The Clinical Potential of the Human Plasma Proteome

Leigh Anderson Ph.D.
Founder & CEO, Plasma Proteome Institute
Board Member, Dade Behring

PPI
The Plasma Proteome Institute
Plasma is the largest, and deepest, version of the human proteome

- Largest = Most proteins
- Deepest = Widest dynamic range
Major Components of the Plasma Proteome

- ~40,000 forms of proteins secreted to function in plasma, most glycoproteins
  - Assume 500 gene products x 2 splice variants x 20 glycoforms x 2 clip forms
- ~500,000 forms of tissue proteins
  - Essentially all tissue proteins x splice and PTM variants
- ~10,000,000 clonal forms of immunoglobulin

Total: the largest version of the human proteome
A New Functional Classification of Proteins in Plasma

1. Secreted proteins that act in plasma (e.g., albumin, fibrinogen)
2. Immunoglobulins (Ig’s A,M,G,D,E)
3. Tissue leakage products (e.g., cardiac Mb)
4. “Distant” receptor ligands (e.g., insulin)
5. “Local” receptor ligands (e.g., IL-8)
6. Aberrant secretions (e.g., PSA in cancer)
7. Temporary passengers (e.g., lysosomal enz.)
8. Foreign proteins (e.g., virus)
Major Plasma Proteins

99% of plasma protein mass

- Albumin
- IgG Total
- IgA Total
- C3 Complement
- Complement Factor B
- C4 Complement
- IgM Total
- Complement Factor H
- Transferrin
- Alpha-1-Antitrypsin
- Factor H
- C8 Complement
- Fibrinogen
- Lipoprotein(a)
- C1q Complement
- C9 Complement
- Alpha-2-Macroglobulin
- Haptoglobin
- Apolipoprotein A-1
- Ceruloplasmin
- C1q Complement
- Apolipoprotein B
- Lipoprotein(a)
- Factor H

0 - 90%
90 - 99%
Proteins Measured Clinically in Plasma Span
> 10 Orders of Magnitude in Abundance

Normal Range Abundances in pg/mL

Classical Plasma Proteins
Tissue Leakage
Interleukins & Cytokines
Other

Dynamic range of single proteomic techniques
Total Cost: $10,695 per sample

PPI Website Oct 2002
© Plasma Proteome Institute

(Human male 50yr - from Specialty Laboratories Books)
Plasma “Proteomics” Began With 2-D Gels (c. 1976)

2-D Electrophoresis
300+ resolved spots
40 identified proteins

PPI Website Oct 2002
© Plasma Proteome Institute
# Short History

Most protein technologies have been applied to plasma rapidly.

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1933</td>
<td>Von Mutzenbacher uses Svedberg’s analytical ultracentrifuge to resolve serum albumin and globulin fractions by molecular weight: demonstrates proteins are not heterogeneous colloids but rather specific structures.</td>
</tr>
<tr>
<td>1937</td>
<td>Tiselius uses his electrophoresis to resolve serum proteins into $\alpha$, $\beta$, and $\gamma$ globulins, establishing a naming convention that still persists.</td>
</tr>
<tr>
<td>1939</td>
<td>Tiselius and Kabat demonstrate that antibodies are components of the $\gamma$ globulin fraction.</td>
</tr>
<tr>
<td>1940’s</td>
<td>Svensson and Longsworth further resolve serum globulins into $\alpha_1$, $\alpha_2$, $\beta_1$, $\beta_2$, $\gamma_1$ and $\gamma_2$.</td>
</tr>
<tr>
<td>1950</td>
<td>Gofman finds lipoproteins float in ultracentrifuge and can be measured clinically.</td>
</tr>
<tr>
<td>1950’s</td>
<td>Paper electrophoresis of serum proteins is widely introduced into clinical chemistry.</td>
</tr>
<tr>
<td>1958</td>
<td>Smithies and Poulk resolve 22 zones of serum proteins using starch gel electrophoresis.</td>
</tr>
<tr>
<td>1960</td>
<td>Grabar and Burtin describe immunoelectrophoresis.</td>
</tr>
<tr>
<td>1964</td>
<td>Ornstein and Davis introduce acrylamide ‘disc’ gel electrophoresis with resolution even higher than starch.</td>
</tr>
<tr>
<td>1965</td>
<td>Laurell introduces crossed immunoelectrophoresis, resolving more than 40 different serum proteins.</td>
</tr>
<tr>
<td>1966</td>
<td>Laurell introduces quantitative “rocket” electrophoresis.</td>
</tr>
<tr>
<td>1977</td>
<td>Anderson and Anderson use 2-D electrophoresis to resolve hundreds of serum protein forms (40 identifications via immunoprecipitation).</td>
</tr>
<tr>
<td>1991</td>
<td>Published 2-D plasma protein database with 49 identified proteins and 727 spots.</td>
</tr>
<tr>
<td>2002</td>
<td>Current SWISS-2D PAGE database with 60 identified proteins and est. 1500 spots.</td>
</tr>
<tr>
<td>2002</td>
<td>Immunosubtraction/chromatography/2-DE with &gt;250 identified proteins and est. 1000-1500 spots (Pieper, et al, manuscript in preparation).</td>
</tr>
</tbody>
</table>
Growth in the Number of Protein Species Observed in Plasma Over Time

- Number of Zones/Spots
- Sequence Identified Unique Proteins

- “1-D”
- “2-D”
- “3+-D”

Log10 Number of Resolved Species

Year


40 49 60 250* 490*

* R. Pieper et al, manuscript in preparation
+ J.N. Adkins, in press

PPI Website Oct 2002
© Plasma Proteome Institute
<table>
<thead>
<tr>
<th>Protein/Molecule</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth Hormone Binding Protein</td>
<td>Growth Hormone Binding Protein</td>
</tr>
<tr>
<td>Neuronal Enolase</td>
<td>Neuronal Enolase</td>
</tr>
<tr>
<td>Neuropeptide-activating peptide 2</td>
<td>Neuropeptide-activating peptide 2</td>
</tr>
<tr>
<td>Osteocalcin</td>
<td>Osteocalcin</td>
</tr>
<tr>
<td>Osteonectin</td>
<td>Osteonectin</td>
</tr>
<tr>
<td>Paraoxonase Parathyroid Hormone</td>
<td>Paraoxonase Parathyroid Hormone</td>
</tr>
<tr>
<td>Parathyroid Hormone-Related Protein</td>
<td>Parathyroid Hormone-Related Protein</td>
</tr>
<tr>
<td>PASH and TSP</td>
<td>PASH and TSP</td>
</tr>
<tr>
<td>Pepsinogen A</td>
<td>Pepsinogen A</td>
</tr>
<tr>
<td>Plasma hyaluronic binding protein</td>
<td>Plasma hyaluronic binding protein</td>
</tr>
<tr>
<td>Plasma deficiencies inhibitor</td>
<td>Plasma deficiencies inhibitor</td>
</tr>
<tr>
<td>Plasminogen</td>
<td>Plasminogen</td>
</tr>
<tr>
<td>Platelet Factor 4</td>
<td>Platelet Factor 4</td>
</tr>
<tr>
<td>Pre-alpha trypsin inhibitor</td>
<td>Pre-alpha trypsin inhibitor</td>
</tr>
<tr>
<td>Pregnancy-associated group protein A</td>
<td>Pregnancy-associated group protein A</td>
</tr>
<tr>
<td>Pregnancy-associated group protein A2</td>
<td>Pregnancy-associated group protein A2</td>
</tr>
<tr>
<td>Pregnancy-specific beta-1 glycoprotein 3</td>
<td>Pregnancy-specific beta-1 glycoprotein 3</td>
</tr>
<tr>
<td>Proactin</td>
<td>Proactin</td>
</tr>
<tr>
<td>Prolyl hydroxylase, alpha</td>
<td>Prolyl hydroxylase, alpha</td>
</tr>
<tr>
<td>Prolyl hydroxylase, beta</td>
<td>Prolyl hydroxylase, beta</td>
</tr>
<tr>
<td>Prostahgtandin-H2 D-Isomerase</td>
<td>Prostahgtandin-H2 D-Isomerase</td>
</tr>
<tr>
<td>Prostatic Specific Antigen</td>
<td>Prostatic Specific Antigen</td>
</tr>
<tr>
<td>Protein C, H</td>
<td>Protein C, H</td>
</tr>
<tr>
<td>Protein C, L</td>
<td>Protein C, L</td>
</tr>
<tr>
<td>Protein Z</td>
<td>Protein Z</td>
</tr>
<tr>
<td>P-selectin, soluble</td>
<td>P-selectin, soluble</td>
</tr>
<tr>
<td>Rantactin</td>
<td>Rantactin</td>
</tr>
<tr>
<td>Retinol Binding Protein</td>
<td>Retinol Binding Protein</td>
</tr>
<tr>
<td>S100 protein</td>
<td>S100 protein</td>
</tr>
<tr>
<td>Secretory inulin</td>
<td>Secretory inulin</td>
</tr>
<tr>
<td>Serum Amyloid A</td>
<td>Serum Amyloid A</td>
</tr>
<tr>
<td>Serum Amyloid P</td>
<td>Serum Amyloid P</td>
</tr>
<tr>
<td>Sex Hormone Binding Globulin</td>
<td>Sex Hormone Binding Globulin</td>
</tr>
<tr>
<td>Tetrabenecitin</td>
<td>Tetrabenecitin</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>Thyroglobulin</td>
</tr>
<tr>
<td>Thyroid Stimulating Hormone</td>
<td>Thyroid Stimulating Hormone</td>
</tr>
<tr>
<td>Thyrotropin-releasing hormone</td>
<td>Thyrotropin-releasing hormone</td>
</tr>
<tr>
<td>Thyroxin Binding Globulin</td>
<td>Thyroxin Binding Globulin</td>
</tr>
<tr>
<td>Tissue Factor</td>
<td>Tissue Factor</td>
</tr>
<tr>
<td>Tissue inhibitor of metalloproteinases-1</td>
<td>Tissue inhibitor of metalloproteinases-1</td>
</tr>
<tr>
<td>Tissue inhibitor of metalloproteinases-2</td>
<td>Tissue inhibitor of metalloproteinases-2</td>
</tr>
<tr>
<td>Tissue Plasminogen Activator Inhibitor</td>
<td>Tissue Plasminogen Activator Inhibitor</td>
</tr>
<tr>
<td>TNS-Alpha</td>
<td>TNS-Alpha</td>
</tr>
<tr>
<td>TNS-binding Protein 1</td>
<td>TNS-binding Protein 1</td>
</tr>
<tr>
<td>TNS-binding Protein 2</td>
<td>TNS-binding Protein 2</td>
</tr>
<tr>
<td>Transcobalamin</td>
<td>Transcobalamin</td>
</tr>
<tr>
<td>Transcoritin</td>
<td>Transcoritin</td>
</tr>
<tr>
<td>Transferrin</td>
<td>Transferrin</td>
</tr>
<tr>
<td>Transferrin (asialo-, tau-, beta-2)</td>
<td>Transferrin (asialo-, tau-, beta-2)</td>
</tr>
<tr>
<td>Transferrin Receptor (Soluble)</td>
<td>Transferrin Receptor (Soluble)</td>
</tr>
<tr>
<td>Triacylglycerol lipase (pancreatic)</td>
<td>Triacylglycerol lipase (pancreatic)</td>
</tr>
<tr>
<td>Triprolin I (cardiac)</td>
<td>Triprolin I (cardiac)</td>
</tr>
<tr>
<td>Triprolin L (skeletal)</td>
<td>Triprolin L (skeletal)</td>
</tr>
<tr>
<td>Triprolin T (cardiac)</td>
<td>Triprolin T (cardiac)</td>
</tr>
<tr>
<td>Trypsine (beta-2)</td>
<td>Trypsine (beta-2)</td>
</tr>
<tr>
<td>Tyrosine hydroxylase</td>
<td>Tyrosine hydroxylase</td>
</tr>
<tr>
<td>Urokinase (High MW kidney type)</td>
<td>Urokinase (High MW kidney type)</td>
</tr>
<tr>
<td>Urokinase (High MW kidney type)</td>
<td>Urokinase (High MW kidney type)</td>
</tr>
<tr>
<td>VCAK-1, soluble</td>
<td>VCAK-1, soluble</td>
</tr>
<tr>
<td>Vitronectin</td>
<td>Vitronectin</td>
</tr>
<tr>
<td>WAP-44</td>
<td>WAP-44</td>
</tr>
<tr>
<td>Zn Alpha-2-glycoprotein</td>
<td>Zn Alpha-2-glycoprotein</td>
</tr>
</tbody>
</table>

**A Provisional Plasma Proteome:**

| 289 Proteins Observed in Plasma (Scientific Literature pre-2002) |

Almost none were discovered by proteomics.

Very little published on prediction of secreted proteins from human genome.
>70% of Proteins in Plasma Are Likely To Be In Complexes

Histogram of Masses of 262 Known Proteins in Plasma
(masses of processed subunits computed from sequence)

18% < 20kd
69% < 60kd
Protein Diagnostics

- How does it compare to the discovery trend in proteomics?
- What are the figures on protein tests approved by FDA (CLIA)?
The Many Diagnostics Available Test For A Small Number of Protein Analytes

6,780 FDA-Approved Assays for 117 Different Protein Analytes in Plasma
Assays for Only 10 New Proteins in Plasma Have Been Approved by FDA Since 1993

- Cancer Antigen 19-9
- BNP
- Lipoprotein(a) (Lp(a))
- Transferrin Receptor
- Troponin T, Cardiac
- Troponin-I (Cardiac)
- Interleukin-2 Receptor
- IGFBP-3

Based on the CLIA (Clinical Laboratory Improvement Amendments) Database maintained by the US Food and Drug Administration.

PPI Website Oct 2002
© Plasma Proteome Institute
The Rate of Introduction of New FDA-Approved (CLIA) Diagnostic Protein Analytes Has Decreased to ~Zero
Some Useful Protein Analytes Were Not Even “Identified” By the Standards of Proteomics: CA-125 Ovarian Cancer Marker

- Discovered in 1984
- Defined by monoclonal Ab(s)
- 2000+ publications on clinical use
- Identified with reference to protein sequence only in late 2001:
  - Tethered mainly-extracellular glycoprotein of 1,269,525 Daltons

∴ Proteomics ID standards not necessarily critical for applied use
Expanding the Diagnostic Proteome

• A declining rate of introduction of new protein analytes contradicts the widespread expectation that genomics and proteomics are rapidly advancing non-genetic diagnostics

• Suggests a major problem in translation of basic research into commercial diagnostics

• PPI seeks to identify and overcome these barriers
Some Barriers Impeding a Major Advance in Protein Diagnostics

- Multivariate marker concept: Not accepted
- Individual baseline concept: Not accepted
- Cost per protein analyte (for multi-protein markers): 100x too high
Multivariate Markers

• Available data indicates that multivariate (multi-protein) markers characterize disease states and drug effects better than single markers

• Examples
  – Acute phase proteins in RA, in CV risk, in bacterial vs viral infections
  – CK-MB, Mb, TnI(T) in MI
  – Rodent tox studies of compound classes

• Despite examples and theory, not enough weight of evidence to convince wide audience
A Co-Varying Set of Protein Markers Yields a Disease Index More Sensitive and Robust Than a Single Protein Assay

Relationship Between RA and Typhoid Vaccination Effects on Human Serum Protein Abundances

Multivariate Protein Markers Resolve Drug Mechanisms As They Do Disease States

Data from Quantitative 2-D Gel Studies in Mouse Liver
With Test Panels Involving > 100 Proteins

Peroxisome Proliferator Treatments
A= Control
B= LY163,443
C= LY171,883
D= DEHP
E= Clofibric acid
F= WY14,643
G= Nafenopin

Each symbol represents the liver protein pattern for 100+ proteins in the liver of an individual mouse.

Peroxisome Proliferators: 6 Compounds Compared Over 107 Selected Protein Spots
The effects of peroxisome proliferators on protein abundances in mouse liver.
Anderson, N.L., Esquer-Blasco, R., Richardson, F., Foxworthy, P. and Eacho, P.
Toxicology and Applied Pharmacology, 137, 75-89, 1996.
Plasma Markers: Monitoring Genetic Risk or Current Health Status?

- Published data for different markers shows a wide distribution from almost total genetic control (unvarying levels) to none
  - Ratio of intra-individual to inter-individual CV’s (epidemiology studies)
  - % variation due to genetics (MZ twin studies)
- Both methods show Lp(a) marker level is genetically determined: i.e., appropriate as a risk factor measure (assay needed one time)
- Average proportion genetic is ~50%
- Genetic component >20% suggests patient is best control for marker changes
- Genetic component >80% suggests patient value will not change
Intra-individual vs Inter-individual Coefficients of Variation for 16 Proteins in Plasma

Coefficient of Variation (CV)

- CVi (Individual)
- CVg (Group)

CVi=CVg: Phenotype
CVi<<CVg: Genotype

(Published values from many sources)
Genetic Component of Variation in Abundance of 15 Proteins in Plasma

Proportion of Variation Attributed to Genetics

- C-peptide
- Factor XIIa
- Leptin
- Insulin-like growth factor I
- IGFBP-3
- Insulin-like growth factor II
- Histidine-rich glycoprotein
- Gc (group-specific component)
- IgE
- Factor XIII activity
- Insulin-like growth factor-I
- PAI-1 antigen
- Lipoprotein(a)

Genotype

Phenotype

* = two discordant studies

(Published values from many sources)

PPI Website Oct 2002
© Plasma Proteome Institute
Individual Variation in CRP Over Time:
Stability of Non-Disease Levels Allows Finer Characterization of Disease

CRP Levels Over Time in Three Individuals

Conclusions

- Diagnostic value of plasma proteome measurements substantially enhanced by
  - Use of more protein markers
  - Use of multiprotein panels
  - Use of patient as self-control
- Current data is convincing but sparse
- A series of demonstration studies is needed to affect change in consensus view
- Regulatory issues need debate
- Integration of the above is timely
Plasma Proteome Institute

- **Purpose**
  Expand the range of protein analytes and indications through application of rapid proteomics quantitation systems to sets of well-characterized clinical samples

- **Aims**
  - Promote multivariate protein tests
  - Promote repeatedly sampling of individuals for detection of trends
  - Advance technologies for routine plasma proteome measurement
PPI
The Plasma Proteome Institute

www.plasmaproteome.org