Library Methods for Production of Soluble and Crystallisable Proteins and Domains

Pär Nordlund

Evitra AB and Division of Biophysics,
Department of Medical Biochemistry and Biophysics,
Karolinska Institutet

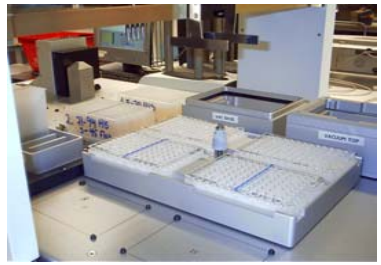
”Our simple (and affordable…) philosophy”

- **Get the maximum out of E.coli - the perfect host when it works !**

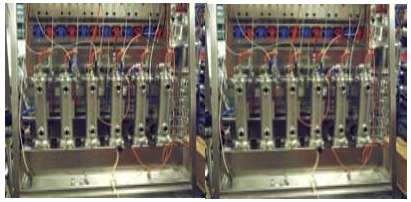
- **Appropriate benchmarking of technologies - large resources are wasted based on rumors and oversold technologies !**

Some platforms established in Stockholm

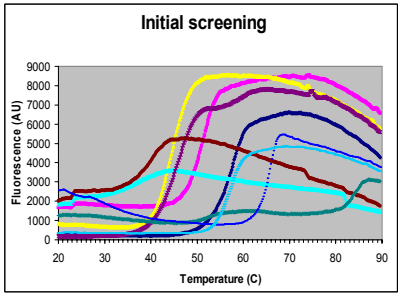
- **Parallel cloning and expression screening** => Gateway cloning, FiDo-screen =>



- **Parallel scale-up** => Belach, Greta Parallel fermentor =>



- **HTP-stability screen** => Thermoflour, TSA =>



Library technologies – a magic box or an ugly sink for resources ?

- **Difficult proteins can often be produced in *E.coli* after exhaustive and sequential efforts. Can take years...**

- **Simultanious screening of large libraries of expression variants could potentially constitute a magic ”one shot solution” ?**

- Different length constructs

- Random mutations

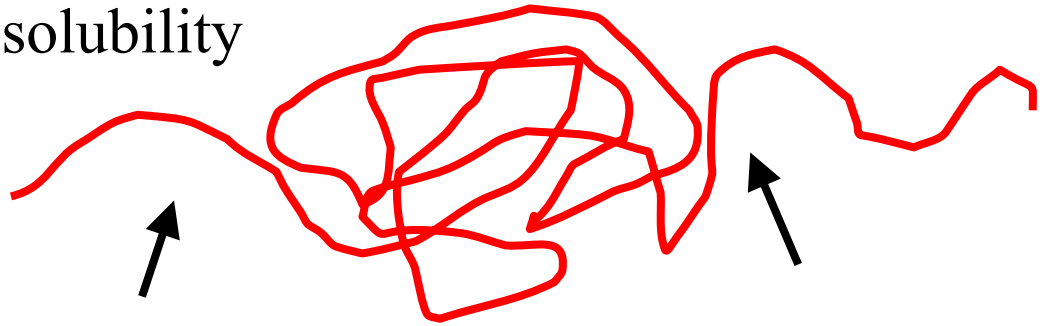
- Mixed pools of vectors, strains etc

- Mixed pools of genes/orthologes etc

Screening construct libraries - Why bother ?

The "right domain borders" can: .

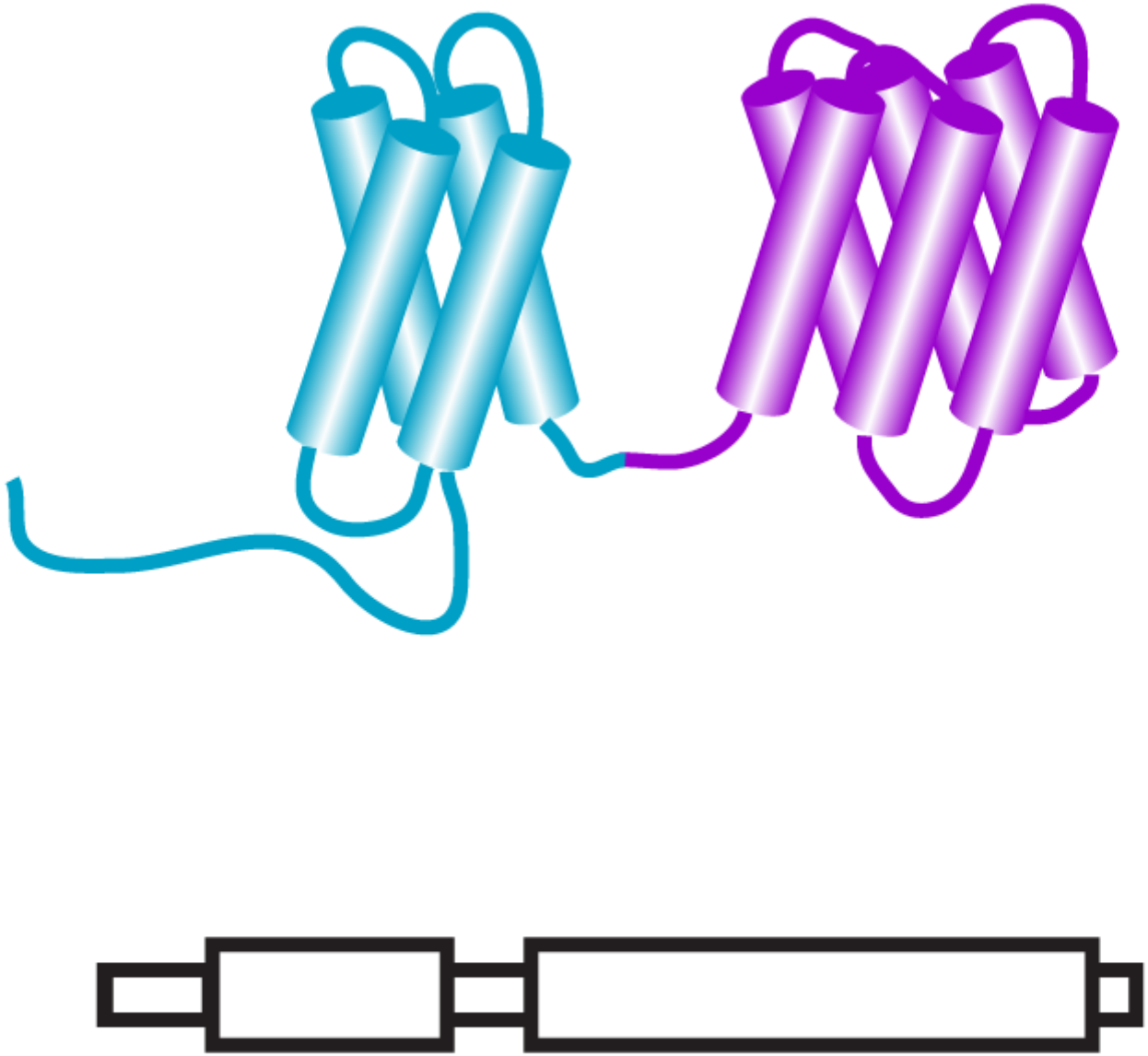
- improve expressibility and solubility
- improve crystallisibility

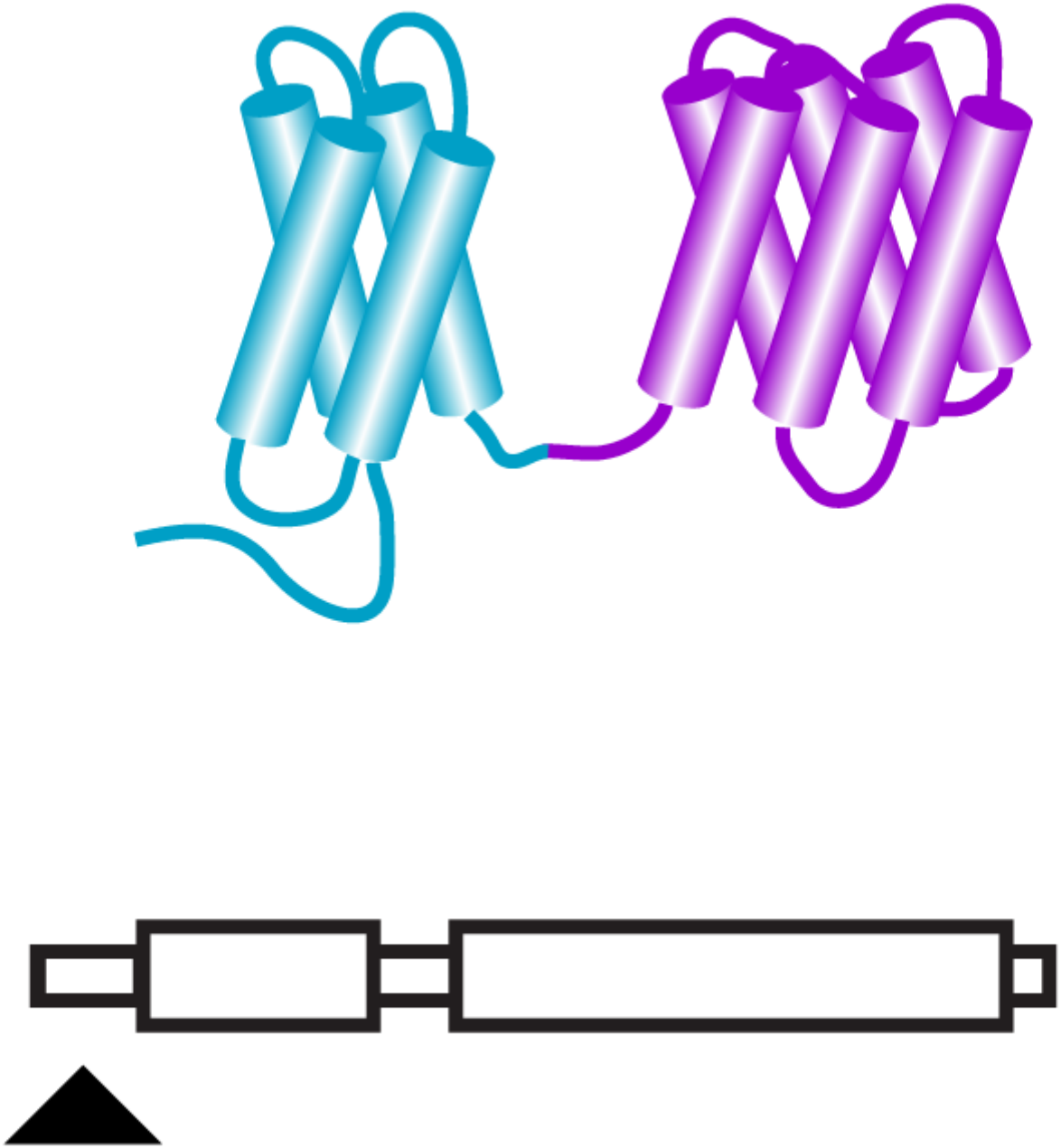


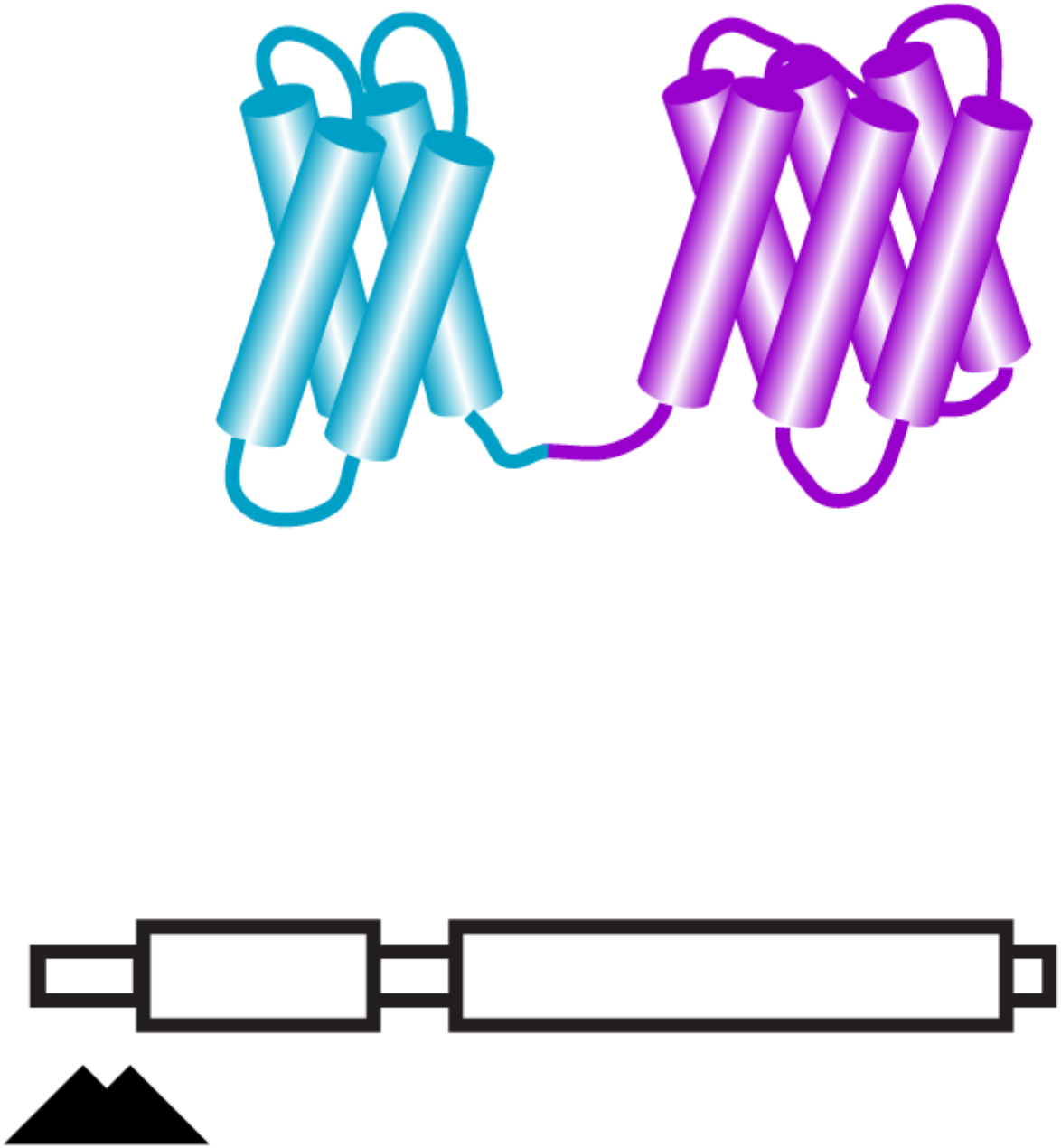
Current main strategies:

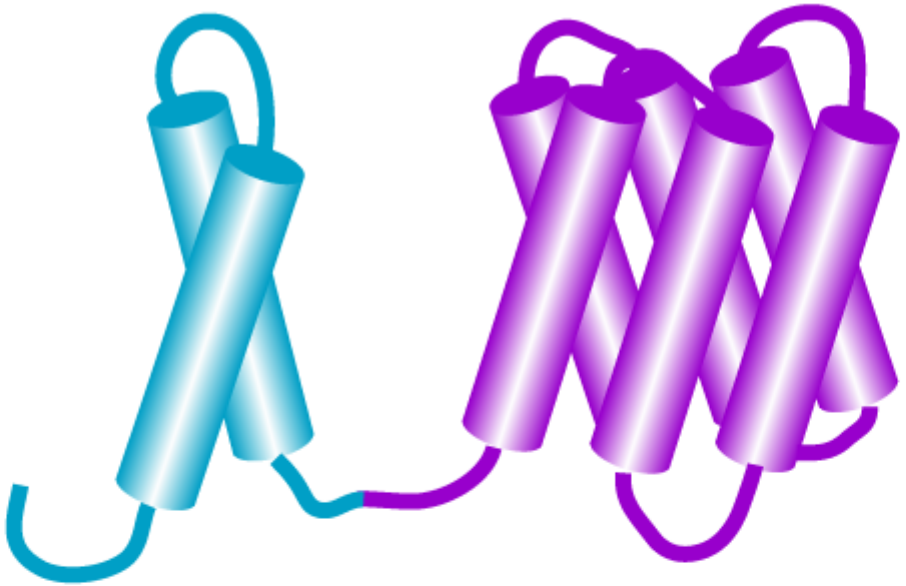
- Bioinformatics based domain border analysis
- Partial proteolysis => Mass Spec => re-cloning
- Often try 5-15 constructs of each protein

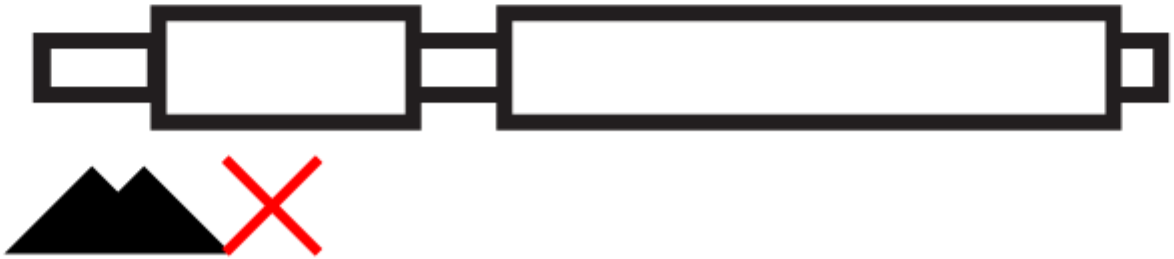
Let's test an N-terminal deletion construct library strategy !

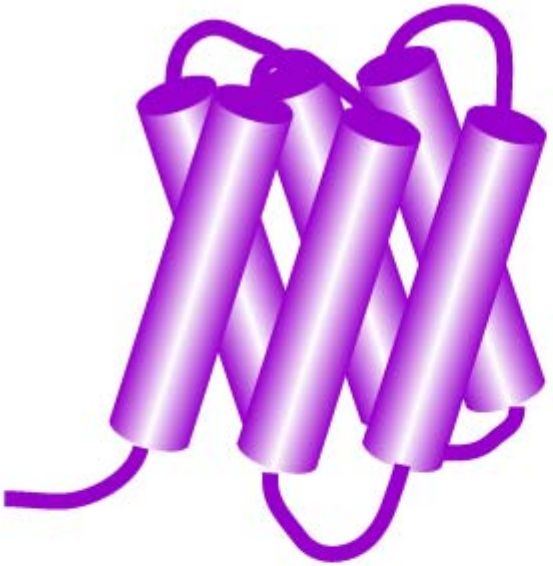


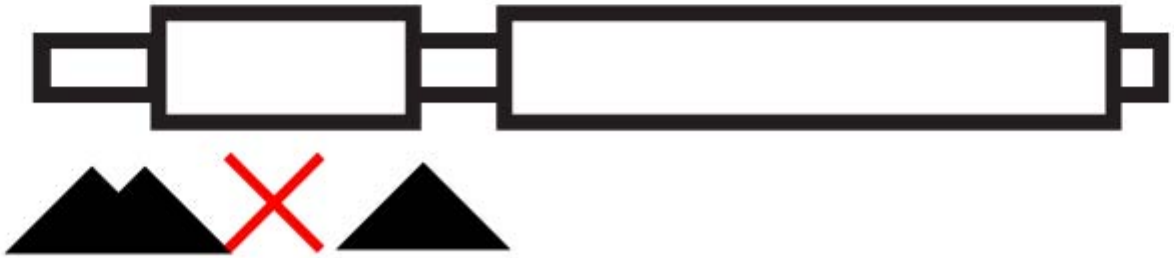


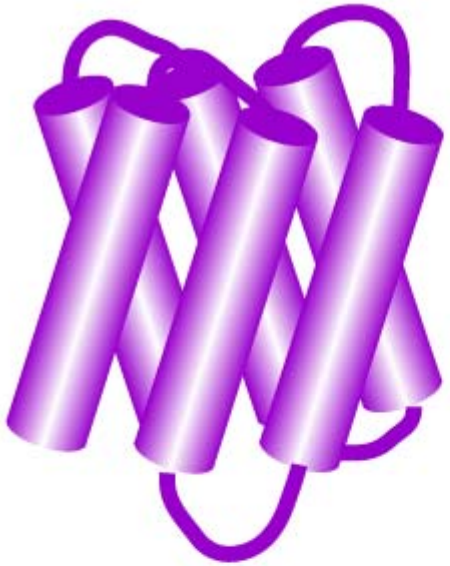


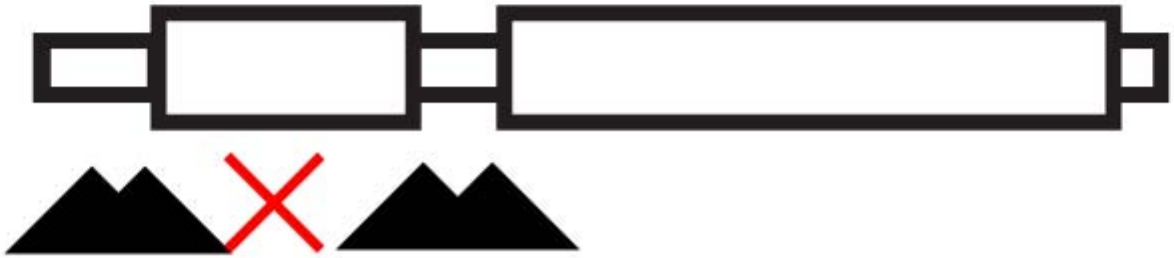


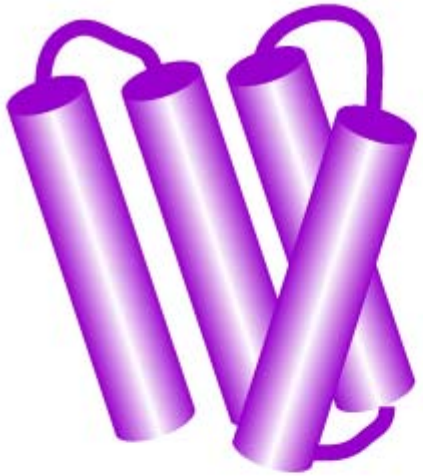


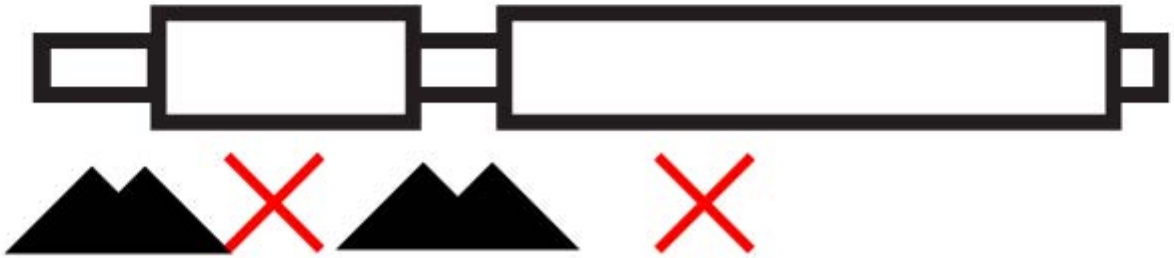


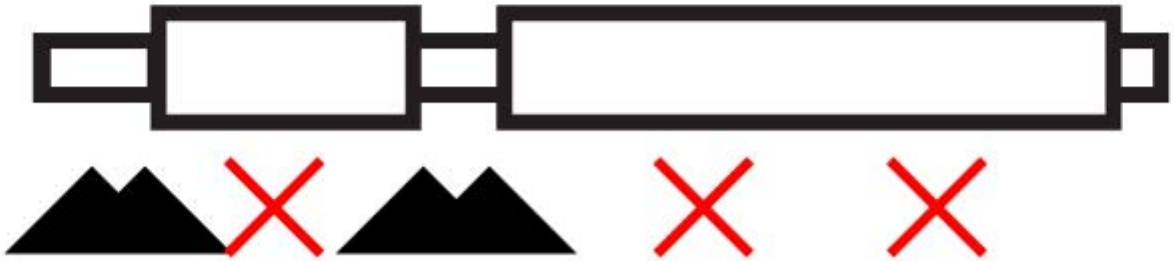




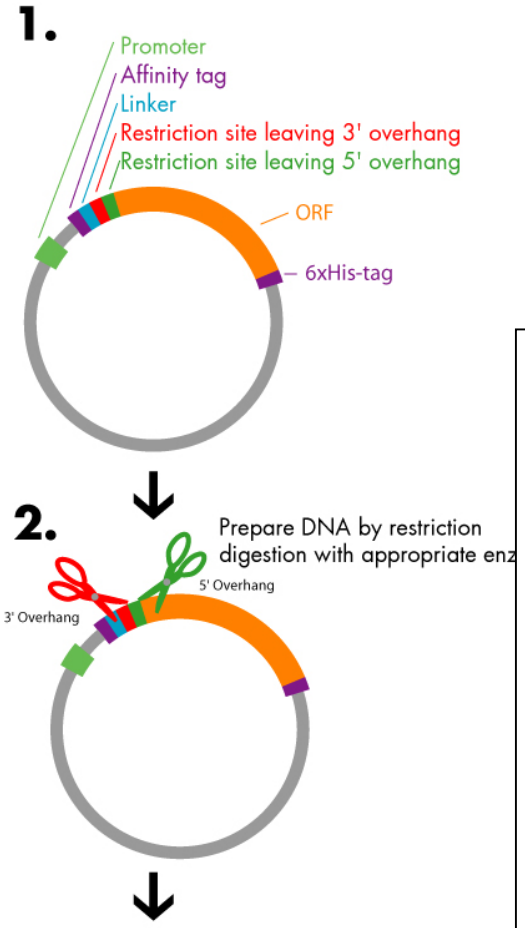


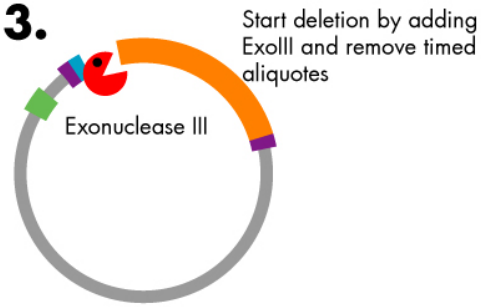




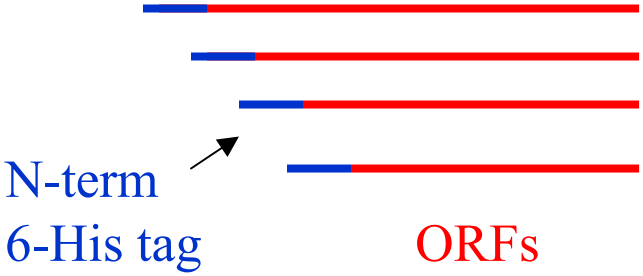


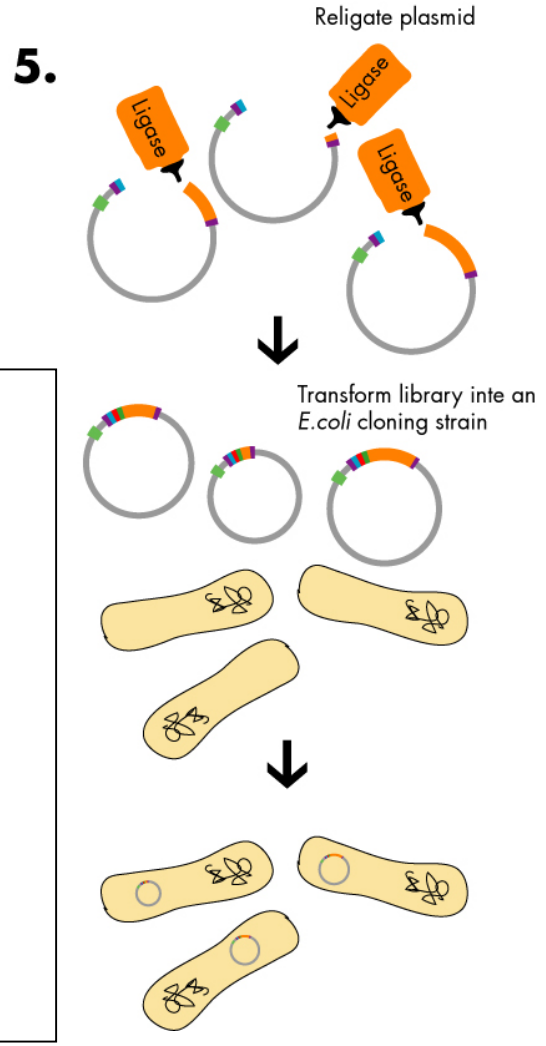
Erase-a-Base (E-a-B) for N-terminal deletion library generation





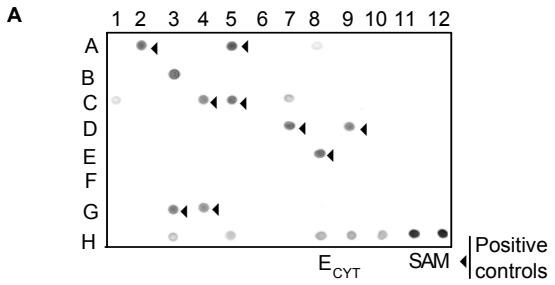
Generate a set of plasmids expressing ORFs with different N-term start



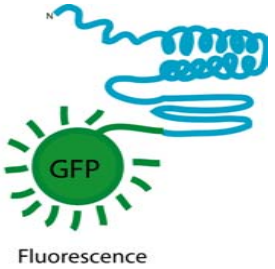


HTP solubility screening methods

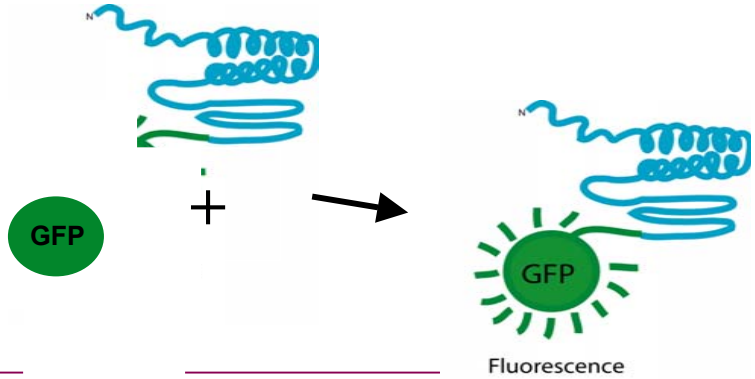
- **Direct small culture 96-well based methods**
  e.g. Ni-NTA affinity, FiDo-screen etc
  Problems:
  - Limited numbers of clones
  - Expensive and requires automation



- **Direct C-terminal fusion proteins as a folding markers**
  e.g GFP, CAT
  Problems:
  - Change solubility of target
  - Solubility can change upon lysis
  - Fusions have to be removed

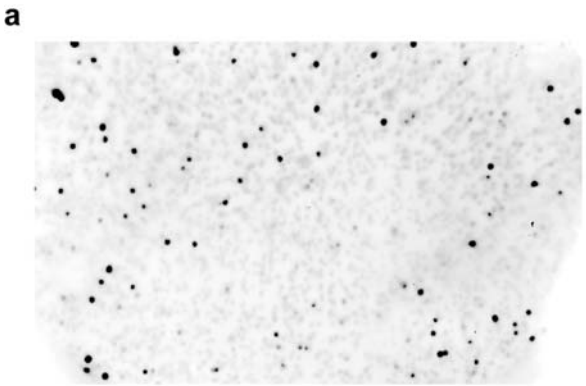


- **Split C-terminal fusion proteins**
  e.g. GFP, Lac-Z
  Problems:
  - Solubility can change upon lysis

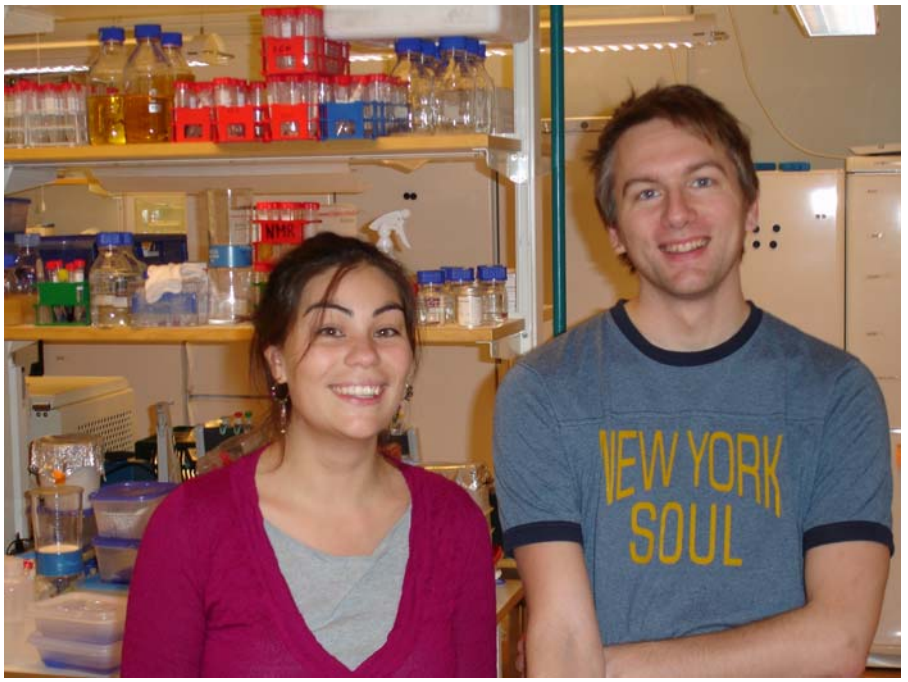


The Colony Filtration blot

Soluble proteins detected on colony level by applying 2-d filtration separation step followed by blotting of His-tag



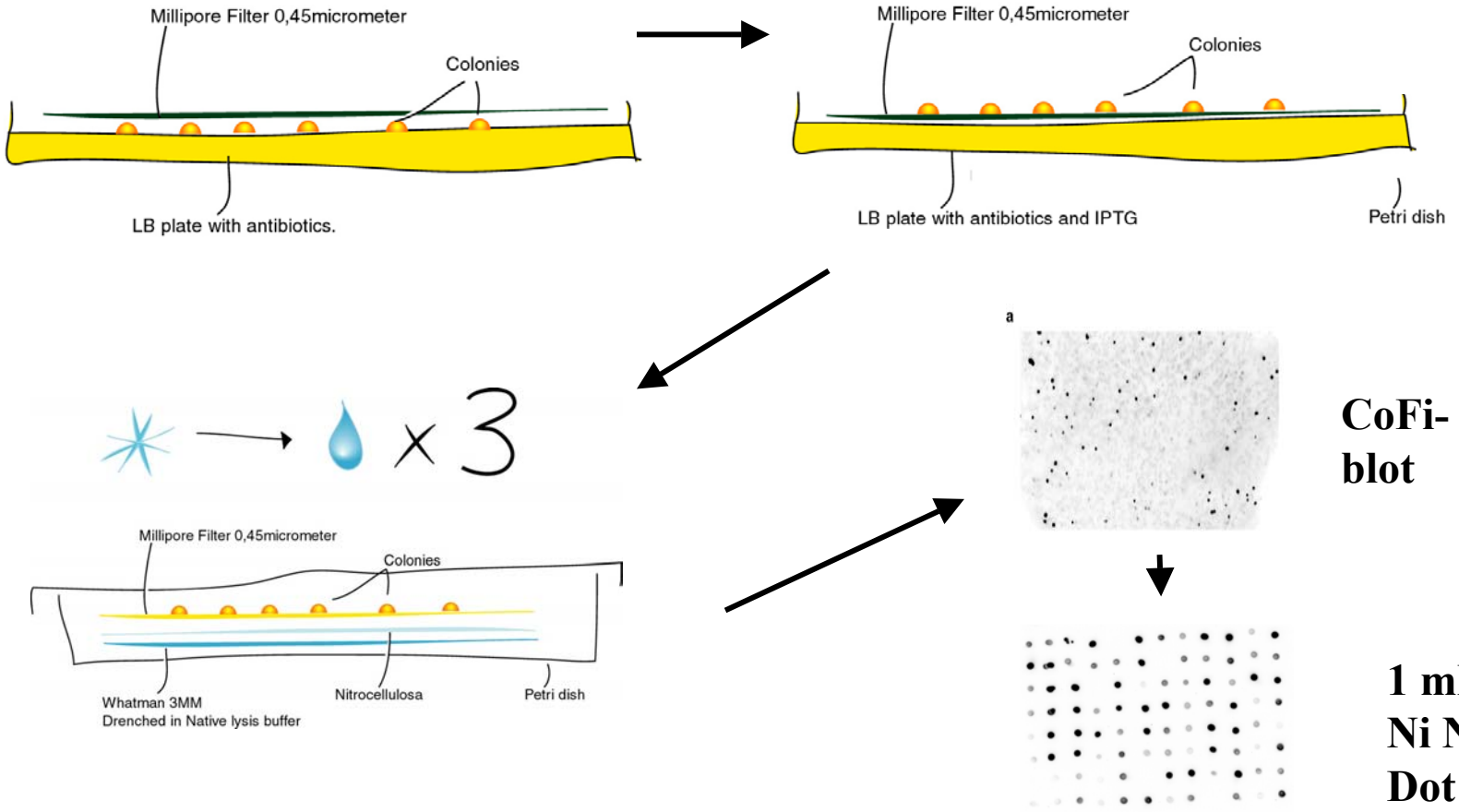
CoFi-blot of a ”deletion library” (Patent pending, www.Evitra.se)



Some applications of CoFi- blots

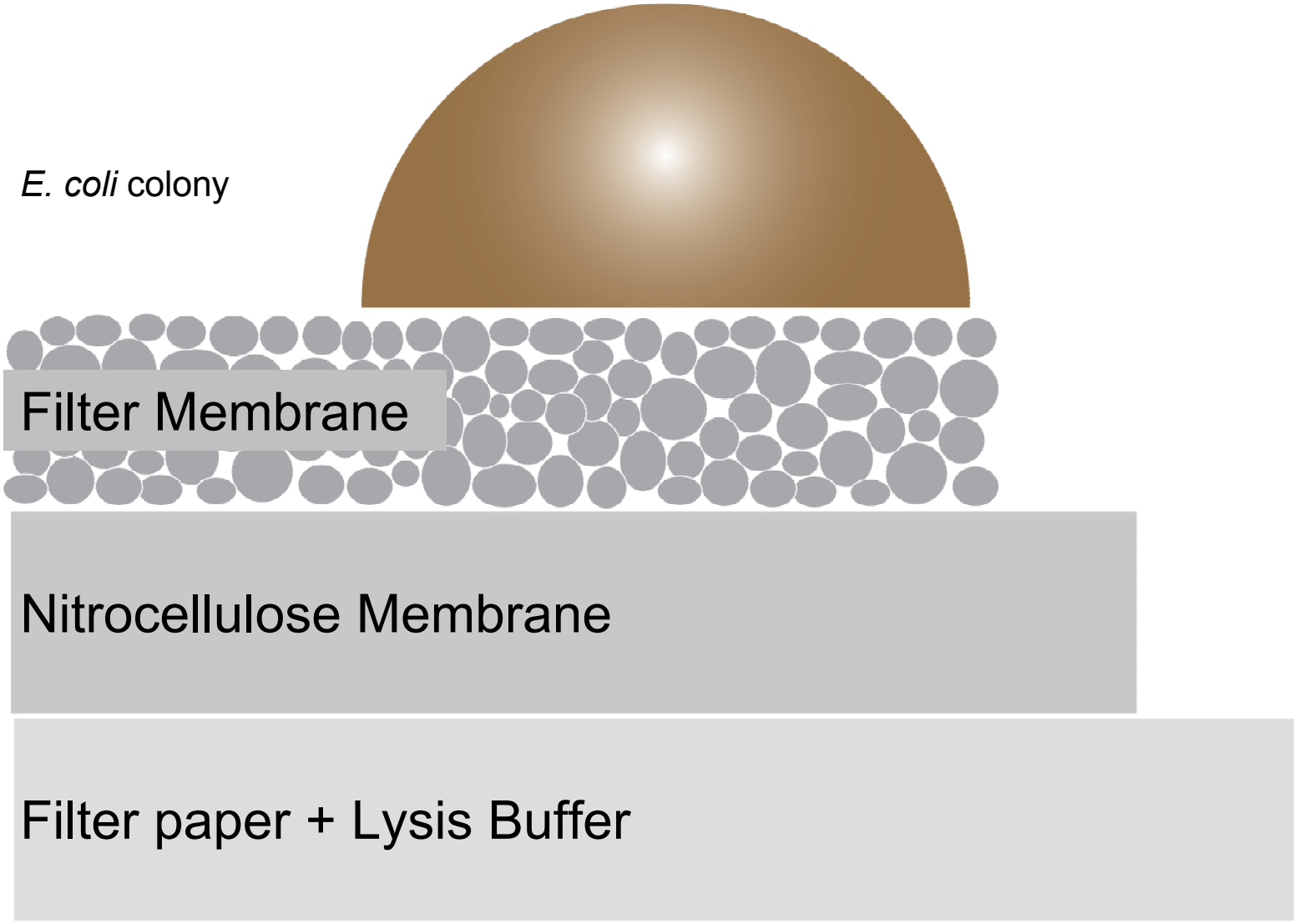
- Deletion libraries
- Random mutagenesis libraries
- Membrane proteins (deCoFi-blot)

The craft of the CoFi blot



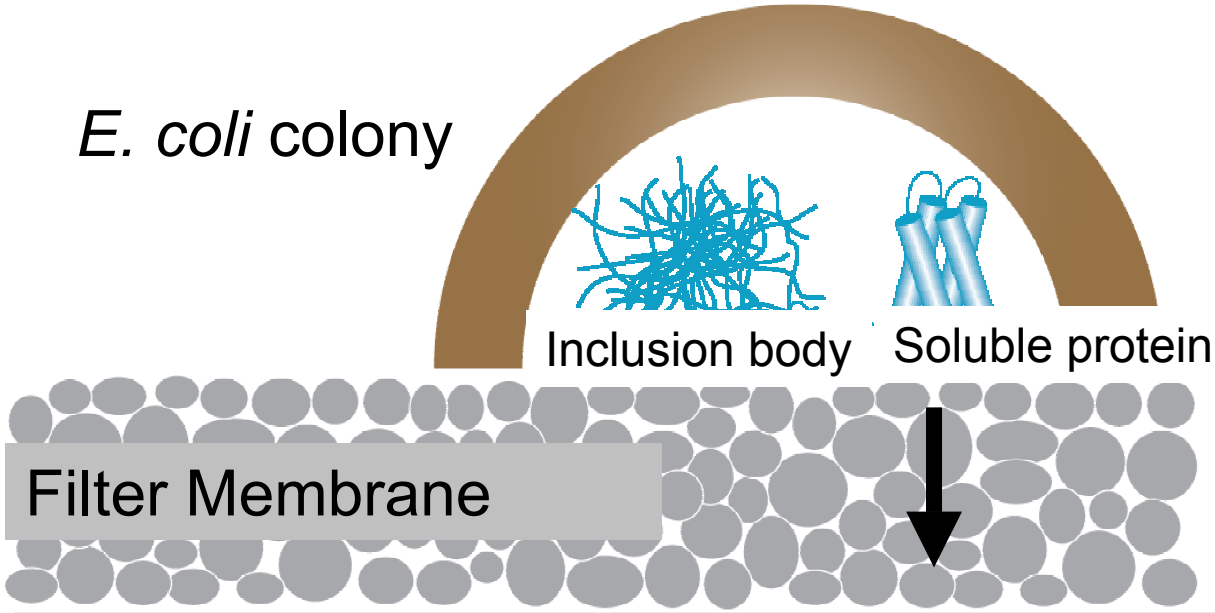
CoFi-blot showing colonies potentially express soluble protein.

Colony Filtration (CoFi) Blot



Evitra

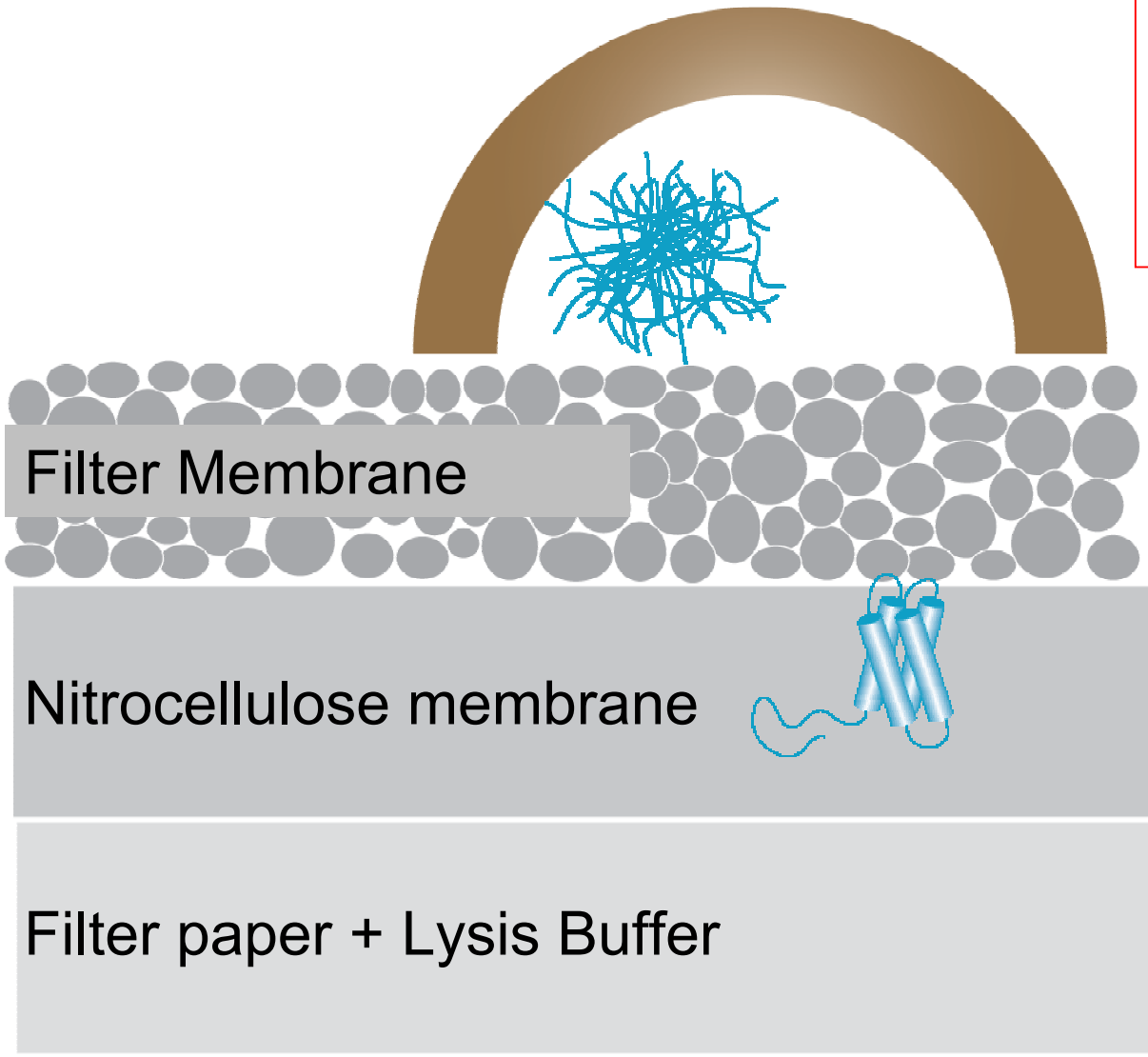
Lysis and filtration



Evitra

Lysis and filtration

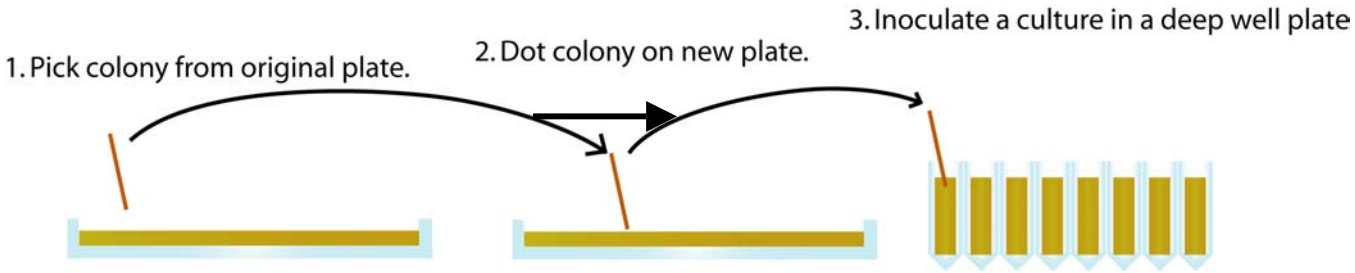
The lysis step and a Ni-conjugate probe simulate the scale-up process !

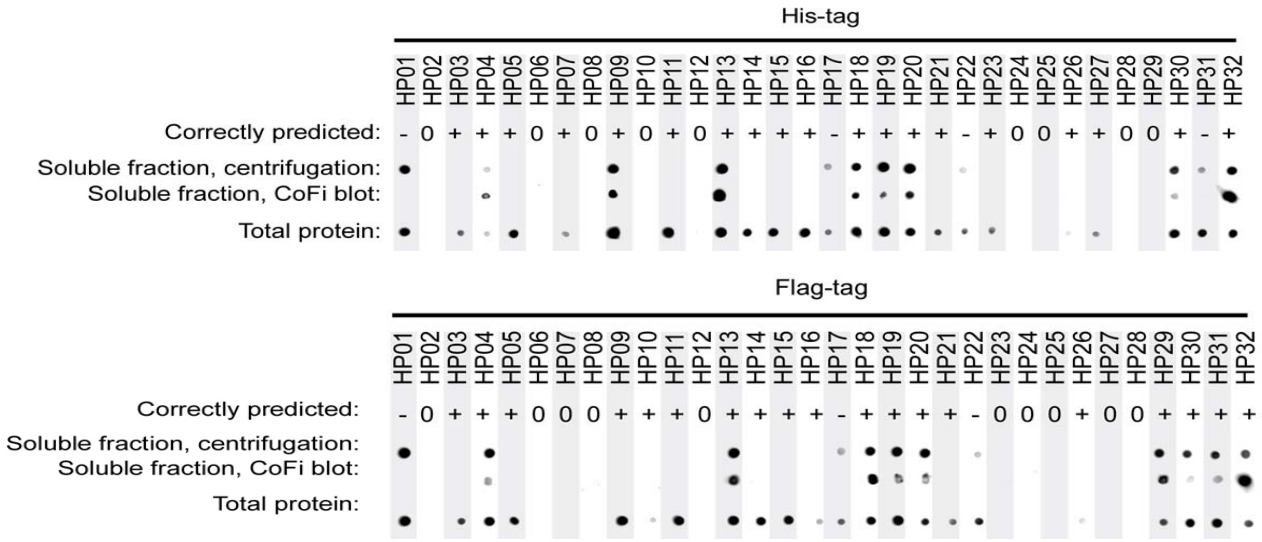


Nitrocellulose membrane is probed with antibody or Ni conjugate to identify colonies that express soluble protein

Evitra

Comparison of CoFi-blot and Centrifugation





Detection using INDIA HIS HRP

- 83 % (38 of 45) proteins with total expression are in agreement between the two methods! In all cases CoFi-blot was a "more stringent criteria" !

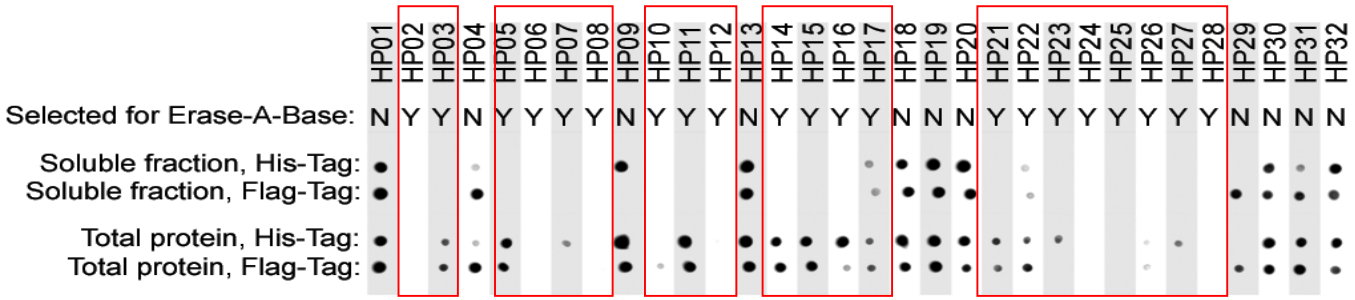
Benchmarking of the E-a-B => CoFi-blot strategy

By Tobias Cornvik
Sue-Li Dahlroth
Audur Magnusdottir
Victoria Lieu &
Monika Ekberg

Success rates before E-a-B => CoFi-blot:

- 32 Human/Mouse proteins used to benchmark the technology

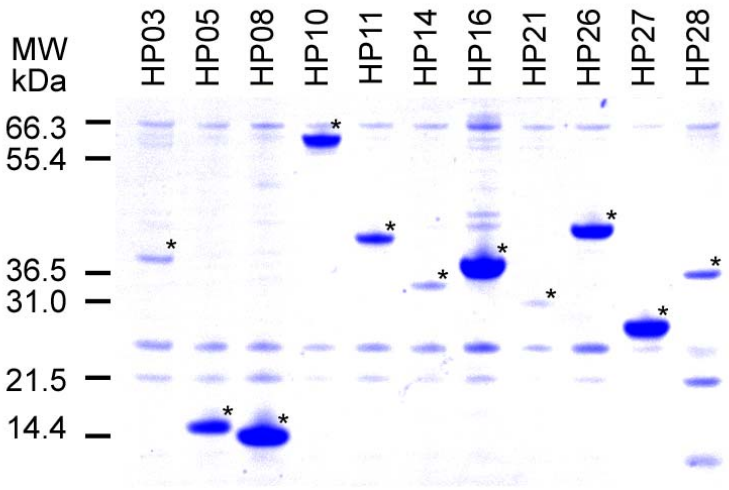
- 19 of 32 proteins do still not express soluble protein in 2 different vectors (N-term Flag and His) - i.e. ”hard to express proteins”



After E-a-B => CoFi-blot – many "hard to express proteins" can now be purified

- 11 of 19 "hard to produce" proteins could be "rescued" by the erase-a-base trick - at least one domain
4 more could be purified at lower level.

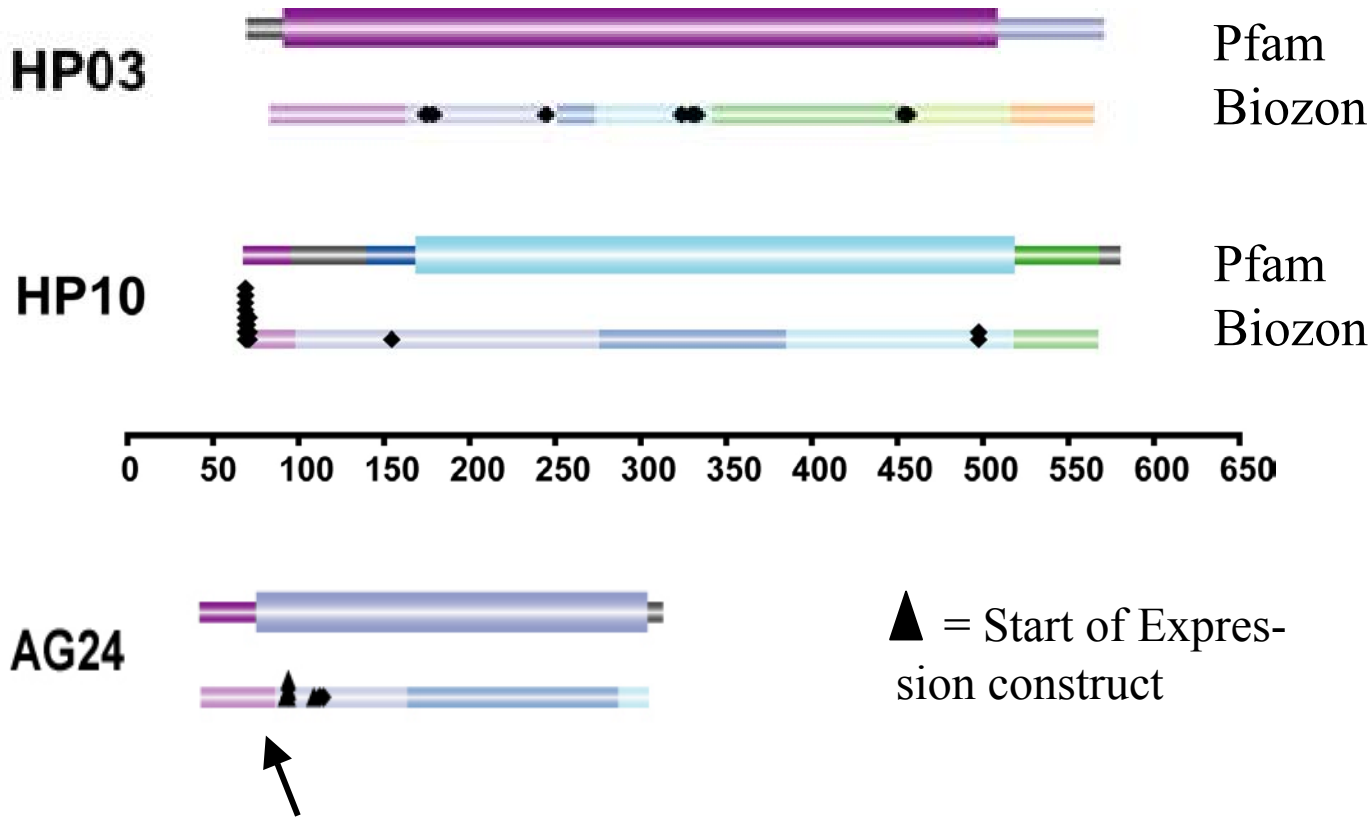
-



After a small-scale IMAC purification

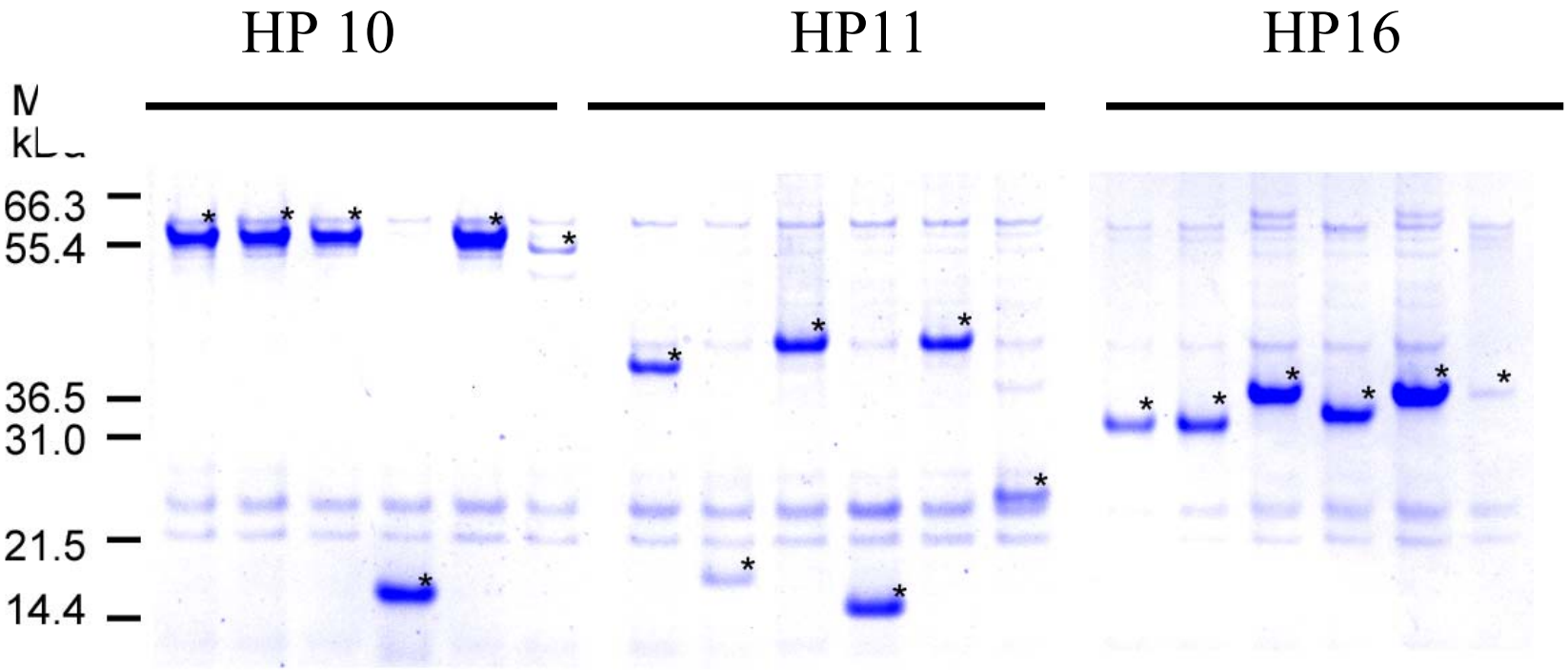
- Purified constructs have no/little proteolysis problems

E-a-B => CoFi-blot - a generic experimental domain foot-printing strategy



N-terminal mitochondrial signalling sequence is removed

A number of different length constructs are directly produced



- ”orthogonal constructs” will improve probability for crystallisation

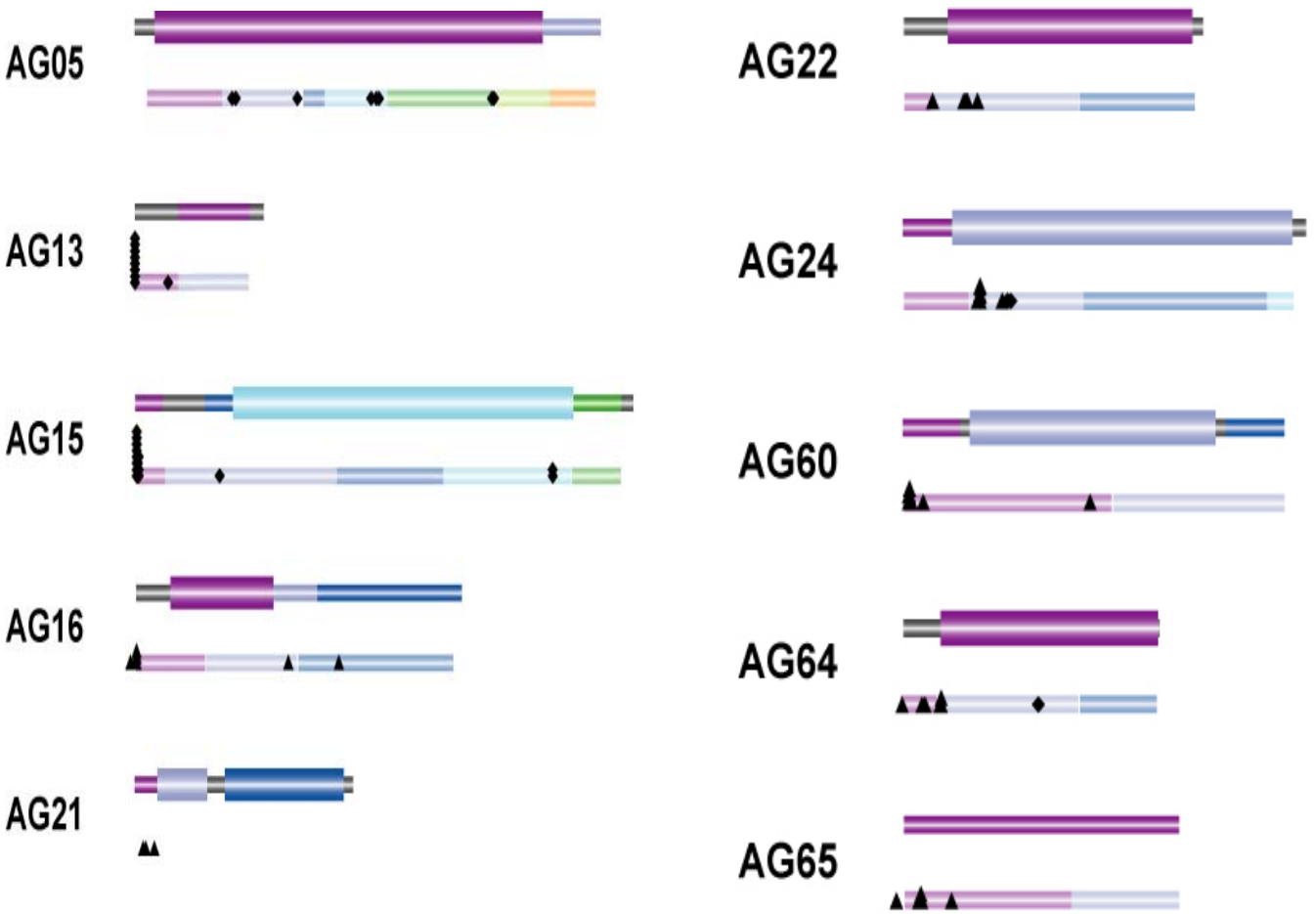
How useful are multi construct for crystallisation ?

SGC-Stockholm study in press in Prot Expr.& Purif (Gräslund et al)

- ~10 different constructs approach approach ⇔ rational 1-2 construct
- 15 proteins with diffracting crystals ⇔ 65 proteins with diffraction crystals

A limited number of construct taken into crytallisation trials give diffracting crystals for > 4 times more proteins

Most Human/Mouse proteins can be expressed close to full-length



Summary of Benchmarking - E-a-B constructs => CoFi-blot strategy

- **> 70 % of the starting 32 Human/Mouse proteins could be purified from *E.coli*.**
- **Most proteins express as nearly full length**
- **Comparison to Multi-Construct methods not made, but the E-a-B => CoFi-blot strategy is likely to work better for less characterised protein families.**
- ***E.coli* can potential produce many more eukaryotic proteins than we have anticipated**

Selection of an intriguing expression construct with a translational frame shift

Karolinska Institutet

By Martin Moche

DNA sequencing and N-terminal peptide sequencing of a positive construct of protein X selected with the E-a-B => CoFi-blot strategy

⇒ 4 bases are not translated by *E.coli* !

ATG AGA GGA TCG CAC CAC CAC CAC CAC G TGG CTG GTC ---
Met Arg Gly Ser His His His His His Trp Leu Val
R G S H H H H H L V

- Yields active enzyme with activity profile equal to wt enzyme !
(In frame reading yields 52 residue peptide)

The detergent adapted CoFI-blot - Selection of ”purifiable domains” of IMPs

Issues:

- How do you define a domain of an IMP ?

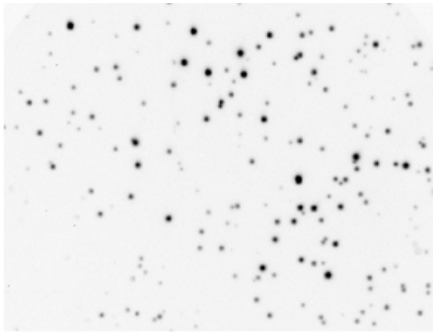
- What does a purifiable domain means when it is covered/protected by a detergent micelle?

Possible domain definitions:

- Single TM-helix

- Compact/”globular” domains

- Set of TM-helices which are “insertion competent”



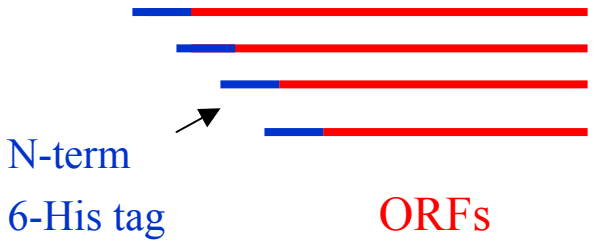
Detergent adapted CoFi-blot devel-oped by Marina I Sabet

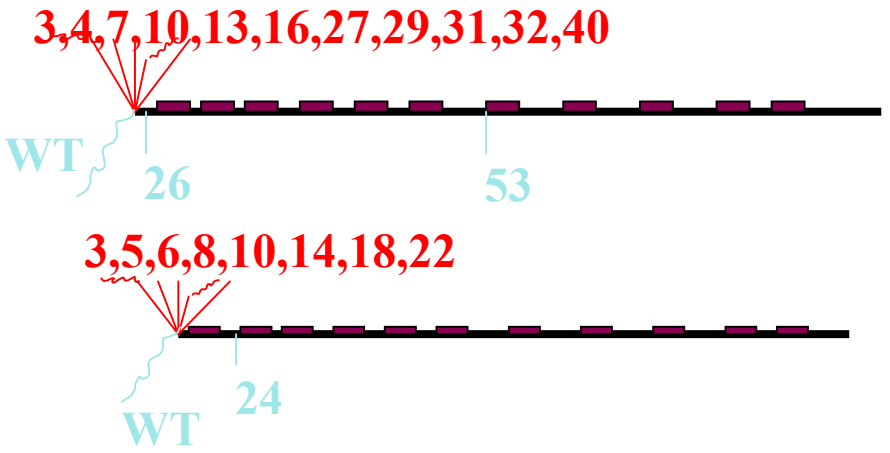
Construct library screening using the detergent adapted CoFi-blot

Karolinska Institutet

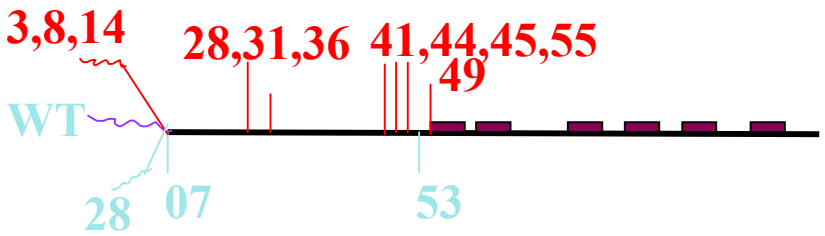
By Marina Igantuschencho-Sabet

Deletion library – ORFs with different N-termal start points



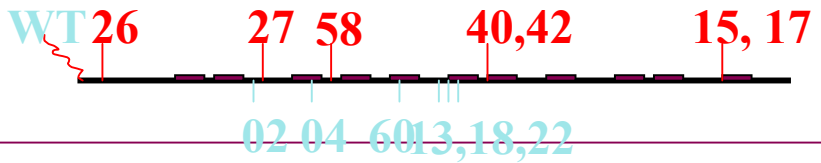


Purifiable IMP constructs selected from large libraries



- **Better expressors**
- **Worse expressors**

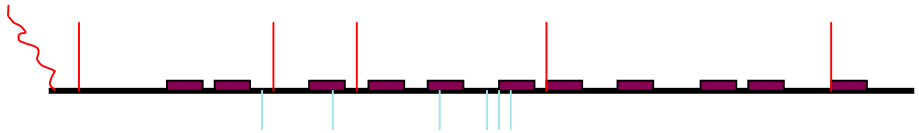
Constructs selected inbetween helices => membrane inserted proteins



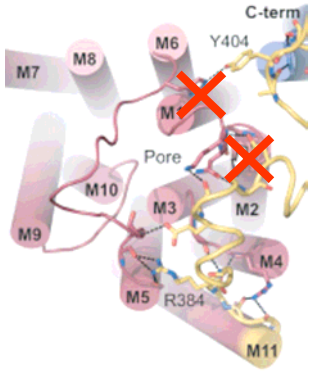
Selected domains of AmtB

WT D0 D1+ D2+ D3+ D4+

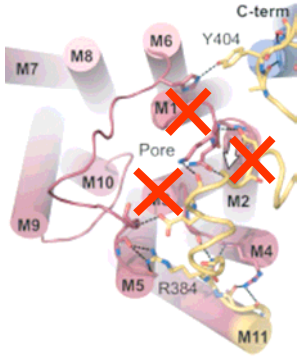
AmtB



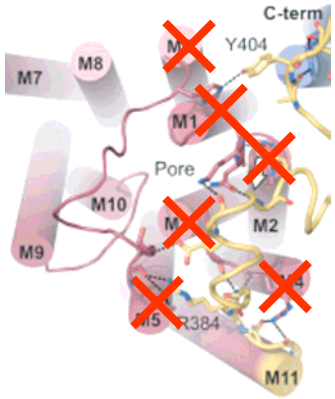
= Deleted helix



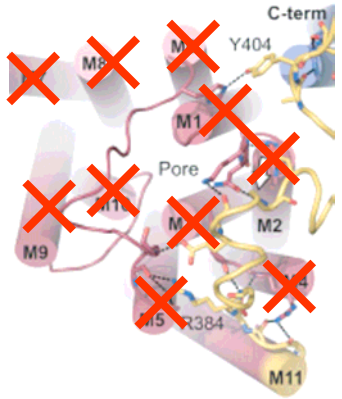
D1+



D2+



D3+

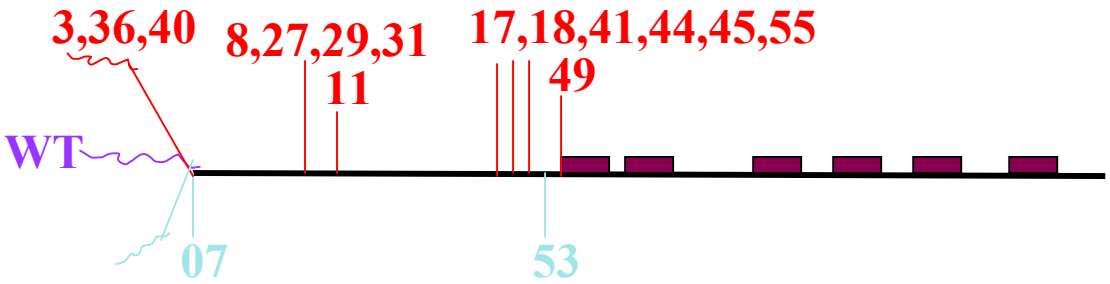


D4+

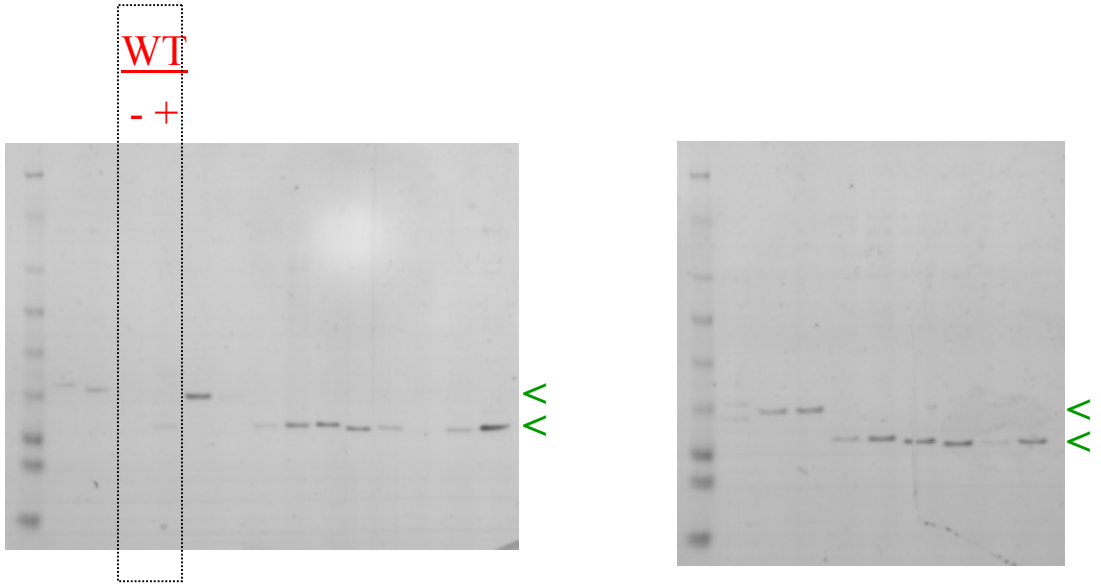
Pictures from
Conroy et al PNAS

- Selected domains show ”some tedency” to be ”globular” as defined from structure?
- Insertion might not be critical factor

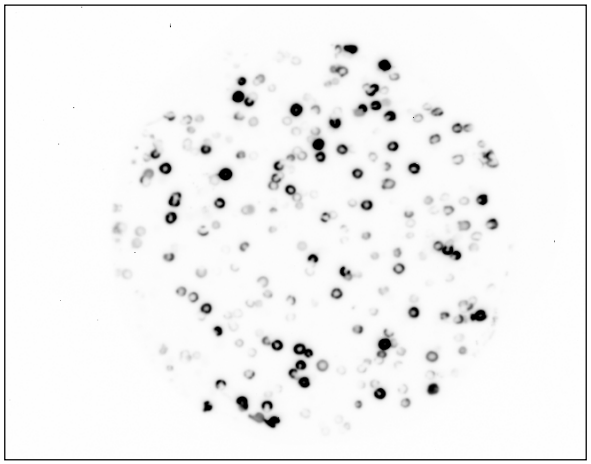
Example: EM29 and *E.coli* transporter



SDS-page from 1 ml cultures after IMAC purification



Rapid screening of solubilizing detergents of a GPCR library using the CoFi-blot



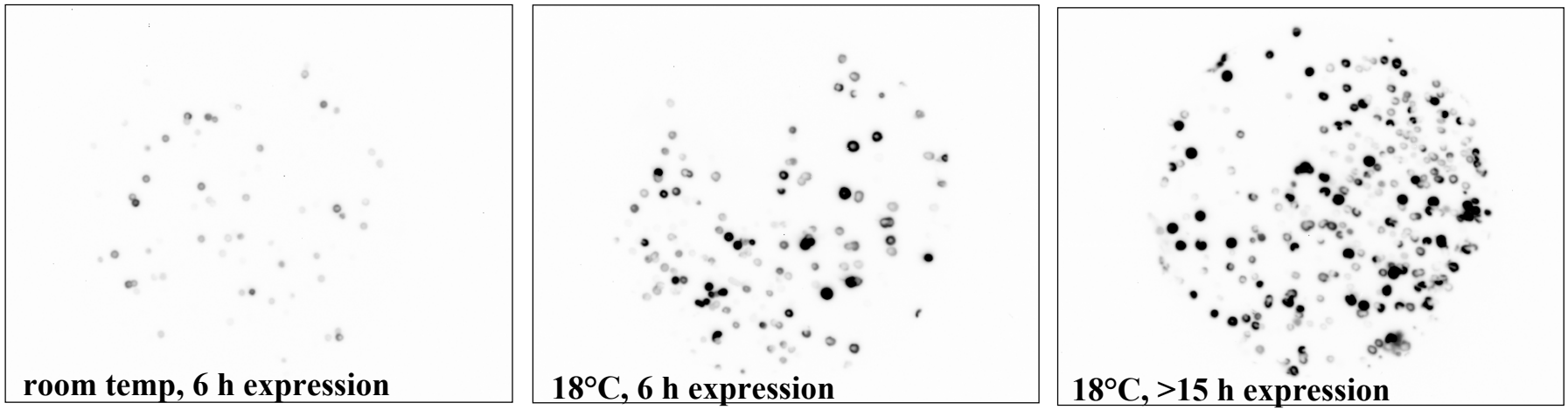
FC12 18°C expression

DM 18°C expression

CHAPS 18°C expression

Effects of different detergents

Rapid screening of expression conditions of a GPCR library in *E. coli*



FC12 in the lysis-buffer

Selection from random mutation libraries – mechanisms for potential improvments

- **Transcription/mRNA – e.g mRNA stability**
- **Translation - e.g. minimize RNA hairpins**
- **Folding - e.g. timing of folding events**
- **Protien stability – e.g. extra salt links**
- **Protein aggregation - e.g hydrophobic surfaces**
- **Protease resistance – e.g mutations in recognition site**

Test of a random mutagenesis => selection approch for IMPs

Karolinska Institutet

By Daniel Martinez-M Tobias Cornvik & Marina Ignatushchenko

- **9 IMPs used in a feasability study:**

  - 3 Ecoli IMPs expressing at medium level

  - 5 Ecoli IMPs expressing at low level or not at all

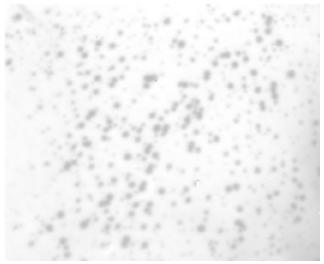
  - 1 Human MGST family member, medium level

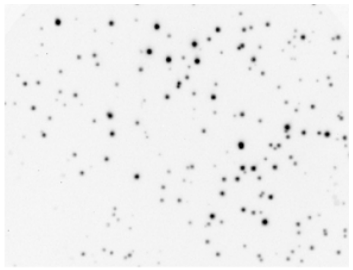
- **Random libraries were generated - 0.6 mutations/ 100 aa**

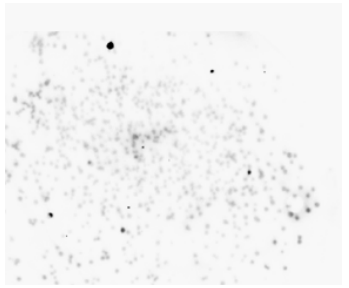
- **Selection made using a detergent adapted CoFi-blot**

Examples of detergent adapted CoFi-blots of IMPs

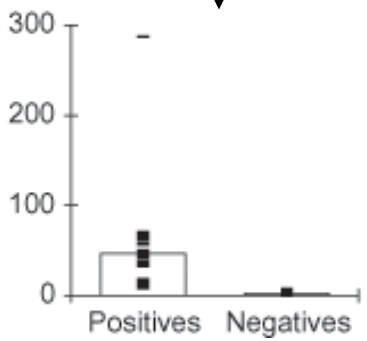
IMP1 **IMP4** **IMP5** **IMP7**

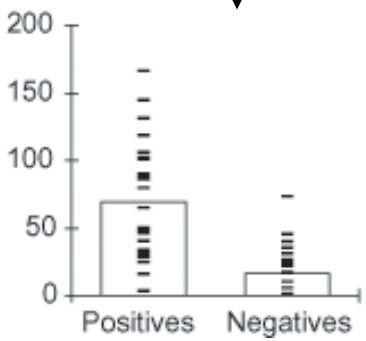


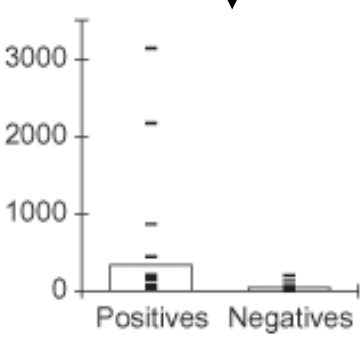




No positive colonies







- Quantitatification of expression using Small-scale IMAC- purification

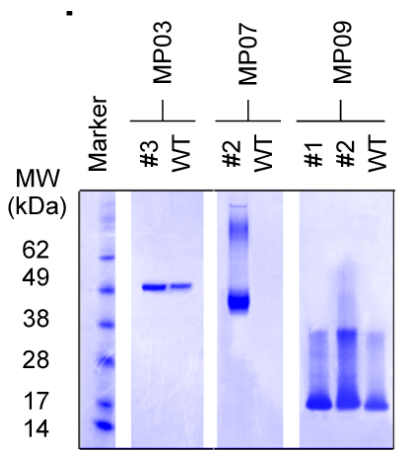
Result - feasability study

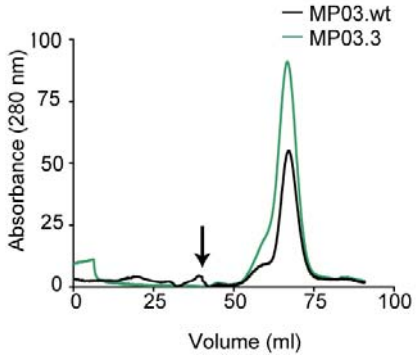
Low/Non expressors:

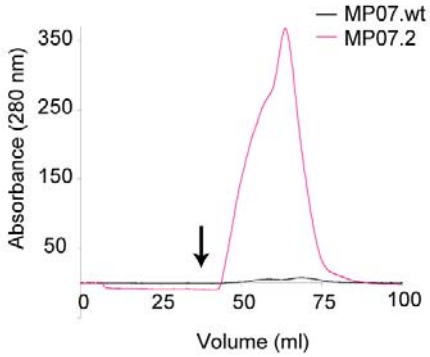
4 No positive colonies
1 Increased expression 4000% fold

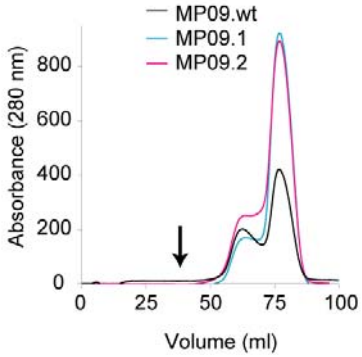
Medium expressers;

5 Increased levels of 40-300 % (one Human)





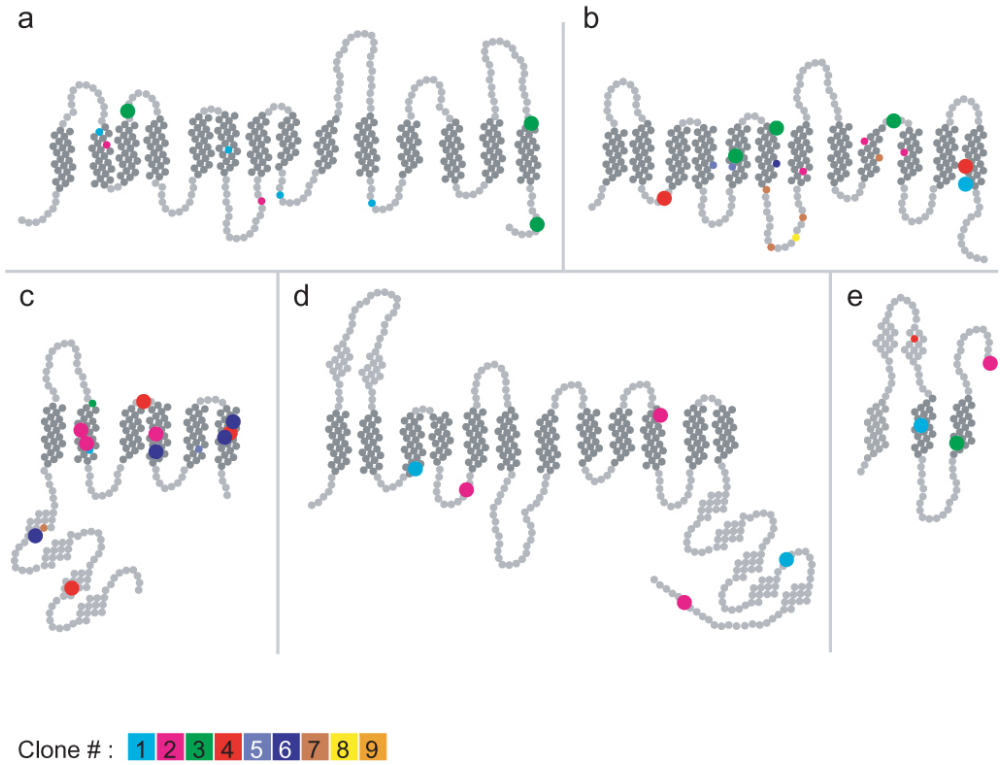


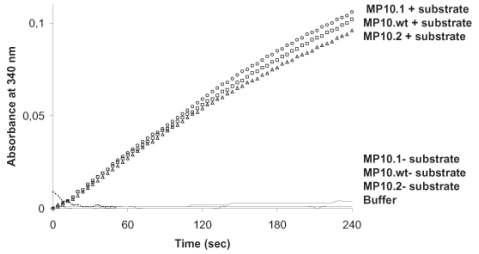


Examples of scale-up for three selected IMPs

What caused the improvment in expression levels ?

- **No strong positional preference for selected mutations**



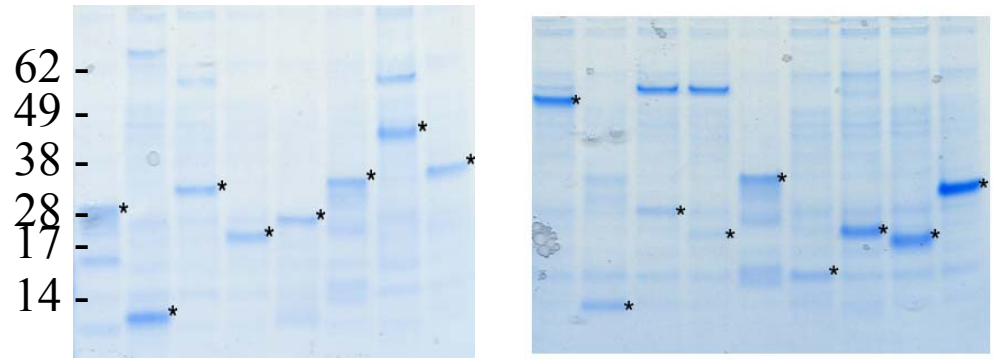


Evolution do not effect activity of the Human MGST family member

EaB + CoFi-blot (= Spotlight) is a generic, robust and relatively fast procedure

Human and mouse proteins	**> 80 processed**	**- > 15 ongoing**
VIZIER – RNA-virus proteins	**> 25 processed**	**- > 60 ongoing**
VIRCIR - Herpes virus proteins	**> 80 processed**	**- > 100 ongoing**

1 ml scale Herpes protein expression after IMAC purification



Summary – library work

- **The CoFi-blot constitutes an efficient HTP tool for screening for soluble protein expression**
- **Requires no automation – ”in everyone's hands”**
- **CoFi-blot ”simulate scale-up” in contrast to fusion reporter blots**
- **Screening of N-terminal expression libraries, yields dramatic improvements of success rates for mammalian proteins (intra cellular)**
- **CoFi-blot works for IMPs !**
- **Allows larger numbers of proteins to be screened => applicable ”genome wide studies”.**

Evitra

Karolinska Institutet

- Offers licensing or Contract Research

- Have completed successful projects for several major pharma and biotech companies.

- Spotlight – soluble proteins
- SpotlightM – membrane proteins

www.evitra.se
info@evitra.se
anders.bratt@evitra.se

People in the HTP developments at Karolinska Institute

Karolinska Institutet

Expression technologies:

Benita Engvall
Marie Hedrén

Direct evolution:

Tobias Cornvik
Sue-Li Dahlroth
Audur Magnusdottir
Victora Lieu
Monika Ekberg (PL)
Lola Herman

Biophysics:

Ulrika Ericsson

Membrane proteins:

Marina Ignatuchshenko
Said Eshaghi
Daniel Martinez
Damian Niegowski

Also thanks to:

Lars Mörtzel et al;
Belach AB
Niek Dekker, A-Z
Scott Leslie, JCSG
Juergen Haas, Munich

Support from:

EU - SPINE
EU - E-MEP
WCN
A-Z
GGs-fundation