Protein isotopic enrichment for NMR studies

Mariapina D'Onofrio

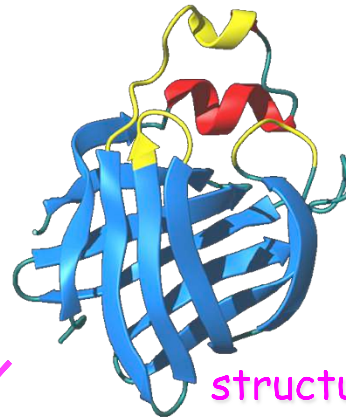
Corso Risonanza magnetica nucleare e imaging

Laurea magistrale LM9 2014

Protein NMR studies

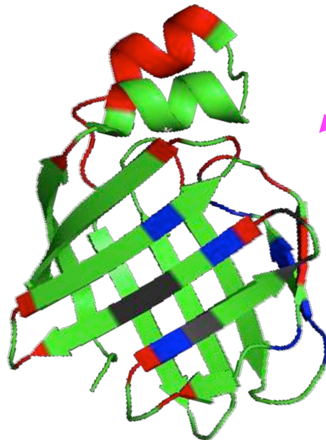
ARTGKYVDES

sequence

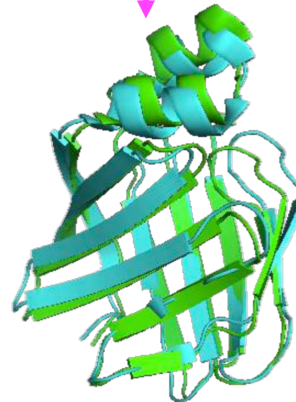


structure

Structure of protein-ligand, protein-protein complexes



Binding of molecules (perturbation mapping)

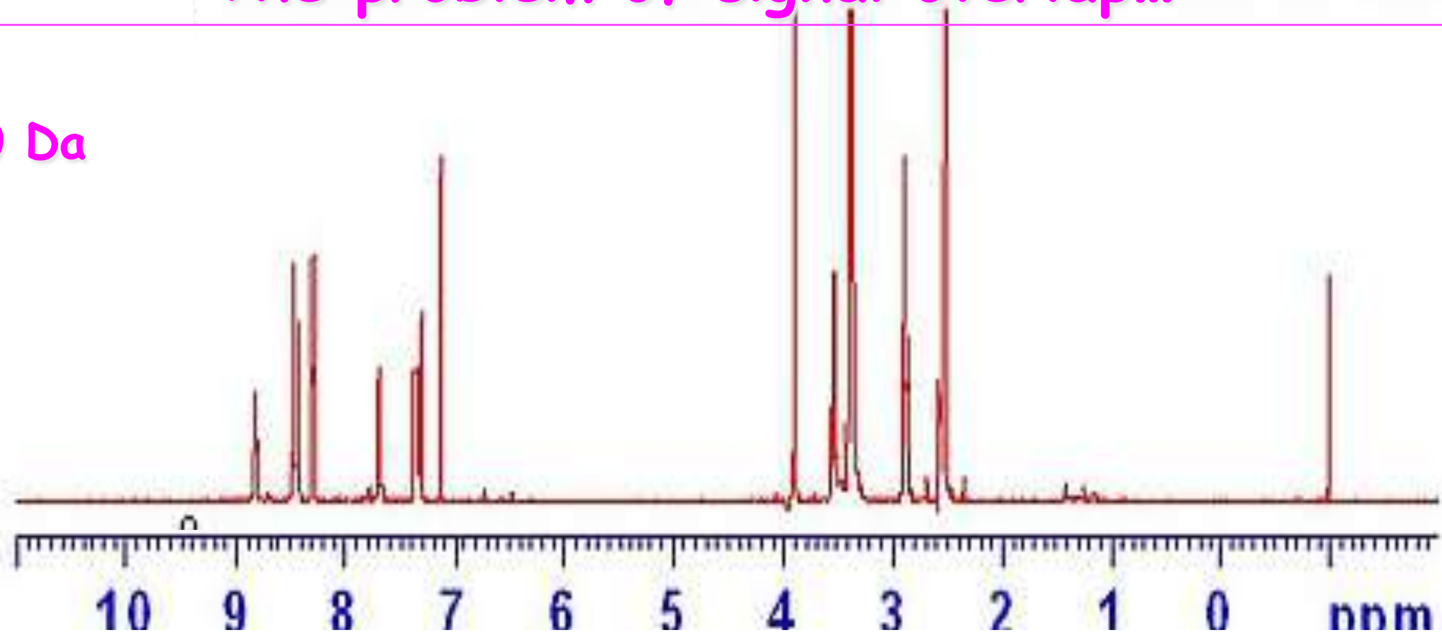


Dynamic properties

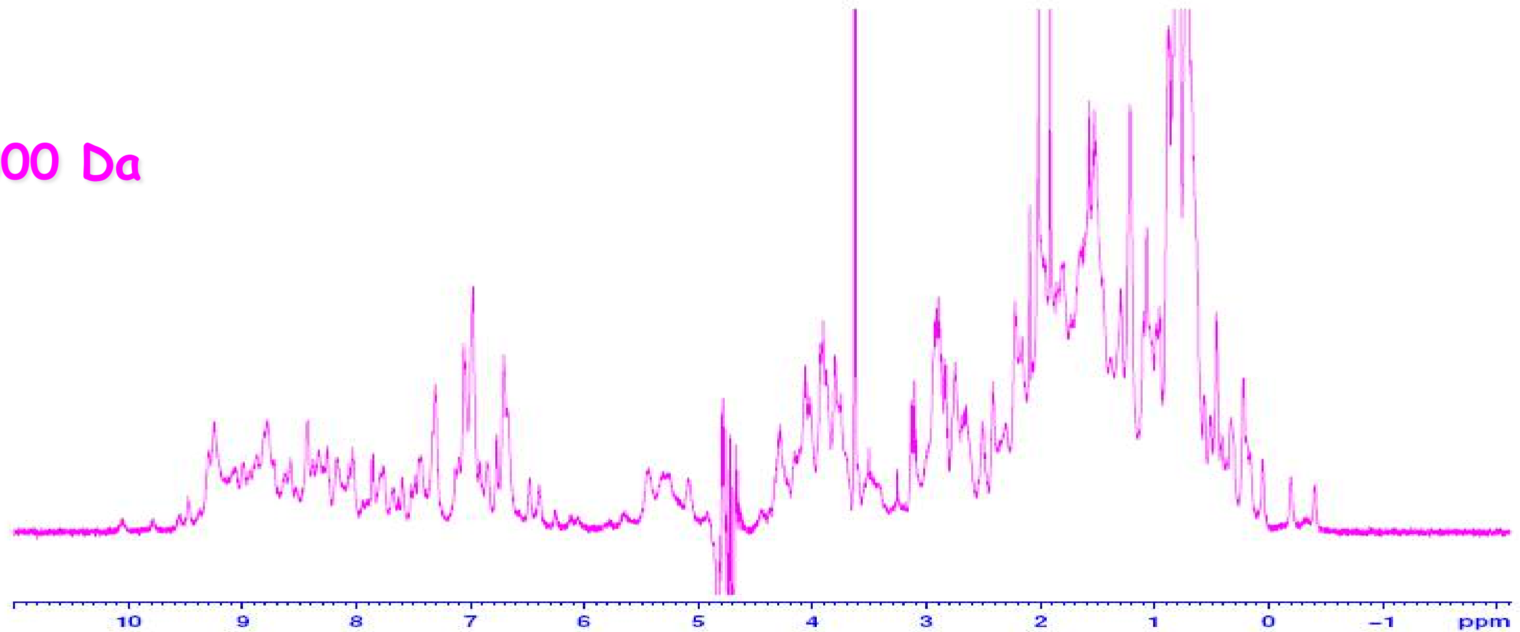
We need to analyze very carefully very complex spectra

The problem of signal overlap…

~300 Da

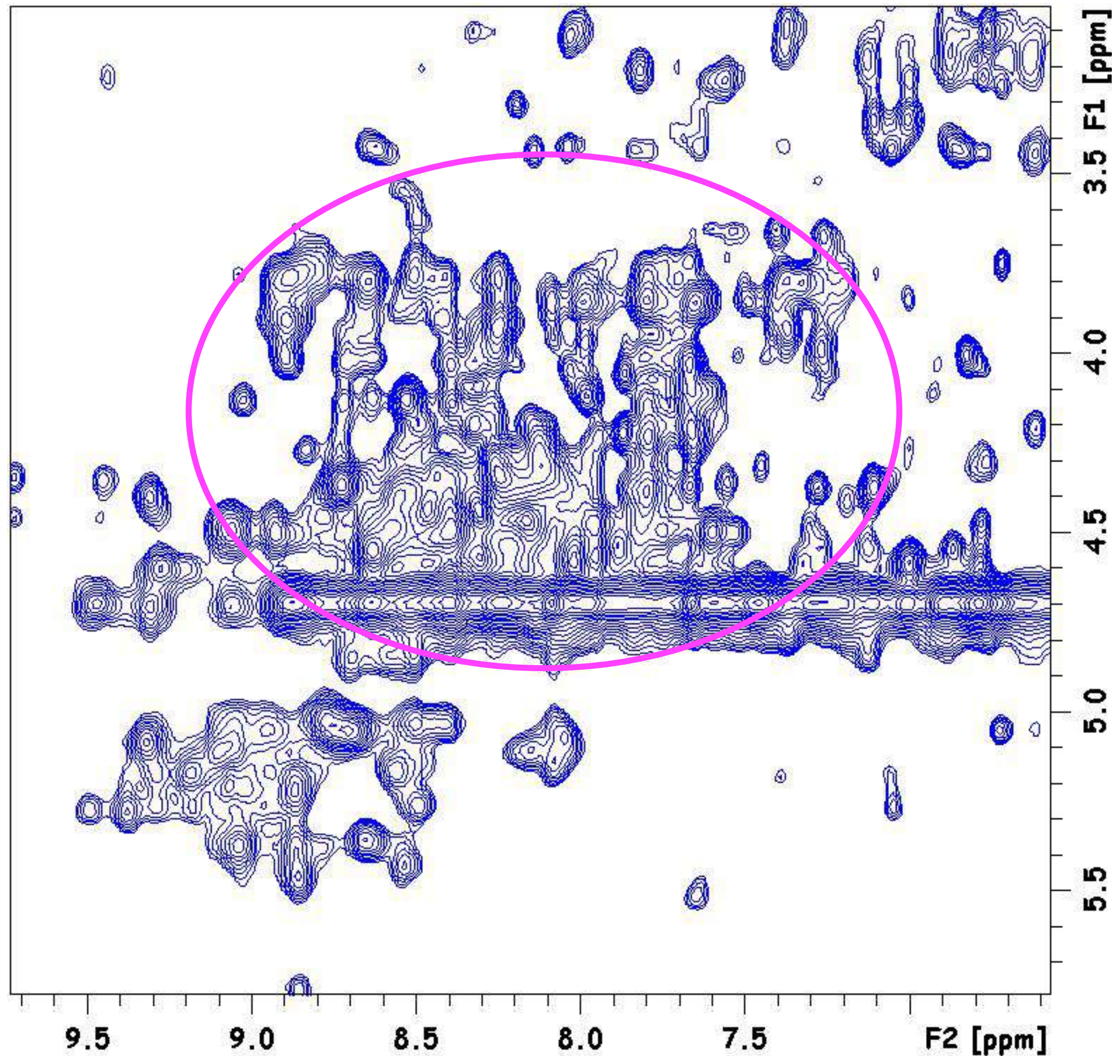


~15000 Da



How can we solve this problem?

Acquiring two dimensional spectra…



Again we can have signal overlap!!!

The limit in size for protein homonuclear NMR is about 100 amino acids…

NMR active nuclei

Isotope	Spin (I)	Natural abundance	Magnetogyric ratio g/10^7 rad $T^{-1}s^{-1}$	NMR frequency MHz (2.3 T magnet)
^{1}H	1/2	99.985 %	26.7519	100.000000
^{2}H	1	0.015	4.1066	15.351
^{13}C	1/2	1.108	6.7283	25.145
^{14}N	1	99.63	1.9338	7.228
^{15}N	1/2	0.37	-2.712	10.136783
^{17}O	5/2	0.037	-3.6279	13.561
^{19}F	1/2	100	25.181	94.094003
^{23}Na	3/2	100	7.08013	26.466
^{31}P	1/2	100	10.841	40.480737
^{113}Cd	1/2	12.26	-5.9550	22.193173

Isotopic enrichment

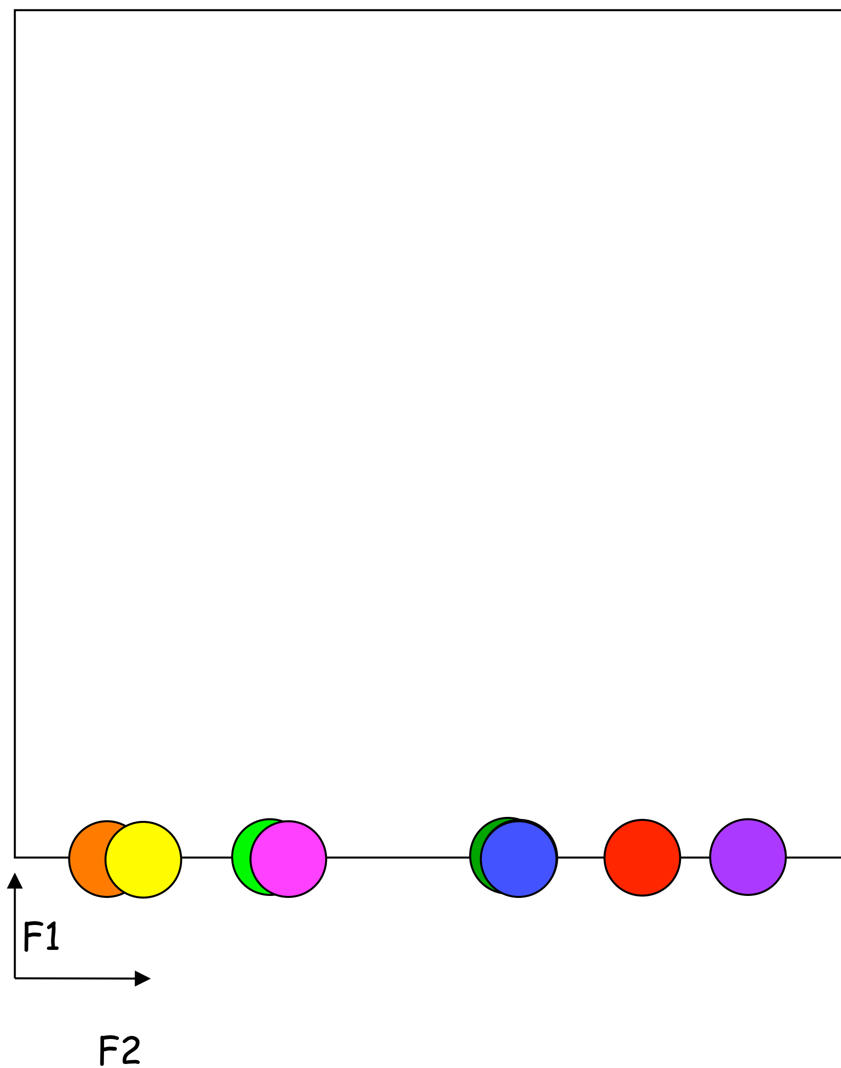
^{1}H (99.98%)
^{13}C (1.108%)
^{15}N (0.37%)

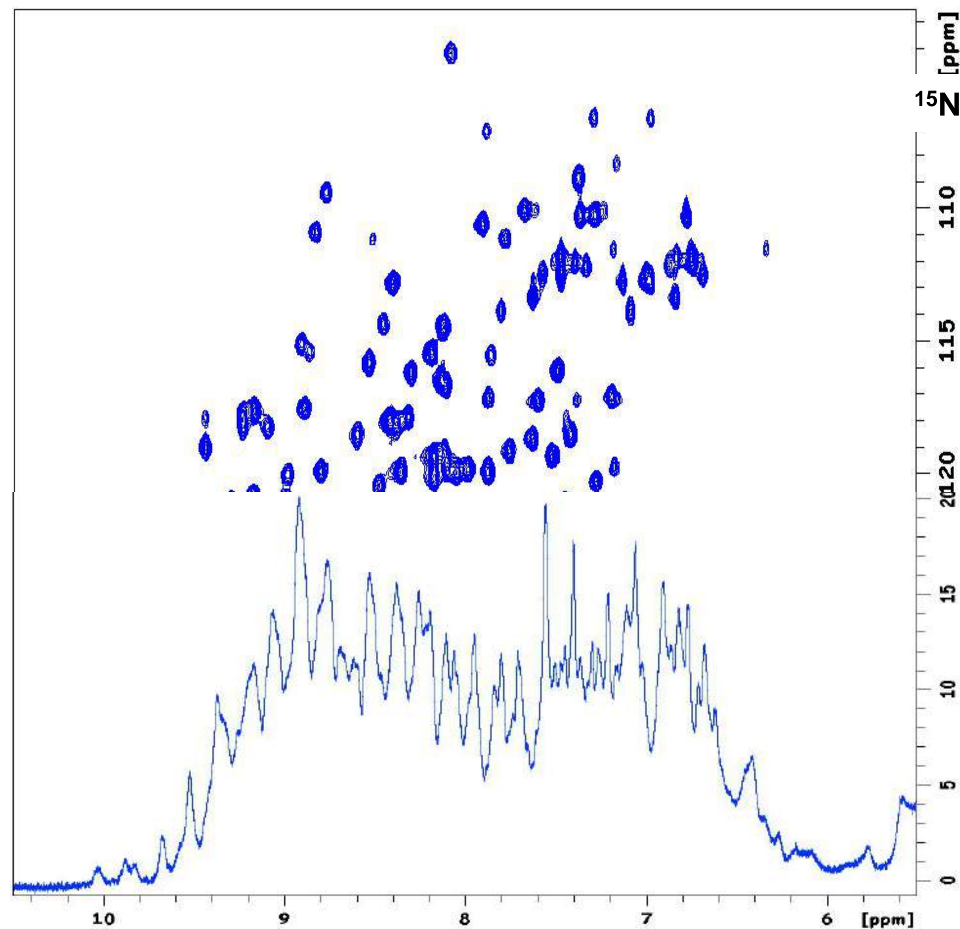
We have additional information in NMR active nuclei in the biomolecules, but there is the limit of the natural abundance.

It is possible to substitute inactive with active nuclei through the isotopic enrichment.

The active nuclei can be used to transfer the magnetization through covalent bonds using heternuclear J coupling

Heteronuclear multidimensional spectra





What it is needed for NMR purposes

Overexpression of the protein to be analyzed (mg of the sample!)

The biological material can be difficult to find

Isotopically enriched samples

Impossible to obtain *in vivo*

The steps to obtain an NMR sample

Cloning

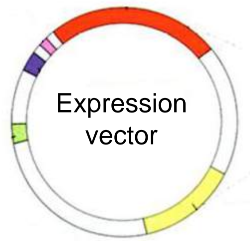
Mutagenesis

Protein expression (unenriched and enriched)

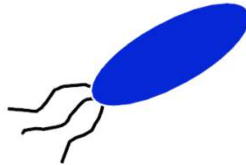
Purification

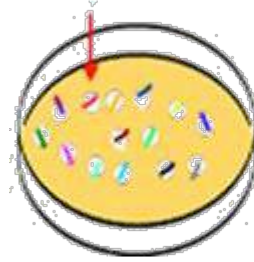
Final steps and NMR sample preparation

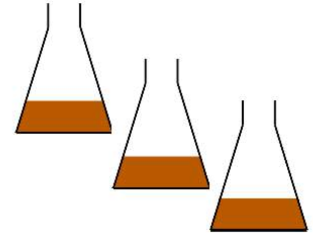
The steps to obtain an NMR sample



+



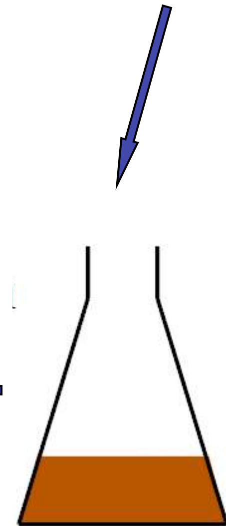




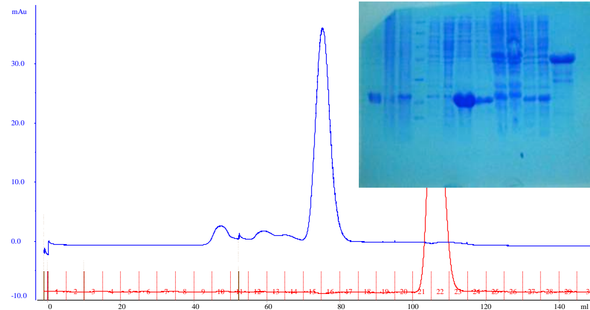
Cloning

Transformation

Expression tests



Scale up expression



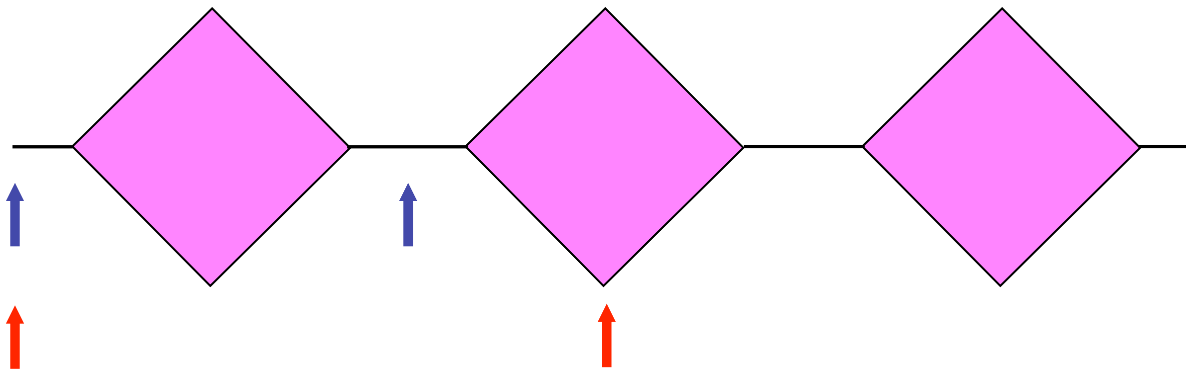
Purification

NMR sample

Cloning

Gene constructs design primers

- Small, globular proteins (full length protein)
- Larger and/or multi-domain proteins
  - -full length protein
  - -choose one or more domains (domains boundaries!)



Well folded

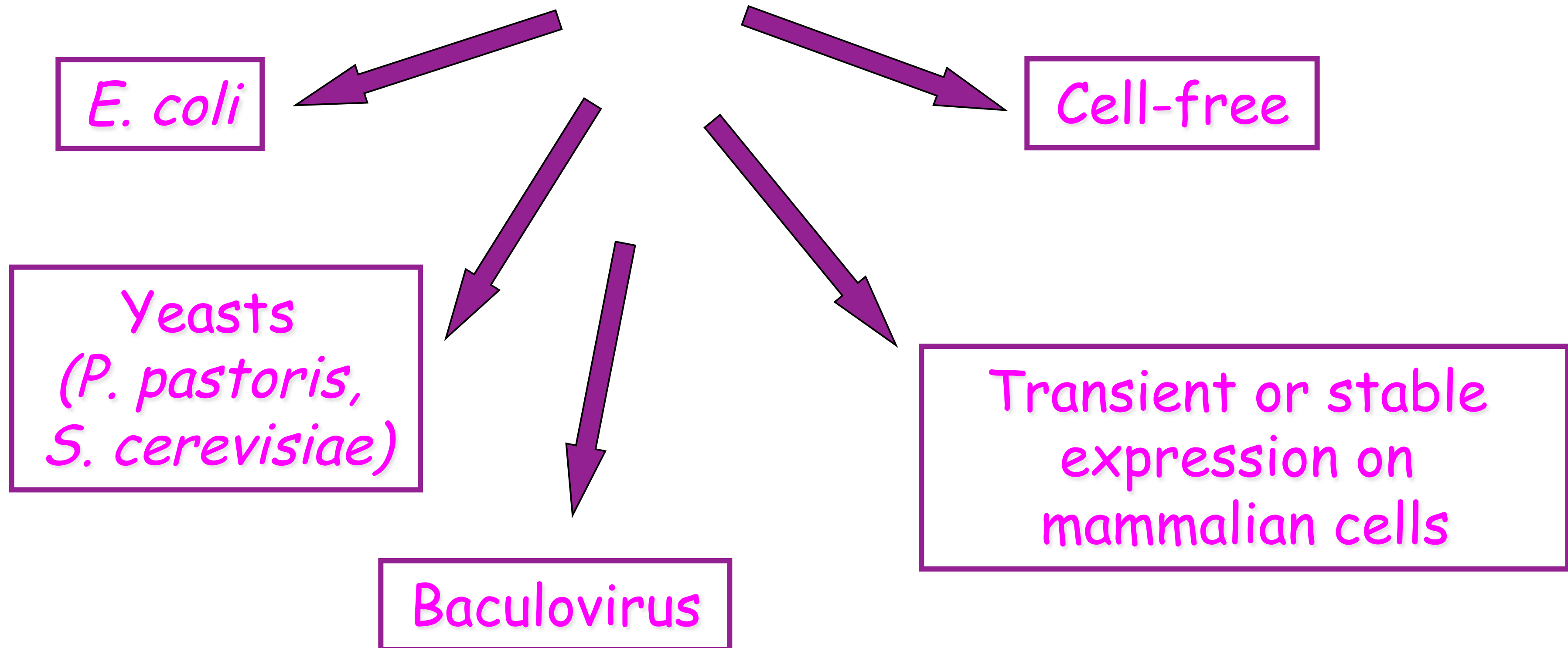
Most likely unfolded

Use structure prediction and multiple sequence alignments to design shorter constructs not to cut secondary structure elements http://bioinf.cs.ucl.ac.uk/psipred/

The steps to obtain an NMR sample

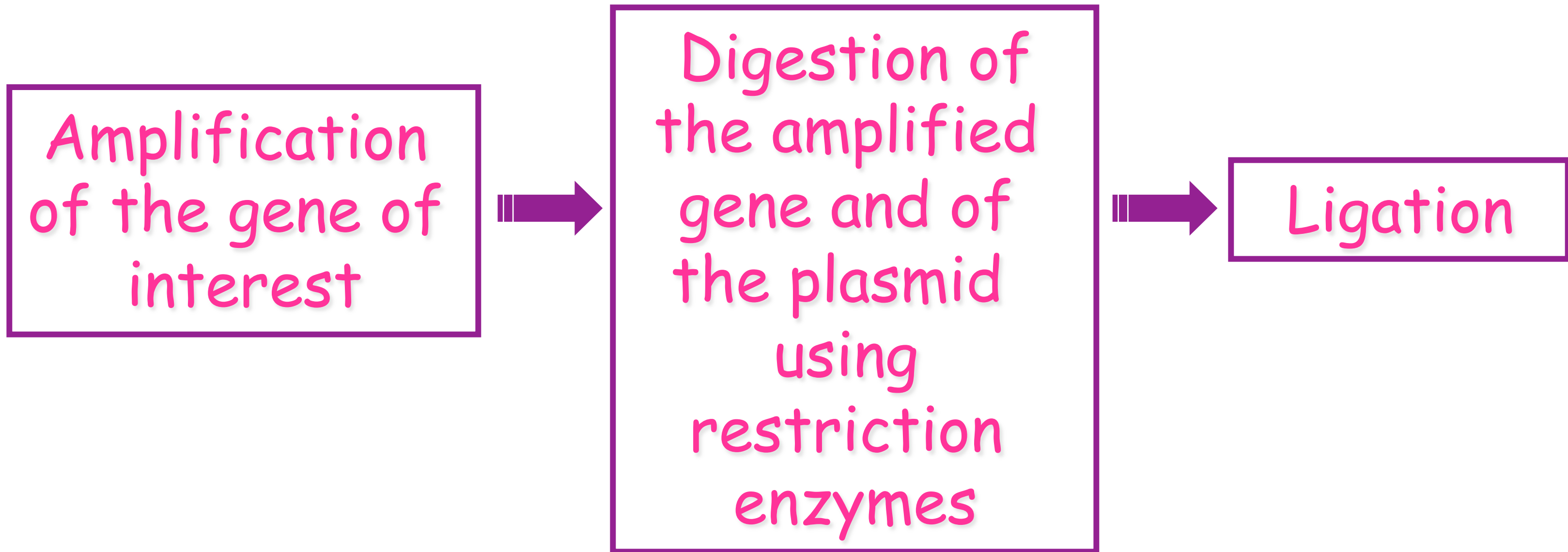
Cloning

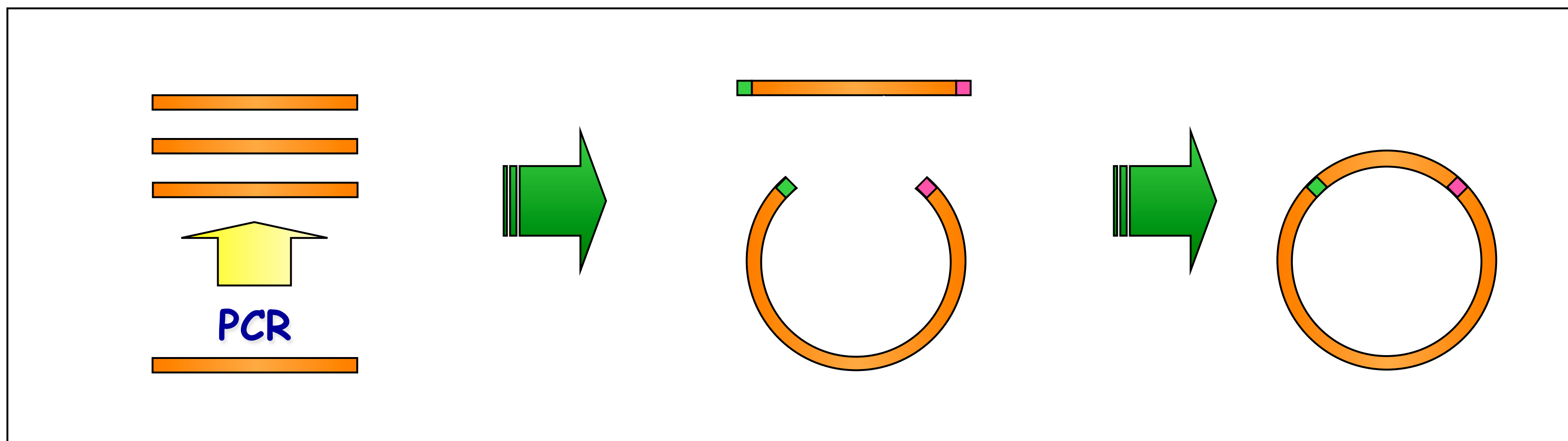
How to choose the expression host



We have to pay attention to the costs of the culture media!

Cloning method: restriction-ligation

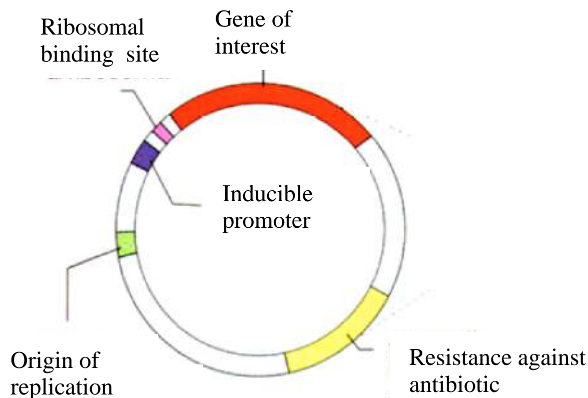




How to choose the plasmid

Depending on how we want to engineer the protein construct

- C or N terminal tag for protein purification (his-tag, GST, etc…)
- Cleavage sites to get rid of the tags (FactorXa, thrombin, TEV)
- choice of the tag (His-tag, fusion proteins, etc…)
- signal peptides for cellular localization



Sequence the plasmid to verify to have successfully cloned the gene of interest.

How to choose the bacterial strain

It is good idea to try different *E. coli* strains to test their ability to survive in the growth media and to express, with a good yield, the protein of interest.

The strains more often used are

BL21DE3, BL21DE3pLysS, Rosetta, Codon plus, Origami, SG…

BL21 strain lacks two proteases

pLysS aviods leaky expression i.e. expression without induction

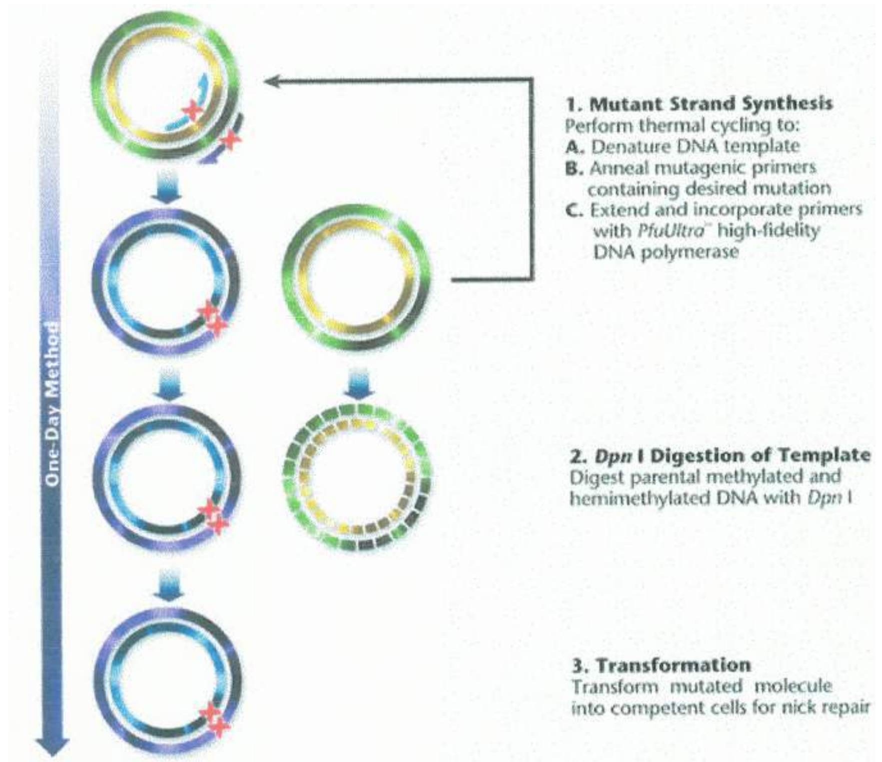
Rosetta and Codon plus strains are designed to enhance the expression of eukaryotic proteins containing rare codons

Origami strain greatly enhances disulfide bond formation in the cytoplasm

The steps to obtain an NMR sample

Mutagenesis

Analysis → Structural
Analysis → Functional



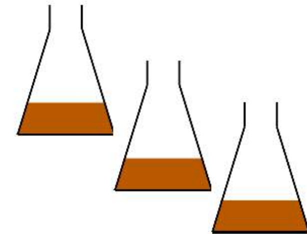
The mutation can be done:

-Substituting the residue with an Ala

-Substituting the residue with one conserved in the family

Expression tests

Expression tests can be performed using small scale cultures (e.g. 10 ml)



Different bacterial strains

Induction conditions

Temperature (20°, 30°, 37°)

Harvesting time (3, 5, 7, 20 hrs)

Amount of IPTG (1mM, 0.5mM, 0.2mM)

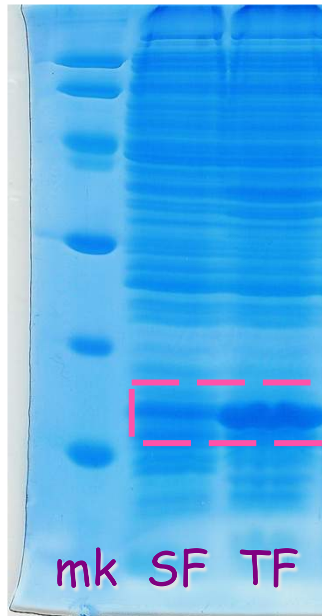
HIGH YIELD OR PROTEIN SOLUBILITY?

Cell lysis Isolation of the soluble/unsoluble fraction (refolding)

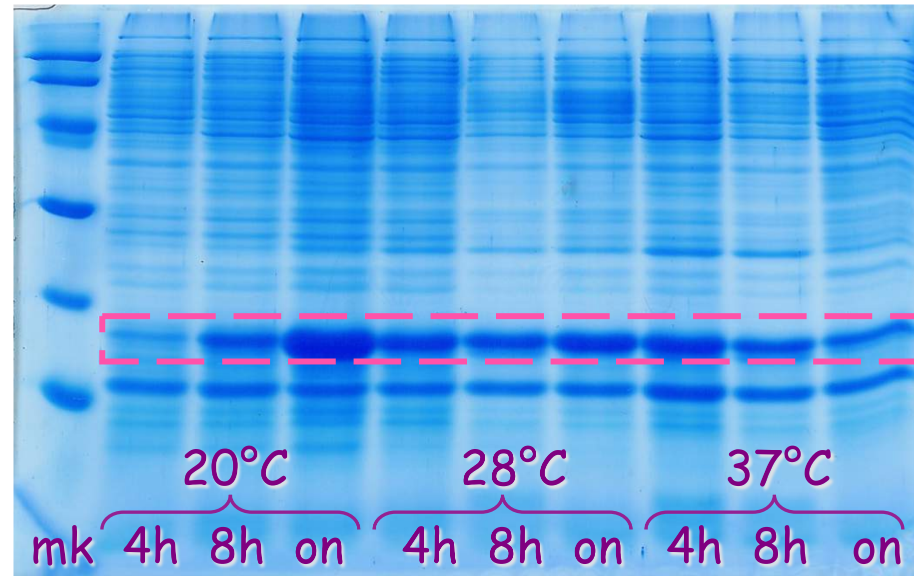
Expression tests

Protein expression can be evaluated on SDS PAGE.

An example



37°C

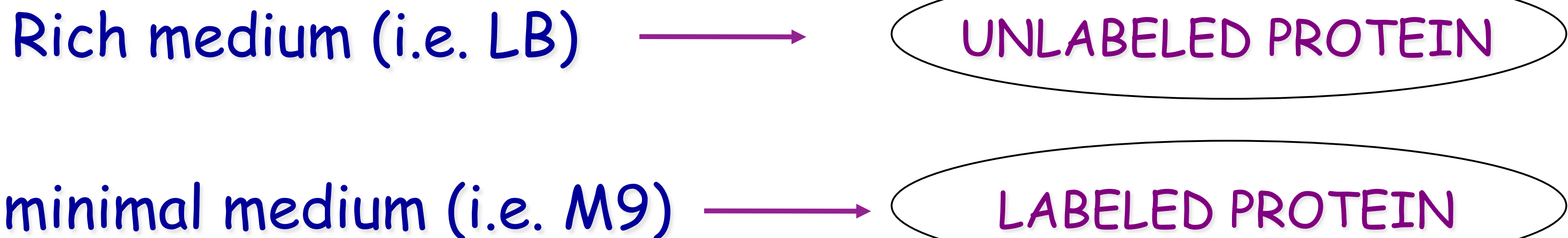


(soluble fractions)

SF soluble fraction
TF total fraction

Recipe 1ml of culture, pellet, add lysis buffer, sonicate 1 min, take the sample of the total fraction, pellet, take the sample of the soluble fraction

Scale up expression: isotopic enrichment procedures



Minimal medium composition

It is a growing broth prepared in the lab starting from simple reagents

It is very important to control the carbon and nitrogen sources

Carbon source: glucose, glycerol, acetate, succinate or methanol

Nitrogen source : NH_4Cl o $(NH_4)_2SO_4$

Salts: NaCl/KCl, $MgSO_4$, $CaCl_2$

Buffer solution: generally phosphate at pH 7.5

^{13}C e ^{15}N have high costs, this implies that each step for sample preparation has to be optimized to obtain as much protein as possible
Tests can be made to reduce the amount of lalebeled salts

Isotopic enrichment procedures

Uniform → the bacteria are grown in a medium containing only sources of the isotope needed. All the atoms of the protein are enriched.

Residue-specific → we enrich only the atoms of the "interesting" residues of the protein

Uniform enrichment

We can prepare minimal media, M9, containing selected labeled nutrients (e.g. $^{15}NH_4Cl$, ^{13}C-glucose), several recipes are available, or buy labeled **ready-to-use medium**.

Bacterial growth is generally higher in ready-to-use media, but minimal media are generally less expensive

Protein expression level has to be checked

To increase the expression level

Classic protocol → the culture is grown and induced directly in the enriched minimal medium

Mixed protocol → the culture is grown in unlabeled rich medium (i.e. LB) and, right before induction, the cells are harvested, washed and resuspended in the enriched minimal medium (Marley J et al. J. Biomol. NMR 2001, 20, 71-75)

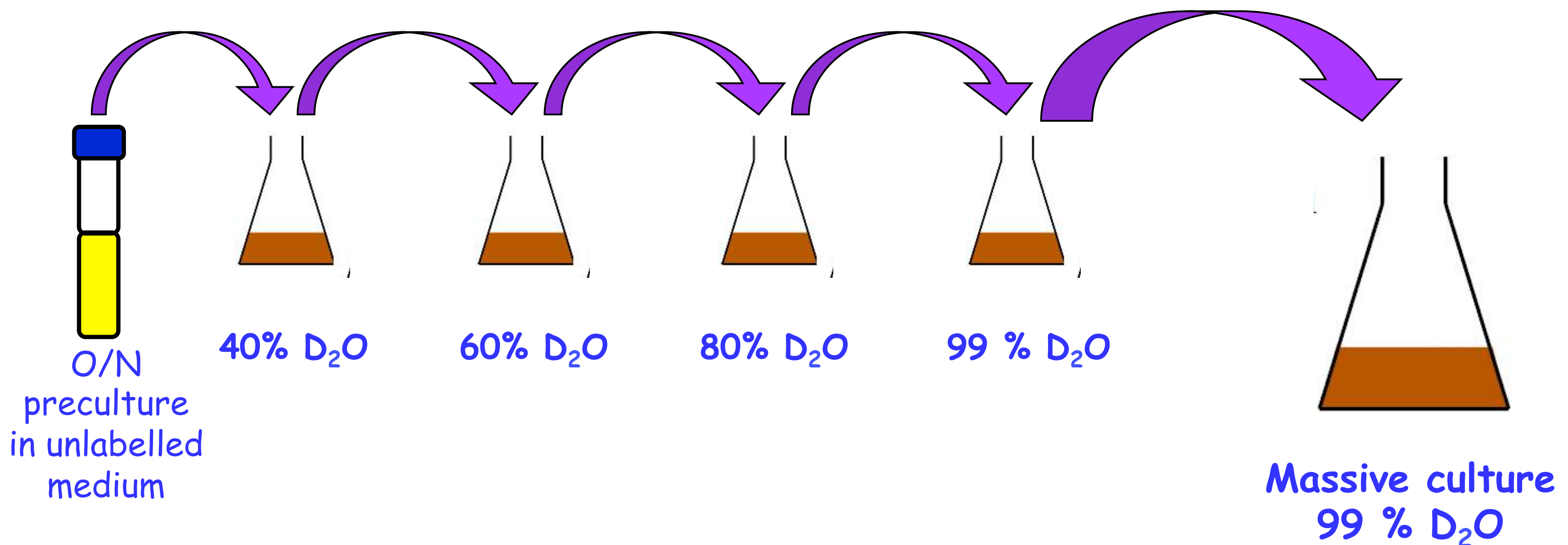
A risk: you could have a mixture of labeled and unlabeled protein!!

Deuterated amino acid side-chains

For large proteins (>250 residues) it is necessary to enrich protein with ^2H on the protons of the side chains, to reduce the relaxation rates of the ^{13}C nuclei

It improves the resolution and sensitivity of NMR experiments

Incorporation of ^2H reduces growth rate of organisms (up to 50%) and decreases protein production as a consequence of the isotopic effect. Bacteria adaptation is required



Isotopic enrichment procedures

Residue-specific enrichment

Auxotrophic strains → we need a specific strain for each residue. Bacteria are grown in the presence of selected labeled amino acids

Prototrophic strains → grow cells with high concentration of aminoacids to inhibit some metabolic pathways

Cell-free → more flexible when using labeled aminoacids

The steps to obtain an NMR sample

Purification

The isolation of the protein is done through several cromatography steps.

The separation is based on one of the following physical properties:

Solubility → i.e. selective precipitation

Charge → i.e. ion exchange

Affinity → i.e. affinity chormatography

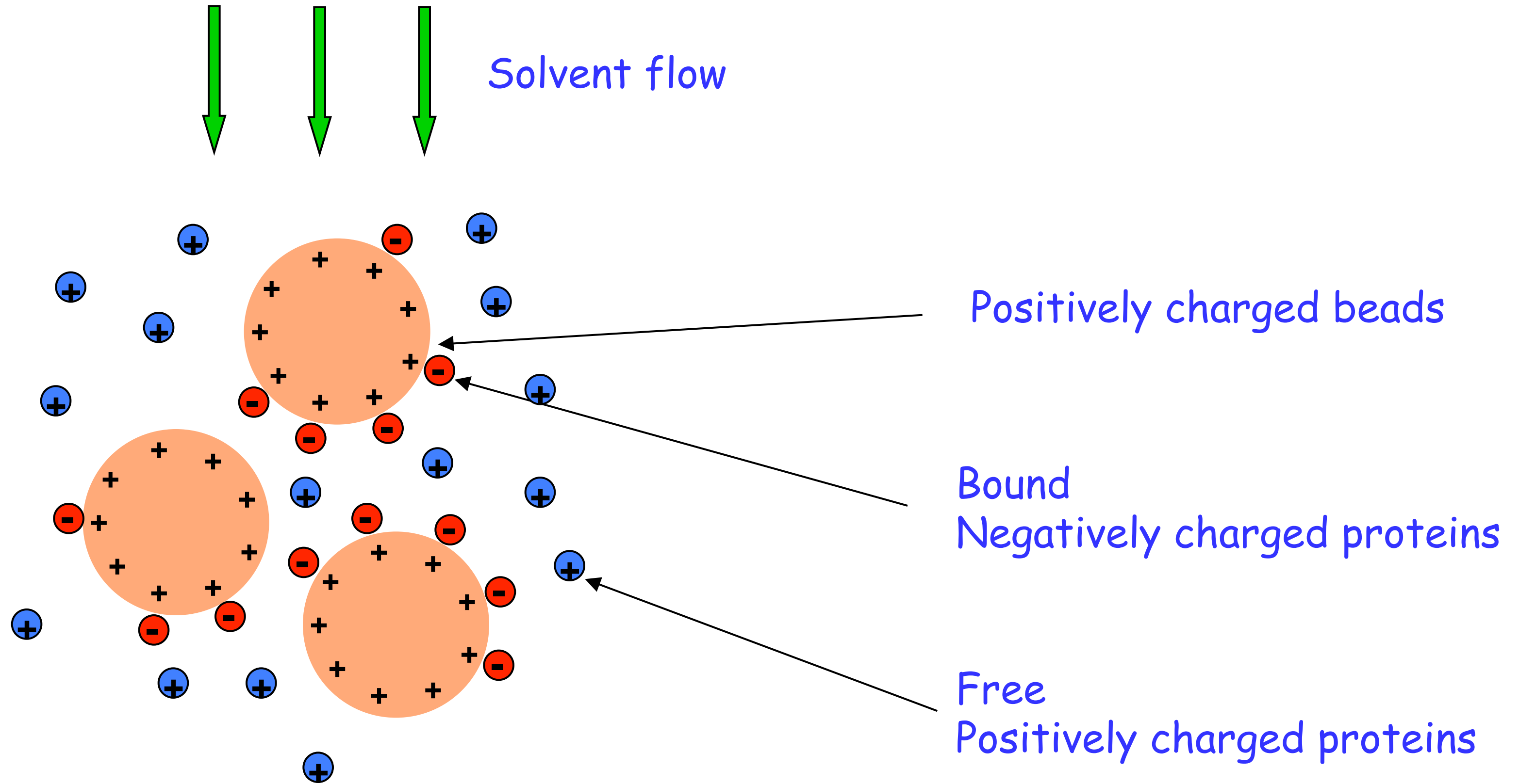
Molecular weight → i.e. gel filtration

The steps to obtain an NMR sample

Affinity chromatography → Tag cleavage using enzymes → Gel-filtration → Delipidation

Ion-exchange → Gel-filtration → Delipidation

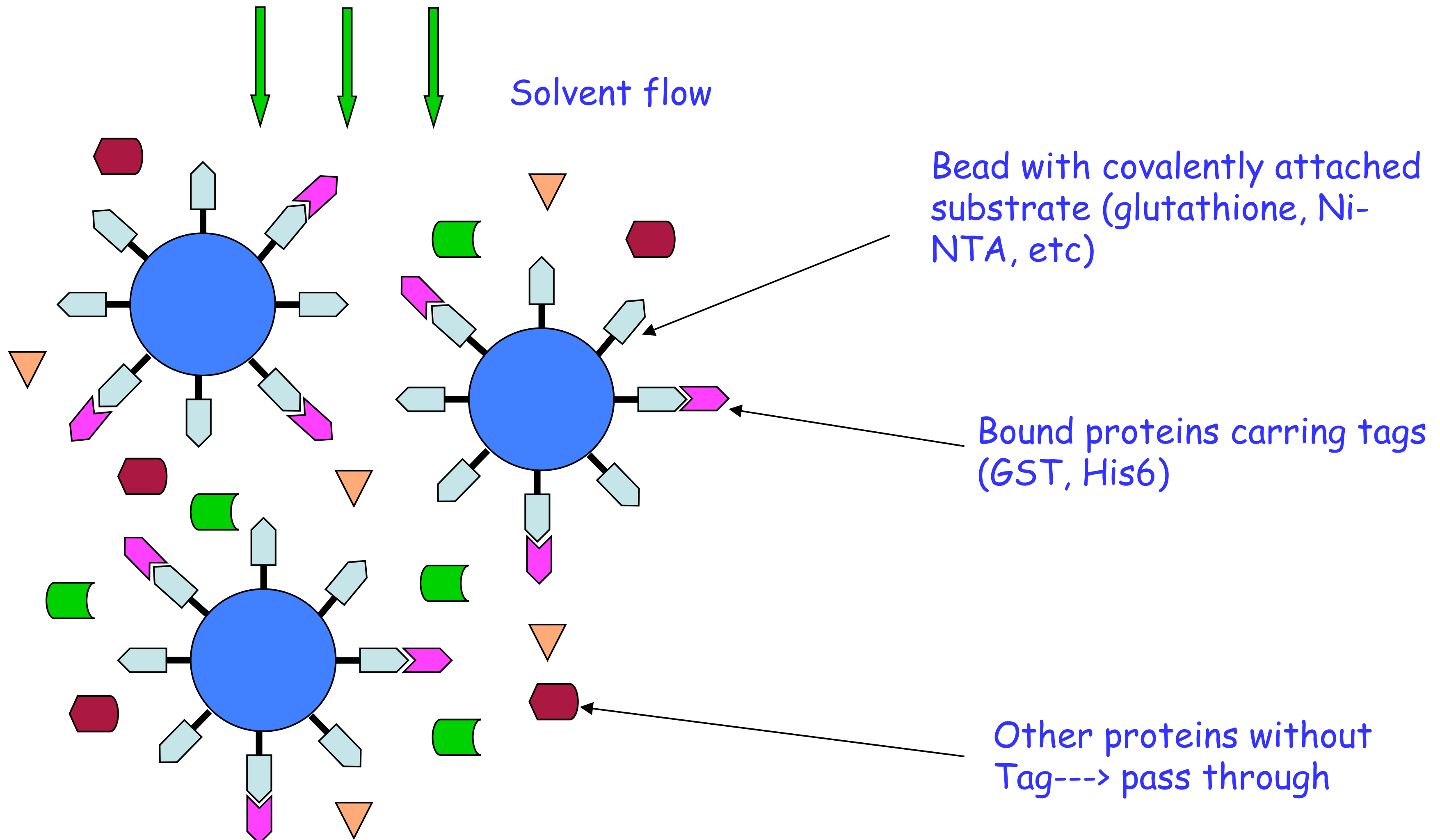
Ion exchange



Usually the elution is done by ionic strength gradients

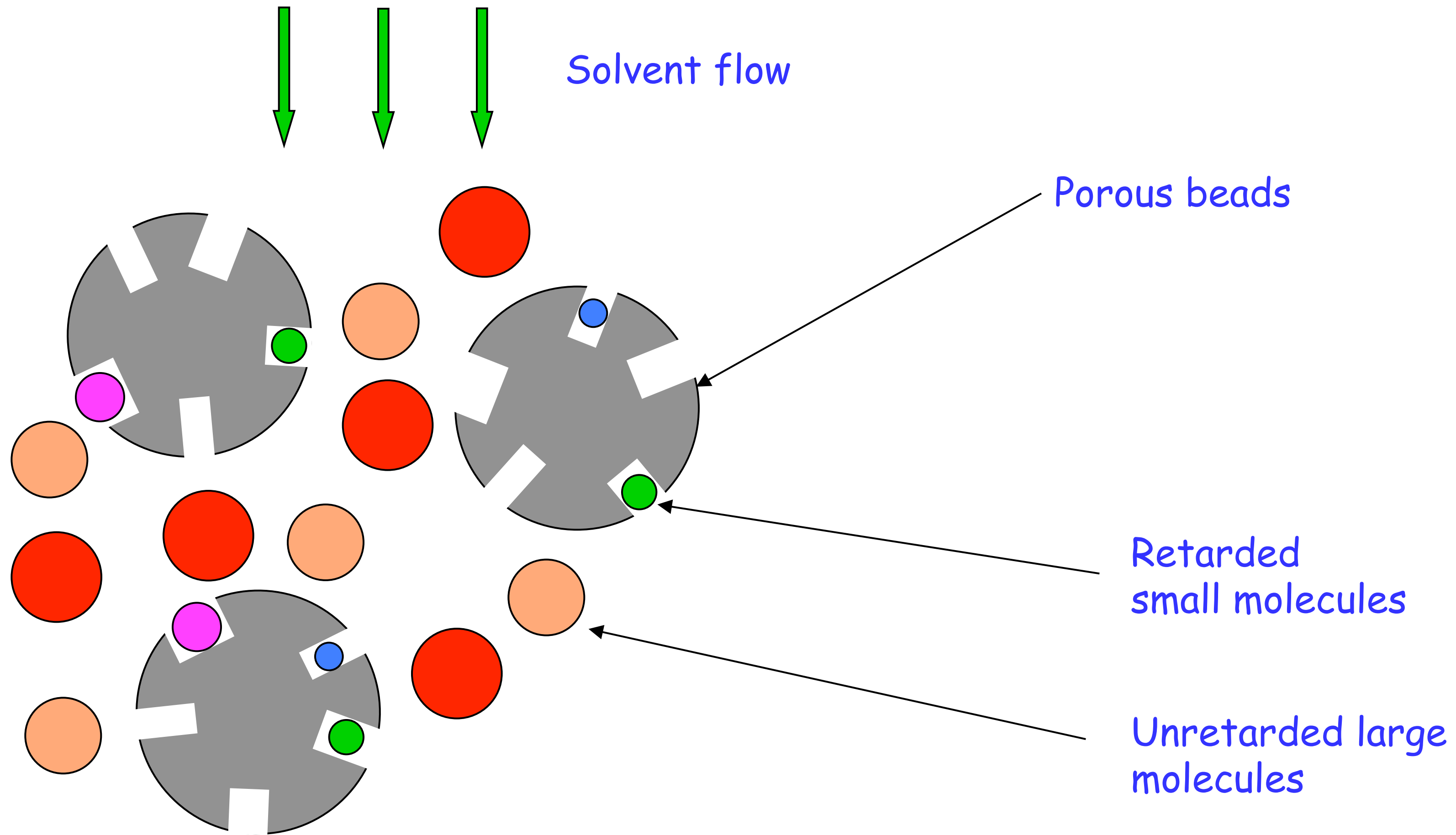
Anion exchange matrices: DEAE-sepharose, Mono-Q, SOURCE-Q, RESOURCE-Q, etc
Cation exchange matrices: SP-sepharose, Mono-S, SOURCE-S, RESOURCE-S, etc

Affinity chromatography



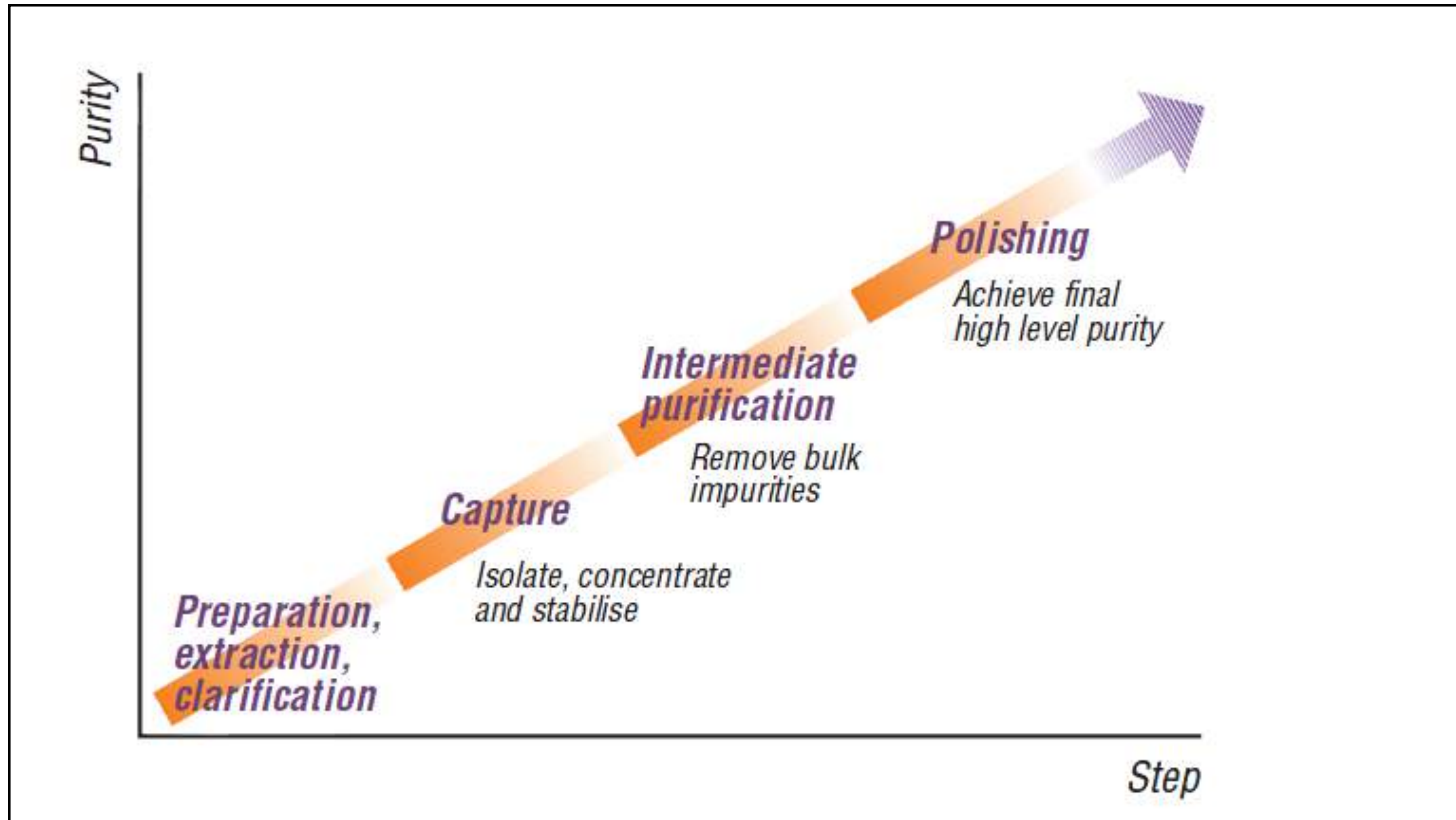
Desorption is usually performed specifically, using a competitive ligand i.e. glutathione for GST, imidazole or EDTA for Ni-NTA, etc...
Tag removal, if present, can be performed after this purification step

Gel filtration



This chromatography is often use to determine molecular weight of the proteins

Three Phase Purification Strategy



The optimum selection and combination of purification techniques is crucial for an efficient purification process and for recovering high amounts of pure protein sample

The steps to obtain an NMR sample

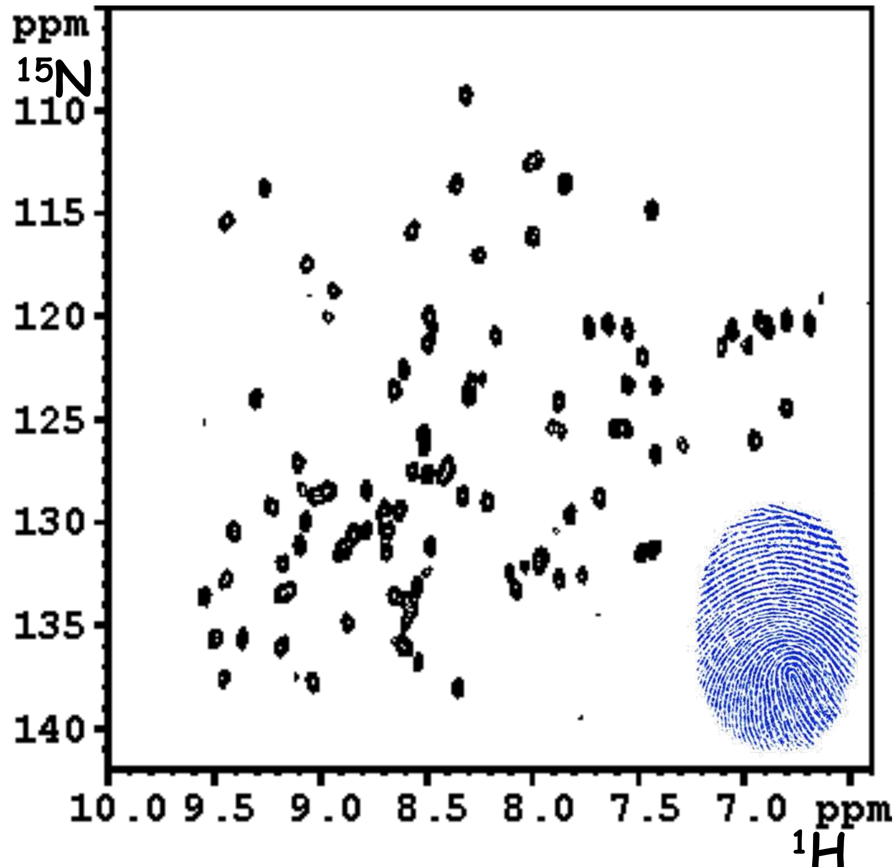
Final steps and NMR sample preparation

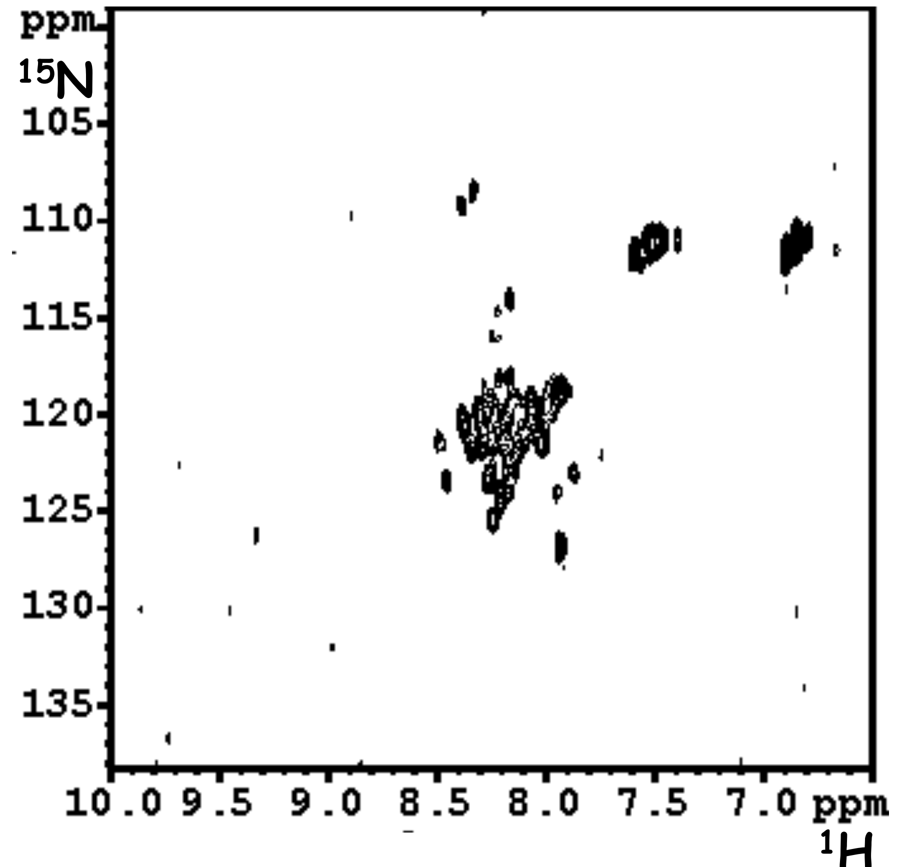
- (Refolding)
- Deglycosylation (not if expressed in *E. coli)*
- Delipidation
- Buffer exchange → Choose buffers suitable for NMR
- Concetrate sample → [P] in the mM range
- Complexes formation
- Storing → i.e. freezing, liophilization,...
- Check protein identity and fold → i.e. mass spectrometry, NMR analysis

Preliminary characterization using NMR

An easy-to-run spectrum lasting 10 minutes is enough to check protein folding!

folded | *not folded*





Even proteins have “fingerprints”!

Once we have a ^{15}N-^{13}C labelled protein…

…we can run a series of three- dimensional (or more!) NMR spectra to assign chemical shifts of all residues of the protein

HNCACB
CBCA(CO)NH
HNCO
HBHA(CO)NH

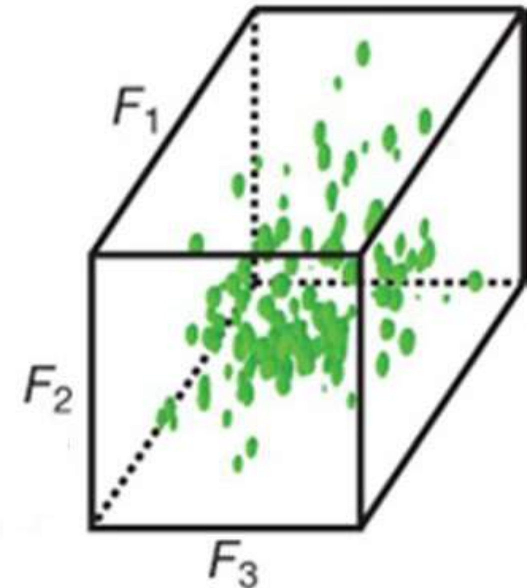
} Backbone atoms assignment

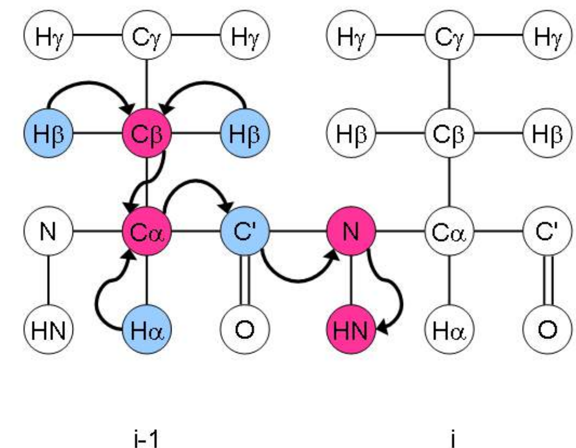
(H)CCH-TOCSY
H(C)CH-TOCSY

} Side chains atoms assignment

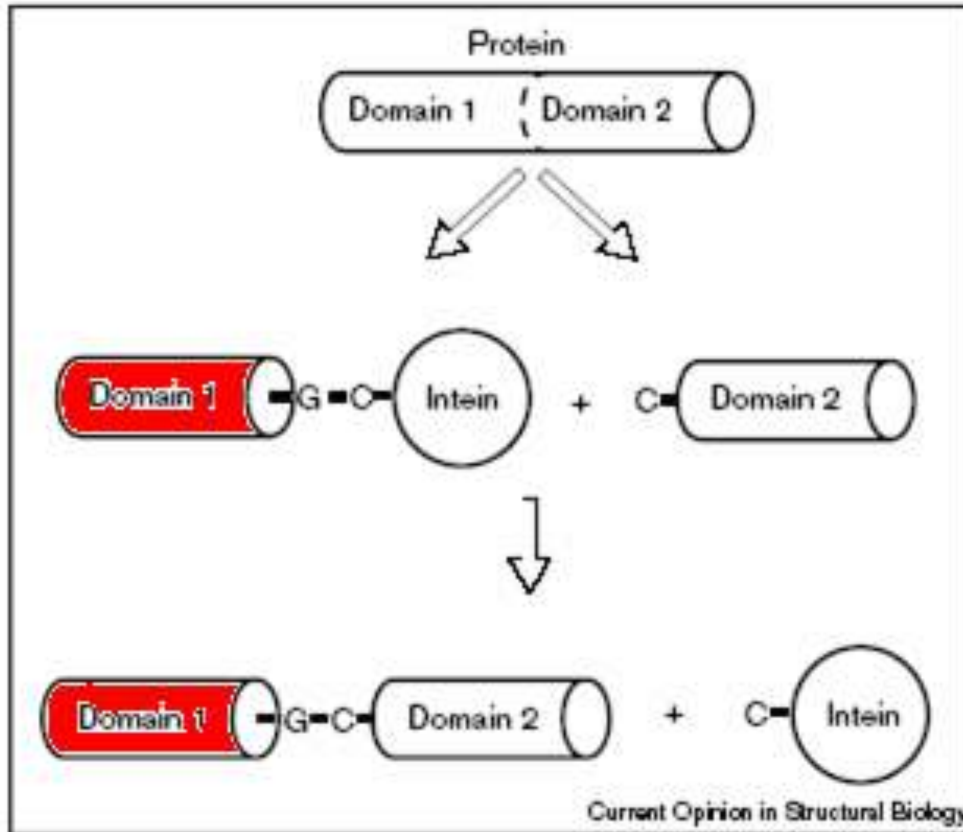
^{15}N-HSQC NOESY
^{13}C-HSQC NOESY

} NOEs assignment for structure calculation





The limit of protein size…some tricks

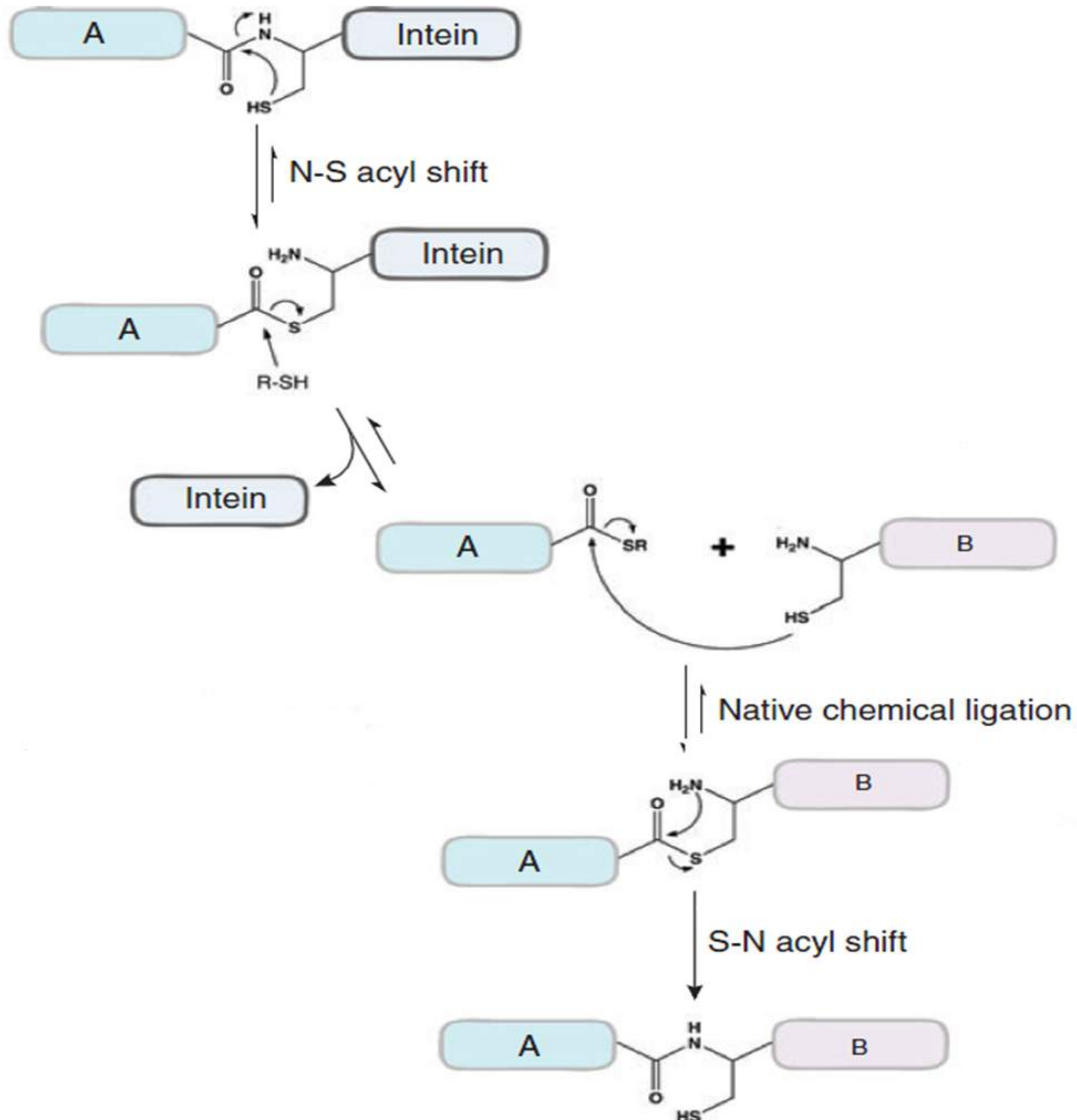


The domain 1 is expressed as fusion protein with intein. This allows a chemical ligation between two domains using mild reagents.

The *in vitro* ligation of the two domains gives us the opportunity to selectively label the domains!

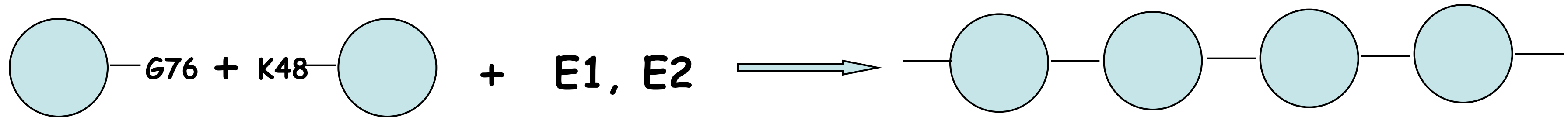
This is one possible solution to overcome size problems

The limit of protein size…some tricks



The limit of protein size…the case of ubiquitin chain

Building up ubiquitin chains and selective labeling



To control the reaction we use a trick!

The two "domains" are equivalent in terms of NMR spectra