

The Human Plasma Proteome

The Human Plasma Proteome: History, Character, and Diagnostic Prospects

N. Leigh Anderson and Norman G. Anderson

The Plasma Proteome Institute, P.O. Box 53450, Washington DC, USA

Correspondence: N. Leigh Anderson

Tel: ( 301) 728-1451

Fax: (202) 234-9175

Email: leighanderson@plasmaproteome.org

Running title: The Human Plasma Proteome



## The Human Plasma Proteome

### Abbreviations:

ECF, extracellular fluid; CSF, cerebrospinal fluid; MRI, magnetic resonance imaging; CT, computed tomography; 2-DE, two-dimensional electrophoresis; IPG, immobilized pH gradient; SDS, sodium dodecyl sulfate; CDGS, carbohydrate-deficient glycoprotein syndrome; SAA, serum amyloid A; hCG, human chorionic gonadotropin; Ig, immunoglobulin; MS, mass spectrometry; SELDI-TOF, surface-enhanced laser desorption ionization-time of flight mass spectrometry; MALDI-TOF, matrix-enhanced laser desorption ionization-time of flight; ESI, electrospray ionization; MudPIT, multidimensional protein identification technology; FTICR, fourier-transform ion cyclotron resonance; GC/MS, gas chromatography/mass spectrometry; ICAT, isotope-coded affinity tag; TTR, transthyretin; FDA, Food and Drug Administration; CLIA, Clinical Laboratory Improvement Amendments; BNP, B-type natriuretic peptide; ASR, analyte-specific reagent; ROC, receiver-operator characteristic; CRP, C-reactive protein

,

## **SUMMARY**

The human plasma proteome holds the promise of a revolution in disease diagnosis and therapeutic monitoring, provided that major challenges in proteomics and related disciplines can be addressed. Plasma is not only the primary clinical specimen, but also represents the largest and deepest version of the human proteome present in any sample: in addition to the classical “plasma proteins”, it contains all tissue proteins (as leakage markers) plus very numerous distinct immunoglobulin sequences; and it has an extraordinary dynamic range, in that more than 10 orders of magnitude in concentration separate albumin and the rarest proteins now measured clinically. Although the restricted dynamic range of conventional proteomics has limited its contribution to the list of 289 proteins (tabulated here) that have been reported in plasma to date, very recent advances in multi-dimensional survey techniques promise at least double this number in the near future. Abundant scientific evidence, from proteomics and other disciplines, suggests that among these are proteins whose abundances and structures change in ways indicative of many, if not most, human diseases. Nevertheless, only a handful of proteins are currently used in routine clinical diagnosis, and the rate of introduction of new protein tests approved by the US FDA has paradoxically declined over the last decade to less than one new protein diagnostic marker per year. We speculate on the reasons behind this large discrepancy between the expectations arising from proteomics and the realities of clinical diagnostics, and suggest approaches by which protein-disease associations may be more effectively translated into diagnostic tools in the future.

## **INTRODUCTION**

Blood plasma is an exceptional proteome in many respects. It is the most complex human-derived proteome, containing other tissue proteomes as subsets. It is collected in huge amounts (millions of liters) for preparation of protein therapeutic products. It is the most difficult protein-containing sample to characterize, on account of the large proportion of albumin (55%), the wide dynamic range in abundance of other proteins, and the tremendous heterogeneity of its predominant glycoproteins. And it is the most sampled proteome, with hundreds of millions of tubes withdrawn every year for medical diagnosis, making it clinically the most important. Proteins in plasma have been studied since before we knew genes existed.

Having lived through the recent superlatives of the human genome effort(s), and the expectations these generated, it might be thought unwise to use such hyperbole in the more modest world of proteins and proteomics. In fact, the exceptional nature of plasma does not lead us to the sin of self-congratulation, insofar as we are in no imminent danger of completing its analysis, or even of making optimal use of its diagnostic possibilities. At this stage, the combination of extreme analytical difficulty with well-founded hopes for radical improvements in disease diagnosis provides a strong case for increased research effort, and in particular some systematic means of accelerating an exploration that has been in process for many decades while so far yielding only a handful of medically useful nuggets.

Molecular biology, including the genome and proteome projects, is revolutionizing the biological and medical sciences, holding out the promise of both fully understanding and effectively treating all human diseases (1). These projects epitomize the ultimate goal of reductionist biology, which is a complete analysis and description of living systems at the molecular level. In the one case quasi-completed so far (the human genome), billions of dollars were raised, tens if not hundreds of thousands of patents filed, and new large integrated laboratories constructed and operated on a crash basis. And yet this is now generally concluded to have been simply laying the foundation for proteomics, a field that requires completely different technologies, sympathy for a very different sort of molecules, and ultimately a very different scale. We are currently in the phase of seeking shortcuts through proteomics analogous to the

path that shotgun sequencing blazed through the genome, but without any guarantee that one exists.

Against this backdrop it may be useful to take a somewhat broader view than might be expected in a review of one particular proteome. Hence we have attempted to survey the larger context of the plasma proteome, as well as the history and status of efforts to explore and use it medically. Finally, we have indulged in some speculation as to the kinds of efforts needed to reach the next stage in the analysis of plasma and its diagnostic applications.

In what follows we use the term “plasma” to embrace all the protein components of the blood soluble phase (excluding cells), and not as a prescription for a specific sample processing technique. We could have referred instead to the “serum” proteome, but chose plasma because it is in a sense the larger, parent collection from which other related samples are derived.

### **DEFINING THE PLASMA PROTEOME**

In his classic series entitled *The Plasma Proteins* (2), Putnam defined true plasma proteins as those which carry out their functions in the circulation, thus excluding proteins that, for example, serve as messengers between tissues (e.g., peptide hormones) or which leak into the blood as a result of tissue damage (e.g., cardiac myoglobin released into plasma after a heart attack). This functional definition correctly emphasized the fact that proteins may appear in plasma for a variety of different reasons, but it also hints at the fact that different methods and approaches were originally responsible for discovery of these classes. The “proteome” (or “protein index”) concept, which stems from analytical advances promising a broad inventory of proteins in biological samples, suggests instead that we should aim for a general analytical foundation for the plasma proteome as a whole, and later extract functional (diagnostic) utility for various proteins based on results of large scale systematic data collection. Hence in this paper we assume that there is a reason to discover, characterize and routinely measure every protein present in human plasma to the limits of detection. This approach is generating something appropriately called the plasma proteome, and distinct from the plasma proteins.

Elaborating on Putnam’s classification from a functional viewpoint, we can classify the protein content of plasma into the following design/function groups:

## The Human Plasma Proteome

Proteins secreted by solid tissues and which act in plasma. The classical plasma proteins, largely secreted by the liver and intestines. A key aspect of plasma proteins is a native molecular weight larger than the kidney filtration cutoff (~45kd), and thus an extended residence time in plasma (albumin, which is just larger than the cutoff, has a lifetime of about 21 days).

Immunoglobulins. Although the antibodies typically function in plasma, they represent a unique class of proteins because of their complexity: there are thought to be on the order of ten million different sequences of antibodies in circulation in a normal adult.

“Long distance” receptor ligands. The classical peptide and protein hormones. These proteins come in a range of sizes, which may indicate a range of timescales for their control actions (i.e., rapid adjustment with small hormones such as insulin and slower adjustments with larger hormones such as erythropoetin).

“Local” receptor ligands. Cytokines and other short distance mediators of cellular responses. In general these proteins have native molecular weights under the kidney filtration cutoff (and hence relatively short residence times in plasma) and appear to be designed to mediate local interactions between cells, followed by dilution into plasma at ineffective levels. High plasma levels may cause deleterious effects remote from the site of synthesis, e.g., sepsis.

Temporary passengers. Non-hormone proteins that traverse the plasma compartment temporarily on their way to their site of primary function, e.g., lysosomal proteins that are secreted and then taken up via a receptor for sequestration in the lysosomes.

Tissue leakage products. Proteins that normally function within cells but can be released into plasma as a result of cell death or damage. These proteins include many of the most important diagnostic markers, e.g., cardiac troponins, creatine kinase or myoglobin used in the diagnosis of myocardial infarction.

Aberrant secretions. Proteins released from tumors and other diseased tissues, presumably not as a result of a functional requirement of the organism. These include cancer markers, which may be normal, non-plasma accessible proteins expressed, secreted or released into plasma by tumor cells.

Foreign proteins. Proteins of infectious organisms or parasites that are released into, or exposed

## The Human Plasma Proteome

to, the circulation.

Given this variety of classes of protein components, how many “proteins” are likely to be present in plasma? A reasonable calculation could be proposed in three stages. First, assume as a baseline that there is a modest number (say 500) of true “plasma proteins” (class I above), and that each of these is present in 20 variously glycosylated forms (since most plasma proteins are heavily glycosylated) and in 5 different sizes (including precursors, “normal” forms, degradation products, and splice variants), yielding a total of 50,000 molecular forms. A second large set of components is contributed by tissue leakage: this is effectively the entire human proteome (say 50,000 gene products), each having (on average) 10 splice variants, post-translational modifications or cleavage products, yielding a further 500,000 protein forms. Finally consider the immunoglobulin class as containing perhaps 10,000,000 different sequences. At least in principal, plasma is thus the most comprehensive and the largest version of the human proteome. In comparison with the genome, its degree of complexity is reminiscent of the real number line as compared to the integers: in other words its complexity is not simply  $n$ -fold that of the genome, but exists on another level entirely. This immense complexity does not doom current efforts to failure, however, because the measurement methods in many cases automatically simplify the picture, collapsing most of the fine variation to yield measurements of all the forms of a protein as one value, or at most as a few special classes.

Serum, the protein solution remaining after plasma (or whole blood) is allowed to clot, is very similar to plasma: prothrombin is cleaved to thrombin, fibrinogen is removed (to form the clot), and a limited series of other protein changes (mainly proteolytic cleavages) take place. We use the term plasma preferentially to refer to the blood’s soluble proteome because it is the parent mixture, and because there may be persuasive reasons to avoid an *in vitro* proteolysis process (which may unexpectedly alter some proteins) as part of the preferred sample acquisition protocol.

At present we do not know much about the detailed relationship between plasma (the routinely-available sample), the much larger extracellular fluid (ECF) compartment (~17 L in the average person, but practically unsampleable), and the lymph derived from ECF. Roughly 2.5L of lymph flows through the thoracic duct into the blood each day, plus another 500mL through other channels, bringing into the

## The Human Plasma Proteome

blood much of the protein output of organs like muscles or the liver. Even though the total protein concentration of thoracic duct lymph is only about half that of plasma (and must generally be so to support the Starling equilibrium governing fluid transport out of the capillaries), it transports a great deal of protein, and in particular contains 5 to 10 times as much lipoprotein as plasma. A comprehensive examination of the relationship between lymph and plasma by proteomics methods remains to be done.

A series of other body fluids including cerebrospinal fluid (CSF), synovial fluid, and urine (the ultimate destination of most of the <60kd protein material in plasma) share some of the protein content of plasma, with specific local additions that reveal interesting clinical information. Unfortunately, these samples are more difficult to obtain in a useful state than plasma: collection of CSF and synovial fluid are invasive procedures involving pain and some risk, while urine is more difficult to process to a useful sample quickly in a clinical setting (centrifugation to remove cells that can lyse if left in suspension, prevention of microbial growth, and concentration).

Is there only one plasma proteome, or are there many: arterial, venous, capillary, capillary in different tissues, etc.? This question is, in many ways, one of timing. Pharmacokinetic studies indicate that there is a central volume of vascular blood that circulates (and is presumably mixed) fairly quickly: the almost immediate appearance in most organs of MRI or CT imaging contrast agents injected as bolus doses into the venous circulation attests to a very rapid (seconds to a minute) homogenization of this volume of blood. Exchange with the larger volume of blood which is not in the major vessels takes longer, and depends on transport of blood through the whole path of arteries to arterioles to capillaries to venules to veins and back through the heart. This process generally has a timescale of minutes to a few hours, depending on the molecular weight of the protein and on the flow of ECF from the site of manufacture either to a nearby capillary or to the lymphatics. We would thus expect the 'immediacy' of a protein marker in plasma to depend on its site of origin, with timescales of minutes to hours possible for different molecules and sites.

A further important feature of the plasma proteome is that it is the furthest removed, among tissue proteomes, from the mRNA level. While many of the major plasma proteins are synthesized in the liver (and comprise many of its most abundant mRNA's (3)), it is known that their plasma levels correlate

only poorly with message abundance in liver (4), and presumably even more poorly for proteins synthesized in smaller organs (individually or collectively). For these reasons, plasma is a biological system that can only be approached with protein methods, and thus remains beyond the scope of DNA- or RNA-based diagnostics.

## **HISTORY OF SYSTEMATIC EXPLORATION**

The exploration of the plasma proteome can be divided roughly into six phases: 1) the earliest investigations, before the nature of proteins was understood; 2) the era of fractionation (chemical methods); 3) the era of enzymes (biochemical methods); 4) the era of monoclonal antibodies (molecular biology); 5) the era of proteomics (separations technologies); and 6) the era of genomics (predictive proteomics). These approaches overlap in time, and are arranged here in only a loosely historical order.

### **Early History and Chemical Methods (Fractionation)**

Blood was first emphasized diagnostically by Hippocrates, who proposed that disease was due to an imbalance of four humours: blood, phlegm, yellow bile, and black bile. The importance of this idea was to propose a physical cause, and not a divine one, for human disease, and it remained basic to medical practice for over a thousand years. With Wohler's synthesis of urea in 1828, the distinction between living matter and chemicals began to disappear, and with the enunciation of the cell theory by Schleiden and Schwann, the question of the location of disease could be productively revisited: Virchow described the cellular (as opposed to humoral) basis of disease, and finally put an end to phlebotomy as general therapy.

Despite not being a humour or "vital principle", plasma remained a subject of interest throughout this period: Liebig and Mulder analyzed a substance called "albumin" in the 1830's, Schmidt coined the term "globulin" for the proteins insoluble in pure water in 1862, and Gurber crystallized horse serum albumin in 1894 (5). Within the last hundred years, two groups revolutionized plasma protein chemistry. One was the group of Cohn and Edsall working on the preparation and fractionation of plasma, who, during the Second World War, generated large amounts of albumin and gamma globulin for therapeutic use (6). The methods they developed are still employed in the plasma fractionation industry described

## The Human Plasma Proteome

below. The second was the Behring Institute, which, using an unusual rivanol precipitation technique, discovered and prepared numerous human plasma proteins, made antibodies against them, and distributed these world-wide (7). This latter work, and the development of simple immunological methods for analysis, meant that researchers around the world could readily discover new correlations between the amounts of specific proteins and disease.

## Enzyme Activities

Enzyme activities were detectable in body fluids long before the enzyme proteins could be isolated and studied (8). Alkaline and acid phosphatase activities were related to bone disease and prostate cancer, respectively, in the decades before 1950, and in 1955 the enzyme now called aspartate aminotransferase was detected in serum following acute myocardial infarction. The attraction of enzymes as analytes is the sensitivity with which the products of an enzymatic reaction can be detected, and the lack of any necessity to fractionate the sample. Serum “chemistry analyzers” represented some of the first fruits of automation in clinical medicine, and made possible the concept of batteries of tests, rather than one or a very few tests ordered by the astute diagnostician. These evolved from the autoanalyzer developed by Leonard Skeggs in the 1950s (9) and commercialized by Technicon, through computer-controlled instruments such as the centrifugal fast analyzer (10), to the very sophisticated integrated instrument/reagent systems of today. Up to the present time, enzyme assays persist for protein markers of liver toxicity because they are so inexpensive relative to immunoassays (whose development requires development of specific antibody reagents instead of simple chemical substrates).

Enzyme assays have the advantage over all the other assay methods that they measure level of function rather than amount of a molecule: unfortunately many of the enzyme activities measured in plasma probably do not have a physiological function there, but rather represent leakage of protein from tissues. Additional drawbacks of enzymatic assays are: the difficulty of obtaining an estimate of the mass of protein involved (since results are in activity units), the difficulty of associating some activities with a single protein and hence with a specific source, and the lack of isotype information unless some electrophoretic or other separation precedes the enzyme detection. In any case, since a large proportion

The Human Plasma Proteome

of proteins in plasma and elsewhere are not enzymes, alternative means are required to discover and measure them.

## **Antibodies and Monoclonal Antibodies**

All proteins have unique surface shapes, and antibodies are nature's answer to accurate shape recognition. It appears to be possible to make a specific antibody to any protein, provided that pure protein is available to immunize an animal (the inverse of the limitation encountered with enzymes). Proteins purified by fractionation are useful as antigens for the preparation of classical rabbit and goat polyclonal antibodies, and these antibodies provide the basis for simple immunochemical tests for each protein (e.g. radial immunodiffusion, rocket electrophoresis, or more recently automated nephelometry) as well as more complex and sensitive sandwich assays with enzymatic or radiochemical detection. These technologies provide a general solution (as demonstrated by the Behring Institute) to the problem of measuring one or more proteins individually in large numbers of samples, provided that relatively pure protein is available in significant quantities as antigen.

The requirement for an isolated antigen was circumvented, however, following the introduction of monoclonal mouse antibodies by Kohler and Milstein (11) in 1975. The general supposition, confirmed in many situations, is that a monoclonal "sees" a specific epitope that is likely to occur on one protein (or potentially on its very close relatives). The antibody thus serves as a ready-made detection reagent for constructing a specific assay, a potential drug (if the epitope is on a therapeutic target), and an immunoadsorbent useful for isolating the protein from plasma. Many different monoclonals can be produced after immunization with a complex antigen mixture (which can be a tissue or a body fluid), and one can test each to determine if it "sees" an antigen that can distinguish diseased samples from normal ones, for example. The approach can be likened to screening a library of chemicals against a protein target to discover most new drugs: it allows one to be lucky, if not smart.

As a protein (or more properly epitope) discovery process, the monoclonal approach was more useful in the discovery of new tissue antigens shed into plasma than in finding new plasma proteins. The reason is presumably that many immunodominant proteins are present at very high abundance in plasma,

## The Human Plasma Proteome

while tissue extracts are not so rich in a few already-known antigens. Thus, for example, many of the newer monoclonal-based cancer detection tests are used clinically before the underlying protein is identified by sequence or gene. A striking example of this phenomenon is cancer antigen-125 (CA-125), used in the diagnosis of ovarian and other cancers. First reported in 1984 (12), this protein marker is in widespread clinical use and has been the subject of more than 2,000 scientific publications - yet its sequence was not elucidated until recently (13), in part because the protein is huge: more than a million daltons. Such monoclonal-based discoveries exist initially outside the boundaries of the current genomics/proteomics space. When markers of this type are identified, it is sometimes the case that two epitopes (for which there may be competing commercial tests) are on the same molecule. The CA27.29, CA15-3 and CASA assays, for example, all recognize antigenic determinants on the MUC1 mucin protein (14,15). Current attempts to invert the monoclonal antibody generation process (i.e., expecting to make good antibodies to a list of specific proteins, rather than embracing whatever the immune system selects as immunogenic in a mixture) have revealed that antibodies in general are idiosyncratic, and not analogous to oligonucleotide probes in their painless generality.

### Profiling and Proteomics

The use of analytical separations to look at the plasma proteome parallels very closely the development of the separations themselves: plasma is always among the first samples to be examined. Shortly after Svedberg, using the analytical ultracentrifuge, found that proteins had unique molecular weights, Tiselius found that serum could be fractionated into multiple components on the basis of electrophoretic mobility. His method of electrophoresis, first in liquid and then later in anti-convective media such as paper, cellulose acetate, starch, agarose and polyacrylamide has dominated the separative side of plasma proteome work until very recently, evolving through a series of one- and two-dimensional systems, and finally combinations with chromatography and mass spectrometry that generalize to n-dimensions. This evolution has resulted in an almost constant exponential increase in resolved protein species for the past 70 years (Figure 1), within which one can distinguish at least three separate phases arising as increasingly complex separations were required to continue forward progress.

## The Human Plasma Proteome

We do not review in detail the “1-D” phase of this development, but begin with the “proteomics” era and 2-D gels, for 20 years the core of proteomics technology and the source of most published work on the plasma proteome.

\*\*\*\* Suggested location of Figure 1 \*\*\*\*

### 2-D Electrophoresis

Soon after the introduction of high-resolution two-dimensional electrophoresis (2-DE) in 1975 by Klose (16), O’Farrell (17) and others, the technique was applied to the plasma proteins by the present authors (18), with the result that the number of resolved species increased to 300 or more. The 2-DE map of human plasma that resulted is recognizably the same as those produced later by many investigators: in contrast to cellular protein patterns, the plasma 2-DE pattern appears basically the same in everyone’s hands, perhaps due to the very high solubility of the proteins involved and the ease with which the distinctive glycosylation trains of specific proteins can be recognized. A more comprehensive database was reported in 1991 in which 727 spots were resolved, of which 376 were identified as 49 different proteins (19). A plasma map using an immobilized pH gradient (IPG) first dimension separation was presented the following year with 40 protein identifications carried out by microsequencing (20), and this has been extended (on the current Swiss2-D Page website<sup>1</sup>) to identify 60 proteins (613 spots identified, a majority by immunodetection). Given the very high abundance of a few proteins in plasma (albumin, transferrin, immunoglobulins, etc.) and the extreme heterogeneity of plasma glycoproteins and immunoglobulins, this may represent the limit in terms of 2-D gel analysis of unfractionated plasma. By combining 2-DE with additional separation steps, however, this limit can be overcome. Recycling immunosubtraction (21) of up to 14 abundant plasma proteins (22) can remove a significant fraction of the protein mass, yielding a corresponding enrichment of the remaining species. Other classical chromatographic separations, (such as size exclusion, ion exchange, lectin binding and hydrophobic interaction) may be applied to further fractionate plasma, generating hundreds of fractions for individual 2-DE analysis. Using this approach, more than 250 different proteins with unique accession numbers are identified by mass spectrometry relative to a non-redundant protein database (R. Pieper, T. Gatlin, S. Steiner and L. Anderson,

## The Human Plasma Proteome

manuscript in preparation). It is estimated that these 250 different proteins correspond to 1000-1500 distinct post-translationally modified protein spots. Given that more fractions can be generated, this does not seem to represent an upper limit to the number of proteins visible in combined LC/2-DE/MS platforms, though the cost and labor involved become a limitation.

The practical utility of 2-DE for studies of the high-abundance plasma proteome has been substantial. Because the procedure's first dimension (isoelectric focusing) is exquisitely sensitive to molecular charge and the second dimension (SDS electrophoresis) to polypeptide length, it is very effective at revealing genetic variants (about 1/3 of which differ in net charge from wild type), proteolytic cleavages, and variations in sialic acid content. Several genetic variants have been discovered by 2-DE (23-25), and a broad survey concluded that heterozygosity estimates obtained by this method were significantly higher than those obtained in cellular protein samples (26), possibly reflecting a lesser selection pressure against structural change in the highly soluble plasma proteins.

Many proteins in plasma show complex combinations of post-translational modifications (particularly involving glycosylation) that can be discriminated by 2-DE. In the haptoglobin  $\beta$  chain, for example, three heterogeneities (attachment of 0,1,2 or 3 glycan structures; a unimodal distribution of sialic acid content; and two size forms) superimpose to create a complex but interpretable pattern of at least 43 resolved spots (27). The mean number of sialic acids present on plasma protein charge trains like Hp  $\beta$  can be summarized by computing a charge modification index (28). The results indicate that sialation remains relatively constant within a healthy person over time, but differs significantly between individuals, as expected if glycosylation (particularly sialation level) is under genetic control (NLA, unpublished). The analysis of glycoforms was extended by Gravel and Golaz (29,30) using lectins and Western blots to probe the glycosylation of plasma proteins revealed in the 2-DE pattern, a technique that clearly demonstrated the changes in plasma protein glycosylation associated with alcoholism and cirrhosis (31). In patients with the genetic disease carbohydrate-deficient glycoprotein syndrome (CDGS) this expectation is confirmed by a systematic loss of highly glycosylated forms of many proteins (32,33), something so characteristic in the 2-DE pattern and so difficult to see by other means that the technique found its first unique clinical use in the diagnosis of this disease. A similar but less extensive loss of glycosylation (transferrin, haptoglobin  $\beta$  and

## The Human Plasma Proteome

$\alpha$ 1-antitrypsin) occurs with alcoholism (34), while the more serious liver pathology of cirrhosis (31) causes declines in the net amounts of haptoglobin and albumin, the liver's most abundant product.

A series of 2-DE studies have examined aspects of the acute phase response, in which many plasma proteins increase or decrease following a range of inflammatory insults. Changes in the plasma concentrations of acute phase proteins such as serum amyloid A (SAA) have been found associated with severe head injury (35,36), meningitis due to infection by *Haemophilus influenzae* type b (37), normal neonatal development (38), sepsis in newborns (39), and viral infections (40). The latter work by Bini *et al* indicates particularly interesting differences between the body's response to bacterial and viral infection: 18 of 18 bacterial infections gave rise to elevated SAA, while only 6 of 16 viral infections did, and while the acute phase changes in other proteins were simultaneous in bacterial infection, they were staggered in viral infection. Other work showed that the plasma proteome changes associated with rheumatoid arthritis are very similar to those provoked by typhoid vaccination (a classic acute phase inducer in humans), and that reversal of these changes could be used to assess the efficacy of non-steroidal anti-inflammatory drugs (41). These results indicate both the generality of the acute phase response and the power of proteomics to subdivide its features in diagnostically useful ways.

Many other disease states and developmental processes have been examined. Tissot *et al* demonstrated characteristic changes in the plasma 2-DE pattern indicative of monoclonal gammopathies, hypergammaglobulinemia, hepatic failure, chronic renal failure and hemolytic anemia (42), as well as progressive changes during fetal development (43). *In vivo* covalent binding of ampicillin to multiple plasma proteins (not simply albumin) was demonstrated (44), providing a general method to probe covalent drug binding in plasma following drug treatment. Other changes have been reported associated with malnutrition (45), haptoglobin in Duchenne muscular dystrophy (46), haptoglobin in Down syndrome (47), Apo A-I during parturition (48), return from extended spaceflight (49) (possibly an acute phase response to resuming 1g), Apo A-I isoforms in heart disease (50), hCG isoforms in patients with trophoblastic tumors (51), apoE isoforms in chronic spinal pain (52), and oxidized plasma proteins in Alzheimer's disease (53). Important parallel work on the rat plasma proteome and its perturbation by experimental treatments (not reviewed in detail here) has been carried out by Gianazza and colleagues (54,55).

## The Human Plasma Proteome

The immunoglobulins represent a special class of plasma proteins because of the tremendous sequence heterogeneity involved in generating their functional specificities: an estimated  $10^6$  different forms may be in circulation in a healthy human at one time. Individual Ig light chain sequences typically appear as single spots in 2-DE (56), thus allowing up to a few hundred “overabundant” sequences to be inspected as indicators of the activity of individual Ig-producing B-cell clones, while the heavy chains of all classes are glycosylated, adding an additional heterogeneity that prevents their being resolved and used as clonal indicators. Thus the position (pI and MW) of a heavy chain is indicative of its class (18,57-60) (i.e., falling in a band associated with  $\mu$ ,  $\gamma$ ,  $\alpha$ ,  $\delta$ , or  $\epsilon$  chains representing IgM, IgG, IgA, IgD or IgE, respectively) and sometimes its  $\gamma$  chain subclass (61), but each sequence appears as a charge train overlapping others. The work on class properties by 2-DE led to successful clinical applications in detecting and classifying myelomas (clonal B-cell tumors in which a single Ig molecule is vastly overproduced, and where prognosis is related to heavy-chain class) because a simple quantitative test for the Ig class is not sufficient and low resolution methods may not be able to determine if the protein is truly clonal. However to follow the overall clonal development of the immune system it has been necessary to characterize the light chains, whose clonal spots occur in two partially overlapping bands (representing the  $\kappa$  and  $\lambda$  classes). Tissot *et al* observed transient “clonal imbalances” in the light chains of patients after immune system reconstitution by bone marrow transplantation (62), and in normal individuals in the 2<sup>nd</sup> to 4<sup>th</sup> months after birth (63), both of which are consistent with an early expansion of selected B-cell clones in response to a few new dominant antigenic stimuli, followed weeks or months later by the emergence of the fully mature, heterogeneous response to all the antigens of daily life. The same phenomenon occurs when mice are removed from the protection of a germ-free facility and exposed to the antigens and infectious agents of the normal world, and a similar result appears in the Ig patterns of some cancer patients (NLA unpublished). One effort (64) has been made to develop a quantitative measure of the “texture” of the complex 2-DE pattern of light chain spots (classifying patterns as monoclonal, oligoclonal or normal), but the more difficult challenge of tracing the timecourse of multiple clones in an individual over time (likely to be particularly relevant in cancer) has not yet been successfully addressed.

## Mass Spectrometry

Mass spectrometry (MS) has solved the problem of identifying proteins resolved by 2-D gel and other methods, and appears poised to provide general solutions to the analysis of complex protein mixtures as well. In the latter category, two general classes of approach can be distinguished: first, the “unbiased” discovery of proteins and peptides achieved via their detection or identification in a sample, and, second, the quantitative measurement of protein or peptides, usually requiring some type of additional standardization. In many cases these methods are just beginning to be applied to the plasma proteome, but offer sufficient promise for future work that their review here is necessary.

MS Discovery: The power of mass spectrometry techniques to discover proteins in complex samples relies, with one notable exception described below, upon the existence of large protein sequence databases generally derived from DNA sequencing efforts. Since these databases are becoming comprehensive, the approach offers, at least in theory, a general solution to protein discovery. So far MS efforts have examined three basic windows into the proteome problem: whole proteins, peptide fragments obtained by digesting proteins *in vitro* (e.g., with trypsin), and naturally occurring peptides (the low molecular weight proteome, or peptidome).

Whole proteins can be analyzed by an approach termed SELDI-TOF (for surface-enhanced laser desorption ionization-time of flight) mass spectrometry, a variant of MALDI-TOF (matrix-assisted laser desorption ionization-time of flight), in which chemical fractionation based on protein affinity for derivatized MS targets is used to reduce sample complexity to a level at which whole-protein MS can resolve a series of fairly reproducible features. This approach has been implemented commercially (e.g., by Ciphergen and Lumicyte) and used to discover patterns of protein features associated with disease (e.g., the ovarian cancer markers discovered by Petricoin et al (65)). A significant disadvantage of the approach is that MS analysis of whole proteins does not directly provide a sequence-based identification (there being many proteins with close to a given mass), and hence the protein peaks discovered as markers are not strictly-speaking identified without significant additional effort. In particular, without a discrete identification, it is not generally possible to demonstrate that a peak is one protein analyte, nor to translate the measurement into a classical immunoassay format. However, as has been clearly

## The Human Plasma Proteome

demonstrated by the success of monoclonal-based assays, this does not pose a significant limitation to clinical use if the technology allows the analysis to be repeated in any interested laboratory (an effort which now appears to be underway).

A more general approach involves digesting proteins (e.g., with trypsin) into peptides that can be further fragmented (MS/MS) in a mass spectrometer to generate a sequence-based identification. The approach can be used with either electrospray ionization (ESI) or MALDI, and is typically applied after one or more dimensions of chromatographic fractionation to reduce the complexity of peptides introduced into the MS at an given instant. Optimized systems of multidimensional chromatography, ionization, mass spectrometry and data analysis (e.g., the multidimensional protein identification technology, or "MudPIT" approach of Yates, also referred to as shotgun proteomics) have been shown to be capable of detecting and identifying ~1,500 yeast proteins in one analysis (66), while a single dimensional LC separation, combined with the extremely high resolution of a fourier-transform ion cyclotron resonance (FTICR) MS identified more than 1,900 protein products of distinct open reading frames (i.e., predicted proteins) in a bacterium (67). In human urine, a sample much more like plasma than the microbial samples mentioned above, Patterson used a single LC separation ahead of ESI-MS/MS to detect 751 sequences derived from 124 different gene products (68,69). Very recently, Adkins et al have used two chromatographic separations with MS to identify a total of 490 different proteins in human serum (70), thus substantially expanding the proteome. Such methods should have the ability to deal with the numerous post-translational modifications characteristic of many proteins in plasma, as demonstrated by its ability to characterize the very complex post-translational modifications occurring in aging human lens (71) (73 modified sites characterized in 11 crystallins). The sensitivity obtainable in such an analysis has been tested by Wu et al (72) using human growth hormone spiked into human plasma at a concentration of 16 ng/ml. Using only a reverse phase separation to resolve a tryptic digest of whole spiked plasma, a single hGH peptide was observed (among 200+ proteins apparently identified), and with additional fractionation processes, additional hGH peptides were seen, confirming its detection. .

Naturally-occurring peptides, typically below the kidney size cutoff and hence usually collected from urine or from blood hemodialysate, provide a complementary picture of many events at the low-

## The Human Plasma Proteome

mass end of the plasma proteome. Thousands of liters of human hemodialysate can be collected from patients with end stage renal disease undergoing therapeutic dialysis (73), and even though it contains only 50ug/ml of protein/peptide material, it provides a large-scale source of proteins and peptides below 45kd.

Such material has been analyzed by combined chromatography and MS approaches to resolve approximately 5,000 different peptides, including fragments of 75 different proteins (74). Fifty-five percent of the fragments were derived from plasma proteins and 7% of the entries represented peptide hormones, growth factors and cytokines.

MS Quantitation: The well-known idiosyncrasies of peptide ionization, which have been major impediments to accurate quantitation by mass spectrometry, can be overcome through the use of stable isotope-labeled internal standards. At least four suitable isotopes ( $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{18}\text{O}$ ) are commercially available in suitable highly enriched (>98atom%) forms. In principle, abundance data as accurate as that obtained in MS measurement of drug metabolites with internal standards (coefficients of variation <1%) should ultimately be obtainable. In the early 1980's  $^{18}\text{O}$ -labeled enkephalins were prepared and used to measure these peptides in tissues at ppb levels (75), and in the 1990's GC/MS methods were developed to precisely quantitate stable isotope-labeled amino acids, and hence protein turnover, in human muscle (76), (77) and plasma (78,79) proteins labeled *in vivo*. The extreme sensitivity and precision of these methods suggested that stable isotope approaches could be applied in quantitative proteomics investigations, given suitable protein or peptide labeling schemes.

Over the past two years, a variety of such labeling strategies have been developed. The most straightforward approach (incorporation of label to a high substitution level during biosynthesis), has been successfully applied to microorganisms (80,81) and mammalian cells in culture (82), but is unlikely to be usable directly in humans for cost and ethical reasons. A related approach (which is applicable to human proteins) is the now-conventional chemical synthesis of monitor peptides containing heavy isotopes at specific positions. Post-synthetic methods have also been developed for labeling of peptides to distinguish those derived from an "internal control" sample from those derived from an experimental sample, with a labeled/unlabeled pair subsequently being mixed and analyzed together by MS. These methods include Aebersold's isotope-coded affinity tag (ICAT) approach (83), as well as deuterated

## The Human Plasma Proteome

acrylamide (84) and N-alkylmaleimides (85) for labeling peptide sulfhydryls, deuterated acetate (86) to label primary amino groups, N-terminal-specific reagents (87), permethyl esterification of peptides carboxyl groups (88), and addition of twin  $^{18}\text{O}$  labels to the C-terminus of tryptic peptides during cleavage (89,90).

In the realm of the plasma proteome, as noted above for MS discovery approaches, the principle applications to plasma have occurred in commercial companies and not yet been published.

Nevertheless we conclude that MS quantitation methods are mature enough to successfully measure tens to hundreds of proteins in plasma on an experimental basis once suitable standards are available.

### Antibody Arrays

Antibody microarrays were developed by Ekins (91) in the mid-1980's (i.e., before DNA microarrays) for measuring proteins in plasma, thus miniaturizing as "microspots" the well-established immunoassay concept. At the time of these developments, the number of proteins of clinical utility was relatively small. Now however, with the expected proliferation of useful protein analytes, there is renewed interest in the concept in a variety of physical formats (recently reviewed in (92)). Detection sensitivity can be very good, as demonstrated by Schweitzer *et al* (93) who recently showed that 75 different cytokines could be measured (using glass-arrayed monoclonal capture antibodies, polyclonal second antibodies, and rolling circle amplification) with sensitivities ranging from 1 to 1,000 pg/mL. This suggests that such systems could be used to measure all of the known plasma proteome components for which high affinity antibody pairs are available. The critical issue is the latter: the generation of antibodies suitable for high sensitivity immunoassays. Even given the resources of large clinical diagnostics organizations, the generation of such antibodies and assays using them is a painstaking process, considerably more difficult than getting an antibody usable in most laboratory research. The extensive effort devoted in the past to making good cytokine antibodies accounts for the fact that most antibody array demonstrations focus on this application, and also indicates why antibody arrays have not so far made a major contribution to the protein discovery in plasma. Nevertheless, as the repertoire of suitable antibodies expands, microarrays provide the most likely path to low-cost, routine measurement of large numbers of plasma markers required for an impact on medical practice.

## The Human Plasma Proteome

Arrays of large numbers of expressed antigens have also been used (94) very effectively to define populations of autoantibodies present in human plasma. This approach improves on the classical gel Western blot by presenting all the antigens in a more native state and at approximately equal abundances, thus enabling a much more informative picture of the immune system's activity.

### The Plasma "Interactome"

As noted below, almost half the known proteins in plasma are too small to persist there as monomers, and many of the larger proteins are known to exist in complexes as well. Hence it is likely that a large majority of plasma proteome exists in multi-protein complexes. In addition there are in plasma a significant number of proteins defined by their ability to bind others and thereby affect function (e.g., the TNF-binding proteins, insulin-like growth factor-binding proteins, etc.). Such interactions have been studied using a variety of classical and modern techniques. The most well-known particulate complexes of plasma are the plasma lipoprotein particles, whose protein components have been enumerated by 2-DE (95). An alternative approach, in which 2-DE-resolved plasma proteins are stained with a specific protein ligand has been used to characterize and extend the set of insulin-like growth factor-binding proteins of rat serum (96). Recently elegant and systematic methods have been devised for enumerating protein complexes in yeast using a combination of molecular tags to capture individual complexes followed by LC/MS/MS (with or without a gel separation step) to analyze them (97). These methods are not directly applicable to plasma (since they require alteration of a gene to introduce the peptide tag during biosynthesis), but specific antibodies can substitute for such tags in many cases, used either for "pulldown" of the complex or binding it to a sensor surface. In one successful example of the latter approach, immobilized antibodies to transthyretin (TTR) captured both TTR and its interaction partner retinol binding protein, as indicated by MALDI-MS of the bound material (98). While the problem of distinguishing possible antibody cross-reactions from weak associations of the target proteins will always need to be addressed, it is nevertheless likely that a systematic plasma interactome will be built using antibodies to define the associations between the hundreds of proteins detected.

## Genomics

## The Human Plasma Proteome

It may be easier to search for and measure a predicted protein at very low abundance in plasma (e.g., via a sensitive immunoassay) than to detect it using an unbiased discovery approach. Hence there is a value in predicting, if possible, all the secreted human proteins as candidates for such a quantitative approach. We expect the *in silico* analysis of human genome sequences to contribute a significant number of previously unknown candidate proteins to be sought in plasma. Several efforts have been made to identify secreted proteins via signal sequence features (either including or excluding the extracellular domains of appropriate transmembrane proteins), and these have variously generated 1,000 (99) to 9,000<sup>2</sup> potential secreted proteins. To date, however, this information has been viewed primarily as a potential commercial source of therapeutic targets, and its relevance to expansion of the plasma proteome from a measurement perspective awaits public release of detailed data.

### **THE KNOWN PLASMA PROTEOME**

In an attempt to survey the currently known components of the plasma proteome, we examined published sources from the literature on electrophoresis, enzymology, clinical diagnostics, as well as public sequence databases. So far this search has produced evidence that almost 300 proteins have been detected in plasma (or serum) to date (Table 1). We have included each subunit of multi-subunit proteins and one or more fragments of a protein where these may have separate biological significance. We did not attempt to tabulate alternate glycoforms (except for asialotransferrin, which is measured clinically) or splice variants (except for fibrinogen alpha), and have generally excluded peptides below approximately 5kd. In the immunoglobulins, we distinguished only the major classes and chains.

\*\*\*\* Suggested location of Table 1 \*\*\*\*

Among the issues involved in tabulating the proteins of plasma are the occasional disparity between the names used at one time or another (what for example would the protein once called alpha-1-T glycoprotein be called now, assuming it was a single protein?), and the difficulty of identifying proteins discovered through epitopes and antibodies. Further work will be required to fully connect the names and properties of these and other proteins, and to develop a better framework for tracing the relationships between precursors, mature forms, proteolytic fragments, and post-translationally modified

## The Human Plasma Proteome

forms. Nevertheless, this list does give a good impression of the breadth of potential disease indicators known to be present in this clinical sample.

One of the revealing features of this collection of proteins is its molecular weight distribution (Figure 2), which shows that a substantial fraction of proteins in plasma (just over 50%) are smaller than the presumed size cutoff of the kidney filtration apparatus (~45kd). This suggests that, to persist in plasma without rapid loss through the kidney, most of these proteins must be part of larger protein complexes or else subject to specific retention mechanisms.

\*\*\*\* Suggested location of Figure 2 \*\*\*\*

## DYNAMIC RANGE OF PROTEINS IN PLASMA

### Scope of Proteins Currently Measured

The quantitative dynamic range currently accessible in the plasma proteome can be assessed by plotting the clinical “reference interval” associated with a sampling of proteins measured in plasma by validated diagnostic assays in commercial clinical laboratories, as shown in Figure 3. At the high abundance end, serum albumin (normal concentration range 35-50 mg/mL, or  $35\text{-}50 \times 10^9$  pg/mL, resulting from ~12 g daily synthesis in the liver and a half-life of about 21 days (100)) is measured clinically as an indication of severe liver disease or malnutrition (101). At the low abundance end, interleukin 6 (IL-6: normal range 0-5pg/mL) is measured as a sensitive indicator of inflammation or infection (101). These two clinically useful proteins differ in plasma abundance by a factor of  $10^{10}$ , and the fact that both can be measured in the hospital laboratory is a dramatic demonstration of the power of current immunoassay technology. Finding a molecule of an analyte present at 10pg/mL among the albumin molecules present at 55mg/mL is like finding one individual human being by searching through the entire world’s population (1 in 6.2 billion).

\*\*\*\* Suggested location of Figure 3 \*\*\*\*

It is sobering to recognize that this immense dynamic range is achieved only by technologies that presuppose the identity of the analyte: available methods for unbiased protein discovery, such as 2-DE

## The Human Plasma Proteome

or LC-MS/MS, have typical dynamic ranges of only  $10^2$  to  $10^4$ , short of the requirement for comprehensive proteome mapping by some six to eight orders of magnitude. Current approaches to the extension of this range combine independent fractionation methods (principally chromatography, immunoaffinity subtraction, preparative IEF, or precipitation) with 2-DE or MS to gain an additional factor of 10 to  $10^2$ , which, though very productive, is still far short of the desired range.

There can be a significant division of opinion over whether the low abundance plasma components are more “interesting” or clinically meaningful than those of high abundance (or vice versa). In fact we know that proteins at all abundance levels prove to be useful, as evidenced by the existence of clinical assays for all those plotted in Figure 3. A more important question is where are the proteins in various abundance classes coming from, and how many of them are there? Here the clinically measured proteins seem to be divided into three major classes: plasma proteins at the high abundance end, tissue leakage proteins in the middle and cytokines at the low abundance end.

Tissue leakage proteins are important because a serious pathology can be detected in a small volume of tissue by measuring release into plasma of a high-abundance tissue protein. Cardiac myoglobin (Mb) is present in plasma from normal subjects at 1-85 ng/mL, but is increased to 200-1,100 ng/mL by a myocardial infarction (102), and up to 3,000 ng/mL by fibrinolytic therapy to treat the infarct (103). The last figure perhaps best represents a complete release of highly soluble (>99% cytosolic (104)) Mb as a result of clearing blocked vessels perfusing the damaged tissue volume. Given an average blood volume of 4.5 L in a 70kg male (105), and an average volume proportion of plasma in blood of 55%, there is approximately 2.5 L of plasma in the average person, containing roughly 250g of plasma protein. A level of 3.0 ug/mL Mb is equivalent to the release and dilution of 7.5 mg of Mb into the whole plasma volume, which, since Mb accounts for 0.28% of total cardiac protein (106), would be equivalent to complete release of Mb from about 2.7g of cardiac tissue. This is clearly an underestimate, since moderate infarct masses are typically ~45g (107), and thus it is probable that something like 5% of the releasable Mb is in circulation at the relevant time point (a marker recovery efficiency of about 5%). A protein such as troponin-T, of which only 8% is cytosolic (104), is released at a 300-fold lower level (~10 ng/mL following reperfusion therapy for MI (108)), indicating a much lower marker recovery efficiency. In future, all tissue proteins will be candidate damage markers,

The Human Plasma Proteome

and they will be sought at even lower abundance levels, raising important questions about the normal background levels of tissue destruction and remodeling in healthy individuals.

Cytokines, which in general act locally (at the site of infection or inflammation), are probably not active at their normal plasma concentrations (or even at the higher levels pertaining after a major local release) because they are diluted from  $\mu\text{L}$  or  $\text{mL}$  volumes of tissue into 17L of interstitial fluid. Hence they are in a sense leakage markers as well, though their presence in plasma does not indicate cell breakage.

The question of sensitivity ultimately required for plasma proteome measurement is not easy to answer. The theoretical limit is clearly calculable as a single protein molecule (of e.g., 60kd) in a practically analyzable plasma sample (e.g.,  $\sim 1\text{mL}$ ): this is  $10^{-7}$  pg/mL ( $\sim 1$  zeptomolar), or 7 to 8 orders of magnitude more sensitive than current immunoassay technologies. A technology to make such measurements would require a dynamic range of more than  $10^{17}$  (between the analyte and albumin), exceeding even the capabilities of accelerator mass spectrometry (109) and approaching some hard physical limits. Nevertheless it is clear that sub-pg/mL sensitivities (low femtomolar concentrations and below) will be needed to allow measurement of many of the tissue proteins that might leak into the plasma. Whether this is compatible with a simultaneous capability to measure many proteins remains to be seen.

### **PLASMA PROTEINS THERAPEUTICS**

Although in this review we focus on the measurement and diagnostic potential of the plasma proteome, it is important to realize that plasma is collected, handled and used on a larger scale than any other material arising in human biology – a scale approaching that of agriculture or brewing. Over 22 million liters of plasma are used worldwide by plasma fractionators each year to generate human therapeutic proteins<sup>3</sup>. The major products prepared, in terms of amount, are human serum albumin and intravenous immunoglobulin, of which, respectively, approximately 50,000 kg and 25,000 kg are distributed in the US each year by six major manufacturers<sup>4</sup>. These amounts dwarf the production levels of recombinant therapeutic antibodies such as Genentech's Herceptin® monoclonal for breast cancer

The Human Plasma Proteome

(produced in CHO cells), of which something like 50 kg will be sold this year.

Additional plasma products include Alpha<sub>1</sub> Protease Inhibitor (a treatment for emphysema caused by genetic deficiency), Factor VIII concentrate (for prophylaxis and treatment of hemophilia A bleeding episodes and von Willebrand disease), Anti-Inhibitor Coagulant Complex (for treatment of bleeding episodes in presence of Factor VIII inhibitor), Anti-Thrombin III (for prevention of clotting and thromboembolism associated with liver disease and anti-thrombin III deficiency), Factor IX Complex (for prophylaxis and treatment of hemophilia B bleeding episodes), Factor XIII (for prophylaxis and treatment of bleeding episodes due to Factor XIII deficiency), Fibrin Sealant (which helps to heal wounds following surgery), as well as a series of specific immune globulin products (e.g., hepatitis B, chicken pox, measles, rabies, tetanus, vaccinia, hepatitis A and cytomegalovirus). In principle, many other components of the plasma proteome could become useful therapeutic products produced by similar methods.

### **PLASMA DIAGNOSTICS**

The attraction of plasma for disease diagnosis lies in two characteristics: the ease by which it can be safely obtained, and the fact that it comprehensively samples the human phenotype – the state of the body at a particular point in time. While other sample types can be obtained (saliva, tears, urine, skin, hair, etc.), each is either a small subset of plasma or else a restricted local sampling of cellular activity. Genetic markers, by virtue of the relative constancy of the genome throughout the cells of the body, can be sampled almost anywhere (skin, lymphocytes, various tissues) with equivalent results, but such markers reveal only the genotype, and hence do not reveal anything of the regulation of biological processes in response to lifestyle, disease, or drugs that is expressed at the mRNA and protein levels. While the mRNA does reveal regulation, it is routinely obtained for clinical purposes only from blood lymphocytes (which are themselves limited in responsiveness), and it is not well correlated with protein expression (3) (and hence function).

The use of plasma and serum for disease diagnosis is thus an obvious approach, and one that has been undertaken with some success for many decades (110). Nevertheless, the potential of the “direct” approach to plasma proteome diagnostics (finding and measuring a single protein that is unambiguously

associated with a disease) seems to be waning. This is both a challenge and an opportunity, requiring us to enquire carefully into the successes and the failures of existing diagnostics, as well as the biological foundation of the protein marker approach.

## The Decline in New Protein Tests

Effective diagnostic tests for proteins in plasma, of which there are many, ultimately appear as commercial products sold to clinical laboratories worldwide. Since 1993, a total of just under 7,000 assays applicable to the determination of 117 proteins or peptides in plasma have been approved by the Food and Drug Administration (FDA) under the Clinical Laboratory Improvement Amendments (CLIA) that govern medical use of clinical tests in the USA. Most of these assay products were in existence before the CLIA regulations came into effect, and since 1994 only ten new protein analytes have been added (Figure 4). Since 1998, only two new protein analytes have been the subject of such an FDA approval: B-type natriuretic peptide (BNP) for use in diagnosing congestive heart failure and a cancer-related epitope (Cancer Antigen 19-9) recognized by a monoclonal antibody. It is true that not all new protein diagnostic tests are implemented as machine-based, FDA-approved assays: some are implemented through “analyte-specific reagents” (ASR’s, typically antibodies) sold to clinical laboratories for implementation of in-house tests. However, since a test based on ASR’s must be labeled on the laboratory result report as “not cleared or approved by the FDA”, they represent a very modest proportion of clinical diagnostics, and the FDA does not provide a public database listing ASR analytes. Thus we believe it is correct to conclude that the pace of introduction of new protein markers through full FDA approval has substantially slowed in recent years (Figure 5), reminiscent of the similar decline in introduction of new therapeutic drug classes. This picture is clearly at odds with our expectation that genomics and proteomics are transforming the clinical landscape through diagnostic application of knowledge on large numbers of new proteins. What accounts for this huge discrepancy?

\*\*\*\* Suggested location of Figure 4 \*\*\*\*

\*\*\*\* Suggested location of Figure 5 \*\*\*\*

Aside from important issues of economics and medical practice, which we return to later, there

## The Human Plasma Proteome

may be good scientific reasons why protein diagnostics are not making more rapid progress. In particular it is likely that human biology is more complex than the simple test paradigm used in current diagnostic practice: a change in the concentration of a single protein analyte is expected to unambiguously diagnose a single disease. This notion, similar in many ways to the expectation that effective drugs could be designed to hit a plethora of novel targets with no side effects, may need substantial revision to take into account the true complexity of biological systems. Below we review the current paradigm for interpreting protein measurements, where the non-disease variation (“noise”) comes from, and two strategies for improving signal-to-noise in the detection of disease: elimination of the effects of genetic differences between individuals, and utilization of multi-protein disease markers instead of single protein analytes.

### **Interpretation of Diagnostic Protein Measurements: What is Considered Abnormal?**

The utility of a protein measurement is currently determined by the definition of what is and what is not in the “reference interval” (a term adopted as a replacement for the sometimes-ambiguous “normal range”). It is conventional practice in clinical chemistry to assign the reference interval based on an analysis of the distribution of measurement results obtained by analyzing a large number of samples from people that satisfy some established criteria making them “normal” with respect to the test in question. The reference interval is typically set as the range (about 2 standard deviations above and below the mean value) within which 95% of the results from the “normal” population lie: 2.5% of the reference population will be classified as above the normal range and 2.5% below.

The concept that an individual’s plasma level of a particular protein should be within  $\pm 2$  standard deviations of the reference population’s mean value is obviously of limited value, a fact which has been recognized by clinical chemists for more than 20 years (111,112). Because of the differences among normal people (due to genetics, age, sex, and life history including disease) there are likely to be significant differences in the normal, healthy levels for many individuals. This has two effects: it broadens the population’s distribution of values (and thus requires larger variations to be treated as significant) and it misclassifies outliers who are healthy. A simple reference interval cutoff will misclassify many of these as abnormal when they are normal for themselves, and miss many abnormal values when disease shifts

## The Human Plasma Proteome

a normal outlier into the population normal range. An example of this kind of data is shown in Figure 6 using published data for C-reactive protein measured every three weeks for six months in a number of individuals (113).

\*\*\*\* Suggested location of Figure 6 \*\*\*\*

A more sophisticated approach, but one which still compares an individual to external populations, is to assign a protein level that best distinguishes people who do not have a specific disease from those that are known to have it (cases vs controls). Using this approach one can obtain a specificity (proportion of non-diseased samples correctly assigned as non-diseased) and a sensitivity (proportion of diseased samples correctly assigned as diseased) for a test, trying obviously to get both as high as possible. These values rarely reach 100% in either case, and some widely used tests have values in the neighborhood of 60%. In a test for serious disease, a failure in specificity is especially important as it may generate tremendous concern among those who receive an incorrect positive result, rendering the net benefit of the test and its commercial success questionable. A more direct evaluation of case-control results can be obtained using “receiver-operator characteristic” (ROC) curves, which plot the false-positive rate against the true-positive rate, compressing the results obtained on all the case and control samples into one graph (114).

There is also a significant theoretical problem with the notion that there should be a single protein in plasma whose level changes in response to one specific disease. If the disease is caused by the protein abnormality, as in insulin-dependent diabetes, the reasoning is clear. However for multifactorial diseases, such as atherosclerosis, or diseases whose cause is likely to be tissue-based, such as osteoporosis, it seems improbable that there should be a very specific single marker.

## Contribution of Genetics to the “Noise” in Protein Measurements, and Its Removal

Many influences affect the concentration of proteins in plasma besides disease, and if these are not taken into account, the detection of medically meaningful changes will be more difficult. Influences can be considered as 1) genetic, 2) non-genetic other than medical treatment, 3) related to medical treatment, such as drug therapy, and 4) related to sample handling. The last is typically a small effect

## The Human Plasma Proteome

since methods of venipuncture typically used in studies of plasma proteins are well standardized, and a recent report demonstrates that extended storage at  $-70^{\circ}\text{C}$  preserves structure and activity of plasma proteins tested (115).

The effects of differing genetic constitutions is evident both at the population level (e.g., race (116,117) differences in plasma protein abundances) and the individual level. Twin studies, in which quantitative protein measurements are compared within and between monozygotic twin pairs, have been carried out for a small number of individually assayed plasma proteins, as well as by simple one-dimensional electrophoresis (118) (Table 2). In aggregate these studies show that 12% to 95% of the quantitative variation in specific plasma protein levels is genetic in origin, with an average of 62% for the proteins shown. Gel patterns of plasma proteins from monozygotic twins are quantitatively almost indistinguishable, whether analyzed by one-dimensional (118) or two-dimensional approaches.

\*\*\*\* Suggested location of Table 2 \*\*\*\*

This level of genetic control is consistent with the observation made frequently in epidemiological studies and reference value investigations that an individual's level of a specific protein can be quite stable over time, showing less variation than the population variance. From a series of published studies (Table 3) the average intra-individual CV (an individual sampled over time) for a series of proteins is ~23% and the average inter-individual CV is ~45%, suggesting that on average the individual's CV may be half that of the population (119-123). The CRP data in Figure 6 provides a clear example of this relative individual constancy, where the average coefficient of variation (CV) for an individual's repeated samples is 37%, while the population overall CV is 94%. In the published studies examined, only one protein (lipoprotein(a)) was examined by both genetic and intra-CV vs inter-CV methods, and in both cases it showed strong evidence of genetic control of plasma abundance.

\*\*\*\* Suggested location of Table 3 \*\*\*\*

Non-genetic, non-treatment effects on plasma protein concentrations are extremely varied. It is known, for example, that the concentration of total protein in plasma varies by up to 10% depending upon the posture of the subject: when standing, the higher arterial pressure forces liquid out of the capillaries and the blood volume is about 10% less than when one lies down (the change occurring over 10-30

## The Human Plasma Proteome

minutes). Since this effect relates to a change in blood volume, it is likely to affect only absolute rather than relative protein concentrations. Similarly, there is a significant and progressive increase in the concentration of proteins with time after venous occlusion (124) by the tourniquet which is always applied before venipuncture, as well as differences between the first and last tube of blood drawn (particularly in proteins related to hemostasis). Other influences known to affect protein composition of plasma are bed rest (short term increase, long term decrease in plasma protein); exercise (125,126); time of day (127,128); smoking (129) four major periods distinguished (newborn, childhood to puberty, adult, and elderly adult); gender (127) (differences generally appearing post-puberty); season (131); diet (effects of both protein (132) and fat (129) content); sleep (133), and weightlessness during spaceflight (134,135). The difference between venous blood (the usual source) and lancet-derived (“finger-stick”) blood (which is a mixture of ECF and plasma) has received little attention, but may evolve into an important issue if serial samples from an individual are to be obtained conveniently. Disposable lancet technology (and the laser puncture equivalent) have been extensively developed for convenient blood sampling by diabetics, often several times per day, and the volumes obtained (from <1 up to 100uL) may be sufficient for some proteomics approaches.

The effects of drugs on protein levels have also been examined extensively, and tabulated periodically by Young (136,137). The effects observed can be more or less direct, as when an intramuscular drug injection causes an increase in muscle enzymes in plasma through leakage or localized tissue toxicity, or when a diuretic alters total protein concentration by changing blood volume. Alternatively they can result from regulatory changes in target organs, as when liver enzymes are elevated (a side effect associated with a substantial proportion of commonly prescribed drugs, albeit in only a subset of patients). Such effects of drugs should not be thought of as noise because they represent real biological effects, and because these effects may (indeed should) be related (ideally inversely) to the disease effects they are meant to counteract.

Of these influences, genetic effects on protein abundance in plasma are usually the largest and in many ways the least useful from the individual’s point of view. Effects of environment or drug administration are either small or else useful as signals of medically relevant events. Taken together, these considerations suggest that the most productive approach to avoid the difficulties of the classical

## The Human Plasma Proteome

reference range problem is to use the individual as his or her own reference, comparing measurements taken over time in order to see changes associated with disease or treatment. Such an approach would eliminate the genetic effects on protein levels, and go some distance towards minimizing many of the non-genetic influences as well, since most life-style changes are made slowly. In particular, it appears clear that individual reference intervals could be decreased on average about 50% if a self-standardized approach were taken, thereby potentially improving the detection of many disease states at an earlier stage.

### Multi-Parameter Indicators

An additional improvement, both in terms of statistical power and breadth of coverage, is to use many protein measurements instead of one or a few. Several streams of scientific effort have generated data supporting this conclusion. More than 20 years ago it became clear that different tumor cell types could be distinguished based on patterns of metabolites analyzed by GC-MS (138), and that a panel of biochemical markers (most of them simple low molecular weight clinical analytes) could “recognize” individuals within a group over periods of years (139) when analyzed by appropriate multivariate statistical methods. There were efforts to use the latter approach to detect disease signatures in then-standard 20-analyte serum chemistry panels, but these met with little success, probably due to the character and small number of the analytes.

Once specific protein measurements were available, a veritable industry was created exploring their correlation with various diseases and treatments. Thousands of scientific papers explore these correlations, but the number of protein analytes employed is typically less than five (often one), and usually one disease or treatment is examined. Particularly valuable are the large epidemiological studies in which a number of proteins have been measured with high precision in large numbers of samples (140).

Proteomics, in the early stages of its use of two-dimensional electrophoresis, also focused on multivariate effect fingerprints. In the early 1980's insect developmental changes (141), mammalian cell type classification (142) and chemical effects on protein expression patterns (142) were examined using multivariate approaches with clustering of large panels of protein markers, methods derived from molecular taxonomy.

## The Human Plasma Proteome

Later, similar methods were applied to analyze the effects of drugs and other toxic agents in rodent tissues, particularly liver. These studies examined the relationships between the effects of compounds within a mechanistic class (143,144), and showed that a mechanism outside the known class could be recognized based on a multivariate protein pattern change (145). Quantitative serum protein proteomics was used to show that a panel of high-abundance acute phase-related proteins could give a better statistical measure of inflammation than the classical marker serum amyloid A (41). Cancer tissue samples could be analyzed to distinguish tumor type and prognosis (146) and a panel of six cancer markers in plasma was found useful (147,148). More recently, mass spectrometry-based proteomics approaches have been used as well to discover patterns of disease-related protein features related to a specific cancer (65).

The concept and utility of multivariate protein markers has thus been established for some time. What requires comment is why this approach has not penetrated significantly into clinical practice. To some extent the barrier has been the difficulty of demonstrating the clinical superiority of multi-analyte measurements in diagnosing and treating disease, attributable in part to the fact that a workable clinical test (something much more robust than 2-D gels (149), for example) is usually required before this demonstration can occur. Each step in the required progression from discovery to routine implementation has been impeded by the technical complexity of multi-protein measurements. While proteomics can demonstrate and sometimes measure many proteins, the techniques required have been difficult to apply to a number of samples large enough to prove a clinical correlation at the research level. Once that hurdle is past, the technology required to do routine clinical measurement has usually not been the same as used in the marker discovery process, requiring an expensive additional effort, typically requiring the skills of another group or institution. To date, a simple commercial calculation has provided more than adequate discouragement: if to measure  $N$  proteins (for use as a single marker), one must invest  $N$  x the resources required to measure a single protein marker, while the potential market size remains the same, then resources are better spent searching for single markers. The proteins shown in Figure 3 can all be measured in a single sample of plasma, but the commercial cost using individual assays is \$10,896.30. Thus in the end, the success of multi-analyte diagnostics is as much a matter of cost as science.

## **TRANSLATING THE PLASMA PROTEOME INTO USEFUL DIAGNOSTIC RESULTS**

Those with long experience can quote a host of reasons why revolutionary improvements in diagnostics will be slow in coming. Nevertheless, it is incumbent on us to be alert to potential watersheds that occasionally permit rapid progress to occur. In the following we have identified six principal barriers to rapid progress in plasma protein diagnostics, and explored how a way might be found through or around the impediment.

Lack of a shared knowledge base. The plasma proteome occupies an important position at the intersection between three primary fields: proteomics (a technology-based research discipline), medicine (a profession serving the ill patient), and the diagnostics industry (a business delivering data to medicine under growing financial and regulatory constraints). Currently, its appearance from these three directions is quite different, and thus communication among the participants required to achieve an advance in plasma-based diagnostics can be challenging when it comes to describing specific proteins and their significance (as was recognized some years ago by Merrill and Lemkin (150)). This problem will become much more troublesome when multiple proteins are involved. A useful model that could be adapted in this context is the shared literature annotation database, pioneered by Garrels in yeast (151,152) and subsequently expanded to cover some aspects of the human proteome (e.g., G-protein-coupled receptors (153)). These databases collect, analyze and abstract the published scientific literature for each protein in the collection, and relate them to external non-text resources such as genome sequences and genetic data. While the present scope of a Plasma Proteome Database is smaller in terms of number of known proteins, it is far larger in terms of the relevant literature to be summarized and will require a concerted ongoing effort.

Discovery of low-abundance proteins and their disease relationships. Most of the proteins and disease markers in plasma remain to be discovered, and doing so will be very challenging because of the technical issues discussed above (especially dynamic range). A deep exploration effort is required, coupled with one or more technologies that can measure newly-found proteins in significant sets of carefully-collected clinical samples. Access to these technologies and samples spans a wide landscape divided by many intellectual property fences, compounding the technical problems.

Acceptance of multivariate protein panels as medical tests. As argued above, we feel that multi-

## The Human Plasma Proteome

protein panels will be required to achieve broad coverage of disease states where single markers do not exist, and that panels are likely to be more reliable even when there is a single marker. A series of convincing demonstrations of this assertion are required, undertaken in a valid clinical context. These demonstrations require samples, identified proteins for test, measurement platforms, and advanced data analysis, all of which are expensive and difficult to assemble. From a commercial viewpoint, the problem is magnified by the breadth of intellectual property that must be assembled: many protein markers are the subjects of existing patents or applications, and it is unclear how this will impact use of those proteins as components of multivariate tests. Acquiring samples and advanced measurement technologies expose an additional series of patent concerns. Finally, there is the issue of regulatory approval: regulatory agencies have been unenthusiastic about multivariate tests in the past, often for good reasons. In future a serious effort will be required to demonstrate that these difficulties can be overcome.

Acceptance of the individual as his/her own reference, with concomitant repeated testing to obtain time-series data. The concept itself is not unfamiliar: many people have annual measurements of blood PSA or lipoprotein cholesterol. However, in most cases these values are still compared to a simple reference interval and not examined as an explicit function of time to detect trends, even when a trend is the purpose of the measurement. How many of us have seen a graph of our cholesterol over time, even those of us who take cholesterol-lowering drugs? When each diagnostic episode is considered in isolation, much of the patient context (the personalization) of the diagnostic is lost. We thus need to look at the balance between the added cost and inconvenience (periodic blood sampling) on the one hand, and the increased ability to detect and manage disease on the other. We suspect that there are major benefits of routine analytical monitoring in the area of chronic disease and aging, areas in which large aggregate expense can be saved by an early intervention. If this is the case, then there may be a major economic incentive to the diagnostics industry in the patient's repeated, rather than occasional, use of a diagnostic system. This benefit comes, however, with an added challenge: because the individual's reference interval is narrower than the population's, the measurements may need to be more accurate to obtain the added value. The scientific challenge is to demonstrate beyond argument the quantitative benefit of using a person as his/her own control.

## The Human Plasma Proteome

Lack of systematic knowledge of the genetic control of diagnostic marker abundance. Unless we know the level of genetic control over the amount of a protein in plasma, we do not know whether an individual's value is a measure primarily of genetics (which affects risk but cannot be changed) or of phenotype (which reflects risk and can be changed). It can be argued that the latter markers are candidates for repeated measurement (for monitoring disease progression and therapy) while the former are not (once is enough).

Cost. As noted above, the cost of protein tests, averaged over all the proteins for which there are now clinical assays, is more than \$100 per protein measurement. If we need another 150 proteins measured by approved clinical tests, will we need the same (huge) aggregate investment that was required to develop the ~150 tests on the market today? Clearly this cannot be: one or more major advances are required that decrease the cost of multi-protein panel measurements by a factor of 10, and preferably 100, while generating data of very high quality and a reasonably profitable business. Major reagent and manufacturing innovations, particularly in the area of antibody arrays and their equivalents, will be required to achieve these goals.

Regulatory issues. The regulatory standards to be met by new protein analytes and new measurement technologies for the clinical laboratory are very high. The measurements must not only be accurate, they must have a concrete medical implication. The cost of demonstrating medical meaning has to date been very high, requiring clinical trials for each analyte in each indication. This tradition would pose a grave problem if panels of new proteins were required to be the subject of trials one at a time. Instead it may be appropriate to explore a more general retrospective trials process (using banks of well-characterized stored samples) that could allow a stream of new protein analytes to be evaluated and added to panels as they demonstrate added statistical value. Changes such as this would require at the outset that regulatory agencies become comfortable with multivariate protein diagnostics treated as panels, and with the new technologies required to make them cost-effective. At present the data needed to produce this comfort is only beginning to appear.

There is an essential, non-commercial, but very focused, research component required to overcome these barriers, one that is addressed tangentially in many institutions but which we believe will

require the commitment of a dedicated team.

## **CONCLUSIONS**

In reviewing the biological and technological challenges embodied in the plasma proteome, one cannot help but be struck by the daunting nature of what needs to be done, and the relatively modest progress that has been made in the last 20 years. It is, from the optimist's viewpoint, an Opportunity. And it is fortunate that, just at the right time, technologies of mass spectrometry and antibody arrays are emerging that may answer the needs of this demanding situation. It is useful to note that methods exist (albeit for measurement of atomic elements) with a dynamic range of  $10^{14}$ , and that other MS methods can routinely quantitate organic compounds (such as drug metabolites) within +/- 1%. Why should we not have such measurements of peptides and proteins, or at least something like them?

Why, in fact, should we not be able to monitor some of the fundamental process parameters of human biology in the way that real engineers monitor complex systems? Why should we not be able to expand the range of monitored diseases to include a much wider range of chronic or slowly-developing conditions? Why should we not have a long-term patient "protein history" showing the slow changes associated with such diseases and tracking various therapeutic options to see which anti-inflammatory drug, for example, best down-regulates the inflammation of rheumatoid arthritis in a particular individual? Clearly we should, and we can, by taking a rational approach to exploration and measurement of the plasma proteome.

### References

1. Yudell, M. and DeSalle, R. The genomic revolution: Unveiling the unity of life. Joseph Henry Press, with the American Museum of Natural History, Washington, DC (2002).
2. Putnam, F.W. The plasma proteins structure, function, and genetic control. Academic Press, New York (1975-1987).
3. Anderson, L. and Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533-7 (1997).
4. Kawamoto, S., Matsumoto, Y., Mizuno, K., Okubo, K., and Matsubara, K. (1996) Expression profiles of active genes in human and mouse livers. *Gene* **174**, 151-8
5. Putnam, F.W. Perspectives - Past, Present and Future. In Putnam, F.W. (ed.) The plasma proteins structure, function, and genetic control. Academic Press, New York (1975-1987). 1-55
6. Cohn, E.J. (1948) The History of Plasma Fractionation. *Advances in Military Medicine* **1**, 364-443
7. Schultze, H.E. and Heremans, J.F. Molecular biology of human proteins with special reference to plasma proteins. Elsevier Pub. Co, Amsterdam, New York (1966).
8. Moss, D.W. and Henderson.A.R. Clinical Enzymology. In Burtis, C.A. and Ashwood, E.R. (eds.) Tietz Textbook of Clinical Chemistry. W. B. Saunders Company, Philadelphia (1999). 617-721
9. Skeggs, L.T. Jr. (2000) Persistence... and prayer: from the artificial kidney to the AutoAnalyzer. *Clin Chem* **46**, 1425-36
10. Anderson, N.G. (1969) Computer interfaced fast analyzers. *Science* **166**, 317-24

## The Human Plasma Proteome

11. Kohler, G. and Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity . *Nature* **256**, 495-7
12. Bast, R.C. Jr, Klug, T.L., St John, E., Jenison, E., Niloff, J.M., Lazarus, H., Berkowitz, R.S., Leavitt, T., Griffiths, C.T., Parker, L., Zurawski, V.R. Jr, and Knapp, R.C. (1983) A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N Engl J Med* **309**, 883-7
13. O'Brien, T.J., Beard, J.B., Underwood, L.J., Dennis, R.A., Santin, A.D., and York, L. (2001) The CA 125 gene: an extracellular superstructure dominated by repeat sequences. *Tumour Biol* **22**, 348-66
14. Devine, P.L., McGuckin, M.A., Quin, R.J., and Ward, B.G. (1994) Serum markers CASA and CA 15-3 in ovarian cancer: all MUC1 assays are not the same. *Tumour Biol* **15**, 337-44
15. Cheung, K.L., Graves, C.R., and Robertson, J.F. (2000) Tumour marker measurements in the diagnosis and monitoring of breast cancer. *Cancer Treat Rev* **26**, 91-102
16. Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231-43
17. O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-21
18. Anderson, L. and Anderson, N.G. High resolution two-dimensional electrophoresis of human plasma proteins. *Proc Natl Acad Sci U S A* **74**, 5421-5 (1977).
19. Anderson, N.L. and Anderson, N.G. A two-dimensional gel database of human plasma proteins. *Electrophoresis* **12**, 883-906 (1991).
20. Hughes, G.J., Frutiger, S., Paquet, N., Ravier, F., Pasquali, C., Sanchez, J.C., James, R., Tissot, J.D., Bjellqvist, B., and Hochstrasser, D.F. (1992) Plasma protein map: an update by microsequencing. *Electrophoresis* **13**, 707-14

## The Human Plasma Proteome

21. Anderson, N.G., Willis, D.D., Holladay, D.W., Caton, J.E., Holleman, J.W., Eveleigh, J.W., Attrill, J.E., Ball, F.L., and Anderson, N.L. Analytical techniques for cell fractions. XIX. The cyclum: an automatic system for cyclic chromatography. *Anal Biochem* **66**, 159-74 (1975).
22. Pieper, R and Anderson N. L. Immunosubtraction Method for Sample Preparation for 2-DGE. International Patent Publication Number WO 02/055654 A2
23. Tracy, R.P., Currie, R.M., and Young, D.S. Two-dimensional gel electrophoresis of serum specimens from a normal population. *Clin Chem* **28**, 890-9 (1982).
24. Harrison, H., Miller, K., Ober, C., Refetoff, S., Dick, M., and Elias, S. Identification of a serum protein polymorphism via two-dimensional electrophoresis. Family and population studies in two genetically isolated groups: North American Hutterites and Australian aborigines. *Am J Hum Genet* **48**, 362-9 (1991).
25. Harrison, H.H., Gordon, E.D., Nichols, W.C., and Benson, M.D. (1991) Biochemical and clinical characterization of prealbuminCHICAGO: an apparently benign variant of serum prealbumin (transthyretin) discovered with high-resolution two-dimensional electrophoresis. *Am J Med Genet* **39**, 442-52
26. Rosenblum, B.B., Neel, J.V., and Hanash, S.M. Two-dimensional electrophoresis of plasma polypeptides reveals "high" heterozygosity indices. *Proc Natl Acad Sci U S A* **80**, 5002-6 (1983).
27. Anderson, N.L. and Anderson, N.G. Microheterogeneity of serum transferrin, haptoglobin and alpha 2 HS glycoprotein examined by high resolution two-dimensional electrophoresis. *Biochem Biophys Res Commun* **88**, 258-65 (1979).
28. Anderson, N.L., Copple, D.C., Bendele, R.A., Probst, G.S., and Richardson, F.C. Covalent protein modifications and gene expression changes in rodent liver following administration of methapyrilene: a study using two-dimensional electrophoresis. *Fundam Appl Toxicol* **18**, 570-80 (1992).

## The Human Plasma Proteome

29. Gravel, P., Golaz, O., Walzer, C., Hochstrasser, D.F., Turler, H., and Balant, L.P. Analysis of glycoproteins separated by two-dimensional gel electrophoresis using lectin blotting revealed by chemiluminescence. *Anal Biochem* **221**, 66-71 (1994).
30. Golaz, O., Gravel, P., Walzer, C., Turler, H., Balant, L., and Hochstrasser, D.F. Rapid detection of the main human plasma glycoproteins by two-dimensional polyacrylamide gel electrophoresis lectin affino blotting. *Electrophoresis* **16**, 1187-9 (1995).
31. Gravel, P., Walzer, C., Aubry, C., Balant, L.P., Yersin, B., Hochstrasser, D.F., and Guimon, J. New alterations of serum glycoproteins in alcoholic and cirrhotic patients revealed by high resolution two-dimensional gel electrophoresis. *Biochem Biophys Res Commun* **220**, 78-85 (1996).
32. Harrison, H.H., Miller, K.L., Harbison, M.D., and Slonim, A.E. (1992) Multiple serum protein abnormalities in carbohydrate-deficient glycoprotein syndrome: pathognomonic finding of two-dimensional electrophoresis? *Clin Chem* **38**, 1390-2
33. Henry, H., Tissot, J.D., Messerli, B., Markert, M., Muntau, A., Skladal, D., Sperl, W., Jaeken, J., Weidinger, S., Heyne, K., and Bachmann, C. (1997) Microheterogeneity of serum glycoproteins and their liver precursors in patients with carbohydrate-deficient glycoprotein syndrome type I: apparent deficiencies in clusterin and serum amyloid P. *J Lab Clin Med* **129**, 412-21
34. Henry, H., Froehlich, F., Perret, R., Tissot, J.D., Eilers-Messerli, B., Lavanchy, D., Dionisi-Vici, C., Gonvers, J.J., and Bachmann, C. (1999) Microheterogeneity of serum glycoproteins in patients with chronic alcohol abuse compared with carbohydrate-deficient glycoprotein syndrome type I. *Clin Chem* **45**, 1408-13
35. Choukaife, A., Visvikis, S., Steinmetz, J., Galteau, M.M., Kabbaj, O., Ferard, G., Metais, P., and Siest, G. Two-dimensional electrophoresis of plasma proteins and high density lipoproteins during inflammation. *Electrophoresis* **10**, 781-4 (1989).
36. Saile, R., Kabbaj, O., Visvikis, S., Steinmetz, J., Steinmetz, A., Ferard, G., Fruchart, J.C., and Metais,

## The Human Plasma Proteome

- P. Variations in apolipoproteins serum amyloid A, A-I, A-II, and C-III in severely head-injured patients. *J Clin Chem Clin Biochem* **28**, 519-25 (1990).
37. Bini, L., Magi, B., Cellesi, C., Rossolini, A., and Pallini, V. (1992) Two-dimensional electrophoresis analysis of human serum proteins during the acute-phase response. *Electrophoresis* **13**, 743-6
38. Bruun, C.F., Sanchez, J.C., Hochstrasser, D.F., Marhaug, G., and Husby, G. (1998) A two-dimensional electrophoretic study of serum amyloid A and C- reactive protein in infants and children. *Electrophoresis* **19**, 776-81
39. Liberatori, S., Bini, L., De Felice, C., Magi, B., Marzocchi, B., Raggiaschi, R., Pallini, V., and Bracci, R. Acute-phase proteins in perinatal human plasma. *Electrophoresis* **18**, 520-6 (1997).
40. Bini, L., Magi, B., Marzocchi, B., Cellesi, C., Berti, B., Raggiaschi, R., Rossolini, A., and Pallini, V. Two-dimensional electrophoretic patterns of acute-phase human serum proteins in the course of bacterial and viral diseases. *Electrophoresis* **17**, 612-6 (1996).
41. Doherty, N.S., Littman, B.H., Reilly, K., Swindell, A.C., Buss, J.M., and Anderson, N.L. Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. *Electrophoresis* **19**, 355-63 (1998).
42. Tissot, J.D., Schneider, P., James, R.W., Daigneault, R., and Hochstrasser, D.F. High-resolution two-dimensional protein electrophoresis of pathological plasma/serum. *Appl Theor Electrophor* **2**, 7-12 (1991).
43. Tissot, J.D., Hohlfield, P., Forestier, F., Tolsa, J.F., Hochstrasser, D.F., Calame, A., Plouvier, E., Bossart, H., and Schneider, P. (1993) Plasma/serum protein patterns in human fetuses and infants: a study by high-resolution two-dimensional polyacrylamide gel electrophoresis. *Appl Theor Electrophor* **3**, 183-90
44. Magi, B., Marzocchi, B., Bini, L., Cellesi, C., Rossolini, A., and Pallini, V. (1995) Two-dimensional

The Human Plasma Proteome

electrophoresis of human serum proteins modified by ampicillin during therapeutic treatment .

Electrophoresis **16**, 1190-2

45. Lonberg-Holm, K., Sandberg, L.B., Doleman, M.S., and Owens, A.J. Classification of malnutrition by statistical analysis of quantitative two-dimensional gel electrophoresis of plasma proteins. *Comput Biomed Res* **19**, 340-60 (1986).
46. John, H.A. and Purdom, I.F. Elevated plasma levels of haptoglobin in Duchenne muscular dystrophy: electrophoretic variants in patients with a severe form of the disease. *Electrophoresis* **10**, 489-93 (1989).
47. Myrick, J.E., Caudill, S.P., Hubert, I.L., Robinson, M.K., Adams, M.J. Jr, and Pueschel, S.M. (1990) Identification of haptoglobin alpha-2FF variants in mid-trimester maternal serum as potential markers for Down syndrome. *Appl Theor Electrophor* **1**, 233-41
48. Del Priore, G., Chatterton, R.Jr., Lee, C., Silver, R., Berg, L., and Lee, M.J. Comparison of mononuclear cell proteins and plasma proteins before and during parturition by two dimensional electrophoresis. *J Perinat Med* **19**, 373-7 (1991).
49. Larina, O.N. Protein composition in human plasma after long-term orbital missions and in rodent plasma after spaceflights on biosatellites "Cosmos-1887" and "Cosmos-2044". *Physiologist* **34**, S94-5 (1991).
50. Burgess-Cassler, A., Johansen, J.J., and Kendrick, N.C. (1992) Two-dimensional gel analysis of serum apolipoprotein A-I isoforms: preliminary analysis suggests altered ratios in individuals with heart disease. *Appl Theor Electrophor* **3**, 41-5
51. Hoermann, R., Spoettl, G., Grossmann, M., Saller, B., and Mann, K. (1993 ) Molecular heterogeneity of human chorionic gonadotropin in serum and urine from patients with trophoblastic tumors. *Clin Investig* **71**, 953-60

## The Human Plasma Proteome

52. Cameron, B.M., VanderPutten, D.M., and Merrill, C.R. Preliminary study of an increase of a plasma apolipoprotein E variant associated with peripheral nerve damage. A finding in patients with chronic spinal pain. *Spine* **20**, 581-9; discussion 589-90 (1995).
53. Choi, J., Malakowsky, C.A., Talent, J.M., Conrad, C.C., and Gracy, R.W. Identification of oxidized plasma proteins in Alzheimer's disease. *Biochem Biophys Res Commun* **293**, 1566-70 (1992).
54. Eberini, I., Miller, I., Zancan, V., Bolego, C., Puglisi, L., Gemeiner, M., and Gianazza, E. (1999) Proteins of rat serum IV. Time-course of acute-phase protein expression and its modulation by indomethacine. *Electrophoresis* **20**, 846-53
55. Haynes, P., Miller, I., Aebersold, R., Gemeiner, M., Eberini, I., Lovati, M.R., Manzoni, C., Vignati, M., and Gianazza, E. (1998) Proteins of rat serum: I. Establishing a reference two-dimensional electrophoresis map by immunodetection and microbore high performance liquid chromatography-electrospray mass spectrometry. *Electrophoresis* **19**, 1484-92
56. Anderson NL . (1981) High resolution two-dimensional electrophoretic mapping of immunoglobulin light chains. *Immunol Lett* **2**, 195-199
57. Harrison, H.H. (1992) Patient-specific microheterogeneity patterns of monoclonal immunoglobulin light chains as revealed by high resolution, two- dimensional electrophoresis. *Clin Biochem* **25**, 235-43
58. Tracy, R.P., Currie, R.M., Kyle, R.A., and Young, D.S. (1982) Two-dimensional gel electrophoresis of serum specimens from patients with monoclonal gammopathies. *Clin Chem* **28**, 900-7
59. Tracy, R.P., Kyle, R.A., and Young, D.S. (1984) Two-dimensional gel electrophoresis as an aid in the analysis of monoclonal gammopathies. *Hum Pathol* **15**, 122-9
60. Stulik, J., Tichy, M., and Kovarova, H. (1993) Two-dimensional gel electrophoresis of four serum samples from patients with IgD myeloma. *Clin Chim Acta* **218**, 149-58

## The Human Plasma Proteome

61. Vu, D.H., Schneider, P., and Tissot, J.D. (2002) Electrophoretic characteristics of monoclonal immunoglobulin G of different subclasses. *J Chromatogr B Analyt Technol Biomed Life Sci* **771**, 355-68
62. Tissot, J.D., Helg, C., Chapuis, B., Zubler, R.H., Jeannet, M., Hochstrasser, D.F., Hohlfield, P., and Schneider, P. (1992) Clonal imbalances of serum immunoglobulins after allogeneic bone marrow transplantation: an analysis by high-resolution two-dimensional gel electrophoresis. *Bone Marrow Transplant* **10**, 347-53
63. Tissot, J.D., Hohlfield, P., Hochstrasser, D.F., Tolsa, J.F., Calme, A., and Schneider, P. (1993) Clonal imbalances of plasma/serum immunoglobulin production in infants. *Electrophoresis* **14**, 245-7
64. Zahnd, A., Tissot, J.D., and Hochstrasser, D.F. (1993) Wave packets analysis of two-dimensional protein maps: a new approach to study the diversity of immunoglobulins. *Appl Theor Electrophor* **3**, 321-8
65. Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., and Liotta, L.A. (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572-7
66. Washburn, M.P., Wolters, D., and Yates, J.R. 3rd. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-7
67. Lipton, M.S., Pasa-Tolic', L., Anderson, G.A., Anderson, D.J., Auberry, D.L., Battista, J.R., Daly, M.J., Fredrickson, J., Hixson, K.K., Kostandarithes, H., Masselon, C., Markillie, L.M., Moore, R.J., Romine, M.F., Shen, Y., Stritmatter, E., Tolic', N., Udseth, H.R., Venkateswaran, A., Wong, K.K., Zhao, R., and Smith, R.D. (2002) Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. *Proc Natl Acad Sci U S A* **99**, 11049-54
68. Davis, M.T., Spahr, C.S., McGinley, M.D., Robinson, J.H., Bures, E.J., Beierle, J., Mort, J., Yu, W., Luethy, R., and Patterson, S.D. (2001) Towards defining the urinary proteome using liquid

## The Human Plasma Proteome

chromatography- tandem mass spectrometry. II. Limitations of complex mixture analyses.

Proteomics **1**, 108-17

69. Spahr, C.S., Davis, M.T., McGinley, M.D., Robinson, J.H., Bures, E.J., Beierle, J., Mort, J., Courchesne, P.L., Chen, K., Wahl, R.C., Yu, W., Luethy, R., and Patterson, S.D. (2001) Towards defining the urinary proteome using liquid chromatography- tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. Proteomics **1**, 93-107
70. Adkins, J.N. and et al. (in press) Toward a human blood serum proteome: Analysis by multidimensional separation coupled with mass spectrometry. Molec Cell Proteomics
71. MacCoss, M.J., McDonald, W.H., Saraf, A., Sadygov, R., Clark, J.M., Tasto, J.J., Gould, K.L., Wolters, D., Washburn, M., Weiss, A., Clark, J.I., and Yates, J.R. 3rd. (2002) Shotgun identification of protein modifications from protein complexes and lens tissue. Proc Natl Acad Sci U S A **99**, 7900-5
72. Wu, S.-L., Amato, H., Biringer, R., Choudhary, G., Shieh, P., and Hancock, W.S. (in press) Targeted Proteomics of Low-Level Proteins in Human Plasma by LC/MS(n): Using Human Growth Hormone as a Model System. Journal of Proteome Research
73. Schepky, A.G., Bensch, K.W., Schulz-Knappe, P., and Forssmann, W.G. (1994) Human hemofiltrate as a source of circulating bioactive peptides: determination of amino acids, peptides and proteins. Biomed Chromatogr **8**, 90-4
74. Richter, R., Schulz-Knappe, P., Schrader, M., Standker, L., Jurgens, M., Tammen, H., and Forssmann, W.G. (1999) Composition of the peptide fraction in human blood plasma: database of circulating human peptides. J Chromatogr B Biomed Sci Appl **726**, 25-35
75. Desiderio, D.M. and Kai, M. (1983) Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. Biomed Mass Spectrom **10**, 471-9

## The Human Plasma Proteome

76. Patterson, B.W., Zhang, X.J., Chen, Y., Klein, S., and Wolfe, R.R. (1997) Measurement of very low stable isotope enrichments by gas chromatography/mass spectrometry: application to measurement of muscle protein synthesis. *Metabolism* **46**, 943-8
77. Yarasheski, K.E., Smith, K., Rennie, M.J., and Bier, D.M. (1992) Measurement of muscle protein fractional synthetic rate by capillary gas chromatography/combustion isotope ratio mass spectrometry. *Biol Mass Spectrom* **21**, 486-90
78. Kulik, W., Meesterburrie, J.A., Jakobs, C., and de Meer, K. (1998) Determination of delta13C values of valine in protein hydrolysate by gas chromatography-combustion isotope ratio mass spectrometry. *J Chromatogr B Biomed Sci Appl* **710**, 37-47
79. Zachwieja, J.J., Bier, D.M., and Yarasheski, K.E. (1994) Growth hormone administration in older adults: effects on albumin synthesis. *Am J Physiol* **266**, E840-4
80. Lahm, H.W. and Langen, H. (2000) Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis* **21**, 2105-14
81. Oda, Y., Huang, K., Cross, F.R., Cowburn, D., and Chait, B.T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **96**, 6591-6
82. Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics* **1**, 376-86
83. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-9
84. Sechi, S. (2002) A method to identify and simultaneously determine the relative quantities of proteins isolated by gel electrophoresis. *Rapid Commun Mass Spectrom* **16**, 1416-24
85. Niwayama, S., Kurono, S., and Matsumoto, H. (2001) Synthesis of d-labeled N-alkylmaleimides and

## The Human Plasma Proteome

application to quantitative peptide analysis by isotope differential mass spectrometry. *Bioorg Med Chem Lett* **11**, 2257-61

86. Ji, J., Chakraborty, A., Geng, M., Zhang, X., Amini, A., Bina, M., and Regnier, F. (2000) Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J Chromatogr B Biomed Sci Appl* **745**, 197-210
87. Munchbach, M., Quadroni, M., Miotto, G., and James, P. (2000) Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal Chem* **72**, 4047-57
88. Goodlett, D.R., Keller, A., Watts, J.D., Newitt, R., Yi, E.C., Purvine, S., Eng, J.K., von Haller, P., Aebersold, R., and Kolker, E. (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun Mass Spectrom* **15**, 1214-21
89. Stewart, I.I., Thomson, T., and Figeys, D. (2001)  $^{18}\text{O}$  labeling: a tool for proteomics. *Rapid Commun Mass Spectrom* **15**, 2456-65
90. Yao, X., Freas, A., Ramirez, J., Demirev, P.A., and Fenselau, C. (2001) Proteolytic  $^{18}\text{O}$  labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* **73**, 2836-42
91. Ekins, R.P. and Chu, F.W. (1991) Multianalyte microspot immunoassay--microanalytical "compact disk" of the future. *Clin Chem* **37**, 1955-67
92. Stoll, D., Templin, M.F., Schrenk, M., Traub, P.C., Vohringer, C.F., and Joos, T.O. (2002) Protein microarray technology. *Front Biosci* **7**, c13-32
93. Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Zhou, Z., Tchernev, V.T., Christiansen, J., Velleca, M., and Kingsmore, S.F. (2002) Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat Biotechnol* **20**, 359-65
94. Cahill, D.J. (2001) Protein and antibody arrays and their medical applications. *J Immunol Methods*

The Human Plasma Proteome

**250**, 81-91

95. James, R.W., Hochstrasser, D., Tissot, J.D., Funk, M., Appel, R., Barja, F., Pellegrini, C., Muller, A.F., and Pometta, D. (1988) Protein heterogeneity of lipoprotein particles containing apolipoprotein A-I without apolipoprotein A-II and apolipoprotein A-I with apolipoprotein A-II isolated from human plasma. *J Lipid Res* **29**, 1557-71
96. Chan, K.C. and Nicoll, C.S. (1994) Characterization of rat serum insulin-like growth factor-binding proteins by two-dimensional gel electrophoresis: identification of a potentially novel form. *Endocrinology* **134**, 2574-80
97. Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelman, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-7
98. Nedelkov, D. and Nelson, R.W. (2001) Delineation of in vivo assembled multiprotein complexes via biomolecular interaction analysis mass spectrometry. *Proteomics* **1**, 1441-6
99. Chen, Y., Luo, J., Yu, P., Gu, X., and Jiang, Y. 2002. Prediction of Secreted Proteins from the Human Genome. Workshop Poster Abstracts from the Human Genome Meeting, April, 2002, Shanghai, China .
100. McFarlane, I., Bomford, A., and Sherwood, R. *Liver Disease & Laboratory Medicine*. ACB Venture Publications, London (2000).
101. Sacher, R.A. and McPherson, R.A. *Widmann's Clinical Interpretation of Laboratory Tests*. F. A. Davis Company, Philadelphia (2000).

The Human Plasma Proteome

102. Cloonan, M.J., Bishop, G.A., Wilton-Smith, P.D., Carter, I.W., Allan, R.M., and Wilcken, D.E. (1979) An enzyme-immunoassay for myoglobin in human serum and urine. Method development, normal values and application to acute myocardial infarction. *Pathology* **11**, 689-99
103. Walter, S., Carlsson, J., Schroder, R., Neuhaus, K.L., Sorges, E., and Tebbe, U. (1999) Enzymatic markers of reperfusion in acute myocardial infarct. With data from the ISAM study. *Herz* **24**, 430-9
104. Voss, E.M., Sharkey, S.W., Gernert, A.E., Murakami, M.M., Johnston, R.B., Hsieh, C.C., and Apple, F.S. (1995) Human and canine cardiac troponin T and creatine kinase-MB distribution in normal and diseased myocardium. Infarct sizing using serum profiles. *Arch Pathol Lab Med* **119**, 799-806
105. In Tietz, N.W. (ed.) *Clinical Guide to Laboratory Tests*. WB Saunders, (1995). 488-489
106. Wodzig, K.W., Kragten, J.A., Hermens, W.T., Glatz, J.F., and van Dieijen-Visser, M.P. (1997) Estimation of myocardial infarct size from plasma myoglobin or fatty acid-binding protein. Influence of renal function. *Eur J Clin Chem Clin Biochem* **35**, 191-8
107. Bleifeld, W., Mathey, D., Hanrath, P., Buss, H., and Effert, S. (1977) Infarct size estimated from serial serum creatine phosphokinase in relation to left ventricular hemodynamics. *Circulation* **55**, 303-11
108. Abe, S., Arima, S., Yamashita, T., Miyata, M., Okino, H., Toda, H., Nomoto, K., Ueno, M., Tahara, M., Kiyonaga, K., and et, a.I. (1994) Early assessment of reperfusion therapy using cardiac troponin T. *J Am Coll Cardiol* **23**, 1382-9
109. Vogel, J.S. and Turteltaub, K.W. (1998) Accelerator mass spectrometry as a bioanalytical tool for nutritional research. *Adv Exp Med Biol* **445**, 397-410
110. Burtis, C.A. and Ashwood, E.R. *Tietz Textbook of Clinical Chemistry*. W. B. Saunders Company,

## The Human Plasma Proteome

Philadelphia (1999).

111. Harris, E.K. (1974) Effects of intra- and interindividual variation on the appropriate use of normal ranges. *Clin Chem* **20**, 1535-42
112. Harris, E.K. (1975) Some theory of reference values. I. Stratified (categorized) normal ranges and a method for following an individual's clinical laboratory values. *Clin Chem* **21**, 1457-64
113. Macy, E.M., Hayes, T.E., and Tracy, R.P. (1997) Variability in the measurement of C-reactive protein in healthy subjects: implications for reference intervals and epidemiological applications. *Clin Chem* **43**, 52-8
114. Zweig, M.H. and Campbell, G. (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* **39**, 561-77
115. Lewis, M.R., Callas, P.W., Jenny, N.S., and Tracy, R.P. (2001) Longitudinal stability of coagulation, fibrinolysis, and inflammation factors in stored plasma samples. *Thromb Haemost* **86**, 1495-500
116. Buckley, C.E. 3rd and Dorsey, F.C. (1971) Serum immunoglobulin levels throughout the life-span of healthy man. *Ann Intern Med* **75**, 673-82
117. Buckley, C.E. and Dorsey, F.C. (1971) Addendum: immunoglobulins and ethnic groups. *Ann Intern Med* **75**, 972
118. Milani-Comparetti, M. and Saccucci, F. (1976) Plasma protein variability in MZ twins. *Acta Genet Med Gemellol (Roma)* **25**, 154-6
119. Cava, F., Gonzalez, C., Pascual, M.J., Navajo, J.A., and Gonzalez-Buitrago, J.M. (2000) Biological variation of interleukin 6 (IL-6) and soluble interleukin 2 receptor (sIL2R) in serum of healthy individuals. *Cytokine* **12**, 1423-5

The Human Plasma Proteome

120. Anesi, A., Rondanelli, M., Trotti, R., and Melzi d'Eril, G.V. (2000) Biological variability of myoglobin in healthy elderly and younger subjects. *Aging (Milano)* **12**, 168-72
121. Tuxen, M.K., Soletormos, G., Petersen, P.H., Schioler, V., and Dombernowsky, P. (1999) Assessment of biological variation and analytical imprecision of CA 125, CEA, and TPA in relation to monitoring of ovarian cancer. *Gynecol Oncol* **74**, 12-22
122. Panteghini, M. and Pagani, F. (1993) Pre-analytical, analytical and biological sources of variation of lipoprotein(a). *Eur J Clin Chem Clin Biochem* **31**, 23-8
123. Huguet, J. and Fuentes-Arderiu, X. (1991) Biological variation in the catalytic concentration of pancreatic alpha- amylase and triacylglycerol lipase in serum. *Scand J Clin Lab Invest* **51**, 735-8
124. Statland, B.E., Bokelund, H., and Winkel, P. (1974) Factors contributing to intra-individual variation of serum constituents: 4. Effects of posture and tourniquet application on variation of serum constituents in healthy subjects. *Clin Chem* **20**, 1513-9
125. Statland, B.E., Winkel, P., and Bokelund, H. (1973) Factors contributing to intra-individual variation of serum constituents. 2. Effects of exercise and diet on variation of serum constituents in healthy subjects. *Clin Chem* **19**, 1380-3
126. Poortmans, J.R. and Haralambie, G. (1979) Biochemical changes in a 100 km run: proteins in serum and urine. *Eur J Appl Physiol Occup Physiol* **40**, 245-54
127. Leonard, P.J. The effect of age and sex on biochemical parameters in blood of healthy human subjects. In Siest, G. (ed.) *Reference Values in Human Chemistry*. Karger, Basel (1973). 134-140
128. Macy, E.M., Meilahn, E.N., Declerck, P.J., and Tracy, R.P. (1993) Sample preparation for plasma measurement of plasminogen activator inhibitor-1 antigen in large population studies. *Arch Pathol Lab Med* **117**, 67-70
129. Siest, G.H.J.S.F. *Interpretation des Examens de Laboratoire*. Karger, Basel (1981).

The Human Plasma Proteome

130. Buckley, C.E. 3rd and Dorsey, F.C. (1970) The effect of aging on human serum immunoglobulin concentrations. *J Immunol* **105**, 964-72
131. Letellier, G. and Desjarlais, F. (1982) Study of seasonal variations for eighteen biochemical parameters over a four-year period. *Clin Biochem* **15**, 206-11
132. Brohult, J. (1969) Effects of high protein and low protein diets on ornithine carbamoyl transferase activity in human serum (S-OCT). *Acta Med Scand* **185**, 357-62
133. Linkowski, P., Spiegel, K., Kerkhofs, M., L'Hermite-Baleriaux, M., Van Onderbergen, A., Leproult, R., Mendlewicz, J., and Van Cauter, E. (1998) Genetic and environmental influences on prolactin secretion during wake and during sleep. *Am J Physiol* **274**, E909-19
134. Larina, O.N. The protein composition of the blood plasma of cosmonauts after prolonged orbital flights]. *Aviakosm Ekolog Med* **26**, 67-9 (1992).
135. Mendlewicz, J., Linkowski, P., Kerkhofs, M., Leproult, R., Copinschi, G., and Van Cauter, E. (1999) Genetic control of 24-hour growth hormone secretion in man: a twin study. *J Clin Endocrinol Metab* **84**, 856-62
136. Young, D.S., Thomas, D.W., Friedman, R.B., and Pestaner, L.C. (1972) Effects of drugs on clinical laboratory tests. *Clin Chem* **18**, 1041-303
137. Young, D.S. (1997) Effects of drugs on clinical laboratory tests. *Ann Clin Biochem* **34 ( Pt 6)**, 579-81
138. Jellum, E., Bjornson, I., Nesbakken, R., Johansson, E., and Wold, S. (1981) Classification of human cancer cells by means of capillary gas chromatography and pattern recognition analysis. *J Chromatogr* **217**, 231-7
139. Robertson, E.A., Van Steirteghem, A.C., Byrkit, J.E., and Young, D.S. (1980) Biochemical individuality and the recognition of personal profiles with a computer. *Clin Chem* **26**, 30-6

The Human Plasma Proteome

140. Pankow, J.S., Folsom, A.R., Cushman, M., Borecki, I.B., Hopkins, P.N., Eckfeldt, J.H., and Tracy, R.P. (2001) Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis* **154**, 681-9
141. Tarroux, P. (1983) Analysis of protein patterns during differentiation using 2-D electrophoresis and computer multidimensional classification. *Electrophoresis* **4**, 63—70
142. Anderson, N.L., Hofmann, J.P., Gemmell, A., and Taylor, J. (1984) Global approaches to quantitative analysis of gene-expression patterns observed by use of two-dimensional gel electrophoresis . *Clin Chem* **30**, 2031-6
143. Anderson, N.L., Giere, F.A., Nance, S.L., Gemmell, M.A. , Tollaksen, S.L., and Anderson, N.G. (1987) Effects of toxic agents at the protein level: quantitative measurement of 213 mouse liver proteins following xenobiotic treatment. *Fundam Appl Toxicol* **8**, 39-50
144. Anderson, N.L., Esquer-Blasco, R., Richardson, F., Foxworthy, P., and Eacho, P. The effects of peroxisome proliferators on protein abundances in mouse liver. *Toxicol Appl Pharmacol* **137**, 75-89 (1996).
145. Anderson, L., Steele, V.K., Kelloff, G.J., and Sharma, S. (1995) Effects of oltipraz and related chemoprevention compounds on gene expression in rat liver. *J Cell Biochem* **Supplement 22**, 108-116
146. Schmid, H.R., Schmitter, D., Blum, P., Miller, M., and Vonderschmitt, D. (1995) Lung tumor cells: a multivariate approach to cell classification using two-dimensional protein pattern. *Electrophoresis* **16**, 1961-8
147. Negishi, Y., Furukawa, T., Oka, T., Sakamoto, M., Hirata, T., Okabe, K., Matayoshi, K., Akiya, K., and Soma, H. (1987) Clinical use of CA 125 and its combination assay with other tumor marker in patients with ovarian carcinoma. *Gynecol Obstet Invest* **23**, 200-7

The Human Plasma Proteome

148. Alaiya, A.A., Franzen, B., Hagman, A., Silfversward, C., Moberger, B., Linder, S., and Auer, G. (2000) Classification of human ovarian tumors using multivariate data analysis of polypeptide expression patterns. *Int J Cancer* **86**, 731-6
149. Young, D.S. and Tracy, R.P. (1995) Clinical applications of two-dimensional electrophoresis. *J Chromatogr A* **698**, 163-79
150. Merrill, C.R., Goldstein, M.P., Myrick, J.E., Creed, G.J., and Lemkin, P.F. The protein disease database of human body fluids: I. Rationale for the development of this database. *Appl Theor Electrophor* **5**, 49-54 (1995).
151. Garrels, J.I. (1996) YPD-A database for the proteins of *Saccharomyces cerevisiae*. *Nucleic Acids Res* **24**, 46-9
152. Garrels, J.I. (2002) Yeast genomic databases and the challenge of the post-genomic era. *Funct Integr Genomics* **2**, 212-37
153. Hodges, P.E., Carrico, P.M., Hogan, J.D., O'Neill, K.E., Owen, J.J., Mangan, M., Davis, B.P., Brooks, J.E., and Garrels, J.I. (2002) Annotating the human proteome: the Human Proteome Survey Database (HumanPSD) and an in-depth target database for G protein-coupled receptors (GPCR-PD) from Incyte Genomics. *Nucleic Acids Res* **30**, 137-41
154. Specialty Laboratories. Directory of Services, and Use and Interpretation of Tests. Santa Monica, CA (2001).
155. Harrela, M., Koistinen, H., Kaprio, J., Lehtovirta, M., Tuomilehto, J., Eriksson, J., Toivanen, L., Koskenvuo, M., Leinonen, P., Koistinen, R., and Seppala, M. (1996) Genetic and environmental components of interindividual variation in circulating levels of IGF-I, IGF-II, IGFBP-1, and IGFBP-3. *J Clin Invest* **98**, 2612-5
156. Verhaeghe, J., Loos, R., Vlietinck, R., Herck, E.V., van Bree, R., and Schutter, A.M. (1996) C-

## The Human Plasma Proteome

- peptide, insulin-like growth factors I and II, and insulin-like growth factor binding protein-1 in cord serum of twins: genetic versus environmental regulation. *Am J Obstet Gynecol* **175**, 1180-8
157. de Lange, M., Snieder, H., Ariens, R.A., Spector, T.D., and Grant, P.J. (2001) The genetics of haemostasis: a twin study. *Lancet* **357**, 101-5
158. Kaprio, J., Eriksson, J., Lehtovirta, M., Koskenvuo, M., and Tuomilehto, J. (2001) Heritability of leptin levels and the shared genetic effects on body mass index and leptin in adult Finnish twins. *Int J Obes Relat Metab Disord* **25**, 132-7
159. Bansal, A., Murray, D.K., Wu, J.T., Stephenson, R.A., Middleton, R.G., and Meikle, A.W. (2000) Heritability of prostate-specific antigen and relationship with zonal prostate volumes in aging twins. *J Clin Endocrinol Metab* **85**, 1272-6
160. Orstavik, K.H. (1990) Genetics of plasma concentration of von Willebrand factor. *Folia Haematol Int Mag Klin Morphol Blutforsch* **117**, 527-31
161. Boomsma, D.I., Hennis, B.C., van Wees, A.G., Frants, R.R., and Klufft, C. (1993) A parent-twin study of plasma levels of histidine-rich glycoprotein (HRG). *Thromb Haemost* **70**, 848-51
162. Daiger, S.P., Miller, M., and Chakraborty, R. (1984) Heritability of quantitative variation at the group-specific component (Gc) locus. *Am J Hum Genet* **36**, 663-76
163. Jacobsen, H.P., Herskind, A.M., Nielsen, B.W., and Husby, S. (2001) IgE in unselected like-sexed monozygotic and dizygotic twins at birth and at 6 to 9 years of age: high but dissimilar genetic influence on IgE levels. *J Allergy Clin Immunol* **107**, 659-63
164. Cesari, M., Sartori, M.T., Patrassi, G.M., Vettore, S., and Rossi, G.P. (1999) Determinants of plasma levels of plasminogen activator inhibitor-1 : A study of normotensive twins. *Arterioscler Thromb Vasc Biol* **19**, 316-20
165. Boomsma, D.I., Kaptein, A., Kempen, H.J., Gevers Leuven, J.A., and Princen, H.M. (1993)

The Human Plasma Proteome

Lipoprotein(a): relation to other risk factors and genetic heritability. Results from a Dutch parent-twin study. *Atherosclerosis* **99**, 23-33

166. Smiles, A.M., Macy, E.M., Sakkinen, P.A., Bovill, E.G., Mann, K.G., and Tracy, R.P. (1998) Variability in plasma prothrombin concentration: implications for use in epidemiology. *Blood Coagulation and Fibrinolysis* **9**, 525-31
167. Valero-Politi, J. and Fuentes-Arderiu, X. (1993) Within- and between-subject biological variations of follitropin, lutropin, testosterone, and sex-hormone-binding globulin in men. *Clinical Chemistry* **39**, 1723-5
168. Maes, M., Cooreman, W., Delanghe, J., Scharpe, S., Wauters, A., Neels, H., D'Hondt, P., Peeters, D., Cosyns, P., and Ranjan, R. (1995) Components of biological variation in plasma haptoglobin: relationships to plasma fibrinogen and immune variables, including interleukin-6 and its receptor. *Clinical Chemistry Acta* **239**, 23-35
169. Gonzalez, C., Cava, F., Ayllon, A., Guevara, P., Navajo, J.A., and Gonzalez-Buitrago, J.M. (2001) Biological variation of interleukin-1beta, interleukin-8 and tumor necrosis factor-alpha in serum of healthy individuals. *Clinical Chemistry Laboratory Medicine* **39**, 836-41

Footnotes

- 1) Swiss Institute of Bioinformatics . SWISS 2D-PAGE, ExPASy Molecular Biology Server at  
<http://us.expasy.org/ch2d/>
- 2) Human Genome Sciences Inc. B Lymphocyte Stimulator (BLyS) at  
[.http://www.hgsi.com/products/BLyS.html](http://www.hgsi.com/products/BLyS.html)
- 3) Plasma Protein Therapeutics Association. Industry Facts at  
[http://www.plasmatherapeutics.org/ppta\\_worldwide/wo\\_industry\\_facts.htm](http://www.plasmatherapeutics.org/ppta_worldwide/wo_industry_facts.htm)
- 4) Georgetown Economic Services. Monthly Industry Aggregation Data.  
[http://www.plasmatherapeutics.org/ppta\\_namerica/0302\\_data.pdf](http://www.plasmatherapeutics.org/ppta_namerica/0302_data.pdf)

## Figure Legends

Figure 1. **The increase in protein species resolved and identified in plasma over a 70 year period.**

Squares show the number of peaks, bands or spots resolved, based on literature data summarized in the text. Diamonds show the number of proteins identified with respect to the external (usually sequence-based) literature. Recent values of 250 (\*) and 490 (+) proteins come from LC/LC/2-DE (R. Pieper, T. Gatlin, S. Steiner and L. Anderson, manuscript in preparation) and LC/LC/MS/MS (70) respectively.

Figure 2. **Histogram of the calculated masses of the processed forms of 289 proteins observed in plasma (Table 1).**

Figure 3. **Reference intervals for 70 protein analytes in plasma.** Abundance is plotted on a log scale spanning 12 orders of magnitude. Where only an upper limit is quoted, the lower end of the interval line shows an arrowhead. The classical plasma proteins are clustered to the left (high abundance), the tissue leakage markers (e.g., enzymes and troponins) are clustered in the center, and cytokines are clustered to the right (low abundance). Hemoglobin is included (far left) for comparison. Data obtained from the Specialty Laboratories publications (154).

Figure 4. **Ten new protein analytes approved by FDA under CLIA regulations since 1993.** For each test, the bar shows the number of versions of the test approved since introduction (left scale), and the line shows the date of approval (right scale).

Figure 5. **The declining rate of introduction of new protein tests.** The data of Figure 4 plotted to indicate rate of introduction of new protein analytes in FDA-approved clinical tests.

Figure 6. **Levels of C-reactive protein (CRP) in three individuals over time.** Selected data from a published dataset (113) for three individuals with different normal values. Samples were taken approximately every 3 weeks for 6 months. The third data point for the topmost individual coincided with a cold (R. P.

The Human Plasma Proteome

Tracy, personal communication).



## TABLES

Table 1. A list of 289 proteins whose detection in plasma or serum is documented in the literature. "Steady-state" forms are listed, but cleavage products formed during activation and disposal are not, nor are splice variants (except fibrinogen gamma) or non-functional proteolytic products (except the cleaved haptoglobin beta chain, and the two major forms of plasminogen: Glu and Lys). Of these 117 are current clinical analytes.

<p>5'-Nucleotidase                      Acid labile subunit of IGFBP                      Acid phosphatase, tartrate-resistant                      Acid phosphatase, prostatic                      Actin beta (from platelets)                      Actin gamma (from platelets)                      Adenosine Deaminase                      Adiponectin                      Alanine aminotransferase (ALT)                      Albumin                      Aldolase (muscle type)                      Alkaline Phosphatase (bone)                      Alpha1,3-fucosyltransferase (FUT6)                      Alpha-1-acid Glycoprotein                      Alpha-1-Antichymotrypsin                      Alpha-1-Antitrypsin                      Alpha-1-B Glycoprotein                      Alpha-1-Microglobulin                      Alpha-2-Antiplasmin                      Alpha-2-HS Glycoprotein                      Alpha-2-Macroglobulin                      Alpha-fetoprotein                      Amylase (pancreatic)                      Angiostatin                      Angiotensin converting enzyme (ACE)                      Angiotensinogen                      Antithrombin III (AT3)                      Apolipoprotein A-I                      Apolipoprotein A-II                      Apolipoprotein A-IV                      Apolipoprotein B-100                      Apolipoprotein B-48                      Apolipoprotein C-I                      Apolipoprotein C-II                      Apolipoprotein C-III                      Apolipoprotein C-IV                      Apolipoprotein D                      Apolipoprotein E                      Apolipoprotein F                      Apolipoprotein H                      Apolipoprotein J (Clusterin)                      Apolipoprotein(a)                      Aspartate aminotransferase (AST)                      Beta Thromboglobulin                      Beta-2-microglobulin                      CA 125                      CA 19-9                      CA 72-4                      CA27.29/15-3 (MUC1 mucin antigens)                      Calreticulin                      Carboxypeptidase N, regulatory                      Carboxypeptidase N, catalytic                      Carcinoembryonic Antigen                      Cathepsin D                      CD5 antigen-like protein                      Ceruloplasmin                      Cholinesterase Plasma                      Chorionic Gonadotropin Beta (hCG)                      Chromogranin A                      Chromogranin B (secretogranin I)                      Coagulation Factor II (Prothrombin)                      Coagulation Factor IX                      Coagulation Factor V                      Coagulation Factor VII, H                      Coagulation Factor VII, L                      Coagulation Factor VIII                      Coagulation Factor X                      Coagulation Factor XI                      Coagulation Factor XII                      Coagulation Factor XIII A                      Coagulation Factor XIII B                      Collagen I c-terminal propeptide                      Collagen I c-terminal telopeptide (ICTP)</p>	<p>Collagen I n-terminal propeptide                      Collagen I n-terminal telopeptide (NTx)                      Collagen III c-terminal propeptide                      Collagen III n-terminal propeptide                      Collagen IV 7S n-terminal propeptide                      Complement C1 Inhibitor                      Complement C1q, A                      Complement C1q, B                      Complement C1q, C                      Complement C1r                      Complement C1s                      Complement C2                      Complement C3A anaphylotoxin                      Complement C3B, alpha'                      Complement C3B, beta                      Complement C4 anaphylotoxin                      Complement C4, alpha                      Complement C4, beta                      Complement C4, gamma                      Complement C4-binding protein, alpha                      Complement C4-binding protein, beta                      Complement C5A anaphylotoxin                      Complement C5B, alpha'                      Complement C5B, beta                      Complement C6                      Complement C7                      Complement C8, alpha                      Complement C8, beta                      Complement C8, gamma                      Complement C9                      Complement Factor B                      Complement Factor B - Bb Fragment                      Complement Factor D                      Complement Factor H                      Complement Factor I                      Connective Tissue Activating Peptide III                      Corticotropin Releasing Hormone (CRH)                      C-reactive Protein                      Creatine Kinase, B                      Creatine Kinase, M                      CRHBP                      Cystatin C                      Elastase (neutrophil)                      Eosinophil granule major basic protein                        E-selectin, soluble                      Ferritin, H                      Ferritin, L                      Fibrin fragment D-dimer                      Fibrinogen extended gamma chain                      Fibrinogen, alpha                      Fibrinogen, beta                      Fibrinogen, gamma                      Fibronectin                      Fibulin-1                      Ficolin 1                      Ficolin 2                      Ficolin 3                      Follicle stimulating hormone                      G-6-PD                      Galactoglycoprotein (Leukosialin)                      Gamma-glutamyl transferase alpha                      Gc-globulin                      GCSF                      Gelsolin                      GHRH                      Glutamate carboxypeptidase II                      Glutathione Peroxidase                      Glutathione S-transferase                      Glycoprotein hormones alpha chain                      GMCSF</p>	<p>Growth Hormone                      Growth Hormone Binding Protein                      Haptoglobin alpha-1                      Haptoglobin alpha-2-chain                      Haptoglobin beta chain                      Haptoglobin beta chain, cleaved                      Haptoglobin-related gene product                      Hemoglobin, alpha                      Hemoglobin, beta                      Hemopexin (Beta-1B-glycoprotein)                      Histidine-rich Alpha-2-glycoprotein                      ICAM-1, soluble                      Ig Kappa light chain                      Ig Lambda light chain                      IgA1                      IgA2                      IgD                      IgE                      IGFBP-3                      IgG1                      IgG2                      IgG3                      IgG4                      IgJ-chain                      IgM                      Inhibin (activin), beta A                      Inhibin (activin), beta B                      Inhibin (activin), beta C                      Inhibin (activin), beta E                      Inhibin, alpha                      Insulin C-Peptide                      Insulin, A chain                      Insulin, B chain                      Insulin-like growth factor IA                      Insulin-like growth factor II                      Inter-alpha trypsin inhibitor, H1                      Inter-alpha trypsin inhibitor, H2                      Inter-alpha trypsin inhibitor, H4                      Inter-alpha-trypsin inhibitorL                      Interferon Alpha                      Interferon Beta                      Interferon Gamma                      Interleukin-1 Beta                      Interleukin-10                      Interleukin-12, alpha                      Interleukin-12, beta                      Interleukin-1receptor antagonist                      Interleukin-2                      Interleukin-4                      Interleukin-5                      Interleukin-6                      Interleukin-8                      IP-10, Small inducible cytokine B10                      Isocitrate dehydrogenase                      Kininogen                      Ksp37                      Laminin, alpha                      Laminin, beta                      Laminin, gamma                      LDH (heart)                      Lecithin-cholesterol acyltransferase                      Leucine-rich Alpha-2-glycoprotein                      LHRH                      Lipase                      Luteinizing hormone (LH), beta                      Mannose-binding Protein                      Matrix metalloproteinase-2                      M-CSF                      Melastatin                      MIP-1 alpha                      MIP-1 beta                      MSE55                      Myelin Basic Protein                      Myoglobin                      N-Acetyl-B-D-Glucosaminidase, alpha                      N-Acetyl-B-D-Glucosaminidase, beta</p>	<p>N-Acetylmuramyl-L-alanine amidase                      Neuron-specific Enolase                      Neutrophil-activating peptide 2                      Osteocalcin                      Osteonectin                      Pancreatic zymogen granule membrane protein GP-2                      Paraoxonase Parathyroid Hormone                      Parathyroid Hormone-Related Protein                      PASP                      Pepsinogen A                      Plasma hyaluronan binding protein                      Plasma kallikrein                      Plasma serine protease inhibitor                      Plasminogen - Glu                      Plasminogen - Lys                      Platelet Factor 4                      Pre-alpha trypsin inhibitor, H3                      Pregnancy-associated plasma protein-A                      Pregnancy-associated plasma protein-A2                      Pregnancy-specific beta-1-glycoprotein 3                      Prolactin                      Prolyl hydroxylase, alpha                      Prolyl hydroxylase, beta                      Prostaglandin-H2 D-isomerase                      Prostate Specific Antigen                      Protein C, H                      Protein C, L                      Protein S                      Protein Z                      P-selectin, soluble                      Rantes                      Renin                      Retinol Binding Protein                      S100 protein                      Secretogranin V                      Serum Amyloid A                      Serum Amyloid P                      Sex Hormone Binding Globulin                      Tetranectin                      Thyroglobulin                      Thyroid Stimulating Hormone                      Thyrotropin-releasing hormone                      Thyroxin Binding Globulin                      Tissue Factor                      Tissue inhibitor of metalloproteinases-1                      Tissue inhibitor of metalloproteinases-2                      Tissue Plasminogen Activator                      Tissue Plasminogen Activator Inhibitor                      TNF-Alpha                      TNF-Binding Protein 1                      TNF-Binding Protein 2                      Transcobalamin                      Transcortin                      Transferrin                      Transferrin (asialo-, tau-, beta-2-)                      Transferrin Receptor (Soluble)                      Transthyretin                      Triacylglycerol lipase (pancreatic)                      Troponin I (cardiac)                      Troponin I, (skeletal)                      Troponin T (cardiac)                      Trypsin, beta-2                      Tyrosine hydroxylase                      Urokinase (High MW kidney type)A                      Urokinase (High MW kidney type)B                      VCAM-1, soluble                      Vitronectin                      Von Willebrand Factor                      Zn-Alpha-2-glycoprotein</p>
---	--	--	--

Table 2. Published Values of Genetic Contributions to Quantitative Variation of Plasma Proteins, with literature sources referenced.

Protein	Proportion Genetic (%)
Insulin-like growth factor binding protein-1 (155)	Very low
C-peptide (156)	12
Factor xiii (157)	38
Insulin-like growth factor I (155)	38
Leptin (158)	34-45
Prostate-specific antigen (159)	45
Fibrinogen, factor VII, factor VIII, PAI-1, tissue plasminogen activator, factor XIII A-subunit and B- subunit, and von Willebrand factor (157)	41-75
Insulin-like growth factor binding protein-3 (155)	60
von Willebrand factor (160)	60
Insulin-like growth factor II (155)	66
Histidine-rich glycoprotein (161)	69
Gc (group-specific component) (162)	70
Immunoglobulin E (163)	78
Factor XIII activity (157)	82
Insulin-like growth factor I (156)	77-93
Plasminogen activator inhibitor type 1 antigen (164)	87
Lipoprotein(a) (165)	93-98

Table 3. Published Values of Intra-Individual and Inter-Individual Coefficients of Variation for Selected Plasma Proteins

<b>Protein</b>	<b>CVi % (individual)</b>	<b>CVg % (group)</b>
Soluble interleukin 2 receptor (sil2r) (119)	5.84	38.81
Prothrombin (166)	7.3	14.5
Lipoprotein(a) (122)	8.6	85.8
Myoglobin (120)	11	24.3
Sex-hormone-binding globulin (167)	12.1	42.7
CEA (121)	13.9	58.3
Follitropin (167)	17.3	36.0
Haptoglobin (168)	20.0	27.9
Interleukin-8 (169)	24	31
Lutropin (167)	24.0	37.0
Interleukin-1beta (169)	30	36
TPA (121)	31.4	62.5
CA 125 (121)	35.5	70.6
C-reactive protein (CRP) (113)	42.2	92.5
Tumor necrosis factor-alpha (169)	43	29
Interleukin 6 (IL-6) (119)	48.5	39.4

# The Human Plasma Proteome

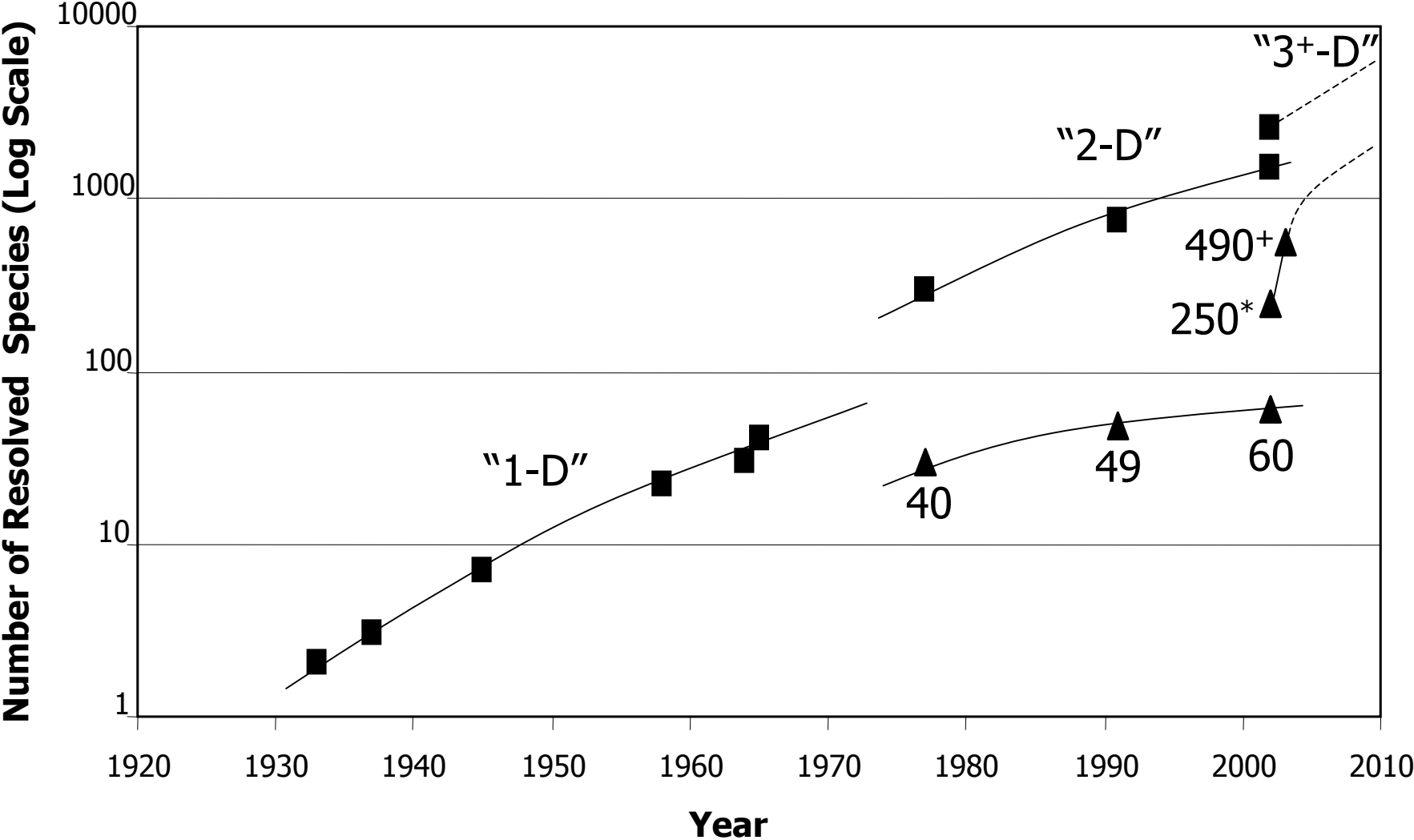


Figure 1

# The Human Plasma Proteome

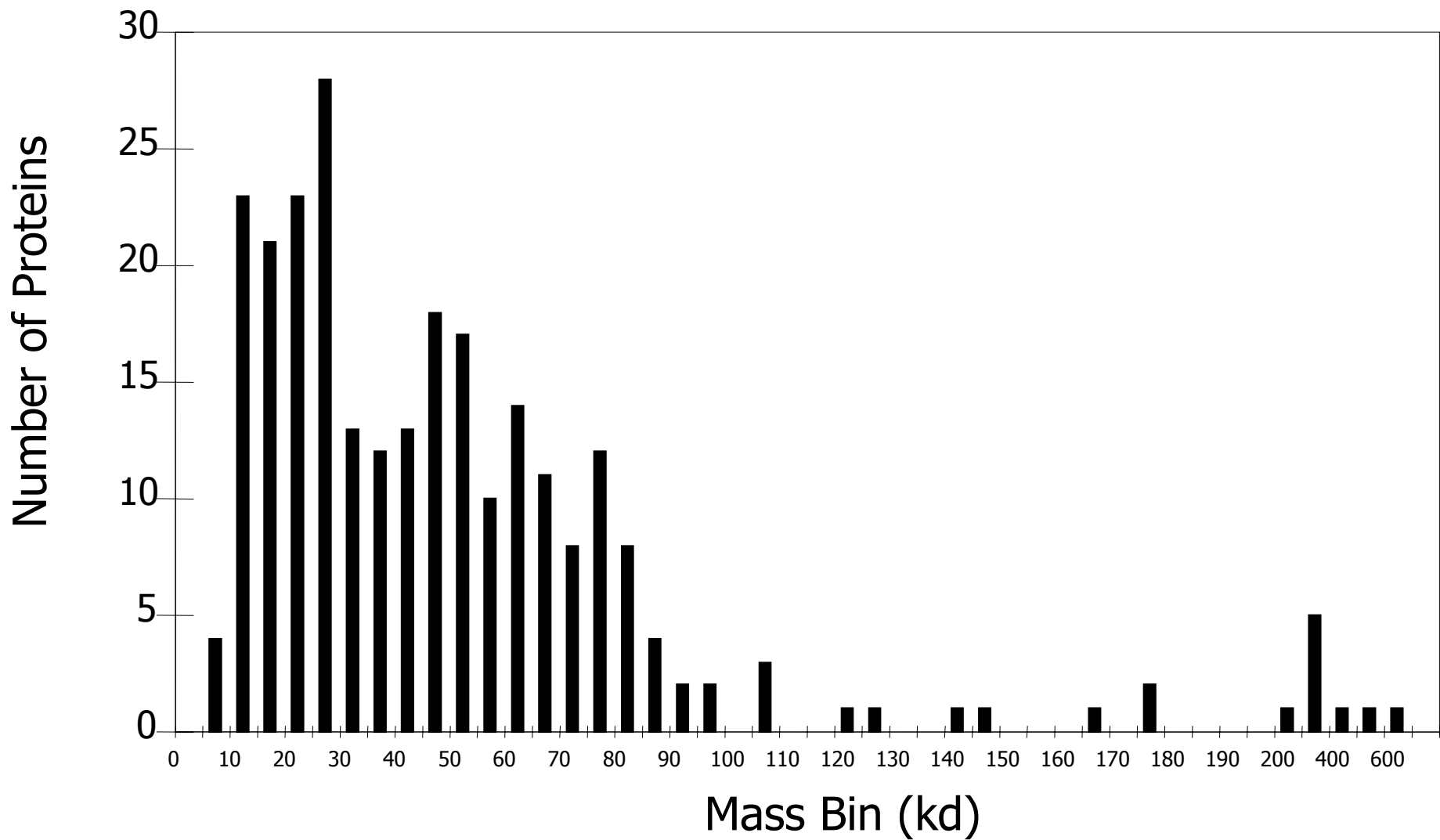


Figure 2

# The Human Plasma Proteome

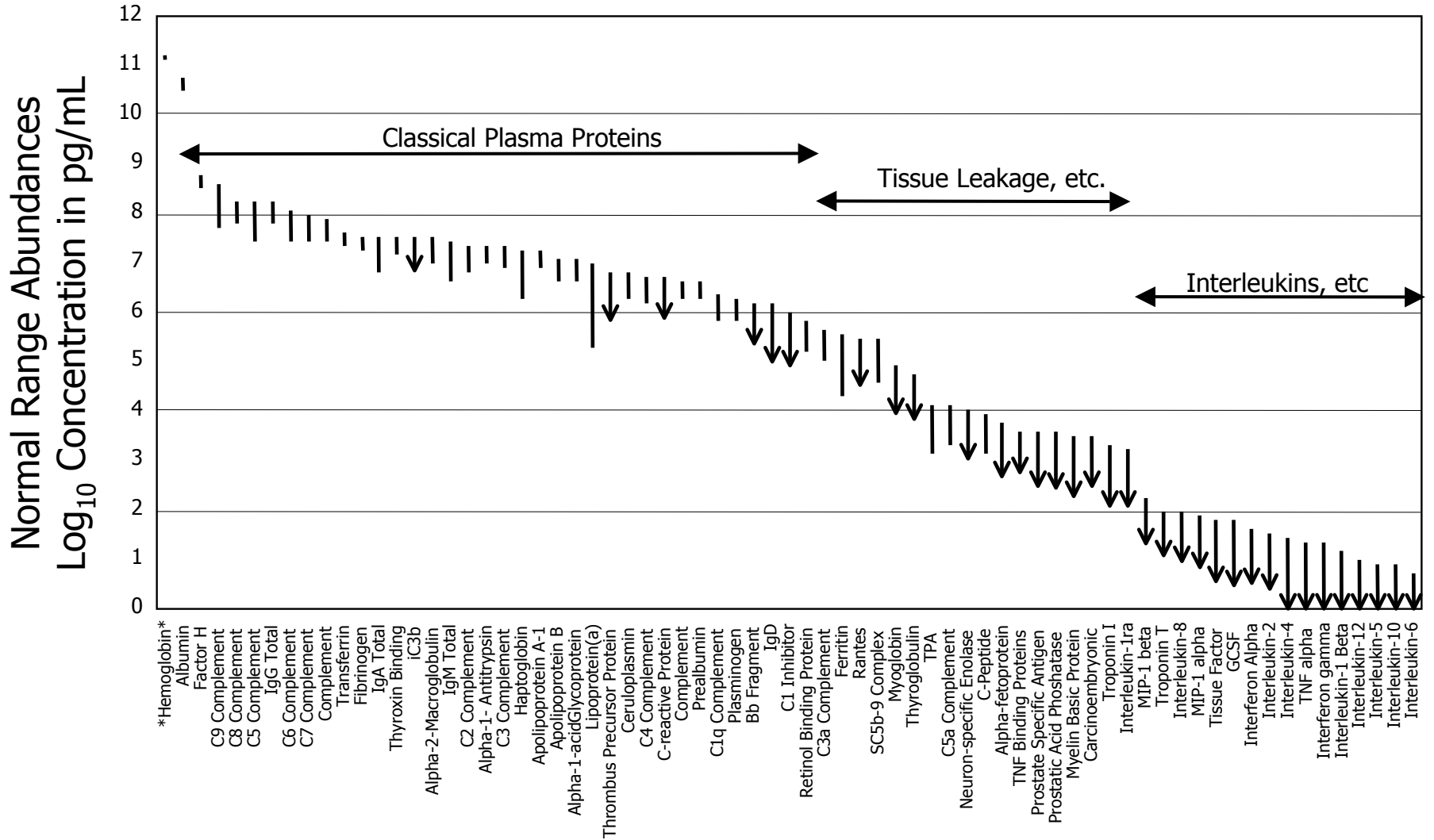


Figure 3

# The Human Plasma Proteome

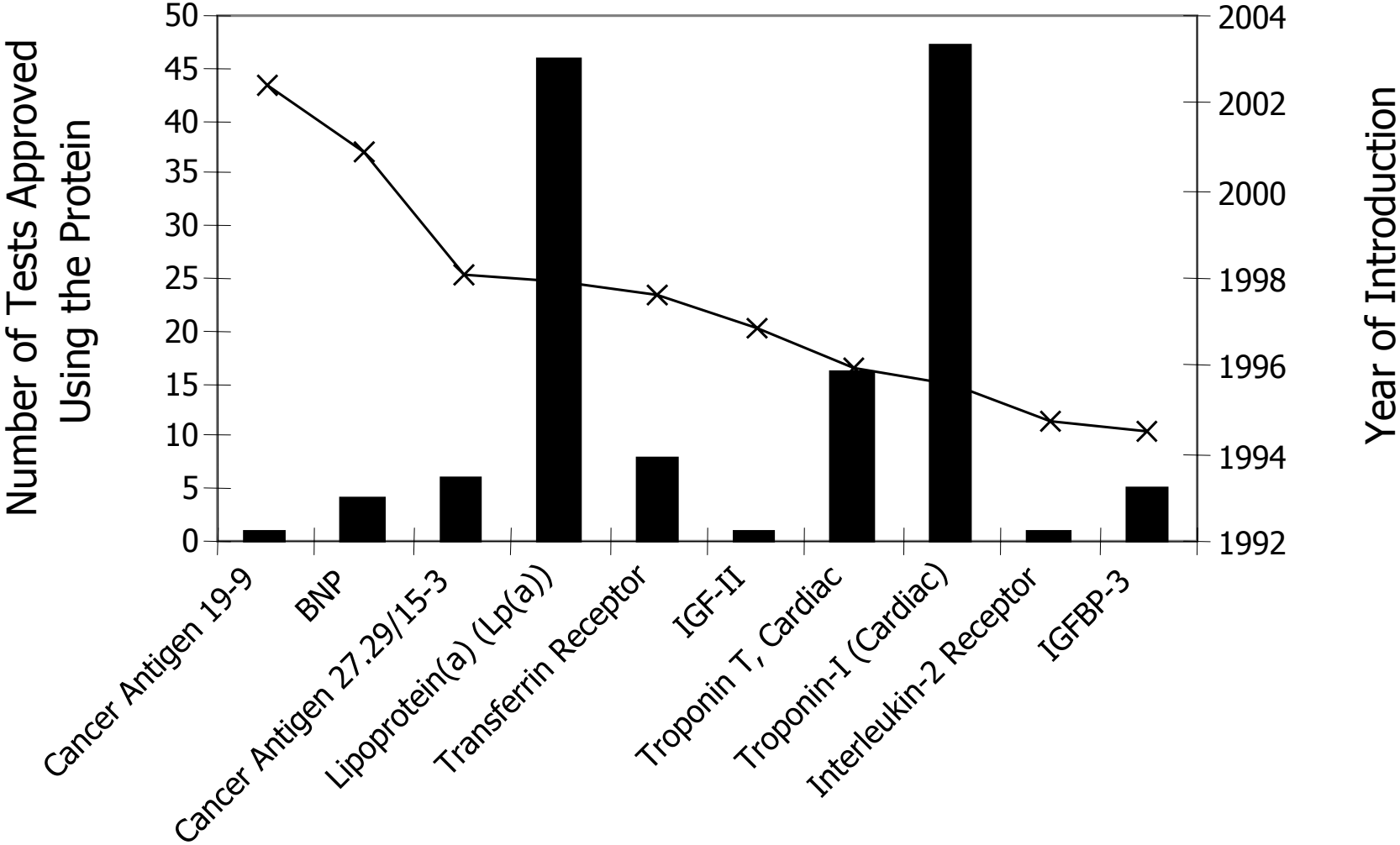


Figure 4

# The Human Plasma Proteome

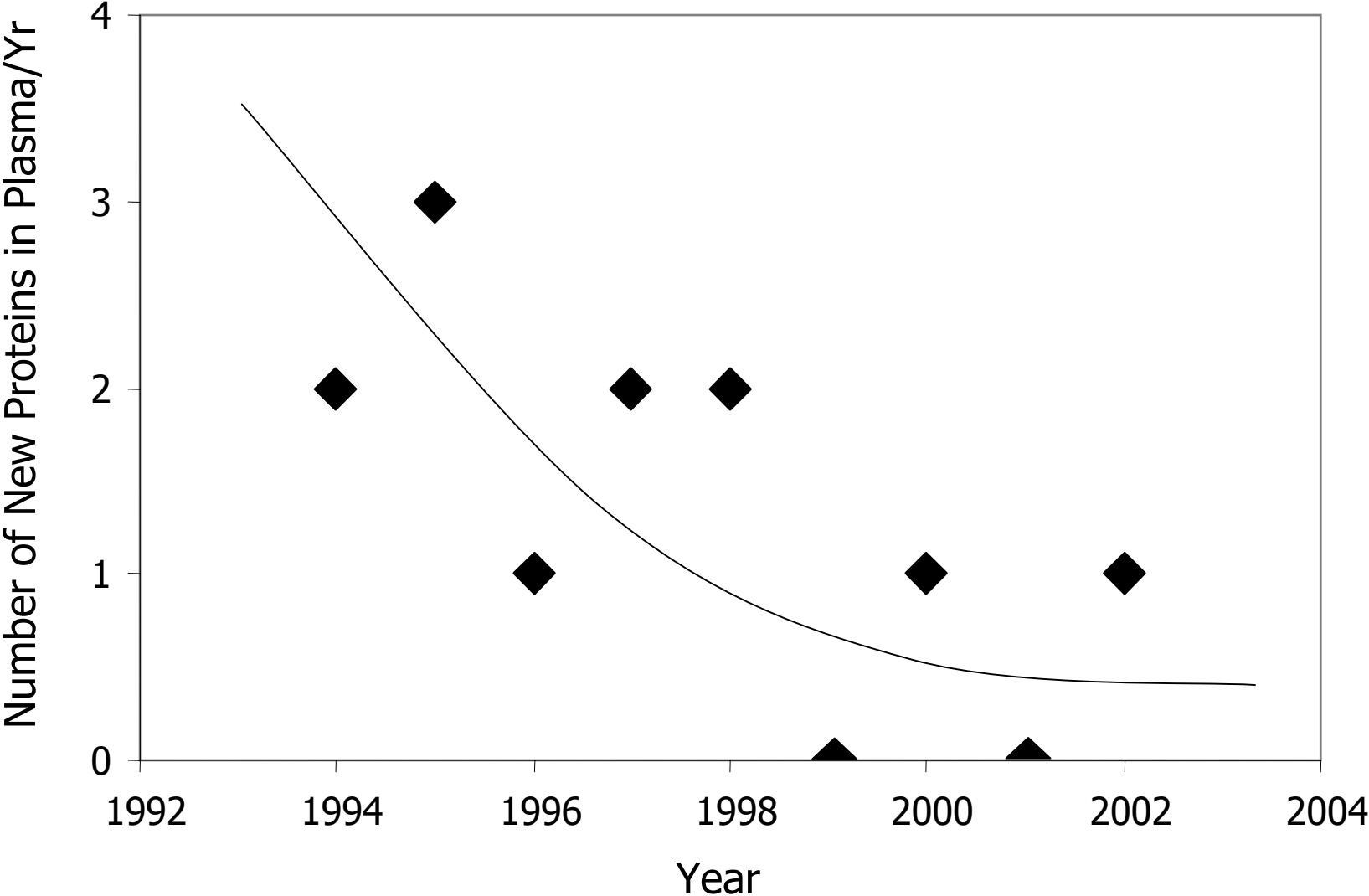


Figure 5

The Human Plasma Proteome

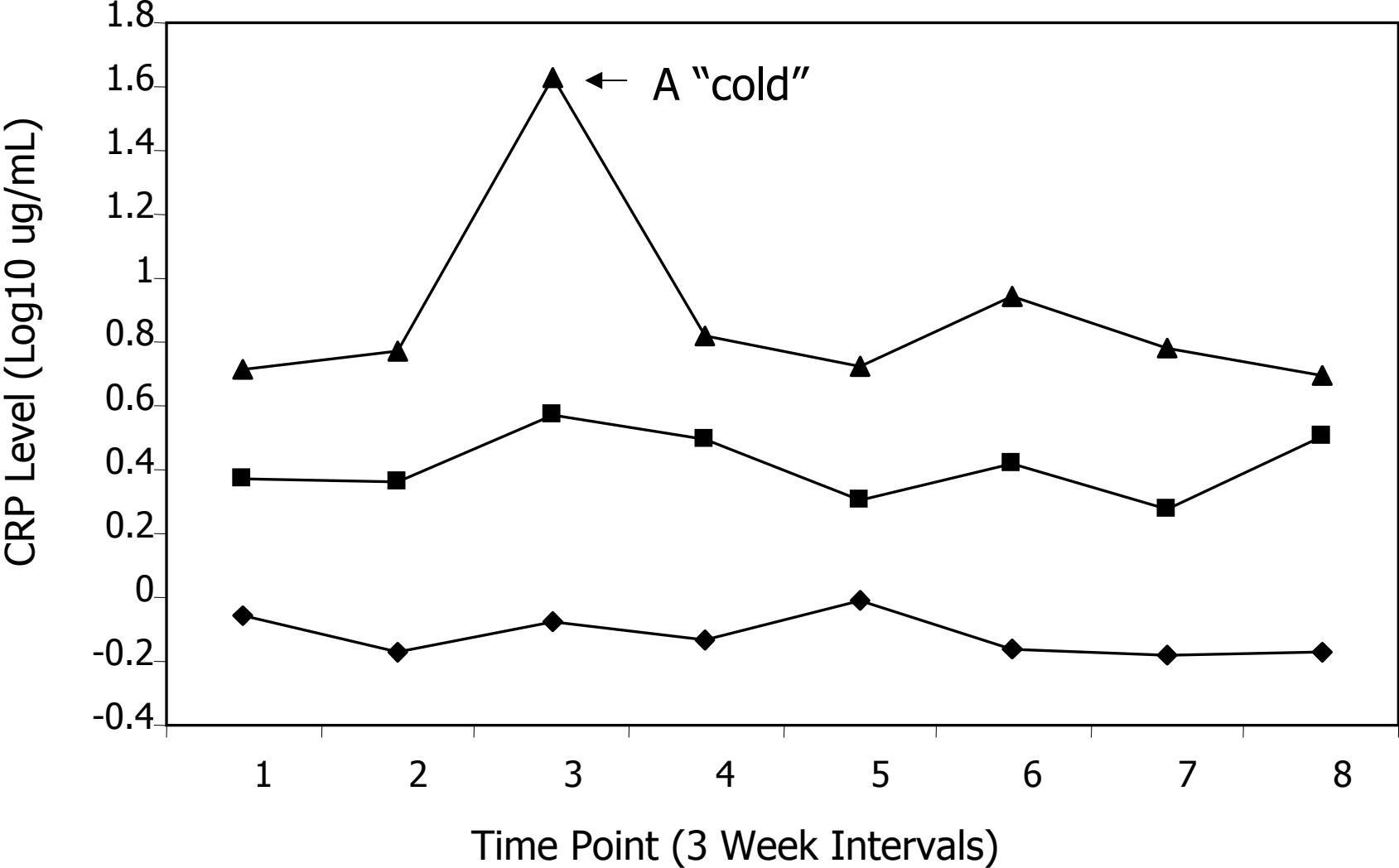


Figure 6