An Oriented Peptide Array Library (OPAL) Strategy to Study Protein-Protein Interactions

Received for publication, October 29, 2003, and in revised form, December 15, 2003
Published, JBC Papers in Press, December 16, 2003, DOI 10.1074/jbc.M311886200

Maria Rodriguez‡, Shawn S.-C. Li§, J. Wade Harper‡¶, and Zhou Songyang‡‡

From the ‡Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030 and the §Department of Biochemistry, Faculty of Medicine and Dentistry, University of Western Ontario, London, Ontario N6A 5C1, Canada

One of the major questions in signal transduction is how the specificities of protein-protein interactions determine the assembly of distinct signaling complexes in response to stimuli. Several peptide library methods have been developed and widely used to study protein-protein interactions. These approaches primarily rely on peptide or DNA sequencing to identify the peptide or consensus motif for binding and may prove too costly or difficult to accommodate high throughput applications. We report here an oriented peptide array library (OPAL) approach that should facilitate high throughput proteomic analysis of protein-protein interactions. OPAL integrates the principles of both the oriented peptide libraries and array technologies. Hundreds of pools of oriented peptide libraries are synthesized as amino acid scan arrays. We demonstrate that these arrays can be used to map the specificities of a variety of interactions, including antibodies, protein domains such Src homology 2 domains, and protein kinases.

Signaling proteins are often composed of distinct protein domains or modules. Some of these serve as protein-protein interaction units, whereas others carry out enzymatic reactions. The elucidation of the genetic codes of various organisms, including humans, has presented an increasing number of protein domain families with unknown functions.

To understand how signaling proteins interact with each other and maintain the fidelity of signal transduction, it is critical to determine the specificities of these protein domains. Protein domains (such as SH2, SH3, and phosphotyrosine binding (PTB) domains) often mediate protein-protein interactions by recognizing short stretches of peptide sequences on target proteins (2, 3). In the past several years, many combinatorial peptide library based approaches have been developed to analyze the binding or substrate specificities of protein domains (reviewed in Ref. 4). These peptide libraries can be classified as either display or pool peptide libraries.

The display libraries employ a variety of matrices for individual peptide presentation, including pin (5), phage (6–8), beads (9), DNA binding proteins (10), or ribosomes (11). Following multiple rounds of affinity enrichment steps, the identities of the selected peptides can be ascertained by DNA or peptide sequencing (4). Display peptide libraries generally require isolation and sequencing of individual peptides to decode a binding motif, which is time-consuming and costly, limiting it to low throughput applications.

In contrast, the pool peptide libraries rely on the comparison of binding affinities of different pools of random peptides to obtain binding consensus motifs (12). Thus, the sequences of individual bound peptides do not need to be resolved. A few years ago, we developed an oriented peptide library method to further simplify and facilitate screening (13). In this method, soluble pools of random peptides are oriented via a central “fixed” amino acid, which reduced the degeneracy of the library and prevented possible phase problems arising from interacting residues being located at different positions on different peptides. Bound peptides or those modified by the protein of interest are then enriched and sequenced as a mixture to deconvolute the consensus motif in a statistical manner. The soluble oriented peptide library approach still depends on peptide sequencing, and a large amount of purified proteins (~100 μg) are generally needed for the assay. In addition, the occasional incomplete cleavage of peptides during Edman degradation sequencing can result in a carryover of amino acids from one sequencing cycle to the next, thereby complicating the results.

We report here an oriented peptide array library (OPAL) approach in which hundreds of pools of oriented peptide libraries are synthesized and arranged as scan arrays. We showed that OPAL could be used as a high throughput method to study the specificities of protein domains.

MATERIALS AND METHODS

**GST Fusion Proteins**—GST-GRB7 and GST-GRB10 fusion constructs were generated by ligating cDNA fragments encoding mouse GRB7 (amino acids 422–535) and GRB10 (amino acids 434–541) into pGEX-4T1 (14). Construction of expression vectors for GST-GRB2, SHC, and SHP-2 SH2 domains was reported previously (15). GST fusion proteins were purified as described previously (13).

**Peptide Library Synthesis**—The peptide libraries were synthesized on cellulose membrane using the ASP222 SPOT robot (16). An extra fine needle tip was used to achieve a density of ~1600 peptides per SPOT membrane (8 × 12 cm). The estimated yield of peptide at each spot was ~5 nmol. The following oriented peptide library arrays were synthesized for phosphorylation dependent interactions: XXXXX[pS/pT]AXXXA; AXXXX[pY]XXXAXX; and AXXXX[pS/pT][QDPF]XXXA; where X is a mixture of 19 amino acids (except Cys), pS is phosphoserine, pT is phosphothreonine, and the brackets ([]) encase the amino acids that were preferred by the protein of interest. To generate oriented peptide library pools, each degenerated

8802

This paper is available online at [http://www.jbc.org](http://www.jbc.org)
position was scanned with any of the 19 amino acids (excluding Cys).

A library with the sequence $AX_1X_2X_3X_4[S/T]X_5X_6X_7A$ was synthesized for protein kinases, where $X_i$ represents any amino acid but not Ser, Thr, and Cys. Ser and Thr are excluded at the degenerate $X_i$ positions; thus, only the fixed Ser/Thr can be phosphorylated. To generate oriented peptide library pools, each degenerate position was substituted with any of the 20 amino acids.

**OPAL Analysis of Antibody and SH2 Domain Specificities**—All steps were carried out at room temperature unless otherwise specified. The OPAL membrane was first blocked with 5% bovine serum albumin in TBST (0.1M Tris-HCl (pH 7.4), 150 mM NaCl, and 0.1% Tween 20) for 2 h. GST fusion proteins (3 μg) were labeled with GSH-horseradish peroxidase (Sigma) (0.6 μg) for 30 min and then added to the array membrane at a final concentration of 1 μg/ml for 30 min (17). The array membrane was subsequently washed three times with TBST for 10 min, and peptide spots that bound the domain of interest were visualized by enhanced chemiluminescence.

To determine the specificities of antibodies, monoclonal antibodies MPM2 (1 μg/ml) (18), 3F3/2 (10 μg/ml) (19), and PY20 (1 μg/ml, Santa Cruz Biotechnology) were first incubated with OPAL membranes for 30 min at room temperature. The membranes were washed three times with TBST for 15 min, probed with anti-mouse horseradish peroxidase, and bound peptide spots were visualized by enhanced chemiluminescence.

**Kinase Assays**—Kinase assays on membranes were performed as described (20). Briefly, the OPAL membrane was blocked overnight in kinase buffer (62.5 mM Mes, pH 6.5, 20 mM MgAc2, 100 mM NaCl, 1 mM dithiothreitol, and 3% bovine serum albumin). The next day, 2 μg of crude protein kinase A (Sigma) in kinase buffer containing 25 μM ATP and 20 μCi of [γ-32P]ATP was added to the OPAL membrane and incubated for 2 h at 30°C. The membrane was then washed five times with 1× NaCl and ten times with 5% phosphoric acid. 32P-labeled peptide spots were visualized and quantitated with PhosphorImager (Amersham Biosciences). His6-Cdc15 was prepared in insect cells and purified using nickel-nitrilotriacetic acid resin (Qiagen).

**Data Base Search**—A weighed data matrix was first generated based on the OPAL analysis result. Data base search of PKA substrates was performed on Scansite (21).

### RESULTS AND DISCUSSION

**The Oriented Peptide Array Library Strategy**—The OPAL method involves synthesis and printing of oriented peptide libraries as spotted arrays on solid supports (Fig. 1). OPAL can be synthesized on solid phase and then cleaved to be spotted on array surfaces such as gold or glass slides. Alternatively, OPAL can be synthesized directly onto solid supports such as SPOT membranes (16). For example, an oriented library $AX_1X_2X_3X_4[pS/\text{pT}]X_5X_6X_7A$ was designed for proteins that recognize phospho-serine or phosphothreonine (Fig. 2A) in which pS and pT are fixed while $X_i$ is random (any natural or amino acid except for Cys). Because there is no pS or pT in the random positions, pS/pT becomes the functional interaction center. The random positions are thus oriented with regard to their distance to the fixed pS/pT. The relative importance of different amino acids at any given position can then be addressed by using a scan library strategy (12, 22). For instance, for $Z$ in $ZX_1X_2X_3X_4[pS/pT]X_5X_6X_7A$ (i.e. position P-4 N-terminal to the pS/pT), 19 pools of libraries were synthesized. Each pool had one of the 19 amino acids (except Cys) fixed at position $Z$ and was spotted as

![FIG. 1. Schematic presentation of the OPAL strategy.](image)

![FIG. 2. Preference of phospho-specific antibodies determined by OPAL Strategy](image)
a single spot on the array. As a result, each pool/spot can, in
theory, contain a half-billion oriented peptides. A total of 152
(19 amino acids/H11003/8 positions) pools of libraries were similarly
spotted for the A
[124]XX
[148]X
[177]X
[201]A library. The OPAL array
essentially scans for the amino acids on the x axis and for the
positions on the y axis. If a library pool at any given spot is
preferred by the protein of interest, this mixture will be de-
tected by antibody blotting and show up as a positive spot.
Consequently, the specificity of a protein domain can be deter-
dined by reading the library array without complicated data
manipulation.

Epitope Mapping of Phospho-specific Antibodies Using
OPAL—The OPAL approach was first tested to determine the
epitopes of phospho-specific antibodies such as MPM2. MPM2
has been shown to react with phosphoproteins in the M phase
of the eukaryotic cell cycle (18). It recognizes phosphopeptides
with the consensus motif [FYW][FW][FLI][pS/pT]P[ILVFM], as
analyzed by the soluble oriented peptide library and expression
cloning approaches (23, 24).

We synthesized a pS/pT fixed OPAL (A
[478]XX
[502]X
[531]X
[555]A,
[317]Fig. 2
A
) directly on SPOT membranes with a density (library pools/cm2) four times that of conventional SPOT arrays.
The MPM2 antibody was incubated with this OPAL, and the
MPM2-bound library spots were then visualized by enhanced
chemiluminescence. As shown in Fig. 2
B
, stronger binding was
observed at the P
[398]1 and P
[441]2 positions, where hydrophobic
amino acids (mainly Ile, Leu, Val, and Phe) were selected.
Hydrophobic amino acids were also preferred (but more
weakly) by MPM2 at the P
[433]2 and P
[475]3 positions. At the P
[475]1 position, we found that Gly was also selected in addition to Pro,
which was previously known to be there. Therefore, MPM2 prefers
sequences containing the motif [FILV][pS/pT][PG][ILVF]. These
findings are in good agreement with published data (23, 24) and indicate that the OPAL approach can be used to analyze the binding specificity of antibodies.

We went on to study the epitopes of two additional phospho-specific antibodies, 3F3/2 and PY20 (Fig. 2, C and D). Monoclonal antibody 3F3/2 was originally developed against *Xenopus* egg extracts that had been supplemented with ATPγS (25). It was later found to recognize an unknown kinetochore phospho-epitope (19). The OPAL analysis determined that 3F3/2 preferred the motif [FY][X][DE][X][pS/pT][X][DE] (Fig. 2C). This specificity is consistent with a known site recognized by 3F3/2 on human topoisomerase II (DFDEKTDDED) (26). Furthermore, the OPAL method can be applied to examine a phosphotyrosine-specific antibody. As shown in Fig. 2D, the monoclonal PY20 antibody recognized almost all the spots equally well except for Asp/Glu at positions P + 1 and P + 3, implying a broad specificity of this antibody. These results further validated the OPAL method.

**Determining the Binding Specificities of Protein Domains**—A large number of both known and new domains that have been identified through the genome projects await the elucidation of their function and targets. Conventional peptide and oriented peptide library techniques have been used extensively to map the binding specificities of protein-protein interaction domains. In many cases, this has yielded novel insights into signaling interactions through data base searches for sequence motifs. In comparison, the OPAL technique holds a clear advantage over the conventional methods, as it requires no peptide sequencing or data manipulation. We decided to verify this with SH2 domains. SH2 domains are a large family of phosphotyrosine (pY)-dependent interaction modules that have been extensively studied.

We started with the SH2 domain of the adaptor signaling molecule GRB2, because its specificity is known (27). As predicted, OPAL analysis indicated that the GRB2 SH2 domain

---

**TABLE I**

| Peptide sequence | $K_m$ | Relative $K_{cat}$ |
|------------------|-------|-------------------|
| ARRGSIGFI        | 23    | 1                 |
| ARRGSIDFI        | 8.8   | 3.5               |
| ARRGSCIFLI       | 9.1   | 2                 |

---

**Fig. 5.** Scansite matrix for searching of potential in vivo substrates of PKA. Based on the amount of $^{32}$P incorporation by PKA in each peptide spot, data in Fig. 4A were transformed into bar graphs to indicate the relative importance of amino acids at each position N- and C-terminal to the fixed Ser/Thr residue. The total values were normalized to 20 (number of possible natural amino acids) such that a value of <1 would denote negative selection.
recognized primarily Asn at the P + 2 position with little or no selectivity for residues N-terminal to the pY (Fig. 3A). Ile and Val or Leu residues were also selected at the P + 1 and P + 3 positions. This motif was determined in a single scan and is in perfect agreement with previous analyses of the GRB2 SH2 domain (13).

We then went on to study three additional SH2 domains, namely SHP-2 C-terminal SH2 (28, 29), GRB7, and GRB10 (14). SHP-2, a phosphatase containing two SH2 domains, is an important regulator of cell proliferation and development (30). The N-terminal SH2 domain has been shown to select a [FY]pY[ETYS]N[ILVPTYS] motif (13), but the preference of the C-terminal SH2 domain is [ILV][ILV][ILVFTY]pY[TILV][IL][ILVP]. This result suggests that the Ile, Leu, and Val residues flanking the phosphotyrosine may be necessary for SHP-CSH2 domain binding. It further offers an explanation for the confusion regarding the known in vivo binding sites of SHP-2, where Ile, Leu, and Val residues flanking the phosphotyrosine may be necessary for SHP-CSH2 domain binding. Indeed, it is random (any natural or modified amino acid except for Ser, Thr, and Cys) (Fig. 4). The random position was subsequently substituted by one of the 20 natural amino acids, including Ser and Thr, and would become fixed. At any given spot, if a library pool was preferred by the kinase of known specificity of PKA (RR\chi4), it would reflect the preference by the kinase for the fixed amino acid at that position. An added advantage of this design is the inclusion of Ser and Thr among the "scanned" amino acids, which should allow us to determine whether a kinase prefers Ser or Thr.

We first examined the kinase domain specificity of PKA using the array. As shown in Fig. 4A, positions relative to Ser/Thr are indicated on the y axis. The amino acids fixed for each spot is indicated on the top x axis. At the P - 2 and P - 3 positions, Arg is preferred by PKA. At the P + 1 position, hydrophobic amino acids are generally favored by PKA, with Ile being the best favored. PKA also preferred Ser over Thr (see positions -1 to +4). These results correspond perfectly with the known specificity of PKA (RRXXS) and yield a clearer picture regarding PKA specificity than was previously obtained using the soluble oriented peptide library approach (33).

Interestingly, the OPAL approach also identified novel sequence preferences of PKA. At the P + 2 site we found that PKA phosphorylated peptides with Cys and Asp 1.9- and 2.7-fold better, respectively, than those with other amino acids (e.g., Gly). To confirm this finding, several peptides that differed only at the P + 2 residue (ARRGSIGFI, ARRGSDFI, and ARRGSCFI, with the P + 2 residue in boldface) were synthesized based on the deduced PKA substrate specificity.
(Fig. 4A). Kinetic analyses demonstrated that the Kemp tide-like peptide was phosphorylated by PKA with a $K_m$ of ~25 $\mu$m (Table I). This value is similar to previously reported data using a LIRASLG peptide as the substrate (34). In comparison, substitution of P + 2 residue to Asp and Cys lowered the $K_m$ value to 8.8 and 9.1 $\mu$m, respectively, while improving the overall $K_m$ by 3.5- and 2-fold. These data are in good agreement with the array analysis result. Our analyses underscored, therefore, the simplicity and power of the OPAL approach to mapping the specificities of protein kinases.

We also expanded the analysis of OPAL for determining protein kinase specificity to a kinase, budding yeast Cdc15, whose specificity had not been previously studied using conventional peptide library approaches. Cdc15 functions in the protein kinase specificity to a kinase, budding yeast Cdc15, mapping the specificities of protein kinases. Therefore, the simplicity and power of the OPAL approach to kinases.

In this report, we have demonstrated that the OPAL method is a highly efficient and informative tool for studying protein-protein interactions. Compared with conventional peptide library methods, several advantages of OPAL analysis stand out. For example, data interpretation becomes easier as the need for peptide sequencing is eliminated and the results are “read” directly from the array. The amount of proteins required for analysis is also far less than that needed for soluble oriented libraries. In particular, if peptides are printed as microarrays on glass slides, the materials utilized will be minimal. For analyzing protein module binding, the arrays can be reused multiple times. Furthermore, by improving synthesis conditions, we were able to synthesize a large number of arrays within a short time (>8,000 peptide pools (12–15-mer) in a week). Therefore, it should be easy to adopt OPAL for high throughput screening of binding or enzymatic specificities of protein domains in the human genome.

Acknowledgment—We thank Dr. Dan Liu for help with the manuscript.

REFERENCES

1. Pawson, T., and Nash, P. (2003) Science 300, 445–452
2. Songyang, Z. (1999) Prog. Biophys. Mol. Biol. 71, 359–372
3. Yaffe, M. B., and Elia, A. E. (2001) Curr. Opin. Cell Biol. 13, 131–138
4. Cabiby, S. (ed) (1998) Methods in Molecular Biology: Combinatorial Peptide Library Protocols, Vol. 87, pp. 87–97, Humana Press, Totowa, New Jersey
5. Maizel, J. V., Bray, A. M., and Geysen, H. M. (1990) J. Immunol. Methods 134, 23–33
6. Devlin, J. J., Panganiban, L. C., and Devlin, P. E. (1990) Science 245, 404–406
7. Songyang, Z., and Smith, G. F. (1997) Science 278, 386–390
8. Cwirala, S. E., Peters, E. A., Barret, R. W., and Dower, W. J. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 6378–6382
9. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) Nature 348, 82–88
10. Cull, M. G., Miller, J. F., and Schatz, P. J. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 1865–1869
11. Matteakis, L. C., Bhatt, R. R., and Dower, W. J. (1994) Nat. Acad. Sci. U. S. A. 91, 9022–9026
12. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T., and Chou, J. H. (1991) Nature 354, 84–86
13. Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnoffsky, S., Lechleider, R. J., Neel, B. G., Biege, R. R., Fujardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., and Cantley, L. C. (1993) Cell 72, 767–774
14. Osi, J., Yajnik, V., Immanuel, D., Gordon, M., Moskow, J. J., Buchberg, A. M., and Marolos, B. (1995) Oncogene 16, 1621–1630
15. Songyang, Z., Shoelson, S. E., McGuire, J. Olivier, P., Lawson, T., Bustin, R. X., Barbaud, M., Saha, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnoffsky, F., Feldman, R. A., and Cantley, L. C. (1994) Mol. Biol. Cell 14, 5777–5785
16. Frank, R. (2002) J. Immunol. Methods 267, 13–26
17. Nollau, P., and Mayer, B. J. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 11531–11536
18. Davis, F. M., Tsao, T. Y., Fowler, S. K., and Rao, P. N. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 2926–2930
19. Campbell, M. S., and Grabsky, G. G. (1997) J. Cell Biol. 129, 1195–1204
20. Teppge, W. J., and Frank, R. (1998) Methods Mol. Biol. 87, 99–106
21. Yaffe, M. B., Leparc, G. G., Lai, J., Obata, T., Volinia, S., and Cantley, L. C. (2001) Nat. Biotechnol. 19, 348–353
22. Geysen, H. M., Melsen, R. H., and Bartling, S. J. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 3998–4002
23. Westendorf, J. M., Rao, P. N., and Gerace, L. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 714–718
24. Yaffe, M. B., Schatkovsky, M., Shen, Z. X., Stukanen, P. T., Rahlff, J. U., Xu, J., Kuang, J., Kirchner, M. W., Fischer, G., Cantley, L. C., and Lu, K. P. (1997) Science 278, 1857–1900
25. Cyert, M. S., Scherson, T., and Kirchner, M. W. (1988) Dev. Biol. 129, 209–216
26. DuRun, J. R., and Grabsky, G. J. (1999) J. Biol. Chem. 273, 30622–30629
27. Pawson, T., and Schlessinger, J. (1993) Curr. Biol. 6, 434–442
28. Freeman, R. M., Jr., Plutzky, J., and Neel, B. G. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 11239–11244
29. Feng, S. G., Hui, C. C., and Pawson, T. (1993) Science 259, 1607–1611
30. Neel, B. G., Hu, H., and Pao, L. (2003) Trends Biochem. Sci. 28, 284–293
31. Han, D. C., Shen, T. L., and Guan, J. L. (2001) Oncogene 20, 6315–6321
32. Thomas, R., Lenartzons, J., Carlberg, M., and Ronnstrand, L. (1999) Biochem. J. 341, 211–216
33. Songyang, Z., Bleichner, S., Hoogland, N., Hoekstra, M. F., Piwnica, W. H., and Cantley, L. C. (1994) Curr. Biol. 4, 973–982
34. Kemp, B. E., Graves, D. J., Benjamini, E., and Kreb, E. G. (1977) J. Biol. Chem. 232, 4888–4894
35. Mah, A. S., Jung, J., and Dshuaia, R. J. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 7225–7230
36. Dooley, C. T., and Houghten, R. A. (1993) Life Sci. 52, 1509–1517
37. Geysen, H. M., Rodda, S. J., and Mason, T. J. (1986) Mol. Immunol. 23, 709–715