

A Strategy for the Rapid Identification of Phosphorylation Sites in the Phosphoproteome*

Justin A. MacDonald^{‡§}, Aaron J. Mackey^{§¶}, William R. Pearson^{||},
and Timothy A. J. Haystead^{‡**}

Edman phosphate (³²P) release sequencing provides a high sensitivity means of identifying phosphorylation sites in proteins that complements mass spectrometry techniques. We have developed a bioinformatic assessment tool, the cleavage of radiolabeled protein (CRP) program, which enables experimental identification of phosphorylation sites via ³²P labeling and Edman degradation of cleaved proteins obtained at femtomole levels. By observing the Edman cycle(s) in which radioactivity is found, candidate phosphorylation sites are identified by determining which residues occur at the observed number of cycles downstream from a peptide cleavage site. In cases where more than one residue could be responsible for the observed radioactivity, additional experiments with cleavage reagents having alternative specificities may resolve the ambiguity. Given a protein sequence and a cleavage site, CRP performs these experiments *in silico*, identifying resolved sites based on user-supplied experimental data, as well as suggesting combinations of reagents for additional analyses. Analysis of the PhosphoBase protein sequence database suggests that CRP data from two cleavage experiments can be used to identify unambiguously 60% of known phosphorylation sites. Data from additional cleavage experiments may increase the overall coverage to 70% of known sites. By comparing theoretical data obtained from the CRP program with ³²P release data obtained from an Edman sequencer, a known phosphorylation site was identified unambiguously and correctly. In addition, our results show that *in vivo* phosphorylation sites can be determined routinely by differential proteolysis analysis and Edman cycling with less than 1 fmol of protein and 1000 cpm. *Molecular & Cellular Proteomics* 1:314–322, 2002.

Proteomic technologies have transformed the manner in which proteins and their contributions to cellular function are viewed (for review see Refs. 1–3). The completed human

genome map represents an invaluable tool that presents a comprehensive account of the structure and sequence of human genes. However, it offers limited insight regarding the various post-translational modifications, such as phosphorylation, that occur on proteins in the cell. Contemporary proteomic analysis uses two-dimensional gel electrophoresis to separate cellular proteins and mass spectrometry to identify the proteins in these gels. The identification of a protein can usually be achieved at the femtomole level with the employment of tandem mass spectrometry to decode the primary amino acid sequences. Numerous advances in proteomics tools, including advances in high sensitivity protein staining and two-dimensional electrophoresis techniques, refinements in ampholytic technology, and the advent of accurate, sensitive, and affordable mass spectrometers including matrix-assisted laser desorption ionization mass spectrometers and quadrupole tandem mass spectrometers has allowed a shift to high throughput analysis of large numbers of candidate proteins. Recent functional proteomic analyses are now being employed specifically to describe the architecture of signal transduction pathways at the level of the individual kinase or phosphatase (4, 5); however, significant barriers still limit the ability to identify individual phosphoproteins and their sites of phosphorylation within a proteome analysis.

Phosphoproteins are often a small fraction of the individual protein concentration and present at low copy number in cells. Prediction of the phosphorylation status of proteins from sequence patterns or more sophisticated neural network motifs has limited sensitivity and greatly lacks specificity (6). Protein phosphorylation must therefore be observed directly. We have developed a bioinformatic assessment tool, CRP,¹ that enables access via ³²P labeling and Edman sequencing to concentrations of phosphorylation sites that are below the femtomole level.

EXPERIMENTAL PROCEDURES

Materials—Mouse submaxillary gland endoproteinase Arg-C and *Staphylococcus aureus* V8 protease (endoproteinase Glu-C) were from Sigma. *Achromobacter lyticus* endoproteinase Lys-C was from

From the Departments of ¶Microbiology and ||Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia 22908 and ‡Department of Pharmacology and Cancer Biology, Duke University, Durham, North Carolina 27710

Received, January 19, 2002, and in revised form, March 22, 2002
Published, MCP Papers in Press, March 25, 2002, DOI 10.1074/mcp.M200002-MCP200

¹ The abbreviations used are: CRP, cleavage of radiolabeled protein; PVM, modified polyvinylidene difluoride membrane; PRP, platelet rich plasma; HPLC, high pressure liquid chromatography.

Wako Bioproducts Inc. (Richmond, VA). The cAMP-dependent protein kinase catalytic subunit and mouse monoclonal anti-Hsp27 antibody were from Calbiochem. C18 Zip-Tips were purchased from Millipore (Bedford, MA). Cyanogen bromide and skatol were obtained from Pierce.

Preparation of ^{32}P -Phosphorylated Recombinant Proteins—Recombinant proteins were prepared and phosphorylated following protocols reported previously. Phosphorylation of proteins was performed at 25 °C for 3 h in 0.5-ml reactions containing 5 mM MgCl_2 , 0.3 mM $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (250 cpm/nmol). Telokin was phosphorylated with the cAMP-dependent protein kinase catalytic subunit as described previously (7).

Identification of Phosphoproteins in Human Platelets by Mixed Peptide Sequencing—Platelet-rich plasma (PRP) was prepared from whole human blood anticoagulated with ACD (sodium citrate, citric acid, dextrose) by differential centrifugation in 50-ml tubes (200g, 20 min, 25 °C). The top layer containing platelet-rich plasma was separated from the red cell layer and used as a source of platelets in all studies. Washed platelets were prepared from PRP by the method of Mustard *et al.* (8). ACD (0.05 volumes), apyrase (7.5 units/ml ADPase activity), and indomethacin (1 $\mu\text{g}/\text{ml}$) were added to PRP. Platelets were then sedimented from PRP (620 g, 20 min, 25 °C) and resuspended in Buffer I (10 mM HEPES, pH 7.4, 0.34 mM Na_2HPO_4 , 140 mM NaCl, 2.9 mM KCl, 5 mM glucose, 2 mM MgCl_2 , 12 mM NaHCO_3). The platelets were treated with 18.5 MBq of ^{32}P orthophosphate for 90 min and harvested by centrifugation ($800 \times g$ for 10 min). Platelets were resuspended in Buffer I and stored at room temperature with gentle rocking until use. Platelet aggregation was stimulated by the addition of 0.05 units/ml of thrombin in the presence of 10 μM calyculin A. Platelets were lysed with the addition of a $5\times$ lysis buffer (200 mM Tris, pH 8.0, 750 mM NaCl, 5% (v/v) Nonidet P-40).

Protein maps were prepared by two-dimensional electrophoresis as described in Ref. 4. The proteins were electroblotted to PVM at 30 V overnight. The transferred proteins were stained with Amido Black. After destaining, the membrane was applied to x-ray film for autoradiography. The autoradiograph from each gel was compared using Melanie software (Bio-Rad). Spots of interest were aligned to the membrane, and the corresponding stained bands were excised. The excised pieces were treated with 200 μl of cyanogen bromide solution (500 mg/ml cyanogen bromide in 70% formic acid) for 90 min. The treated piece of PVM was placed in an Applied Biosystem 494 protein sequencer, and 8–18 cycles of pulsed liquid chemistry were carried out. The mixed peptide sequences generated were sorted and matched against the yeast protein databases by the FASTF algorithm (9).

Sample Preparation and Edman Sequencing—Phosphorylated protein samples from the *in vitro* phosphorylation of telokin were incubated overnight at 37 °C with the endopeptidase of choice. Cleavage with cyanogen bromide (0.1 mg/ml) was completed overnight at 5 °C in 70% (v/v) formic acid. Cleavage with skatol (0.1 mg/ml) was completed at 70 °C in 70% acetic acid for 3 h in the dark. The digests were acidified by the addition of trifluoroacetic acid and spun through a Zip-Tip equilibrated previously in 0.1% trifluoroacetic acid. The tip was washed three times with 0.1% trifluoroacetic acid (5 μl) before peptides were eluted with 70% acetonitrile/0.1% trifluoroacetic acid (5 μl).

Peptides were immobilized to a $2.5 \times 2.5\text{-mm}$ piece of Immobilon membrane (Millipore) following the manufacturer's instructions. The membrane was washed sequentially with 1 volume of 100% trifluoroacetic acid and 50 ml of 0.1% trifluoroacetic acid in distilled H_2O . The membrane piece was placed into a 494 Procise sequencing cartridge. Vapor phase amino acid sequencing was performed using an instrument that allowed for the collection of the Edman sequencing reactions, in our case an Applied Biosystems 494 cLC protein sequencer.

Phosphorylated residues were located by determining the cycles in which ^{32}P was released when samples were subjected to sequential Edman degradation under conditions that optimized recovery of ^{32}P (10).

Phosphorylated Hsp27 was immunoprecipitated from the *in vivo* ^{32}P -labeled platelet lysate. The lysate was pre-cleared with protein G-agarose beads (1 h at 5 °C). Platelet lysate was incubated overnight with 10 μg of monoclonal mouse anti-Hsp27 followed by harvest with protein G-agarose. Immunoprecipitated proteins were eluted from protein G with 50 mM glycine, pH 2.0. The pH of the eluate was quickly neutralized with 1 M phosphate buffer, pH 8. The protein samples were incubated overnight at 37 °C with the protease of choice; the resulting peptides were processed for Edman sequencing as described above.

CRP Program World Wide Web Interface—The acquired ^{32}P release data were interpreted using the CRP program, accessible at fasta.bioch.virginia.edu/crp/. The program allows for the input of protein sequence data in the same fashion as a normal BLAST query, either a raw sequence or a FASTA-formatted sequence or via a unique sequence identifier (see www.ncbi.nlm.nih.gov/blast/html/search.html for details). As an example, the sequence for myelin basic protein, a commonly used protein kinase substrate protein, was processed by the CRP program (Fig. 1A). CRP-generated theoretical cleavage data were obtained by selecting the specific carboxyl-terminal amino acid at which to cut, in this case, at Arg residues. CRP displays results of the *in silico* cleavage as a table of the Edman cycles in which radioactivity might be observed, listing each associated potential site (Fig. 1B). The percent coverage, or the cumulative number of observable potential phosphorylation sites, is also provided. With myelin basic protein and proteolysis with endoproteinase Arg-C, 21 cycles are required to ensure 100% coverage of 35 possible phosphorylation sites. The program highlights sites that agree with known phosphorylation target consensus sequences and provides a link to the EXPASY PROSITE database (11), where information pertaining to the consensus site is available.

If radioactivity was observed in a cycle containing multiple potential phosphorylation sites, the determination becomes ambiguous; by selecting all cycles (✓) in which radioactivity was actually observed (including any unambiguous, as well as the ambiguous cycles), CRP provides a second table (Fig. 1C), detailing the new cycle positions of the potential phosphorylation residues when the original protein is cut at different cleavage sites. These data permit the selection of proteases for second, and if necessary, third cleavages that could yield unambiguous data.

RESULTS

Theoretical Potential for the CRP-based Analysis of Phosphoproteins—We analyzed the ability of the CRP technique to determine the phosphorylation site(s) of proteins *in silico*, using either the SwissProt (11) or PhosphoBase (12) protein sequence databases. We selected an arbitrary Edman cycle cutoff limit of 25 cycles; any residues appearing after 25 cycles were considered unresolved by the experiment. Residues that appeared within 25 cycles but were found in a cycle containing other residues were also considered unresolved. Therefore, within a single cleavage experiment, any residue found alone within a cycle below the cutoff was considered resolved; *i.e.* if radioactivity were observed in that cycle, it would identify unambiguously that residue as phosphorylated.

The *left panel* of Fig. 2 demonstrates the percent coverage

A

Cleaved Radioactivity of Phosphopeptides

Enter an accession number or other unique identifier:

Enter your protein sequence in fasta format:

masqkrpsqrhgskylatastmdharhgflprhrdtgildsigrffggdrgapkrgsgkvp
wlkpgrsplspharsqpglcnmykdshhpatahygslpqkshgrtgdenpvvhffkni
prtpppsggkgrglslsrfswgaegqrpqgyggrasdyksahkgfkgvdaqgtlskifl
ggrdsrsgpmarr

Cleave at which residue?

R - Arg

Submit Query

Clear

Reset

B

Histogram of expected 'hot spots'

Cycle #	Potential Phosphorylation Sites						# in Cycle	Coverage
☒ 1	S068	S076	T093	T107	T125	S190	6	17%
☒ 2	S008	T036	S057	S142	S159	S188	6	34%
☐ 3	S003	S013	S137	S192			4	45%
☐ 4	Y096	Y161					2	51%
☐ 5	Y015	S072	S129	S139	Y154		5	65%
☐ 6	S098	S163					2	71%
☐ 7	S041						1	74%
☐ 8	T018						1	77%
☐ 9	Y084						1	80%
☐ 10	S020						1	82%
☐ 11	T021	S103					2	88%
☐ 12	S087						1	91%
☐ 13								
☐ 14								
☐ 15								
☐ 16	T122						1	94%
☐ 17								
☐ 18								
☐ 19	T176						1	97%
☐ 20								
☐ 21	S178						1	100%

Process selected cycles -->

Submit Query

1	11	21	31	41
MASQKRPSQR	HGSKYLATAS	TMDHARHGFL	PRHRDTGILD	SIGRFFGGDR
51	61	71	81	91
GAPKRGSGKV	PWLKPGRSPL	PSHARSQPGL	CNMYKDSHHP	ARTAHYGSPL
101	111	121	131	141
QKSHGRGTQDE	NPVVHFFKNI	VTPRTPPPSQ	GKGRGLSLSR	FSWGAEGQRP
151	161	171	181	191
GFGYGGRASD	YKSAHKGFKG	VDAQGTLSKI	FKLGGRDSRS	GSPMARR

Fig. 1

C

Potential Phosphorylation Residues' Cycle Position When Cut at Specific Cleavage Sites

Cleavage Site	S008	T036	S057	S068	S076	T093	T107	T125	S142	S159	S188	S190
A	6	11	5	16	2	2	13	31	48	1	15	17
C	8	36	57	68	76	12	26	44	61	78	107	109
D	8	1	8	19	27	7	21	16	33	50	1	3
E	8	36	57	68	76	93	107	15	32	13	42	44
F	8	7	11	22	30	47	61	8	1	7	7	9
G	8	8	1	2	10	14	2	20	7	3	3	5
H	8	3	24	35	3	4	3	10	27	44	23	25
I	8	36	15	26	34	51	65	5	22	39	8	10
K	3	22	3	4	12	8	5	7	10	27	6	8
L	8	6	18	5	6	13	8	26	4	21	5	7
M	7	14	35	46	54	10	24	42	59	76	105	107
N	8	36	57	68	76	11	25	6	23	40	69	71
P	1	5	4	3	5	3	7	2	14	9	38	40
Q	4	27	48	59	67	16	6	17	12	11	14	16
R	2	2	2	1	1	1	1	1	2	2	2	1
S	5	16	16	11	4	6	4	22	3	17	10	2
T	8	15	21	32	40	57	14	3	17	34	12	14
V	8	36	57	8	16	33	47	4	21	38	17	19
W	8	36	57	6	14	31	45	63	80	16	45	47
Y	8	21	42	53	61	9	11	29	46	5	27	29

Cleavage Residue: F

Potential P-Sites :	S142	T36 S159 S188	S8	T125	S190	S57	S68	S76	T93	T107
Cycle#	1	7	8	9	11	22	30	47	61	

Cleavage Residue: R

Potential P-Sites :	S68 S76 T93 T107 T125 S190	S8 T36 S57 S142 S159 S188
Cycle#	1	2

Cleavage Residue: K

Potential P-Sites :	S8 S57	S68	T107	S188	T125	T93 S190	S142	S76	T36	S159
Cycle#	3	4	5	6	7	8	10	12	22	27

Cleavage Residue: M

Potential P-Sites :	S8	T93	T36	T107	S57	T125	S68	S76	S142	S159	S188	S190
Cycle#	7	10	14	24	35	42	46	54	59	76	105	107

Cleavage Residue: W

Potential P-Sites :	S68	S8	S76	S159	T93	T36	T107 S188	S190	S57	T125	S142
Cycle#	6	8	14	16	31	36	45	47	57	63	80

FIG. 1. **The CRP program interface.** A, initial data entry screen allowing accession number or protein sequence inputs. The sequence of myelin basic protein is shown in the protein sequence input box, and cleavage with endoproteinase Arg-C has been selected. By submitting the query, a histogram of expected hot spots is presented (B). The theoretical data displayed include the following: *Cycle #*, corresponds to the Edman sequence cycle number; *Potential Phosphorylation Sites*, any Ser, Thr, or Tyr residues present in the protein; *# in Cycle*, the number of phosphorylatable residues present in a particular cycle number; *Coverage*, the cumulative percentage of phosphorylatable residues included in a particular number of Edman sequencing cycles or cycle numbers. By selecting (✓) one or more particular cycle number(s) and clicking on "Submit Query," a new screen showing a table of the cycle position of the potential phosphorylation residues when cut at specific cleavage sites is obtained (C). These data show all the phosphorylation sites within the selected cycle number(s) along the *top* of the table and all amino acids along the *side* of the table. By orienting a particular residue with a particular cleavage site, the cycle number in which ^{32}P release should occur is obtained. For ease of use, the potential phosphorylation sites for ascending cycle numbers for each cleavage site are also presented (D).

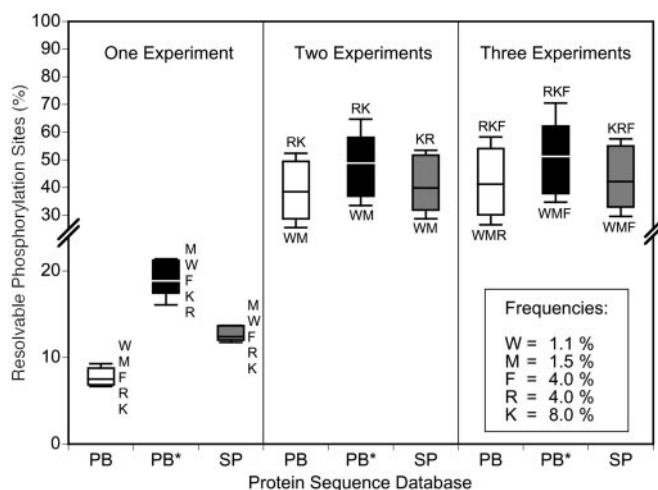


FIG. 2. Theoretical ability of CRP analysis to uniquely identify phosphorylation sites. For all of the potential phosphorylation sites (all Ser (S), Thr (T), and Tyr (Y) residues) in SwissProt (SP) and PhosphoBase (PB) or only PhosphoBase-annotated known phosphorylation sites (PB*), the percentage of sites that could be resolved uniquely with only 25 Edman cycles was measured. Cleavage sites used were Met (M), Trp (W), Phe (F), Lys (K), and Arg (R); the frequencies at which these residues occur in SwissProt are listed in the inset. Shown are boxplots of the coverage exhibited by each experimental methodology; the *midline* of each boxplot indicates the median (50th percentile), whereas the boxed area demarcates the 25th and 75th percentiles. The lines *above* and *below* the boxes represent the 10th and 90th percentiles. Single residue labels in the *left panel* represent the order in which the data occur. Labels in subsequent panels represent the worst and best combination of cleavage sites.

from theoretically obtained data using each of five cleavage agents that target methionine (M), tryptophan (W), phenylalanine (F), lysine (K), or arginine (R). Coverage is measured for all potential phosphorylation sites in SwissProt (SP) and PhosphoBase (PB), as well as only in terms of known phosphorylation sites as described by PhosphoBase annotations (PB*). In SwissProt, around 14% of Ser, Thr, or Tyr sites are resolvable by a single cleavage experiment. Approximately 20% of the known phosphorylation sites from PhosphoBase are resolvable. The relative performance of the various cleavage agents reflects the frequencies of their cleavage sites. Sites that occur more frequently in a protein will lead to a larger number of shorter fragments, increasing the coverage achievable within 25 cycles, but also increasing the chance of ambiguity. Sites that occur less frequently will lead to a smaller number of longer fragments. Therefore, for single cleavage experiments, use of a cleavage reagent with rare-cutting properties increases the number of unambiguous assignments but reduces the total coverage.

Additional cleavage experiments with different endopeptidases may be performed to disambiguate cycles containing multiple residues. For double or triple cleavage experiments (where two or three separate and unique cleavage experiments are performed), we continued to define an unresolved site as either a phosphorylation site that could be seen within

25 cycles in a previous experiment, but ambiguities left the site unresolved, or one that would occur after 25 cycles. Performing additional cleavage experiment increases dramatically the theoretical coverage for both databases (Fig. 2, *center*), identifying 50–60% of known phosphorylation sites. In contrast to the single cleavage experiment, here cleavage reagents that cut more frequently are more effective, as the number of potential sites that fall within 25 cycles is greater, and, in combination, they are more informative. Additional cleavage experiments (Fig. 2, *right*) can improve coverage marginally. However, most of the remaining sites remain unresolved because of stretches of sequence longer than 25 residues that cannot be cleaved, leaving the phosphorylation sites above the 25-cycle threshold. Increasing the cycle threshold to 40, for instance, increases dramatically coverage at all levels of measurement (achieving nearly 90% coverage with some triple cleavage experiments).

For the sake of simplicity, these theoretical experiments assume that only one residue in the protein is phosphorylated. If radioactivity is found in more than one cycle, then additional experiments would produce more candidate residues, increasing the possibility of continued ambiguity. However, a lower limit on coverage may be easily calculated as simply the sum of residues resolvable by each single cleavage; for a triple cleavage experiment, the coverage could be as low as 30%. This represents the worst-case scenario; all of the potential sites of a protein are phosphorylated.

To increase the probability of uniquely identifying phosphorylation sites, a phosphoamino acid analysis can be completed on an aliquot of the phosphoprotein prior to Edman cycle analysis to determine whether the phosphorylation site is a phospho-Ser, -Thr, or -Tyr. By limiting the total number of residues under consideration, this information reduces dramatically the complexity of the CRP results and further resolves assignment ambiguities, increasing the theoretical coverage to nearly 100% in most triple cleavage experiments (data not shown).

Application of CRP Analysis in the Identification of Phosphorylation Sites—To test our strategy, we took advantage of the *in vitro* phosphorylation of telokin documented previously (7). Telokin is a small acidic protein (17 kDa) with a serine/threonine-rich amino terminus and contains substrate recognition sequences for a variety of kinases, including cAMP-dependent protein kinase. Using conventional methods, we have identified previously a single site of *in vitro* phosphorylation on telokin by cAMP-dependent protein kinase as Ser-13 (7).

As shown in Fig. 3A, when peptides from an endopeptidase Lys-C digest were applied to the Edman sequencer, ^{32}P release was observed in cycle 2; when peptides from a cyanogen bromide digest were applied to the Edman sequencer, ^{32}P release was observed in cycle 10. The theoretical results for an endopeptidase Lys-C digest of telokin are displayed in Fig. 3B. In theory, ^{32}P release in cycle 2 could be the result of

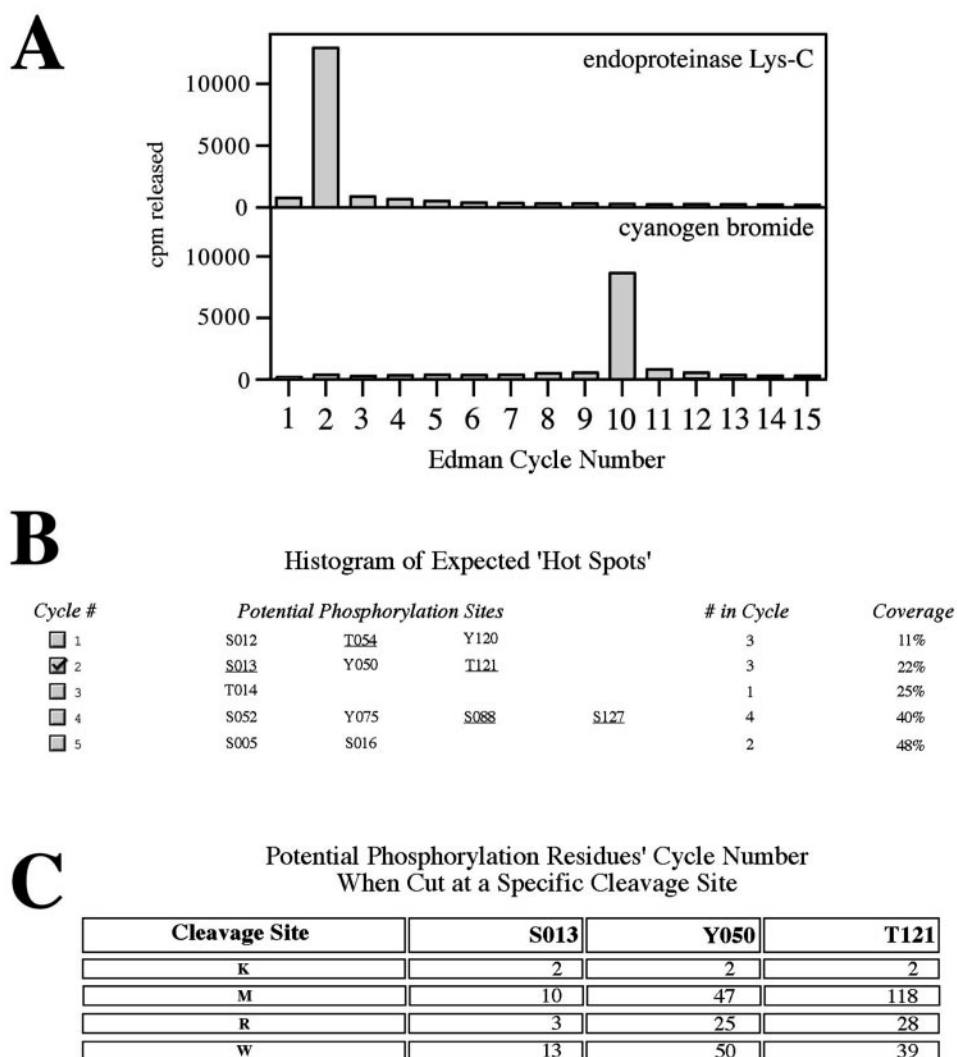


FIG. 3. **Identification of phosphorylation sites from telokin using the CRP analysis program.** A, recombinant telokin was phosphorylated with cAMP-dependent protein kinase to a stoichiometry of 0.9 mol phosphate/mol of protein prior to endoproteinase Lys-C or CnBr digestion. The results represent the amount of ^{32}P released in each cycle when the peptide digests ($\sim 25,000$ cpm) were subjected to sequential Edman solid-phase sequencing. B, the theoretical results obtained from the CRP-program for an endoproteinase Lys-C digest of telokin are presented. C, CRP-program display of the new cycle positions of the potential phosphorylation residues when telokin is cut with different proteases.

phosphorylation at three different sites, Ser-13, Tyr-50, and/or Thr-121. By selecting cycle 2 (✓) a new Table is created (Fig. 3C) that displays the new cycle position of the potential phosphorylation residues when the original protein is cut with a different protease. In this example, cleavage of phosphotelokin with cyanogen bromide would result in ^{32}P release for residues Ser-13, Tyr-50, and Thr-121 in cycle 10, 47, and 118, respectively. By comparing the theoretical data in Fig. 3C with the ^{32}P release data obtained from the sequencer (Fig. 3A), the identity of the phosphorylation site is assigned unambiguously.

Identification of Thrombin-sensitive Phosphorylation Sites on Hsp27 *in Vivo*—We used human platelets as a model system to test the CRP method of phosphorylation site identification on *in vivo* target proteins. The cellular ATP stores of

platelets can be spiked easily by [^{32}P]orthophosphate isotope labeling, and there is a large and rapidly growing body of information on the different phosphoproteins that emerge within platelets from agonist stimulation (13, 14). A number of proteins exhibit increased phosphorylation in response to thrombin stimulation (Fig. 4). Fourteen phosphoproteins were selected for mixed-peptide sequencing, and one of those proteins, Hsp27, was selected as a candidate phosphoprotein for CRP analysis.

We immunoprecipitated Hsp27 from ^{32}P -labeled platelets treated with thrombin. All of the potential phosphorylation sites and the theoretical results of an endoproteinase Arg-C digest of Hsp27 were identified using the CRP program (Fig. 5A). This analysis showed that 25 rounds of Edman sequencing were sufficient to cover 87% of all serine, threonine, and

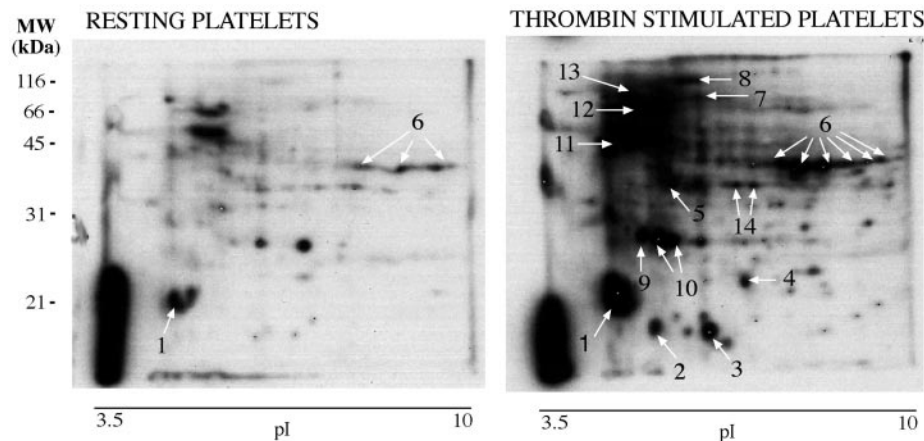


FIG. 4. **Identification of phosphoproteins in human platelets by mixed peptide sequencing.** Platelets from 1.0 ml of human blood were ³²P-labeled for 90 min and then treated with thrombin (0.2 units) in the presence of calyculin A (10 μM) for 1 min. Preliminary experiments determined that calyculin treatment alone did not enhance significantly the level of phosphorylation. Phosphoproteins were characterized on two-dimensional SDS-PAGE gels. The separated proteins were transferred to PVM, stained with Amido Black, and autoradiographed. ³²P-Labeled proteins that corresponded to stained proteins were cut from the PVM, treated with CnBr solution, and sequenced by mixed peptide sequencing. Proteins that were identified ranged from 2 to 0.2 pmol. Isoelectric values (pI) are theoretical values based on the primary sequence of the identified proteins and were determined using the EXPASY program (expasy.hcuge.ch/ch2d/pi_tool.html). 1, myosin light chain 20, pI 5.09, 19.8 kDa; 2, Rho-guanine nucleotide dissociation inhibitor (Rho GDI), pI 5.1, 23.0 kDa; 3, neurogranin, pI 6.53, 7.4 kDa; 4, RAP1A, pI 6.39, 20.9 kDa; 5, tubulin, pI 4.75, 49.7 kDa; 6, Pleckstrin, pI 8.32, 40.1 kDa; 7, GPIIIa-II, pI 5.3, 84.6 kDa; 8, Rho-GAP, pI 5.91, 105 kDa; 9, ESTAA059313; 10, HSP27, pI 5.98, 22.7 kDa; 11, c-Raf oncogenic fusion protein, pI 5.15, 47.0 kDa; 12, amphiphysin homologue, pI 4.95, 64.0 kDa; 13, HSP90, pI 4.94, 84.6 kDa; 14, pleckstrin fragments.

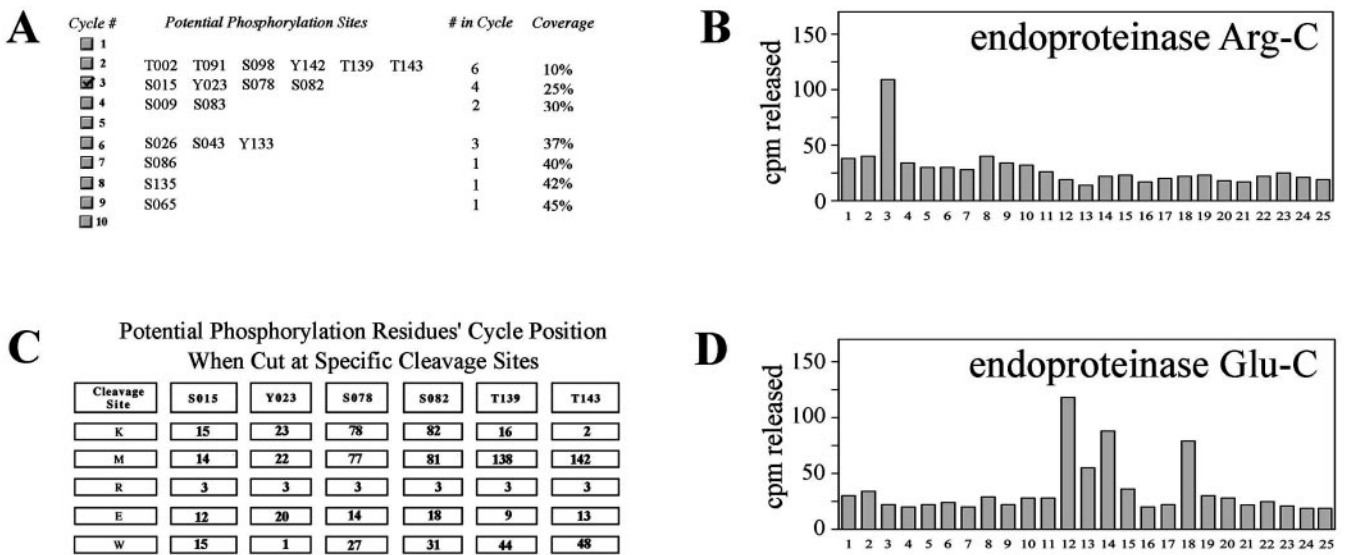


FIG. 5. **Identification of thrombin-sensitive phosphorylation sites on Hsp27 by CRP analysis.** A, theoretical results were obtained from the CRP program for an endoproteinase Arg-C digest of Hsp27. B, Hsp27 was immunoprecipitated from ³²P-labeled platelets treated with thrombin and calyculin A for 5 min. Peptides (~1000 cpm) obtained from endoproteinase Arg-C digestion were cross-linked to Immobilon P (Perspective Biosystems), and the membrane was placed in an Applied Biosystems Procise 494cLc automated sequencer. Twenty-five cycles of Edman degradation chemistry were carried out, and the released phenylthiohydantoin (PTH) amino acids were collected, and their radioactivity was determined by Cerenkov counting. C, CRP program display of the new cycle positions of the potential phosphorylation residues when Hsp27 is cut with different proteases. D, the results of ³²P release from peptides (~2500 cpm) obtained from an endoproteinase Glu-C digest of immunoprecipitated Hsp27 are presented.

tyrosine residues present in Hsp27. Twenty-five rounds of Edman degradation chemistry were carried out on the ³²P-labeled peptides obtained from the endoproteinase Arg-C digest of the immunoprecipitated Hsp27. This analysis yielded

radioactivity only in cycle 3. Inspection of the analysis shown in Fig. 5 shows that six potential phosphorylation sites would produce a signal in cycle 3. Further analysis of these sites by the CRP program algorithm indicated that digestion at Glu

followed by Edman cycle analysis would identify uniquely the position of the thrombin-stimulated *in vivo* phosphorylation site(s). Indeed, Edman degradation chemistry carried out on the peptides obtained from the endoproteinase Glu-C digest yielded radioactivity in cycles 12, 14, and 18. Thus, Ser-15, Ser-78, and Ser-82 were identified unambiguously as major *in vivo* thrombin-sensitive phosphorylation sites in Hsp27.

DISCUSSION

Traditionally, protein phosphorylation sites are located by enzymatic cleavage of a ^{32}P -radiolabeled phosphoprotein into peptides followed by HPLC C18 reverse phase chromatography or two-dimensional thin layer chromatography to isolate and separate the ^{32}P phosphopeptide. The sequence of the phosphopeptide is then obtained by Edman sequencing. The need for a large amount of starting material (more than pmol amounts of protein) and the length of time to completion has made this procedure prohibitive to high throughput studies of the phosphoproteome. We have developed a strategy that permits identification of phosphorylation sites from *in vivo* or *in vitro* ^{32}P -labeled proteins of known sequence at the sub-femtomolar level.

From our initial assessment of the thrombin-stimulated platelet phosphoproteome (Fig. 4), we selected Hsp27 as a phosphoprotein for CRP analysis. In human platelets, increased phosphorylation of Hsp27 through the p38/mitogen-activated protein kinase-activated protein kinase 2 pathway in response to thrombin treatment has been observed previously (15). Hsp27 phosphorylation is associated with platelet aggregation and regulation of microfilament organization. Activation of p38/mitogen-activated protein kinase-activated protein kinase 2 pathway after thrombin stimulation leads to a marked shift from the 27-kDa unphosphorylated form to at least three major phosphorylated forms. The phosphorylation sites on Hsp27 (16, 17) have been mapped previously using conventional methods, *i.e.* proteolytic digestion and fractionation of the peptides by reverse phase HPLC followed by Edman sequence analysis. The sites phosphorylated by mitogen-activated protein kinase-activated protein kinase 2 after *in vivo* thrombin treatment were identified as Ser-15, -78, and -82 (16, 17). The present study confirms Hsp27 as a target of phosphorylation during thrombin stimulation but more importantly demonstrates the ability of the CRP analysis to ascertain multiple sites of phosphorylation on a target phosphoprotein isolated from an *in vivo* source.

Some limitations to the CRP methodology do exist; phosphorylation sites directly adjacent to Lys or Arg may not cut because of steric occlusion of the protease, the 3° structure of the protein may prevent proteolysis at every site hence giving rise to missed cleavages, and ragged cuts at frequently occurring Lys-Lys or Arg-Lys motifs may lead to ambiguous results. However, with experience and careful consideration of the amino acid sequence of the phosphoprotein, these limitations are circumventable. We have had much success using the CRP methodology for the identification of *in vitro*

phosphorylation sites. Although telokin is at the lower end of degree of difficulty, we have successfully used the methodology to determine multiple phosphorylation sites on proteins with molecular masses of up to 130 kDa. With minimal starting product (*i.e.* <10 fmol of protein) we have identified multiple sites of phosphorylation on the protein kinase C phosphatase inhibitor protein, CPI-17 (18), the myosin targeting subunit (MYPT1) of smooth muscle myosin phosphatase,² and the transcriptional co-activator cAMP-response element-binding protein-binding protein (CBP)/p300.³ To increase the probability of uniquely identifying phosphorylation sites, a phosphoamino acid analysis can be completed on an aliquot of the phosphoprotein prior to Edman cycle analysis to determine whether the phosphorylation site is a phospho-Ser, -Thr, and/or -Tyr. By limiting the total number of residues under consideration, this information reduces dramatically the complexity of the CRP results and resolves further assignment ambiguities, increasing the theoretical coverage to nearly 100% in most triple cleavage experiments. The CRP analysis methodology complements existing mass spectrometry techniques (19–21) for phosphorylation site identification, because it presents an alternative method of identification in situations where peptides are unable to be resolved by mass spectrometry.

Acknowledgment—We thank Applied Biosystems for supporting the Haystead laboratory.

* This work was supported in part by a Natural Sciences and Engineering Research Council of Canada post-doctoral fellowship (to J. A. M.), by National Library of Medicine Grant LM04969 (to W. R. P. and A. J. M.), and by National Institutes of Health Grant HL19242-24 (to T. A. J. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Contributed equally.

** To whom correspondence should be addressed: Dept. of Pharmacology and Cancer Biology, Box 3813, Duke University, Durham, NC 27710. Tel.: 919-613-8606; Fax: 919-668-0977; E-mail: hayst001@mc.duke.edu.

REFERENCES

1. Godovac-Zimmermann, J., and Brown, L. R. (2001) Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom. Rev.* **20**, 1–57
2. Gygi, S. P., and Aebersold, R. (2000) Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* **4**, 489–494
3. Jung, E., Heller, M., Sanchez, J. C., and Hochstrasser, D. F. (2000) Proteomics meets cell biology: the establishment of subcellular proteomes. *Electrophoresis* **21**, 3369–3377
4. Alms, G. R., Sanz, P., Carlson, M., and Haystead, T. A. J. (1999) Reg1p targets protein phosphatase 1 to dephosphorylate hexokinase II in *Saccharomyces cerevisiae*: characterizing the effects of a phosphatase subunit on the yeast proteome. *EMBO J.* **18**, 4157–4168

² M. A. Borman, J. A. MacDonald, A. Muranji, D. Hartshorne, and T. A. J. Haystead (2002) Smooth muscle myosin phosphatase-associated kinase induces Ca^{2+} -independent contraction via myosin phosphatase inhibition. *J. Biol. Chem.*, submitted for publication.

³ E. E. Corcoran, J. D. Joseph, J. A. MacDonald, T. A. J. Haystead, and A. R. Means, unpublished results.

5. MacDonald, J. A., Borman, M. A., Muranji, A., Somlyo, A. V., Hartshorne, D. J., and Haystead, T. A. J. (2001) Identification of the endogenous smooth muscle myosin phosphatase-associated kinase. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2419–2424
6. Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362
7. MacDonald, J. A., Walker, L. A., Gorenne, I., Nakamoto, R., Somlyo, A. V., Somlyo, A. P., and Haystead, T. A. J. (2000) Phosphorylation of telokin by cyclic nucleotide kinases and the identification of *in vivo* phosphorylation sites in smooth muscle. *FEBS Lett.* **479**, 83–88
8. Mustard, J. F., Perry, D. W., Ardlie, N. G., and Packham, M. A. (1972) Preparation of suspensions of washed platelets from humans. *Br. J. Haematol.* **22**, 193–204
9. Damer, C. K., Partridge, J., Pearson, W. R., and Haystead, T. A. J. (1998) Rapid identification of protein phosphatase 1-binding proteins by mixed peptide sequencing and data base searching. Characterization of a novel holoenzymic form of protein phosphatase 1. *J. Biol. Chem.* **273**, 24396–24405
10. Russo, G. L., Vandenberg, M. T., Yu, I. J., Bae, Y. S., Franza, B. R., Jr., and Marshak, D. R. (1992) Casein kinase II phosphorylates p34cdc2 kinase in G1 phase of the HeLa cell division cycle. *J. Biol. Chem.* **267**, 20317–20325
11. Bairoch, A., and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48
12. Kreegipuu, A., Blom, N., and Brunak, S. (1999) PhosphoBase, a database of phosphorylation sites: release 2.0. *Nucleic Acids Res.* **27**, 237–239
13. Wadman, I. A., Virdee, K., Fernandez, D. S., Wasunna, C. L., and Farndale, R. W. (1996) *Platelets: A Practical Approach* (Watson, S. P., and Authi, K. S., eds) pp. 173–198, Oxford University Press, New York
14. Clark, E. A., and Brugge, J. S. (1996) *Platelets: A Practical Approach* (Watson, S. P., and Authi, K. S., eds) pp. 199–216, Oxford University Press, New York
15. Mendelsohn, M. E., Zhu, Y., and O'Neill, S. (1991) The 29-kDa proteins phosphorylated in thrombin-activated human platelets are forms of the estrogen receptor-related 27-kDa heat shock protein. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 11212–11216
16. Landry, J., Lambert, H., Zhou, M., Lavoie, J. N., Hickey, E., Weber, L. A., and Anderson, C. W. (1992) Human HSP27 is phosphorylated at serines 78 and 82 by heat shock and mitogen-activated kinases that recognize the same amino acid motif as S6 kinase II. *J. Biol. Chem.* **267**, 794–803
17. Stokoe, D., Engel, K., Campbell, D. G., Cohen, P., and Gaestel, M. (1992) Identification of MAPKAP kinase 2 as a major enzyme responsible for the phosphorylation of the small mammalian heat shock proteins. *FEBS Lett.* **313**, 307–313
18. MacDonald, J. A., Eto, M., Borman, M. A., Brautigan, D. L., and Haystead, T. A. J. (2001) Dual Ser and Thr phosphorylation of CPI-17, an inhibitor of myosin phosphatase, by MYPT-associated kinase. *FEBS Lett.* **493**, 91–94
19. Ficarro, S. B., McClelland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305
20. Zhou, H., Watts, J., and Aebersold, R. A. (2001) A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19**, 375–378
21. Oda, Y., Nagasu, T., and Chait, B. (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379–382