Mammalian proteins are frequently synthesized in bacterial, yeast, insect, or mammalian cell cultures using recombinant DNA technology. One of the concerns is that proteins thus synthesized are pure and folded properly. Host protein impurities would be immunogenic while improperly folded proteins, even though they are pure, could be of diminished biological activity, immunogenic, or form aggregates that are immunogenic. Another concern is that the protein is deamidated at some site or a methionine residue is oxidized during production. These are typical problems that must be addressed in the production of therapeutic proteins.

How would you go about convincing a federal regulatory agency that the therapeutic protein you produced
   a) is pure, i.e. it contains no immunogenic host proteins,
   b) has the correct (native) conformation,
   c) is not deamidated at any of the asparagine residues in the protein,
   d) has no oxidized methionine residues, and
   e) has no aggregates.

Describe how you would go about answering the questions above. You can use either bioassays or chemical assays to address these questions.

CRIB

The questions above ask you to describe analytical methods that would confirm the purity and safety of human therapeutic proteins manufactured in cell cultures by recombinant DNA technology. It is generally the case that multiple methods could be used, but one method is generally superior to others. Depending on which method you selected you could have received either partial or full credit. I have listed analytical solutions to the questions in what I believe to be descending order of preference.

ESTABLISHING PRODUCT PURITY, ALONG WITH THE ABSENCE OF HOST PROTEINS.

It is critical in human therapeutic proteins that will be administered to patients multiple times that they are pure. The presence of structural isoforms of the therapeutic protein or host proteins could be immunogenic. Chronic use of the therapeutic protein would trigger an immune response in subjects taking the protein preparation.

2-Dimensional gel electrophoresis, using isoelectric focusing in the first dimension and SDS-gel electrophoresis in the second dimension is a good method to separate all the proteins in the preparation. A large amount (≈100 μg) of the sample
would be administered to the gel to enable detection of trace amounts of impurities. Proteins on the 2-D gel runs should be visualized with Sypro Ruby and Silver staining in repeat runs, both of which can detect slightly less than a ng of protein. Most, but not all the proteins in the sample will be revealed with this approach.

Next, you must be concerned with identifying the spots in the gel. The major spot on the gel should be excised and characterized by matrix assisted laser desorption ionization (MALDI) mass spectrometry in two ways. One would be by direct mass analysis to confirm that the protein has approximately the correct molecular weight to within ±10 atomic mass units. This of course depends on the mass accuracy of the instrument. [Aggregates can also be detected by MALDI-MS, but with more difficulty and not with 2-D gels as will be discussed below.] The second approach would be to tryptic digest the protein in a gel spot (excised from the gel) and analyze the peptide fragments by tandem mass spectrometry (MALDI-MS/MS in this case with our ABI 4800 or a Bruker instrument) to identify their sequence. Commonly one would expect to identify 10-20 peptides from the protein with a confidence level ≥98%.

Still another mode of identification would be Western blotting. Obviously the major spot would be identified, but there is a high probability isoforms would be recognized as well. If there are multiple isoforms spots in the gel, the protein preparation would not be acceptable for human use. There is the concern these isoforms would be immunogenic.

All other spots in the 2-D gel would likely be from the cellular host. In such a case preparative gel electrophoresis would be in order to obtain enough protein to allow MALDI-MS/MS analysis of these proteins. Proteins would again be tryptic digested in the gel and peptides eluted from the gel and applied to a MALDI plate. MALDI-MS/MS (described in greater length below along with ESI-MS/MS) would allow some of the peptides to be sequenced, identified, and traced back to their genomic origin through a search of DNA database. If they come from the host this will be revealed in a database search for the genomic origin of parent protein.

2-Dimensional liquid chromatography would be a second approach to resolving the components in the therapeutic protein sample. Size exclusion chromatography (SEC) and reversed phase chromatography (RPC) would be the chromatographic methods of choice. Detection in this case would be by absorbance at 214 nm. Roughly a 100 mg sample could be used and the detection limit at 214 nm is 1 ng. Fractions collected from the RPC column would again be tryptic digested and examined by MALDI-MS/MS or ESI-MS/MS in a fashion almost identical to 2-D gel electrophoresis. Again an immunological assay would be used as well to confirm the presence of isoforms. Host proteins would be identified by MS/MS. This is probably not as good a method as gel electrophoresis.

Shotgun proteomics is a third approach which would be quick, but far less desirable. The sample would be tryptic digested immediately and the peptide cleavage fragments resolved by capillary reversed phase chromatography before peptide sequencing with tandem mass spectrometry. Because this approach has a dynamic range approaching 60,000 it would be possible to detect both the therapeutic protein and host proteins in the same analysis. What it would not do well is detect isoforms of the therapeutic protein.
CONFORMATIONAL ANALYSIS, I.E. IS THE STRUCTURE NATIVE.

Actually, federal regulatory agencies have never found an answer to how best to examine protein conformation. If the proteins has some biological activity that can be assayed this one way conformation can be established. It is generally the case that a molecule has to have native conformation to be active.

Optical methods such as circular dichroism, Raman spectroscopy, and infrared spectroscopy can see major conformational differences in proteins, but not small differences in conformation.

Chromatographic methods such as ion exchange and hydrophobic interaction chromatography frequently separate proteins that are not conformationally identical, but not always. Native gel electrophoresis does that same thing, but does not always see conformational differences.

Polyclonal antibodies generally bind to discontinuous epitopes. That means that conformational changes can cause the loss of an epitope and one of the polyclonal antibody species will not bind. This fact has been developed into an assay for changes in protein conformation, but is doesn't detect all conformational changes.

Although there are multiple ways to examine protein conformation there is no single definitive method.

Finding Deamidated Residues in a Protein

Some asparagine residues in proteins deaminate during purification to yield an aspartate residue, generating an isoform of the therapeutic protein. This alteration changes the isoelectric point of the protein and can also be immunogenic.

Reversed phase chromatography-MS/MS is widely used to assess the quality of therapeutic proteins. This is similar to the shotgun proteomics method described above. The sample is first tryptic digested. All the peptide fragments derived from a protein are then resolved by reversed phase chromatography (RPC). A good RPC column will produce 300 peptide peaks. This will almost resolve all the peptides in a sample, but not quite. Some will coelute. This is not a problem because the first dimension of MS can resolve up to 100 coeluting peptides in most cases. Peptides eluting from the RPC are electrospayed into the first dimension of a tandem mass spectrometer. During the electrospay process peptides acquire a positive charge ranging from +1 to +4. In the first dimension of mass spectrometry peptides are resolved by their mass to charge ratio and then directed into a collision cell where they collide with a gas such as helium and are fragmented into a series of ions. Ions that have been designated as b⁺ and y⁺ arising from fragmentation in the collision cell are related to the sequence of the peptide and are transported into a second dimension of mass spectrometry. These fragment ions are again separated on the basis of their mass to charge ratio, but this time the difference of mass to charge ratio of the ions reveals the sequence of a peptide fragmented in the collision cell. Data available from this process for identification of a peptide is RPC retention time, molecular weight, and sequence. Moreover, this sequence is easily compared with the expected sequence.
from a DNA database. A deamination site would be immediately apparent. First it would have a retention time different than that of the asparagine bearing peptide. Second it would have the wrong molecular weight by 1 atomic mass unit. And third, it would have an aspartic acid in the sequence that the DNA database would list as an asparagine residue. RPC-MS/MS can pick up less that 1% deamidation. If a carbon-13 labeled internal standard of the deamidated peptide were added to samples it would be possible to see 0.01% deamidation.

Reversed phase chromatography (RPC) along with cation exchange (CEX) chromatography could be used to identify deamination although the method would be inferior to RPC-MS/MS. Again proteins would be tryptic digested immediately. Peptide standards for the asparagine and aspartic acid containing peptides would have to be synthesized to establish their CEX and RPC retention times. Peptides from the tryptic digest would be separated by CEX first at pH ~3 using an ionic strength gradient and fractions from the CEX column then resolved by RPC using 0.1% TFA and an acetonitrile gradient ranging from 0 to 70% acetonitrile.

Isoelectric focusing (IEF) would pick up the fact that the isoelectric point of the protein had changed as well. The problem with IEF is that it could identify the site of modification or the fact the change was from an asparagine to aspartate transformation.

FINDING OXIDIZED METHIONINE RESIDUES IN A PROTEIN.

Methionine oxidization would be detected in the same way as deamidation. The first choice would be RPC-MS/MS. The second would be RPC. IEF really wouldn’t work in this case.

Reversed phase chromatography-MS/MS would be applied to the analysis of a tryptic digest of the protein in the same manner as described above. There are so few methionine residues in a protein that tryptic peptide standards would be synthesized to establish the retention time and fragmentation pattern of the unoxidized peptide. The internal standard peptide would then be oxidized with hydrogen peroxide to convert methionine to methionine oxide and analyzed again to establish the new retention time and fragmentation properties of the oxidized peptide. Retention time would increase after methionine oxidization. Again all peptides in the sample would be examine by electrospray ionization along with RPC and tandem mass spectrometry to establish retention times and fragmentation patterns. Using carbon-13 labeled internal standards that were added to samples after proteolysis it would be possible to see 0.01% oxidation of methionine.

Reversed phase chromatography (RPC) along with cation exchange (CEX) chromatography could be used again in exactly the same way as in the search for deamidation.

RECOGNIZING AGGREGATED STATES OF A PROTEIN.

There are two types of protein aggregation. One is aggregation of the protein with itself. The other is aggregation of the protein with another protein. The linkage between proteins in the aggregate can either be non-covalent or covalent.

Size exclusion chromatography (SEC) is a very good, fast method for examining the hydrodynamic volume (which is roughly related to molecular weight) of a protein. Separation in SEC occurs by differential permeation into the pores in a porous
chromatographic matrix. Ideally the pores are of a very narrow pore distribution. The pore diameter for aggregate analysis should be of roughly 300 angstroms (exclusion limit is a little over 10^6 daltons with this pore size), but with smaller proteins 150 angstrom would work better (exclusion limit is roughly 600,000 daltons with the pore size). SEC generally resolves to baseline proteins varying two fold from each other in size. This means that the monomer, dimer, and tetramer of a protein can be resolved. The trimer can not be completely resolved from the dimer and tetramer, but it can be seen.

When the effluent from the SEC column is fed to a multiple low angle light scattering (MALLS) detector with simultaneous refractive index detection it will estimate the hydrodynamic volume and molecular weight of a protein by a completely different mechanism. MALLS/RI determines molecular weight by light scattering while RI estimates concentration. Molecular weight is related to light scattering at a given angle. Simultaneous analysis at multiple angles improves molecular weight estimation accuracy.

The more complicated case of aggregate analysis is when a smaller protein is aggregated to the therapeutic proteins. The molecular weight will be larger but not by much and worse yet, it could be a host protein. The good news in this case is that you would already have picked this second protein up in the 2-D gel electrophoresis analysis described above.

Covalent crosslinking would also have been detected by the proteomics methods described above, especially electrospray ionization-MS/MS. The point of covalent crosslinking would have been detected above because two of the peptides from the protein would have appeared together in a dimer instead of individually. This peptide dimer would have had the unusual property of having two amino-termini amino acids and two carboxy-terminal amino acids. When tryptic digestion is carried out in H_2^{18}O the basic C-terminal amino acid acquires two ^18O and the mass of the peptide increases four atomic mass units. Using this approach the peptide dimer would increase by eight atomic mass units. This is a totally unique property of crosslinked peptides.

**Native gel electrophoresis** could also be used. Actually it is of higher resolution than SEC but is slow and does not couple as easily to MALLS.

**Ultracentrifugation** is another method for estimating molecular weight. With long centrifugation times aggregates can be separated. Fractions from ultracentrifugation can be analyzed by MALLS.

**Field flow fractionation** is yet another method for aggregate analysis. It has the resolution of SEC and is easily coupled to MALLS and RI.
1. (30 points) The porphyrin complex of iron(II), Fe(Por), reacts with NO according to the following chemical equation:

\[
\text{Fe(Por)} + \text{NO} \xrightleftharpoons[k_{\text{off}}]{k_{\text{on}}} \text{Fe(Por)(NO)}
\]

(a) Describe how the rate constants \( k_{\text{on}} \) and \( k_{\text{off}} \) can be determined experimentally.

\[-d[\text{Fe(Por)}]/dt = k_{\text{on}} [\text{Fe(Por)}][\text{NO}] - k_{\text{off}} [\text{Fe(Por)(NO)}] \]

Use [NO] in excess and Fe(Por) limiting. This simulates pseudo-first order conditions. Follow reactants or products by UV-vis (or another convenient spectroscopic method). According to integrated form of the rate law above after substitution of \([\text{Fe(Por)(NO)}]\)
gives,

\[[\text{Fe(Por)}]_t = [\text{Fe(Por)}]_0 - [\text{Fe(Por0)}]_0 \exp[-(k_{\text{on}} [\text{NO}] + k_{\text{off}})t] \]

Therefore, \( k_{\text{obs}} = k_{\text{on}} [\text{NO}] + k_{\text{off}} \).

Collect data at different [NO]. Plot of \( k_{\text{obs}} \) versus [NO] afford \( k_{\text{on}} \) as the slope and \( k_{\text{off}} \) as the intercept.

(b) What are the units for \( k_{\text{on}} \) and \( k_{\text{off}} \) if the reaction is run in aqueous solution?

\( k_{\text{on}} \text{ L mol}^{-1} \text{ s}^{-1} \) or \( \text{M}^{-1} \text{ s}^{-1} \)

\( k_{\text{off}} \text{ s}^{-1} \)

(c) Describe the bonding interaction between the nitrosyl ligand and iron.

NO is a sigma donor and pi acceptor. It could be linear or bent.

2. (30 points) The reaction kinetics for the above reaction in question (1) have been measured at different temperatures. The dependence of the rate constants \( k_{\text{on}} \) and \( k_{\text{off}} \) on temperature was found to adhere to transition-state theory:

\[
k = \frac{k_B T}{h} \exp\left(\frac{-\Delta H^\ddagger}{RT} + \frac{\Delta S^\ddagger}{R}\right).
\]

(a) Define \( \Delta H^\ddagger \) and \( \Delta S^\ddagger \).

\( \Delta H^\ddagger \) enthalpy of activation, kJ mol\(^{-1}\)

\( \Delta S^\ddagger \) entropy of activation, J mol\(^{-1}\) K\(^{-1}\)
(b) If $\Delta S^\circ$ is a large positive value, what does this imply about the reaction mechanism and why?

The transition state is less ordered than the reactants, implies dissociative mechanism.

3. (30 points) Dihydrogen complexes are coordination compounds containing intact H$_2$ as ligand. The prototypical complex is M(CO)$_3$(PR$_3$)$_2$(H$_2$) (see structure below).

![Structure of M(CO)$_3$(PR$_3$)$_2$(H$_2$)](image)

$R = Cy$ or iPr
$M = Mo$ or W

(a) Describe the orbital interactions between the metal and H$_2$ that would account for this coordinate bond.

H$_2$ is a sigma donor, the two electrons in $\sigma^b$ for H$_2$ act as 2-electron donor. Pi-backbonding from the filled d orbitals on the metal into $\sigma^*$ of H$_2$.

(b) Give two general synthetic methods for the preparation of dihydrogen complexes.

(a) From H$_2$ and the metal complex
(b) Protonation of metal hydride.

4. (10 points) Pick two of the following names of inorganic chemists and describe briefly one of their important contributions:

(a) Geoffrey Wilkinson
(b) Alfred Werner
(c) Henry Taube
(d) Stephen Lippard
(e) Albert Cotton

Google any of these names.
A (25 pts each). Suggest a metathesis-based synthesis (including the structures of all reagents and catalysts, approximate conditions, stoichiometry and mechanisms for preparation of compounds 1 and 2. Be sure to include synthesis of the precursor of the metathesis step.

1. 

2. 

B (40 pts). A long-standing limitation of the catalytic metathesis reaction has been the inability to control the stereochemistry of the resulting olefin. Discuss recent solutions to this problem with respect to catalyst design and provide a mechanism that explains how this catalyst favors highly selective formation of Z-olefins.

C. (10 pts). Who won the Nobel prize for the metathesis reaction and at what university(uni versus(ses)) is(are) he/she(they) located?

A1. JACS 2008, 130, 11297

A2. JACS 2009, 131, 15090

B. JACS 2009, 131, 3844 & 16630

C. 2005 – Yves Chauvin (Institut Franais du Pétrole), Robert H. Grubbs (Cal Tech), Richard R. Schrock (MIT)