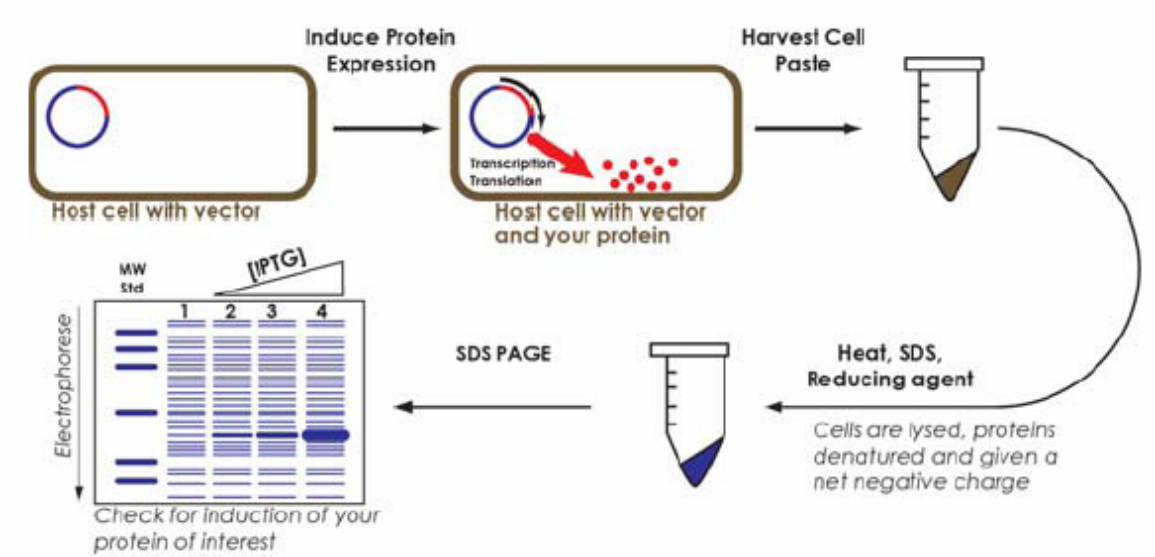
Protein Expression and Purification

TECHNIQUES & STRATEGIES IN MOLECULAR MEDICINE

Dec 2007, TCD

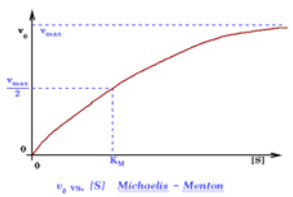
Purification



Outline

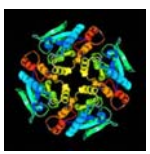
- Applications
- Protein Classes
  - Secreted, Intracellular
  - Integral membrane, membrane associated
- Construct design considerations
  - Protein class
  - Domains – identification and defining endpoints
  - Tags, fusions and cleavage sites
- Expression systems
  - E. coli, Yeast
  - Baculovirus infected insect cells
  - Mammalian cells
  - In vitro transcription/translation
- Purification

Applications requiring protein expression and purification

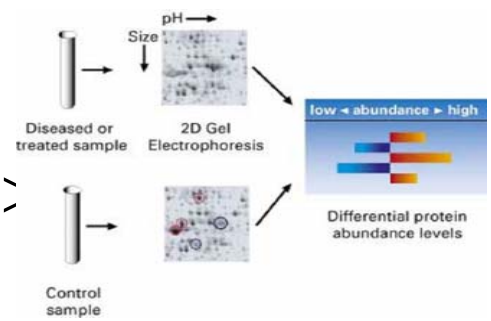


– Enzymology/functional studies (mg)

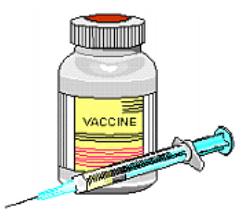
– Antibody production (µg-mg)



– Structure determination (50 mg or >



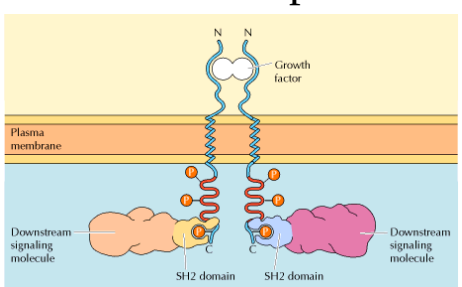
– Commercial/therapeutic uses (g)



  – *e.g.* Vaccines, immunomodulators

Types of protein to be expressed and purified

- Soluble – cytoplasmic
  - Easiest to work with
  - Usually no S=S bonds, usually well expressed
- Secreted (stable)
  - Proteases, antibodies, cytokines…
  - Frequently contain PTMs
    - Glycosylation, S=S, signal sequence removed
- Membrane associated
  - Intracell or extracell
    - Frequently expressed without membrane anchor
    - Extracell and intracell domains expressed and studied separately

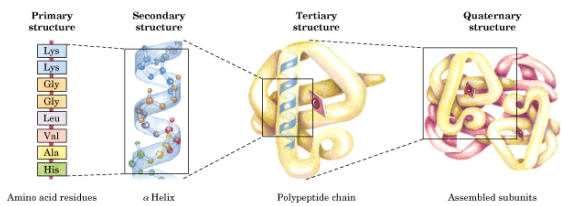


Before you start…..

•Check the literature

•Any useful predictive bioinformatic information available?

•www.expasy.org/ **E**xpert **P**rotein **A**nalysis **Sy**stem



•Is the protein membrane-bound or water soluble?

•Is it single domain or multi-domain?

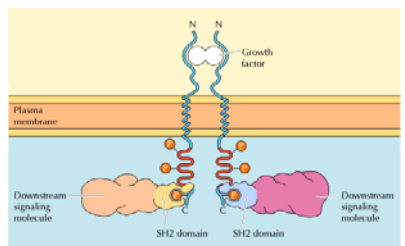
•How to purify and how to identify?

•Do you expect/require PTMs ?

•Do you have an assay for your target protein?

Before you start…..

- Construct design
  - Which cloning protocols?
    - Identify domains, select domains for expression
    - e.g. kinase domain from RTKs for assays and structure based drug discovery
    - Fusion tags ?
  - Which host cell system?
  - Which expression vector?



Construct design

- Search literature for precedent with similar/related protein
  - PepcDB
    Protein **E**xpression **P**urification and **C**rystallization **D**ata**b**ase
    http://pepcdb.pdb.org/index.html
  - Protocols for cloning, expression, purification are stored and are searchable

Protein expression: which vector ?

N- or C-terminal fusions of proteins to short peptides (tags) or to other proteins (fusion partners) offer several potential advantages:

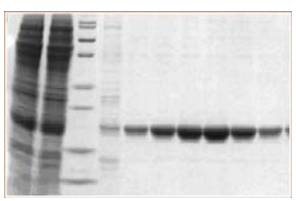
- **Improved expression**

- **Enhanced solubility/folding**
  - E.g. Trx, MBP.
- **Easy detection**

  e.g. Western blot analysis, GFP by fluorescence –

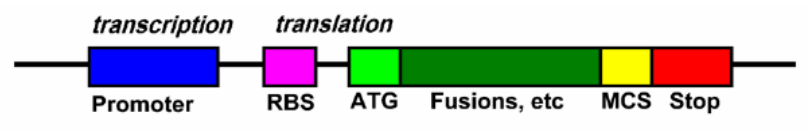
- **Simple purification.**

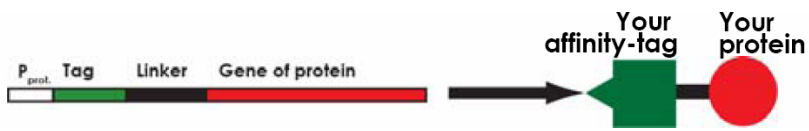
  - fusion partner/tag bind specifically to affinity resins



- Protect from degradation

Tags and fusions





Tag	Length (aa)	Ab detection	Purification
Poly –His	5-10	Yes	Ni-affinity
FLAG	8	Yes	Ab
S-tag	15	Yes	RNase S
cMyc	11	Yes	Ab
Strep	8	Yes	Ab
T7	11-16	Yes	Ab
V5	14	Yes	Ab
CBP	26	No	Calmodulin
SBP	38	No	strepavidin

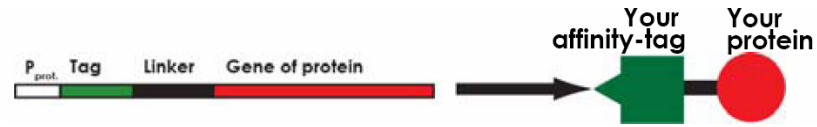
Tags and Fusions

Fusion	Size	Purification
GST	26 kDa	Glutathione Sepharose
MBP	40 kDa	Amylose resin
Trx	12 kDa	Thiobond

Hunt I, (2005) Protein Exp & Purif 40, 1-22 From gene to protein: a review of new and enabling technologies for multi-parallel protein expression.

Construct design – fusion cleavage sites

- often included to allow removal of tags/fusions



Enzyme	Recognition site	Comments
Thrombin	LVPR/GS	Less specific
Factor Xa	IEGR/	Less specific
Enterokinase	DDDDK/	Very specific

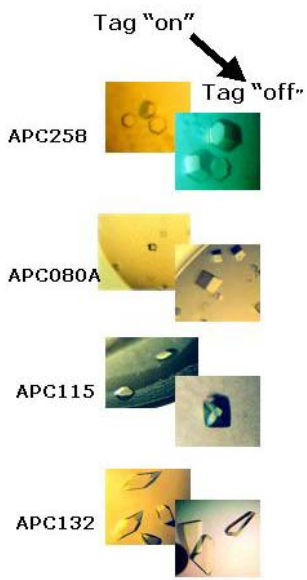
Waugh DS (2005) Trends Biotechnol.23(6):316-20 Making the most of affinity tags

Disadvantages of using fusion partners:

• Cleaving/removing the fusion partner requires expensive proteases (Factor Xa, enterokinase)
• cleavage rarely complete – low-moderate yield
• solubility not guaranteed

- his-tag in target protein does not prevent the crystallization
- for obtaining the bigger and better diffracting crystals the "tag-off" protein is needed

target	diffraction Å + His tag	diffraction Å - His tag
APC236	1.85	1.1
APC132	2.4	2.0
APC115	2.8	1.2
APC182	3.0	2.7
APC135	2.5	1.8
APC258	4.0	2.8

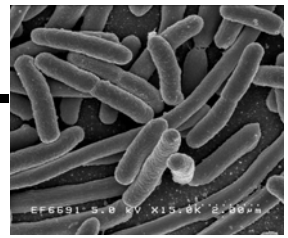


Which host cell expression system?

- *E. coli*
- Yeast
- Insect cells
- Mammalian cells
- Cell-free

Prokaryotic expression systems-

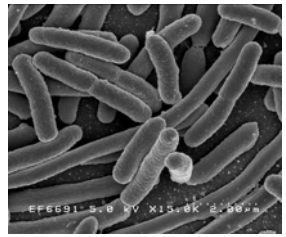
advantages V disadvantages



- e.g. *Escherichia coli*, *Lactococcus lactis*, *Bacillus* species….
- Widely used for expression of recombinant proteins
- Easy manipulation/transformation, rapid growth, simple nutrient requirements
- Many commercial vectors (*e.g.* Invitrogen, Novagen, Stratagene) available with a variety of N- and C-terminal tags to facilitate purification (*e.g.* His tag, GST fusions, Trx fusions)
- Well suited for proteins to be used for Ab production, structural, functional studies

- Many proteins expressed in inclusion bodies
- No post-translational modifications
- Improper folding of disulphide linked proteins
- Occasional problems with removal of fusion partner
- High endotoxin content with G-ve
  - (Reichelt et al. (2005) Single step protocol to purify recombinant proteins with low endotoxin content. Prot. Expr. Purif.

E. coli

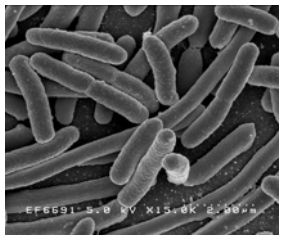


- Try different hosts when optimising expression
  - (e.g. protease negative)
- Expression levels can vary depending on strain choice

Strain	Derivation	Key Feature(s)	Antibiotic Resistance	Available as Competent Cells
Rosetta	Tuner™ (B)	BL21 *lacYZ* deletion, Lacks *lon* and *ompT* proteases	Cam	yes
Rosetta(DE3)			Cam	yes
Rosetta(DE3)pLysS			Cam	yes
Rosetta(DE3)pLacI			Cam	yes
RosettaBlue™	NovaBlue (K-12)	*recA, endA, lacI*^{q} High transformation efficiency	Tet + Cam	yes
RosettaBlue(DE3)			Tet + Cam	yes
RosettaBlue(DE3)pLysS			Tet + Cam	yes
RosettaBlue(DE3)pLacI			Tet + Cam	yes
Rosetta-gami™	Origami™ (K-12)	*trxB/gor* mutant, greatly facilitates cytoplasmic disulfide bond formation	Kan + Tet + Cam	yes
Rosetta-gami(DE3)			Kan + Tet + Cam	yes
Rosetta-gami(DE3)pLysS			Kan + Tet + Cam	yes
Rosetta-gami(DE3)pLacI			Kan + Tet + Cam	yes

Novagen

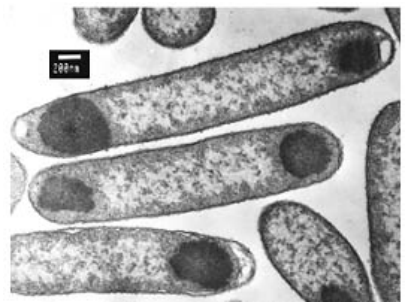
E. coli



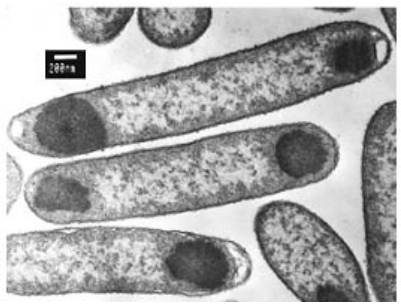
- Fast, easy, cheap
- 1-4 ml expression trials
- Most frequently used method for intracell proteins
- For secreted proteins (requiring S=S)
  - Target to periplasm
  - Express in *gor* or *trxB* mutants

Expression problems

- Protein is insoluble
  - Collect inclusion bodies and refold protein
  - Reduce growth temperature
  - Use heat shock to induce chaperones
  - Use a low/moderate copy number plasmid vector
  - Fuse a periplasmic targetting sequence to N-terminus
  - Co-express chaperones/foldases (PDI…)
  - Try a different fusion partner (e.g. MBP)



Refolding proteins from inclusion bodies



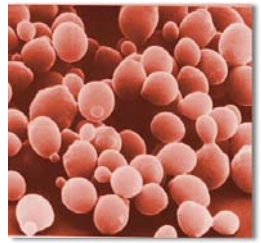
- Commercially available kits
  - e.g. iFOLD (Novagen), Pro-Matrix Protein Refolding Kit **(**Pierce)
  - 96-well protein refolding buffer matrix
  - Refolding conditions based on extensive literature review and REFOLD database **refold**.med.monash.edu.au/

Protein expression in bacteria

Cloning of a gene of interest into a bacterial expression vector	•Amplification/isolation of gene •Sequence gene •Mini-prep	**1-2 weeks**
Generation and identification of colonies expressing target protein	•Transformation of recombinant constructs into bacterial expression strain •Mini-induction to overexpress target protein •Test for expression of recombinant protein	**1-2 weeks**
Large scale culture	1L culture Induce Harvest	**2 days**
Purification	Protein purification using affinity chromatography: Ni-NTA (for His-tags) Glutathione (for GST-tags)	**1 week**

Yeast expression systems

(e.g.*Pichia pastoris*, *Saccharomyces cerevisiae*)



- Advantages
  - Can be grown to high densities
  - Simple fermentation and transformation
  - Well characterised systems
  - Lacks detectable endotoxins
- Disadvantages
  - Limited post-translational modifications
  - Glycosylation not identical to mammalian systems
  - Gene expression less easily controled than bacteria

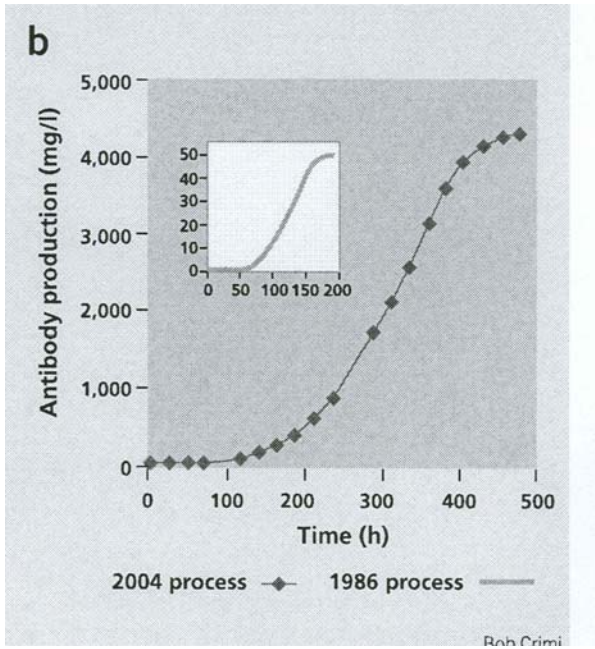
Cultured insect cells – *Baculovirus* vector

- Advantages
  - Facilitates glycosylation and formation of disulphide bonds
  - More complex post-translational modifications achieved
  - Proper folding of proteins of mammalian origin
  - Bv received FDA approval for clinical trials
  - Yields good
- Disadvantages
  - Expensive
  - More complex fermentaion required
  - Different glycosylation patterns to mammalian cells
  - Product not always fully functional

Protein *N* Glycosylation in the Baculovirus–Insect Cell Expression System and Engineering of Insect Cells to Produce "Mammalianized" Recombinant Glycoproteins (2006) Advances in Virus Research 68,159-191

Mammalian cells

- Dominant system for production of 60-70% proteins for therapeutic use
  - CHO, mouse myeloma (NSO), baby hamster kidney (BHK), human embryo kidney (HEK-293), human retinal cells – regulatory approval granted
- All PTM, proper folding, assembly
- Improvements in process developemnt, media composition, vector and host cell engineering have resulted in good yields (g/l)



Nature Biotech (2004) 22: 1393-8 Production of recombinant protein therapeutics in cultivated mammalian cells.

Other expression systems

- In vitro transcription/translation

  Host-less – cell free

  Several lysate sources available

  E. coli, yeast, wheat germ, insect cell, rabbit reticulocyte

  Advantages

  Small scale, parallel

  rapid testing

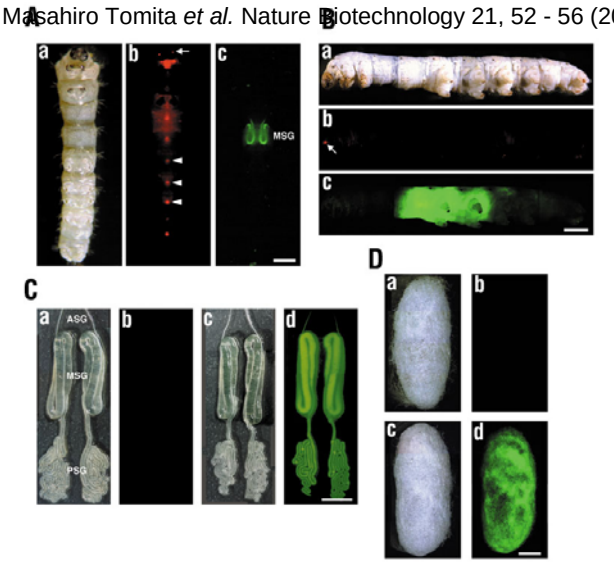
  Toxic proteins tolerated

  Amenable to isotopic labeling

- Others
  - Xenopus oocytes
  - Milk of transgenic animals
  - Transgenic plants
  - Transgenic Silkworm (5 kg p.a./1.5 m)

Transgenic silkworms produce recombinant human type III procollagen in cocoons

Masahiro Tomita *et al.* Nature Biotechnology 21, 52 - 56 (2003)

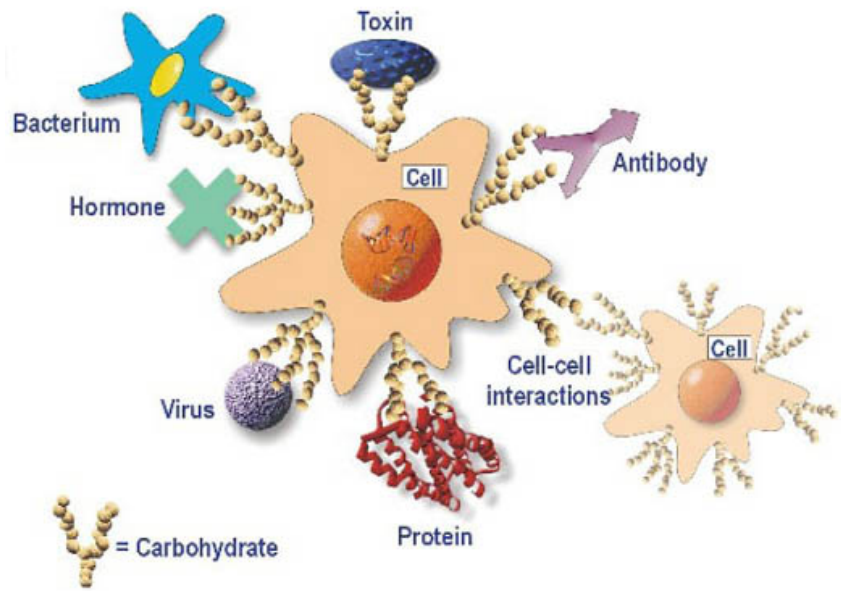


Expression systems: for functional membrane protein production

Processing	E. coli	Yeasts	Insect	Mammalian	Cell-free
Folding	+/=	+	++	+++	+/-
Glycosyl-ation	-	+/-	+	++	
Acylation	-	+	+	+	
Disulfide link	-	+	++	++	
Yield (mg/ml)	<0.5	<0.2	<0.02	<0.02	<10
% functional	1-10	1-50	10-90	10-90	1-10

State-of-the-art in protein expression

- Protein expression by engineering of yeast, plant and animal cells (Fussenegger and Hauser (2007) Current Opinion in Biotechnology 18, 385)



'Glycosylation engineering in yeast: the advent of fully humanized yeast' (387-392)

Table 1 The potential role and effect of the glycocomponent of glycoproteins

Role/effect	Comment
Protein folding	Glycosylation can affect local protein secondary structure and help direct folding of the polypeptide chain.
Protein targeting/trafficking	The glycocomponent can participate in the sorting/directing of a protein to its final destination.
Ligand recognition/binding	The carbohydrate content of antibodies, for example, function in antibody binding to monocyte Fc receptors and interaction with complement component $C1_q$.
Biological activity	The carbohydrate side chain of gonadotrophins (specifically the α-subunit N^{52} side chain) is essential to the activation of gonadotrophin signal transduction.
Stability	Sugar side chains can potentially stabilize a glycoprotein in a number of ways including enhancing its solubility, shielding hydrophobic patches on its surface, protecting from proteolysis and directing participation in intrachain stabilizing interactions.
Regulates protein half-life	Large amounts of sialic acid can increase a glycoprotein's plasma half-life. Exposure of galactose residues can decrease plasma half-life by promoting uptake through hepatic galactose residues. Yeast N-glycosylation is of a high-mannose type, driving rapid removal from circulation through mannose receptors.
Immunogenicity	Some glycosylation motifs characteristic of plant-derived glycoproteins (often containing fucose and xylose residues) are highly immunogenic in mammals.

Nature Biotechnology (Oct 2006) 24, 1241-52

Characteristic	Bacteria	Yeast	Baculovirus	Mammalian
Cell growth	Rapid (30min)	Rapid (90min)	Slow (18-24h)	Slow (24h)
Med. Complexity	Minimum	Minimum	Complex	Complex
Cost	Low	Low	High	High
Expression	High	Low-High	Low-High	Low-Medium

Post-translational modifications (PTMs)

Protein folding	Not reliable	Usually reliable	Very reliable	Very reliable
N-linked glycosylation	None	High mannose	Simple, no sialic acid	Complex
O-linked glycosylation	N	Y	Y	Y
Phosphorylation	N	Y	Y	Y
Acetylation	N	Y	Y	Y
Acylation	N	Y	Y	Y
γ-carboxylation	N	N	N	Y

Bottlenecks in Protein expression

Protein Expression Research Group (PERG)

Gene cloning	3
Bacterial expression	15
Insect cell expression	18
Yeast expression	3
Mammalian expression	21
Purification	42

Most frequent problem encountered with bacterial expression

Low/No expression	25%
Solubility/ Inclusions	75%

- How many expression constructs does your facility design and test for a given protein?

1	13%
2-5	49%
6-12	16%
12-24	10%
➢24	13%

How many expression conditions do you test to optimize expression from a given construct ?

1	6%
2-5	42%
6-12	35%
➢12	16%

Recombinant expression systems: the obstacle to helminth vaccines?

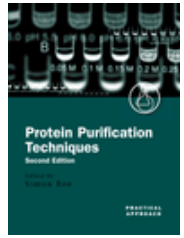
- Geldhof P et al. (2007) Trends Parasitol 23, 528
- Advocate thorough analysis of native antigens prior to testing recombinants

  – Immunogenicity -
  – Glycan structure -

Summary

- Consider protein class in selecting expression system
  - Secreted V intracell V membrane associated
  - PTMs required?
- If targeting a domain(s)
  - Align seq with orthologs and homologs
    - (Nayeem et al (2006) Protein Science 15, 808-24
  - Search for structures of related proteins
  - Select endpoints of domains
- Select tag/fusion partner, cleavage site
- Clone into appropriate vectors
- Test and optimise expression/purification

References



- Protein purification techniques: a practical approach (IRL press)
- Hellwig et al (2004) Plant cell cultures for the production of recombinant proteins. *Nat. Biotech*. 22, 1415-1422
- Gerngross T. (2004) Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat. Biotech*. 22, 1409-1414.
- Gerngross TU. Advances in the production of human therapeutic proteins in yeast and fungi (2004) *Nat. Biotech* 22, 1409
- Ulmer et al. (2006) Vaccine manufacturing: challenges and solutions. *Nat Biotech* 24, 1377
- Walsh G. (2006) Post-translational modifications in the context of therapeutic proteins. *Nat Biotech* 24, 1241.
- Kost TA et al. (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotech*. 23, 567.
- Sorensen (2005) Advanced genetic strategies for recombinant protein production in E. coli. *J. Bacteriol*. 115, 113-28.
- Recombinant protein folding and misfolding in *E. coli* (2004) *Nat. Biotech*. 22, 1399-1408
- Butt TR et al. (2005) SUMO fusion technology for difficult to express proteins. *Prot. Expr. Purif*. 43, 1-9
- Burckstummer et al (2006) An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells *Nature Methods* - **3**, 1013 - 1019 (2006)