

## PROFILE

# Keck Foundation Biotechnology Resource Laboratory, Yale University

Kathryn L. Stone, Robert D. Bjornson, Gregory G. Blasko, Can Bruce, Renee Cofrancesco, Nicholas J. Carriero, Christopher M. Colangelo, Janet K. Crawford, J. Myron Crawford, Nancy C. daSilva, Joseph D. Deluca, James I. Elliott, Margaret M. Elliott, P. John Flory, Ewa J. Folta-Stogniew, Erol Gulcicek, Yong Kong, TuKiet T. Lam, Ji Y. Lee, Aiping Lin, Mary B. LoPresti, Shrikant M. Mane, Walter J. McMurray, Irina R. Tikhonova, Sheila Westman, Nancy A. Williams, Terence L. Wu, Hongyu Zhao, and Kenneth R. Williams\*

*Keck Foundation Biotechnology Resource Laboratory, Yale University, 300 George Street, New Haven, Connecticut*

## HISTORY

The W.M. Keck Foundation Biotechnology Resource Laboratory (<http://keck.med.yale.edu/>) began in 1980, when its precursor (the Protein Chemistry Facility, PCF†) was founded by Drs. Kenneth Williams and William Konigsberg in the Department of Molecular Biophysics & Biochemistry. Today, the Keck Lab provides more than 175 state-of-the-art genomic, proteomic, biostatistical, bioinformatics, and high performance computing technologies to hundreds of Yale and non-Yale investigators whose research programs otherwise

may not benefit from the highly sophisticated and expensive instrumentation upon which biological and biomedical research is increasingly dependent.

The genesis of the Keck Lab arose from the two years (1978-9) needed by Williams and other staff to sequence the 301 amino acids of a single-stranded DNA binding protein (gp32) from bacteriophage T4 [1]. During this time, Williams was a postdoctoral associate with Konigsberg, and his research required a complete sequence of gp32 for continued progress. Although a neighboring group of investigators had acquired an auto-

---

\*To whom all correspondence should be addressed: Kenneth R. Williams, PhD, Keck Foundation Biotechnology Resource Laboratory, Yale University, 300 George Street, New Haven, Connecticut. Tele: 203-737-2206; E-mail: [kenneth.williams@yale.edu](mailto:kenneth.williams@yale.edu).

†Abbreviations: PCF, Protein Chemistry Facility; MudPIT, Multi-dimension Protein Identification Technology; ICAT, Acid-Labile Isotope Coded Affinity Tag; iTRAQ, Isobaric tagging technology for relative and absolute quantitation; MRM, Multiple Reaction Monitoring; DIGE, Differential 2D Fluorescence Gel Electrophoresis; SILAC, Stable isotope labeling by amino acids in cell culture; N/A, Not applicable.

The Keck Laboratory currently or within the last year has been partially supported by NIH grants and contracts: U24 NS 051869, DA018343, N01-HV-28186, U54 AI057158, CA016359, RR019895, RR 022419, RR023004, RR023748, RR024617 and UL1 RR024139.

mated peptide sequencer via a National Institutes of Health (NIH) Program Project grant, little time was available on this instrument for the gp32 project. Instead, most of the gp32 sequence had to be determined by manual Edman degradation, with much of this work carried out by a staff member, Mary LoPresti. Each 15 residue tryptic peptide required three weeks to generate the 15 resulting amino acid derivatives that would then be identified by reverse phase (RP) HPLC. The experience of spending two years manually sequencing peptides that could have been sequenced “automatically” on a nearby instrument that operated unattended and continuously once the sample and reagents were loaded forever imbued in Williams the desire to bring biotechnology instrumentation within equal and sufficient access of all Yale investigators.

While virtually every university now has one or more biotechnology core laboratories, there were few such cores in 1980, with the Keck Lab being among the first and developing into one of the largest. The PCF began in a small section of Konigsberg’s laboratory. Initially, the PCF was equipped with a Beckman 121M Amino Acid Analyzer with the apt serial number “007,” which contained seemingly endless miles of capillary tubing that often led “nowhere” — as the instrument had been extensively “customized.” The first two PCF staff members were LoPresti (1980) and Kathy Stone (1982), who both still work in the Keck MS/Proteomics Resource. Since 1980, the Keck Lab has grown to 12 Resources (Table 1), each with its own director and budget.

When the Keck Lab needs to expand into a new area of biotechnology, it does so to the maximal possible extent by bringing together and building on existing Yale expertise and infrastructure. Rather than start its own oligonucleotide synthesis resource, the Keck Lab merged in 1988 with the Oligonucleotide Synthesis Core in Genetics, directed by John Flory, PhD, who continues to direct this resource and is one of two associate directors of the Keck Lab. In 1998, the Keck Mass Spectrometry (MS) Resource, founded in 1993 by Kathy Stone, merged with the Yale Cancer Center MS Shared Resource directed

by Walter McMurray, PhD. The resulting YCC/Keck Proteomics Resource brought a wealth of complementary MS experience to bear on biological and biomedical research. McMurray’s experience included analysis of lunar samples from the Apollo 11 mission.

## TECHNOLOGIES PROVIDED

### *Survey suggests an unusually wide range of services*

The Keck Laboratory helps investigators compete for grants by providing access to state-of-the-art biotechnologies, including many that are seldom offered by academic core laboratories, which increases the competitiveness of grant applications proposing use of these technologies. According to a 2006 proteomics survey by Keck staff of 25 core laboratories at institutions similar to Yale or having large biotechnology cores, the Keck Laboratory provides competitive service charges and a very wide range of technologies. Of the 20 major proteomics/MS services surveyed (all of which are available from the Keck Lab), the average non-Yale academic core lab provided four services — with a range of 0 to 12. Besides the Keck Lab, no other core surveyed offered SEC/LS determination of the native MW of proteins or FT-ICR MS; only two other cores offered iTRAQ protein profiling; and only two other cores offered DIGE profiling with MALDI-MS/MS protein identification. The median service charge for the 20 proteomics technologies surveyed was \$128 for the Keck Lab as compared to \$139 for all 25 core labs; only the Keck Lab offered all 20 services surveyed.

While the Keck Laboratory was built on the foundation provided by established proteomics technologies (e.g., amino acid analysis), it expanded to include established genomics technologies (e.g., oligo synthesis) and several emerging technologies. Table 1 and the following sections provide brief descriptions of technologies available from the 12 Resources that are the Keck Lab. Table 2 lists the major instruments in each Resource. A listing of a few recent publications illustrating the use of selected Keck technologies in research is at

**Table 1:** Brief Overview of Resources within the Keck Laboratory

Resource	Major Technologies	Notes
Amino Acid Analysis (1980) <sup>a</sup>	Hydrolysis, amino acid analysis of peptides & proteins	Limit of detection for determining [protein] is about 0.5 µg or 50-200 pmol for individual amino acids, best approach for quantifying [peptide] & [protein].
Bioinformatics (2008)	Free or low cost 24/7 access to software for sequence, microarray, genotyping, MS, & pathway data analysis; pipelining data processing; and molecular modeling using Lasergene, Gene Construction Kit, Genespring GX, Partek Genomics Suite, Gene-Spring GT, HelixTree, Ingenuity Pathway, Meta-Core, Sybyl, Mol-CAD, Pipeline Pilot, VIBE, GPMW.	Two PhD staff provide training on software use; fee-based bioinformatics consultation services; and collaborate on projects requiring longer-term commitment of time and effort, the latter is often funded by an appropriate % effort charged to a grant.
Biostatistics (2002)	Statistical analysis of Affymetrix, Illumina, and in-house spotted expression & SNP arrays as well as of MALDI-MS, iTRAQ, ICAT, DIGE, peptide disease biomarker discovery, and other proteomics data using open source and commercial software and in-house developed programs.	Developed novel statistical methods to analyze microarray gene expression data using pathway-based tests [2]; SNP association data; and a customized, Random Forest disease biomarker discovery algorithm for analysis of MALDI-MS data acquired on sera [3,4]
Biophysics (1999)	Microcalorimetry, HPLC SEC/laser light scattering (LS), dynamic LS, stopped flow absorption/fluorescence kinetic analysis, surface plasmon resonance, & spectrofluorometry	Quantitative analysis (e.g., kinetics, stoichiometry, thermodynamics & binding affinities) of interactions between biomolecules; many instruments available for direct use by investigators with training and support for projects' design, implementation, & data interpretation.
DNA Sequencing (1989)	Single tube & 96 well sequencing; primer walking, sequence assembly & editing, fragment analysis; high throughput (Solexa) sequencing available through Microarray Resource	Avg read length is 650 bp, but can be as high as 700-750 bp. CY 2007 throughput = 206,387 seq. runs. Sequence 6-10 x 96-well plates/day with 24-hour turnaround and some same day sequencing.
High Performance Computing (HPC) (2005)	Two PhD computer scientists support users of Yale HPC Centers by optimizing, parallelizing, and de novo writing of codes and by use of: message passing (MPI), Linda-like parallel languages, batch queuing (PBS) & other methodologies.	The Yale clusters & other HPC instrumentation are used to carry out many analyses: large scale BLAST comparisons, Pseudogene searches, Electron microscopy image processing, Gaussian evaluation, molecular dynamics simulations of proteins, primer design for gene chips, & use of Matlab, Perl, Python and R to support distributed computing.
Mass Spectrometry (MS)/Proteomics (1993/1980)	Peptide, protein, oligo, small molecule MS; protein identification (ID) by MS/MS peptide sequencing; FT-ICR MS exact mass; ID of posttranslational modifications; LC/MS; differential 2D gel electrophoresis (DIGE), 8-plex Multiplexed Isobaric Tagging Technology (iTRAQ), & LC-MS disease biomarker discovery protein profiling.	DIGE uses Cy-2, Cy-3, Cy-5 in vitro labeling of control vs experimental protein extracts prior to mixing the samples and subjecting to 2D PAGE. The dynamic range of 10 <sup>4</sup> and ability to quantify expression of >1,500 protein spots/2D gel surpasses that of other profiling technologies. DIGE and iTRAQ are the two most powerful and usually quite complementary protein profiling technologies.
Microarray (1999)	Custom glass slide microarrays printed with oligos, cDNAs, & proteins (e.g., Ab and cell lysates). Affymetrix, Illumina, & Nimblegen gene expression profiling & SNP genotyping. High-throughput Solexa DNA sequencing & 384 well plate quantitative PCR	Custom & generic oligo & cDNA microarrays printed for human, mouse, rat, Arabidopsis, Drosophila, and other species. Custom protein array printing. Full service training & project support. >130 publications acknowledging Microarray Resource support.
Oligo Synthesis (1988)	DNA oligo synthesis (40nmol, 200 nmol & 1.0 micromole scales), a wide range of modified DNA oligos can be synthesized – will incorporate most commercially available and some custom synthesized modified amidites; RNA oligo synthesis (1.0 micromole); gel purification.	CY 2007 throughput was 31,980 chargeable oligos. Current charge for a 40 nmol 50-mer is \$28 as compared to \$48 in 1998. Turn-around for normal, unpurified DNA oligos is typically <24 hours for <60-mers.
Peptide Synthesis: Large Scale (1986)	Peptides are made using tBOC chemistry at the 0.5 mmol scale, yields for normal peptides <40 residues are "guaranteed" at >50 mg & >90% purity. Many different unusual amino acids may be incorporated during synthesis	Peptides usually can be made up to 40 residues and often, depending on sequence, up to 70 residues. All peptides that can be purified are chromatographed on a preparative C-18 or C-4 RP-HPLC.
Peptide Synthesis: Small Scale (1994)	Peptides made with Fmoc chemistry at 25, 50 & 100 µmol scales with or without RP-HPLC purification. Avg. yield from 100 µmol synthesis of 5-25mer is 20-200 mg crude, 1-10 mg purified. Many different unusual amino acids may be incorporated during synthesis.	Service is designed for 7-30 residue peptides. Since the difficulty in synthesizing peptides is length, composition, and sequence dependent, some peptides within this range will be difficult to synthesize. Conversely, many peptides can be synthesized that are >30 residues.
Protein Sequencing (1980)	Edman (chemical) sequencing of proteins/peptides blotted onto PVDF membranes (less than 15 mm <sup>2</sup> ) or submitted in <0.1 ml volatile solvent. Max. sample load is <1 nmol, with best results obtained on 10-100 pmol sample.	Limit of detection is 50 fmol of an individual amino acid. In general, 1 to 10 pmol is sufficient to sequence from 5 to >25 residues, with the quality of data and length of sequence that can be assigned increasing with the amount of sample.

<sup>a</sup>(Founding year) for the individual Keck Resource.

[http://keck.med.yale.edu/pdfs/Publications\\_usi ng\\_Keck\\_technologies\\_032808.pdf](http://keck.med.yale.edu/pdfs/Publications_usi ng_Keck_technologies_032808.pdf)

## ESTABLISHED PROTEOMICS TECHNOLOGIES

### *Amino Acid Analysis Resource*

The first Keck Resource established uses cation exchange HPLC, which can tol-

erate reasonable amounts of many non-volatile salts and detergents, and external calibration to quantify amino acids in acid hydrolysates of cell/tissue extracts, proteins, and peptides, as well as in similar undigested samples. Separated amino acids are quantified with post-column ninhydrin derivatization for detection at 570 nm and 440 nm. Amino acid analysis of hydrolyzed samples determines protein/peptide concentra-



or protein. The resulting phenylthiohydantoin derivative of each amino acid derivative is identified and quantified by on-line HPLC. We believe the Applied Biosystems (AB) Procise 494 cLC instrument used is the most sensitive, commercially available instrument. Common uses of this technology are to confirm the N-termini of recombinant proteins, identify limited proteolytic cleavage sites, find sites of post-translational modification (e.g., radiochemical sequencing with [<sup>32</sup>P]-peptides) or of cross-linking (e.g., to DNA or RNA), and to obtain peptide sequences from novel proteins to design oligo-primers for DNA sequencing of the corresponding genes. While trypsin digestion followed by MS/MS analysis of the resulting peptides is the method of choice for protein identification, Edman sequencing often is the best approach for determining N-terminal sequences of peptides/proteins that are > 3,000 Da (the approximate upper limit for “standard” MS/MS-based peptide sequencing). Exceptions are intact proteins from higher eukaryotes such as mammals where ~80 percent of the proteins are N-terminally acetylated, which blocks Edman sequencing [6]. In contrast, N-terminal acetylation rarely occurs in prokaryotic proteins and eukaryotic proteins expressed in bacteria.

*Peptide Synthesis Resource:  
Small Scale*

Small-scale Fmoc peptide synthesis is used by this Resource to synthesize > 1,000 custom peptides annually and is generally suitable for seven to 30 residue peptides. Since the degree of difficulty in synthesizing peptides is length, composition, and sequence dependent, some peptides within this range will prove to be difficult to synthesize. Conversely, many peptides longer than 30 residues can be synthesized. A wide range of modified amino acids may be incorporated into small-scale synthetic peptides, with the primary limitation being the commercial availability of the required Fmoc derivative. Synthesized peptides are confirmed by mass spectrometry (MS) to have the expected MW and are accompanied by an analytical RP-HPLC profile.

*Peptide Synthesis Resource:  
Large Scale*

Synthetic peptides routinely are made up to 40 residues and often, depending on their sequence, up to 70 residues. Peptides that can be purified are chromatographed by RP-HPLC and then are delivered as a lyophilized material. Yields for “normal” peptides (which are made at the 0.5 mmole scale with tBOC chemistry) under 40 residues are > 50 mg at + 90 percent purity. For the incorporation of unusual amino acids, such as non-radioactive isotopes, the appropriate tBOC and side-chain protected material must be supplied by the submitter at a level of 2 mmoles/residue. The presence of the synthesized peptide with the expected mass is confirmed by MS.

## **EMERGING PROTEOMICS TECHNOLOGIES**

*Mass Spectrometry and Proteomics  
Resource*

A broad spectrum of MS based techniques, HPLC, and chemistries are used to separate, characterize, and quantify analytes from complex biological samples. The 11 staff members include four PhD-level appointments with > 100 years of MS and protein chemistry experience. This Resource has nine state-of-the-art tandem mass spectrometer systems, including TOF-TOF, Quadrupole-TOF, Triple-Quadrupole, FT-ICR, and LTQ-Orbitrap type analyzers with either matrix assisted laser desorption ionization (MALDI) or electrospray ionization (ESI) sources coupled to HPLCs (Table 2). MS samples include proteins, oligonucleotides, lipids, carbohydrates, synthetic peptides, and many small molecules. In 2006, for example, 14,077 MS analyses were completed for 24 Yale departments and 86 outside institutions.

To address the rapidly increasing interest, this Resource offers nine proteome profiling technologies (Table 3). The methods depend upon either protein or tryptic peptide separation and quantitation. Protein separations use either differential 2D gel elec-

**Table 3:** Overview of Proteome Profiling Technologies Available from the Keck Laboratory MS and Proteomics Resource at Yale University

Technology	Requires Labeling?	Detect PTM's	Max. # Proteins Detected	Comments
MudPIT [7]	No	Yes	872 [8]	2D LC approach used to catalogue proteins from complex samples
ICAT	C <sup>12</sup> /C <sup>13</sup> cleavable ICAT reagent at Cys	No	491 [9]	2-plex; only detects Cys-containing proteins, cannot generally detect protein post-translational modifications
iTRAQ	Isobaric tags at N-terminus & epsilon-N of Lys of peptides	Yes	685 [10]	4-plex & 8plex MS/MS quantitation based on intensity of 114 -117 reporter ions
SILAC	Stable isotope labeled peptides	Yes	5,111 [11]	Protein samples are mixed after cell harvest; any number of MS based fractionation steps can be used with minimal quantitation error
DIGE [12, 13]	In vitro with Cy fluoro-phores at primary amines OR at cysteines	Yes	1500-2000+ [14]	3 samples compared in same gel; very useful for protein isoform detection
PF2D Automated 2D Protein Fractionation	No	Yes	2,500 [15]	Uses 2D Chromato-focusing/ and reversed phase HPLC to quantify pairwise RP fractions by UV
Label Free Quantitation [1]	No, see [16]	Yes	N/A	LC-MS based disease biomarker discovery tool; follow up required for protein ID
MRM based Targeted proteomics	stable isotope labeled peptides	Yes	100 [17]	Means for Absolute Quantitation; internal peptide standards are synthesized or made via ICAT or iTRAQ
Phospho-proteomics	No	Yes	2,328 [18]	Enrichment step is required; phosphorylation sites are determined; label & label-free techniques are used for quantitation

trophoresis (DIGE) with MALDI-MS/MS identification of tryptic digests of proteins in spots of interest or automated 2D HPLC. While fluorescence labeling allows DIGE to analyze up to three proteins in the same gel, 2D HPLC analyzes one protein extract per analysis with quantitation based on the A<sub>210</sub> nm absorbance of the RP-HPLC second dimension. The remaining seven technologies (Table 3) involve analysis of tryptic or other digests of protein extracts. These “bottom-up” approaches rely on tandem mass spectral peptide identification and can be quantitative with chemical or stable-isotope tagging of proteins or peptides (e.g., iTRAQ). The web-based YPED platform transmits profiling data to users and also archives the data and is being equipped with tools to integrate, analyze, and visualize the results. Two new technologies are phosphoproteome profiling and quantitative analysis of pre-selected, poten-

tial biomarker proteins. We also are launching a label-free quantitation technology for disease biomarker discovery using FT-ICR LC-MS to analyze trypsin digests of complex cell and tissue extracts.

## ESTABLISHED GENOMICS TECHNOLOGIES

### Conventional DNA Sequencing

Four technologies are provided: single tube and high-volume, 96-well plate DNA Sequencing; fragment analysis; and primer walking. A Tecan Genesis workstation aliquots reagents and samples, while a Biomek NX system carries out post-cycle sequencing cleanup via Agencourt's CleanSEQ methods. We plan to implement new services that would allow researchers to submit samples in 384 well format or in

**Table 4:** Overview of Keck Technologies Available to Study Genetic Events

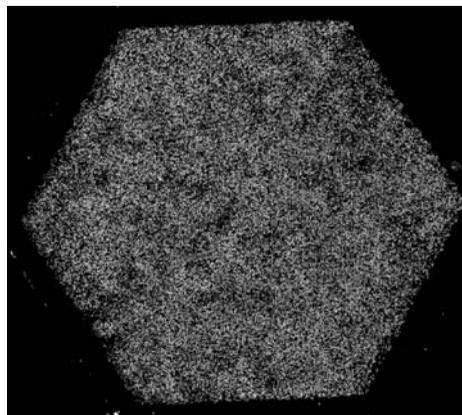
Genetic Events	Common Detection Methods	Applicable technology/platforms
Genetic Susceptibility	Linkage and Association	Affymetrix, Illumina, Sequenom, Nimblegen and Solexa
Somatic and germ line mutations	SNP identification and genotyping	Affymetrix, Illumina, Sequenom, Nimblegen and Solexa
Gene amplification and deletions	Comparative genome amplification analysis	Affymetrix, Illumina, Sequenom, Nimblegen and Solexa
	Gene and Exon-level amplification analysis	Affymetrix, Illumina, Sequenom, Nimblegen and Solexa
Epigenetic modifications	Methylation, PCR/bisulfite sequencing, ChIP	Illumina, Sequenom, Nimblegen, Solexa and Affymetrix (tiling array)
Differential gene expression	Gene and Exon-level analysis	Affymetrix, Illumina, Sequenom, Nimblegen and Solexa
Alternative splicing	Exon level analysis	Affymetrix
Gene regulation	ChIP	Affymetrix, Nimblegen, Solexa
SNP discovery	Direct sequencing	Solexa
miRNA discovery	Direct sequencing	Solexa
miRNA Analysis	Microarray and sequence analysis	Exiqon miRCURY arrays, Illumina and Solexa
Sequencing and resequencing	Direct sequencing	Solexa

four 96 well plates to have them combined into a 384 well plate. Use of the 384 well plate would reduce reaction volumes from 10  $\mu$ l to 5  $\mu$ l, which would reduce reagent costs and lower service charges. We also would hope to offer PCR product cleanup for researchers submitting PCR samples in high volume plates.

### *Oligonucleotide Synthesis*

Nucleic acid syntheses use beta-cyanoethyl chemistry on 13 instruments (Table 2). Oligo synthesis is offered at three DNA scales (50, 200, and 1,000 nmol) with lengths up to 250-mers and a wide range of synthesis services, including derivatized DNA oligos containing a large variety of structures; RNA and 2'-O-methyl RNA and chimeras of these and DNA, plus derivatives as above; phosphorothioation; and gel purification of DNA oligos. New modifications are offered as they become available, and procedures are optimized for each. DNA oligos are delivered fully deprotected, lyophilized, and unpurified with hydroxyl groups at both ends. Exceptions are:

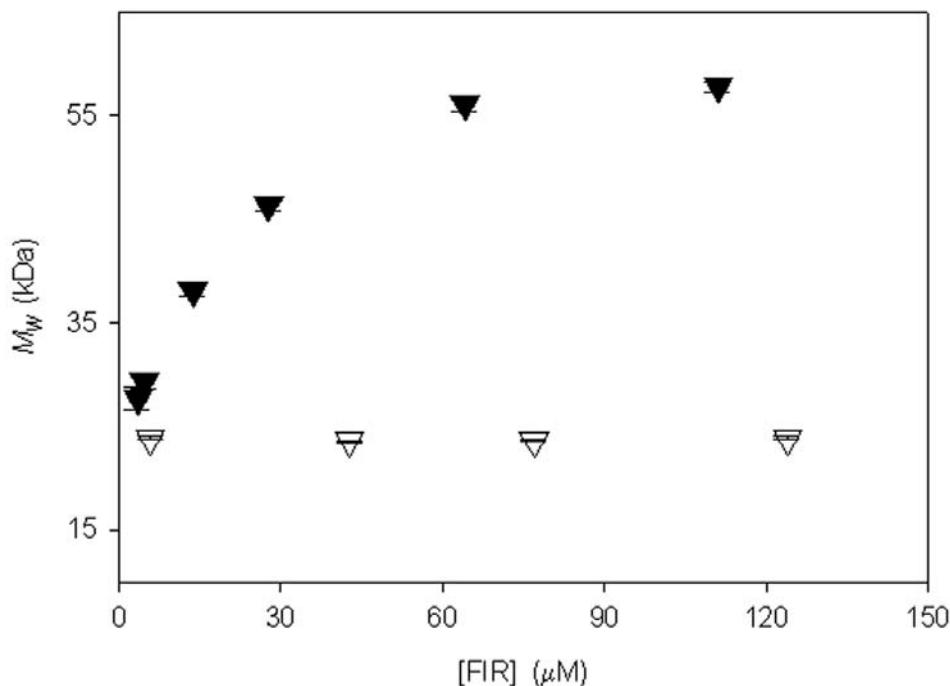
- Trityl-On oligos are supplied on the support without cleavage or deprotection;



**Figure 1.** Image of an Illumina 1,500 custom SNP panel.

- RNA oligos are delivered with a 2'-protecting group to prevent degradation;
- Modified oligos requiring UltraMild deprotection (i.e. TAMRA-dT, etheno-dA, cy3, cy5 etc) are supplied cleaved and deprotected in 3 mls potassium carbonate/MeOH/TEAA and must be desalted before use.

The quality of unpurified products is maximized by using high quality reagents and optimizing synthesizer cycles. Oligo syntheses are monitored by trityl yields and by analyzing about 20 percent of oligos by



**Figure 2.** DNA induced dimerization of FIR-protein. Oligomeric State of FIR alone and of the FIR-DNA complex from SEC-UV/LS/RI analysis. Weight average molecular weights determined from SEC-UV/LS/RI analyses are plotted as filled triangles for FIR-DNA complex and open triangles for FIR protein alone. Averages and standard deviations were calculated from 25 MW determinations for the top 0.2 ml of the eluting peaks for which the concentration is within 5 percent of the plotted value.

capillary electrophoresis. The quality of our unpurified oligos is sufficiently high that they usually are used directly (or after desalting) for sequencing or PCR. Although the website provides anticipated DNA yields/base, we urge users to determine oligo DNA concentrations before use. For sequences > 100 bases, we recommend 200 nanomole or 1.0  $\mu\text{mole}$  scales. The turn-around time for normal DNA oligos is typically < 24 hours for < 60-mers.

## EMERGING GENOMICS TECHNOLOGIES

### Microarray Resource

This full-service Resource (Tables 1 and 4) is dedicated to providing RNA expression profiling, DNA genotyping (Figure 1), high throughput DNA sequencing, and microRNA analysis services using Affymetrix, Illumina, NimbleGen, Solexa, Sequenom,

and Applied Biosystems 7900 instrumentation, as well as in house spotted arrays. A supplement to the Yale Cancer Center NCI Core Grant helped fund the instrumentation needed to initiate the spotted glass section of the Microarray Resource in 1999. The Affymetrix platform was established in 2001 to support carrying out large-scale gene expression studies utilizing commercially available microarrays. The Microarray Resource has 10 staff and occupies 5,350 square feet on the second floor at 300 George Street and contains Class 100 clean rooms for printing and slide processing. From July 2004 to June 2007, this Resource provided about 18,000 services to 537 researchers from Yale and 313 from 160 other institutions. To our knowledge, there have been about 140 publications utilizing its services. It has emerged as one of the leaders in the identification of disease-causing genetic factors as evidenced by recent publications in *Science* identifying genes associated with

age-related macular degeneration [19] and coronary disease and metabolic risk factors [20]. In addition to the NIH Center and High End instrumentation grants mentioned below, this Resource also obtained an administrative supplement that partially funded its Illumina microarray system.

## SUPPORTING BOTH GENOMICS AND PROTEOMICS

### *Biophysics Resource*

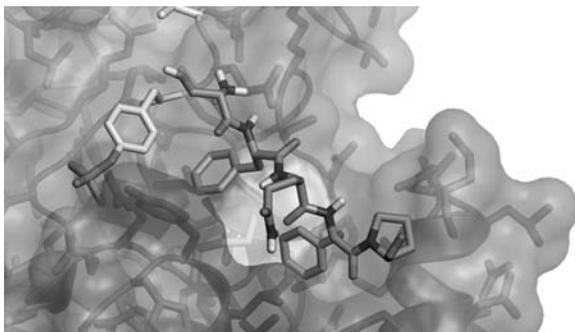
This Resource provides technologies that allow characterization of interactions between biomolecules, including the oligomeric state of the interacting species, thermodynamic parameters that govern the interactions including binding constants, the enthalpic and entropic contributions to complex formation, and kinetic information:  $k_{\text{on}}$  and  $k_{\text{off}}$  rates. The instrumentation in the resource (Table 2) allows determination of the oligomeric state of the interacting species and of the resulting complex utilizing size exclusion chromatography/laser light scattering (SEC/LS) or dynamic laser light scattering (DLS). As illustrated in Figure 2 for a DNA-binding dependent dimerization of FIR protein [21], binding constants are determined using isothermal calorimetry (ITC), surface plasmon resonance (SPR), or an SLM 8000C spectrofluorometer; kinetics using stopped-flow and SPR; and the enthalpy and entropy of binding reactions using ITC — which is an almost universal technology for studying macromolecular interactions that is based on the heat that interactions give off or take up upon complex formation. Thus, this approach does not require labeling of the interacting species. Similarly, (SPR) also is used to study label-free macromolecular interactions. SPR detects binding in real time by monitoring changes in mass concentration at the chip surface; the association and dissociation rate constants are determined from the reaction traces. Samples ranging from small molecules to crude extracts, lipid vesicles, viruses, bacteria, and eukaryotic cells can be studied in real time with little or no sample preparation.

The best methods for each project are recommended based on the questions being addressed, the amount of available sample, and the spectroscopic properties of the interacting molecules. Samples are accepted for analysis on a “fee-for-service basis” or through “open access,” whereby investigators have *direct* use of the needed instrumentation. The resource provides instrument training and support for each project’s design, execution, and interpretation of the resulting data. This approach allows investigators with little knowledge of biophysics to successfully complete advanced biophysical analyses that have been designed to best address their research challenges.

A new NIH SIG will fund Asymmetric Flow Field-Flow Fractionation (AFFF) and Automated Composition Gradient (CG) syringe delivery systems that will share Light Scattering (LS) and other detectors. AFFF is a single phase chromatography technique that can separate samples from 1nm to > 20 microns. High-resolution separation by size is achieved within a very thin channel against a perpendicular flow force. The entire separation is gentle, rapid, and non-disruptive — without a stationary phase that may degrade, bind, or otherwise alter the sample. The combination of AFFF fractionation and light scattering (LS) detection allows sample fractionation and determination of size and molar mass in a single experiment. The coupling of the LS measurement to a fractionation step provides the ability to determine the molar masses and oligomeric states of very diverse macromolecules (e.g., proteins and their complexes, nucleic acids, liposomes, and polysaccharides). In contrast to ultracentrifugation, fractionation and sizing on AFFF/LS allows the facile collection of fractions for further analyses.

### *Biostatistics Resource*

This Resource provides state-of-the-art statistical data analysis of genomics, genetics, and proteomics research using open source, commercial software, and in-house programs. The Biostatistics Resource has collaborated with the Keck Microarray and Proteomics Resources and the Center for Medical Infor-



**Figure 3.** Molecular model for DEF site-MAPK interactions. Docking of a capped pentapeptide ligand representing the DEF sequence (acetyl-SFQFP-amide) to the active form of ERK2 was modeled using AutoDock. This is a joint research project involving Dr. Y. Kong of the Bioinformatics Resource with Dr. B. Turk and Dr. D. Sheridan of the Department of Pharmacology, Yale University.

matics to build the institutional, Web-based Yale Microarray Database (YMD) and Yale Protein Expression Database (YPED). The Biostatistics Resource also collaborates with the Yale Center for Statistical Genomics and Proteomics to develop pathway and protein interaction database and visualization tools and collaborates with and complements the Bioinformatics Resource. Finally, the Biostatistics Resource works with the High Performance Computing (HPC) Resource on those challenges amenable to an HPC solution. While fee-based services are provided for well-defined projects, this Resource also is involved in biostatistical research and developing novel statistical methods to analyze gene expression data [2] and genetic association and proteomics data [3,4]. Where success requires deeper scientific involvement, the Biostatistics Resource will work collaboratively. In fiscal year 2007, the Biostatistics Resource carried out 61 services for 25 Yale users from 14 departments and centers as well as 26 services for 15 users at 14 non-Yale institutions. The Biostatistics Resource currently has research collaborations with Drs. Judy Cho (Internal Medicine), Joel Gelernter (Psychiatry), and Zoran Zimolo (Psychiatry) at Yale University and also with groups of non-Yale investigators at Intrinsic Bioprobes, Inc.

### *Bioinformatics Resource*

This Resource's mission is to provide bioinformatics support at three levels. First, free or subscription-based 24/7 access is provided to many commercial and open source bioinformatics programs. Some software is loaded on Resource PCs, while other soft-

ware can be used remotely, either through client programs or with a Web browser. Software provided by the Resource are loaded on Windows and Linux workstations available 24/7 in the Resource; a few programs are subscription, Web-based services accessible with a Web browser; and some are remotely accessible after installation of a client program on the user's PC. The available software covers a wide range of applications, including analysis of DNA/protein sequences, microarray expression and genotyping data, pathway and network analysis, protein structure modeling and docking (Figure 3). Bioinformatics staff will recommend the most appropriate software to meet individual research objectives and provide training in the use of these resources. Second, the staff will provide fee-based consultation services for well-defined bioinformatics analyses. Third, the staff will collaborate on projects requiring a longer commitment of time and effort, which often will be charged to an appropriate grant.

The Bioinformatics Resource should very positively leverage the value of the technologies provided by the Biostatistics and several other Keck Resources, including Microarray, MS/Proteomics, and DNA and Protein Sequencing. Finally, the clusters and other instrumentation in the Yale Biomedical High Performance Computing (HPC) Center and the computer scientists in the HPC Resource work with the Bioinformatics Resource on tasks amenable to an HPC solution.

### *High Performance Computing (HPC) Resource*

The HPC Resource plays a critical role in the new Yale University Biomedical Cen-

ter for High Performance Computing (see below). The HPC Resource is co-directed by two Ph.D.-level computer scientists. Together, they provide the parallel programming knowledge for bringing HPC to bear on research. The interaction between users and the Resource varies widely, depending upon the interest and ability of each user. In some cases, experienced investigators are provided with accounts and proceed on their own. Usually, however, the HPC center's staff consult with users to some degree. Typically, one of the center's computer scientists examines and benchmarks user's codes, often finding ways to improve the serial performance or to parallelize the code or both.

## **EVOLUTION OF CHALLENGES AND GENERAL OPERATING POLICIES**

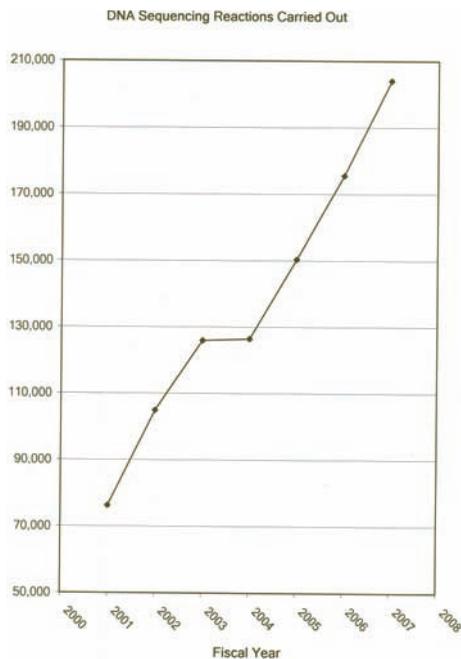
### *Charging for services*

To avoid having its financial stability depend completely upon grants, the Keck Lab adopted a different model in which user fees play a major role. In 1980, the idea of each user paying for the actual cost of carrying out its biotechnology analyses and syntheses was quite counter to the prevailing mechanism of obtaining access to expensive biotechnology analyses and syntheses by collaborating with faculty who had the needed instrumentation. Williams encountered such strong opposition to charging for the cost of carrying out biotechnology services that he was one of the six founders of the Association of Biotechnology Resource Laboratories (ABRF, <http://www.abrf.org/>). Soon after its incorporation in 1988, the ABRF carried out a survey of core laboratories [5]. A significant finding was that the extent of subsidization of core lab operating expenses varied over a wide range. This makes it quite misleading to simply compare web-posted service charges at different core laboratories. The ABRF continues to play a valuable role as "an international society dedicated to advancing core and research biotechnology laboratories through research, communication, and education."

### *Expansion creates a continuing financial challenge*

The expansion of the Keck Laboratory from 1980 to 2008 created a continuing financial challenge. While several factors enter into the Keck Laboratory's decisions about which new technologies to provide — with the potential positive impact on research being paramount — one guideline is to focus on services requiring instrumentation that is too expensive to be purchased by individual research laboratories and requires a high level of expertise and large number of samples/requests to keep the instrumentation operating at its maximum capabilities. Our experience shows that most instruments — especially those with fluidics and valves — have the lowest malfunction rate when operated continuously. The latter is difficult for a single research laboratory to achieve — where demand for individual technologies tends to ebb and flow. Since the Keck Lab provides technologies to hundreds of Yale and non-Yale laboratories at hundreds of institutions, it is often able to maintain a backlog of non-Yale requests that it utilizes during those times when demand by Yale investigators is below average. A major challenge created by continuing expansion of the Keck Lab is that the purchase value of its instrumentation is now about \$17 million. Assuming an average instrument life span of seven years requires that the Keck Lab obtain \$2.4 million dollars annually just to replace obsolete and worn-out equipment — and that does not include purchasing equipment needed to expand to meet increasing demand for existing technologies nor to provide new technologies.

An important source of instrument funding for the Keck Lab has been the annual NIH Shared Instrumentation Grant (SIG) program, which funds instrument systems in the \$100,000 to \$500,000 range, and a "High End" variant of this program that funds instruments in the \$1 to \$2 million dollar range. However, even if the Keck Lab were able to obtain a "standard" SIG each year, this would provide only 20 percent of the \$2.4 million dollars per year needed to sustain *existing* biotechnologies. Since 1981,



**Figure 4.** “Conventional” DNA sequencing reactions carried out from fiscal year 2001–2007 by the Keck DNA Sequencing Resource. The Yale University fiscal year extends from July 1 through June 30.

the Keck Lab has submitted 25 NIH/NSF SIG applications with 21 (84 percent) being funded. Among them have been three “High End” awards that funded an FT-ICR MS, most of the instrumentation in the Biomedical High Performance Computing Center, and “next generation” DNA sequencing equipment, including a Solexa instrument. Taken together, the 21 grants have covered about 48 percent of the purchase cost of instruments that are “online” in the Keck Lab. Other sources have included institutional funds and user fees.

#### *Advantages of providing support for instrumentation in core laboratories*

The institutional advantages of having a well-equipped biotechnology core laboratory are numerous. Although impossible to quantify, we believe state-of-the-art biotechnology support from core laboratories leads to increased research productivity, publications, grant funding, and indirect costs for the institution. Additionally, “cutting edge” core

laboratories can enhance recruitment and retention of the very best faculty, research staff, and students. Core laboratories also provide access to well-trained staff, who typically have many years of expertise with a wide range of sample types/requests. By contrast, research laboratory staff may not have the knowledge and experience needed to optimally prepare samples for a given technology while also minimizing sample loss; may not have the advantages gained from having attended training courses given by manufacturers of biotechnology instrumentation; may lack the experience needed to operate advanced instrumentation at the limits of its capability; and may fail to quickly detect subtle indicators of impending malfunctions of instrumentation. In comparison to the approximate \$436,200 cost of equipping each laboratory in a university that needs access to a quantitative protein profiling technology such as Multiplexed Isobaric Tagging Technology (iTRAQ™), it is less expensive for universities to equip a core laboratory with the instrumentation needed to bring this technology within equal reach of all interested investigators. Although there are some instances in which it is worthwhile to equip an individual laboratory with expensive biotechnology instrumentation (e.g., if they wish to devote considerable instrument time to teaching — perhaps by integrating it into laboratory courses, are carrying out research on the instrumentation itself, have unusually large and continuing demand for the instrument funded by multiple grants, or have specialized requirements unique to their laboratory that require significant modifications of the instrument), most often we believe this option is not best for the institution.

#### *Commercial vs. institutional core laboratories*

The Keck Lab is sometimes asked why it offers technologies such as “conventional” DNA sequencing and oligonucleotide synthesis, which are offered at lower fees by some commercial laboratories. The simple answer goes back to the foundation upon which the Keck Lab was built: to meet the needs of the Yale scientific community. Over the last seven

years, the demand for DNA sequencing has increased about 18 percent annually (Figure 4). In fiscal year 2007, the Keck Lab carried out 204,024 DNA sequencing analyses with > 88 percent of the requests coming from 376 Yale investigators. It seems likely that the increasing demand for this Keck service results from faster turnaround, higher quality data, higher success rate, more personalized service, more responsive staff, and more extensive assistance with analysis of multiple datasets, compared with the services provided by commercial DNA sequencing companies. Similarly, in fiscal year 2007, the Keck Lab synthesized 32,051 custom oligonucleotides, with > 86 percent of these requests coming from Yale, which represented a 10 percent increase from 2006. By using three overlapping staff shifts, this Keck Resource usually is able to provide < 24 hour turnaround for unpurified, 40 nanomole DNA oligos that are < 60-mers. Another factor contributing to the increased use may be the very broad range of modified oligonucleotides (many of which may not be available commercially) that have been synthesized successfully by this Resource. In the case of state-of-the-art biotechnologies, such as the phosphoproteome profiling technology recently introduced by the Keck MS/proteomics Resource, we have not been able to find any commercial vendor that provides a comparable technology. Other biotechnologies (e.g., the SEC/laser light scattering technology from the Biophysics Resource) appear to be more expensive from commercial vendors than from the Keck Lab.

### *Key operating policies*

To maximize its positive impact on research, the Keck Lab strives to provide as many high-quality services to as many investigators as possible. Although priority always is given to Yale investigators, accepting requests from scientists from across the United States and around the world helps ensure the backlog required to maintain high productivity. This policy, which benefits all users, minimizes operating costs by increasing productivity, maximizes the contribution of the Keck Lab and Yale University to research, and also contributes to the high success rate the lab has

had at obtaining SIGs. Virtually every Study Section review of a Keck SIG application contains statements such as: “The Keck Foundation Biotechnology Resource Laboratory at Yale University is the premier biotechnology resource center in the world and is a model for such facilities on a scale that most universities cannot even contemplate. The contribution of this lab to biomedical research in the U.S. has been and will continue to be enormous. They are a role model for how a core lab should work.” In keeping with this philosophy, most analyses and syntheses are provided as services, since it would not be feasible for the Keck Lab to collaborate with even a small fraction of the 993 investigators from 280 institutions in 22 countries who utilized 255,559 Keck services in fiscal year 2007. Whether services are carried out on a service or collaborative basis, however, each Yale user is given the same first-come, first-served priority and turnaround. Unless an instrument malfunctions or the sample is being used to test or optimize a new procedure, the cost of carrying out each service is charged either to a user or a Keck grant.

### *Space: The Ultimate Challenge*

The Keck Lab received its current name in 1989, when Yale was awarded a grant from the W.M. Keck Foundation to build 3,350 square feet of customized biotechnology space in the Boyer Center for Molecular Medicine. The volume of services continued to increase, and over the next decade, the Keck Lab was forced to borrow additional space from departments scattered throughout the School of Medicine. By 2001, it had become clear that the Keck Lab needed a new home, and the School of Medicine renovated another approximate 25,000 square feet of custom-designed laboratory and support space at 300 George Street — space that now houses 50 Keck staff and about 100 instrument systems.

## **FUNDING OF THE KECK LAB**

### *Important role for center grants*

The value of the Keck Laboratory’s instrumentation has been very positively leveraged by the awarding of several center

and other grants, which provide funding for biotechnological research and subsidized access for center investigators to new technologies. As new and improved technologies and databases are developed by these centers, they are rapidly published, offered as services, and made available to users of the Keck Lab. Examples include implementation and development of many new protein profiling technologies (see above) and two major institutional databases built in collaboration with the Yale Center for Medical Informatics:

- **Yale Microarray Database:** YMD is an Oracle database that archives microarray data and provides tools to retrieve and analyze spotted microarray data. We are in the process of archiving data into YMD that has been generated by the Affymetrix GeneChip microarray platform.

- **Yale Protein Expression Database (YPED):** YPED is an interoperable protein expression database being built to archive, manage, and analyze the very large data sets generated by Keck, other proteomics centers, and investigators to quantify the relative levels of expression of thousands of proteins in hundreds of samples annually.

The Keck Lab contains or is very closely associated with several NIH Centers:

- **Yale/NHLBI Proteomics Center** is one of 10 centers established in 2002. This center supports 19 projects that use protein/phosphoprotein profiling and the development of cell-permeable, synthetic biotechnologies for blocking specific protein:protein and protein post-translational modifications *in vivo*.

- **Northeast Biodefense Center (NBC)** is one of 10 Regional Centers of Excellence established in 2003. The NBC Proteomics Core encompasses six Keck proteomics resources and supports basic and clinical biodefense research programs.

- **Yale/NIDA Neuroproteomics Research Center** is one of two centers established in 2004 that brings together 14 Yale research programs in proteomics and signal transduction in the brain with MS/Proteomics Resource and other Keck staff to identify adaptive changes in protein signal-

ing that occur in response to substance abuse.

- The instrumentation in the Yale Biomedical Center for High Performance Computing (HPC) was funded primarily by a 2004 NIH Instrumentation Grant. Yale ITS provides systems administration and computer scientists in the Keck HPC Resource work with researchers to optimize codes, develop parallel variants, explore new formulations, and support new genomic (e.g., Solexa DNA sequencing) and proteomic technologies as they are brought online by the Keck and other laboratories. Seventy-nine users from 33 laboratories logged 2.25 million CPU-hrs on this center's clusters in 2007.

- The Yale Microarray Center for Research on the Nervous System was established in 2005 as one of four centers to provide DNA microarray services at lower cost to approximately 10,000 neuroscientists funded by 15 NIH Blueprint Institutes, thus supporting a broad range of research, and is located primarily within the Keck Lab.

- **Yale Cancer Center (YCC)** shares the Keck Proteomics/Biophysics and Microarray Resources, both of which were rated as outstanding during the 2007 review of the successful YCC competing grant renewal.

## **TRAINING, EDUCATION AND COMMUNITY SERVICE**

The Keck Lab views education and training of users and the Yale community as one of its most important functions. This education includes training workshops, individual training, user groups, seminars, Web-based training, newsletters, and publications. To help users take maximum advantage of its resources, the Keck Web pages provide information on the biotechnologies it offers and on interpreting the resulting data. Keck staff have trained core laboratory staff from as far away as Argentina and South Korea. Since 2004, Keck staff have presented many posters and have given about 20 seminars at Yale and 18 invited talks at scientific meetings. Responding to the need for more minorities in scientific re-

search, the MS/Proteomics Resource sponsored two minority undergraduate students for a 10-week summer research and science mentorship program funded by Yale BioSTEP. Multiple Keck staff have served as judges in New Haven's annual high school science fair and offered science education presentations and Keck Laboratory tours for local middle school science students and teachers. Keck Resource directors and staff periodically visit and teach after school science classes to area elementary school students. In 2006, Shrikant Mane, PhD, provided hands-on training for GeneChip expression analysis to a Hamden High School science teacher. Similarly, Dr. Tukiet Lam provided scientific outreach to 20 Albertus Magnus College students and teachers by demonstrating FT-ICR MS instrumentation and technologies. The dedication of the Keck Lab to education was mentioned in the summary statement from one of our current NIH SIG awards (RR024617): "The laboratory and staff are also committed to graduate, undergraduate, and high school level educations and it is clear that the availability of an (LTQ-Orbitrap) 'CSI'-type instrument may encourage students to pursue a career in science." And lastly, the Biostatistics Resource has engaged in many education and training activities that included Dr. Zhao organizing the Genome-wide Association Conference (2006) that was well attended.

## CONCLUSIONS

The genomics "revolution" has succeeded in sequencing the human and many other genomes and was made possible by key discoveries in molecular biology (e.g., restriction enzymes) and the amazing rate at which *major* biotechnological breakthroughs were and are continuing to be made in this broad field. As anticipated, the ever-growing knowledge about the human and other genomes as well as the new and very powerful genomics biotechnologies are giving rise to impressive achievements that span from taxonomy to criminal investigations to uncovering genes associated with

human disease. Realizing the importance of spurring a similar revolution in proteomics, NIH has funded biotechnology centers and grants to develop more powerful proteomics technologies. Generally, however, the expectations for proteomics have exceeded the clinical accomplishments. Some of the driving forces for a proteomics "revolution" are the renewed appreciation that the biological effector molecule generally is the protein and not its encoding mRNA; the inability of predicting the occurrence of many important protein post-translational modifications (PTM) such as phosphorylation (which is thought to occur on as many as one-third of human proteins and often plays a key role in modulating protein function) from genomics data; and the frequent lack of agreement between mRNA vs. protein expression data (e.g., see [22]).

Several challenges stand in the way of a true proteomics "revolution," and they are illustrated in human plasma, which is the most complex human proteome but also the most useful as it potentially contains virtually the entire human proteome due to tissue "leakage" and is the most readily available clinical specimen. While there are probably only a relatively modest number of true plasma proteins (e.g., about 500 secreted by the liver and intestines), each is present in an average of perhaps 100 forms due to differential glycosylation, splicing, proteolytic processing, and PTMs [23]. Added to these 50,000 protein variants are perhaps almost 21,000 other human proteins [24] that may leak into the plasma and may each be present in about 50 variant forms (e.g., five resulting from alternative splicing/promoter usage and 10 from the addition of > 200 different PTMs), thus adding another 1,000,000 potential "plasma" proteins that are then mixed with perhaps another 10 million different immunoglobulin sequences [23]. Adding considerably to the challenge is the 10 order of magnitude range in protein concentrations in plasma, which is *many* orders of magnitude above the dynamic range of any current biotechnology such as iTRAQ, DIGE, or immunological platforms. Three approaches used to try to address the

wide dynamic range of individual plasma proteins are depletion of abundant proteins such as serum albumin, enrichment (by immunological or other means) of classes of proteins of interest (e.g., phosphoproteins), and multi-dimensional/multistep approaches used to fractionate plasma prior to proteomic analysis. One need not look very far into the Keck Web pages to discern that while genomics can interrogate the relative level of expression of approximately 38,000 human transcripts on a single array and a single chip can analyze 1 million human SNPs, proteomics is limited to the range of about 500 to 700 (with optimal iTRAQ samples) to 1,000 to 2,000 proteins/sample (with optimal DIGE samples). We believe that to reach the expectations anticipated for proteomics, this several order of magnitude difference between the capabilities of contemporary genomic and proteomics technologies must be closed, and currently, it is not clear if any of the available proteomics technologies have the inherent capability to do so. While we believe that closing this biotechnological gap is one of the most difficult of all biotechnological challenges, we also believe the rewards for doing so will prove to be well worth the needed effort and funding. In our opinion, however, no government agency has yet made the high level of sustained commitment that will be needed to bring the goal of "routine" interrogation of the human proteome within reach.

**Acknowledgment:** We would like to acknowledge the very strong support from the Yale School of Medicine, Deputy Dean Carolyn Slayman, Richard Lifton (Genetics), William Konigsberg (Molecular Biophysics & Biochemistry), Yale University, NIH/NCRR, the many departments who provided temporary space and other support throughout the last 28 years, and the tens of thousands of investigators from hundreds of institutions who have entrusted the Keck Laboratory with their samples and requests for syntheses.

## REFERENCES

- Williams KR, et al. Amino acid sequence of the T4 DNA helix-destabilizing protein. *Proc Natl Acad Sci USA*. 1980;77:4614-7.
- Pang H, et al. Pathway analysis using random forests classification and regression. *Bioinformatics*. 2006;22:2028-36.
- Wu B, et al. Comparison of statistical methods for classification of ovarian cancer using a proteomics dataset. *Bioinformatics*. 2003;19:1636-43.
- Yu W, et al. MALDI-MS Data Analysis for Disease Biomarker Discovery in *Methods in Molecular Biology*. In: Nedelkov D, Nelson R. editors. *New and Emerging Proteomics Techniques*. New Jersey: Humana Press; 2006. p. 199-216.
- Williams KR, et al. The size, operation, and technical capabilities of protein and nucleic acid core facilities. *FASEB J*. 1988;2:3124-30.
- Brown JL, Roberts WK. Evidence that approximately eighty per cent of the soluble proteins from Ehrlich ascites cells are N $\alpha$ -acetylated. *J Biol Chem*. 1976;251(4):1009-14.
- Wolters DA, Washburn MP, Yates JR. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem*. 2001;73:5683-90.
- Washburn MP, et al. Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem*. 2002;74:1650-7.
- Han DK, et al. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol*. 2001;19:946-51.
- Ross PL, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004;12:1154-69.
- Graumann J, et al. SILAC-labeling and proteome quantitation of mouse embryonic stem cells to a depth of 5111 proteins. *Mol Cell Proteomics*. Nov 2007; doi:10.1074/mcp.M700460-MCP200.
- Tonge R, et al. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*. 2001;1:377-96.
- Gygi SP, et al. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci*. 2000;97:9390-5.
- Görg A, Weiss W, Dunn M. Current two-dimensional PAGE technology for proteomics. *Proteomics*. 2004;4:3665-85.
- Betgoveguez E, Simonian MH. Reproducibility and Dynamic Range Characteristics of the Proteome PF 2D System. *Beckman Coulter Application*. 2003; Information Bulletin A-1964A.
- Bondarenko PV, Chelius D, Shaler TA. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal Chem*. 2002; 74(18):4741-9.
- Kalkum M, Lyon GJ, Chait BT. Detection of secreted peptides by using hypothesis-driven multistage mass spectrometry. *Proc Natl Acad Sci USA*. 2003;100:2795-2800.

18. Villén J, et al. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci USA*. 2007; 104(5):1488-93.
19. Klein R, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005; 308:309-452.
20. Mani A, et al. LRP6 mutation in a family with early coronary disease and metabolic risk factors. *Science*. 2007; 315(5816):1278-82.
21. Crichlow GV, et al. Dimerization of FIR upon FUSE DNA binding suggests a mechanism of c-myc inhibition. *The EMBO Journal*. 2008;27:277-89.
22. Greenbaum D, et al. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*. 2003; 4:117.1-117.8
23. Anderson NL, Anderson NG. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol Cell Proteomics*. 2002;1:845-67.
24. Clamp M, et al. Distinguishing protein-coding and non-coding genes in the human genome. *Proc Natl Acad Sci*. 2007;104(49):19428-19433.