

Over-expression of proteins for NMR studies

Below are general comments on NMR sample conditions, protein expression systems, and two recipes for over-expressing proteins on minimal media for incorporating ^{15}N and ^{13}C . Both recipes have been used at labs here at UCHC and both work very well giving protein expression levels on par with LB for most systems. The vitamin and trace element stock solutions for both protocols have been prepared and can be borrowed. See Mark or Li for details.

Note that for some stubborn proteins we have used BioExpress cell growth media from Cambridge Isotope Laboratories (CIL), but the cost is expensive for $^{15}\text{N}/^{13}\text{C}$ labeling (\$1,700 per liter list). It is advisable to always deal with the CIL sales rep when purchasing labeled compounds, as you will generally get a better price than list.

NMR sample conditions

- **Concentration**
 - Protein concentrations 200 μM and higher are desired for structure determination, although this is just a rule of thumb and is dependent on how well the protein behaves.
 - Protein concentrations of 50 μM or higher are needed to run a ^{15}N HSQC, although at 50 μM it may need to be an overnight experiment.
- **pH**
 - The optimal pH of NMR samples to achieve the highest sensitivity (for NH protons) is around 4.5. This is due to NH exchange which is base catalyzed. pH values from 4.5 – 7.5 are very doable, with sensitivity dropping as the pH goes up (for experiments which measure NH protons; carbon bound protons have the same sensitivity at any pH). pH values above 7.5 are possible, but more sample may be needed.
- **Buffers**
 - Historically NMR spectroscopists have used buffers with no detectable protons by NMR, such as phosphate, or have used deuterated buffers. However, with ^{15}N and ^{13}C labeling and modern pulse sequences any buffer may be used.
- **Ionic strength**
 - The sensitivity of NMR experiments drops as ionic strength increases. It is therefore advisable to use as low an ionic strength as possible to keep your protein happy.
 - Ionic strengths of 50 mM are routine and much higher ionic strengths are possible as long as the protein concentration is not too low. We have worked at 1M NaCl concentrations, but with significant sensitivity loss.
 - 3 mm NMR tubes can be used for high salt samples to reduce the salt effect responsible for signal loss, although the protein concentration needs to be higher for the smaller tube.
 - The new 800 MHz spectrometer will have a salt tolerant probe that should help with higher ionic strength samples. For samples with higher than 125 mM ionic strength a higher signal will be obtained when dual 2.5 mm NMR tubes are used. There is an added advantage that a reduced amount of sample is needed with the dual 2.5 mm tubes.
- **Volume**
 - Conventional 5 mm NMR tubes require around 600 μl of sample.
 - Shigemitsu tubes can be used with as little as 300 μl of sample, although around 320 μl is typically better to account for some loss. The dual 2.5 mm tubes require a volume of less than 300 μl and Shigemitsu versions of the 2.5 mm tubes exist to further reduced the sample volume necessary.
- **Locking and D_2O**

- All NMR samples need around 5% or more of a deuterated buffer added. For most biological samples it is typical to use 95% H₂O / 5% D₂O. For other solvents such as DMSO, chloroform, methanol, etc it is best to use 100% deuterated solvents.
- For some samples it is fine to add a few µl of D₂O to your final protein samples to create the 5% D₂O. However, for proteins that need a certain level of ionic strength to keep them in solution, adding straight D₂O can cause the sample to precipitate around the D₂O as it is added to the sample. To avoid this it is often desirable to lyophilize 1 ml of NMR buffer (assuming it is a non-volatile buffer) and re-suspend in 1 ml D₂O. This D₂O buffered solution can then be added to create the 95% H₂O / 5% D₂O sample.
- For experiments that do not detect NH protons it may be desirable to use 100% D₂O rather than 95% H₂O / 5% D₂O.
- **Other buffer components & sterilization**
 - The following are often added to NMR samples to keep them happy
 - EDTA (to bind paramagnetic metal ions)
 - Protease inhibitors such as AEBSF and protease inhibitor cocktails (to avoid protease degradation)
 - DTT or TCEP (to break disulfides)
 - NaN₃ (as a preservative)
 - Samples are typically filter sterilized to remove precipitate and sterilize sample.
- **Molecular weight**
 - **Quality:** NMR spectra quality deteriorates as the molecular weight (MW) increases due to a slowing of the rotational correlation time. There are two types of relaxation in NMR; T₁ and T₂.
 - **T₁:** T₁ is the time it takes for the magnetization to re-equilibrate and dictates the length of time we must wait before we can measure the signal again. T₁ times range from a few hundred msec for smaller proteins to seconds for large proteins. Thus as the MW increases we must wait longer times between scans, thus reducing the amount of signal averaging that can be accomplished (S/N increases as the square of the number of signal averaged scans).
 - **T₂:** T₂ relaxation times measure how fast coherence of the NMR signal is lost and T₂ times get faster as tumbling slows. Once the coherence is lost the signal is gone and can no longer be detected. The linewidths in NMR signals are equal to $(\pi T_2)^{-1}$, thus as MW increases the linewidths broaden and the sensitivity drops. For large MW systems the signal becomes so broad that it simply vanishes into the baseline. For modern 3D NMR experiments there are a series of pulses and delays. For MW systems > 30 kDa the T₂ times are so fast that the signal loses coherence during these delays and before we detect the signal. This is why ~ 30 kDa is the limit for solving a protein structure with conventional NMR experiments.
 - **Rules of thumb:** As a general rule of thumb, well behaved systems < 20 kDa are straightforward to solve by NMR. Systems between 20 kDa and 30 kDa are possible, but require much more effort. Systems > 30 kDa will likely fail for structural studies with conventional NMR experiments (see deuteration below). However, certain experiments such as the HSQC can work for larger systems.
- **Deuteration**
 - While conventional experiments fail above 30 kDa due to short T₂ times it is possible to increase T₂ times by replacing protons with deuterons in the protein. This can be accomplished by expressing protein with [U-¹³C, U-²D]-glucose. NMR can study systems of hundreds of kDa with deuteration, although there are many practical issues that must be overcome with these systems. It is beyond the scope of this document to discuss these issues so please see me to discuss possible large MW systems you may like to study.

▪ Vectors

- Vector choice is critical for obtaining high expression levels of proteins. Expression vectors based on the t7 promoter (pET) are excellent and there are some good systems based on the *tac* promoter (pGEX) as well. In house we have used pET-23a that has no tag (S. King lab), pET-16b with N-terminal His tag cleavable with Factor Xa (S. King lab), pMAL-c2 with Maltose Binding Protein (MBP) tag cleavable with Factor Xa (S. King lab), and pTEV home built vectors from the B. Hao lab. The Hao lab has several vectors with N- and C-terminal His tags, and GST tags. The His tag versions are based on t7 promoter (pET) vectors and the GST versions are based on pGEX *lac* promoter vectors. They have all been modified to provide a TEV protease cleavage site to remove the tag. TEV protease has higher sequence specificity than factor Xa or thrombin and is easily overproduced so no commercial proteases need to be purchased. It recognizes the epitope ExxYxQ(G/S) with cleavage after the Q.

▪ Tags

- It is common to use N- and C-terminal tags to aid in protein purification. However, many tags can cause problems with NMR spectra.
 - **Large tags:** The quality of NMR spectra will deteriorate as the MW increases. Increasing the MW due to the presence of a large protein tag will certainly cause problems.
 - **His tags:** Sometimes His tagged proteins will give high quality spectra. Other times they seem to cause some type of polymerization. When this occurs the MW increases to such a size that the NMR signal vanishes. This has happened for several proteins that we have studied even at 1 mM protein concentrations. Whatever the cause of this it does not appear to affect solubility, as the samples are always very soluble. We have also noted that in the cases where the NMR signal vanishes with a His tag that the tag will be un-cleavable even when it has a protease cleavage site.
 - **Maltose Binding Protein:** MBP is a very soluble protein and has been shown by many groups to increase expression levels and help keep over expressed proteins in the soluble fraction when expressed as a MBP fusion protein. MBP can also be used as an affinity tag for purification with amylose resin.
 - **Cleaving tags:** When possible it is always best to cleave affinity tags and attempt to design the clone so that there are no (or as few as possible) extra residues left behind after the cleavage. Commercial enzymes for cleaving proteins are expensive and often not highly purified (many groups purify these enzymes after purchasing them). Thrombin and factor Xa are also not extremely specific and can cause unintentional cleavage. TEV protease is a nice alternative to thrombin and factor Xa (see above).

▪ Unstructured residues

- **Dynamic range:** As discussed above, as the MW increases the rotational correlation time slows causing the linewidths to increase, thus reducing sensitivity. However, if there is a flexible loop or N- or C-terminal unstructured region these residues will have increased flexibility and generate sharp, strong NMR signals as compared to the folded region of the protein. This will cause a dynamic range issue where the flexible residues will have a much higher signal strength obscuring the smaller folded region peaks. These peaks will lie in the middle of the spectrum where many other peaks will be located causing significant overlap problems.
- **Finding the correct clone:** If you find that there is a N- or C-terminal unstructured region of significant size (> 7 residues) it will generally be less work to re-clone the protein without the unstructured residues then to deal with them in the spectra. If the flexible residues are located in a loop there is not much that can be done.

General notes on cell growths

▪ Cells

- Use BL21 cell lines. HMS174 cells are okay, but typically not as good as BL21 cells. Never use mm294 cells for protein expression.
- BL21(DE3) cells are needed for T7 promoter based systems (pET) to produce the T7 RNA polymerase. Any BL21 cell line will work for *tac* promoter systems (pGEX). For expression of proteins that may be toxic to the cell BL21(DE3)-pLysS or BL21(DE3)-pLysE may be used which produce low (pLysS) or high (pLysE) levels of T7 lysozyme protein which inhibits T7 RNA polymerase shutting down any leaky T7 RNA polymerase before induction. pLysS has been shown to have no affect on protein production as it is expressed at too low a level and is overwhelmed by induction, but pLysE will reduce protein yield and should only be used when expressed protein toxicity is extreme.

▪ Induction

- It is wise to create a growth curve without induction first to find the optimal OD₆₀₀ to induce. Inducing at OD₆₀₀ of 0.8 is common, but induce too early and you will reduce protein expression levels. Induce to late when the cells are no longer in log phase and protein production will be reduced as well.

▪ Induction temperature

- Typical bacterial cell growth is at 37°C. However, increased expression levels are often achieved if the protein induction is at a lower temperature.
 - Induction at 30°C, 25°C, 20°C, or even 15°C can lead to greater protein production. For proteins that are produced in inclusion bodies the lower protein induction temperature may keep more of the protein in the soluble fraction.
 - Some groups grow their cells at 37°C and then chill them to the desired temperature for induction while others grow their cells at the desired induction temperature throughout.
 - Note that if you grow the cells at a reduced temperature the cell growth times will be longer, but the same final OD₆₀₀ values should be reached as a growth at 37°C. As a rule of thumb every 7°C will cause a 2X increase in cell growth or induction time.

Protocol 1 for minimal medium cell growths

Preparation of stock solutions

• Solution 1: FeCl₂ solution (100 ml)

- 8 ml concentrated HCl
 - 5 g FeCl₂•4H₂O [198.8 g/mol]
 - 184 mg CaCl₂•2H₂O [147.0 g/mol]
 - 64 mg H₃BO₃ [61.83 g/mol]
 - 40 mg MnCl₂•4H₂O [197.9 g/mol]
 - 18 mg CoCl₂•6H₂O [237.9 g/mol]
 - 4 mg CuCl₂•2H₂O [170.5 g/mol]
 - 340 mg ZnCl₂ [136.3 g/mol]
 - 605 mg NaMoO₄•2H₂O [241.95 g/mol]
- Make up to 100 ml with H₂O. The solution is green and you will need to stir for several hours before everything dissolves.
 - Store at RT

• Solution 2: Vitamin solution (1000 ml)

- 1.1 mg biotin stored at 4°C
 - 1.1 mg folic acid* stored at RT
 - 110 mg PABA (para-aminobenzoic acid) stored in 4°C
 - 110 mg riboflavin* stored at RT
 - 220 mg pantothenic acid stored in 4°C
 - 220 mg pyridoxine HCl* stored at RT in dessicator
 - 220 mg thiamine HCl* stored at RT
 - 220 mg niacinamide stored at RT
- Note that these vitamins are light sensitive.
 - Add 500 ml H₂O + 500 ml high-purity ethanol, then filter sterilize. The solution will be bright yellow.
 - Store at 4°C in a container designed for light sensitive material or cover the container with aluminum foil.

• Solution 3: "SBM" solution

- 16.5 g KH₂PO₄
 - 87.5 g K₂HPO₄
 - 18.25 g NaCl
- Add H₂O to 500 ml and autoclave
 - Store at RT

• Solution 4: "O" solution

- 28.8 g MgCl₂•6H₂O
 - 10 ml FeCl₂ stock solution
- Add H₂O to 500 ml and filter sterilize
 - Store at RT

• Solution 5: "S" solution

- 4.8 g K₂SO₄
- Add H₂O to 100 ml and then autoclave
 - Store at RT

- **Solution 6: Thiamine solution (1 mg/ml)**
 - Make up 10 ml of a 1 mg/ml solution of thiamine in H₂O and filter sterilize.

Preparation of 1 liter of minimal media - Protocol 1

- 940 ml H₂O
- 40 ml "SBM" solution
- 1 ml "S" solution
 - Mix and autoclave
- 2 ml "O" solution
- 1 ml vitamin solution
- 1 ml thiamine solution (1 mg/ml)
 - Mix and filter sterilize into the autoclaved growth medium of "SBM" and "S" solutions.
- Dissolve 1 g ¹⁵NH₄Cl or (¹⁵NH₄)₂SO₄ into 5 ml H₂O
 - Filter sterilize into the growth medium
 - Unlabelled compounds can also be used
- Dissolve 2 - 4 g [U-¹³C]glucose into 10 ml H₂O
 - Filter sterilize into the growth medium
 - Unlabelled compounds can also be used
 - We have found that 3.5 g glucose will give optimal yields. However, for expensive growths you should test what the optimal amount of glucose is for your system.

Cell growth

- Inoculate 5 ml LB medium with a single colony from a fresh LB plate
 - Grow culture for around 6 hours
- Use 200 µl of LB growth to inoculate 25 ml of minimal media starter culture for each liter of cells you intend to grow.
 - For example inoculate 200 µl LB into 25 ml starter for 1 L culture, 400 µl LB into 50 ml starter for 2 L culture, 50 µl LB into 6.25 ml starter for 250 ml culture, etc.
 - Minimal media is prepared as above, but with unlabeled nitrogen and carbon sources
- Grow starter culture at 37°C overnight. Culture should grow to an OD₆₀₀ of 2-3
- Spike 1 L of minimal media with 25 ml of the overnight minimal medium starter culture.
 - Grow and induce cells as normal.
 - Vary the amount of starter culture depending on the size of the cell growth.
- Do not use more than 1 L of culture in a 5 L flask.
 - Aeration is critical for good growth.
- IPTG-inducible transcription is repressed by glucose. Therefore higher levels of IPTG may be needed for inductions below OD₆₀₀ ~ 0.8. Typically by OD₆₀₀ ~ 0.8 glucose has been consumed.

Reference

Weber DJ, Gittis AG, Mullen GP, Abeygunawardana C, Lattman EE, and Mildvan AS. "NMR docking of a substrate into the X-ray structure of staphylococcal nuclease." *Proteins* (1992) **4**, 275-287.

Protocol 2 for minimal medium cell growths

Preparation of stock solutions

- **Solution 1: Metal (micronutrient) solution (1000 ml)**
 - 3.70 mg $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}\cdot 4\text{H}_2\text{O}$ [1235.9 g/mol]
 - 24.7 mg H_3BO_3 [61.83 g/mol]
 - 7.14 mg $\text{CoCl}_2\cdot 6\text{H}_2\text{O}$ [237.93 g/mol]
 - 2.50 mg $\text{CuSO}_4\cdot 5\text{H}_2\text{O}$ [249.68 g/mol]
 - 15.8 mg $\text{MnCl}_2\cdot 4\text{H}_2\text{O}$ [197.91 g/mol]
 - 2.88 mg $\text{ZnSO}_4\cdot 7\text{H}_2\text{O}$ [287.53 g/mol]
 - Make up to 1 L with H_2O . You may need to stir for several hours before everything dissolves and filter sterilize.
 - Store at RT

- **Solution 2: Vitamin solution (1000 ml)**
 - 400 mg choline chloride stored at RT
 - 1000 mg biotin stored at 4°C
 - 500 mg folic acid* stored at RT
 - 50 mg riboflavin* stored at RT
 - 500 mg pantothenic acid stored in 4°C
 - 500 mg pyridoxine HCl (pyridoxal HCl)* stored at RT in dessicator
 - 500 mg thiamine HCl* stored at RT
 - 500 mg niacinamide (nicotinamide) stored at RT
 - 1000 mg myo-inositol stored at RT
 - Note that these vitamins are light sensitive.
 - Add 500 ml H_2O + 500 ml high-purity ethanol, then filter sterilize. The solution will be bright yellow.
 - Store at 4°C in a container designed for light sensitive material or cover the container with aluminum foil.

- **Solution 3: 1M MgSO_4 (100 ml)**
 - 24.65 g $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$ [246.47 g/mol]
 - Make up to 100 ml with H_2O and filter sterilize.
 - Store at RT

- **Solution 4: 1M CaCl_2 (10 ml)**
 - 1.47 g $\text{CaCl}_2\cdot 2\text{H}_2\text{O}$ [147.0 g/mol]
 - Make up to 10 ml with H_2O and filter sterilize.
 - Add a bit of acid if needed to solubilize
 - Store at RT

- **Solution 5: 0.1M $[\text{NH}_4]_2[\text{Fe}][\text{SO}_4]_2$ (1 ml) Prepared fresh**
 - 39.2 mg $[\text{NH}_4]_2[\text{Fe}][\text{SO}_4]_2\cdot 6\text{H}_2\text{O}$ [392.14 g/mol]
 - Add 1 ml of H_2O to ammonium iron(II) sulphate in an eppendorf tube and mix
 - Add a bit of acid if needed to solubilize
 - Store at RT

- **Solution 6: Thiamine solution (1 mg/ml)**
 - Make up 10 ml of a 1 mg/ml solution of thiamine in H_2O and filter sterilize.

Preparation of 1 liter of minimal media – Protocol 2

- 980 ml H₂O
- 6 g Na₂HPO₄
- 3 g KH₂PO₄
- 500 mg NaCl
 - Mix, pH to 7.2, and autoclave

- 2 ml MgSO₄ solution
- 0.1 ml CaCl₂ solution
- 0.1 ml [NH₄]₂[Fe][SO₄]₂ solution
- 1 ml vitamin solution
- 1 ml metal (micronutrient) solution
- 1 ml thiamine solution (1 mg/ml)
 - Mix and filter sterilize into the autoclaved growth medium

- Dissolve 1 g ¹⁵NH₄Cl or (¹⁵NH₄)₂SO₄ into 5 ml H₂O
 - Filter sterilize into the growth medium
 - Unlabelled compounds can also be used

- Dissolve 2 - 4 g [U-¹³C]glucose into 10 ml H₂O
 - Filter sterilize into the growth medium
 - Unlabelled compounds can also be used
 - We have found that 3.5 g glucose will give optimal yields. However, for expensive growths you should test what the optimal amount of glucose is for your system.

Cell growth

- Inoculate 5 ml LB medium with a single colony from a fresh LB plate
 - Grow culture for around 6 hours
- Use 200 µl of LB growth to inoculate 25 ml of minimal media starter culture for each liter of cells you intend to grow.
 - For example inoculate 200 µl LB into 25 ml starter for 1 L culture, 400 µl LB into 50 ml starter for 2 L culture, 50 µl LB into 6.25 ml starter for 250 ml culture, etc.
 - Minimal media is prepared as above, but with unlabeled nitrogen and carbon sources
- Grow starter culture at 37°C overnight. Culture should grow to an OD₆₀₀ of 2-3
- Spike 1 L of minimal media with 25 ml of the overnight minimal medium starter culture.
 - Grow and induce cells as normal.
 - Vary the amount of starter culture depending on the size of the cell growth.
- Do not use more than 1 L of culture in a 5 L flask.
 - Aeration is critical for good growth.
- IPTG-inducible transcription is repressed by glucose. Therefore higher levels of IPTG may be needed for inductions below OD₆₀₀ ~ 0.8. Typically by OD₆₀₀ ~ 0.8 glucose has been consumed.