

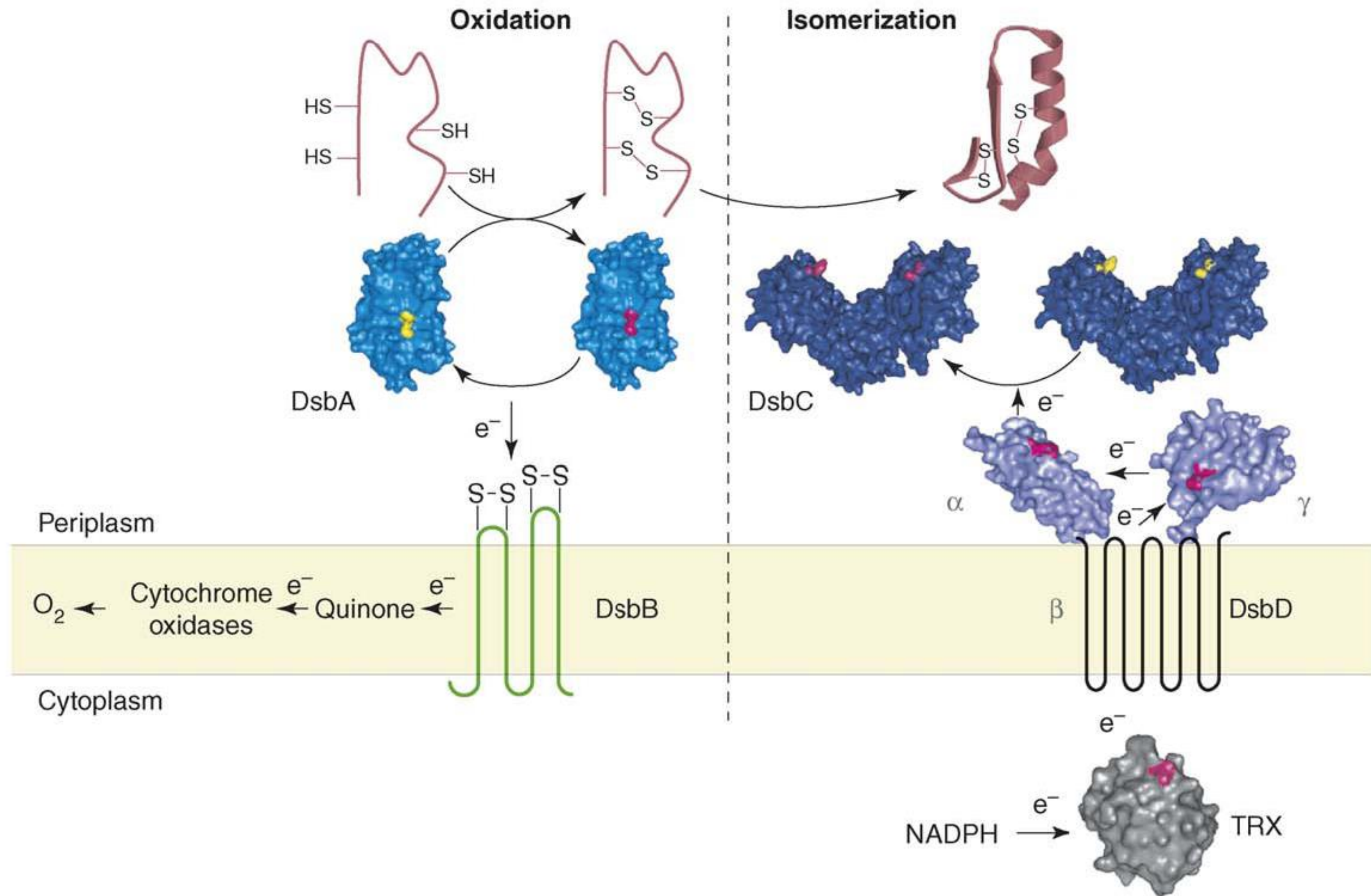


Dsb – Disulfide bond proteins

- Dsb-family proteins in prokaryotes regulates the forming of disulfide bonds in the periplasmic space like PDI does in the ER, but...
- There are two pathways:
  - oxidation pathway: disulfide bonds are introduced by DsbA
  - isomerization pathway: rearrangement of incorrect disulfide bonds by dsbC (or dsbG)

- DsbA is a monomer (like PDI)
- DsbC and DsbG are homodimers

Dsb – Disulfide bond proteins



oxidation pathway isomerization pathway

Comparison of PDI and dsbC, dsbG

- All of them have a chaperone activity, which is independent of their redox-properties, as they do not require the catalytic cysteines
- PDI is a monomer with 4 TRX-like domains, 2 of them with catalytic active CXXC-motifs
- DsbC and DsbG are homodimers, thus having also 2 catalytic active CXXC-motifs
- PDI is a multifunctional enzyme: It is able to function as a a disulfide oxidase as well as a chaperone and a disulfide isomerase
- DsbC and DsbG work as chaperones and isomerases

Glycosylation of Proteins in *P. pastoris*

First steps common in lower and higher eukaryotes

Hyperglycosylation in yeasts

Missing reactions in yeast (sialylation)

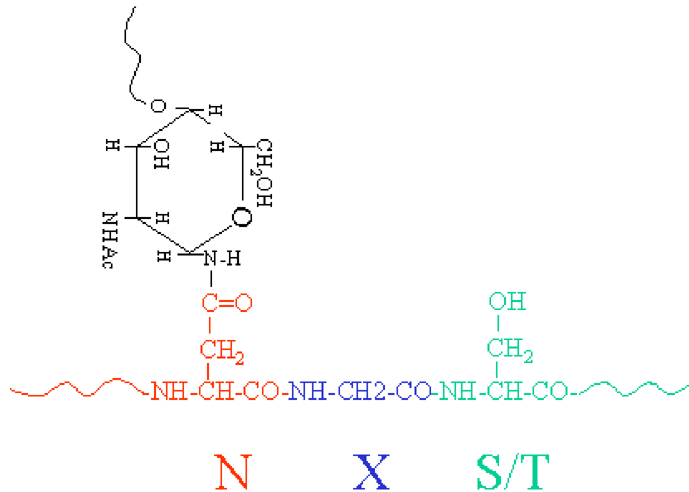
Problem for production of human therapeutic proteins

- activity determined by glycosylation
- stability and fate in human body

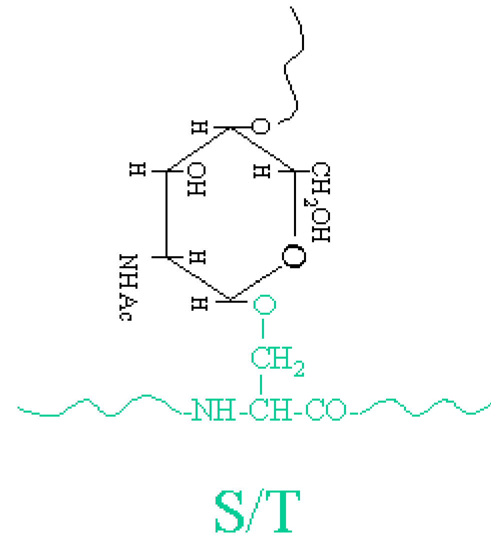
Benefit for industrial enzymes

N- and O-linked Protein Glycosylation

"N-Linked "



"O-Linked "



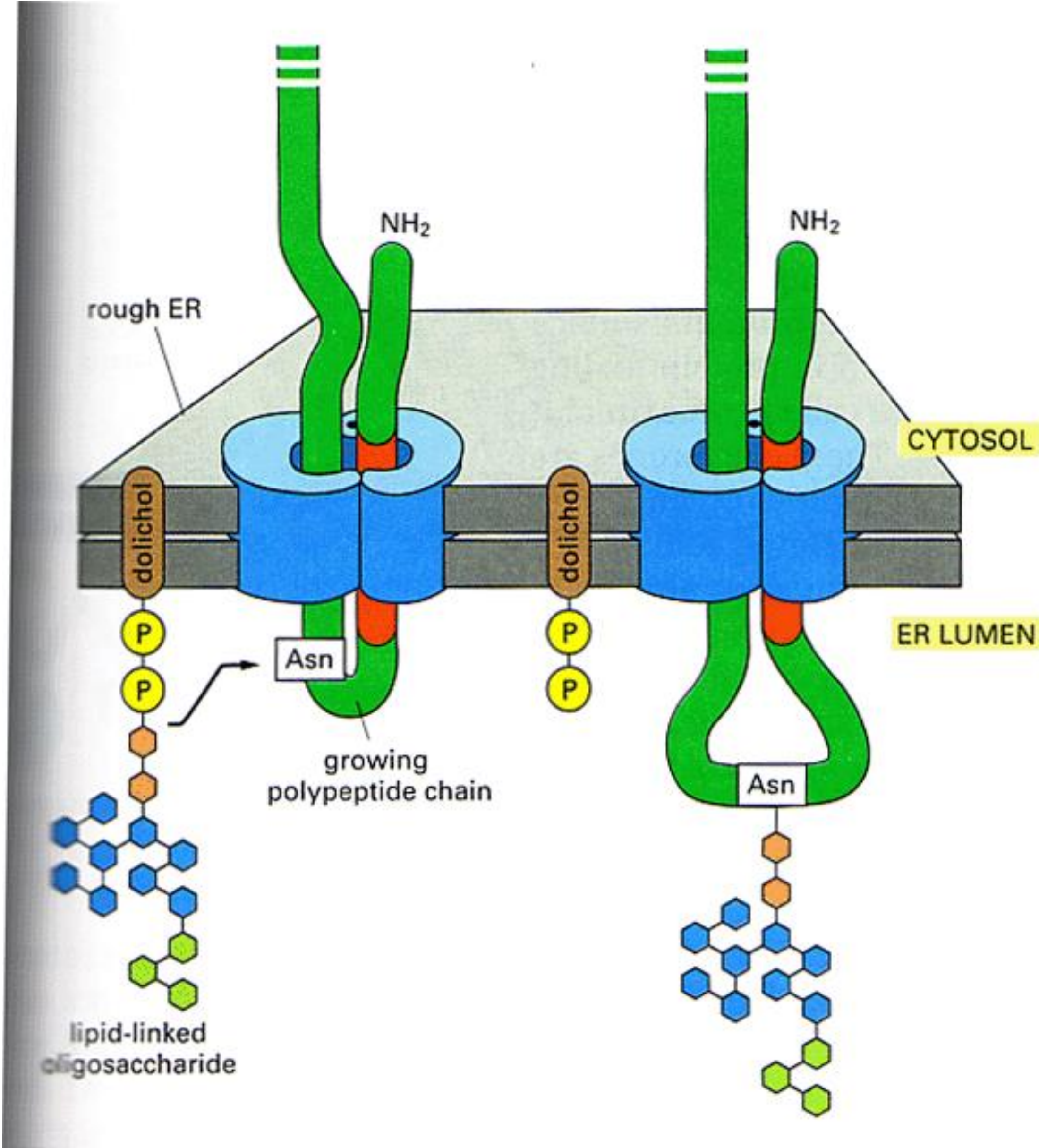
All N-linked carbohydrates are linked through **N-Acetylglucosamine** and the amino acid asparagine

The N-linked amino acid consensus sequence is Asn-any AA- Ser or Thr. The middle amino acid can not be proline (Pro).

Most O-linked carbohydrate covalent attachments to proteins involve a linkage between the monosaccharide N- Acetylgalactosamine and the amino acids serine or threonine.

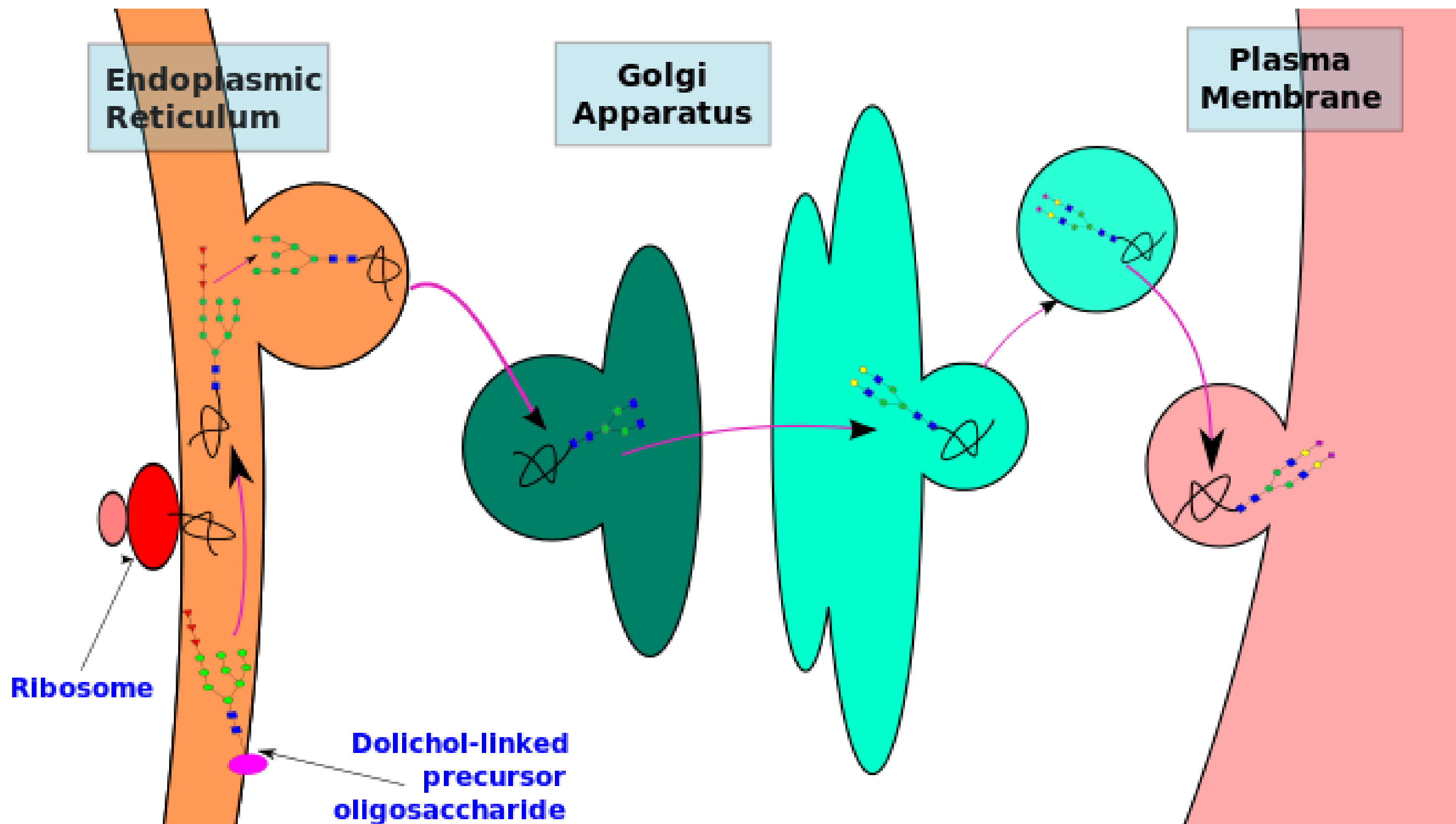
No consensus sequence defined for O-linked.

N-Glycosylation

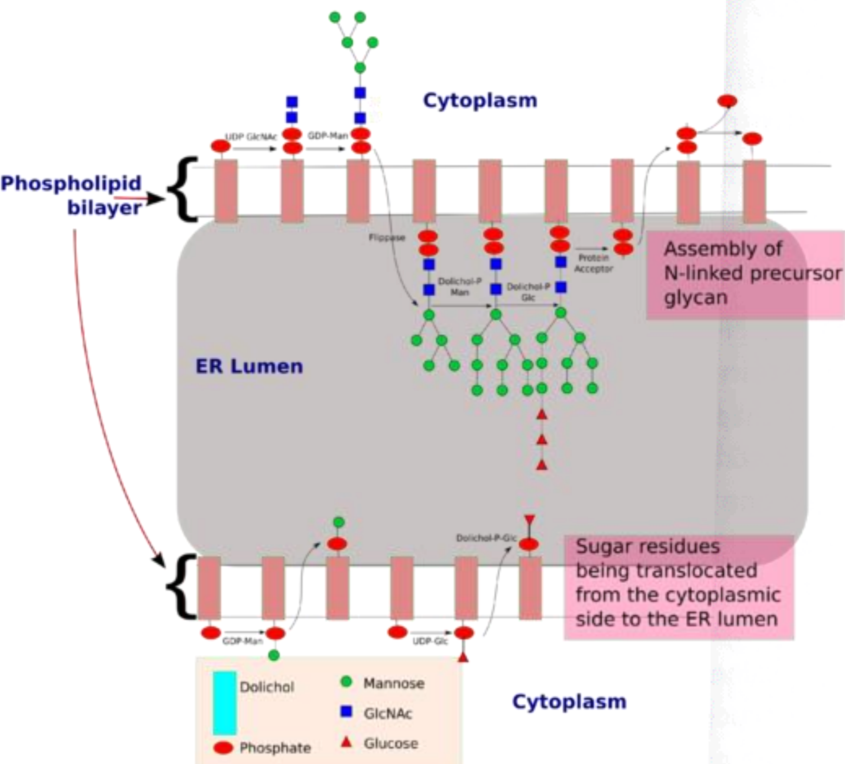


N-Glycosylation

General pathway in eukaryotes



https://en.wikipedia.org/wiki/N-linked_glycosylationhttps://en.wikipedia.org/wiki/N-linked_glycosylation



Formation of Glycosyl structures

lipid bilayer of ER membrane

ER LUMEN

CYTOSOL

dolichol

CTP

CDP

dolichol – P

2 UDP – GlcNAc

UDP + UMP

dolichol – P – P – $(GlcNAc)_2$

5 GDP – Man

5 GDP

dolichol – P – P – $(GlcNAc)_2(Man)_5$

"FLIPPING" IN MEMBRANE

$(Man)_5(GlcNAc)_2$ – P – P – dolichol

4 X

Man – P – dolichol

P – dolichol

mannose donor made from dolichol phosphate and GDP-mannose

$(Man)_9(GlcNAc)_2$ – P – P – dolichol

3 X

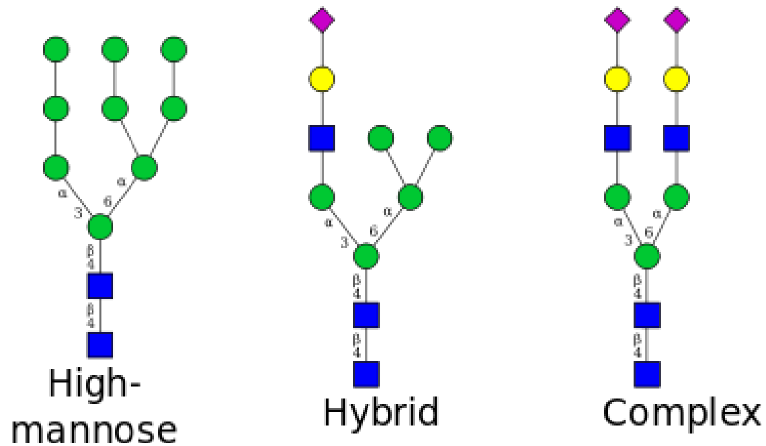
Glc – P – dolichol

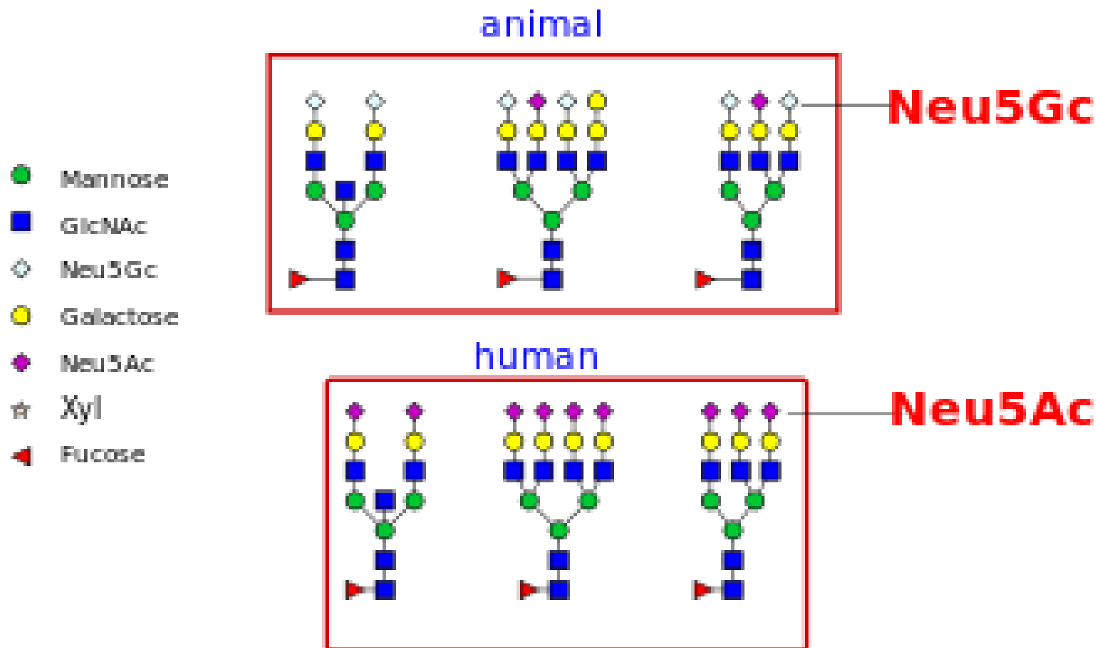
P – dolichol

glucose donor made from dolichol phosphate and UDP-glucose

$(Glc)_3(Man)_9(GlcNAc)_2$ – P – P – dolichol

https://en.wikipedia.org/wiki/N-linked_glycosylationhttps://en.wikipedia.org/wiki/N-linked_glycosylation

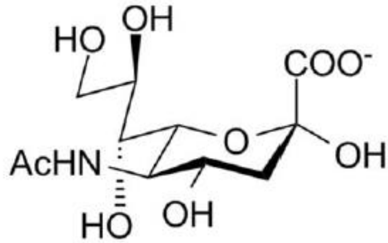




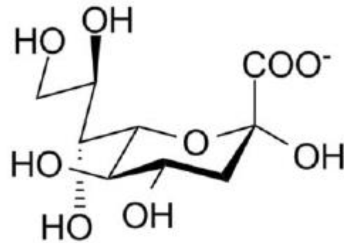
Non-human mammalian expression systems such as CHO or NS0 cells have the machinery required to add complex, human-type glycans. However, glycans produced in these systems can differ from glycans produced in humans, as they can be capped with both N-glycolylneuraminic acid (Neu5Gc) and N-acetylneuraminic acid (Neu5Ac), whereas human cells only produce glycoproteins containing N-acetylneuraminic acid.

https://en.wikipedia.org/wiki/N-linked_glycosylationhttps://en.wikipedia.org/wiki/N-linked_glycosylation

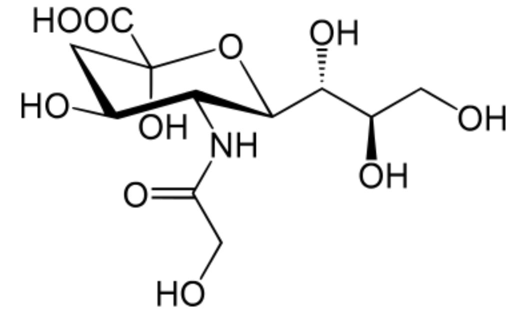
Sialic Acids



N-Acetylneuraminic acid
Neu5Ac

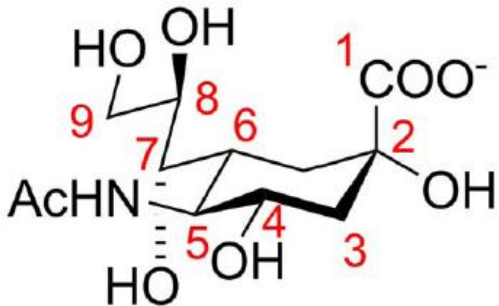


2-Keto-3-deoxynonic acid
Kdn

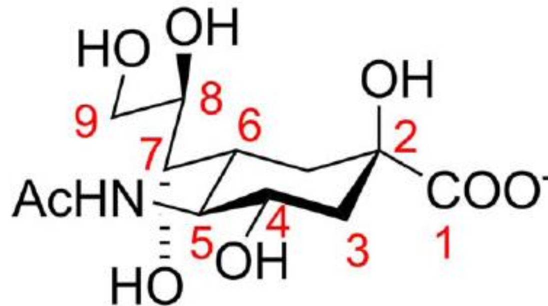


N-Glycolylneuraminic acid

The two most common sialic acid derivatives are Neu5Ac and Kdn.



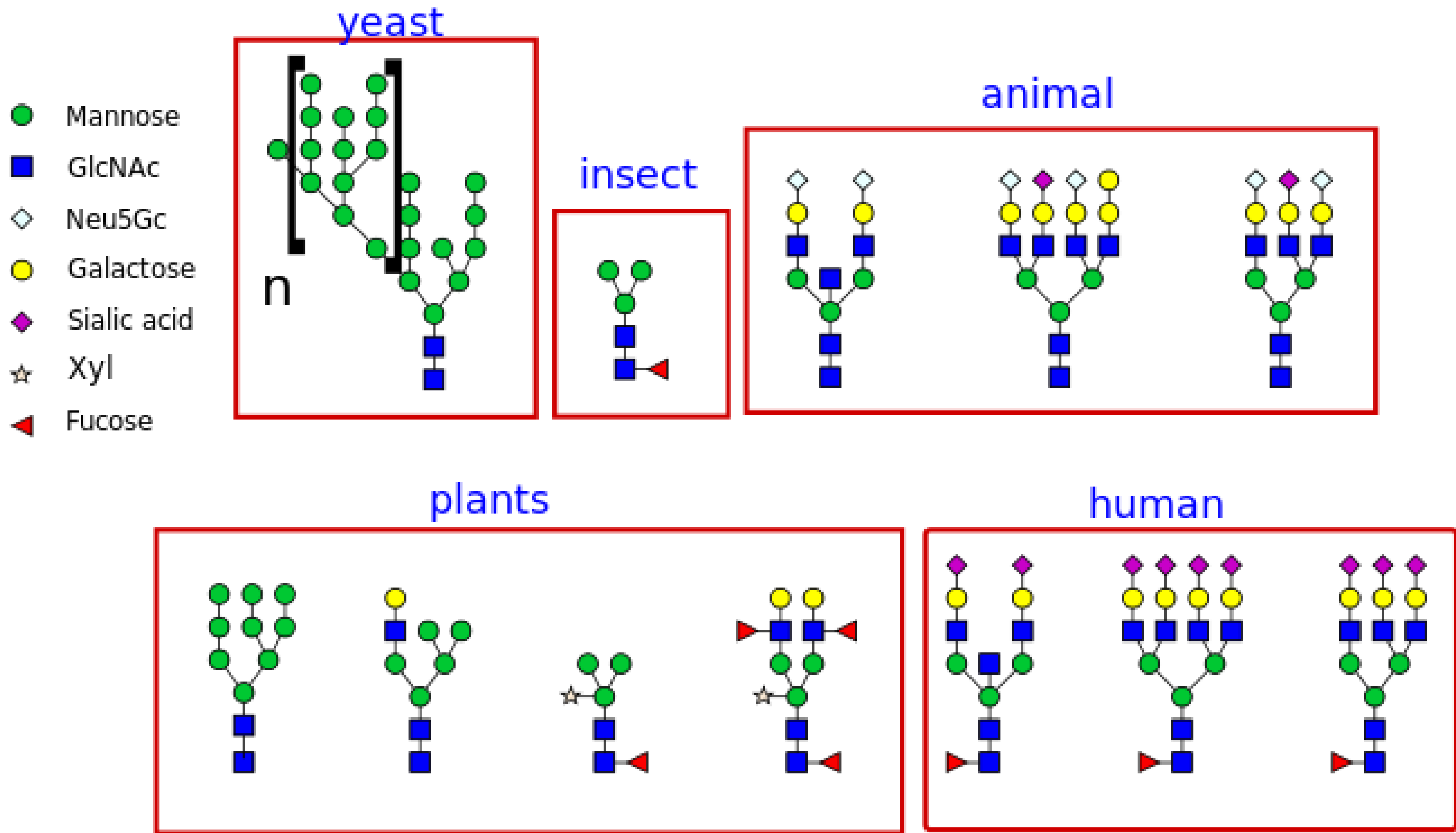
α-anomer



β-anomer

Glycans produced in animal cell systems (e.g. CHO) can differ from glycans produced in humans, as they can be capped with both N-glycolylneuraminic acid (Neu5Gc) and N-acetylneuraminic acid (Neu5Ac), whereas human cells only produce glycoproteins containing N-acetylneuraminic acid

N-linked glycosylation



The N-linked glycosylation process occurs in eukaryotes and widely in archaea, but very rarely in bacteria

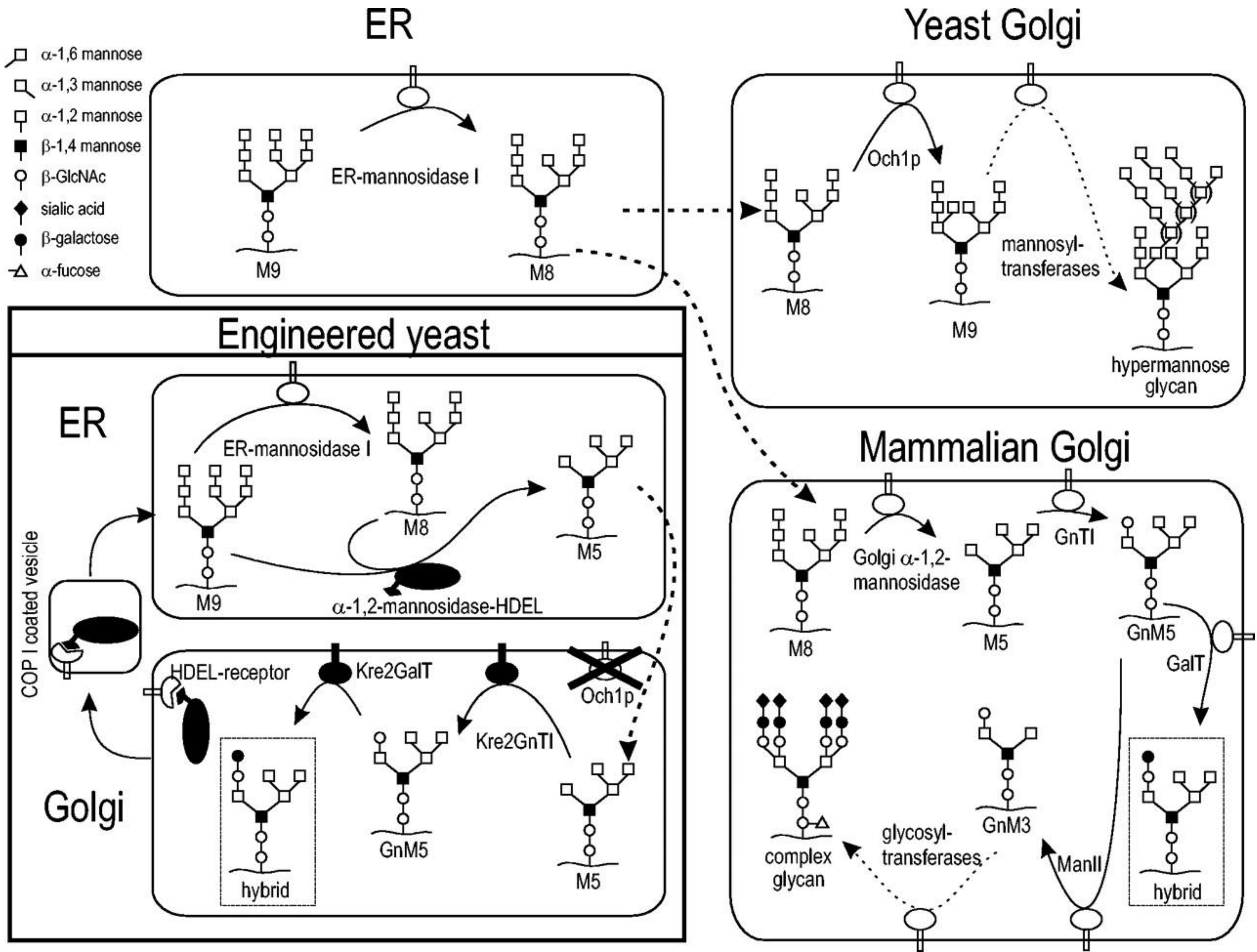
https://en.wikipedia.org/wiki/N-linked_glycosylationhttps://en.wikipedia.org/wiki/N-linked_glycosylation

Enzymes for Analysis of Glycoproteins

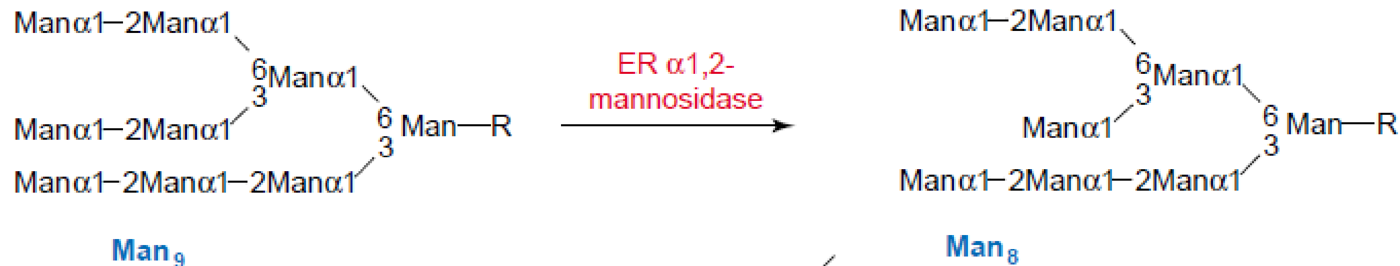
These are just a few of the enzymes available for carbohydrate analysis. Abbreviations are as follows: Asn--Asparagine, Gal--Galactose, GlcNAc--N-acetylglucosamine, GalNAc--N-acetylgalactosamine, and NeuAc--N-acetylneuraminic acid.

Enzyme	Type of enzyme	Specificity
Endoglycosidase D	Endo	Cleaves various high mannose glycans
Endoglycosidase F	Endo	Cleaves various high mannose glycans
Endoglycosidase H	Endo	Cleaves various high mannose glycans
β-galactosidase	Exo	Removes terminal galactosides from Gal-β1,3-GlcNAc, Gal-β1,4-GlcNAc or Gal-β1,3 GalNAc.
Peptide:N-Glycosidase F	Endo	Glycoproteins between Asn and GlcNAc (removes oligosaccharides)
Sialidases (Neuraminidases) *Vibrio cholerae* *Clostridium perfringens* *Arthobacter ureafaciens* Newcastle disease virus	Exo	NeuAc-α2,6-Gal, NeuAc-α2,6-GlcNAc or NeuAc-α2,3-Gal

Engineering of Glycosylation



Synthesis Route for Core Glycan Structure

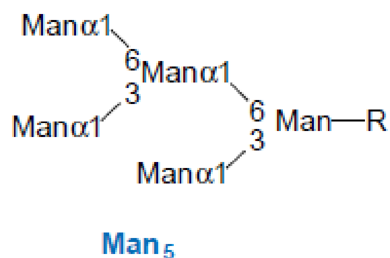


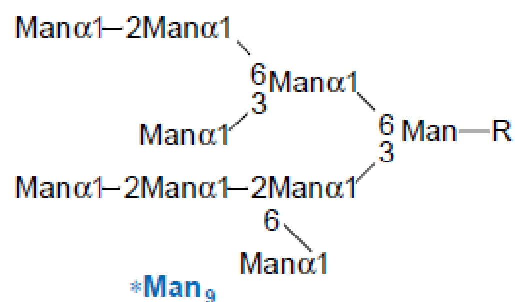
Man9

Man8

α1,2-
Mannosidase

α 1,6-
Mannosyltransferase
(Ochlp)

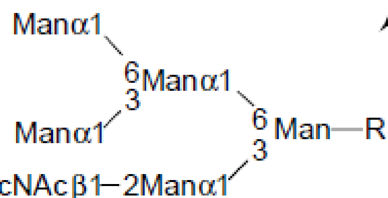




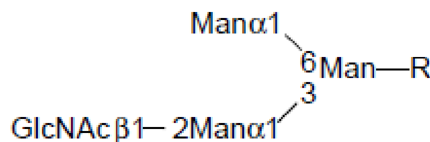
Man5

*Man9

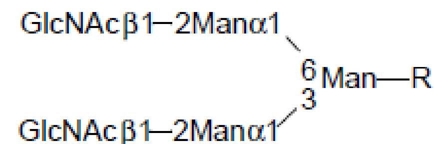
GlcNAc transferase I



α 1,3/1,6-
mannosidase



GlcNAc
transferase II



GlcNAcMan5

GlcNAcMan3

GlcNAc2Man3

TRENDS in Biotechnology

Engineering of Glycosylation

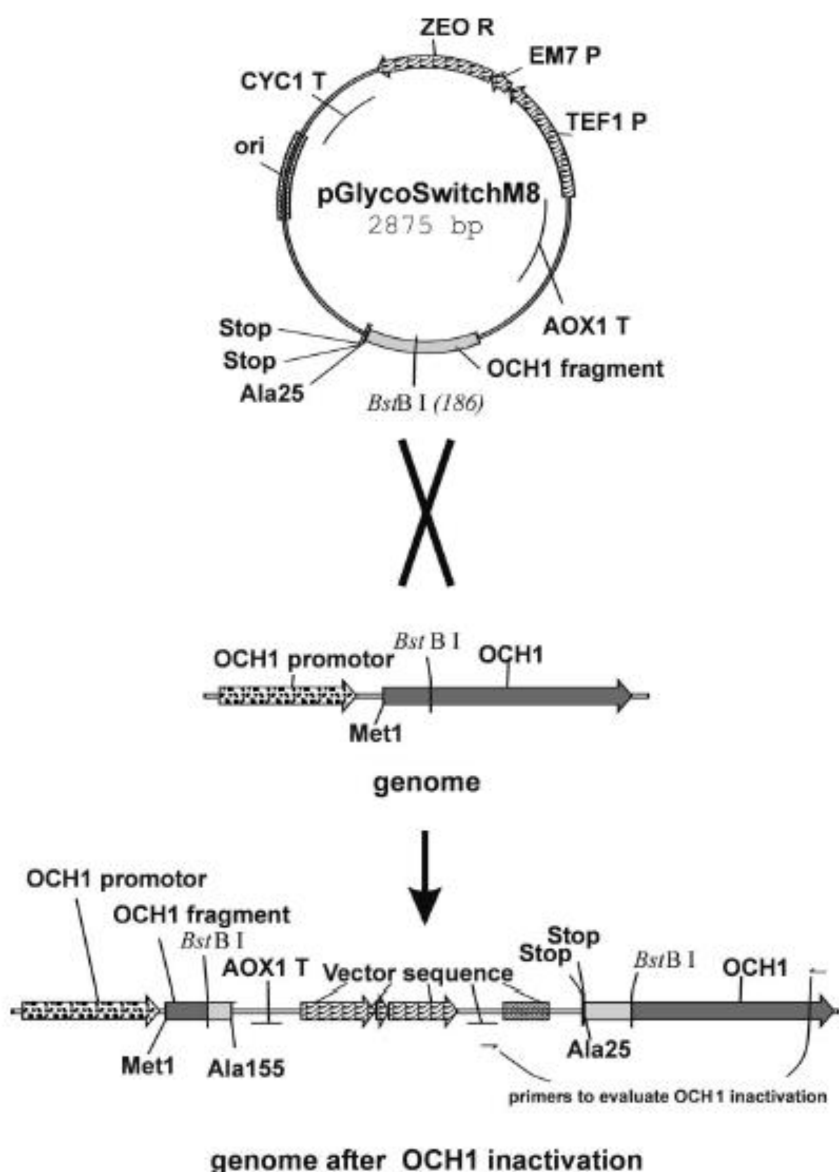


FIG. 2. *OCH1* inactivation vector. Upon digestion of pGlycoSwitchM8 with BstBI and transformation in *P. pastoris*, the construct integrates at the *OCH1* locus. This results in a short *OCH1* fragment that does not translate to a functional *OCH1* gene and a promotorless fragment that cannot be translated because of the absence of a promoter and the presence of two in-frame nonsense codons.

Deletion of *OCH1*

Och1 Protein:Mannosyltransferase of the cis-Golgi apparatus, initiates the polymannose outer chain elongation of N-linked oligosaccharides of glycoproteins

Engineering of Glycosylation

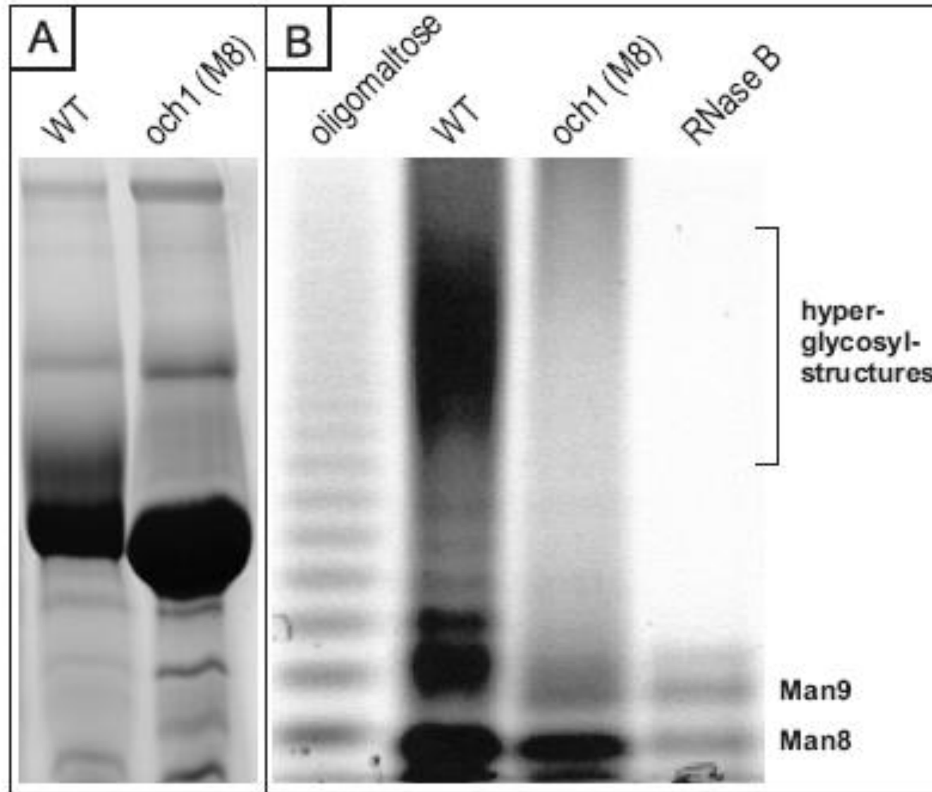
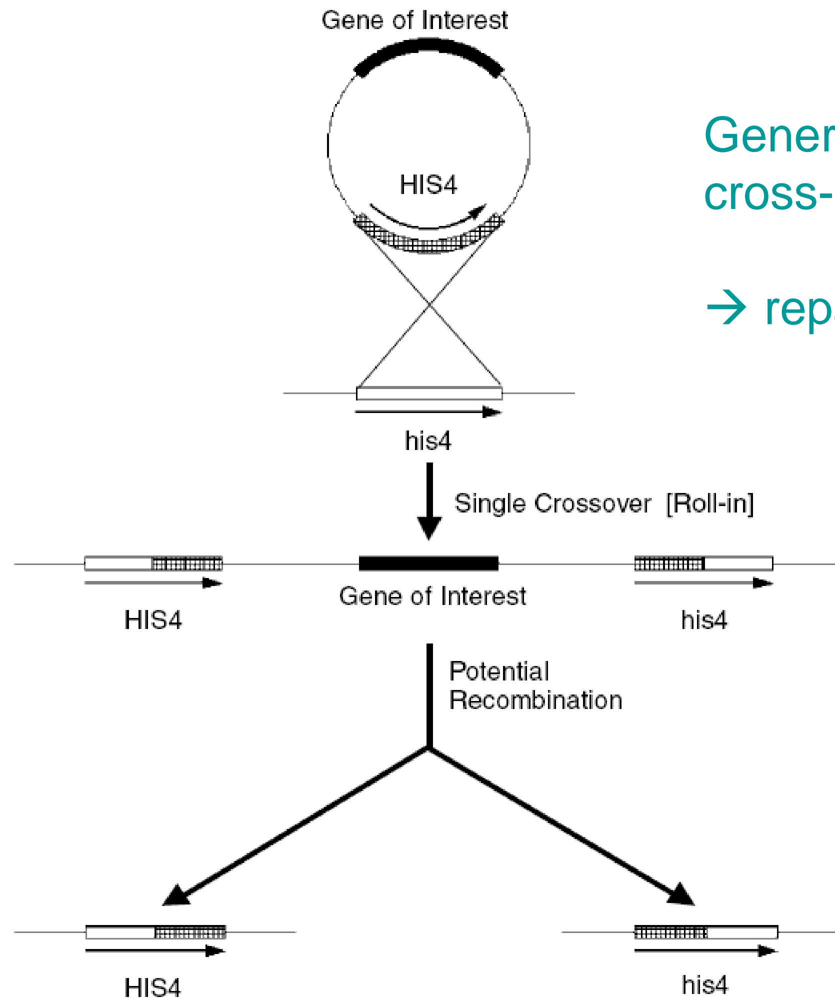


FIG. 5. Evaluation of hyperglycosylation after inactivation of *P. pastoris OCH1*. (A) Coomassie brilliant blue-stained SDS-PAGE gel containing supernatants of *P. pastoris* strains secreting *T. reesei* mannosidase. For the nonengineered strain (WT) a clear smear is visible, whereas this smear is not present for the strain with *och1* inactivated [och1 (M8)]. (B) FACE analysis of N-glycans derived from mannosidase secreted by a nonengineered strain (WT) and a strain with *och1* inactivated [och1 (M8)]. The bands with greater electrophoretic mobility are the Man8 and Man9 bands and represent core N-glycan structures. The hyperglycosyl structures are slowly migrating sugars. They are not present in the strain with *och1* inactivated.

Deletion of *OCH1*



General problem of single cross-over integration

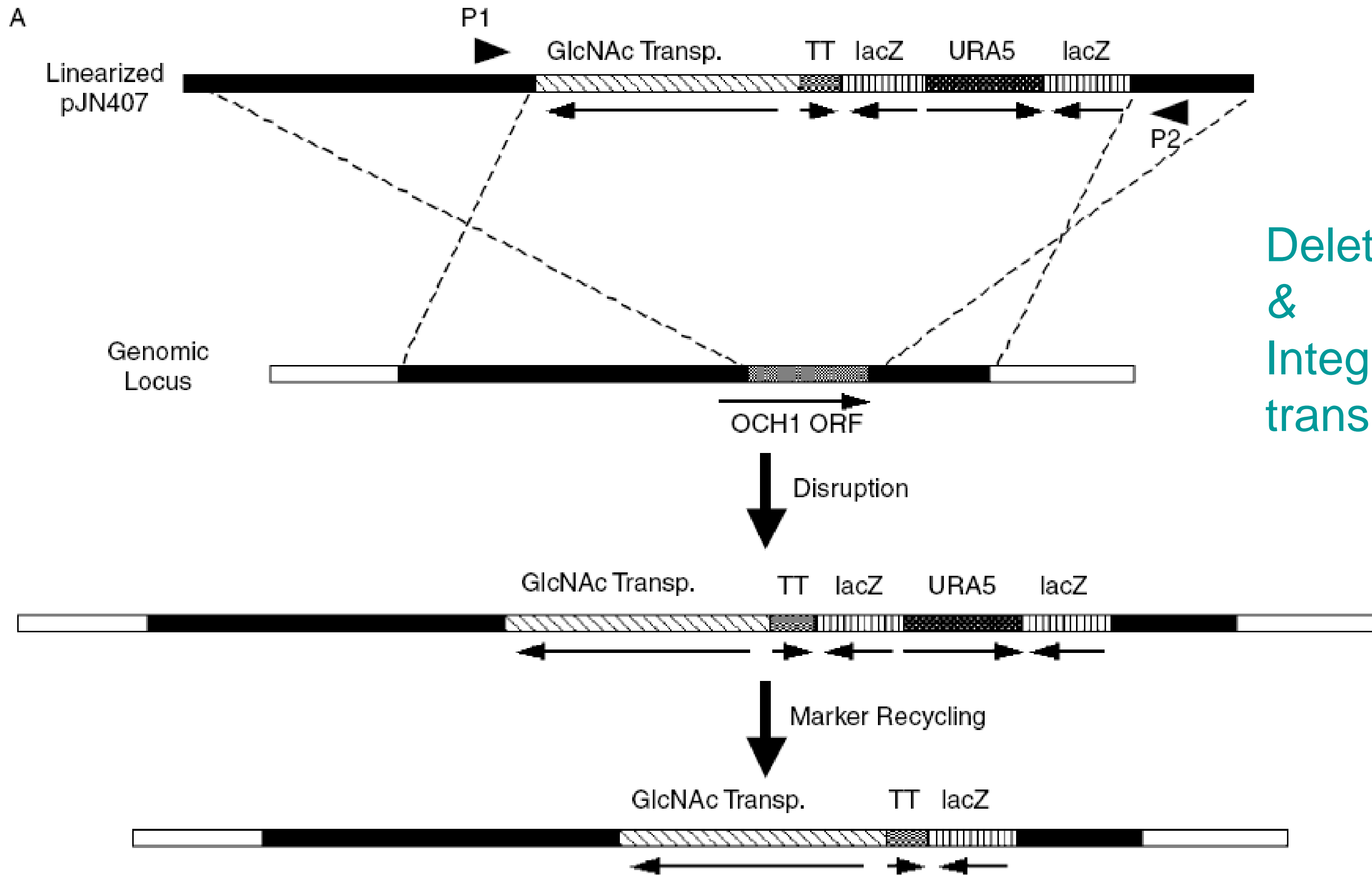
→ repair by reverse recombination

Figure 5. (A) Stable integration of the gene for the *K. lactis* UDP-GlcNAc-transporter into the *OCH1* locus of *P. pastoris*. Plasmid pJN407 is linearized with *Sfi*I and integrated into the *P. pastoris* genome by double cross-over (knock-in). After marker recycling by selection on 5FOA, the gene of interest and a '*lacZ* scar' are stably retained. (B) Integration of a gene of interest into the genome by single cross-over (roll-in). Because the roll-in method leads to duplication of the marker locus, a potential recombination event can lead to loss of the gene of interest and repair of the auxotrophic marker

Copyright © 2003 John Wiley & Sons, Ltd.

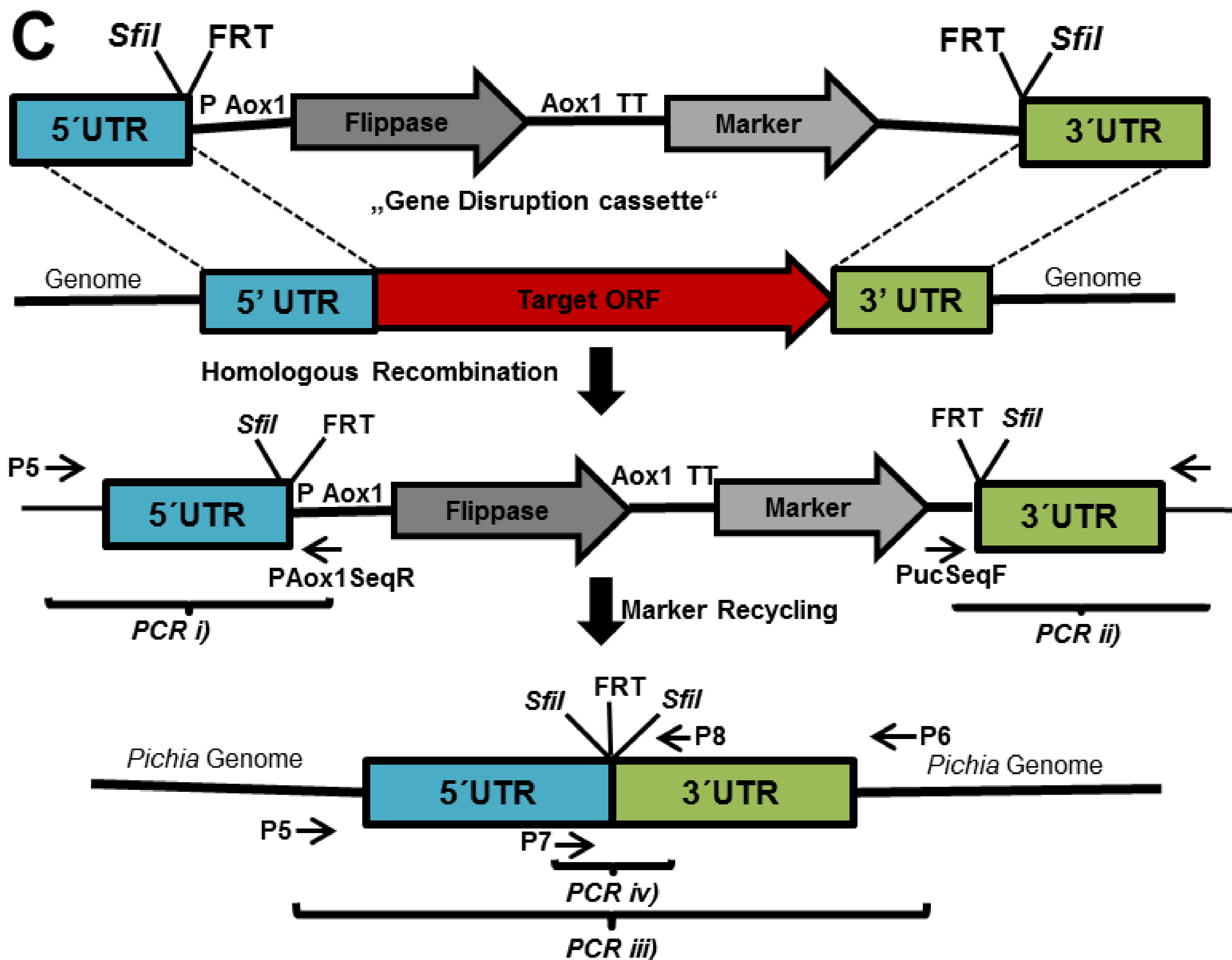
Yeast 2003; **20**: 1279–1290.

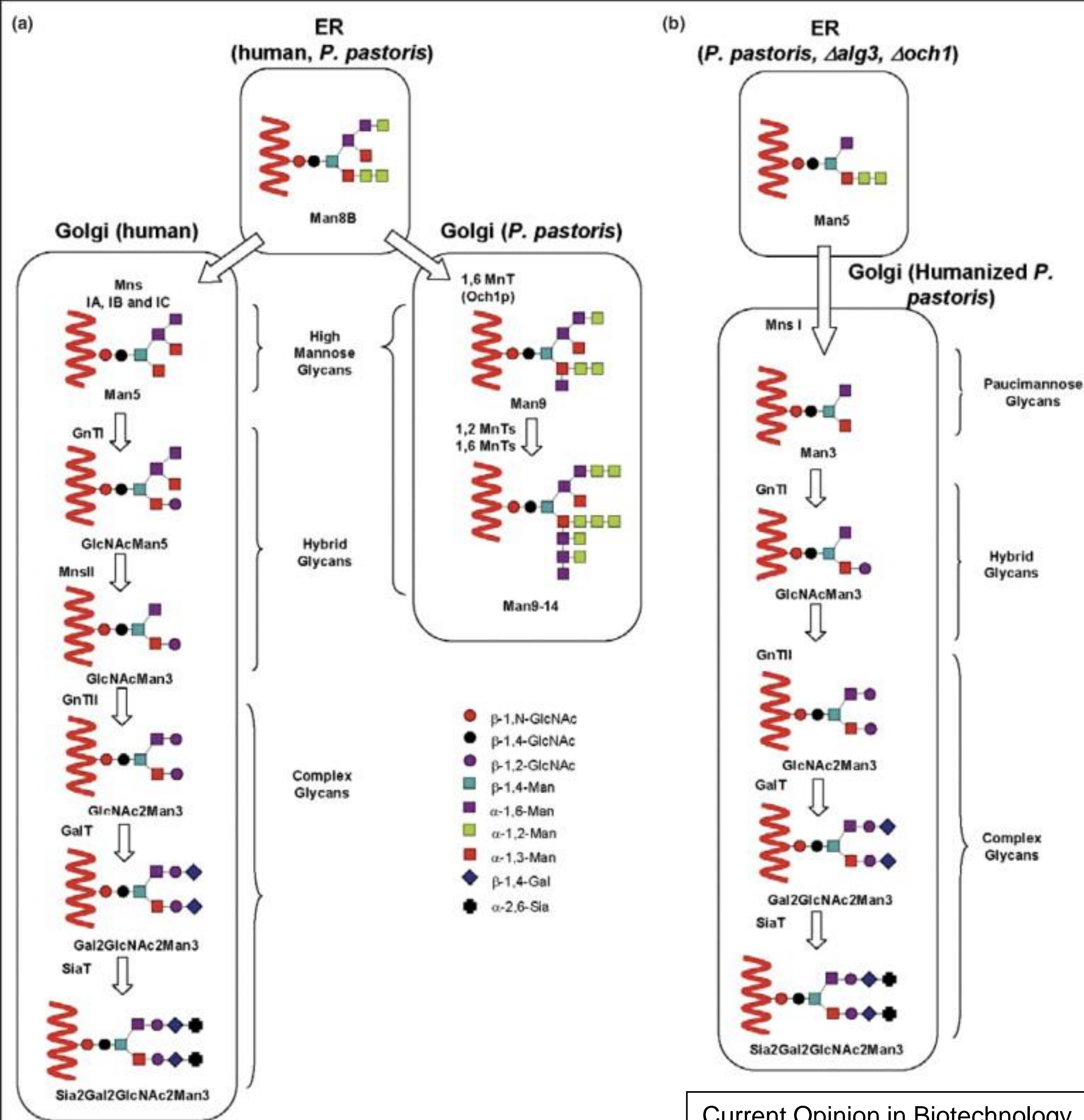
Engineering of Glycosylation



Deletion of *OCH1*
&
Integration of GlcNac-transporter

Marker-free knock-out



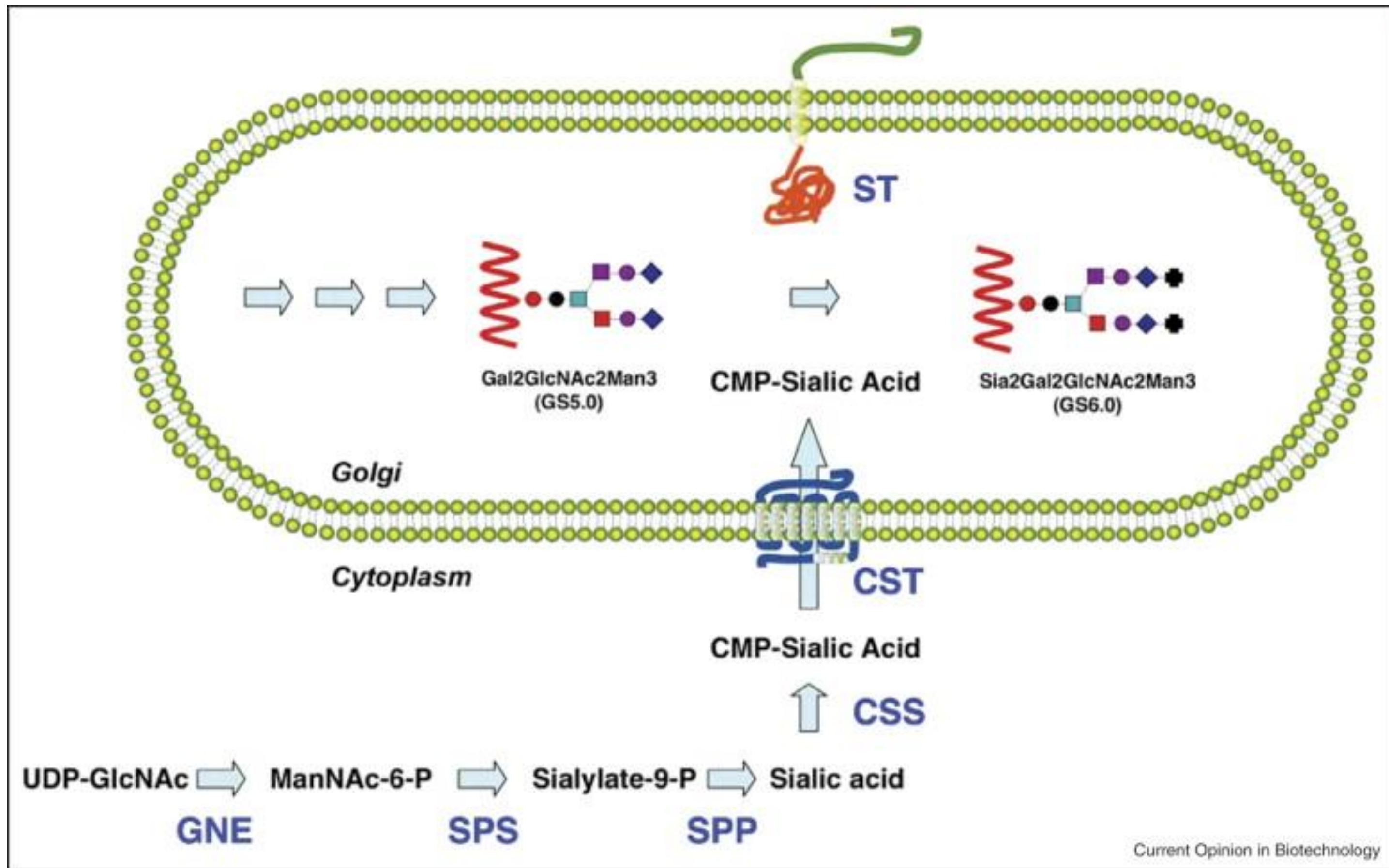


N-linked glycosylation pathways in humans and yeast. Representative N-linked glycosylation pathways in humans and *P. pastoris* **(a)**. An alternative humanized N-linked glycosylation pathway in *P. pastoris* **(b)**.

Mns; α 1,2- mannosidase, MnsII; mannosidase II, GnTI; β 1,2-*N*-acetylglucosaminyl-transferase I, GnTII; β 1,2-*N*-acetylglucosaminyltransferase II, GalT; β 1,4-galactosyl-transferase, SiaT; α 2,6-sialyltransferase, MnT; mannosyltransferase.

For simplicity the two GlcNAc residues present at the reducing end of all glycans have been omitted.

Current Opinion in Biotechnology, Volume 18, Issue 5, 2007, 387–392

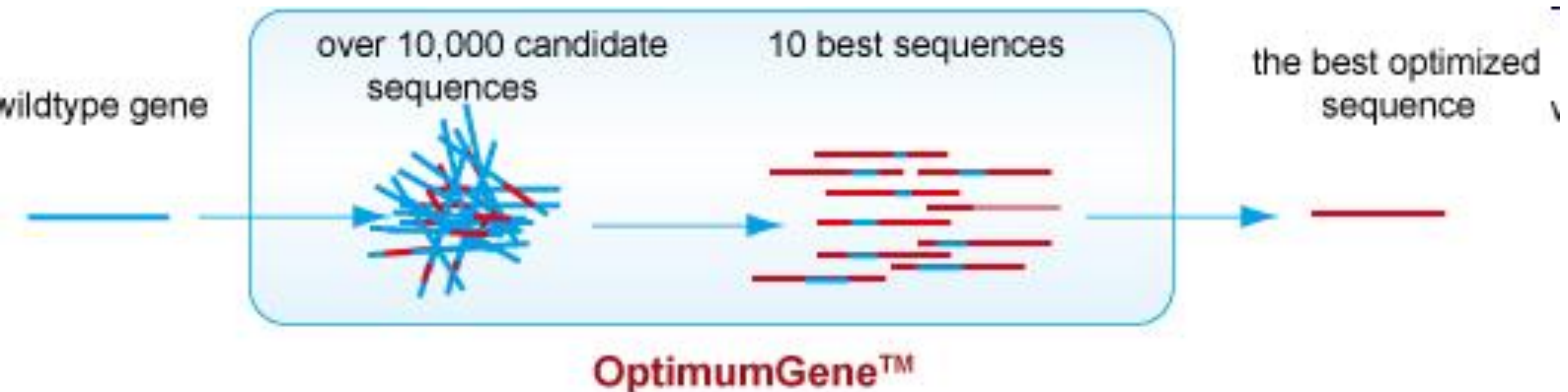


Glycoengineering steps required for sialic acid transfer in the yeast Golgi. Endogenous UDP-GlcNAc, present in the yeast cytoplasm, is converted to CMP-sialic acid by UDP-*N*-acetylglucosamine-2-epimerase/*N*-acetylmannosamine kinase (GNE), *N*-acetylneuraminate-9-phosphate synthase (SPS), sialylate-9-P phosphatase (SPP) and CMP-sialic acid synthase (CSS). Subsequently, the product is translocated into the Golgi by the CMP-sialic acid transporter (CST) and sialic acid is transferred onto the acceptor glycan by sialyltransferase (ST). Enzymes are indicated by blue text and metabolic intermediates by black text. For simplicity the two GlcNAc residues present at the reducing end of all glycans have been omitted.

Current Opinion in Biotechnology, Volume 18, Issue 5, 2007, 387–392

Gene Design

DNA
mRNA
Transcription:
Antiviral Motifs
Chi-sites
Polymerase slippage sites
TATA box
Cryptic splicing sites
Terminal signal etc
-G
-C
-U
-G
-G
-U
-G
-G
A
Ribosome
Protein folding:
Codon context
Translation pause sites
etc
Deviant folding
Natural folding
Ser
Asn
Val
Ala
Arg
Arg
CGA
GCU
GCA
CGG
GCA
CGU
mRNA
Ribosome
Translation:
Codon usage bias
mRNA secondary structure
GC content
Negative cis-acting elements etc



Optimization Parameters

parameters that are critical to transcription, translation and protein folding:

Transcriptional Efficacy:

GC content
CpG dinucleotides content
Cryptic splicing sites
Negative CpG islands
SD sequence
TATA boxes
Terminal signal

Translation Efficiency:

Codon usage bias
GC content
mRNA secondary structure
Premature PolyA sites
Internal chi sites and ribosomal binding sites
RNA instability motif (ARE)
Inhibition sites (INS)
Stable free energy of mRNA

Protein Refolding:

Codon usage bias
Interaction of codon and anti-codon
Codon-context
RNA secondary structures