PLAYING THE PROTEIN POLYPHONY

Orchestrating research on protein structure and function requires the accompaniment of almost every major spectroscopic instrument available.

BY MARK S. LESNEY

It is no longer the harmony of the spheres, but the symphony of the sequences that enraptures science today. Analyzing the structure and function of proteins can be likened to the analysis of a complex piece of music rather than a simple detailing of chemical bonds. The changing rhythms of structure—from amino acid sequences to complex and chaperone-directed three-dimensional folding mediated by sulfur bridges and glycosylation—coupled to the variations on a multitude of themes induced by co-protein- and ligand-binding, cell pH, and phosphorylating signal transducers, all influence the final protein performance. To be listened to, these cellular symphonies must be played on a variety of instruments—not on mandolins and harpsichords, but on MS and NMR. The final scores must be transcribed not by quill pens onto five-lined parchment, but by software scribes into databases, and performed at last by bioinformatics conductors inspired by the twin muses of biology and XML.

Structural motifs

Form is foremost in modern proteomics. Sequence analysis allows the opportunity to speculate on function based on similarities to known proteins. Molecular modeling based on an understanding of amino acid sequence can point to potential ligand binding sites, potential sulfide bridges, membrane attachment regions, and regulatory sites for glycosylation and/or phosphorylation. Since the heyday of X-ray crystallography (see box, "Classics in crystal"), a variety of spectroscopic instruments have been assembled to play out the various levels of protein form, from sequence determination to three-dimensional topologies. These include NMR spectrometry, MS, and Raman spectroscopy.

NMR describes the type and location of paramagnetic nuclei such as ^{1}H, ^{13}C, and ^{15}N in complex molecules, so it is ideal for protein structural analysis. Because surrounding nuclei influence the immediate magnetic environment of a given nucleus, sensitive measurements can pinpoint very subtle differences in neighboring chemical structures. After complex mathematical analysis and modeling, these experimentally determined patterns can be identified by comparing spectral fingerprints against a rapidly growing library of chemical moieties. By alter-

ILLUSTRATION: TONY FERNANDEZ

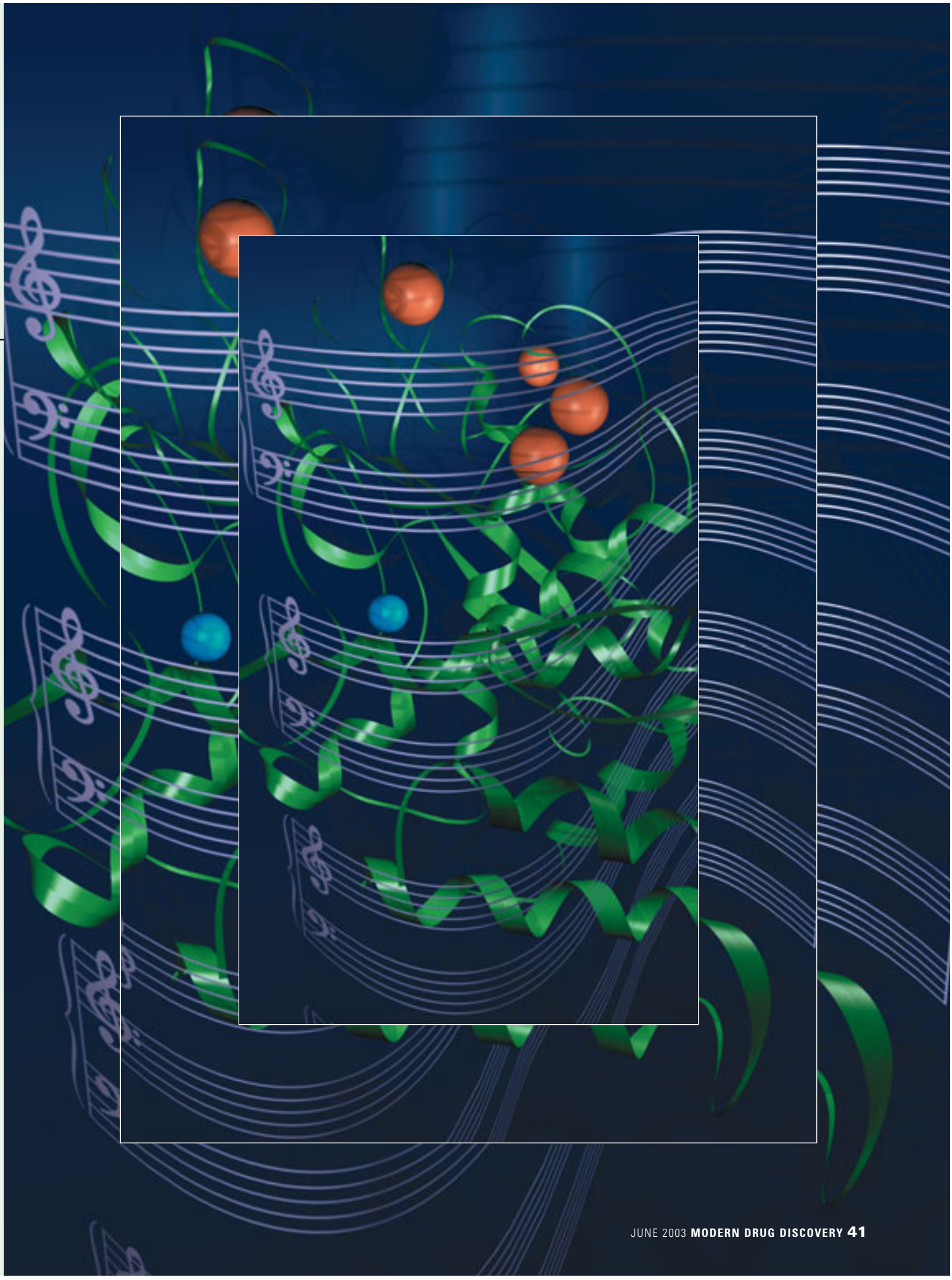
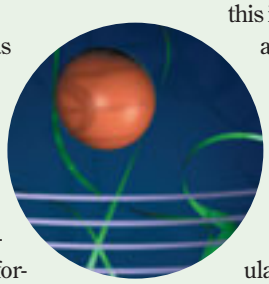


ILLUSTRATION: TONY FERNANDEZ

ing experimental parameters, extremely fine differentials can be found between entities in highly similar, but not identical, magnetic environments. This is especially important in examining molecular interactions such as ligand binding.

Structural analysis of complex biomolecules such as proteins requires the use of multidimensional NMR techniques with isotopic labeling to provide added signal differentiation. Multidimensional NMR provides information on structure by detailing "through bond" spin–spin interactions (which gives information on connected nuclei) and "through space" magnetic-dipole-mediated spin–spin interactions (which gives information on spatially relevant nuclei). This form of analysis provides highly detailed information on the regional environment of the various nuclei that can be mathematically translated into structural information. For example, researchers at the University of Washington solved the solution structure of the SH3 domain (a site of herpesvirus protein binding) of a T-cell-specific tyrosine kinase using multidimensional NMR (*1*).



A WIDER RANGE OF SPECTROSCOPIC INSTRUMENTS AS WELL AS MODIFIED TECHNIQUES ARE REQUIRED TO REVEAL THE FUNCTIONS BEHIND THE FORMS.

Mass spectrometry has achieved broad utility for structural studies of proteins, whether singly or in tandem, MALDI or ESI (see *Today's Chemist at Work*, October 2001, pp 30–35). Much of the current research is focusing on methods of adapting and improving the sensitivities of existing methods for difficult proteins or complex mixtures. For example, proline-rich proteins prove difficult to sequence by MS because of high-efficiency cleavage at the amide bond on the N-terminal of proline residues, resulting in low relative abundance of fragments arising from other amide bonds. Researchers at the Boston University School of Medicine and the Goldman School of Dental Medicine compared collision-induced dissociation (CID) and electron-capture dissociation (ECD) methods for producing mass spectra. They found that ECD spectra of peptides containing more than 30% proline residues are simpler and easier to interpret than CID spectra, but ultimately determined that a complementary approach using both decomposition methods provided the most complete information, yielding >93% sequence coverage for a salivary protein they studied (*2*).

Multidimensional MS (MS^N) is particularly useful for analyzing complex proteins. Unlike multidimensional NMR, MS^N relies on sequential analysis (each representing a new dimension) of a selected fragment peak from the previous MS run. For proteins, this iterative technique is capable not only of providing amino acid sequence information, but also of localizing and providing the three-dimensional structure of glycan side chains. For example, researchers at Thermo Finnigan described a high-throughput system for structural analysis of glycosylated proteins using multiple-tandem ion-trap mass spectrometry up to MS^4 (*3*).

Raman spectroscopy creates a fingerprint of molecular vibrations observed by the measurement of a weak-intensity specific-frequency region of light scattering produced by a laser-probed sample. Raman is particularly useful in the study of proteins; for example, it is easily able to recognize disulfide bonds. Identifying the precise frequency (occurring at 500–550 cm^{-1}) can give the exact conformation of the protein region in which the bridges occur, such that β-pleated sheets can be identified and their gauche versus trans linkages distinguished.

Researchers at Tohoku University in Japan used comparative resonance Raman analysis to examine the structure of the heme-PAS binding domains of several signal-transducing heme proteins (in vivo, these proteins detect and respond to the presence of heme ligands). Because the Raman spectral bands observed were highly sensitive to the coordination structure and spin state of the heme iron, they were able to determine that the amino acids (of those available from the known protein sequence) that most likely coordinated to the heme pocket region were a specific arginine and an FG-loop methionine. The entire heme was organized between five antiparallel β-strands (a β-sheet) and an α-helix (*5*).

Function con moto

It is in their performance, however, that proteins are best understood. Variations in enzyme and ligand-binding kinetics, much less cytological localization, aren't easily played out in crystals and ionized fragments but rather in fluid motion—solutions found in solution. Thus, a wider range of spectroscopic instruments as well

Classics in crystal

Historically, beginning in the glory days of the 20th century in the laboratories of John Kendrew and Max Perutz at Cambridge, X-ray crystallography has been the most powerful method for analyzing protein structure. And although this is rapidly changing, especially with the increasing versatility demonstrated by NMR and MS instruments, this classical technique is by no means off the charts. Protein X-ray crystallography relies on the way a crystal lattice can scatter an X-ray beam in regularly spaced patterns that give information on the underlying structure of the crystal. The addition of marker atoms in the crystals, which greatly scatter X-rays, provides an additional means of elucidating structural information. Modern crystallography even has the capability to study structure–function relationships. Researchers at the University of Toronto recently characterized the active site of the enzyme orotidine 5´-monophosphate decarboxylase by examining the enzyme in the presence of its inhibitor, product, substrate complexes, and during denaturation with guanidine hydrochloride. They were able to detect the stabilizing influence of a specific asparagine residue and the requirement for charge–charge repulsion at the active site as part of the enzyme's normal activity (*4*). In the future, the availability of increasingly intense X-rays from synchrotrons promises even more rapid and detailed structural analysis from this time-honored technique.

as modified techniques are required to reveal the functions behind the forms. In most cases, mixing and matching of multiple instruments leads to the broadest possible analysis.

NMR is particularly useful for more than just simple structural analysis and has proven one of the most valued instruments in the functional orchestra. One common technique is that of NMR titration, which can be used to determine what portions of a complex protein are involved in ligand binding. In a titration experiment, the ligand is added to the protein solution and chemical shifts from the unbound pattern are examined to determine what regions are most interactive. In the case of the tyrosine kinase/herpesvirus study mentioned earlier, titration demonstrated that the most significant shifts occurred in three stretches of the polypeptide chain corresponding to two known loops and a helical turn connecting two β-strands—the components generally known to form the binding surface of proline-rich polypeptides (*1*).

UV–vis, fluorescence, and IR spectroscopies are the classical instruments for the analysis of enzyme kinetics and overall biomolecular interactions. Natural changes in the absorbed wavelengths and/or fluorescence of biomolecules frequently occur during protein–ligand and enzyme–substrate interactions and have proved extraordinarily useful for kinetic and pharmacological analysis. But "natural" changes are not required, as demonstrated by the many cases in which appropriate labeling techniques (either colorimetric or fluorescent tagging of the protein or its ligands) have proved valuable in increasing the sensitivity and the types of information that can be obtained. Unlabeled fluorescent spectroscopy was used to determine the binding affinities of the SH3 domain protein described above with a variety of nonreceptor (signal-transduction) tyrosine kinases as part of a broad-based structure–activity relationship study (*1*).

Table 1

Selected suppliers of protein-analysis instruments

Company	Website	Product area
ABB Bomem	www.bomem.com	IR
Bio-Rad	www.bio-rad.com	MS
Bruker AXS	www.bruker-axs.com	X-ray
Bruker Biospin	www.bruker-biospin.com	NMR
Bruker Daltonics	www.bdal.com	MS
Chromex, Inc.	www.chromexinc.com	Raman
Digilab	www.digilabglobal.com	IR, Raman
Douglas Instruments	www.douglas.co.uk	X-ray
Hampton Research	www.hamptonresearch.com	X-ray
Huber	www.xhuber.com	X-ray
Jasco	www.jascoinc.com	CD, IR, UV–vis, Raman
JEOL	www.jeol.com	MS, NMR
JY Horiba	www.jyhoriba.com	Raman, fluorescence
Micromass	www.micromass.co.uk	MS
Olis	www.olisweb.com	CD, UV–vis, Raman, NIR, fluorescence
Oxford Cryosystems	www.oxfordcryosystems.co.uk	X-ray
Oxford Instruments	www.oxford-instruments.com	X-ray, NMR
PerkinElmer Instruments	www.perkinelmer.com	IR, UV–vis, fluorescence
PANalytical	www.panalytical.com	X-ray
Remspec	www.remspec.com	IR
Rigaku MSC	www.rigakumsc.com	X-ray
Shimadzu Biotech	www.shimadzu-biotech.com	MS
Shimadzu Scientific Instruments	www.shimadzu.com	UV–vis
Stanford Research Systems	www.srsys.com	MS
Thermo Finnigan	www.thermofinnigan.com	MS
Thermo Nicolet	www.thermonicolet.com	IR, Raman
Varian	www.varianinc.com	NMR, MS, UV–vis, IR

Similarly, the IR spectrum has proved useful for protein analysis. Like Raman, IR spectroscopy relies on vibrational analysis for its information. Researchers at the Albert-Ludwigs Universität in Freiburg, Germany, were able to use a combination of UV–vis, fluorescence, and FTIR spectroscopy to examine various conformational changes in the eye pigment rhodopsin induced by pH and salinity changes, giving them insight into physiologically relevant structural shifts (*6*).

CD spectroscopy uses circular dichroism to determine protein structure information, relying on the property of chiral molecules to differentially absorb separated right- and left-handed circularly polarized light. It is especially useful for observing changes during ligand–protein interactions. In the functional analysis of the interaction of the SH3 domain with the herpesvirus protein, far-UV CD spectroscopy demonstrated that a stable polyproline helix of the viral protein formed before binding with the kinase rather than during binding, as had been proposed (*1*). The combination of NMR, CD, and fluorescence spectroscopies described shows the routine benefits of using multiple instruments to obtain a complete analysis.

MS has proved a valued accompanist to functional analysis of protein interactions, especially for drug development. It is easy to imagine that the destructive nature of MS would be counterproductive to binding studies, but not so. Researchers at Wyeth Research (Cambridge, MA, and Pearl River, NY) coupled size exclusion chromatography (SEC), MS, and NMR to screen small

organic molecules for their ability to bind to proteins. The proteins were mixed with a combinatorial library, isolated by SEC, and analyzed with ESI-MS. The appearance of "extraneous" material in the peptide patterns indicated a positive binding event. These samples were subsequently analyzed using NMR (*7*).

Similarly, to study protein-binding interactions in complex cellular systems, researchers at North Carolina State University (Raleigh) coupled SEC to micro-ESI-MS for direct analysis of complexes present in biological mixtures containing the high salt concentrations often needed for complex formation and stability. Their method required minimal samples and easily demonstrated changes in the size of complexes relative to individual components (*8*).

Symphonic denouement

Obviously, no single instrument can play the full symphony of proteomics on its own. Collaborative effort is required, synthesizing information between a variety of spectroscopic devices chosen for the particular venue. This is especially true for gathering the kind of information needed to coordinate protein structure with function, from simple ligand-binding kinetics to the analysis of complex pharmacological interactions and signal transduction mechanisms. Perhaps the most telling need is for better bioinformatics—developing and maintaining in a usable fashion libraries for each of the techniques, recording the broadest possible series of protein structural motifs to act as references—key notes, if you will—to maximize the utility of all of the instruments, perhaps pointing to the far distant day when completely automated analysis of proteins will be possible.

But until then, protein analysis will remain something of an art, and the better the researcher's "ear", the greater his or her ability to gain the maximal performance from the variety of technologies available—so much so that it is unlikely that any of these instruments will soon disappear from the orchestra; rather, they will become more and more integrated for a greater overall performance.

References

(1) Schweimer, K.; et al. *Biochemistry* **2002,** *41,* 5120–5130.
(2) Leymarie, N.; et al. *Anal. Chem.* **2002,** *74,* 4124–4132.
(3) Wu, S.-L.; et al. Structural analysis of glycosylated peptides in complex mixtures with ion trap MSn. ThermoFinnigan Application Report No. 300; www.thermofinnigan.com (registration required for access).
(4) Wu, N.; Gillon, W.; Pai, E. F. *Biochemistry* **2002,** *41,* 4002–4011.
(5) Tomita, T.; et al. *Biochemistry* **2002,** *41,* 4819–4826.
(6) Vogel, R.; Siebert, F. *Biochemistry* **2002,** *41,* 3536–3545.
(7) Moy, F. J.; et al. *Anal. Chem.* **2001,** *73,* 571–581.
(8) Cavanagh, J.; et al. *Biochemistry* **2002,** *41,* 7859–7865.

This article is reprinted from *Today's Chemist at Work,* October 2002, pp 36–41.

Mark S. Lesney is a senior associate editor of *Today's Chemist at Work* and *Modern Drug Discovery*. Send your comments or questions regarding this article to mdd@acs.org or to the Editorial Office address on page 3. ■

KEY TERMS: automation, cell biology, clinical, high throughput, imaging, informatics, modeling, proteomics, screening, technique