Kód předmětu: Bi8980

MASARYKOVA UNIVERZITA

Protein expression and purification

XI. Protein labelling

Lubomír Janda, Blanka Pekárová and Radka Dopitová

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Isotope labelling

Isotopes are different types of atoms of the same chemical element, each having a different number of neutrons:

$^{12}_{6}C$ $^{13}_{6}C$ stable isotopes

$^{14}_{6}C$ radioactive isotope

↓

Stable isotope labelling – a very powerful tool in NMR studies of proteins: spectral overlap reduction → ↑ spectral resolution → facilitates the study of the structure and dynamics of the proteins and protein complexes

NMR spectroscopy – solution of protein is placed in strong magnetic field; then bombarded with radio waves - hydrogen nuclei generate NMR signals (spectrum) indicating distances between atoms

Obtaining isotopically labelled proteins

1. Strategies

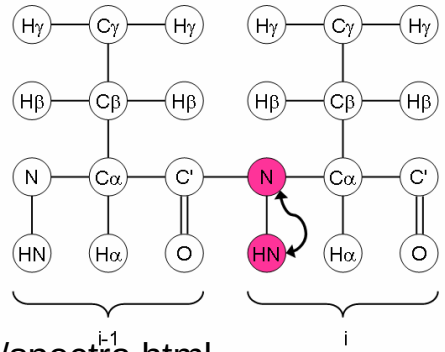
1. uniformly labelled
2. selectively labelled
3. segmental isotope labelling

2. Approaches

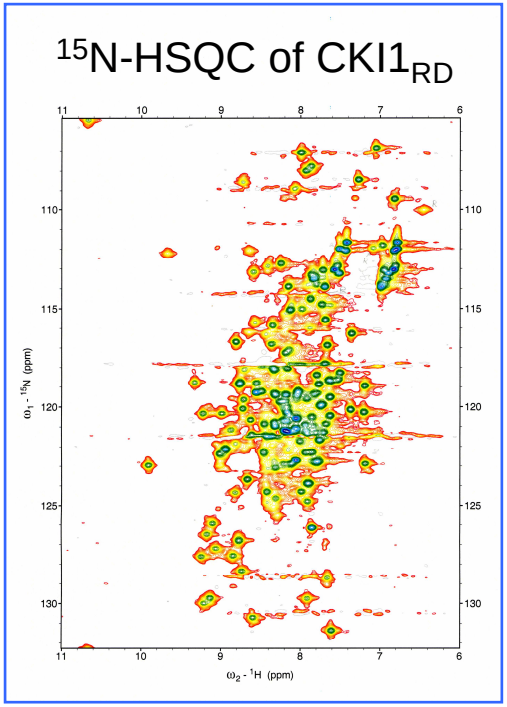
1. *in vivo*: by expressing the corresponding gene in host organisms, which grow on isotope-enriched media
2. *in vitro*: cell-free synthesis system

1.1.2. *^{15}N* uniformly labelled

- The simplest and the cheapest labelling
- Nitrogen source: $^{15}NH_4Cl$ or $(^{15}NH_4)_2SO_4$
- Applications:
  - standard solution-NMR HSQC (heteronuclear single quantum coherence) experiment → spectrum: folded x unfolded protein
  - dynamics experiments
  - titrations with ligands forming complex
  - small proteins (≤ 150 aa) ^{1}H and ^{15}N backbone assignment using ^{15}N-NOESY and ^{15}N-TOCSY



↔



http://www.protein-nmr.org.uk/spectra.html

1.1.3. ^{15}N, ^{13}C (double labelling)

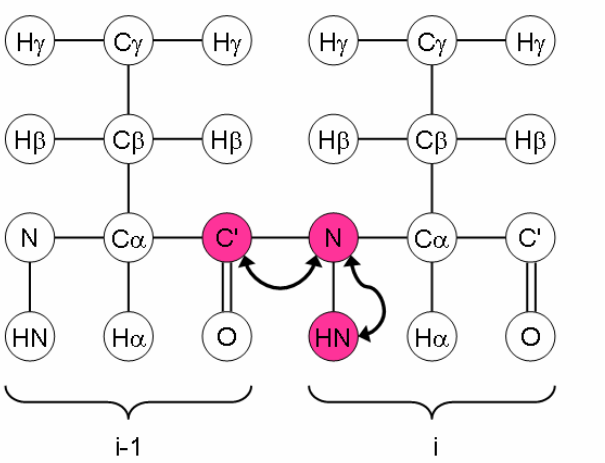
- Carbon source: ^{13}C-glucose
- Applications
  - a very common form of labelling
  - assignment of both the backbone and side-chain ^{1}H, ^{13}C and ^{15}N atoms using triple-resonance spectra (structure determination of proteins ~ 20 kDa)

1.1.4. ^{15}N, ^{13}C, ^{2}H (triple labelling)

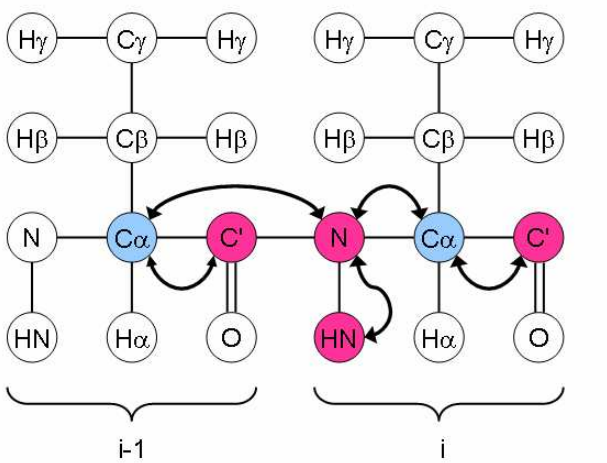
- Medium in D_2O instead of H_2O
- Applications
  - mainly used for structural studies of large proteins (20-80 kDa): by deuterating the protein and thus removing most ^{1}H atoms (protons), the relaxation properties are improved → ↑ sensitivity and spectral resolution

1.1.3. ^{15}N, ^{13}C – triple resonance experiments (3D)

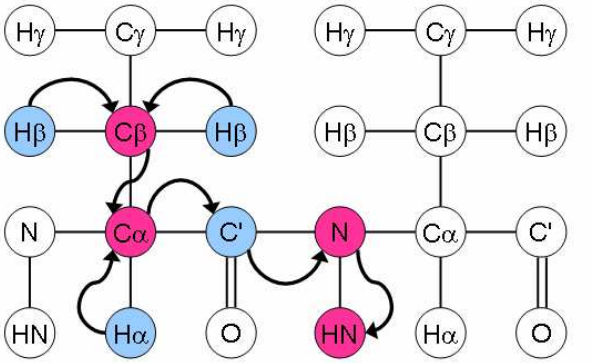
HNCO



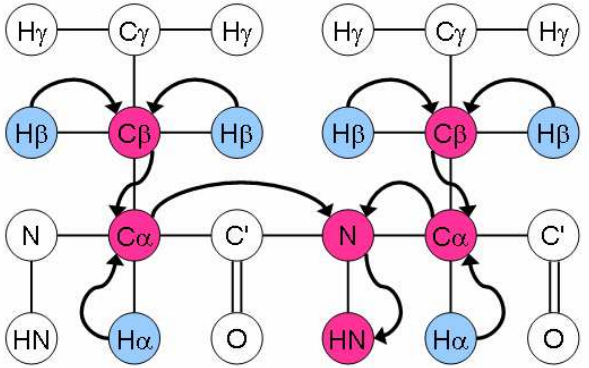
HN(CA)CO



CBCA(CO)NH / HN(CO)CACB

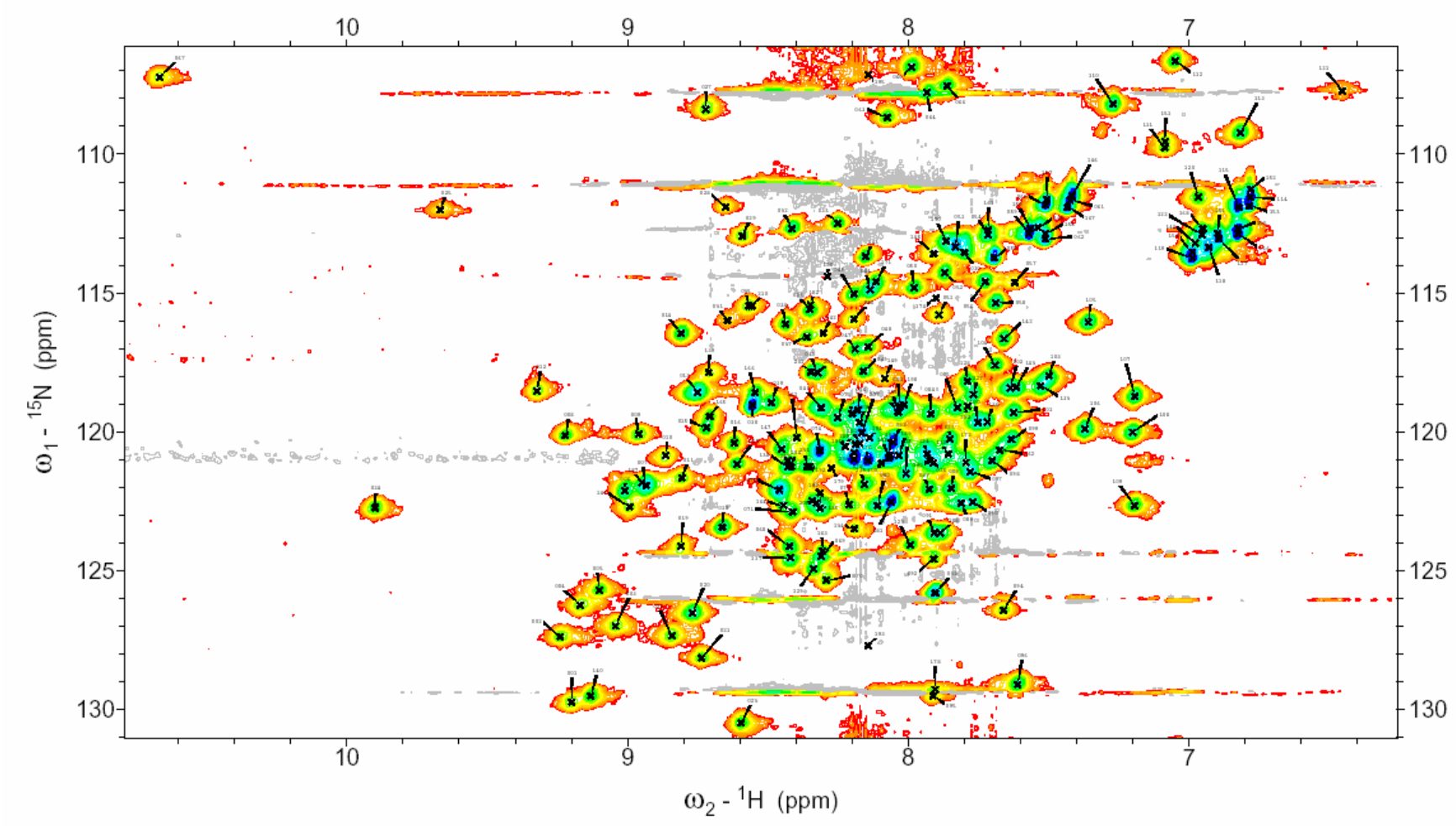


CBCANH / HNCACB



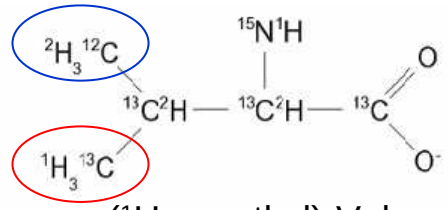
http://www.protein-nmr.org.uk/spectra.html

2D spectrum

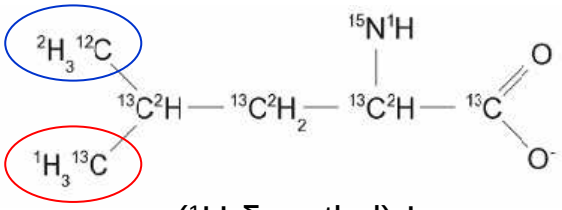


1.2. Selective isotope labelling
1.2.1. Ile, Val and Leu side-chain methyl groups

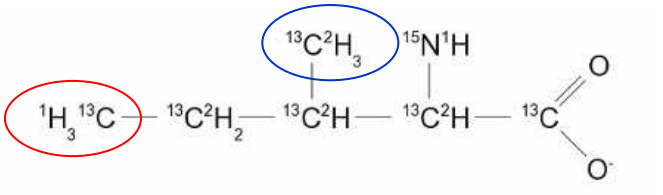
- To improve structure calculations of large proteins
- The IVL labelling produces uniformly ^{2}H,^{13}C,^{15}N-labelled proteins, except for the Val, Leu, and Ile side-chains which are labelled as follows:



(^{1}H-γ methyl)-Val

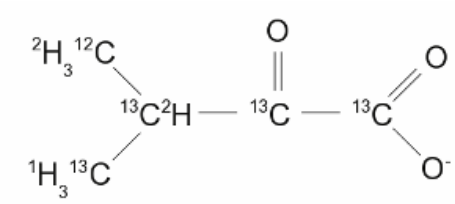


(^{1}H-δ methyl)-Leu

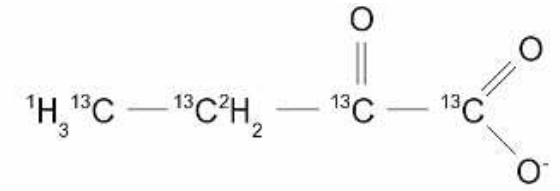


(^{1}H-δ1 methyl)-Ile

- The protein is produced by expression from bacteria which are grown on MM in D_2O using ^{13}C,^{2}H-glucose and $^{15}NH_4Cl$. One hour prior to induction, labelled α-ketobutyrate and α-ketoisovalerate are added.



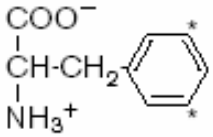
[3-^{2}H] α-ketoisovalerate → Val, Leu



[3,3-$^{2}H_2$] α-ketobutyrate → Ile

1.2. Selective isotope labelling

Chemical structures of metabolites involved in selective isotope labeling strategies.

Labeling agent	Chemical structure	Incorporated as
[3-^{2}H] α-ketoisovalerate	$^{*}CH_3$-$^{*}CD$-$^{*}CO$-$^{*}COO^-$ $^{*}CH_3$	(^{1}H-δ methyl)-Leu (^{1}H-γ methyl)-Val
[3,3-$^{2}H_2$] α-ketobutyrate	$^{*}CH_3$-$^{*}CD_2$-$^{*}CO$-$^{*}COO^-$	(^{1}H-δ1 methyl)-Ile
[ε-^{13}C] L-phenylalanine		[ε-^{13}C] Phe
[2-^{13}C] or [1,3-$^{13}C_2$] glycerol	CH_2-OH $^{*}CH$-OH CH_2-OH or $^{*}CH_2$-OH CH-OH $^{*}CH_2$-OH	^{12}C-^{13}C-^{12}C pattern
^{13}C pyruvate	$^{*}CH_3$-$^{*}CO$-$^{*}COO^-$	(^{1}H-δ methyl)-Leu (^{1}H-γ methyl)-Val (^{1}H-γ2 methyl)-Ile (^{1}H-β methyl)-Ala
[3-^{13}C] pyruvate	$^{*}CH_3$-CO-COO^-	(^{13}C-δ methyl)-Leu (^{13}C-γ methyl)-Val (^{13}C-γ2 methyl)-Ile (^{13}C-β methyl)-Ala

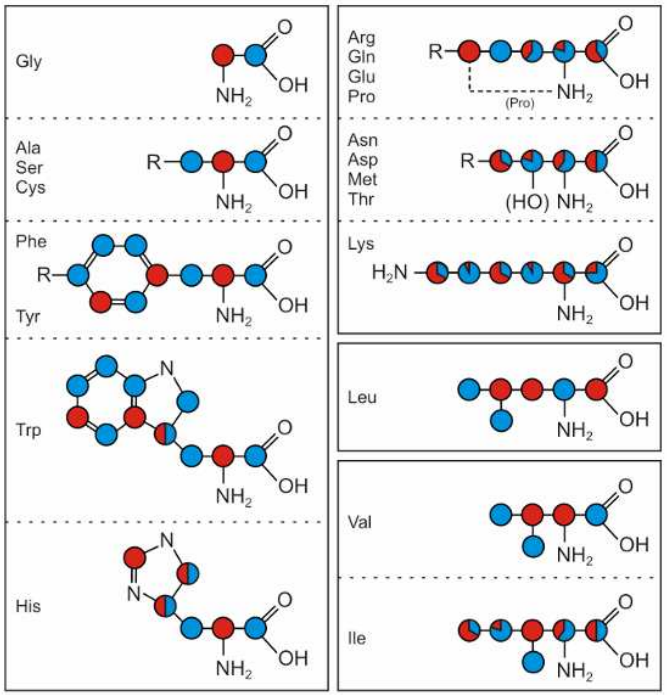
* Indicates positions labelled by ^{13}C

(Goto & Kay, 2000)

1.2.2. *1,3-^{13}C-* and *2-^{13}C-glycerol*

- Source: $^{15}NH_4Cl$ and either 1,3-^{13}C- or 2-^{13}C-labelled glycerol

- Applications
  - to measure relaxation rates → to study internal dynamics of side-chains within proteins (Thioredoxin: LeMaster and Kushlan, 1996)
  - for protein structure determination in solid-state MAS (magic-angle-spinning) NMR (Castellani et al., 2002)



blue: 1,3-^{13}C glycerol labelling
red: 2-^{13}C glycerol labelling

http://www.protein-nmr.org.uk/labelling.html#aaspecific

1.2.3. Selective isotope labelling – other strategies

- Expression from bacteria which are grown on MM supplemented with small amounts of $^{15}NH_4Cl$ and ^{13}C-glucose as well as labelled and unlabelled amino acids.

- **Reverse labelling –** The protein is produced by expression from bacteria which are grown on MM supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose as well as unlabelled amino acids.

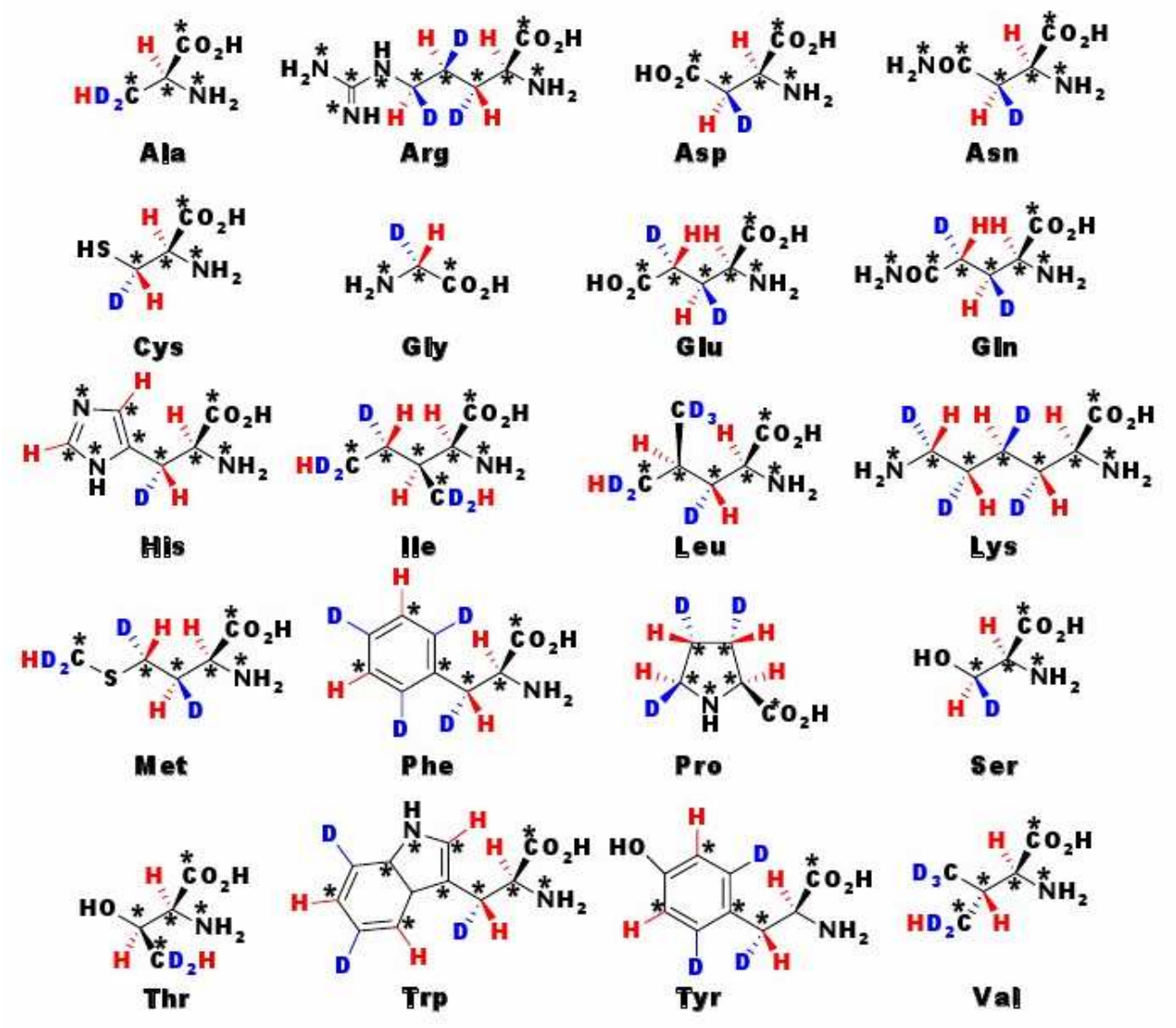
  This suppresses the labelling of these amino acids and only those which have not been added unlabelled will be synthesised by the bacteria using the ^{13}C-glucose as the carbon source.

1.2.4. SAIL – stereo-array isotope labelling (Kainosho et al., *Nature* 2006)

- SAIL-labelled proteins are prepared using cell-free technology.
- The amino acids used for the protein production are prepared using chemical and enzymatic syntheses.
- Their labelling is guided by the following principles:
  - in each methylene group, one of the ^{1}H atoms is stereo-selectively replaced by a ^{2}H atom;
  - in each methyl group, two of the ^{1}H atoms are replaced by ^{2}H atoms;
  - the prochiral methyl groups of Leu and Val are stereo-selectively $^{12}C^{2}H_3$ and $^{13}C^{1}H^{2}H_2$ labelled;
  - six-membered aromatic rings have alternating $^{12}C^{2}H$ and $^{13}C^{1}H$ moieties.

- For the structure calculation of large proteins
- For the investigation of side-chain motions
- Very expensive

20 SAIL amino acids

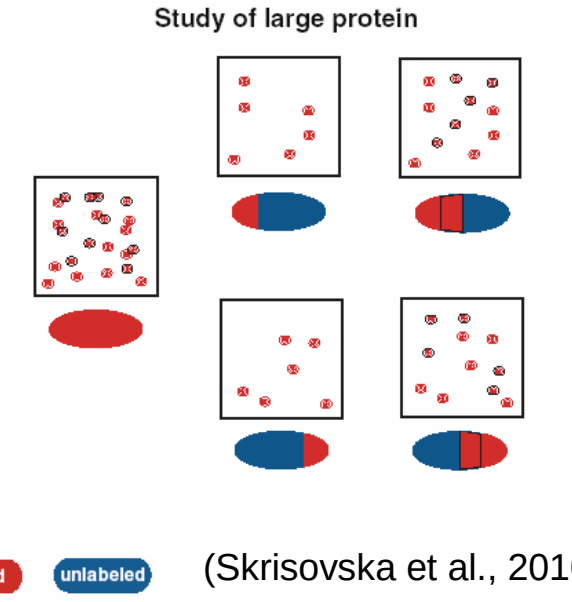


http://www.sailnmr.org/wiki/index.php/File:SailAminoAcids.jpg

1.3. Segmental isotope labelling

- Applications
  - To investigate interdomain interactions within multidomain proteins
  - To study conformational changes after ligand binding
  - To help resonance assignment of large proteins
  - To facilitate protein structure determination

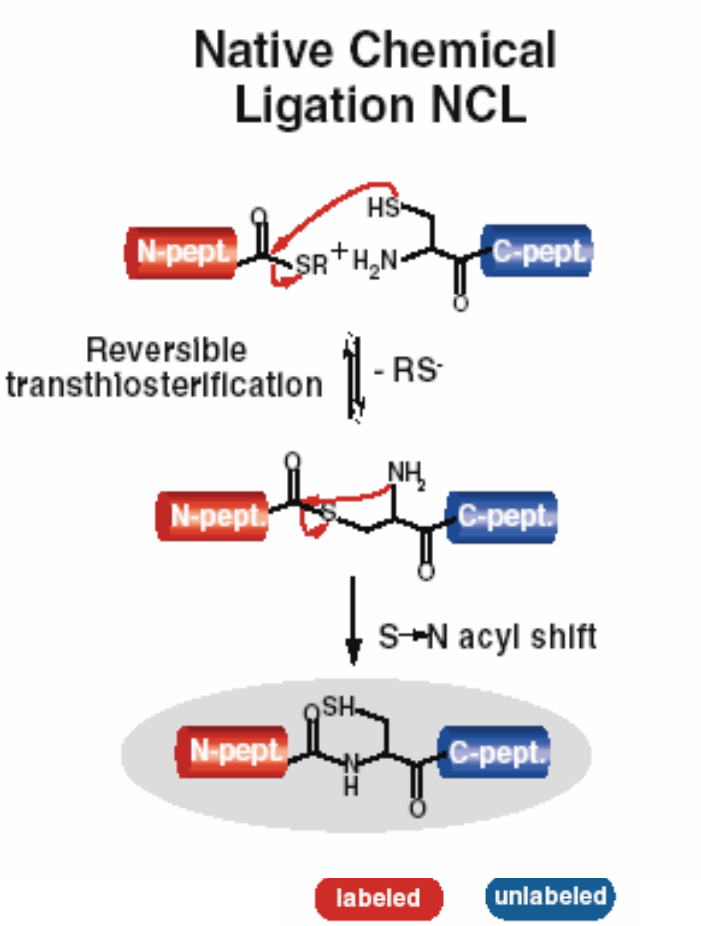
- Approaches
  1. Native chemical ligation
  2. Expressed protein ligation
  3. Protein trans-splicing



(Skrisovska et al., 2010)

1.3.1. Native chemical ligation

- Based on an interaction of two synthetic peptides (~ 50 aa), one containing C-terminal α-thioester and the other N-terminal α-cysteine → formation of native peptide bond
- one peptide can be labelled and one unlabelled



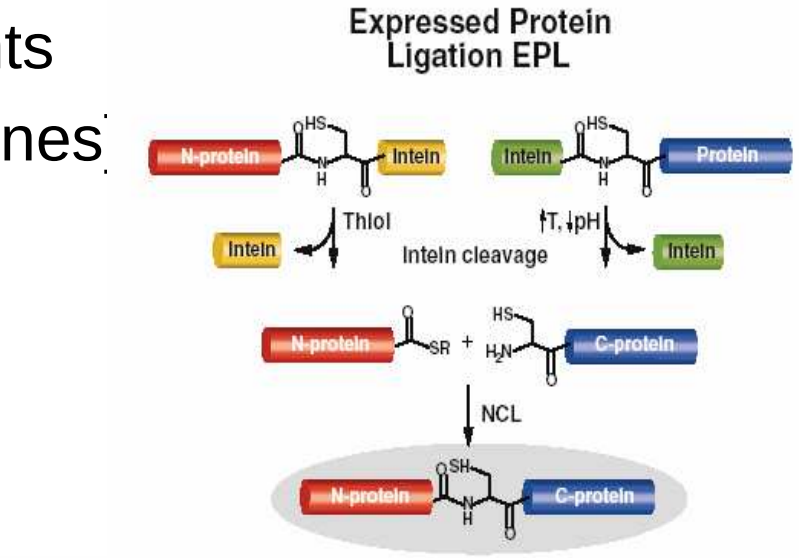
(Skrisovska et al., 2010)

1.3.2. Expressed protein ligation

- = intein-mediated protein ligation
- Based on native chemical ligation and inteins properties
- Involves recombinant expression of one or both protein fragments [IMPACT™ system (NEB): set of bacterial vectors allowing recombinant

  production of protein fragments

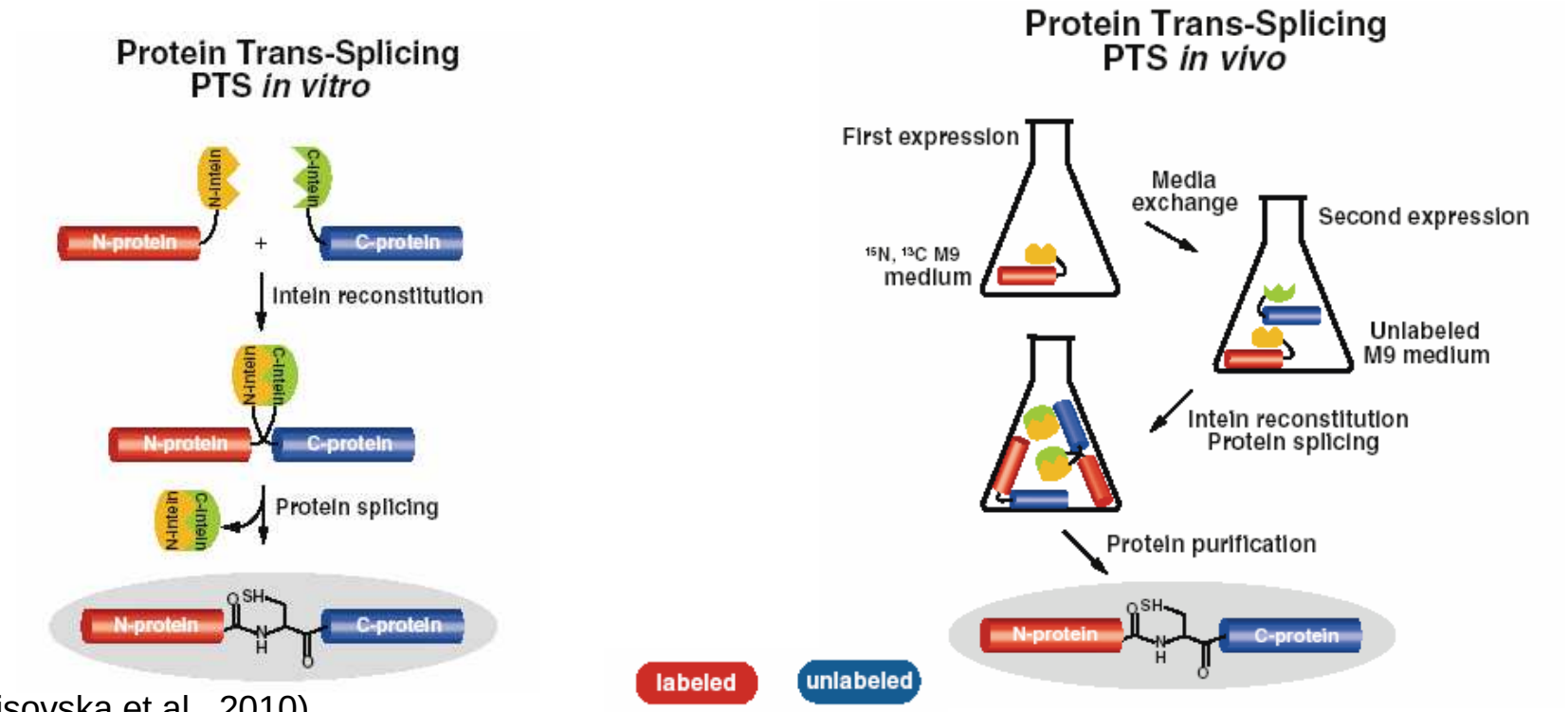
  with α-thioesters and α-cysteines



(Skrisovska et al., 2010)

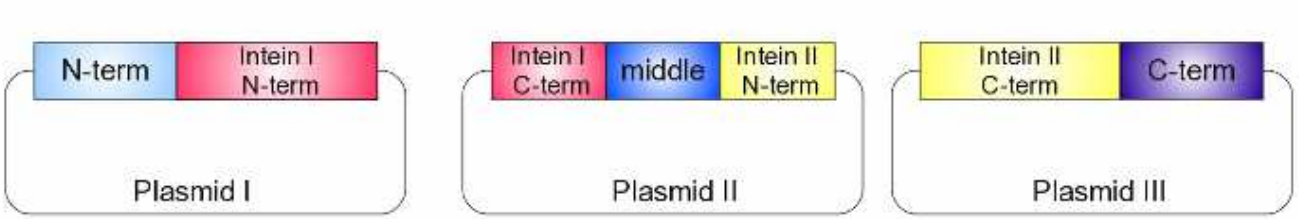
1.3.3. Protein trans-splicing

- Based on protein trans-splicing process = Inteins are fragmented in two inactive parts. After their association, they reconstitute into active intein which performs a splicing reaction resulting in ligation of their fusion protein fragments.

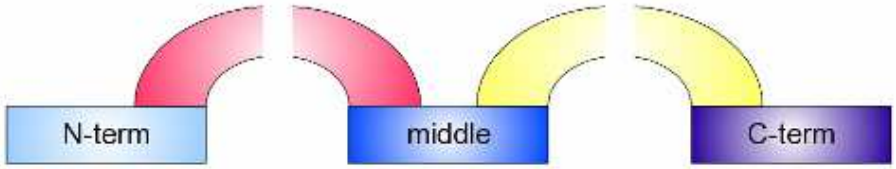


(Skrisovska et al., 2010)

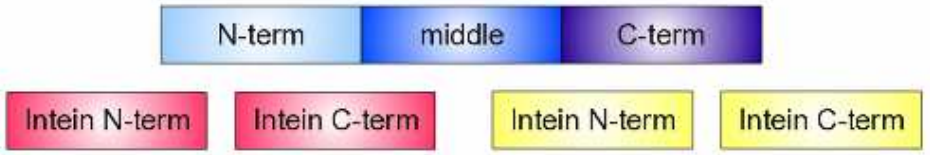
1.3.3. Protein trans-splicing in multidomain protein



1. Express separately using different isotopic labelling



2. Split inteins fold into functional inteins



3. N-, middle and C-terminal segments of target protein are joined together covalently and split from the intein fragments

http://www.protein-nmr.org.uk/labelling.html

2. Protein expression systems

2.1. in vivo

- *Escherichia coli*
- Methylotrophic yeast *Pichia pastoris*
- Baculovirus expression system (BVES)
- Mammalian cells
- Slime mold
- Hybridoma cells
- Plants

2.2. in vitro

- Cell-free expression system

2.1.1. *Escherichia coli*

- Advantages
  - the most economical system (easy genetic manipulation, easy culturing conditions, rapid population growth, high-level protein production)

- Disadvantages
  - unwanted metabolic conversion to other amino acids (scrambling, e.g. ^{14}N Val → ^{15}N Ala, ^{15}N Leu → ^{15}N Glu)
  - lack of intracellular organelles (Golgi system, endoplasmic reticulum)
  - limited number of molecular chaperones
  - absence of post-translational modification

↓

Some eukaryotic proteins (containing disulfide bonds) cannot be folded correctly and are expressed insolubly as inclusion bodies.

2.1.1. *Escherichia coli*

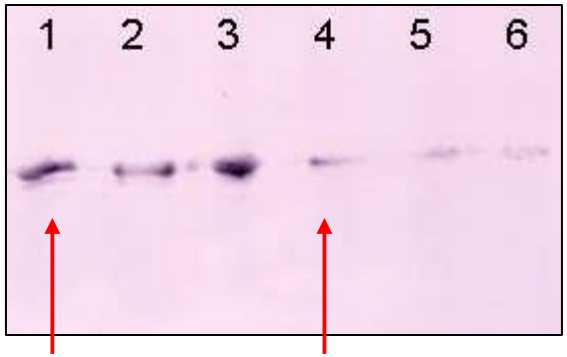
Strategies to overcome some limitations:

- Scrambling:
  - the use of auxotrophic strains and/or suppression of transaminase activity
  - inhibition with a 10-fold excess of unlabelled amino acids in medium relative to the labelled aa

- Insolubility:
  - ↓ temperature of expression
  - adding highly soluble tags (GST, TrxA, …) to target proteins
  - the use of other expression systems:
    - » *Corynebacterium glutamicum* – expression of soluble *Streptoverticillium mobaraense* transglutaminase (Shinagawa et al., 2005)
    - » *Brevibacillus choshinensis* (Tanio et al., 2009; Udaka & Yamada, 1993)
    - » other non-bacterial systems

2.1.1. *Escherichia coli*

Expression of $CKI1_{RD}$



1. TB medium, 37℃ → 28℃, 3 h

2. M9 minimal medium, 37℃ → 28℃, 15 h
3. M9 minimal medium, 37℃ → 25℃, 15 h
4. M9 minimal medium, 37℃ → 28℃, 3 h
5. M9 minimal medium, 29℃ → 22℃, 3 h
6. M9 minimal medium, 22℃ → 22℃, 3 h

Purification of 1 L of bacterial culture: M9 medium → 7–10 mg of protein
TB medium → 20–25 mg of protein

2.1.2. *Pichia pastoris*

- Advantages
  - easy genetic manipulation
  - yeast grows to high cell density
  - high yield of secreted protein (100–500 mg/L when using a fermenter)
  - capable of post-translational modifications (glycosylation, proteolytic processing, disulfide bond formation)
  - stable isotope labelling of secreted proteins

- Disadvantages
  - inability of some post-translational modifications: prolyl hydroxylation, certain types of phosphorylation, high mannose glycosylation (N- and O-linked glycosylation patterns are different from higher eukaryotes)
  - *P. pastoris* uses alcohol oxidase 1 promoter induced by methanol; due to toxicity, methanol must be strictly controlled during cultivation → *Kluyveromyces lactis* (promoter LAC4 induced by galactose)

2.1.3. Baculovirus expression system

- For the expression of mammalian proteins such as kinases and membrane proteins.
- Based on the infection of insect cells (Sf9 cells) with a recombinant baculovirus carrying target gene and the subsequent expression of target protein by the insect cells.

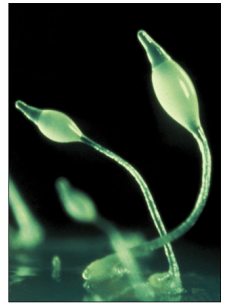
- Disadvantages
  - insect cells don’t grow in minimal medium → expensive labelled rich media
  - slow growth
  - low yields
  - eukaryotic cells cannot survive in deuterium oxide media (toxic for them) → combination of amino acid-type selective ^{15}N labelling and several basic triple resonance experiments

2.1.4. Mammalian cells: Chinese hamster ovary (CHO) cells, HEK 293 cells

- For production of mammalian proteins
- Advantages
  - Contain a mammalian N-glycosylation system → production of “authentic” mammalian glycoproteins
  - Amino acids are directly incorporated from the medium into the expressed protein → no scrambling of isotope labels

- Disadvantages
  - Slow growth
  - Low protein yield
  - Cells cannot grow on the isotopically enriched minimal media, mostly require serum → uniform isotope labelling is expensive → serum free medium supplemented with labelled amino acids purified from the bacterial or algal hydrolysate (Hansen et al., 1992)
  - eukaryotic cells cannot survive in deuterium oxide media (toxic for them) → combination of amino acid-type selective ^{15}N labelling and several basic triple resonance experiments

2.1.5. Slime mold *Dictyostelium discoideum*



- Promising eukaryotic expression system
- Advantages
  - Rapid cell growth
  - Simple media
  - Good yields (~ 9 mg from 10 g ^{13}C glycerol (Cubeddu et al., 2000)

2.1.7. Hybridoma cells

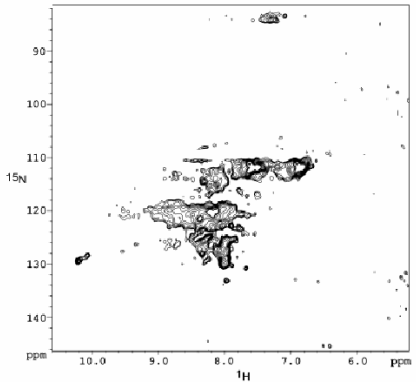
- For production of uniformly $^{15}N/^{13}C$ labelled antibodies using serum free media
- Yields 20–40 mg/l cell culture

2.1.8. Plants

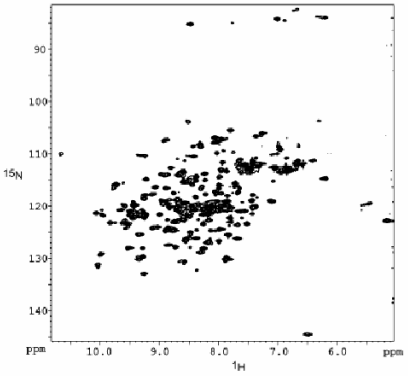
- Advantages
  - all constituting proteins are labelled and become available as functional, post-translationally modified, correctly folded proteins
  - Nitrogen source: $K^{15}NO_3$,$^{15}NH_4{}^{15}NO_3$
  - Carbon source: ^{13}C sucrose

Uniformly ^{15}N-labelled (>98%) potato plants (Ippel et al., 2004)

Hydroponic system ($K^{15}NO_3$)



Potato tuber lysate



Purified protease inhibitor PSPI:6.5

2.2. Cell-free expression system

- *In vitro* protein expression:

  DNA or mRNA for the target protein is added to the cell lysate (derived from *E. coli* or wheat germ) containing all the cellular components for protein expression (transcription and translation machinery) along with 20 amino acids, nucleoside triphosphates, several enzymes as well as buffers, salts, etc.

- Advantages
  - Can incorporate variety of reagents (e.g. detergents, protease inhibitors, chaperones, ligands) → may facilitate protein synthesis, folding, post-translational protein stability → useful for producing cytotoxic, integral membrane proteins, proteins containing multiple disulfide bonds (elimination of cellular transport and toxicity)

2.2. Cell-free expression system

- Advantages
  - The target protein is the only protein synthesized and labelled
  - Efficient technique for selecting labelling of certain aa and for specific aa position
  - Incorporation of non-natural aa (Fluoro-tryptophan [Neerathilingam et. al., 2005], L-3,4-dihydroxyphenylalanine [DOPA; Ozawa et al., 2005])
  - Reduces reaction volumes (µl-ml), quantities of expensive and unusual labelled aa, isotopic scrambling (transaminase activities are suppressed)
  - Commercial cell-free expression kits (expensive for large scale production – 1 ml ~ $350)

- Disadvantages
  - Expensive equipment: Roche RTS Proteomaster reaction device
  - Low yield of protein
  - Not all proteins are synthesized in vitro

Labelling for X-ray crystallography

- Incorporation of **selenomethionine** into proteins in place of methionine aids the structure elucidation of proteins using multi–wavelength anomalous diffraction (MAD).
  - The incorporation of heavy atoms such as selenium helps solving the phase problem in X-ray crystallography.

- **Neutron protein crystallography** provides a powerful complement to X-ray crystallography by enabling key hydrogen atoms to be located in biological structures that cannot be seen by X-ray analysis alone.
  - The availability of a fully deuterated protein eliminates the hydrogen incoherent scattering contribution to the background and brings further ~10-fold improvements in the signal to noise ratios.
  - The neutron Laue diffractometer LADI, run jointly by EMBL and ILL at the ILL high flux reactor in Grenoble, is a dedicated facility for neutron protein crystallography at high-resolution (1.5 Å).

Literature

- Staunton et al. (2006): Cell-free expression and selective isotope labelling in protein NMR. *Magn. Reson. Chem.* 44: S2-S9.

- Skrisovska et al. (2010): Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. *J. Biomol. NMR* 46: 51-65.

- Takahashi & Shimada (2010): Production of isotopically labeled heterologous proteins in non-E. coli prokaryotic and eukaryotic cells. *J. Biomol. NMR* 46: 3-10.

- Goto & Kay (2000): New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Opin. Struct. Biol*. 10: 585-592.