Biochemistry 412

Protein Mass Spectrometry

February 20th, 2009
“Proteomics”

The study of the complete complement of proteins found in an organism
“Degrees of Freedom” for Protein Variability

Covalent Modifications in Proteins

- Post-translational modifications (e.g., phosphorylation, glycosylation, etc.)
  - more than 200 such modifications are known, and they can occur at multiple sites in a single protein

- Alternative splicing of a primary transcript
  - in extreme cases, a single gene can produce tens of thousands of different mRNAs!

- Proteolytic processing

- Protein aging

Thus, there are probably many millions of different proteins in our bodies!!
More Reality Therapy re Proteins

• They have “personalities”: each behaves differently

• They exist in different concentrations, ranging over a million-fold

• It will be extremely difficult to even identify them all (see previous slide)

Take-home message:

Proteomics presents challenges that are orders-of-magnitude more difficult than those presented by genomics!
**“Divide and Conquer” Proteomics**

Identification by fractionation, fragmentation, mass analysis, & comparison with database(s) of virtual peptides

---

*Note: “MudPIT” stands for multidimensional protein identification technology.*

There are two major types of “soft ionization” methods of mass spectrometry for use in analyzing samples of biological origin:

- **Matrix-assisted laser desorption**
- **Electrospray**
Matrix-Assisted Laser Desorption Ionization (MALDI) Mass Spectrometry

Mann et al (2001)
*Annu. Rev. Biochem.* 70, 437.

**Figure 1** Schematic of MALDI process and instrument. (A) A sample cocrystallized with the matrix is irradiated by a laser beam, leading to sublimation and ionization of peptides. (B) About 100–500 ns after the laser pulse, a strong acceleration field is switched on (delayed extraction), which imparts a fixed kinetic energy to the ions produced by the MALDI process. These ions travel down a flight tube and are turned around in an ion mirror, or reflector, to correct for initial energy differences. The mass-to-charge ratio is related to the time it takes an ion to reach the detector; the lighter ions arrive first. The ions are detected by a channeltron electron multiplier.
Electrospray Ionization (ESI) Mass Spectrometry

Figure 1 Features of the ESI interface (upper panel) and schematic representation of the ESI process (lower panel)

A high positive potential is applied to the capillary (anode), causing positive ions in solution to drift towards the meniscus. Destabilisation of the meniscus occurs, leading to the formation of a cone and a fine jet emitting droplets with excess positive charge. Gas-phase ions are formed from charged droplets in a series of solvent evaporation–Coulomb hysteresis cycles. With the continual emission of positively charged droplets from the capillary (green), to maintain charge balance, oxidation occurs within the capillary. If the capillary is metal, evaporation of the metal may occur at the liquid/metal interface:

\[ \text{M}^{2+} \rightarrow \text{M}^{3+} \text{(aq)} + 2e^- \quad \text{(in metal)} \]

or alternatively negative ions may be removed from solution by electrochemical oxidation:

\[ 4\text{OH}^- \rightarrow \text{O}_2 \text{(g)} + 2\text{H}_2\text{O} + 4e^- \quad \text{(in metal)} \]

where "aq." is aqueous phase and "g." is gas phase. The upper panel is modified from [39] and the lower panel is modified from [40]; both with permission © John Wiley & Sons Inc., New York.

Griffiths et al (2001)
Biochem. J. 355, 545.
Attractive Features of Mass Spectrometry

- High resolution (can measure mass differences of ~0.1 Da)
- High sensitivity (10^{-13} to 10^{-19} moles)
- Ability to work with complex samples (MALDI-MS)
- Ability to interface with other separation techniques (e.g., LC)
- “Reasonably” affordable
Example of an ESI Mass Spectrum

![Example of an ESI Mass Spectrum](image)

Griffiths et al (2001)
Biochem. J. 355, 545.

Figure 2  Positive-ion ES mass spectrum of a PEG mixture
The spectrum was recorded on an AutoSpec double-focusing magnetic-sector instrument. Peaks corresponding to \([\text{H}2\text{O}]_n\) are annotated.

Note: mass-charge ratio
Simple summary of mass spec-based proteomics

Virtually every mass spectrometry–based proteomic workflow consists of three distinct stages: (i) Protein samples are isolated from their biological source and optionally fractionated. The final protein sample is then digested and the resulting peptide sample is further fractionated. (ii) The peptides are subjected to qualitative and quantitative mass-spectrometric analysis. (iii) The large data sets generated are analyzed by suitable software tools to deduce the amino acid sequence and, if applicable, the quantity of the proteins in a sample. The peptide identity is assigned to the MS/MS spectra through database searching (21), which is performed according to established guidelines to generate consistent results (22). A subsequent statistical analysis of the search results is critical to ensure confidence in the identifications (23).
Mass spectrometry-based proteomics strategies

Fig. 2. Proteomics strategies. (A) Identification of simple protein (prot.) mixtures from 2D gel electrophoresis or pull-down experiments is carried out by enzymatic (enz.) digestion and by mass spectrometry analysis of the resulting peptides (pep.) (in ESI or MALDI mode). Peptide masses allows their identification (and that of the parent proteins) using peptide mass fingerprinting (PMF). Additional MS/MS data are also used for the peptide identification. (B) Random protein identification and quantification, also referred to as shotgun proteomics, couples identification and quantification of specific peptides in a sample. Selected peptides are subjected to product ion scanning (Fig. 1A) in a tandem mass spectrometer. The precursor ions are selected randomly, and typically only a fraction of the precursor ions detected are selected (undersampling). The ion intensities in MS1 are used to quantify the analytes by relating the signal intensity of the selected analyte to the signal intensity of a suitable reference molecule (frequently, a reference peptide labeled with heavy stable isotopes).

2-stage ("tandem") mass spectrometry with CID ("collision-induced dissociation") between MS1 and MS2

Combines mass determination with amino acid sequence determination

=> unique peptide identification

*Fig. 1.* Schematic representation of various types of tandem mass spectrometry experiments. (A) Product ion scanning is the most common MS/MS experiment in proteomics. Its purpose is the generation of fragment ion spectra for the identification of the amino acid sequence of specific peptides. In this experiment, the first analyzer (MS1) is set to a value that selects one specific precursor ion at a time. The selected ion undergoes CID in the collision cell, and the resulting fragments are analyzed by the second analyzer (MS2). This process is repeated for different precursors. (B)

Quantitation in mass spectrometry is a challenge!

**Fig. 3.** Strategies for quantitative peptide analyses. (A) Quantification using isotope dilution is widely used and accepted in the proteomics community. It is based on the incorporation of a stable isotope signature into all of the proteins of one sample and the incorporation of a different stable isotope signature in all proteins of a second sample. The samples are then combined to serve as mutual references. Stable isotope incorporation has been achieved by chemical modification of proteins using suitable isotope coded labeling reagents (26), metabolic labeling (35), or by enzyme reactions (36). The method is schematically illustrated here. (B) Quantification using tandem mass tags relies on variants of stable isotope labeling reagents (37, 38). They consist of two isotopically labeled elements, which have an overall constant mass. Currently, these reagents can be multiplexed to four channels. Quantification is performed in the MS/MS mode by measuring the relative intensity of the reporter group attached at the N terminus and observed in low mass range of the CID spectrum. (C) Quantification using internal standards is a variant of isotopic dilution in which a subset of isotopically labeled peptides is added to the sample at defined concentrations to perform precise quantification using calibration curves. Although it is more demanding in terms of sample preparation, this method is likely to gain importance in the future in the more directed approach indicated above for quantifying proteins in a larger number of samples. It may also be a more effective way to perform hypothesis-driven studies by screening for known or putative proteins (i.e., peptides) present in samples.

Figure 6 | Identification of candidate ATM and/or ATR substrates involved in the DNA-damage response. Cells treated with light-isotope-labelled amino acids were exposed to ionizing radiation, and cells treated with heavy-isotope-labelled amino acids were maintained under control conditions. Candidate ATM and ATR substrates were then identified by trypsin digestion of whole-cell proteomes, followed by immunoprecipitation with antibodies specific for the consensus ATM and ATR phosphorylation motif phospho-Ser/Thr-Gln, and then LC–MS/MS analysis. Phosphoproteins produced in response to irradiation were identified by ratiometric analysis of mass signals from light-isotope-labelled cells and heavy-isotope-labelled cells. Many of these proteins were found to have important roles in the DNA-damage response.

Figure 7 | Discovery of DAF-2-regulated protein pathways that modulate longevity in Caenorhabditis elegans. a, A quantitative proteomic analysis of changes in protein abundance was carried out in daf-2 mutant C. elegans, by using metabolic labelling and MudPIT analysis. Shown are representative examples of proteins that either decreased (EFT-2) or increased (TAX-6) in abundance in daf-2 mutants. b, Follow-up studies on differentially expressed proteins identified cases in which RNAi-mediated knockdown of the corresponding mRNAs decreased (EFT-2) or increased (TAX-6) the lifespan of worms. These results suggest that the proteins participate in compensatory pathways that limit the effects of daf-2 mutation on longevity and dauer formation. c, A model of how DAF-2-regulated proteins participate in compensatory pathways that affect longevity was assembled from the results of these experiments.

For some interesting recent applications of biological mass spectrometry please browse the following slides on your own and refer to the assigned paper, Cravatt et al (2007) *Nature* **450**, 991.
A “classic” proteomics use of mass spectrometry:

Inventorying the complete complement of proteins coded for by a genome

Example:
The proteome of *Plasmodium falciparum* (the malaria parasite)

Methodology:
*Multidimensional Protein Identification Technology* (‘MudPIT’*)

“MudPIT” (Washburn et al. [2001] Nature Biotechnol. 19, 242-247) is a highly parallel, automated method for carrying out proteomics inventories of expressed proteins in a complex sample. It is comparable to microarray experiments that are used to inventory expressed mRNAs in a tissue.

MudPIT involves digesting complex mixtures of proteins isolated from whole organisms with proteases, followed by automated liquid chromatography, tandem mass spectrometry, and ‘look up’ searches of databases of predicted proteins (inferred from DNA sequencing data) to identify peptides predicted to have originated from the expressed proteins contained in the sample.
A proteomic view of the *Plasmodium falciparum* life cycle


* Department of Cell Biology, The Scripps Research Institute, SR-11, 10550 North Torrey Pines Road, La Jolla, California 92037, USA
† Department of Proteomics and Metabolomics, Torrey Mesa Research Institute, Syngenia Research & Technology, 3115 Merryfield Row, San Diego, California 92121-1125, USA
‡ Infection and Immunity Section, Department of Biological Sciences, Imperial College of Science, Technology & Medicine, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ, UK
§ Naval Medical Research Center, Malaria Program (IDD), 503 Robert Grant Avenue, Room 3A10; and ¶ Department of Immunology, Walter Reed Army Institute of Research, Silver Spring, Maryland 20910-7500, USA
¶ The Division of Parasitology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
☆ Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA
☆☆ Malaria Research and Reference Reagent Resource Center, American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110-2209, USA
†† The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

The completion of the *Plasmodium falciparum* clone 3D7 genome provides a basis on which to conduct comparative proteomics studies of this human pathogen. Here, we applied a high-throughput proteomics approach to identify new potential drug and vaccine targets and to better understand the biology of this complex protozoan parasite. We characterized four stages of the parasite life cycle (sporozoites, merozoites, trophozoites and gametocytes) by multidimensional protein identification technology. Functional profiling of over 2,400 proteins agreed with the physiology of each stage. Unexpectedly, the antigenically variant proteins of *var* and *rif* genes, defined as molecules on the surface of infected erythrocytes, were also largely expressed in sporozoites. The detection of chromosomal clusters encoding co-expressed proteins suggested a potential mechanism for controlling gene expression.
Vaccine Targets in Malaria

Transmission-blocking vaccines
Stimulate immune response against sexual stages of parasite to prevent development of infectious sporozoites in salivary glands of mosquitoes, breaking chain of transmission

Blood-stage (erythrocytic) vaccines
Stimulate immune response against surface proteins of merozoites, reducing incidence and severity of clinical disease

Liver-stage (pre-erythrocytic) vaccines
Stimulate immune response against sporozoites or schizont-infected liver cells, preventing liver infection or release of merozoites from liver

Susan Okie, M.D.
357, 1877-81.
November 3, 2005
Figure 2: Expression patterns of known stage-specific proteins. a. Cell surface, organelle, and secreted proteins are plotted as a function of their known subcellular localization. b. stvor, var set of polymorphic surface variants are plotted as a function of the chromosomes encoding their genes. The matrices are colour-coded by sequence coverage measured in each stage (proteins not detected in a stage are represented by black squares). Locus names associated with these proteins are listed in Supplementary Table 2. Spz, sporozoite; mrc, merozoite; tsp, tachyzoite; gmt, gametocyte.

Generation of Distance Constraints for Protein Structure Determination Using a Combination of Protein Cross-Links and Mass Spectrometry
High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry

Malin M. Young*†‡, Ning Tang*†, Judith C. Hempel*, Connie M. Oshiro*, Eric W. Taylor§, Irwin D. Kuntz*, Bradford W. Gibson*¶, and Gavin Dollinger§¶

*Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446; and §Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608

We have used intramolecular cross-linking, MS, and sequence threading to rapidly identify the fold of a model protein, bovine basic fibroblast growth factor (FGF)-2. Its tertiary structure was probed with a lysine-specific cross-linking agent, bis(sulfosuccinimidyl) suberate (BS3). Sites of cross-linking were determined by tryptic peptide mapping by using time-of-flight MS. Eighteen unique intramolecular lysine (Lys-Lys) cross-links were identified. The assignments for eight cross-linked peptides were confirmed by using post source decay MS. The interatomic distance constraints were all consistent with the tertiary structure of FGF-2. These relatively few constraints, in conjunction with threading, correctly identified FGF-2 as a member of the β-trefoil fold family. To further demonstrate utility, we used the top-scoring homolog, IL-1β, to build an FGF-2 homology model with a backbone error of 4.8 Å (rms deviation). This method is fast, is general, uses small amounts of material, and is amenable to automation.
Young et al (2000)

**Fig. 1.** Experimental Design: (Top) Cross-linking of FGF-2. (Middle) Purification of monomeric FGF-2. (Bottom) Proteolytic digestion, HPLC separation, and mass spectrometry.
Fig. 2. Mass spectrometry. (A) MALDI-TOF spectrum from tryptic digest of BS\textsuperscript{3} cross-linked FGF-2. Cross-linked peptides are identified by using the program ASAP and are denoted with an asterisk (9). (B) MALDI-PSD spectrum of cross-linked peptide E45–R60 (M + H\textsuperscript{+} = m/z 2059.08). (C) MALDI-PSD spectrum of cross-linked peptides L23–R33 and E45–K52 (M + H\textsuperscript{+} = m/z 2465.25). Fragments from peptide L23–R33 are labeled with an α subscript; those from peptide E45–K52 with a β subscript. Fragments due to amide bond cleavages (b-ions, y-ions, and immonium ions) are labeled according to their single letter abbreviations. The most abundant fragments arose from cleavages at cross-linked lysines and aspartic acid. Peaks at m/z 696 and 1771 correspond to fragmentation at y\textsubscript{6β} and/or y\textsubscript{6α}, and b\textsubscript{2β} and/or b\textsubscript{5α}, respectively.

Table 1. BS\textsuperscript{3} Cross-linked tryptic peptides from FGF-2

<table>
<thead>
<tr>
<th>Lys-Lys cross-link</th>
<th>Cross-linked peptide(s)</th>
<th>Observed M + H\textsuperscript{+}</th>
<th>Error, ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>21–26</td>
<td>19–33</td>
<td>1952.09</td>
<td>39</td>
</tr>
<tr>
<td>21–125</td>
<td>19–26, 121–129</td>
<td>2151.27</td>
<td>45</td>
</tr>
<tr>
<td>21–135</td>
<td>19–22, 130–145</td>
<td>2327.30*</td>
<td>3</td>
</tr>
<tr>
<td>26–46</td>
<td>23–33, 45–52</td>
<td>2465.25*</td>
<td>2</td>
</tr>
<tr>
<td>46–52</td>
<td>45–60</td>
<td>2059.08*</td>
<td>1</td>
</tr>
<tr>
<td>46–119</td>
<td>40–52, 111–120</td>
<td>2889.41</td>
<td>12</td>
</tr>
<tr>
<td>46–125</td>
<td>40–52, 120–129</td>
<td>2764.48</td>
<td>13</td>
</tr>
<tr>
<td>46–145</td>
<td>45–52, 136–146</td>
<td>2267.24</td>
<td>0</td>
</tr>
<tr>
<td>77–86</td>
<td>67–97</td>
<td>3839.79</td>
<td>5</td>
</tr>
<tr>
<td>77–110</td>
<td>73–81, 110–119</td>
<td>2463.32</td>
<td>22</td>
</tr>
<tr>
<td>77–119</td>
<td>73–81, 111–120</td>
<td>2508.24</td>
<td>1</td>
</tr>
<tr>
<td>77–125</td>
<td>73–86, 121–129</td>
<td>2739.47*</td>
<td>2</td>
</tr>
<tr>
<td>77–135</td>
<td>73–81, 130–145</td>
<td>2894.41</td>
<td>41</td>
</tr>
<tr>
<td>110–119</td>
<td>110–125</td>
<td>2116.08*</td>
<td>20</td>
</tr>
<tr>
<td>110–135</td>
<td>110–119, 130–145</td>
<td>3056.61</td>
<td>17</td>
</tr>
<tr>
<td>119–125</td>
<td>111–120, 121–129</td>
<td>2401.26*</td>
<td>18</td>
</tr>
<tr>
<td>119–129</td>
<td>111–120, 126–135</td>
<td>2392.37*</td>
<td>25</td>
</tr>
<tr>
<td>125–129</td>
<td>121–135</td>
<td>1697.92*</td>
<td>0</td>
</tr>
</tbody>
</table>

*MALDI-PSD spectra were obtained for these peaks.
\textsuperscript{1}Mass calibrations were made with internal calibrations using previously identified unmodified tryptic peptides.
Young et al (2000)

Fig. 3. The 15 nonlocal throughspace distance constraints generated by the chemical cross-links (yellow dashed lines) superimposed on the average NMR structure of FGF-2 (1BLA). The 14 lysines of FGF-2 are shown in red.
Hydrogen-Deuterium Exchange and Mass Spectrometry to Probe Protein Dynamics, Solvent-Accessibility, and Protein-Protein Interactions
Domain Organization of d-AKAP2 Revealed by Enhanced Deuterium Exchange-Mass Spectrometry (DXMS)

Yoshitomo Hamuro1†, Lora L. Burns2†, Jaume M. Canaves2
Ross C. Hoffman2,3, Susan S. Taylor2,3 and Virgil L. Woods Jr1*

1Department of Medicine
University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92030-0656, USA
2Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92030-0656, USA
3Howard Hughes Medical Institute, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92030-0654, USA

Dual specific A-kinase anchoring protein 2 (d-AKAP2) is a scaffold protein that coordinates cAMP-mediated signaling complexes by binding to type I and type II protein kinase A (PKA). While information is unfolding regarding specific binding motifs, very little is known about the overall structure and dynamics of these scaffold proteins. We have used deuterium exchange-mass spectrometry (DXMS) and limited proteolysis to probe the folded regions of d-AKAP2, providing for the first time insight into the intra-domain dynamics of a scaffold protein. Deuteration on-exchange revealed two regions of low deuterium exchange that were surrounded by regions of high exchange, suggestive of two distinctly folded regions, flanked by disordered or solvent accessible regions. Similar folded regions were detected by limited proteolysis. The first folded region contained a putative regulator of G-protein signaling (RGS) domain. A structural model of the RGS domain revealed that the more deuterated regions mapped onto loops and turns, whereas less deuterated regions mapped onto α-helices, consistent with this region folding into an RGS domain. The second folded region contained a highly protected PKA binding site and a more solvent-accessible PDZ binding motif, which may serve as a potential targeting domain for d-AKAP2. DXMS has verified the multi-domain architecture of d-AKAP2 implied by sequence homology and has provided unique insight into the accessibility of the PKA binding site.

© 2002 Elsevier Science Ltd. All rights reserved.
Figure 1. (a) The predicted domain organization of human d-AKAP2. d-AKAP2 contains two putative regulators of G protein signaling (RGS) domains, a PKA kinase binding (AKB) domain, and a PDZ binding motif at the C terminus (TKL). A consensus PKA phosphorylation site is indicated between RGS B and the AKB domains. The bold, black line indicates the region of d-AKAP2 used in the current study. (b) A Clustal W sequence alignment of RGS4, RGS A and RGS B of d-AKAP2. An asterisk (*) indicates the identical residues and the residues comprising the hydrophobic core are in red. The secondary structure corresponding to RGS4 is indicated above the alignment.

Figure 3. A pepsin digestion map of d-AKAP2. A grey line indicates a peptide that was identified and a black line indicates a peptide that was used for DXMS analysis.

Hydrogen - Deuterium Exchange and Protein Conformational Fluctuations

\[
\begin{align*}
A-H + OD^- & \xrightarrow{k_{\text{m}}[OD^-]} A^- + H-OD \\
A^- + D_3O^+ & \xrightarrow{} D_2O + A-D
\end{align*}
\]

\[b\]

\[c\]

Figure 1 Mechanism of hydrogen exchange. (a) Amide exchange at neutral pH involves base catalyzed proton abstraction and acid catalyzed transfer of deuterium from solvent. Measurable isotope effects on the amide hydrogen and a lack of a solvent isotope effect indicate that proton abstraction is rate limiting. (b) Hydrogen exchange of a buried amide is facilitated by different mechanisms, involving small-amplitude fluctuations (upper pathway) on one extreme, and complete unfolding (lower pathway) on the other. The observed rate of exchange (\(k_{\text{obs}}\)) for small-amplitude fluctuations is a function of the rate of structural opening (\(k_{\text{op}}\)), the rate of structural closing (\(k_{\text{cl}}\)), and the chemical rate of exchange (\(k_{\text{ch}} = k_{\text{int}} [\text{catalyst}]\)), where catalyst is OH\(^{-}\) or buffer. In native proteins, the rate of opening is assumed to be much slower than the rate of closing, which results in a simplified rate expression (upper equation, far right). The observed rates of small-amplitude fluctuations lie on a continuum described by EX1 and EX2 conditions, as described in the text.

Hydrogen - Deuterium Exchange: Schematic Experimental Design

Figure 3  Apparatus for HX-MS by fragment separation. Proteins are incubated with $D_2O$, quenched in acid and lower temperature, and digested with pepsin. Peptides are separated with reversed-phase HPLC prior to analysis by ESI-MS. Parts of the HPLC, including the injection syringe, solvent precooling loop, sample loop, injector, and capillary column, are all immersed in ice to minimize back-exchange.
Figure 4. Deuteration levels of D-AKAP2 after on-exchange at 0 °C. The deuteration level is indicated below the sequence for D-AKAP2 as a colored bar (see the inset). The on-exchange time is listed to the left of the sequence. The residues corresponding to the split RGS A, RGS B, AKB and PDZ binding motif are indicated. Black arrows indicate limited proteolysis cleavage sites after one hour. White arrows indicate limited proteolysis cleavage sites after 24 hours. The asterisk (*) above S267 indicates an in vitro PKA phosphorylation site.

Figure 5. (a) A backbone structural alignment of the modeled RGS domain of D-AKAP2 and RGS4. (b) The modeled RGS B domain of D-AKAP2 with deuteration levels after 3000 seconds on-exchange at 0 °C are mapped onto the structure. (c) Deuteration levels of RGS B domain after 3000 seconds on-exchange at 0 °C. Black bars above the sequence indicate the structure predicted in the modeling. Red characters indicate the conserved hydrophobic residues in Figure 1(b).

Fragment-Based Drug Discovery
Site-directed ligand discovery


*Sunesis Pharmaceuticals, Incorporated, 3696 Haven Avenue, Suite C, Redwood City, CA 94063; and ‡Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0448

We report a strategy (called “tethering”) to discover low molecular weight ligands (≈250 Da) that bind weakly to targeted sites on proteins through an intermediary disulfide tether. A native or engineered cysteine in a protein is allowed to react reversibly with a small library of disulfide-containing molecules (≈1,200 compounds) at concentrations typically used in drug screening (10 to 200 μM). The cysteine-captured ligands, which are readily identified by MS, are among the most stable complexes, even though in the absence of the covalent tether the ligands may bind very weakly. This method was applied to generate a potent inhibitor for thymidylate synthase, an essential enzyme in pyrimidine metabolism with therapeutic applications in cancer and infectious diseases. The affinity of the untethered ligand ($K_i≈1$ mM) was improved 3,000-fold by synthesis of a small set of analogs with the aid of crystallographic structures of the tethered complex. Such site-directed ligand discovery allows one to nucleate drug design from a spatially targeted lead fragment.
High local concentration of bound “monophore” (molecule shown in red) stabilizes the disulfide bond.

Fig. 1. (A) Schematic illustration of the tethering approach: a cysteine-containing protein is equilibrated with a disulfide-containing library in the presence of a reducing agent such as 2-mercaptoethanol. Most of the library members will have little or no inherent affinity for the protein, and thus by mass action the equilibrium will lie toward the unmodified protein. However, if a library member does show inherent affinity for the protein, the equilibrium will shift toward the modified protein. (B) Schematic illustration of a generic disulfide library derived from carboxylic acids. Other functional groups have also been converted to disulfide libraries, as described in Materials and Methods. In the present case, 1,200 compounds were screened against TS in pools of 8 to 15 compounds.
Fig. 2. (A) Two representative tethering experiments. The protein TS is present at a concentration of 15 μM, and each of the 10 disulfide library members in each pool is present at 200 μM. The buffer contains 25 mM potassium phosphate (pH 7.5) and 1 mM 2-mercaptoethanol, and the samples were allowed to equilibrate at ambient temperature for 1 h before analysis. (B) Three tethering experiments in which the concentrations of 2-mercaptoethanol were varied as stated. The pool of disulfides is the same as in A Right, and the conditions (other than 2-mercaptoethanol concentration) were the same as above. (C) Three tethering experiments in which the pool size and concentrations were varied as stated. All other conditions were the same as above.

Fig. 3. Compounds either selected by covalent tethering (Left) or present in the disulfide library but not selected (Right). Except in the case of \(N\)-tosyl-\(d\)-proline, all compounds tested were racemic unless otherwise indicated. For \(N\)-tosyl-\(d\)-proline, both stereoisomers were screened separately in different pools, and both were identified as hits, although the \(d\)-isomer appeared to be selected slightly more strongly than the \(L\)-isomer.

The Monophore Binds in the Same Relative Orientation, Regardless of Attachment Site

Fig. 4. Overlay of three crystallographically determined structures. The structure in green was determined after soaking $N$-tosyl-$d$-proline (free acid) into crystals of unmodified TS. The structure in red is TS covalently modified by $N$-tosyl-$d$-proline disulfide-bonded to C146 (the active-site cysteine). Finally, the structure in blue is mutant TS (C146S/L143C) covalently modified by $N$-tosyl-$d$-proline disulfide bonded to L143C.

Fig. 5. Grafting a glutamate residue onto N-tosyl-D-proline improves the affinity 50-fold, and adding a negatively charged appendage further increases the affinity by an additional 70-fold.
Pharmacology Using Imaging Mass Spectrometry
Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections
Pierre Chaurand, Sarah A Schwartz and Richard M Caprioli*

MALDI MS imaging mass spectrometry can be used to map the distribution of targeted compounds in tissue sections with a spatial resolution currently of about 50 µm, providing important molecular information in many areas of biological research. After matrix application, a raster of a section by the laser beam yields ions from compounds in a tissue mass-to-charge range from 1000 to over 100 000. Two-dimensional intensity maps can then be reconstructed to provide specific molecular images of a tissue.

Addresses
Mass Spectrometry Research Center and Department of Biochemistry, Vanderbilt University, Nashville, TN 37232-6400, USA
*e-mail: r.caprioli@vanderbilt.edu

Current Opinion in Chemical Biology 2002, 6:676–681
MALDI Imaging of a transversal rat brain section. (a) Survey profile with data acquisition taken randomly across the section. (b) Optical image of the section before matrix application. The general outlines of the section as well as several features visible in the section have been delineated. (c)–(g) Ion density maps obtained at different m/z values. The section was imaged with a resolution of 180 μm.