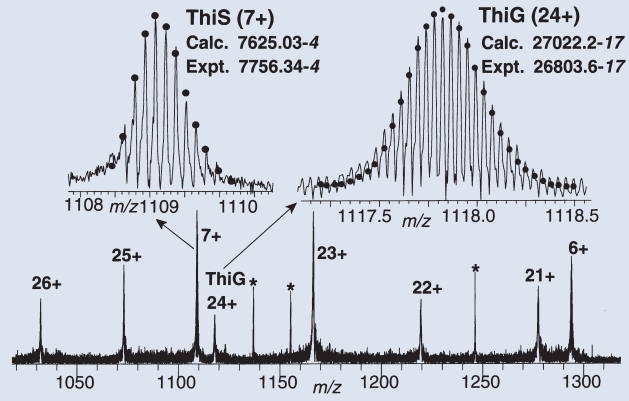
Beyond the Expected: Proteins in Complex Mixtures

Proteomics requires the identification of proteins in evermore complex mixtures. The most popular of the "top-down" methods that use MS for structural analysis relies heavily on the power of comparison. Proteins are identified in MS from their breakdown products, which are matched with calculated results derived from DNA-predicted amino acid sequences in gene databases. However, the match is not always a good one. Original errors in DNA sequencing incorporated in the database can give anomalous calculated results, as can any number of post-translational modifications that influence the sizes or charges of breakdown products.

Fred W. McLafferty and colleagues at Cornell University (Ithaca, NY) recently reported using electron capture dissociation (ECD) to enhance the ability of an electrospray ionization Fourier transform MS (ESI-FTMS) system to analyze complex mixtures from *E. coli* cell extracts (*J. Am. Chem. Soc.* **2002,** *124,* 672–678). They were looking for specific proteins involved in three separate systems: the biosynthesis of thiamine (GoxB, ThiS, ThiG, and ThiF); Coenzyme A biosynthesis (CoaBC); and the hydroxylation of proline residues in proteins (prolyl 4-hydroxylase). When the researchers used ESI-FTMS, none of the enzymes involved in thiamine biosynthesis were found to correspond to the 102 accurate molecular weights that could be determined. MS/MS analysis of one ion species, however, identified it as ThiS on the basis of predictions from DNA sequences (despite the discrepancies in the molecular weight measurements). Subsequent ECD analysis of the ion species demonstrated that it was actually a mixture of ThiS and ThiG, with two modifications: the addition of an N-terminal methionine on the 8-kDa ThiS and the removal of an N-terminal methionine and serine from the 27-kDa ThiG.



Reducing complexity. A broadband spectrum of the purified ThiS/ThiG mixture. (Adapted from McLafferty, F.W.; et al. (*J. Am. Chem. Soc.* 2002, *124,* 672–678.)

Analysis of the Coenzyme A system showed similar results. The enzyme was not found where predicted by the DNA sequence information; it was only located after ECD demonstrated that the protein's mass was 131 Da lower than predicted due to the removal of the N-terminal methionine. Analysis of the proline hydroxylation system demonstrated further divergence from predictions, which resulted from the production of multiple ions due to noncovalent phosphate adducts. As well, a mass discrepancy of 135 Da was found, which was the result of removal of the initiation methionine (131 Da) and the formation of disulfide bonds at two locations (2 × 2 Da). The authors suggest that without ECD, correct "top-down" FTMS in these cases would have been difficult if not impossible.

Protein Blots, India Ink, and MS

Proteomics has traditionally relied on 2-D electrophoresis gels, in-gel proteolytic digestion of the protein spots, and peptide analysis using MS. Silver staining is a sensitive method for visualizing the protein spots, but recent studies indicate that it may not be the optimal choice because it might chemically modify the proteins and result in poor yields of some peptides. Silver staining also suffers from being incompatible with the electroelution of proteins from a gel onto membrane supports such as nitrocellulose.

India ink, however, offers an alternative. It has been used extensively to stain membrane-immobilized proteins and has a detection limit in the low- to mid-nanogram level, putting it on a par with silver. Because it is based on noncovalent interactions between proteins and colloidal carbon, it should not modify the proteins and thereby interfere with MS analysis. But Klaus Klarskov and Stephen Naylor, then at the Mayo Clinic (Rochester, MN), wanted to determine whether this was true (*Rapid Commun. Mass Spectrom.* **2002,** *16,* 35–42).

Klarskov and Naylor, now at the University of Sherbrooke (Quebec) and Beyond Genomics (Waltham, MA), respectively, compared the MALDI-TOF mass spectra of proteins blotted onto nitrocellulose, stained with India ink, and proteolytized with those of proteins stained in-gel with silver or Coomassie blue and digested in-gel. They found that the India ink-treated proteins yielded the highest number of peptides. The researchers reasoned that more complete proteolytic digestion might have taken place on the membrane than in the gel.

The researchers then tried the system in a "real world" setting, looking for proteins covalently modified during treatment with the nonsteroidal anti-inflammatory drug zomepirac. Serum proteins were isolated from a rat that had been treated with the drug, separated on a 2-D gel, blotted, and stained with India ink. The blot was then probed with antibodies directed against the drug, a process that highlighted three modified protein spots. These were excised, digested, and analyzed by MS.

When the resulting peptide sequences were com-

pared with those in the National Center for Biotechnology Information's protein database using the search algorithm Protein Prospector (*http://prospector.ucsf.edu*), the peptides mapped to the α-, β-, and γ-chains of fibrinogen. The specific reasons why zomepirac modifies these proteins were unclear to the researchers.

Crystallization under Pressure

As a downstream step in large-scale protein purification, crystallization offers the advantages of high purity and good appearance. But the general use of crystallization has been hindered by the lack of methods for determining the thermodynamic and kinetic properties of crystal formation.

Previous experiments suggested that pressure might exert as much influence on crystallization as temperature, supersaturation, and solvent ionic strength. It is not known, however, whether its influence is through the direct compression of the amino acid residues as they pack or through alterations in the hydration state of the residues. Further complicating the issue is the fact that pressure effects vary from protein to protein, so that one protein will crystallize faster under high pressure than another.

Charles Glatz and his colleagues at Iowa State University (Ames) studied the effects of pressure on the nucleation rate of Properase, a variant of subtilisin (*Cryst. Growth Des.* **2002,** *2,* 45–50). By examining the reaction and activation volumes and crystal structures of the protein, they hoped to better understand the roles of surface properties and hydration in crystal nucleation.

Under increasing pressure, Properase crystals nucleated less frequently and the protein's overall solubility increased. The change in protein solubility was used to calculate the change in the reaction volume (ΔV), which is also the sum of the changes in the protein's intrinsic volume (V_M; essentially its van der Waals volume), the thermal volume (V_T; due to molecular vibration), and the interaction volume (V_I; effectively the volume of the hydration shell).

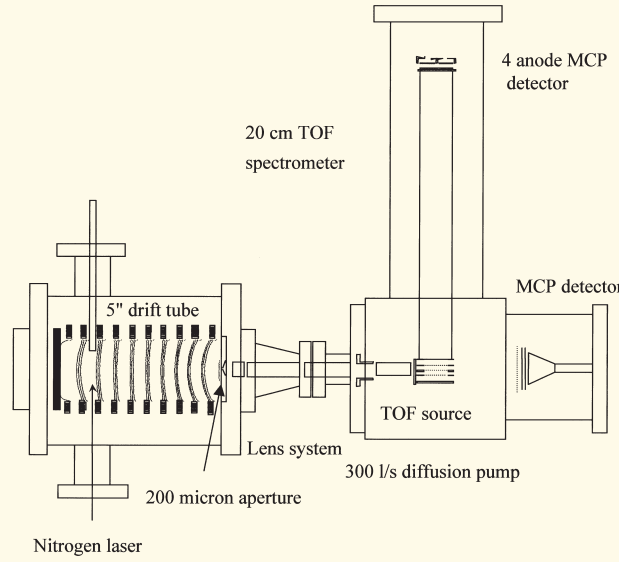
The researchers assumed that ΔV_M was zero and used the protein's crystal structure to calculate the ΔV_T. This allowed them to calculate ΔV_I. From this value for ΔV_I, they inferred the role of solvent water molecules, which would have to be removed from the interface between protein molecules during crystal nucleation.

V_I increased with increasing pressure, which Glatz and his colleagues argued was due to the burial of previously solvent-exposed polar or charged groups and the loss of water molecules from the original hydration shell. Similar pressure-induced changes in the hydration of the crystallization transition state (activation volume) also affected crystal nucleation.

By no means a finished project, this experiment is another step in understanding the physicochemical behavior of crystal nucleation.

Ion Mobility MS and Peptide Interactions

MS has proven to be an essential tool for the identification of isolated proteins, but in cells, proteins interact with other macromolecules that are present. Thus, it was critical that MS technologies be developed for the detection and analysis of noncovalent complexes. ESI-MS has typically been the method of choice for this purpose, but peptide–peptide interactions are often studied in the presence of lipids, detergents, or salts that suppress ESI. In such cases, MALDI appears to be the more suitable ionization choice.



An eye on mobility. A schematic of the MALDI-IM-o-TOF apparatus developed by Kent Gillig and colleagues (*Anal. Chem.* 2000, *72,* 3965–3971).

Amina Woods and her colleagues had previously shown that MALDI-MS could be used to observe noncovalent complexes that involved a salt bridge between acidic and basic peptides and that this association was strongly pH-dependent (*J. Am. Soc. Mass Spectrom.* **2001,** *12,* 88–96). Recently, however, Woods extended this analysis by incorporating ion mobility (IM) into the process (*J. Am. Soc. Mass Spectrom.* **2002,** *13,* 166–169).

Desorption from MALDI into the high-pressure IM drift cell creates a different set of ionization conditions from standard high-vacuum MALDI, and recent evidence suggests that the buffer gas might stabilize ions and noncovalent complexes. Woods' group studied this possibility by looking for evidence of complex formation between the acidic Mini Gastrin I peptide and the basic dynorphin 1–7 peptide.

The researchers found a near-linear relationship between *m*/*z* and mobility drift time that allowed them to distinguish signals from each of the peptides, the dynorphin/Mini Gastrin complex, and a Mini Gastrin homodimer, the latter of which was not observed using ESI. The complex was, as expected, pH-dependent. Interestingly, the complex also showed a small degree of fragmentation, which was indicated by a peak that had the mobility of the complex but the *m*/*z* of dynorphin. Future research will focus on applying the technology to biological samples to look for biologically relevant complexes.

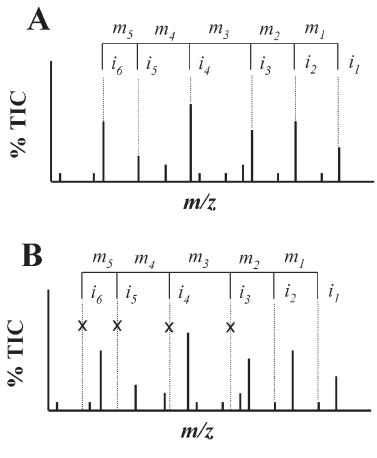
Identifying Peptides with SALSA

Algorithms that identify peptides from MS data typically work by comparing actual MS data to predicted data based on known sequence information. But this sequence-based mechanism can hamper the accuracy of these programs, because proteins can be altered by post-translational modifications, gene mutations, and splice variants that can't always be predicted from their sequences. If a peptide is modified, its relative peak position or fragmentation pattern might be altered. To accommodate this problem, Daniel Liebler and his colleagues at the University of Arizona (Tucson) developed SALSA (*S*coring *AL*gorithm for *S*pectral *A*nalysis; *Anal. Chem.* **2002,** *74,* 203–210).

In SALSA, a predicted ion series is defined as a group of ions (i_1, i_2, i_3,…i_n) separated by specific m/z values (m_1, m_2, m_3,…m_n) or the masses of the next amino acids along the probe peptide chain. The first predicted ion (i_1) is lined up with the highest m/z in the actual MS/MS spectra, and the algorithm then looks for the next ion (i_2) in the series, connecting the ions sequentially. The user can choose to align the spectra based on either the primary (y- or b-) or secondary (y″- or b″-) ions or a paired combination thereof. SALSA then scores the alignment, taking into account factors such as the search strategy used, the length of the search motif, the number of ions that match the series, and the intensity of the scored ions.

The researchers tested the algorithm first against mass spectra of bovine serum albumin (BSA), comparing the results with those obtained using another algorithm, Sequest. Overall, SALSA detected the same peptides as Sequest but then outperformed the other algorithm by identifying peptides that had undergone modification through processes such as cysteine oxidation or N-terminal carbamylation. This capability arises from the fact that modified peptides still follow the same ion patterns as those described above, with the sole exception that most or all of the series are shifted along the m/z axis by the mass of the modification.

SALSA was then tested against the spectra of a mixture of BSA and human serum albumin (HSA) peptide fragments. Many of these peptides differ by only one or two amino acids. When probed with a BSA peptide sequence, SALSA identified both of the appropriate BSA and HSA peptides, but the latter had a much lower score. Probing with the HSA sequence, however, gave the opposite result, proving SALSA's merit as another tool in the proteomic arsenal.



Peak positioning. An alignment that matches often (A) scores well, while one that misses peaks (B) scores poorly. (Adapted from Liebler, D.; et al. *Anal. Chem.* 2002, *74,* 203–210.)

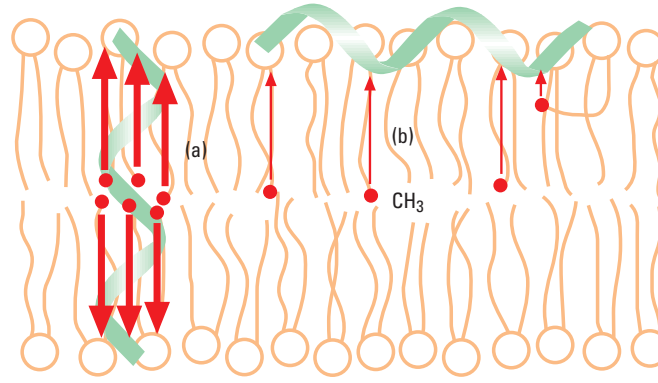
Protein Topology on Membranes

The in situ orientation of a membrane-bound protein offers important information about its function. Unfortunately, the properties that allow the protein to interact with the membrane (e.g., hydrophobicity) can also complicate biophysical analysis. Solution-state NMR experiments on such proteins typically use extrinsic paramagnetic probes and detergent micelles, which mimic the membrane environment, but the tight curvature and monolayer characteristics of these micelles limit their usefulness. And although solid-state NMR can make use of actual lipid bilayers, traditional experiments have required that the proteins be aligned along a single axis—a challenge for large proteins.

Recently, however, researchers explored the use of proton spin diffusion to analyze membrane-bound proteins by solid-state NMR. Unlike previous experiments performed at low temperatures, where the only source of proton magnetization was the surrounding water molecules, the experiments of Mei Hong and colleagues at Iowa State University (Ames) were performed at room temperature and relied on the transfer of magnetization from lipid methyl protons to ^{13}C-labeled protein amino acids (*J. Am. Chem. Soc.* **2002,** *124,* 874–883). The experiments were based on the theory that proton magnetization transferred directly to ^{13}C atoms within the lipid bilayer would give a stronger signal than if the magnetization had to diffuse among protons in the fatty acid chain before transferring to a ^{13}C atom outside of the membrane.

To test their method, the Iowa researchers compared the topology of the channel-forming domain of the protein colicin 1A with that of membrane-bound but surface-oriented DNA. Compared with the DNA ^{13}C chemical shifts, which showed weak magnetization transfer because of their location on the membrane surface, those of the protein residues in the predicted transmembrane segments had strong signals.

This result suggested that these peptides did indeed enter the lipid bilayer. The finding that a membrane-imbedded protein can also facilitate magnetization transfer from lipid methyl protons to ^{31}P in the lipid headgroup in the absence of protein labeling further expands the applicability and reduces the cost of this technology.



The orientation express. Magnetization transfer from lipid protons to protein residues is faster when the protein is imbedded in the membrane (a) than when it sits on the bilayer surface (b). (Adapted from Hong, M.; et al. *J. Am. Chem. Soc.* 2002, *124,* 874–883.)