HUPO 2009 World Congress

Seminar Notes, Observations, and Thoughts

Lee S Parsons

October 2009

HUPO Conference Structure

- Saturday & Sunday - “pre-congress”
  - Saturday “education day”
  - Sunday “clinical day”
- Monday, Tuesday, Wednesday - “congress”
  - Two morning, Two evening plenary speakers per day
  - 5 parallel tracks throughout each day
  - Unscheduled poster sessions each day, posters up for one day only

Saturday - “education day”

- Several “Initiative Workshops” took place simultaneously.
- Also: “The 1st International Forum of Proteomics”, including several award lectures
- Registration stuff – bag, flash drive, pens, publications, vendor advertisements, schedule, maps, name tag

Education Day, Theirry Rabilloud

- “Critical Analysis of the Current Proteomics Platform”
- Proteins have dynamic range on the order of 10^10
- Undersampling is a major problem in proteomics – a unique problem within the *omics fields
- Sampling in proteomics generally in the range of 10^4

Education Day, Theirry Rabilloud

- Range example: T4 virion
  - Total 37 different proteins; expressed from 3 – 954 copies per virion
  - Best current shotgun approaches can identify around 30 of them in a single approach
- Proteomics bottlenecks:
  - Dynamic Range
  - MS speed (missing other molecules during ms^n)
  - Peptide separation

Education Day, Theirry Rabilloud

- Shotgun proteomics is intensive in MS time
- Blind-cutting gels is useful but has reproducibility issues
- 2d gels save MS time but can lose proteins
- Future proteomics approaches may need to specifically focus on protein modifications and combinations of modifications

Education Day, Shabaz Mohammed

- Utrecht University, The Netherlands
- “Liquid Chromatography”
- Generally, peptide response is proportional to the inverse of the flow rate – sweet spot around 10-20 nL/min with tip diameter ~1um
- This needs to be set to match the diameter of the column to avoid drift
- Optimal linear flow for nanoLC around .2cm/s
  - 20nL/min for 15um column

Education Day, Shabaz Mohammed

- Dramatic change in signal occurs as column ID goes down from 75 to 15um
- RP C18 chemistry elutes least hydrophobic peptides first; most elute between 10-50% ACN
- LC → MS transition requires a conductive needle
  - Generally Ag or Pt coated Si needle, teflon sleeve
  - Watch for dead volume in sleeve
- Non-volatile salts can crystallize in needle!

Cardiovascular and Blood Plasma Proteome

- Workshop running simultaneously to education day, in adjacent building
- Chaired by Peipei Ping (though not seen) and Gil Omenn

Plasma – Anthony Gramolini

- U Toronto Dept of Physiology – multiple Orbis
- “Mouse Cardiac Proteomics”
- Ca+ cycling in cardiac cells via sarcoplasmic reticulum
- Ca+ ATPase (SERCA)
- Dounce homogenization?

Plasma – Anthony Gramolini

- Cardiac Proteome
  - ~4906 proteins
    - 3666 in specific fractions
      - 1234 cytoplasmic
      - 939 micro?
      - 647 mitochondrial matrix
      - 846 mitochondrial membrane
- “We know a lot about very few proteins”
  - 2019 unannotated / poorly annotated of cardiac relevance
  - 30% of top 200 had no localization information

Plasma – Anthony Gramolini

- Much of the progress so far has selected **against** membrane proteins in the cardiac proteome
  - How to correct this trend?
- Schnitzer, Jan:
  - Nature 429 629-635 (2004)
  - Nature Biotech 22 (2004)

Plasma – Jennifer Van Eyk

- Cardiac dissection in both Marfans and non-Marfans patients
- Marfans – Fibrillin 1 defect (FBN1 allele)
  - C1039G mutation?
- An “aged” cardiac cytoskeleton, “primed to do something” (rat study, same mutation)
  - Serum amyloid A goes up with age
  - Reduced by Losartan (angiotensin II inhibitor?)

Plasma – Jennifer Van Eyk

- Aortic dissection is the most common cause of death amongst Marfans patients
- Acute dissection can happen to non-Marfans patients as well
  - ~1% of chest pain patients in ER are presenting with acute dissection
  - Difficult to diagnose w/o imaging
  - Mortality >1% per hour; ~50% by 24 hours untreated

Plasma – Jennifer Van Eyk

- Skeletal proteome proteins present in plasma due to high turnover rate (esp vs cardiac)
- MESOSCALE – electrochemiluminescence detection
- “thoracic aortic aneurysm”
- GenTac dataset – tracking any cardiac condition: TGF-Beta increases very common in this set

Plasma – MingMing Ning

- Mass. General Hospital
- “Brain → Heart interaction … heart → brain interaction”
- Patent (patient?) foramen ovale (PFO) – direct passage of blood between L and R atria
  - PFO patients 5-10x more likely to have a stroke
  - Clot can pass back to atrial side
  - Resolution for PFO is not clear
  - Migraines are common in PFO patients

Plasma – MingMing Ning

- “Voodoo death and economy class stroke”
- Some number of patients see normal cardiac EKG after brain death...
- Bidirectionality of brain → heart interaction?
- PFO patients can face stroke risk from flights of 5 or more hours if not properly mobile and hydrated
  - On a long flight you should drink enough fluids to need to use the bathroom at least once en route

Plasma – MingMing Ning

- 50% of ischemic strokes originate with the heart – one hypothesis includes clots passing through PFO
- Alcohol as a stroke risk factor?
- 85-90% of PFO patients w/o clot-forming conditions end up with embolisms - why?

Plasma – Terry Farrab

- ISB (Aebersold?) Computational Biologist
- “2009 Human Plasma Peptide Atlas”
- Plasma Proteome Project (PPP) 3,020 proteins in 2005 atlas - www.peptideatlas.org
- Currently has 2052 human plasma proteins
- Spectrum libraries available (part of Reudi's big project?)
- Also available as MySQL DB dump

Plasma – Terry Farrab

- PeptideProphet – probability of a match? To be released “soon”
  - Based on sorting peptide families to a “canonical” protein
  - Canonical should have fewer than 80% of its peptides with other canonicals
  - Not clear on how the first canonical is chosen...
- How to estimate False Discovery Rate (FDR)?

Plasma – Juan Antonio Vizcaino

- EMBL-EBI
- “Pride: ProteinExchange and HUPO 2009 PPP2 Update”
- “pride converter” (Java) code.google.com/p/pride-converter

Plasma – Bernd Wollscheid

- ETH Zurich (the other group Aebersold works with - “Institute of Molecular Systems Biology”)
- Focus on Glycoproteomics
  - Biomarker relevant sub-proteome
  - 34 of 42 currently used clinical biomarkers are known to be glycosylated
- Triple quad SRM to monitor glycosylation
- MRM atlas: mrmatlas.org / srmatlas.org

Plasma – Bernd Wollscheid

- “Artificial proteomes”
  - SPOT synthesis
  - 5000 human / 4000 mouse
  - Peptides in 96 well plate
  - Mix and dissolve, then shoot as control
    - Validated coordinates for later samples
- What is a proteotypic peptide?

Plasma – Michael Kuzyk

- Christoph Borcher's group
- “Plasma MRM assay development & Application to a pilot cardiovascular disease (CVD) patient cohort”
- Plasma trypsin digest kinetics
- Previous work; 177 CVD putative markers, 45 moderately high abundance (Anderson 2006)
- +2 vs +3 precursor ions; ~5x signal loss for same peptide

Plasma – Michael Kuzyk

- Try to match internal standard (IS) 1:1 to experimental peptide
  - Signal within 10 fold is OK but closer is better
- Multiplexed study
  - Individual biomarkers of CVD have minimal value currently
- Microtiter plate (MTP)
- Support vector machine (SVM)

Kidney - Satish P RamachanandraRao

- Capillary Electrophoresis – Mass Spectrometry (CEMS) for urinary analysis
- Diabetic / renal complication patients
- Dihazi H, et al; Clin Chem 2007 53; 1635-1645
  - SELDI analysis of urine proteins
- HKUPP Human Kidney & Urine Proteome Project
- Siriraj Hospital, Bangkok

Kidney – Barbara Seliger

- Medical Immunology, Martin Luther U Germany
- Identification of novel biomarkers in renal cell carcinoma
- “Proteomex” - serum incubated on film to compare before / after
- Cellular distribution associated with function

Education – Bernd Kuster

- “Quantification by MS”
- Hard MS question – how much protein is present?
- Much easier – how does it change by time or condition?
- Cellular protein expression spans 6-8 orders of magnitude
- Serum spans more than 10

Education – Bernd Kuster

- Some labs report analytical data for 300-3000 proteins
- Technical reproducibiilty 50-90%
- Relative vs Absolute quantification
  - Absolute most often includes an IS
- Direct MS signal quantification
  - Signal intensity proportional to amount
    - Peptides vary in bottom-up proteomics
      - Must compare the same **peptide** in a digest

Education – Bernd Kuster

- Old, et al, MCP 2005
  - Comparison of label-free methods for quantifying human proteins by shotgun proteomics
- Quantification accuracy (in MS2) increases with increasing number of peptides

Education – Bernd Kuster

- Spectrum counting (label free quantitation)
  - Protein level, not peptide level
  - Compare one protein across samples, differs by protein
  - Needs many spectra for accuracy
  - Metabolic labeling of cells is best for technical accuracy
    - Early labeling ensures all populations treated equally downstream

Education – Bernd Kuster

- SILAC:
  - Ong, et al, MCP 2002
- TMT
  - Tandem mass tag (Thermo)
- Absolute quantification of one sample (AQUA)
  - Isotopically labeled peptide std spiked in
- MRM
  - Applied when the product protein is known: **not** a discovery experiment

Education – Bernd Kuster

- Data analysis: don't use simple thresholds or cutoffs!
- Global view of variation for proteins doesn't really tell up/down regulation
- Population proteomics studies are inherently difficult
- Top 3 peptides as a single datum for spectra counting
  - Careful to avoid comparing **within** a single spectra

Education – Ole Noerregaard-Jensen

- “Beyond Simple Translation: The Challenge of Modified Peptides”
- Human Proteome Project:
  - Now trying to characterize all the products of any gene (mRNA and protein variants)
- PO4: 80 Da mass shift
- MS3 can find sites of labile PTMs (that do not survive MS2)
- MR Larsen, et al, Biotechniques 2006

Education – Ole Noerregaard-Jensen

- Whole cell lysate isn't very good for looking for phosphopeptides
- Phosphopeptide separation: Orthogonal separation prior to MS
- Shave-and-conquer technique for membrane proteins
- www.hupo.org/educational/courses

Saturday evening Workshop

- “Gene-Centric Human Proteome Project”
  - Strategies and Specific Inputs for HUPO Human Proteome Project
  - Coordination and Collaboration across the Initiatives, including use of ProteomExchange
- Open forum with powerpoint...

Saturday Workshop

- Opening statements (speaker ?)
- Open issues:
  - Drowning in data, Starving for Meaning
  - Who's buying?
  - What's for sale?

Saturday workshop

- Is the Human Proteome Project (HPP) “Investable?”
  - Creating enduring value
  - Need positive reasons for investors
  - NIH is built on disease-specific initiviatives
  - Need an endpoint to encourage funding...
  - Define deliverables in order to reach where the genome project is today
- Leaders have a short attention span

Saturday workshop

- Pierre Legrain
- What should HPP achieve in next 10 years?
- Goal:
  - Detect any protein in any sample, accurately
- Objective:
  - DB for reference, “certified” MS
  - Linkage to ProteinAtlas, UniProt

Saturday workshop

- Legrain, continued
- Final Output (not the deliverable):
  - Phenotype or disease status related to protein or protein expression
- Centering on proteins / genes instead of biological sample
- MS as a standard for analysis
  - Best known proteins as a check
  - If you can't find them, then check your technique!

Saturday workshop

- Legrain, continued
- Three categories of human proteins
  - A: Very well known, characterized, annotated
  - B: Partially characterized / annotated
  - C: Poorly annotated or completely un-annotated
- One year's post-doc work should suffice to move 1 protein from C to B
- We should have MS signatures for **all** proteins of **all** categories

Saturday workshop

- Legrain, continued
- HUPO Agenda:
  - 2010 – Working group with funding
  - 2011 – Launch HPP pilot phase (2-3 years)
  - 2013 – Prep full scale HPP
  - 2014 – Full scale HPP launch
- HUPO role (as an organization):
  - Initiate working group formation, negotiate funding

Saturday workshop

- Tommy Nilsson (?)
- 120 person-years into annotating human protein-encoding genome
  - 20,331 proteins
    - 14,000 isoforms 46,000 SAPs, 21,000 to disease
    - 60,000 post-translational modifications (PTMs)
- MS is robust
- Antibodies are robust

Saturday workshop

- Kenny? (Juan Antonio Vizcaino?)
- ProteomExchange Consortium
- Is raw data safeguarded?
- Will it be available from Tranche later?
- 1 Jan 2010: MCP journal will require raw data submission for publication

Saturday workshop

- Reudi Aebersold
- US – 10 year $70M proteomics investment
- 3 assumptions shared between genomics and proteomics:
  - “know the parts, know the biology”
  - “space is finite”
  - “complete is good”

Saturday workshop

- Aebersold, continued
- Proteomics & Genomics, compared:
  - Genome map was an endpoint
  - Proteome map is only a starting point
    - Proteome needs to be measured **quantitatively**
- Sensible goal:
  - Affinity reagents, robust assays, antibodies
  - MS assays, robust assays, SRM
    - For all human proteins!

Saturday workshop

- Aebersold, continued
- 2 year, $1M US project
- Reference spectra, Ref SRM assays from synthesized peptides
- Available for researchers to use as reference in their experimentation
- (Peptides synthesized to match selected parts of human proteins)

Saturday workshop

- Erik Deutsch – Peptide Atlas (ISB)
- Benefits of centralized analysis
  - More complete identification
  - Promote complete raw data accessibility

Saturday workshop

- EGFR as a starting point?
- Disease focus a a logical starting point for funding?

- Tanaguchi (HUPO award)
  - >50% of human proteins are glycosylated
  - Functional glycomics as a starting point
  - Glycosylation has no chromosome bias

Saturday workshop

- Albert Heck
  - Phosphorylation studies on hESCs...
- Gene Centric HPP (GC-HPP)
  - Young-Ki Paik

Sunday – the plan

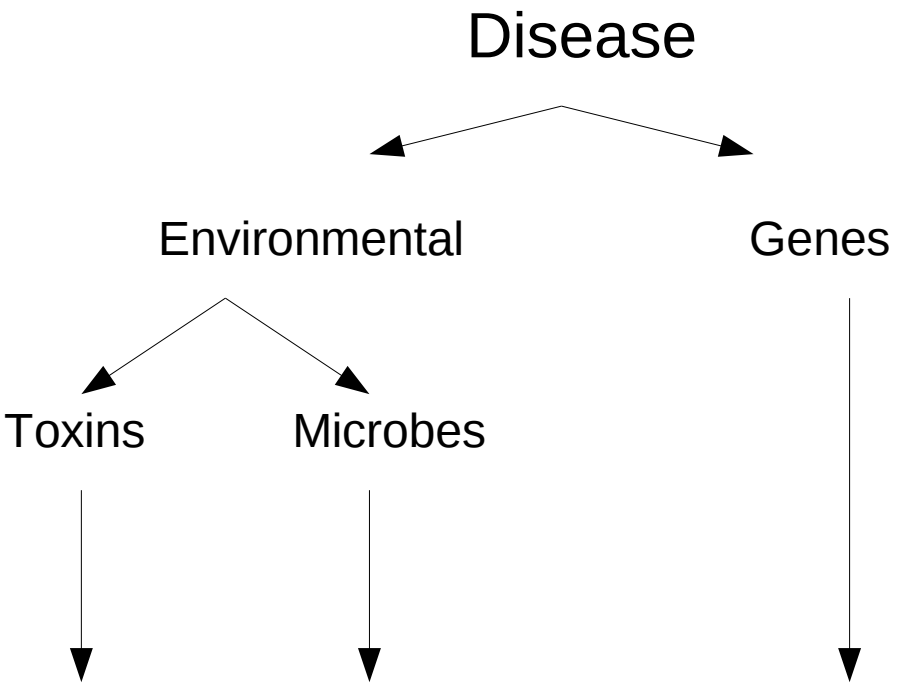
- Clinical proteomics 9 – 10:40
- Tech devel 10:45 – 12:00
- Clinical 12 – 12:30
- Lunch 12:30 – 1:30
- Clinical 1:30 – 3:00
- Tech Devel 3:15 – 4:30
- Bruker Dinner Cruise at 6:30

Sunday Morning; Clinical Proteomics

- Hochstrasser (“hawk-straw-sir”); Geneva University & University Hospital
- “Clinical Toxicology & Proteomics: Opportunities and Challenges”
  - Later: MALDI-TOF for bacterial identification
    - Much faster than traditional methods – up to 24hours faster
- Toxins as the primary threat to humans
- Swiss Center of Applied Human Toxicology (SCAHT.org)

Hochstrasser, continued

- Human toxicology paradigm
- Environmental factors
  - Toxins
  - Microbes
- SCAHT
  - 21 research groups
  - Swiss fertility: down 50% in 10 years in men



All disease caused by 1 or more of these three

Disease categories:
Acquired, predisposed, or multifactorial

Hochstrasser, continued

- Interactive periodic table
  - Pubmed/SwissProt references for elements
    - Element → peptide → protein
  - MALDI-ICP-MS for elemental analysis of tissue
    - 50um sections imaged
- Arsenic makes mice more susceptible to H1N1?

Hochstrasser, continued

- Proteomics perspective
  - MS biomarkers for detecting exposure signatures (drugs and toxicants)
    - Stomach: ~6k genes
    - Liver: ~8k
    - Brain: ~10.5k
    - Testes: ~12k
- *Ex vivo* approaches
  - “mini brain” - 3d growth shows EEG
  - Testes in a plate
  - Swiss hospital removes ~40 testes / year from patients

Hochstrasser, continued

- MS2 standards for toxicology labs...
- Similar problems for small molecules as with peptides
- Small molecules often benefit from negative ion MS rather than positive
- How to screen 300k molecules?
- Infectious disease (a la CDC) versus toxic exposure (SCAHT)

Clinical proteomics, continued

- Goha, University Health Network (Toronto), Labatts Brain Tumor Center
- “Primary Malignant Brain Tumors”
- Current and Potential Impact of Proteomic Based analysis
- Brain tumors are the 2nd most common adolescent cancer
- 2-3 of 100,000 adults get grade IV brain tumor
  - Prognosis usually 12 months or less
  - Surgery, reduce ICP, decrease symptoms

Goha, continued

- Brain abscesses can be treated with antibiotics
- Surgery cannot cure glioblastoma on its own
  - Low grade glioma responds well to surgery
- Adult brain does show limited plasticity
- Gliadel Wafers to surgery cavity – deliver drugs to site after surgery
- Balancing biomarker search requirements with the reality of the patient
  - Important to minimize sampling frequency of tumor

Goha, continued

- Biological imaging; more than just the lump
- www.braintumorbank.ca
- EGFR in cancer
  - Both N and C terminal mutations in patients
  - Though reg/exp changes are more common
  - N term second most common (often late start and novel glycine residue)
  - C term 3rd most common (often early termination)
  - Mutant EGFR can dimerize with wild type

Goha, continued

- ProteomeRes (Goha, Siu, et al?)
  - Proteome Identification in Cerebral Spinal Fluid (CSF)

Clinical Proteomics, continued

- Dario Neri – ETH (Swiss Federal Institute of Technology, Zurich)
- “Translating Proteomics to the Clinics”
  - Neri & Bicknell Nat Rev Cancer 5 436-446
- Vascular Tumor Targeting – concentrate drugs in tumors
  - Rybak, et al, Nature Methods 2, 291-298
- Finding vascular targets
- Clinical applications
- Looking for proteins that are unique to tumors (absent in healthy tissue)

Neri, continued

- Fibronectin: alt splice form in tumors found by MS2
  - Metastatic liver tumor
  - Raise Abs to the alt form to base treatment on
- Chemical proteomic analysis of neo-vasculature in lymphoma
- Clinical applications
  - Fibronectin: EDA EDB extra domains in tumors
  - Schliemann et al, Blood 2009 113, 2275-2283
  - Glucose consumption as tracking method for tumor
  - Normal: most glu consumption in heart, brain, urine?

Technology Development Workshop

- Albert Heck
- “Enabling Technologies for Phosphoproteomics”
  - 2002: Hunt detects 383 phosphorylation sites by MS
  - Now: Over 12,000 sites shown in one study
    - Mostly Ser/Thr; Tyr sites are more difficult

Heck, continued

- Challenges of Phosphoproteomics
  - Highly dynamic
  - Very low abundance
  - Cell phosphatases may remove phosphorylation during or after lysis
  - Peptide detection more difficult than normal
- SCX/SAX, then TiO2 or IMAC for enrichment
- Agilent phosph chip: TiO2 + RP “microfabricated system”

Heck, continued

- 5,222 proteins; 1,280 phosphoproteins; 3,201 phosphopeptides
  - 2,466 Serine; 612 Threonine; 50 Tyrosine
- Improving Phosph-Tyr detection
  - Currently: need a lot of material (~2 mg)
  - Phosph-Tyr antibodies?
- How to reduce cost vs ITRAQ – dimethylene labeling
  - Formaldehyde labeling of peptides (Nat Met paper)
    - “about 10,000 times cheaper”
  - In-line reaction cell (LC)

Heck, continued

- Lys-N cleavage (instead of Lys-C Trypsin)
  - Different cleavage pattern – charge stays with Lys
  - MS/MS shows C series ions
  - Very helpful for phosphoproteomics
- Phosphoproteome is less than 1-20% complete

Technology Development, continued

- Neil Kelleher, UIUC (moving to Northwestern)
- Top-Down Proteomics: 2009
  - Bottom-up is more common (trypsin digest)
- MS/MS with accurate mass to start both bottom-up and top-down
  - Top-down starts with intact protein, then does collision after ion trap
- New MS instruments have high resolution on both MS1 and MS2 – previous instruments lost resolution on MS2.

Kelleher, continued

- Fragment ions at FTMS resolution
- Multiplexing: chimeric mass spectra
- “ProSight PC” - “software for precision proteomics” (Thermo?)
- Three pillars of top-down proteomics
  - Front end (sample prep, LC)
  - Automation (MS instrument)
  - Back end (software)

Kelleher, continued

- Shotgun annotation of histone modifications
- “ProSight PTM” - web version of ProSight?
  - 7T LTQ, upgraded to 12T LTQ
    - 14.5T LTQ to come online later?
  - “Top-down proteomics on a chromatographic time scale”
- “GELFREE”
- New Objective PicoTips with columns inside?

Kelleher, continued

- SDS Removal procedure
  - Methanol based
  - Anal. Biochem. 1984, 138 141-143
  - Wessel D, Flugge U

Tech Devel, continued

- Lilijana Pasa Tolic; PNNL
- “Integrated Top-down and Bottom-Up”
- Bottom-up: problem with incomplete sequence coverage
- Top-down: problem with severe undersampling
- Orbitrap good to around 15-20 kDa mass
  - Problems with mid-to-high mass
- Farnesylation: 204.1 Da PTM

Tolic, continued

- High pressure on-line digestion system (pepsin based)
  - Capture run: proteins
  - Replay run: peptides
- Environmental Molecular Sciences Lab (EMSL) at PNNL
  - 12T MS
  - 900 Mhz Wide Bore NMR

Afternoon Tech Devel

- Pierre Chaurand: U Montreal (formerly Vanderbilt)
- “MALDI Imaging Mass Spec (MALDI-IMS)”
- Profiling vs. Imaging
  - Profiling – low density, droplet arrays
  - Imaging – homogenous, matrix coating
- High density matrix array spotting
  - Preventing delocalization at the expense of resolution?
  - 40um resolution

Chaurand, continued

- Whole Body Imaging
  - 24 hours MALDI time to image 1 day old mouse pup
- On-Tissue trypsin digest
  - Print trypsin with spotter first, then print matrix
- Formalin Fixed Paraffin Embedded (FFPE) tissue
- Continuous scanning 5khz laser can do 1 day old mouse pup in 2 hours
  - 3-5 um resolution (lab of Rich Caprioli)

Afternoon tech devel

- Michael Becker (Bruker?)
- High Definition MALDI Imaging (HDMI)
  - “Technologies for top-down biomarker discovery and clinical histology”
- MIRAX - “Google Earth for Pathologists”

Afternoon tech devel

- Bruno Domon – IMSB ETH (Aebersold's group?)
- “Global Proteome Analysis: Detection, Identification, and Quantification of Peptides in Large-Scale SRM Experiments”

Sunday Evening Talks

- Carol V Robinson
- “Use of Mass Spectrometry For Studying Membrane and Soluble Complexes”
- Opening question - “Why couldn't we study membrane protein complexes?”
  - ~100 detergent molecules per hydrophobic molecule to form a micelle
  - The “micelle mountain” in mass spec results
    - Large high-signal hump in lower m/z range pushes down signal of the complex itself

Robinson, continued

- Can the detergent be stripped in the gas phase instead?
  - Previous approaches were to remove the detergent while still in liquid phase
  - Spectra looks better when detergent is stripped in gas phase!
- Micelles seem to **enable** MS of macromolecular complexes
- Lipid binding can be apparent in spectra

Robinson, continued

- Thermus Thermophilus V1VO ATP Synthase
- Psuedo-atomic modeling via ion mobility separation (IMS)
  - IMS can give information on the cross-sectional diameter of a molecule
- Thermostable complexes behave better in MS than others...

Sunday evening Talks

- Richard Smith – PNNL
- “Improved Proteomics Measurements Based Upon Ion Mobility Separation Combined with Mass Spectrometry”
- Transmitted current in a mass spec – up to 10^11 elementary charges per second
  - Though analysis capabilities are on the order of 10^6

Smith, continued

- IMS can be used as an additional dimension for separation
  - Though it is not completely orthogonal to MS
- 50us IMS time for a good separation?
- IMS dimension can help to identify additional features beyond what traditional LC/MS can analyze
-

Sunday evening

- Andrew Emili (University of Toronto)
- “Targeted Pathway Monitoring Using Synthetic Peptide Arrays”
- “New Tools For Biology”
- “Next Generation Protein Sequencing”
  - Global expression profiling w/ single molecule sensitivity
- Proteomics phenotype + Genomics genotype = personalized medicine

Emili, continued

- Mass Spec is orders of magnitude less sensitive than genomics technology
- Protein sequencing by (Edmann) degradation:
  - From N terminus
  - Easily falls victim to phasing asynchronicity
  - How to identify the AA of cleavage?
    - HPLC not currently capable – lacking resolution
- Changing DNA sequencing chemistry for AA sequencing?

Emili, continued

- Single cell proteomics
  - Discovery, diagnostics, screening
  - Resolving 100,000,000,000 molecules per cell