## THEORY/PRINCIPLES

# SOLUBILITY AS A FUNCTION OF PROTEIN STRUCTURE AND SOLVENT COMPONENTS

Schein, C. H. (1990) Nature Biotechnology 8: 308. (Adapted)

### 1. Introduction

This review deals with ways of stabilizing proteins against aggregation and with methods to determine, predict, and increase solubility. Solvent additives (osmolytes) that stabilize proteins are listed with a description of their effects on proteins and on the solvation properties of water. Special attention is given to areas where solubility limitations pose major problems, as in the preparation of highly concentrated solutions of recombinant proteins for structural determination with NMR and X-ray crystallography, refolding of inclusion body proteins, studies of membrane protein dynamics, and in the formulation of proteins for pharmaceutical use. Structural factors relating to solubility and possibilities for protein engineering are analyzed.

It is generally known that proteins must be stored in an appropriate temperature and pH range to retain activity and prevent aggregation.

Proteins are often most soluble in solution conditions mimicking their natural environment. Serum proteins are soluble in a pH and salt range where mature insulin, which is stored in acidic granules in the cell, precipitates. Bacterial proteins may prefer buffers containing glutamate or betaine, compounds that accumulate in response to high concentrations of $Cl^-$ in the medium.[2] Caseins and other $Ca^{2+}$-associated proteins may require small amounts of the ion to maintain their native structure during purification.[4,5] The stability of lactase (β-galactosidase) is greatly increased in the presence of milk proteins.[6] But for most proteins, experimental determination of the solution properties can help in solvent design.

Low solubility in aqueous solvents is often regarded as an indication that a protein is "hydrophobic" as aggregation of integral membrane proteins after transfer to a hydrophilic environment is a well-described phenomenon.[7] But all proteins are to some extent hydrophobic, with tightly packed cores that exclude water.[8,9] As native, properly folded structures aggregate less than unfolded, denatured ones, there is an intimate relationship between solubility and stability. The free energy of stabilization of proteins in aqueous solution is very low (ca. 12 kcal/mole at 30°C);[10] consequently, proteins are on the verge of denaturation.[10,11] Protein stability can be increased by solvent additives or by alteration of the protein structure itself.

### 2. The Properties of Proteins in Solution

*2.1 Defining Solubility.* The chemist"s definition of solubility, parts purified substance per 100 parts of pure water, is not useful in a biological frame, as proteins in nature are never found in pure water. Blood and eukaryotic cytoplasm contain on the order of 0.15 M salt, with large quantities of trace metals, lipids, and other proteins. The cytoplasm of bacteria is more variable, with salt concentration ranging from 0.3 to 0.6 M.[2] The solubilizing effects of small molecules and even other proteins means that protein solubility does not correlate with purity.[12]

Operationally, solubility is the maximum amount of protein in the presence of specified cosolutes that is not sedimented by 30,000 x *g* centrifugation for 30 minutes.[13] An even stricter criterion, function retained after centrifuging for 1 hours at 105,000 x *g*, has been suggested for membrane proteins.[14] If one has a pure, lyophilized protein or a salt precipitate, one can determine solubility by adding increasing amounts of weighed solid, centrifuging, and measuring the protein content of the supernatant. Dissolved protein should reach a maximum (solubility) and level off. However, in the food industry, solubility is defined by sediment (in ml)

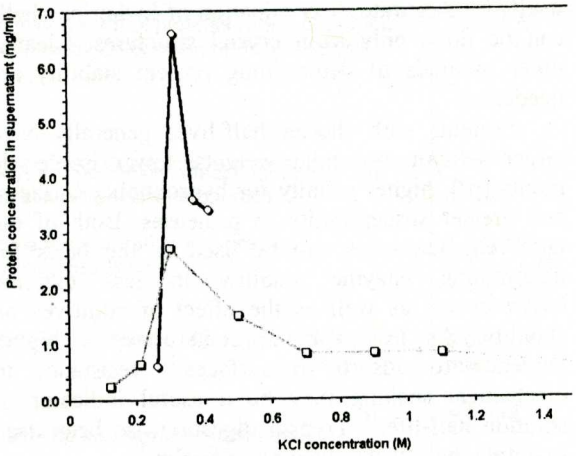


**Figure 1** - Solubility of T7 RNA polymerase as a function of salt concentration in 10 mM cacodylate buffer, 1 mM DTT, and 0.02 mM PMSF. The polymerase solution (*ca.* 1 mg/ml in 0.1 M KCl, 20 mM Tris-HCl, pH 7.9, 5% glycerol) was diluted in 1:10 with the indicated buffers and each sample was individually concentrated in 30-kDa MW cutoff "Centricons" (Amicon). The protein concentrating in the supernatant (measured by the Coomassie blue assay) after concentration is indicated (Protein A, squares). The top curve (Protein B, circles) is from a second measurement using a finer salt gradient and more protein per sample.

remaining after centrifuging; the solubility index is thus inverse to the actual solubility.[13]

The method described in Figure 1 allows definition of the solubility range of a protein in solution. A protein solution is diluted into a buffer series and the samples are centrifuged in microconcentrators. As one can conveniently concentrate about 50-fold, a relatively small amount of protein is sufficient for the estimation.

*2.2 Measuring Stability.* Methods for determining the thermodynamic stability of proteins use pH and temperature extremes, or high concentrations of denaturants.[10] Although useful for discerning changes in the structural stability of mutant proteins that are not clear from activity data, they are not directly correlated with the half-life of proteins in solution. Since aggregation occurs at temperatures well below the $T_d$ for proteins; additives that stabilize proteins against aggregation may not necessarily affect the $T_d$.[15]

The major problem with using thermodynamic measurements is their failure to account for the kinetic effects that lead to aggregation. Both the enthalpy ($\Delta H$) and entropy ($\Delta S$) of hydration vary greatly with temperature, but they cancel to give a relatively small measured free energy ($\Delta G$) of hydration that seems to vary little with temperature. Most of the temperature-dependent kinetic contribution, which is the more important in explaining hydrophobic effects, dissipates in alterations of the solvent structure around the protein and reversible deformation of the protein structure itself.[10,16] Accurate discrimination of hydration shells can be done only from crystal structures. Clearly, other methods of determining protein stability are needed.

Proteins with shorter half-lives generally have larger subunit molecular weights, lower isoelectric points (p*I*), higher affinity for hydrophobic surfaces, and greater susceptibility to proteases. Both of the latter characteristics can be used as the basis for determining enzyme stability in less extreme environments as well as the effect of additives on stability. As less stable proteins have a higher tendency to adsorb to surfaces,[17] resistance to mechanical shaking may be a useful indicator of solution half-life.[18] Trypsin digestion has been used to define the salt stabilization of hyalin.[5]

*2.3 Determining Surface Charge.* Isoelectric focusing gives the p*I*, the pH at which the protein shows no net charge in isoionic conditions. However, due to the binding of salt, one cannot assume that a protein in solution will be negatively charged at pHs above its p*I* (e.g., acidic caseins bind $Ca^{2+}$ and appear positively charged at pH 7.4). At pH 7.5 and 50 mM salt, most proteins will bind to DEAE-coupled resins if they are negatively charged and to phospho- and other negatively charged resins if they are positively charged. The charge strength can be estimated from the salt concentration required for elution. Gel methods for following the changes in surface charge during protein folding and aggregation have also been developed.[19]

Generally, charged proteins can be "salted in" by counterions. Binding of salts to proteins decreases bound water as well as the net charge at the surface. The solubility of lysozyme, a positively charged protein, was shown to vary more with the anion added than the cation; the anion dependence followed the Hofmeister series.[20]

The solubility of caseins with p*I* between pH 3 and 5 varies with the cation: sodium, potassium, and ammonium caseinates are all more soluble than those prepared with calcium or aluminum.[4,15]

*2.4 Determining Hydrophobicity.* Binding to resins coupled with hydrophobic groups, like Phenylsepharose (Pharmacia), indicates the presence of hydrophobic residues at the protein surface. Proteins are applied in high salt (0.7-1 M ammonium sulfate), which furthers hydrophobic interactions, and then eluted with a decreasing salt gradient. Most proteins elute between 0.5 and 0.1 M salt; very hydrophobic proteins will not elute into low salt buffer unless the polarity is decreased by adding ethylene glycol. If a protein does not bind to phenylsepharose, it either has a very hydrophilic surface (e.g., RNase A) or it is aggregated.

One can determine the hydrophobicity of a purified protein or follow changes in exposure of hydrophobic groups during folding by measuring interaction with a hydrophobic dye or radioactive tracer (e.g., 1-anilino-8-naphthalenesulfonate[21] or $^{125}$I-TID, 3-(trifluoromethyl)-3-(*m*-[$^{125}$*I*]iodophenyl) diaz-irine[3]).

*2.5 Aggregation and Precipitation.* Precipitation via any agent can be:

*2.5.1 Reversible*, as after precipitation with salts or large organic molecules like polyethylene glycol (PEG). Because PEG molecules are excluded from the surface of the protein, a two phase system develops and the protein is concentrated into a smaller volume, where its chances of interacting with another protein molecule to form an aggregate are increased ("excluded volume" model).[22,23] When the precipitant is removed, the water layer around the original molecule can reform and the protein molecules separate into soluble monomers.

The protein structure does not significantly change during reversible aggregation. A plot of protein in solution versus the concentration of the precipitant should look the same whether it is made with increasing precipitant (to precipitation) or decreasing precipitant (to solubility). Reversibility is assumed for most mathematical models of salting out[12] as well as some recent models of low salt aggregation phenomena.[25,26]

**Table 1** - Protein Cosolutes

| Compounds | Mode of Action | Amount Used |
|---|---|---|
| A. Osmotic stabilizers | Generally have little direct interactions with proteins; they affect the bulk solution properties of water. | |
| i. Polyols and sugars<br>• glycerol, erythritol, arabitol, sorbitol, mannitol, xylitol, mannisidomannitol (Man-Man), glucosylglycerol, glucose, fructose, sucrose, trehalose, isofluoroside | These stabilize the lattice structure of water, thus increasing surface tension and viscosity. They stabilize hydration shells and protect against aggregation by increasing the molecular density of the solution without changing the dielectric constant. | 10 – 40% |
| ii. Polymers<br>• dextrans, levans, polyethylene glycol | Polymers increase the molecular density and solvent viscosity thus lowering protein aggregation in a single phase system. At high polymer concentration, a two phase system develops and the protein aggregates in the phase where its concentration is the highest. | 1 – 15% |
| iii. Amino acids and derivatives<br>• glycine, alanine, proline, taurine, betaine, octopine, glutamate, sarcosine, α-aminobutyric acid, trimethylamine N-oxide (TMAO) | Small amino acids with no net charge, like Gly and Ala, have weak electrostatic interactions with proteins. Octopine is a derivative of Arg that is less denaturing to proteins. TMAO stabilizes proteins, even in the presence of denaturants like urea. Most of these compounds increase the surface tension of water. | 20 – 500 mM |
| B. Ionic compounds | These affect enzyme reactions and their stabilizing effects on protein occur in a much narrower concentration range than the above compounds. | |
| i. Stabilizing<br>• citrate, sulfates, acetate, phosphates, quaternary amines | Larger anions shield charges and can stabilize proteins at low concentrations. They lead to precipitation by competing for water molecules. | 20 – 400 mM |
| ii. Destabilizing<br>• chlorides, nitrates, thiocyanates | These are generally less stabilizing than large ions but are also useful in charge shielding at lower concentrations. | 20 – 400 mM[1] |
| C. Denaturing chaotrophs<br>• urea, guanidinium salts, trichloroacetates, cetylmethylammonium salts, organic solvents | Denatures; either stabilize the unfolded state of proteins (urea) or perturb protein structure by interfering with hydrogen bonding or disturbing the hydration shell. | 0.2 – 8 M |
| D. Other common additives (mostly nonphysiological) | Either interact directly with proteins or specifically affect impurities in the buffer but do not change the bulk solvent physical properties. | |
| • 2 – mercaptoethanol, dithiothreitol (DTT) | Reductants, protect free sulfhydryls from oxidation and prevent intermolecular glutathione sulfhydryl cross-linking. | 1 – 5 mM<br>0.1 – 1 mM[2]<br>1 – 4 mM |
| • phenylmethylsulfonyl fluoride (PMSF), benzamideine | Inhibit serine proteases by reacting with the active site serine hydroxyl group. | 0.02 – 0.05 mM[3] |
| • leupeptin, peptides | Protect from protease attack by serving as alternate substrates. | < 1 mM |
| • ethylenediaminetetraacetic acid (EDTA), ethylene-bis(oxyethylenenitrilo)tetra-acetic acid (EGTA) | Chelate divalent metal ions which may react with proteins; inhibit metalloproteases; EDTA at 5-20 mM aids in the lysis of G-bacteria by lysozyme. | 0.01 – 0.1 mM for buffers |
| • catechols, phenolics, $NaN_3$ | Bacteriocides | < 0.05% |

[1]Higher concentrations may be used when working with enzymes from halophilic organisms.
[2]DTT is a potential denaturant of proteins at higher temperatures and has limited solubility in high salt. The concentrations indicated should not be exceeded.
[3]Dissolve PMSF to 20 mM in isopropanol. The indicated concentrations represent the maximum solubility in aqueous buffers.

*2.5.2 Partially reversible*, a behavior frequently seen in pH-induced precipitation. Proteins precipitate around their pI and resolubilize as the pH is adjusted upward or downward. But during the pH adjustment, residues may change orientation. When the pH is readjusted, they may not be able to regain their former position and a mixture of structures (isomers) results. Even primary structure can change if a protein is held at acid pH for long periods of time, as for example the deamidation of as asparagine 21 of insulin.[1] A plot of protein in solution as a function of pH will depend on whether the protein has already precipitated. Kinetic modeling of pH-dependent aggregation has been attempted by linear regression.[27] Models could also use *hysteresis* in the folding-unfolding equations (thermodynamic cycles do not exactly connect.)

*2.5.3 Irreversible*, which is usually initiated by extreme changes in the solvent leading to protein denaturation. But some proteins (Figure 1) also precipitate irreversibly when concentrated above their maximum solubility in a given buffer. Inactive flakes of protein form and remain insoluble even on redilution of the sample or transfer to a buffer of the correct salt concentration. The nature of this tight intermolecular binding is not easy to study, as the aggregates arise from *many-body interactions* potentially involving all parts of the protein. The initiation could be direct interaction of surface hydrophobic residues, or, as aggregation shows cooperativity, partial disturbance of the hydration sheath, or unfolding of the protein structure allowing interaction between normally "buried" residues. Irreversible protein denaturation is not easily modeled. Thermal denaturation curves are calculated at very low protein concentration to avoid aggregation terms in the equations.[10]

## 3. Buffer Design for Maximizing Solubility

*3.1 The Properties of Water as a Solvent.* Water"s high dielectric constant and its tendency to solvate ions make it an active copartner in enzyme reactions. When NaCl crystals are added to water, the atoms attract each other with about 1/80th of the force in the dry state and the crystal dissolves. Analogously, dissolved proteins are coated with a "hydration shell" around charged and polar groups that prevents self-binding. This bound water does not freeze (some proteins are even efficient antifreezes)[28] and has different properties than in the surrounding solvent molecules.[29] Bulk water molecules and the protein are in continual fluctuation, which leads to instability in the system.[11,30,31]

3.2 *Protein Stability in the Solid State vs. Solution.* On the other hand, a protein completely stripped of its hydration shell is difficult to redissolve, as intermolecular hydrophobic forces must be broken. Lyophilization and other drying methods should thus be used with caution and osmotic stabilizers added where necessary to ensure that the protein can be rehydrated. The water content of dry milk powder is a compromise between shelf life, which decreases with increasing content of water, and solubility, which increases with hydration index.[13]

Proteins in the solid state have different levels of reactivity depending on the water content. Dried protein with a water content below 22-25%, the minimum required for conformational flexibility and activity, is thermostable. *Glycerol*, which stabilizes proteins in solution, acts as a humectant on the powder and causes decomposition (as indicated by the Maillard browning reaction) at much lower water content (5-15%). Conversely, *sorbitol* competes for the hydration water of the protein and does not enhance denaturation.[31]

*3.3 Solvent Additives.* There are many potential stabilizing cosolutes for proteins (Table 1). Buffers are described in several excellent reviews[32,33] and will not be covered here. Table 1 is separated into groups of compounds that have varying effects on the solvation properties of water: (1) *dielectric constant*, (2) *chemical potential*, (3) *viscosity*, and (4) the clathric tendency (*surface tension*). The first two qualities are related to protein *polarity*; the last two relate to the *diffusion* of the protein, its *partial molar volume*, and to *hydrophobic hydration.*

*3.4 Osmolytic Stabilizers.* The first group of compounds is osmolytes, which are not strongly charged and have little effect on enzyme activity up to at least 1 M concentration.[34] Their major effects are on the viscosity and surface tension of water, and hence on solvent ordering. Many of these compounds are used *in vivo* to control the osmotic pressure of eukaryotic and bacterial cells.

Osmolytes can be polyols, sugars, polysaccharides, neutral polymers, amino acids and their derivatives, and large dipolar molecules like TMAO. Glycerol is the most commonly used osmolyte, as it is easily removed by dialysis and does not interfere with ion-exchange chromatography. It does not alter the dielectric constant of the medium significantly and its stabilization effect on proteins seems to be due to its ability to enter into and strengthen the water lattice structure. High concentrations of glycerol decrease the diffusivity and the partial molar volume of proteins,[15] thus lowering the rate of aggregate-producing solute interactions.

Glycerol has major drawbacks, however, especially for large-scale work, as it is an excellent substrate for bacteria. Xylitol, a potential substitute, is not degraded well by bacteria and can be recycled from buffers by alcohol precipitation. PEG can be added to *in vitro* systems for nucleic acid and protein

synthesis, where sufficient molecular density but low ionic strength is needed.

*3.5 Ionic Stabilizers.* Ionic compounds and salts can stabilize protein structure by shielding surface charges. Salts are also considered as osmolytes and are used to some extent as such *in vivo*. *E. coli* transiently accumulated $K^+$ and glutamate after osmotic shock, but within 30 minutes switches to carbohydrates as osmoprotectants.[2] Most ionic compounds will affect the dielectric constant and the chemical potential of the solvent and the protein at concentrations well below where they affect the other bulk properties of the solution. Normal bacterial and mammalian enzymes function at a rather low salt concentration and are inhibited by high salt. *Halophilic* organisms, which can accumulate as much as 7 molal $K^+$ intracellularly, have adapted their enzymes to function in very high salt concentrations.[34]

There is no general rule on *salting in* of proteins; models that work for one protein are not necessarily applicable to another.[12] The salt concentration for maximum solubility frequently falls within a very narrow range. As shown in Figure 1, a 50 mM change in salt concentration gave as much as a 20-fold increase in dissolved T7 RNA polymerase. The solubilizing effect of ions is dependent on the size and charge distribution, but because polar groups on proteins are so diverse, it is hard to say a priori which ion will be best. Large ions are generally better at stabilizing proteins than small ones; in general, the more electronegative the ion, the more it interacts with and destabilizes protein structure.

The finest experimental work on the effects of salts on protein solubility (usually during salting out) has been done by crystallographers.[20,35,36] The assumed mechanism for salting out by small molecules is that they compete for water molecules until the concentration is too low to maintain the hydration sheath around the protein.[35]

*3.6 Divalent Cations.* These components have extremely pronounced solubility effects at very low concentrations. Even 1 mM $Ca^{2+}$ induces a conformational change characterized by insensitivity to trypsin in sea urchin hyalin, and $Ca^{2+}$ in the range of 1-20 mM encourage self-association.[5] $Zn^{2+}$ aids in insulin solubilization as well as crystallization.[1] As even tiny amounts of Cu, Zn, and Mn (among others) can also induce aggregation, chelators are often added to buffers.

*3.7 Denaturants, Chaotrophs, Cryoprotectants and Other Additives.* One can solubilize almost any protein (usually at the expense of its activity) by chemical denaturation with perturbing ions. Urea stabilizes the unfolded states of proteins because essentially all protein parts, from the backbone to the tryptophan side chains, are more soluble in 6 M urea than water as evidenced by the free energy of transfer into this solvent.[37] Another class of denaturants, "chaotrophs," like guanidinium, cetyltrimethylammonium salts, trichloroacetate, and thiocyanate ions, disrupt hydrogen bond formation and disturb the hydration shell around proteins.[15] Detergents, amphiphilic compounds that lower the surface tension of water, bind to hydrophobic areas of proteins.

Another class of denaturants, organic solvents, lowers the dielectric constant of water. The denaturing activity of hydrophobic solvents is due to a limited detergent effect and a competing interaction for the intramolecular hydrophobic interactions responsible for a stable tertiary structure. Some proteins are remarkably resistant to the denaturing effects of protic, hydrophilic organic solvents. The original method for isolation of insulin and human interferon-α from tissue and bacteria used extraction with acidic ethanol;[38] crambin can be crystallized from 60% ethanol.[39]

Two organic solvents frequently used as cryoprotectants, dimethyl sulfoxide (DMSO) and ethylene glycol, can also denature proteins. DMSO encourages unfolding by favoring peptide N-H • O=S solvent bonds over peptide N-H • O=C peptide bonds.[40] Ethylene glycol, by reducing solvent polarity, weakens structural hydrophobic interactions.

Unless a protein is to be used *in vivo*, it is general practice to include protease inhibitors, sulfhydryl reductants, bacteriocides, and chelating agents in small amounts to all protein solvents.

*3.8 Concentrating Proteins.* Limits on the maximum protein concentration one can achieve are due to the structure of the protein, the buffer components, and the purity of the protein preparation. Overloading the preparation on SDS acrylamide gels may not detect proteases that cause damage during concentration or storage. To minimize this contamination, during purification from bacterial extracts, the protein should completely change buffer at least three times. Suitable *transfer methods* are salt precipitation and dissolving in fresh buffer, binding to an affinity resin or HPLC column and elution, or gel filtration. Dialysis, flow-through affinity steps, and redissolving lyophilized samples do not count as buffer transfers. All purification buffers should be made with ultrapure water and HPLC grade chemicals where possible, and sterilized to avoid the reintroduction of bacterial contaminants.

The most commonly used *methods for concentration* are salt precipitation, affinity chromatography, ultrafiltration, and occasionally, chromatofocusing, electrofocusing, and freeze condensation (for cryoresistant proteins). Very stable proteins and peptides can be lyophilized or spray-dried and redissolved. One should get the preparation to as high a concentration as possible by judicious elution during the last affinity step.

The easiest method for concentrating proteins that cannot be lyophilized is ultrafiltration. Micro-concentrators are useful for volumes up to 10 ml. Stirred pressure cells (Amicon, Millipore, or equivalent) are available for volumes between 10 and 500 ml, and membrane type can be selected according to the size and hydrophobicity of the protein. Pressure cells did not work for Mx protein or T7 RNA polymerase, however, as aggregation at the membrane surface was too high. The stir rate should be kept to a minimum as concentrated protein solutions are shear sensitive. For T7 RNA polymerase, losses were lowest with the Sartorius vacuum dialysis system, where concentrations up to 40-50 mg/ml were obtained in 0.2 M ammonium sulfate (pH 7).

Hollow fibers or parallel plate continuous flow systems can be scaled up to any size. The Minitan system from Millipore is a good intermediate size for lab use. Protein loss on the membranes is significantly higher than the maximum predicted by the manufacturers.

## 4. Situations Where Protein Solubility Becomes Limiting

*4.1 Refolding Inclusion Body (IB) Proteins.* IBs behave like protein that has been irreversibly precipitated. To obtain active protein, high concentrations of chaotrophic agents in the presence of sulfhydryl reducing agents are used to unfold the chains, which must then be refolded during removal of the denaturants. The primary refolding problem is aggregation of partially unfolded protein. In one study, the maximum protein concentration for efficient refolding was only 20 μg/ml;[41] for interleukin-2 the maximum was only 1 μg/ml.[42] Concentration by ultrafiltration after refolding is possible, but losses due to proteolysis, aggregation of isomers, and membrane binding are frequently very high. For tissue plasminogen activator (t-PA), the folding to intermediate states is rapid but the proper disulfide bonds formed much more slowly. As the close to folded forms are relatively soluble, timed addition of more unfolded protein concentrate (a sort of "feedback') can allow much higher final concentration of the extract.[41a] Residual denaturant can also stabilize the native state of the protein; its optimal concentration in the final extract should also be determined. TMAO may be a useful osmolyte when refolding proteins from urea solution.[34]

Every protein contaminant present during refolding increases the total dilution necessary to avoid aggregation. In addition, partially unfolded proteins are excellent protease substrates. Thus, one of the major advances in IB protein refolding has been the development of purification steps that can be used in the presence of the denaturant. These include gel filtration, certain types of affinity chromato-graphy,[43] and a new method based on the interaction between a poly-histidine peptide fused to the protein of interest and a nickel chelate column.[44]

Alternate methods for refolding, such as binding denatured protein to thiol-Sepharose columns or other affinity matrices and eluting with denaturant-free buffers,[45] are also being explored. It is possible that activated thiol-Sepharose mimics the structure of protein disulfide isomerase.[84] Serine proteinases[46] and interleukin-4[47] refolding yields were greatly improved by pretreatment with glutathione. Interleukin-2 was renatured by dilution and autooxidation in the presence of $Cu^{2+}$.[42]

Appropriate choice of buffer during the refolding step can also improve yields at higher concentrations of protein.[11] As optimal refolding conditions vary with the protein, one should either dilute the denatured sample into or dialyze it against many different buffers, and measure active or soluble protein after centrifugation.

*4.2 Solubilization and Reconstitution of Membrane Enzyme Systems.* Difficulties in solubilizing proteins from membranes have greatly limited structure and function studies.[48] Membrane proteins function in an amphiphilic environment and fold differently from cytoplasmic proteins: they turn their hydrophobic sites outward rather than inward. This probably accounts for why computer programs developed from soluble proteins predict the opposite of the known X-ray structures for membrane proteins.[49] This structural difference also accounts for the failure of detergents to solubilize IB proteins.

The only way to isolate most *integral membrane proteins* is to extract them from their lipid environment with bulky detergents (typically Triton X-100 or Emulphogen BC-720). The protein is integrated into a detergent micelle with detergent replacing phospholipids or proteins that were previously in contact with the hydrophobic surfaces.[14] Even if the protein is not inactivated by this treatment, *low critical micelle concentration (CMC) detergents* interfere with protein concentration (by giving a gel), functional assays, and further purification steps (as the detergent"s properties dominate the protein"s).

Thus, proteins are transferred after the initial extraction to less harsh detergents forming smaller micelle[48] via gel filtration. For detergents with CMCs too low to allow for efficient dilution into monomers, one may need to use highly polar micelle dispersing agents like ethanediol or bile salts.[50] NMR structural studies of small membrane proteins in micelles[51] are possible.

A major advance in membrane protein crystallization is the use of "small amphiphiles" to replace detergents binding to the face of the protein. One can thus prevent some of the problems caused by phase separation at higher salt and protein

precipitation as the detergent in the micelles becomes too concentrated.

*Osmolytic stabilizers* (20% glycerol), or high salt (0.3-0.4 M KCl) added before the detergent, may stabilize the tertiary structure of the protein during extraction and dilution into proteoliposomes.[52] Glycerol or PEG is needed for efficient elution of membrane proteins from chromatofocusing columns. *In vitro* assays of transport systems from bacteria,[24,52] signal peptidase from yeast,[54] and the tamoxifen binding protein from a breast cancer cell line[55] were only possible by judicious control of the salt concentration during detergent extraction of the membrane.

As there is some evidence that high salt concentrations can stabilize secondary structural elements even during tertiary structure disruption,[21] the need for osmolytic stabilization may indicate that membrane proteins can undergo a transition phase *"molten globule" state* during solubilization. This state is defined for soluble proteins as an intermediate during reversible unfolding which retains comparable structure and a CD spectrum similar to the native state, but shows other evidence (e.g., increased binding of a hydrophobic dye) of a nonnative tertiary structure.

*4.3 Very Concentrated Protein Solutions and NMR Work.* As growth factors and enzymes are so active, one generally works with solutions containing less than 1 μg/ml. But much more concentrated solutions are required for microinjection into cells, for clinical trials of drugs, and for analytical studies of protein structure. There are many references on preparing proteins for X-ray studies.[1,35] As it has only recently been shown to be a general method for protein structure determination,[56] less has been written on preparing proteins for NMR. The major requirement for good spectra is absolutely pure protein at high concentration (1-20 mM).

Most structural determination by $^1$H NMR use solutions in $D_2O$ and $H_2O$ at acid pH. Acid conditions encourage aggregation and protein unfolding, which shortens sample life. Solvent protons can significantly obscure regions of interest in the protein spectrum ($C_\alpha$ protons), so buffers are usually phosphate or deuterated Tris. Some groups prefer to work without ionic stabilizers, as they can blur peak profiles and cause excessive heat-up of the sample during measurements. These stringent requirements obvious-ly limit the proteins that can be studied by the technique to small, stable ones.

Assuming the solubility requirements are met, structure analysis for up to 80 amino acid proteins is almost routine.[57] The recent descriptions of well resolved (but very complex) 2-D NOESY and COSY spectra for urokinase (54 kDa; solution was 1.5 mM in $D_2O$ at pH 4.5),[58] as well as the interaction of pepsin (35 kDa) with its $^{15}$N-labeled inhibitor[59] show that investigation of even larger proteins is possible. Isotope-edited NMR spectroscopy, which selectively detects protons bound to labeled ($^{15}$N,$^{13}$C) nuclei, allows larger proteins to be analyzed and widens the choice of noninterfering buffers.[33,55-61]

The solution should be stable during the measurement, which for target proteins means addition of some salt. Staphylococcal nuclease was solubilized in 0.3 M NaCl at pH 7.6, ovomucoid domains (55 amino acids) were soluble to 12-15 mM in 0.2 M KCl at pH 8,[62] and yeast phosphoglycerate kinase substrate binding was studied in 0.1 M Na $^2$H-acetate buffer at pH 7.1 (unspecified enzyme concentration).[63] Narrowest line widths were obtained for a solution of thrombin (35 kDa) concentrated to 0.5 mM in 0.2 M KCl at neutral pH. Significant line broadening [population heterogeneity] was seen if the protein concentration was increased or at lower salt concentrations at the same pH (Gerhard Wagner, personal communication).

## 5. Protein Engineering to Increase Solubility

*5.1 Amino Acid Solubility and Water Affinity.* Individual amino acids vary greatly in solubility and affinity for water (Table 2). Protein solubility is based on the ability of soluble, polar residues to interact with water in such a way that the rest of the protein can maintain an active structure. According to the "hydrophobic collapse" model of protein folding, the driving force for folding is hydrophobic amino acid clustering to avoid water, with the eventual secondary and tertiary structure further stabilized by hydrogen bonding and electrostatic interactions.[8,9] The distri-bution of polarity toward the surface is so typical that it has been used as a criterion for protein design.[64]

The data in Table 2 show that the tendency of residues to be "buried" in a protein (definitions range from less than 5% of the residue surface exposed to solvent[65] to up to 30%)[9] agrees with these generalizations. Most positively charged and amide side chain residues (His, Lys, Arg, Gln, Asn) were on the surfaces of the proteins studied, and the interiors were primarily composed of the aliphatics Gly, Ala, Ile, Leu, Val and the aromatic Phe. But only 23% of the Trp residues and 13% of the Tyr in the structures were nonaccessible to solvent, similar to that of the negative polar residues Glu (20%) and Asp (14.5%). One could argue that the large volume that Trp and Tyr residues occupy makes them difficult to completely bury in a small protein, but more likely the aqueous affinity of the tryptophan imidazole ring and the hydroxyl group of tyrosine were underestimated by early hydrophobicity measure-ments.

*5.2 Peptide Solubility.* There is also a great difference in solubility in secondary structural

elements, as illustrated by peptides designed to adopt one conformation or another. For peptides of more than 8 amino acids, sequences favoring α-helix/random coil structures are more soluble in polar solvents than those forming β-sheet structures. The sum of the *Chou-Fasman coil index* (Table 2) for individual amino acids correlated with the solubility of a series of peptides. The tendency of peptides to form β-sheets could be significantly reduced by the strategic positioning of tertiary peptide bonds (protected residues or prolines) at intervals in the sequence.[40]

The small membrane-interacting protein melittin has a positively charged end, which makes it soluble in water, but the protein spontaneously forms a tetramer through interactions at the hydrophobic end.[26] For other peptides, insertion of arg-$NO_2$ residues, or replacement of hydrophobic residues, improved solubility and lowered aggregation tendencies.[66]

*5.3 Primary Structure Alterations.* Small changes in protein primary structure can have drastic effects on stability and solubility. Replacement of the hydrophobic (-EGNFFGKIIDYIKLMFHHWFG) car-boxy-terminal amino acids of *E. coli* penicillin-binding protein 5 with a shorter hydrophilic sequence (-IRRPAAKLE) made the protein water soluble and allowed crystallization.[67] A 13 residue deletion (EVLNENLLRFFVA) in α-casein makes the molecule more soluble.[4] Note that both of the deleted sequences contained FF. *Phenylalanine residues are likely to self-interact and are frequently found at subunit interfaces.*[68]

A series of point mutations altered the stability and solubility of insulin without significantly affecting the biological activity.[1] In particular, it was possible to replace the *asparagine* at position 21, which *deamidates in acid solution and leads to dimer formation*, with Gly, Ser, Thr, Asp, His, and Arg. Similarly, the tendency of yeast cytochrome *c* to autoreduction and dimerization was eliminated by substituting a Thr for Cys-107.[69] A hybrid interferon-α protein precipitated at low salt, unlike either of the parent molecules.[70]

The fragility of protein structure is the major limiting factor on the industrial use of enzymes. Thus the question of what makes proteins stable at high temperatures and in organic solvents and whether the two correlate is not purely academic. Specific sequence changes in proteins from *thermophilic organisms* show a tendency to r*eplace lysine and glutamic acid with arginine and aspartic acid*, and a preference for the hydrophobic amino acids Phe, Val, and Ile over Leu, Ala, and Met.[71] *Most* of these *changes occur in α-helical regions and increase the net hydrophobicity of the residue.*[72] Crambin, a plant toxin that is extremely stable in polar organic solvents, contains no Met and has a higher content of Phe, Val, and Ile residues than the hydrophilic plant toxins to which it is related.[93]

*5.4 Post-Isolation Alterations.* One can alter the solubility of isolated proteins *in vitro* by coupling to polyethylene glycol. Such modifications have been shown to significantly increase the activity, aqueous solubility, and *in vivo* half-life of interleukin-2.[76] Native lipase M from *Candida rugosa*, which acts on nonpolar substrates, is soluble up to 12% in water but insoluble in benzene. PEG 5000-lipase dissolved rapidly and was active in benzene, toluene, chloroform, and trichloroethane.[77]

*5.5 Designer Proteins.* As site-directed mutagenesis is relatively straightforward for recombinant proteins, one might simply replace surface hydrophobic amino acids with acidic residues when aggregation problems arise. But which residues are at the surface and what will the changes do to the tertiary structure and the enzymatic activity? Obviously, the problem of designing soluble proteins is greatly dependent on the ability to predict protein structure.

The *Chou-Fasman rules*, like most programs used to predict secondary structure from primary sequence data, are based on the study of known structures and the pattern of amino acid usage discerned from them.[78] The learning capabilities of *neural networks* may be the basis for the next generation of predictive programs.[79] Although the determined "code" can predict where α-helices are likely to occur, β-sheets and turns are less easy to locate. Faster computing techniques have allowed the development of local energy minimization of conformations to predict stable structures,[80] but the problem of dealing with solvent energies remains. Further, many intermediate secondary structures disappear before the native state is reached,[81] and no program in use today correctly predicts tertiary structures. Thus, [directed] mutation is still [somewhat] guesswork.

Mutation of proteins like T4 lysozyme[69] or RNase A (unpublished) may aid in structure-based stability design. For example, conversion of a single Thr residue near the carboxy terminus of T4 lysozyme to Ile, Gln, Ser, Arg, or His lowered the stability of the molecule compared to the wild type.[69] Such mutations are rather easy to produce but time-consuming to characterize; even selective mutagenesis may deplete the graduate student supply long before all the possibilities are exhausted. It may be easier to design a soluble protein from scratch, and make "designer proteins" from designer genes.[82] The potential usefulness of this approach was recently demonstrated by the production in *E. coli* of an α-helical protein designed from "first principles." The tetramer was soluble in the bacteria and seems to be both α-helical (by its CD spectrum) and very stable

(-22 kcal/mol). Betabellin, a predominantly β-sheet engineered protein, which is being made synthetically, may also be coming into solution.

These proteins show that although the folding language is not understood, a primitive but internally consistent translation is available. If this subcode really works, the next molecules should be stable, soluble, and active.

*References*:


1. Markussen, J., Diers, I., Hougaard, P., Langkjaer, L., Norris, K., Snel, L., Sorensen, A. R., Sorensen, E., and Voigt, H. O. 1988. Soluble, prolonged-acting insulin derivatives. 111. Degree of protraction, crystallizability, and chemical stability of insulins substituted in positions A21, B13, B23, B27 and B30. Protein Eng. 2:157-166.
2. Dinnbier, U., Limpinsel, E., Schmid, R., and Bakker, E. P. 1988. Transient accumulation of potassium glutamate and its replacement by trehalose during adaptation of growing cells of *Escherichia coli* K-12 to elevated sodium chloride concentrations. Arch. Microbiol. 150:348-357.
3. Mitchell, R. D., Simmerman, H. K. B., and Jones, L. R. 1988. $Ca^{2+}$-binding effects on protein conformation and protein interactions of


**Table 2** - The Aqueous Solubility and Affinity of the Amino Acids, Their Relative Tendency to Exist in a Coil Conformation ($P_c$), and Accessibility to Solvent in Protein Crystal Structures (Percent Buried). The Calculations in the Last Two Columns are Based on Table II of Ref. 65. The Amino Acid Names are Followed by the One Letter Codes in Parentheses.

| Amino Acid | Solubility[a] | F[b] | W | $P_c$[c] | % Buried[d] | VR($A_3$)[e] |
|---|---|---|---|---|---|---|
| i. Aliphatics | | | | | | |
| glycine (G) | 25.0 | 0 | 2.39 | 1.5 | 37% (10) | 66 |
| alanine (A) | 16.7 | 0.91 | 1.94 | 0.7 | 38% (12) | 92 |
| isoleucine (I) | 4.1 | 1.8 | 2.15 | 0.66 | 65% (12) | 169 |
| leucine (L) | 2.4 | 1.7 | 2.28 | 0.68 | 41% (10) | 168 |
| valine (V) | 8.9 | 1.22 | 1.99 | 0.62 | 56% (15) | 142 |
| ii. Aromatics | | | | | | |
| phenylalanine (F) | 2.97 | 1.79 | -.076 | 0.71 | 48% (5) | 203 |
| tryptophan (W) | 1.14 | 2.25 | -5.88 | 0.75 | 23% (1.5) | 240 |
| tyrosine (Y) | 0.045 | 0.96 | -6.11 | 1.06 | 13% (2.2) | 203 |
| iii. Hydroxy/sulfur | | | | | | |
| serine (S) | 5.0 | -0.004 | -5.06 | 1.82 | 24% (8) | 99 |
| threonine (T) | s | 0.26 | -4.88 | 1.07 | 25% (5.5) | 122 |
| methionine (M) | 3.4 | 1.23 | -1.48 | 0.58 | 50% (2) | 171 |
| cystine | 0.01 | | | | | |
| cysteine (C) | s | 1.54 | -1.24 | 1.18 | 47% (8) | 106 |
| iv. Proline | | | | | | |
| proline (P) | (160) | 0.72 | NA | 1.59 | 24% (3) | 129 |
| hydroxy-l-proline | 36.1 | | | | | |
| v. Charged/Amides | | | | | | |
| aspartic acid (D) | 0.5 | -0.77 | -10.95 | 1.2 | 14.5% (5) | 125 |
| glutamic acid (E) | 0.86 | -0.64 | -10.20 | 0.83 | 20% (2) | 155 |
| asparagine (N) | 3.1 | -0.6 | -9.68 | 1.35 | 10% (2) | 135 |
| glutamine (Q) | 3.6 | -0.22 | -9.98 | 0.86 | 6.3% (2.2) | 161 |
| histidine (H) | 4.2 | 0.13 | -10.27 | 1.06 | 19% (1.2) | 167 |
| lysine (K) | s | -0.99 | -9.52 | 0.98 | 4.2% (0.1) | 171 |
| arginine (R) | 15 | -1.01 | -19.92 | 1.04 | 0 | 225 |

[a]Solubility of the amino acids in g/100 g water at 25°C. Source: *CRC Handbook of Chemistry and Physics*, 68th Edition (1987-88) and *Lang"s Handbook of Chemistry* (12th edition). Sigma l-proline was not soluble at more than 1 g/ml, even at 40°C; s = freely soluble.

[b]Two different scales are shown. F is the hydrophobicity scale of Fauchere *et al.*,[73] which is based on the partition coefficient of the Na-acetyl-amino acid amides in octanol/water relative to glycine. **W** is the hydration potential (water affinity) of the amino acid side chain as calculated from the free energy of transfer of the side chain (e.g., methane for A) from the vapor phase to water (see ref. 74 for details). Note that both of these scales differ from the frequently used Nozaki and Tanford[75] scale which assigns values only to residues considered hydrophobic (A:0.5; I:1.8; L:1.8; V:1.22; M:1.3; C:0.5; F:2.5; Y:2.3; W:3.4; all other amino acids: 0).

[c]Coil conformation parameter based on Chou-Fasman data.[40] The parameter is based on the frequency with which a residue is present in a coil relative to its overall occurrence in the 29 proteins studied.

[d]This column represents the tendency of an amino acid to be buried (less than 5% of residues available to solvent) in the interior of a protein, and is based on the structures of 9 proteins [total of about 2000 individual residues studied, with 587 of these (29%) buried]. The first number indicates how often each amino acid was found buried, relative to the number of residues of this amino acid found in the proteins. The number in parentheses indicates the number of buried residues of this amino acid found relative to all buried residues. For other calculation methods with similar results, refs. 9 and 74a.

[e]Average volume of buried residues, calculated from the surface area of the side chain (refs. 29, 64).

canine cardiac calsequestrin. J. Biol. Chem. 263:11376-11381.
4. Farrell, H. M., Kumosinski, T. F., Pulaski, P., and Thompson, M. P. 1988. Calcium-induced associations of the caseins: a thermodynamic linkage approach to precipitation and resolubilization. Arch. Biochem. Biophys. 265:146-158.
5. Robinson, J. J. 1988. Roles for $Ca^{2+}$, $Mg^{2+}$ and NaCl in modulating the self-association reaction of hyalin, a major protein component of the sea-urchin extraembryonic hyaline layer. Biochem. J. 256:225-228.
6. Mahoney, R., Wilder, T., and Chan, B. S. 1988. Substrate-induced thermal stabilization of lactase (*E. coli*) in milk. Ann. N.Y. Acad. Sci. 542:274-278.
7. McCloskey, M. and Poo, M. 1984. Protein diffusion in cell membranes: some biological implications. Int. Rev. Cytol. 87: 19-81.
8. Wright, P. E., Dyson, H. J., and Lerner, R. A. 1988. Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. Biochemistry 27:7167-7175.
9. Rose, C. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. 1985. Hydrophobicity of amino acid residues in globular protein. Science 229:834-838.
10. Privalov, P. L. 1979. Stability of proteins, small globular proteins. Adv. Protein Chem. 33:167-241.
11. Jaenicke, R. 1988. Stability and self organization of proteins. Naturwissenschaften 75:604-610.
12. Arakawa, T. and Timasheff, S. N. 1985. Theory of protein solubility. Methods Enzymol. 114:49-77.
13. Kinsella, J. E 1984. Milk proteins: Physicochemical and functional properties. CRC Crit. Rev. Food Sci. Nut. 21:197-262.
14. Hjelmeland, L. M. and Chrambach, A. 1984. Solubilization of functional membrane proteins. Methods Enzymol. 104:305-318.
15. Gekko, K. and Timasheff, S. 1981. Thermodynamic and kinetic examination of protein stabilization by glycerol. Biochemistry 20:4677-4686.
16. Huot, J. Y. and Jolicoeur, C. 1985. Hydrophobic effects in ionic hydration and interactions, P. 417-471. In: *The Chemical Physics of Solvation.* (Dogonadze, R. R. et al., Eds.). Elsevier Science Publications, Amsterdam and New York.
17. Horbett, T. A. and Brash, J. L. 1987. Proteins at interfaces: Current issues and future prospects. In: *Proteins at Interfaces: Physiochemical and Biochemical Studies.* (Brash, J. L. and Horbett, T. A., Eds.) Am. Chem. Soc., Washington, D.C.
18. Mann, D. F. and Shah, K., Stein, D., and Snead, G. A. 1984. Protein hydrophobicity and stability support the thermodynamic theory of protein degradation. Biochim. Biophys. Acta 788:17-22.
19. van den Oetelaar, P. J. M., de Man, B. M., and Hoenders, H. J. 1989. Protein folding and aggregation studies by isoelectric focusing across a urea gradient and isoelectric focusing in two dimensions. Biochim. Biophys. Acta 995:82-90.
20. Ries-Kautt, M. M. and Ducruix, A. F. 1989. Relative effectiveness of various ions on the solubility and crystal growth of lysozyme. J. Biol. Chem. 264:745-748.
21. Goto, Y. and Fink, A. L. 1989. Conformational states of β-lactamase: molten globule states at acidic and alkaline pH with high salt. Biochemistry 38:945-952.
22. Zimmerman, S. B. and Trach, S. O. 1988. Effects of macromolecular crowding on the association of *E. coli* ribosomal particles. Nucleic Acids Res. 16:6309-6326.
23. Ingham, K. C. 1984. Protein precipitation with polyethylene glycol. Meth. Enzymol. 104:351-356.
24. Hanada, K., Yamato, I., and Anraku, Y. 1988. Solubilization and reconstitution of proline carrier in *Eschericia coli*; quantitative analysis and optimal conditions. Biochim. Biophys. Acta 939:282-288.
25. Brenner, S. L. Zlotnick, A., and Griffith, J. D. 1988. RecA protein self-assembly. Multiple discrete aggregation states. J. Mol. Biol. 204:959-972.
26. Schwarz, G. and Beschiaschvili, G. 1988. Kinetics of melittin self-association in aqueous solution. Biochemistry 27:7826-7831.
27. Zimmerle, C. T. and Frieden, C. 1988. Effect of pH on the mechanism of actin polymerization. Biochemistry 27:7766-7772.
28. Yang, D. S. C., Sax M., Chakrabartty, A., and Hew, C. L. 1988. Crystal structure of an antifreeze polypeptide and its mechanistic implications. Nature 333:232-237.
29. Richards, F. M. 1977. Areas, volumes, packing, and protein structure. Annu. Rev. Biophys. Bioeng. 6:151-176.
30. Eisenberg, D., Wilcox, W., and McLachlan, A. D. 1986. Hydrophobicity and amphiphilicity in protein structure. J. Cell. Biochem. 31:11-17.
31. Hageman, M. J. 1988. The role of moisture in protein stability. Drug Dev. and Ind. Pharm. 14:2047-2070.
32. Good, N. E. and Izawa, S. 1972. Hydrogen ion buffers. Methods Enzmol. 24:53-68.
33. Blanchard, J. S. 1984. Buffers for enzymes. Methods Enzymol. 104:404-414.
34. Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D., and Somero, G. N. 1982. Living with water stress; evolution of osmolyte systems. Science 217:1214-1222.

35. McPherson, A. 1982. *Preparation and Analysis of Protein Crystals*. John Wiley and Sons., N.Y.
36. Feher, G. and Kam, Z. 1985. Nucleation and growth of protein crystals; general principles and assays. Methods Enzymol. 114:77-111.
37. Kamoun, P. P. 1988. Denaturation of globular proteins by urea: breakdown of hydrogen or hydrophobic bonds? TIBS 15:424-425.
38. Schellekens, H., de Reus, A., Boluis, R., Fountoulakis, M. Schein, C., Easodi, J., Nagata, S., and Weissmann, C. 1981. Comparative antiviral efficiency of leukocyte and bacterially produced human α-interferon in rhesus monkeys. Nature 292:775-776.
39. Arnold, F. H. 1988. Protein design for nonaqueous solvents. Protein Eng. 2:21-25.
40. Narita, M., Ishikawa, K., Chen, J. Y., and Kim, Y. 1984. Prediction and improvement of protected peptide solubility in organic solvents. Int. J. Peptide Protein Res. 24:580-587.
41. Janicke, R. and Rudolph, R. 1989. Folding proteins, P. 191-223. In: *Protein Structure: a Practical Approach.* (T. E. Creighton, Ed.). Oxford University Press, U.K.
41a. Rudolph, R. and Fisher, S. 1987. Verfahrung zur Renaturierung van Proteinen. Eur. Patent Appl. 0241-022.
42. Weir, M. P. and Sparks, J. 1987. Purification and renaturation of recombinant human interleukin-2. Biochem. J. 245:85-91.
43. Marston, F. A. O. 1986. The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. Biochem. J. 240:1-12.
44. Hochuli, E. Bannarth, W., Dobeli, H., Gentz, R., and Stuber, D. 1988. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Bio/Technology 6:1321-1325.
45. Smith, D. C. and Hider, R. C. 1988. Thiol exchange catalyzed refolding of small proteins utilizing solid-phase supports. Biophys. Chem. 31:21-28.
46. Light, A., Duda, C. T., Odorzynski, T. W., and Moore, W. G. I. 1986. Refolding of serine proteinases. J. Cell. Biochem. 31:19-26.
47. van Kimmenade, A., Bond, M. W., Schumacher, J. H., Laquoi, C., and Kastelein, R. A. 1988. Expression, renaturation and purification of interleukin-4 from *Escherichia coli*. Eur. J. Biochem. 173:109-117.
48. Kuhlbrandt, W. 1988. Three-dimensional crystallization of membrane proteins. Quart. Rev. Biophys. 21:429-477.
49. Wallace, B. A., Cascio, M., and Mielke, D. L. 1986. Evaluation of methods for the prediction of membrane protein secondary structures. Proc. Natl. Acad. Sci. U.S.A. 81:9423-9427.
50. Furth, A. J., Bolton, H., Potter, J., and Priddle, J. D. 1984; Separating detergent from proteins. Methods Enzymol. 104:318-328.
51. Lee, K. H., Fitton, J. E., and Wuthrich, K. 1987. Nuclear magnetic resonance investigation of the conformation of δ-haemolysin bound to dodecylphosphocholine micelles. Biochim. Biophys. Acta 911:144-153.
52. Maloney, P. C. and Ambudkar, S. V. 1989. Functional reconstitution of prokaryotic and eukaryotic membrane proteins. Arch. Biochem. Biophys. 269:1-10.
53. Welte, W. and Wacker T. 1989. Protein-detergent micellar solutions for the crystallization of a membrane protein. Some general approaches and experiences with the crystallization of pigment-protein complexes from purple bacteria. In: *Membrane Protein Crystallization.* (Michel, H., Ed.). CRC Press, Inc., Boca Raton, FL.
54. Yadeau, J. T. and Blobel, G. 1989. Solubilization and characterization of yeast signal peptidase. J. Biol. Chem. 264:2928-2934.
55. Fargin, A., Faye, J. C., le Maire, M., Bayard, F., Potier, M., and Beauregard, C. 1988. Solubilization of a tamoxifen-binding protein. Biochem. J. 256:229-236.
56. Kline, A. D., Braun, W., and Wuthrich, K. 1988. Determination of the complete three-dimensional structure of the α-amylase inhibitor tendamistat in aqueous solution by nuclear magnetic resonance and distance geometry. J. Mol. Biol. 204:675-724.
57. Montelione, G. T., Wuthrich, K., Nice, E. C., Burgess, A. W., and Scheraga, H. A. 1987. Solution structure of murine epidermal growth factor determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry. Proc. Natl. Acad. Sci. U.S.A. 84:5226-5230.
58. Oswald, R. E., Bogusky, M. J., Bamberger, M., Smith, R. A. G., and Dobson, C. M. 1989. Dynamics of the multidomain fibrinolytic protein urokinase from two-dimensional NMR. Nature 337:579-582.
59. Fesik, S. W. 1988. Isotope-edited NMR spectroscopy. Nature 332:865-866.
60. Oh, B. H., Westler, W. M., Darba, P., and Markley, J. L. 1988. Protein carbon-13 spin systems by a single two-dimension nuclear magnetic resonance experiment. Science 240:908-911.
61. Senn, H., Eugster, A., Otting G., Suter, F. and Wuthrich, K. 1987. $^{15}N$-labeled P22 c2 repressor for nuclear magnetic resonance studies of protein-DNA interactions. Eur. J. Biophy. 14:301-306.

62. Markley, J. L. 1987. One- and two-dimensional NMR spectroscopic investigation of the consequences of amino acid replacements in proteins, P. 15-33. In: *Protein Engineering* (Oxender, D. L. and Fox, C. F., Eds.), Alan R. Liss, Inc., New York.
63. Fairbrother, W. J., Hall, L., Littlechild, J. A., Walker, P. A., Watson, H. C., and Williams, R. J. P. 1988. Probing the 3-phosphoglycerate-binding site of yeast phosphoglycerate kinase using site-specific mutants and $^1$H nuclear magnetic resonance spectroscopy. Biochem. Soc. Proc. 16:724-725.
64. Baumann; G., Frommel, C., and Sander, C. 1989. Polarity as a criterion in protein design. Protein Eng. 2:239-334.
65. Richards. F. M. 1986. Protein design: are we ready. UCLA Symp. Mol. Cell. Biol. 39:171-196.
66. Toniolo, C., Bonora, G. M., Moretto, V., and Bodanszky, M. 1985. Self-association and solubility of peptides. Int. J. Peptide Protein Res. 25:425-430.
67. Ferreira, L. C. S., Schwarz, U., Keck, W., Charlier, P., Dideberg, O., and Ghuysen, J. M. 1988. Properties and crystallization of a genetically engineered, water-soluble derivative of penicillin-binding protein 5 of *Escherichia coli* K12. Eur. J. Biochem. 171:11-16.
68. Argos, P. 1988. An investigation of protein subunit and domain interfaces. Protein Eng. 2:101-113.
69. Shaw, W. V. 1987. Protein engineering. The design, synthesis and characterization of factitious proteins. Biochem. J. 246:1-17.
70. Le, H. V., Syto, R., Schwartz, J., Nagabhushan, T. L, and Trotta, P. P. 1988. Purification and properties of a novel recombinant human hybrid interferon, δ4 α2 / α1. Biochim. Biophys. Acta 957:143-151.
71. Zuber, H. 1988. Temperature adaptation of lactate dehydrogenase. Structural, functional and genetic aspects. Biophys. Chem. 29:171-179.
72. Menendez-Arias, L. and Argos, P. 1989. Engineering protein thermal stability. Sequence statistics point to residue substitutions in α-helices. J. Mol. Biol. 206:397-406.
73. Fauchere, J., Charton, M., Kier, L. B., Verloop, A., and Pliska, V. 1988. Amino acid side chain parameters for correlation studies in biology and pharmacology. Int. J. Peptide Protein Res. 32:269-278.
74. Wolfenden, R., Andersson, L., Cullis, P. M., and Southgate, C. C. 198l. Affinities of amino acid side chains for solvent water. Biochemistry 20:849-855.
74a. Janin, J. 1979. Surface and inside volumes in globular proteins. Nature 277:491-492.
75. Nozaki, Y. and Tanford, C. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. J. Biol. Chem. 246:2221-2227.
76. Knauf, M. J., Bell, D. P., Hirtzer, P., Luo, Z. P., Young, J. D., and Katre, N. V. 1988. Relationship of effective molecular size to systematic clearance in rats of recombinant interleukin-2 chemically modified with water soluble polymers. J. Biol. Chem. 263:15064-15070.
77. Baillargeon, M. W. and Sonnet, P. E 1988. Lipase modified for solubility in organic solvents. Ann. N.Y. Acad. Sci. 542:244-49.
78. Kikuchi, T., Nemethy, G., and Sheraga, H. A. 1988. Prediction of probable pathways of folding in globular proteins. J. Prot. Chem. 7:491-507.
79. Holley, L. H. and Karplus, M. 1989. Protein secondary structure prediction with a neural network. Proc. Natl. Acad. Sci. U.S.A. 86:152-156.
80. Karplus, M. 1987. The prediction and analysis of mutant structures, P. 55-44. In: *Protein Engineering* (Oxender, D. L. and Fox, C. F., Eds.), Alan R. Liss, Inc. New York.
81. Creighton, T. E. 1988. On the relevance of non-random polypeptide conformations for protein folding. Biophys. Chem. 31:155-162.
82. DeGrado, W. F., Wasserman, Z. R., and Lear, J. D. 1989. Protein design, a minimalist approach. Science 243:622-628.
83. Kim, P. S. 1988. Passing the first milestone in protein design. Protein Eng. 2:249-250.
84. Schein, C. 1989. Production of soluble recombinant proteins in bacteria. Bio/Technology 7:1141-1149.