

A photograph of the Golden Gate Bridge in San Francisco, California. The bridge's iconic orange-red towers and suspension cables are prominent against a clear blue sky. The bridge spans across the water, with hills and a few sailboats visible in the distance. The text is overlaid on the upper left portion of the image.

Progress in Proteomics: What's Working and What's Not

R. A. Bradshaw, R. J. Chalkley, J. Biarc, and A. L. Burlingame
Mass Spectrometry Facility
& Department of Pharmaceutical Chemistry
UCSF

“This is a standard pattern in the history of science: when tools for measuring increase their precision by orders of magnitude, new paradigms often emerge, because the newfound accuracy reveals anomalies that had gone undetected.”

“This is a standard pattern in the history of science: when tools for measuring increase their precision by orders of magnitude, **new paradigms** often emerge, because the newfound accuracy reveals anomalies that had gone undetected.”

“This is a standard pattern in the history of science: when tools for measuring increase their precision by orders of magnitude, **new paradigms** often emerge, because the newfound accuracy reveals anomalies that had gone undetected.”

“The Invention of Air”, Steven Johnson (2008)
in reference to Joseph Priestley’s
experiments on oxygen

Innovation (or “revolution”) in science

- Scientific progress does NOT occur in a linear fashion as a series of facts, determined one after another (“development-by-accumulation” model).
- Science instead works within an established paradigm, which is governed by certain rules that give definition to terms, the collection of data and the boundaries of inquiry.
- New paradigms form when anomalies (data that doesn’t fit and questions that can’t be answered with the existing tools) cause some scientists to begin reaching outside the old boundaries and establish new rules and conventions.
- Ultimately a new paradigm is created as the old one collapses, i.e. a “paradigm shift”.

From “The Structure of Scientific Revolutions”, Thomas Kuhn, 1962

Landmark Proteomic Developments – “paradigm shifts”

- 2D gel analyses
- Arrays (nucleic acid/protein)
- ESI and MALDI MS
- Human genome sequence

What Has Proteomics Accomplished?

- Protein identification/expression
- More extensive mapping of PTMs
- Protein networks (protein/protein interactions)
- Protein machines, e.g. nuclear pore complex
- Technical advances in sensitivity and quantitation
- Clinical proteomics

What Hasn't Proteomics Accomplished?

- The functional identification of a significant percentage of the human proteome
- A useful integration (understanding) of proteomic data with that of transcriptomics and metabolomics
- The reproducible identification and quantification of biologically significant post-translational modifications
- The biomarker debacle

What Hasn't Proteomics Accomplished?

- The functional identification of a significant percentage of the human proteome
- A useful integration (understanding) of proteomic data with that of transcriptomics and metabolomics
- The reproducible identification and quantification of biologically significant post-translational modifications
- The biomarker debacle

Characteristics of post-translational modifications

- Both signal dependent and independent
- Cannot be accurately predicted - though consensus sequences have been reported for some of them
- Can be tissue- or location-specific
- Stable (irreversible) or dynamic (transient)
- Variable stoichiometry
- Alter biological activity *and* physical properties
- May alter the immune response

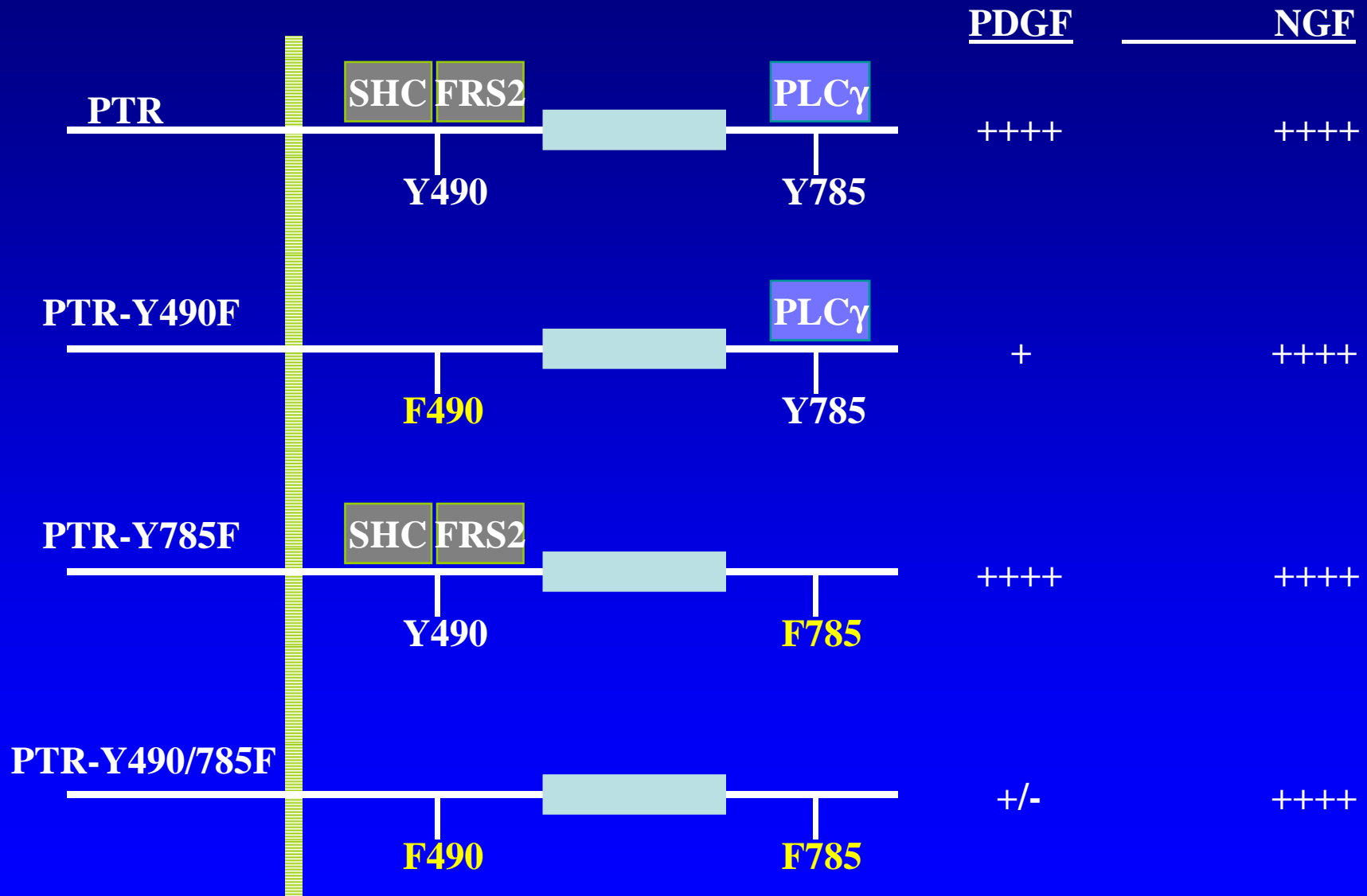
Phosphorylation Facts

- ~1/3 of all proteins contain ≥ 1 phosphate
- pSer:pThr:pTyr is 1800:200:1
- Phosphorylation of Tyr more tightly regulated
- Heterogenous phosphorylation of individual proteins (>1 site/protein)
- ~2% of human genome encode protein kinases and phosphatases (~500 and ~100)

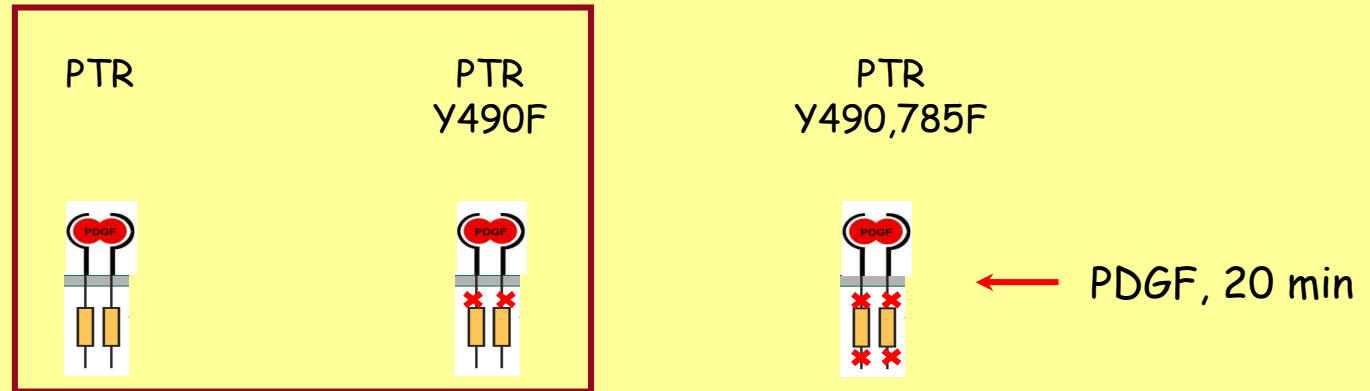
Cell Phosphorylation Kinetics

- Ligand-induced tyrosine phosphorylation is very rapid and short-lived (largely back to baseline by 5')
- Ligand-induced serine/threonine phosphorylation is slower and more prolonged

PDGF/TrkA Receptor (PTR) Constructs Neurite Response



Workflow



Purification of proteins with trizol/BCA quantification

Reduction, alkylation, & digestion by trypsin

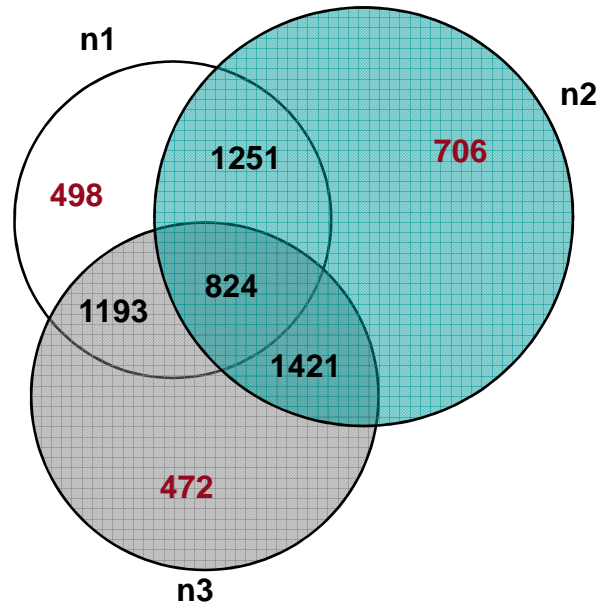
Phosphopeptide enrichment by TiO_2 column

SCX fractionation on polyLC

LC-MS/MS analysis on Q-star

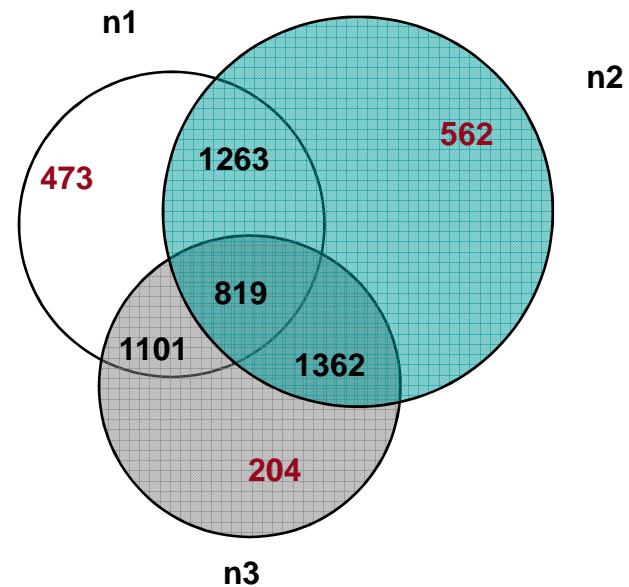
Overlap phosphopeptides n1/n2/n3

Unique phosphopeptides PTR



Total peptides	824 peptides
n1: 2942	n1: 28%
n2: 3378	n2: 24%
n3: 3086	n3: 27%

Unique phosphopeptides PTR Y490F



Total peptides	819 peptides
n1: 2837	n1: 29%
n2: 3187	n2: 26%
n3: 2667	n3: 31%

Red: total unique phosphopeptides identified in each experiment

black: unique phosphopeptides in common in 2 or 3 experiments

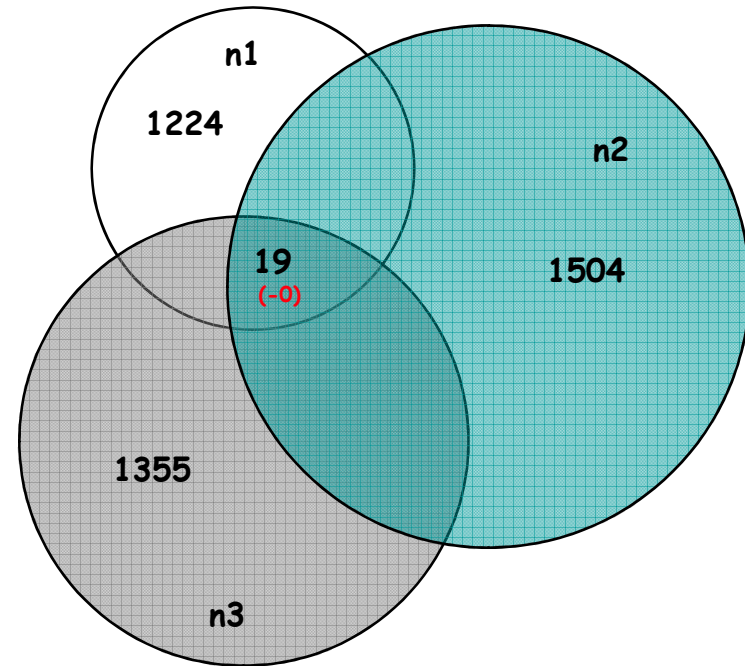
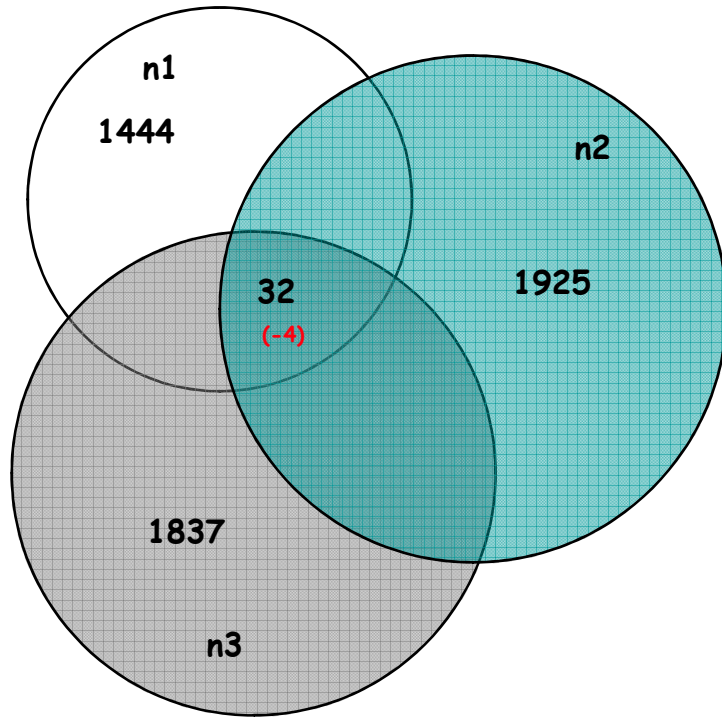
Overlap specific peptides n1/n2/n3

Specific Peptides PTR

Specific Peptides PTR Y490F

black: peptides specific unique

(-red): non phosphopeptides specific unique



Total peptides
n1: 1476 (246)
n2: 1957 (390)
n3: 1869 (222)

32 peptides
n1: 2.2%
n2: 1.6%
n3: 1.7%

Total peptides
n1: 1243 (118)
n2: 1523 (197)
n3: 1374 (146)

19 peptides
n1: 1.5%
n2: 1.2%
n3: 1.4%

Issues in Phosphoproteomic Studies

- Establishing which modifications are relevant – particularly ones of partial stoichiometry
- Determining which kinase actually carries out which phosphorylation
- Crosstalk with other PTMs

Issues in Phosphoproteomic Studies

- Establishing which modifications are relevant – particularly ones of partial stoichiometry

Solution: differential signaling systems

- Determining which kinase actually carries out which phosphorylation
- Crosstalk with other PTMs

Issues in Phosphoproteomic Studies

- Establishing which modifications are relevant – particularly ones of partial stoichiometry

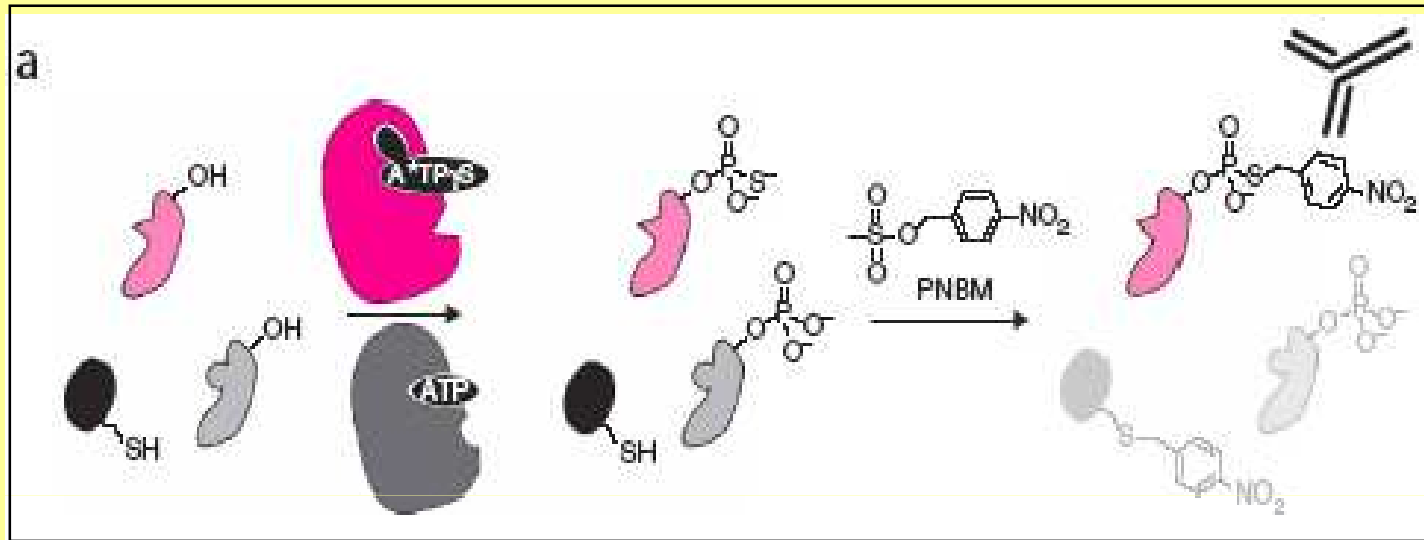
Solution: differential signaling systems

- Determining which kinase actually carries out which phosphorylation

Solution: Shokat modifications

- Crosstalk with other PTMs

General principle: Shokat lab



Allen et al., Nature Methods, 2007

1. Single mutation in the kinase domain
use of a modified ATP γ S analog (as kinase)
2. Alkylation of thiophosphorylated Y-PNBM
3. Immunoprecipitation of the thiophosphate alkylated
4. LC-MS/MS analysis

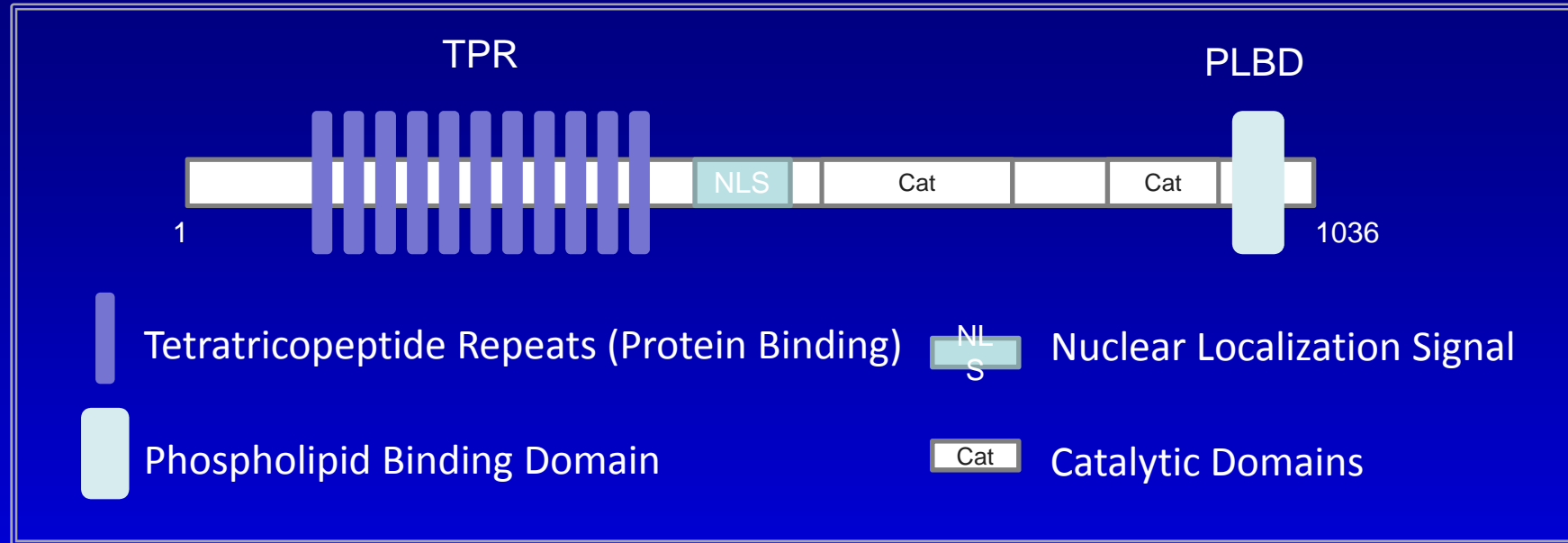
Issues in Phosphoproteomic Studies

- Establishing which modifications are relevant – particularly ones of partial stoichiometry
Solution: differential signaling systems
- Determining which kinase actually carries out which phosphorylation
Solution: Shokat modifications
- Crosstalk with other PTMs
Solution: Expand analyses to other PTMs

O-GlcNAcylation

- Addition of a single sugar residue: N-Acetylglucosamine (GlcNAc) to serine or threonine residues of nuclear and cytoplasmic proteins.
- Different from 'conventional' glycosylation:
 - Inside the cell
 - Transient modification
 - Enzymes responsible for addition and removal of modification
 - O-GlcNAc transferase and O-GlcNAcase
- Modification is involved in cellular response to nutritional and other stresses
 - UDP-GlcNAc as modification donor
 - UDP-GlcNAc levels high in 'well fed' state => Increased O-GlcNAcylation
 - Also involved in response to other stresses.
- Clear links to Diabetes and Alzheimer Disease.

O-GlcNAc Transferase (OGT)



Single Transferase, but many targets.

Localization: transported to cellular locations by interactions with other molecules:

NLS: Mostly nuclear; target transcription factors.

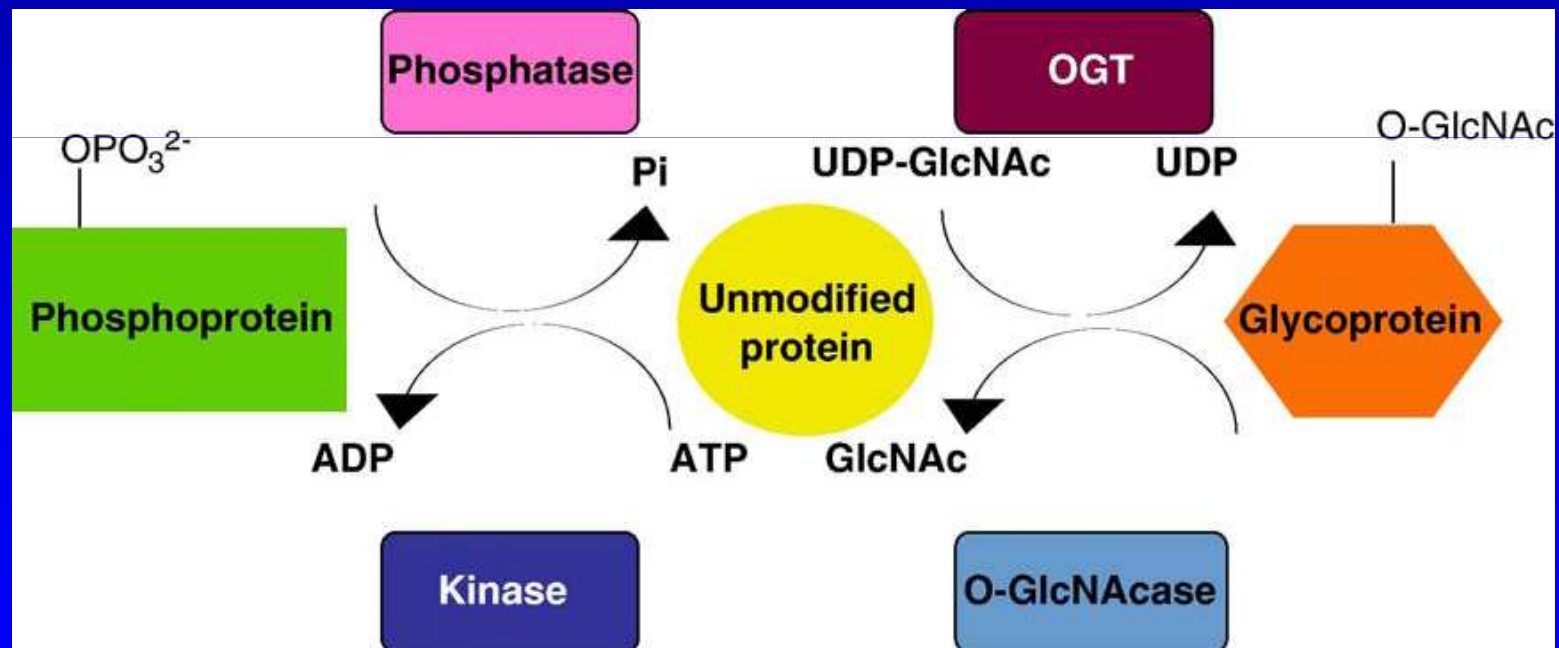
TPR: Protein binding domains.

PLBD: Phospholipid binding domain. This domain has been shown to allow transporting of enzyme to plasma membrane, where it can target signal transduction pathways; e.g. upon insulin stimulation.¹

¹Yang X. et al. *Nature* (2008) **451** 964-969

O-GlcNAcylation and Phosphorylation

- O-GlcNAc modified proteins are also potential phosphoproteins
 - Many examples where the same or neighboring residues can be either GlcNAcylated or phosphorylated.
 - Multiple experiments have shown the two modifications interact/affect each other

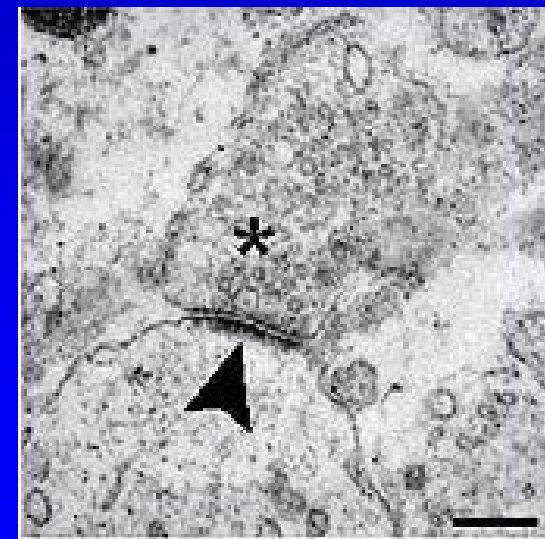
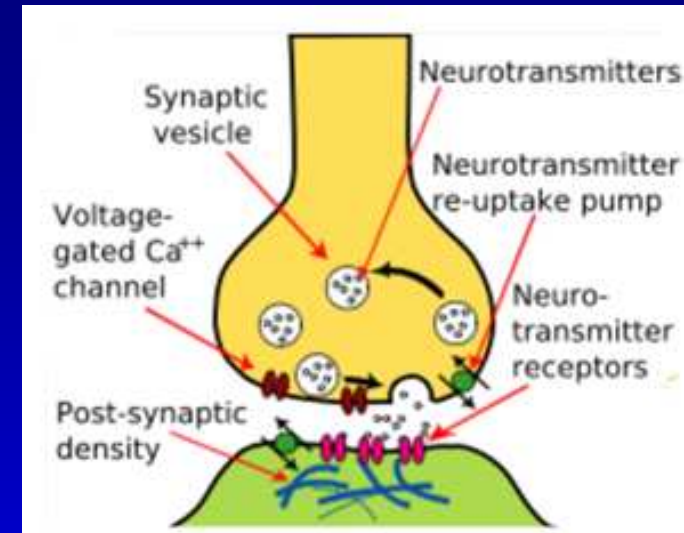


O-GlcNAcylation in the Post-Synaptic Density (PSD)

- There is a dense pseudo-organelle structure associated with the post-synaptic membrane of neurons known as the 'post-synaptic density' (PSD).
- The brain is known to contain large amounts of O-GlcNAcylation.
- The PSD can be purified relatively easily by density centrifugation.
- We, and others, have identified O-GlcNAc modification in PSD.

This Experiment:

- 2mg PSD purified
- Digested with Trypsin
- Modified peptides enriched using LWAC
- Peptides analyzed using LTQ-Orbitrap with CID and ETD.



Arrowhead points to PSD
Asterisk indicates pre-synaptic vesicles

Phosphorylation in the PSD

- The PSD is also a heavily researched compartment for phosphorylation analysis
- Studies within our lab have identified over 1500 phosphorylation sites in the PSD¹
- Would be interesting to relate phosphorylation and O-GlcNAcylation sites
- Is there evidence of cross-talk between the modifications?

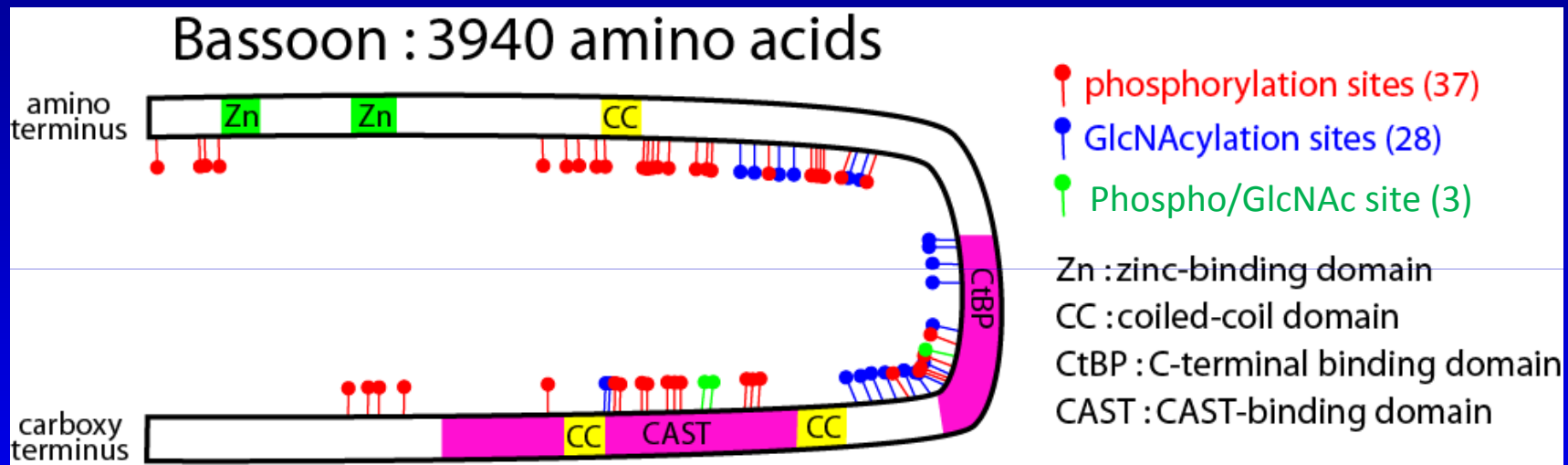
¹Trinidad, J.C. et al. Mol Cell Proteomics 7: 684-696 (2008)

58 O-GlcNAc Modification Sites Identified in Mouse PSD

Accession Number	Protein	O-GlcNAc Modification Site/s
O88737	Protein Bassoon	1354, 1395, 1418, 1445, 1517, 1537, 1657, 1666, 1707, 1772, 1962, 2027, 2029, 2058, 2067, 2068, 2070, 2091, 2141, 2188, 2295, 2317, 2318, 2694, 2700, 2703, 2941, 2945
Q9QYX6	Protein Piccolo	2634, 2639, 2656, 2930, 2948, 3873
Q69ZX8	<u>Actin-binding LIM protein 3</u>	383, 419, 423, 534, 546, 547
Q8K4G5	<u>Actin-binding LIM protein 1</u>	496, 499
P08553	<u>Neurofilament medium polypeptide</u>	37, 46, 430
P08551	<u>Neurofilament light peptide</u>	48, 414
O70511	<u>270 kDa Ankyrin G isoform</u>	1520
P97836	Disks large-associated protein 1	525, 526
Q8BMB0	Protein EMSY	499
O35927	<u>Catenin delta-2</u>	447, 453
Q9WV69	<u>Dematin</u>	285
Q60974	<u>Nuclear receptor corepressor 1</u>	1496
Q8CHP6	<u>Polyhomeotic-like protein 3</u>	238
Q3UHF7	Human immunodeficiency virus type-1 Enhancer-binding protein 2	1271
Q9EQZ7	Regulating synaptic membrane <u>exocytosis protein 2</u>	1528

Bassoon is Heavily O-GlcNAcylated and Phosphorylated

- Bassoon is a major component of the cytomatrix in the presynaptic active zone.
- Involved in spatial and temporal control of neurotransmitter release.



- Why is this protein so heavily post-translationally modified?
- What are the modifications doing?
 - Regulating protein-protein interactions?

O-GlcNAc Consensus Sequence?

- It has previously been noted that sites of O-GlcNAcylation often contain a proline two or three residues prior to the site of modification
- No systematic analysis of modification sites has been performed before, probably due to the relative paucity of known modification sites.

Summary of the Sequence around Modification Sites Found in this Study



What Hasn't Proteomics Accomplished?

- The functional identification of a significant percentage of the human proteome
- A useful integration (understanding) of proteomic data with that of transcriptomics and metabolomics
- The reproducible identification and quantification of biologically significant post-translational modifications
- The biomarker debacle

Sources of Biomarkers

- Fluids
 - plasma, urine, synovial fluid, CSF etc
- Tissue
 - healthy vs diseased

Sources of Biomarkers

- Fluids
 - plasma, urine, synovial fluid, CSF etc
- Tissue
 - healthy vs diseased

Protein Biomarkers in Plasma Have Shown Substantial Clinical Value

Since when?	Decades
How many tests?	> 10 million/yr
What instruments?	Ca. 100,000 machines in hospitals, labs
How accurate?	CV~5-10% worldwide at $\geq 100\text{pg/ml}$

Cardiac damage	Tnl, CK-MB, Mb, MPO, BNP
Cancer	PSA, CA-125, Her-2
Inflammation	CRP, SAA, cytokines, RF
Liver Damage	ALT, ALP, AST, GGT (enzyme assays)
Coagulation	AT-III, proteins C&S, fibrinogen, VWF
Allergy	IgE against various antigens
Infectious disease	HIV-1, Hepatitis BsAg

The Proteomics Approach

- Extensive programs in both industry and academia (including government) as well as some alliances between them
- Financial analysts estimate ~\$10⁹ spent or invested in companies with business models based on proteomic-driven strategies for the identification of drug targets and/or biomarkers

The Proteomics Biomarker Track Record

- Few, if any, new biomarkers resulting from proteomic identification, validation and clinical approval are being found
- Actual number of new protein targets is declining (proteomics is NOT changing this)
- In fact, only a handful of proteins are effective as markers; instead there are whole families of commercial products targeting the same small group

Why is the record so bad?

- Technology not yet able to overcome problems of dynamic range and sample complexity?

Why is the record so bad?

- Technology not yet able to overcome problems of dynamic range and sample complexity?
- Too much effort spent on identification and not enough on validation (easier to hype the former than the latter)?

Why is the record so bad?

- Technology not yet able to overcome problems of dynamic range and sample complexity?
- Too much effort spent on identification and not enough on validation (easier to hype the former than the latter)?
- Bad luck?

Why is the record so bad?

- Technology not yet able to overcome problems of dynamic range and sample complexity?
- Too much effort spent on identification and not enough on validation (easier to hype the former than the latter)?
- Bad luck?
- There are no significant numbers of new markers to find (with any technology)?

“Thinking outside the box...”

Human Proteome Detection and Quantitation Project: hPDQ

N. L. Anderson, N. G. Anderson, T. W.
Pearson, C. H. Borchers, A. G. Paulovich,
S. D. Patterson, M. Gillette, R. Aebersold
and S. A. Carr

Mol. Cell. Proteomics 8: 883 - 886 (2009)



Molecular & Cellular PROTEOMICS

[Guidelines for Preparing Manuscripts Describing Research in
Clinical Proteomics - Published](#)

Open Access

[Submit Manuscripts Online](#)

[Subscriptions](#)

[Librarian's Resource](#)

[Resources for Authors](#)

[Editorial Policies](#)

[Advertising](#)

[Copyright Permission](#)

[Editorial Board](#)

[Code of Ethics](#)

[Related Resources](#)

[Email Alerts](#)

[Net Tips](#) [Help](#)

[Feedback](#)



**The American Society for
Biochemistry and Molecular**

Biology publishes Molecular
& Cellular Proteomics.

RSS

[Subscribe to MCP feed](#)



Stanford University Libraries

 **MCP Papers
in Press** FREE!

▶ [Electronic Letters to the
Editor](#)

▶ [Mission Statement](#)

▶ [Search the MCP](#)

▶ [Search all HighWire
Journals](#)

▶ [Select an Issue from the
Archive](#)

▶ January 2002 - March
2009

▶ [Article Statistics](#)
[Most Read](#) / [Most Cited](#)



Current Issue:
March 2009



RECENT EDITORIALS:

[New Important
Change for
NIH-funded
Authors](#)

[New Guidelines
for Clinical
Proteomics
Manuscripts](#)

[NEW MCP
Blog!](#)

[MCP Calendar](#)

Waters

THE SCIENCE OF WHAT'S POSSIBLE.™

Advertisement

CLINICAL PROTEOMICS RESEARCH

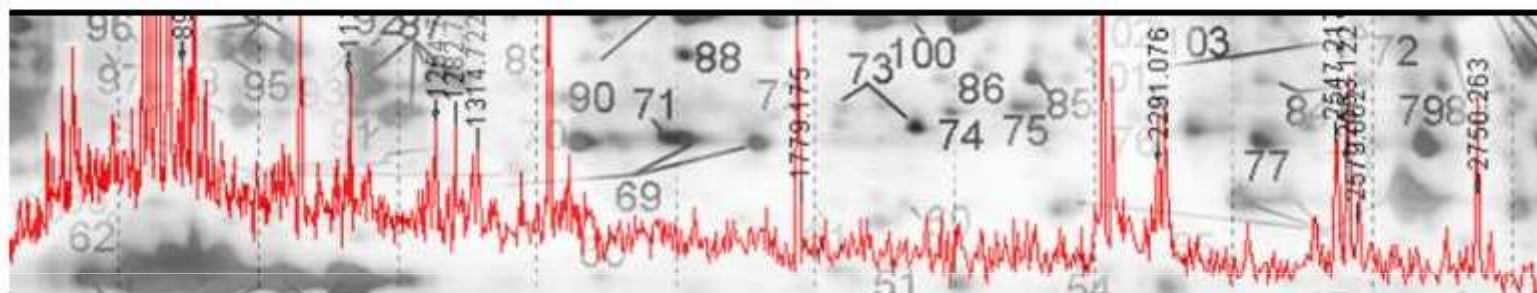
Molecular & Cellular Proteomics



Molecular & Cellular Proteomics Blog

[FRONT PAGE](#)

[ABOUT MCP BLOG](#)



A human proteome detection and quantitation project: hPDQ

January 15, 2009 • 2 Comments

N. Leigh Anderson, Norman G. Anderson, Terry W. Pearson, Christoph H. Borchers, Amanda G. Paulovich, Scott D. Patterson, Michael Gillette, Ruedi Aebersold, and Steven A. Carr

The lack of sensitive, specific, multiplexable assays for most human proteins is the major barrier impeding development of candidate biomarkers into clinically useful tests. Recent progress in mass spectrometry-based assays for proteotypic peptides, particularly those with specific affinity peptide enrichment, offers a systematic and economical path to comprehensive quantitative coverage of the

To search, type and hit enter

RECENT POSTS

- [A human proteome detection and quantitation project: hPDQ](#)
- [Welcome to the MCP Blog](#)

TOP POSTS

- [A human proteome detection and quantitation project: hPDQ](#)

The hPDQ Project

The hPDQ would construct a suite of assays using two peptides, enriched by individual specific antisera, from the protein product of each of the ~20,500 human genes, followed by identification by mass spectrometry. This would lay a quantitative foundation for subsequent efforts to define the larger universe of splice variants, post-translational modifications, protein-protein interactions and tissue localization (which is not, however, a part of the present proposal).

hPDQ Required Resources:

- A comprehensive database of proteotypic (protein-unique) optimized peptides for each of the 21,500 human proteins
- At least two synthetic, isotope labeled, proteotypic peptides for each protein for use as internal quantitative standards
- High affinity anti-peptide antibodies, preferably monoclonal, specific for the same two peptides.
- Robust and affordable instrument platforms (triple-quadrupole mass spectrometers coupled with nanoflow (~300-600 nl/min) LC systems) for quantitative analysis of small (amol to fmol) amounts of tryptic peptides and for sample preparation. MALDI platforms may provide similar capabilities.

hPDQ costs

- Initial phase: assays for 2000 proteins of 'interest' in 2 years - $\$50 \times 10^6$
(funding through existing academic and commercial resources appropriately networked)
- Phase 2: assays for 18,500 proteins in 5 years - $\$250 \times 10^6$
(will require large scale government and philanthropic support)

Concluding thoughts...

- Is the PTM problem (physiologically significant vs. background) solvable with the present technologies?
- Has too much translational emphasis (vested interests?) subverted proteomics?
- How will proteomic standards be maintained?
- Where is the next 'paradigm shift' coming from?

Acknowledgements

A photograph of the Golden Gate Bridge in San Francisco, California. The bridge's iconic red-orange towers and suspension cables are prominent against a clear blue sky with light, wispy clouds. The bridge spans across the water, with a small red buoy visible in the foreground. In the distance, the city of San Francisco and the surrounding hills are visible under a bright sky.

A. L. Burlingame

Robert J. Chalkley
Jordane Biarc

Ralf Schoepfer
Agnes Thalhammer

Peter Baker
Aenoch Lynn

<http://prospector.ucsf.edu>

NIH NCRR P41 RR001614