Comparison of Peptide Array Substrate Phosphorylation of c-Raf and Mitogen Activated Protein Kinase Kinase Kinase 8

Kaushal Parikh[1]*, Sander H. Diks[1], Jurriaan H. B. Tuynman[2], Auke Verhaar[3], Mark Löwenberg[3], Daan W. Hommes[3], Jos Joore[4], Akhilesh Pandey[5], Maikel P. Peppelenbosch[1]

1 Department of Cell Biology, Section Immunology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, 2 Laboratory for Experimental Internal Medicine, Academic Medical Center, Amsterdam, The Netherlands, 3 Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, The Netherlands, 4 Pepscan Presto, Lelystad, The Netherlands, 5 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, United States of America

Abstract

Kinases are pivotal regulators of cellular physiology. The human genome contains more than 500 putative kinases, which exert their action via the phosphorylation of specific substrates. The determinants of this specificity are still only partly understood and as a consequence it is difficult to predict kinase substrate preferences from the primary structure, hampering the understanding of kinase function in physiology and prompting the development of technologies that allow easy assessment of kinase substrate consensus sequences. Hence, we decided to explore the usefulness of phosphorylation of peptide arrays comprising of 1176 different peptide substrates with recombinant kinases for determining kinase substrate preferences, based on the contribution of individual amino acids to total array phosphorylation. Employing this technology, we were able to determine the consensus peptide sequences for substrates of both c-Raf and Mitogen Activated Protein Kinase Kinase Kinase 8, two highly homologous kinases with distinct signalling roles in cellular physiology. The results show that although consensus sequences for these two kinases identified through our analysis share important chemical similarities, there is still some sequence specificity that could explain the different biological action of the two enzymes. Thus peptide arrays are a useful instrument for deducing substrate consensus sequences and highly homologous kinases can differ in their requirement for phosphorylation events.

Competing Interests: Jos Joore is employed by Pepscan Systems and adheres to all the PLoS ONE policies on sharing data and materials.

* E-mail: k.parikh@med.umcg.nl

Introduction

Phosphorylation by protein kinases is involved in many facets of cellular regulation and plays an integral part of almost all signalling pathways by bringing about the transduction and amplification of various upstream signals [1–3]. Manning *et al* identified 518 putative protein kinase genes in humans, half of which were found to map to disease loci or cancer amplicons [4]. Most of these kinases are yet to be characterized and a substrate profile for each of these kinases would not only help decipher the complexity of these signalling cascades, but also enable the discovery of drug candidates to regulate their enzymatic activity.

Various methods have been described to predict phosphorylation sites by specific kinases: a database created by Kreegipuu *et al* from annotated phosphorylation sites found in literature[5] and Blom *et al* have used an artificial neural network method to predict eukaryotic phosphorylation sites [6]. Brinkworth *et al* have made use of the available crystal structures, molecular modelling and sequence analyses of kinases and substrates to predict the optimal substrate peptides [7]. Songyang *et al* have made use of an oriented peptide library to predict optimal substrates of protein kinases [8]. In this method, the kinase of interest was added to a soluble

mixture of two and a half billion distinct peptides and then the phosphorylated peptides were separated from the bulk of non-phosphorylated peptides and sequenced to determine an optimal sequence for the kinase. Although a powerful and precise strategy, it is a very expensive and time consuming method.

Recent developments in array technology have now made it possible to make protein chips to study protein substrate interactions, and peptide chips for ligand-receptor interactions and enzymatic activities [9–15]. Very recently, Diks *et al* designed a novel peptide array to make descriptions of total cellular kinase activity [16]. In this approach, kinase substrates described in Phosphobase [17] were spotted on glass and incubated with cell lysates and radio active ATP. Subsequent phosphorylation of the peptides provided substrate phosphorylation profiles of LPS (lipopolysaccharide) -stimulated monocytes [16] and was also instrumental for the discovery of Lck (p56Lck) and Fyn (p59Fyn) as early targets of glucocorticoids [18]. It was also observed that many peptides were differentially phosphorylated. Many other studies using these arrays have been recently reported in the literature [19–22].

Importantly, in the study of Diks *et al* [16], purified PKA (Protein Kinase A) was employed for peptide array phosphoryla-

tion, and the results obtained suggested that peptide array phosphorylation was indeed capable of extracting the known optimal phosphorylation motif for PKA, although this possibility was not investigated in detail in this study. Encouraged by these results, we decided to explore the usefulness of peptide arrays for predicting optimal substrate sequences for kinases with as yet unknown substrate preferences. To this end, we used smaller arrays to study enzyme kinetics and determine experimental conditions of peptide array phosphorylation by purified kinases. These arrays, which were kindly provided by Pepscan Systems (The Netherlands), have 192 peptides spotted in duplicates. Diks et al have described the design of this array in great detail [16]. Subsequently, we employed a commercially available array, exhibiting 1176 Phosphobase database substrates to characterize the effects of two different MAP kinase kinase kinases (MAP3K). MAP3Ks form a part of a module which is classically activated by G-proteins. MAP3Ks on activation phosphorylate and activate a MAP kinase kinase (MAP2K; e.g., MEK) and finally activate a MAP kinase (MAPK; e.g., ERK). Thus, this MAP3K-MAP2K-MAPK module represents critical intermediate effectors that either positively or negatively propagate extracellular stimuli into cellular responses, such as differentiation, proliferation, and apoptosis. Two members from the MAP3K family, namely, c-Raf AND MAP3K8 (Mitogen activated kinase kinase kinase 8/c-Cot/Tpl-2) were used in this study.

c-Raf is a kinase important in human pathology, for instance, as a mediator of oncogenic Ras [23–26] or as an oncogene in its own right [27–29]. More recently c-Raf was also implicated as an essential mediator in chronic inflammation [30]. Analysis of the contribution of the individual amino acids in substrate peptides to total phosphorylation patterns enabled us to deduce a substrate consensus sequence for c-Raf. We were able to validate our results by using a different array containing 1024 peptide sequences derived from motifs in human proteins that are known to be phosphorylated. Analysis of the *in vitro* phosphorylation of this array yielded an almost identical preferential substrate sequence for c-Raf. Furthermore, we decided to exploit the possibility to use peptide arrays to predict kinase consensus sequences for deducing the preferential substrate peptide sequence of MAP3K8, a kinase which is homologous to c-Raf, but has a completely different function in cellular physiology, prompting the question whether both kinases share the same substrate preference or whether, despite the similarity in sequences both kinases have sufficient substrate specificity to account for the differences in biological function. MAP3K8 has been shown to participate in the transcriptional regulation of several important genes, including those for tumour necrosis factor alpha and IL-2 (Interleukin 2) [31–33]. MAP3K8 is also an integral component of signalling pathways that control the proteolytic processing of the NF-κB1 p105 protein [34] and is able to stimulate NF-κB-dependent transcription through the interaction and activation of the NF-κB-inducing kinase (NIK) [35]. Our study shows that peptide arrays are useful for deducing substrate consensus sequences and highly homologous kinases can differ in their requirement for phosphorylation events.

Materials and Methods

Reagents

Truncated constitutively active human MAP3K8 kinase and c-Raf were purchased from Upstate Biotechnology (Upstate Biotechnology, Lake Placid, NY). ^{33}P-γ-ATP was purchased from Amersham Biosciences (Amersham Biosciences AB, Uppsala, Sweden). MEK$^{Ser218/222}$/MEK2$^{Ser222/226}$ antibodies were purchased from

Upstate Biotechnology (Upstate Biotechnology, Lake Placid, NY). Lysis buffer was purchased from Cell Signaling Technology. Lysis buffer was supplemented with protease and phosphatase inhibitors, including 1 mM sodium fluoride, 1 μg/ml leupeptin, 1 μg/ml aprotinin, and 1 mM PMSF (Phenylmethanesulfonyl fluoride).

Peptide Array design

The trial arrays consisting of 192 peptides were kindly provided by Pepscan Systems (Lelystad, The Netherlands). The full list of these substrates is listed elsewhere by Diks et al [16]. The array consisting of 1176 substrates was purchased from Pepscan systems (Lelystad, The Netherlands) and the design is described in detail on their website: http://www.pepscanpresto.com/index.php?id = 27. We used a second array consisting of 1024 peptides based on known phosphorylated motifs in human proteins found in the Human Protein Reference Database (HPRD) [36] spotted in triplicate again made available by Pepscan Systems Briefly, a panel of known, phosphorylation motifs derived from different signalling cascades were selected from the proteins annotated in HPRD. Full list of peptides is available under license from the manufacturer's website. This is in contrast to the 1176 array, which used an unbiased set of amino-acid motifs that could be phosphorylated.

Enzyme kinetics

Trial peptide arrays consisting of 192 peptide substrates were used to test enzyme kinetics. 5 μg/ml of purified active MAP3K8 was incubated with trial peptide arrays for 1, 5, 10, 20, 30, 45, 60 and 120 minutes.

In vitro kinase assays

In vitro kinase assays were used according to the instructions of the manufacturer (Upstate Biotechnology, Lake Placid, NY). Active MAP3Ks were diluted in an Mg/ATP mixture and recombinant inactive MEK was added and in vitro kinase assays were performed at 30°C for 20 minutes. Samples were dissolved in sample buffer, incubated at 95°C for 5 minutes, and analyzed on Western blot using an anti-phospho-MEK$^{Ser218/222}$/MEK2$^{Ser222/226}$ Antibody.

Kinase profiling

Peptide arrays with 1176 different kinase pseudo-substrates were incubated with active c-Raf and MAP3K8 incubation mix (end concentration of 5 μg/ml active MAP3K8 kinase and 2 μg/ml of active c-Raf, 8% glycerol, 0.5 mM ATP, 10 mM MgCl$_2$, 0.05% v/v Brij-35, 25 μg/ml BSA (Bovine Serum Albumin)) and 30 μCi ^{33}P-γ-ATP, at 37°C for 60 minutes in a humidified oven. The slide was then washed twice with PBS (+0.1% Triton X-100), 2M NaCl and demineralised H$_2$O and dried with N$_2$ gas. Only active c-Raf was incubated with the peptide array consisting of 1024 pseudo-substrates and the same washing steps were carried out.

Peptide Array analysis

After drying, the glass slides were exposed to a phosphor imager plate for 72 hours. Acquisition of the peptide array was performed using a phosphor-imager (Storm^{TM}, Amersham Biosciences, Uppsala, Sweden). The level of incorporated radioactivity, which corresponds to the phosphorylation status, was quantified by Scanalyze. (http://rana.lbl.gov/EisenSoftware.htm) and exported to a spreadsheet program (Microsoft Excel 2002, Microsoft, Redmond, WA, USA). The relative contribution of each individual amino acid at each individual position was calculated and corrected for the relative abundance of that amino acid at that position relative to the central serine, threonine or tyrosine residues and the respective consensus sequences were generated.

Difference between the two array designs

In order to determine how different the two arrays were from each other, we created a single sequence of all the peptide substrates on the 1024 array, separating each by ten gaps (the letter 'X' was used to create gaps). A set of hundred random peptides from the 1176 array design were generated using Microsoft Excel. (Microsoft Excel 2002, Microsoft, Redmond, WA) and each of these peptides were aligned individually against the single sequence obtained from the 1024 array design using MultAlin [37] with their default parameters. The number of identical amino acids were calculated for each of these hundred peptides and averaged to obtain an approximate estimation of similarity between these two array designs.

Results

### Enzymatic characteristics of peptide array phosphorylation by purified enzymes

We set out to evaluate the usefulness of peptide arrays for deriving consensus substrate sequences for kinases. To extract useful information from the phosphorylation of peptide arrays by kinases, it is important to ensure that such phosphorylation conforms to the Michaelis-Menten laws of enzyme kinetics. Hence, we decided to perform a series of initial experiments using active MAP3K8 on trial arrays consisting of 192 peptide substrates investigating array phosphorylation in the temporal domain. Figures 1A and IC show that a subset of substrate peptides displays increased phosphorylation at 1, 5, 10, 20, 30, 45, 60 & 120 minute time points, when incubated with 5 µg/ml of MAP3K8. A steady increase in phosphorylation intensities is seen till it reaches a steady state at the 60 minute time point, indicating that the peptide substrate levels only become a limiting factor after a 1 hour treatment. Figure 1B depicts a plot of the time coefficient deduced from the temporal results for peptide phosphorylation. The results suggest that a minimum of 30 minutes is required for efficient phosphorylation. Interestingly, the majority of substrates did not appear to be capable of undergoing phosphorylation by MAP3K8 at all, thus the MAP3K8 enzyme is not capable of catalyzing the phosphorylation of any given peptide and phosphorylation by this enzyme appears to exhibit qualitative characteristics: even prolonged incubation times do not yield detectable phosphorylation of unfavourable peptides.

A

B.

C.

1 min   30 mins   120 mins   Array layout
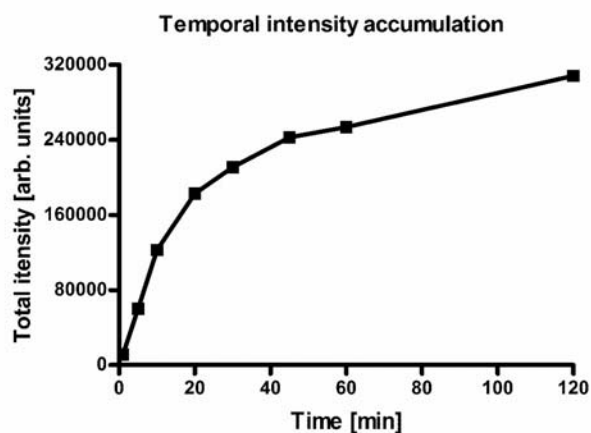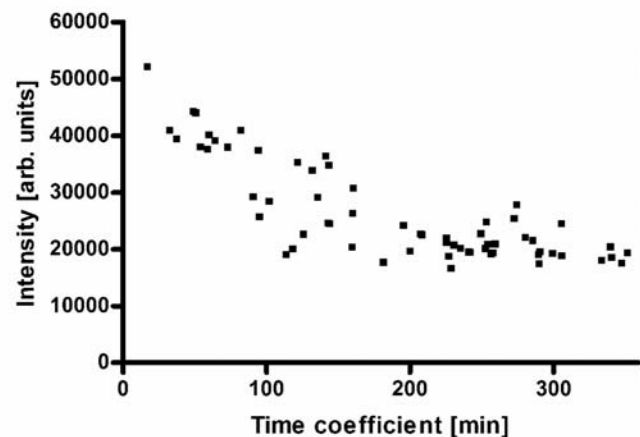
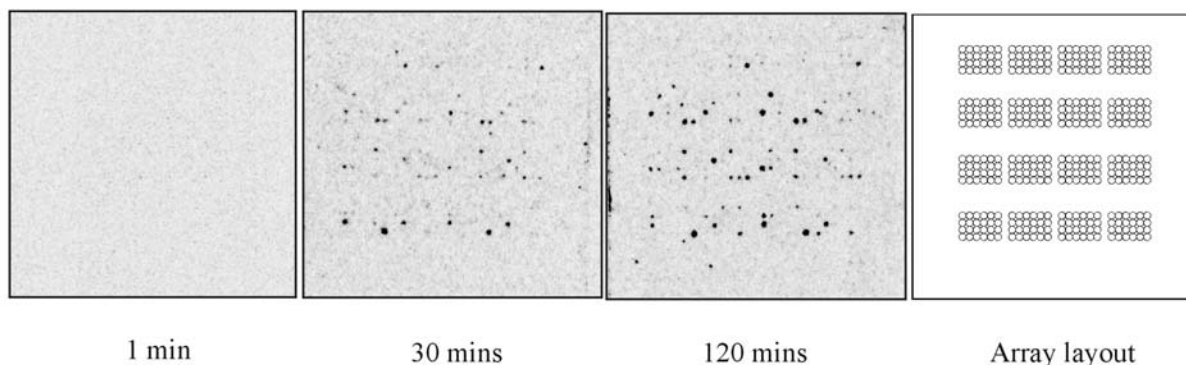**Figure 1. Determination of Enzyme Kinetics (MAP3K8).** A. Plot showing the phosphorylation of various substrates at 1, 5, 10, 20, 30, 60 & 120 min. time points. B. Time coefficient plotted against intensity. C. Images of trial arrays at 1 minute, 30 minute and 120 minute time-points showing the steady increase in phosphorylation level. An overlay for array orientation is also shown.
doi:10.1371/journal.pone.0006440.g001

Generation of a putative c-Raf substrate consensus sequence

The capacity of c-Raf for *in vitro* phosphorylation studies was examined by incubating it with MEK (Mitogen Activated Protein Kinase Kinase), a well established substrate. As evident from figure 2A, our c-Raf preparation was highly active on MEK and we decided to test its ability to phosphorylate peptides immobilized in an array format containing 1176 phosphobase-derived peptides (see materials and methods). A one hour incubation with c-Raf resulted in extensive peptide phosphorylation, with different peptides incorporating wildly different amounts of ^{33}P, demonstrating that peptide sequences confer specificity to c-Raf-dependent phosphorylation (figure 2B). Subsequent analysis was performed to see whether the primary sequence of the peptides employed revealed information as to the amino acid preferences of this enzyme for substrate phosphorylation.

It is possible that a peptide could be phosphorylated at more than one residue, which would mean that a peptide that, for instance, is phosphorylated at two serine's adjacent to each other could result in a higher intensity than a peptide phosphorylated on one serine and this would mask that peptide which could have been left out of the analysis. Hence, only those peptides which had a single phosphorylable residue were considered, i.e. only those peptides which had a single serine, threonine or tyrosine residue at the central position. Of the 1176 peptides, 353 peptides which had a single serine, threonine or tyrosine residue were selected. (Supplementary data S1). These peptides were then aligned manually relative to the centrally fixed serine, threonine or tyrosine residue and ranked on the mean intensity of the duplicates for each spot. For deriving the consensus sequence using arrays with 1176 substrates, we considered only positions −3, −2, −1, 0, 1, 2 & 3 because not all peptides were 9 amino acids in length and also while aligning them based on a fixed central phosphorylation site, we did not have an equal distribution of amino acids at the −4 and +4 positions.

Furthermore, we have only selected peptides with cut-off intensities within 50% of the peptide with the maximum intensity (peptide LRRASLRG with intensity of 37482.5 arbitrary units) and the relative contribution of each individual amino acid at each
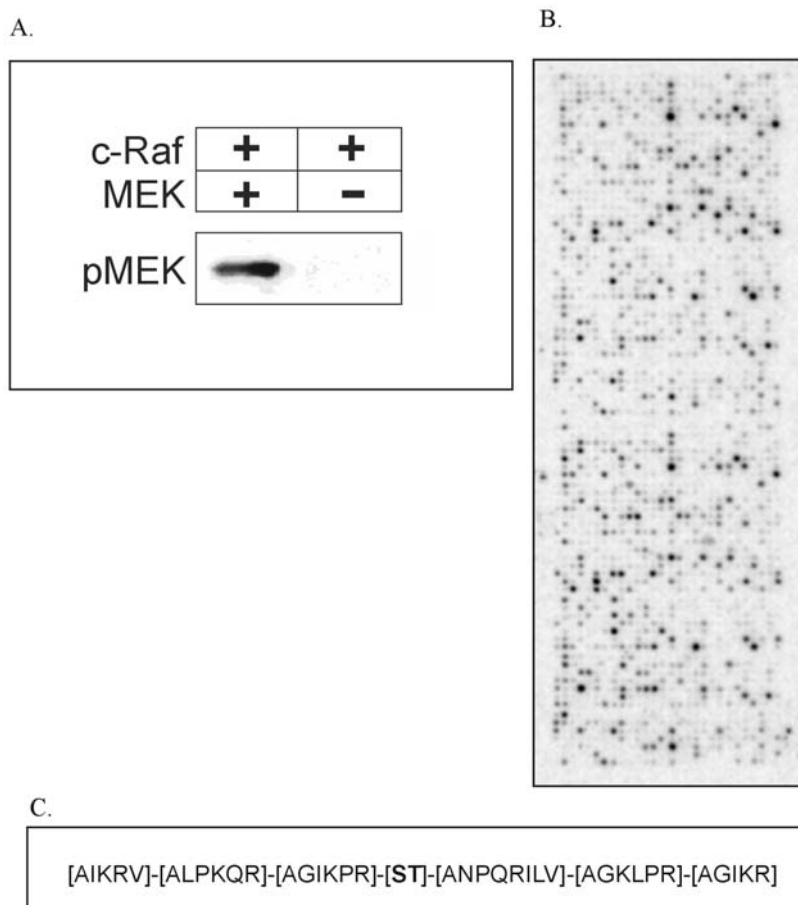
Figure 2. Analysis of phosphorylation of 1176 peptide array by c-Raf. A. *In vitro* phosphorylation of MEK by c-Raf. The capacity of purified c-Raf for in vitro phosphorylation studies was examined by incubating purified c-Raf with MEK and detected using MEK[Ser218/222]/MEK2[Ser222/226] antibodies. B. c-Raf phosphorylation of 1176 peptide array. Phosphorylation of the 1176 peptide array, spotted in duplicate, on incubation with c-Raf and ^{33}P-γ-ATP for one hour shows differential phosphorylation of the various substrate peptides demonstrating that peptide sequences confer specificity to c-Raf-dependent phosphorylation. Further analysis was carried out to determine whether the primary sequence of the peptides employed revealed information as to the amino acid preferences of this enzyme for substrate phosphorylation. C. Consensus sequence of c-Raf substrates using 1176 array design. Consensus sequence determined for c-Raf substrates on analysis of peptide array data shows a strong selection for both hydrophobic and basic residues at the −3 position. A strong preference for leucine is seen at the −2 position. Proline and arginine are strongly preferred at the −1 position. An examination of the amino acid preferences C-terminal to the fixed phosphorylated residue reveals a bias towards asparagine compared to other residues at the +1 position. Also, acyclic and hydrophobic amino acids are seen at the +1 position with no preference for any particular group of amino acid at the +2 position. The +3 position shows a strong preference for basic residues.
doi:10.1371/journal.pone.0006440.g002

individual position was calculated and corrected for the relative abundance of that amino acid at that position relative to the central serine, threonine or tyrosine. The resulting putative c-Raf consensus sequence is shown in figure 2C, whereas table 1 lists the detailed results of the contribution of each amino acid at each position. For c-Raf, a strong selection for both hydrophobic and basic residues is seen at position −3, namely isoleucine and lysine. A strong preference for leucine is seen at position −2 with some preference for proline and arginine. There seems to be a strong selection for proline and arginine at the −1 position. An examination of the amino acid preferences C-terminal to the fixed phosphorylated residue reveals a bias towards asparagine compared to other residues at the +1 position. Also, acyclic and hydrophobic amino acids are seen at the +1 position. There is no preference for any particular group of amino acids at the +2 position. The +3 position also shows a strong preference for basic residues.

Verification of the consensus sequence on a different array

If analysis of the contribution of each amino acid at each position in peptide array phosphorylation patterns yields meaningful results, it should follow that the analysis of phosphorylation of an array containing totally different substrate peptides, which on determination showed only five amino acids to be common, on average, between the two designs, should give a similar result. To test this hypothesis, we analyzed c-Raf-dependent phosphorylation of an array: consisting of 1024 peptides (figure 3A). Similarly, for the 1024 peptide array, peptides with single phosphorylation sites were selected for further analysis. (Supplementary data S2). The resulting putative c-Raf consensus sequence is shown in figure 3B, whereas table 2 lists the detailed results of the contribution of each amino acid at each position. There seems to be a strong preference

for arginine at −1, −4 and the −5 position while the −2 position shows a strong preference for a hydrophobic residue and no distinctive preference is seen at the −3 position. Analysis of the C-terminal position relative to the centrally fixed phosphorylated residue shows a very high preference for methionine besides an equal preference for other basic and hydrophobic amino acids at the +1 position. Arginine is preferred at the +2 and +4 positions while methionine and proline along with arginine are preferred at the +3 position. Hydrophobic residues are preferred over basic residues at +5. Thus, totally different array designs yield similar c-Raf substrate consensus sequences, suggesting that this type of analysis is a valid tool for deducing kinase substrate preferences.

c-Raf and MAP3K8 kinases are highly homologous but have substantially different substrate preferences

Subsequently, we addressed the question whether the various MAP3Ks, which share substantial sequence homology (figure 4A) in their kinase domain, have identical or different substrate specificities. To this end, the 1176 array was incubated with active MAP3K8. As evident from figure 4B, again specific incorporation of radioactivity into different peptides was observed. Figure 4C shows a correlation plot between substrate phosphorylation of c-Raf and MAP3K8, which indicates that despite the highly similar primary sequences both enzymes have different substrate preferences.

Analysis of MAP3K8 substrate preferences

Subsequent analysis, however, of the importance of the individual amino acids relative to the central residue shows that the substrate preference of both kinases also displays substantial similarities, with preferences for arginine, isoleucine, lysine and alanine at −3, −2, −1 and +1 positions which are also preferred by c-Raf. However, a major difference between the two is the

Table 1. Relative weights of amino acids for c-Raf using 1176 array design.

Amino Acid	Position								
	−4	−3	−2	−1	0	1	2	3	4
A	1.09	0.75	0.75	0.87	*	1.60	1.13	1.30	0.00
D	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00
E	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00
F	3.69	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00
G	0.00	0.00	0.00	1.66	*	0.00	1.07	1.40	0.00
H	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00
I	0.00	1.73	0.00	0.32	*	1.38	0.00	1.78	0.00
K	0.57	1.66	0.88	0.83	*	0.00	0.39	1.42	1.79
L	0.69	0.00	2.27	0.00	*	1.62	1.66	0.00	1.13
M	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00
N	0.00	0.00	0.00	0.00	*	2.55	0.00	0.00	0.00
P	0.00	0.00	1.36	2.03	*	1.54	0.94	0.00	0.00
Q	0.00	0.00	0.73	0.00	*	0.91	0.00	0.00	0.00
R	2.11	1.10	1.38	1.85	*	0.96	1.75	2.05	2.60
S	*	*	*	*	0.91	*	*	*	*
T	*	*	*	*	1.58	*	*	*	*
V	2.35	1.39	0.00	0.00	*	1.17	0.00	0.00	1.03
W	*	0.00	0.00	0.00	*	0.00	0.00	*	*
Y	*	*	*	*	0.36	*	*	*	*

[*]replaces all divided/0 values.
doi:10.1371/journal.pone.0006440.t001

A.
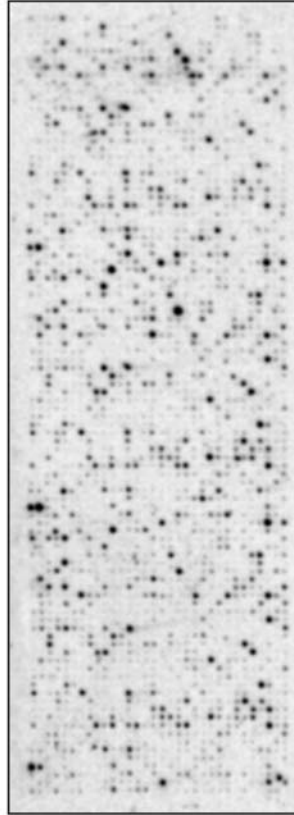
B.

[GKLMR]-[AGKLR]-[AKLQR]-[AGIKLPQR]-[AGKLPR]-[**ST**]-[AGLMPR]-[AKPR]-[DEKLMPR]-[ADLNRV]-[FKLQR]

Figure 3. Analysis of phosphorylation of 1024 peptide array by c-Raf. A. c-Raf phosphorylation of 1024 peptide array. Phosphorylation of the custom made 1024 peptide array, spotted in triplicate, on incubation with c-Raf and ^{33}P-γ-ATP for one hour shows differential phosphorylation of the various substrate peptides. B. Consensus sequence of c-Raf substrates using 1024 array design. Consensus sequence obtained from the 1024 peptide array shows a strong preference for arginine at −1,−4 and the −5 position while the −2 position shows a strong preference for a hydrophobic residue and no distinctive preference is seen at the −3 position. Analysis of the C-terminal position relative to the centrally fixed phosphorylated residue shows a very high preference for methionine besides an equal preference for other basic and hydrophobic amino acids at the +1 position. Arginine is preferred at the +2 and +4 positions while methionine and proline along with arginine are preferred at +3 positions. Hydrophobic residues are preferred over basic residues at +5.
doi:10.1371/journal.pone.0006440.g003

strong preference for phenylalanine at positions −1, −2, −3, +1 in MAP3K8. Another important difference is the absence of proline at the −2 and +1 positions in MAP3K8. (Figure 4D and Table 3).

Discussion

The predominance of phosphorylation as a regulator of cellular metabolism makes it of utmost importance to know kinase substrates for proper understanding of cellular physiology. Unfortunately, our understanding of kinase action does not yet permit the determination of kinase substrates based on the primary sequence of proteins. Indeed, if kinases with highly similar amino acid composition have similar or distinct substrate preferences remain unknown. Hence, empirical methods for determining kinase substrate sequences remain essential. We describe here a novel methodology for predicting kinase substrates, which makes use of a library of peptides, known to serve as phosphorylation motifs to determine a kinase substrate consensus phosphorylation

sequence and employ this methodology for comparing substrates for c-Raf and MAP3K8 enzymatic activity respectively. c-Raf and MAP3K8 are two serine/threonine kinases which are associated with cellular transformation, but which are suggested to have divergent functions in cellular physiology despite their high sequence homology. Employing peptide arrays we derived consensus sequences for substrate phosphorylation based on the relative importance of multiple amino acids (except serine, threonine and tyrosine) carried out at each position and this sequence could then be used to search databases and predict possible substrates. We considered only peptides which had a single phosphorylation site as one of the concerns we had was that some peptides could also be phosphorylated on more than one spot and would thus add to the intensity generated from that peptide. However, as suggested by Diks et al [16], it is still not known whether two phosphorylation sites on a peptide are phosphorylated simultaneously by the kinase.

Confidence in our results was bolstered by the observation that two peptide arrays displaying different peptides yielded similar

Table 2. Relative weights of amino acids for c-Raf using 1024 array design.

Amino Acid	Position										
	−5	−4	−3	−2	−1	0	1	2	3	4	5
A	0.00	1.17	1.58	1.79	2.28	*	1.30	2.36	0.00	1.09	0.00
D	0.00	0.00	0.00	0.00	0.00	*	0.00	0.00	1.72	1.24	0.00
E	0.00	0.00	0.00	0.00	0.00	*	0.00	0.00	0.79	0.00	0.00
F	0.00	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00	3.17
G	1.56	1.78	0.00	1.45	2.02	*	1.38	0.00	0.00	0.00	0.00
H	0.00	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00	0.00
I	0.00	0.00	0.00	2.54	0.00	*	0.00	0.00	0.00	0.00	0.00
K	2.30	1.86	1.92	1.45	0.84	*	0.00	1.17	1.07	0.00	0.96
L	0.70	0.64	1.27	1.60	0.68	*	1.68	0.00	0.74	1.16	2.20
M	1.60	0.00	*	0.00	0.00	*	3.12	0.00	2.36	0.00	0.00
N	0.00	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	1.51	0.00
P	0.00	0.00	0.00	0.97	1.33	*	1.09	1.45	2.12	0.00	0.00
Q	0.00	0.00	1.47	1.68	0.00	*	0.00	0.00	0.00	0.00	1.14
R	2.78	3.34	1.32	0.41	2.38	*	2.26	2.83	2.08	3.11	1.64
S	*	*	*	*	*	0.99	*	*	*	*	*
T	*	*	*	*	*	1.89	*	*	*	*	*
V	0.00	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	1.15	0.00
W	*	*	*	0.00	*	*	0.00	0.00	0.00	*	0.00
Y	*	*	*	*	*	0.00	*	*	*	*	*

*replaces all divided/0 values.

doi:10.1371/journal.pone.0006440.t002

results. On comparing the consensus sequence derived from the 1176 array design:

[AIKRV]-[ALPKQR]-[AGIKPR]-[**ST**]-[ANPQRILV]-[AGK-LPR]-[AGIKR] with that derived from the 1024 array design:

[GKLMR]-[AGKLR]-[AKLQR]-[AGIKLPQR]-[AGKLPR]-[**ST**]-[AGLMPR]-[AKPR]-[DEKLMPR]-[ADLNRV]-[FKLQR] one can see that practically the same group of amino acids are seen at positions −3 to −1, except at positions +1 and +3 with methionine appearing instead of asparagine and glycine, and methionine in place of glycine respectively. Surprisingly, with the 1176 array, two peptide motifs with a central tyrosine residue are among the substrates favoured by c-Raf, showing that serine/threonine restriction of this kinase is not necessarily absolute. Maybe this observation has physiological relevance. Phosphorylation of Y_{340} in c-Raf itself is important for association to MEK, it's most important substrate, thus auto-phosphorylation of c-Raf at this residue may contribute to its physiological function, but obviously further studies are essential for answering this question. We feel, however, that weighing individual amino acids in their contribution to overall peptide array phosphorylation seems a valid tool for determining consensus sequences.
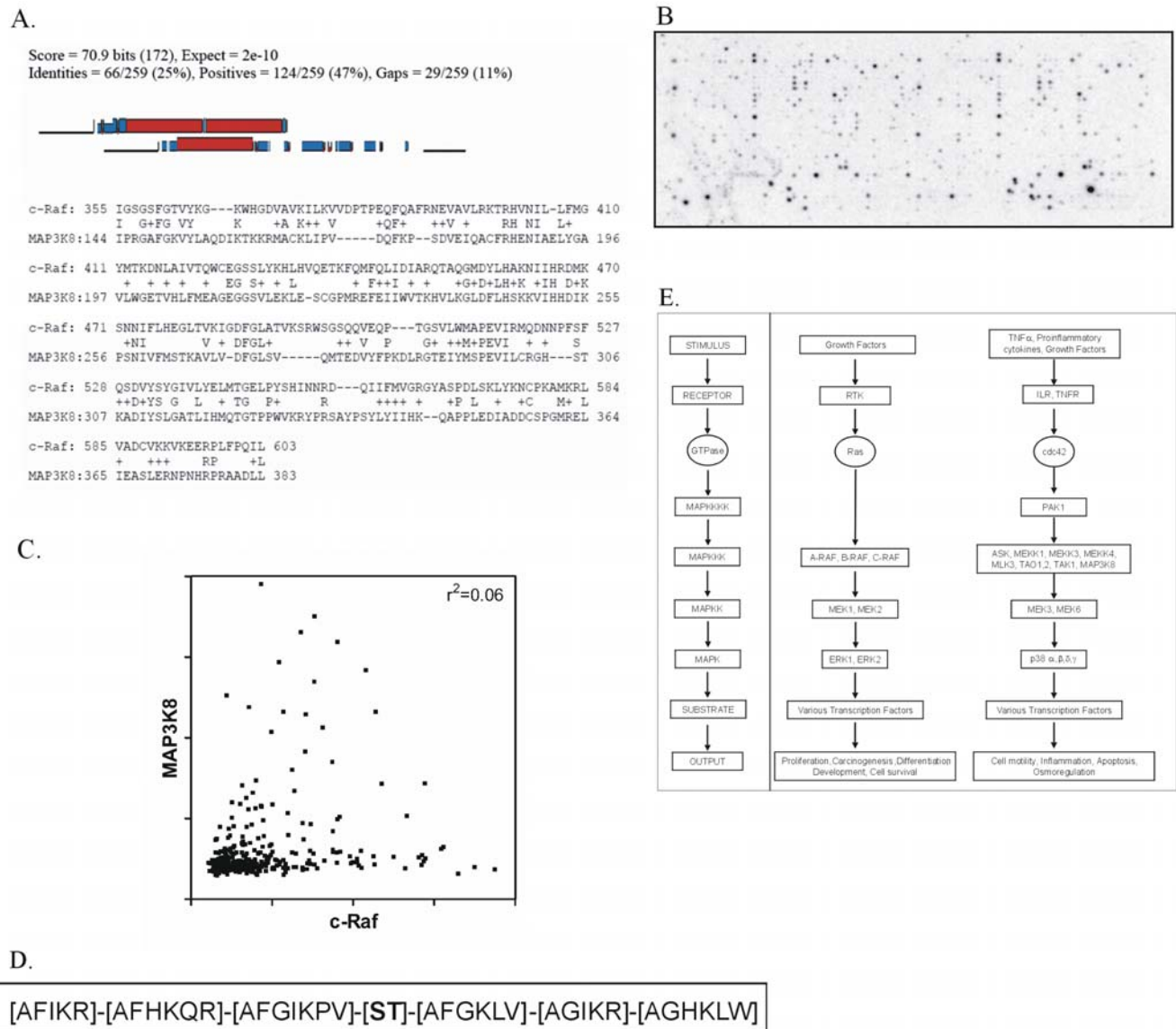
Comparison of c-Raf with MAP3K8 ([AFIKR]-[AFHKQR]-[AFGIKPV]-[**ST**]-[AFGKLV]-[AGIKR]-[AGHKLW]) shows the same set of amino acids dominating except for the strong presence of phenylalanine at all 3 positions N terminal and the +1 position C terminal to the centrally fixed serine/threonine residue. Another difference observed is the dominance of lysine over arginine at all positions where basic residues are seen. Although these two kinases show very similar consensus/scaffold sequences, they seem to phosphorylate completely different sets of substrates as seen in figure 4C, showing that the few amino acids in the consensus sequences which differ between the two kinases are indeed important in defining the substrate profile for that kinase. Thus,

our results show that the similarity in the primary sequences of both kinases is reflected in a chemically similar substrate preference, but the gradual differences that remain, when combined are substantial enough to ensure substrate specificity for the kinases in cellular physiology. In close agreement, divergent downstream targets in living cells for these two kinases have been described in literature e.g. MEK for c-Raf [38–40] and NIK for MAP3K8 [41].

c-Raf seems to have a strong preference for basic residues at the −3 position and hydrophobic residues at the +1 position relative the phosphorylated serine/threonine residue. The Protein kinase C (PKC), AKT kinase (Protein Kinase B), mammalian AMP-activated protein kinase, SNF1 (sucrose non-fermenting kinase 1), calcium/calmodulin-dependant kinase, phosphorylase kinase have similar preferences for basic residues at the −3 position and hydrophobic residues at the +1 position [42–46]. As these kinases are phylogenetically considered to be close to c-Raf it seems that, a common evolutionary origin of kinases has consequences for substrate specificity.

With the consensus sequences determined, a broad choice of possible substrates remains for both c-Raf as well as MAP3K8 in the human genome. To test this, we performed kinase restricted pattern searches using Scanprosite [47], employing patterns consisting of only the most weighted amino acids from tables 1 and 3 for c-Raf, namely, [IKRV]-[LPR]-[GPR]-[ST]-[AILNPV]-[AGLR]-[AGIKR] and MAP3K8, namely, [FIK]-[AFHKQ]-[AFIKV]-[ST]-[AFGV]-[IKR]-[AHKLW] and found 33 and 6 hits respectively. Many of these proteins are not likely to represent true intracellular substrates for these kinases. However, we did find some interesting candidates such as Ephrin type-B receptor 4, which is known to be phosphorylated but the kinase remains unknown and could thus be an interesting substrate for c-Raf. We also found sites within the Mast/stem cell growth factor receptor

A.

Score = 70.9 bits (172), Expect = 2e-10
Identities = 66/259 (25%), Positives = 124/259 (47%), Gaps = 29/259 (11%)

```
c-Raf:   355  IGSGSFGTVYKG---KWHGDVAVKILKVVDPTPEQFQAFRNEVAVLRKTRHVNIL-LFMG  410
              I  G+FG VY     K   +A K++ V    +QF+  ++V +   RH NI  L+
MAP3K8:  144  IPRGAFGKVYLAQDIKTKKRMACKLIPV-----DQFKP--SDVEIQACFRHENIAELYGA  196

c-Raf:   411  YMTKDNLAIVTQWCEGSSLYKHLHVQETKFQMFQLIDIARQTAQGMDYLHAKNIIHRDMK  470
              +  + + + +  EG S+ + L      + F++I + +   +G+D+LH+K +IH D+K
MAP3K8:  197  VLWGETVHLFMEAGEGGSVLEKLE-SCGPMREFEIIWVTKHVLKGLDFLHSKKVIHHDIK  255

c-Raf:   471  SNNIFLHEGLTVKIGDFGLATVKSRWSGSQQVEQP---TGSVLWMAPEVIRMQDNNPFSF  527
              +NI      V + DFGL+      ++ V P    G+ ++M+PEVI  + +     S
MAP3K8:  256  PSNIVFMSTKAVLV-DFGLSV----QMTEDVYFPKDLRGTEIYMSPEVILCRGH---ST  306

c-Raf:   528  QSDVYSYGIVLYELMTGELPYSHINNRD---QIIFMVGRGYASPDLSKLYKNCPKAMKRL  584
              ++D+YS G  L + TG P+      R     ++++ +   +P L + +C  M+ L
MAP3K8:  307  KADIYSLGATLIHMQTGTPPWVKRYPRSAYPSYLYIIHK--QAPPLEDIADDCSPGMREL  364

c-Raf:   585  VADCVKKVKEERPLFPQIL  603
              +   +++   RP   +L
MAP3K8:  365  IEASLERNPNHRPRAADLL  383
```

C.

B.

E.

D.

[AFIKR]-[AFHKQR]-[AFGIKPV]-[**ST**]-[AFGKLV]-[AGIKR]-[AGHKLW]

Figure 4. Analysis of phosphorylation of 1176 peptide array by MAP3K8 and comparison against c-raf. A. Alignment of c-raf against MAP3K8. Alignment of c-Raf with MAP3K8 using BLAST shows considerable homology between their kinase domains. B. Phosphorylation by MAP3K8 of 1176 peptide array. Specific incorporation of radioactivity into different peptides was observed after incubating the 1176 peptide array with MAP3K8 and ^{33}P-γ-ATP. C. Comparison of c-Raf and MAP3K8 peptide phosphorylation. A correlation plot between substrate phosphorylation of c-Raf and MAP3K8 (using the 1176 peptide array design), which indicates that despite the highly similar primary sequences both enzymes have different substrate preferences. D. Consensus sequence of MAP3K8 substrates using 1176 array design. Consensus sequence for MAP3K8 shows preferences for phenylalanine at −3, −2, −1 and +1 positions and lysine at all six positions, which is similar to c-raf that has a preference for arginine instead of lysine at all positions as shown in figure 2C. E. Flow chart depicting the MAPK cascade and the roles of various kinases including c-Raf and MAP3K8 within these cascades.
doi:10.1371/journal.pone.0006440.g004

and ribosomal protein S6 kinase alpha 1 which are known sites of phosphorylation for PKC and ERK and could well be putative c-Raf substrates. Confidence in our method was further bolstered when the most well established substrate for c-Raf, MEK1 phosphorylated at Threonine 292 was found back with the pattern search. For the six hits found with the MAP3K8 substrate pattern, we found two interesting candidates, both similar to each other, namely, Myristoylated alanine-rich C-kinase substrate (MARCKS) and the other named MARCKS-related protein. Both these substrates were phosphorylated on serines by various kinases including PKCβ and Serine Threonine Protein Kinase N. This is an interesting result because MARCKS is known to be phosphor-

ylated by a MAPK family member on the matched site, although the exact upstream kinase remains unknown [48]. It would prove interesting to see if MAP3K8 could directly phosphorylate this protein *in vivo*. Moreover, these substrates also had phenylalanine at the +1 position C terminal to the centrally fixed serine residue (FKKS**F**KL for MARCKS and KKFS**F**KK for MARCKS-related protein), which was one of the important differences seen between the c-Raf and MAP3K8 consensus sequences as described above. It is also interesting to note that BLAST [49] searches with the consensus substrates often yield many matches in which the central serine/threonine is replaced by a phosphorylation incapable amino acid, possibly a reflection of evolutionary pressure to avoid non-

Table 3. Relative weights of amino acids for MAP3K8 using 1176 array design.

Amino Acid	Position								
	−4	−3	−2	−1	0	1	2	3	4
A	2.37	0.76	1.40	1.57	*	2.52	0.80	1.50	0.00
D	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00
E	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00
F	0.00	3.21	2.81	2.13	*	3.36	0.00	0.00	0.00
G	0.00	0.00	0.00	0.57	*	1.12	0.62	0.97	0.00
H	0.00	0.00	2.42	0.00	*	0.00	0.00	4.87	2.18
I	0.00	2.25	0.00	1.54	*	0.00	1.31	0.00	0.00
K	2.65	2.10	2.15	2.31	*	1.01	2.32	2.27	2.45
L	0.58	0.00	0.00	0.00	*	0.58	0.00	1.80	0.70
M	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	2.86
N	0.00	0.00	0.00	0.00	*	0.00	0.00	0.00	0.00
P	0.00	0.00	0.00	1.03	*	0.00	0.00	0.00	1.84
Q	0.00	0.00	1.88	0.00	*	0.00	0.00	0.00	0.00
R	1.40	0.40	0.38	0.00	*	0.00	1.67	0.00	1.10
S	*	*	*	*	1.10	*	*	*	*
T	*	*	*	*	1.14	*	*	*	*
V	0.00	0.00	0.00	1.22	*	1.47	0.00	0.00	0.00
W	*	0.00	0.00	0.00	*	0.00	0.00	5.72	*
Y	*	*	*	*	0.00	*	*	*	*

*replaces all divided/0 values.
doi:10.1371/journal.pone.0006440.t003

regulatory phosphorylation events. In addition, many other factors (such as the presence of scaffolding proteins, adaptor proteins or intracellular localization) will *in vivo* influence the extent to which a motif is really subject to phosphorylation by c-Raf or MAP3K8. In this sense, the consensus sequence determination is more suitable for determining the possible upstream regulatory kinases when an amino acid is known to undergo phosphorylation, rather than providing insights in the downstream targets of a given kinase. Nevertheless, analysis of possible downstream targets may be useful for hypothesis generation. In conclusion, we have developed a new method to generate consensus sequences using peptide arrays based on the relative intensity of amino acids seen at all the positions N and C terminal to the centrally fixed serine/threonine/tyrosine residue. The general utility of this method would be identification of new substrates and it also has an edge over the oriented and combinatorial peptide microarrays as it has peptides with known phosphorylation sites for different kinases.

References

1. Graves JD, Krebs EG (1999) Protein phosphorylation and signal transduction. Pharmacol Ther 82: 111–121.
2. Blume-Jensen P, Hunter T (2001) Oncogenic kinase signalling. Nature 411: 355–365.
3. Chang L, Karin M (2001) Mammalian MAP kinase signalling cascades. Nature 410: 37–40.
4. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298: 1912–1934.
5. Kreegipuu A, Blom N, Brunak S, Jarv J (1998) Statistical analysis of protein kinase specificity determinants. FEBS Lett 430: 45–50.
6. Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294: 1351–1362.
7. Brinkworth RI, Breinl RA, Kobe B (2003) Structural basis and prediction of substrate specificity in protein serine/threonine kinases. Proc Natl Acad Sci U S A 100: 74–79.
8. Songyang Z, Blechner S, Hoagland N, Hoekstra MF, Piwnica-Worms H, et al. (1994) Use of an oriented peptide library to determine the optimal substrates of protein kinases. Curr Biol 4: 973–982.
9. Lueking A, Horn M, Eickhoff H, Bussow K, Lehrach H, et al. (1999) Protein microarrays for gene expression and antibody screening. Anal Biochem 270: 103–111.
10. Arenkov P, Kukhtin A, Gemmell A, Voloshchuk S, Chupeeva V, et al. (2000) Protein microchips: use for immunoassay and enzymatic reactions. Anal Biochem 278: 123–131.
11. MacBeath G, Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. Science 289: 1760–1763.
12. Zhu H, Snyder M (2001) Protein arrays and microarrays. Curr Opin Chem Biol 5: 40–45.
13. Wenschuh H, Volkmer-Engert R, Schmidt M, Schulz M, Schneider-Mergener J, et al. (2000) Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides. Biopolymers 55: 188–206.

Supporting Information

Data S1 Peptides with single phosphorylatable residues considered for 1176 array analysis
Found at: doi:10.1371/journal.pone.0006440.s001 (0.34 MB DOC)

Data S2 Peptides on 1024 array with single phosphorylation sites that were used for analysis
Found at: doi:10.1371/journal.pone.0006440.s002 (0.13 MB DOC)

Author Contributions

Conceived and designed the experiments: KP DWH JJ MPP. Performed the experiments: KP. Analyzed the data: KP SHD JT MPP. Contributed reagents/materials/analysis tools: SHD JT AV ML JJ AP. Wrote the paper: KP MPP.

14. Falsey JR, Renil M, Park S, Li S, Lam KS (2001) Peptide and small molecule microarray for high throughput cell adhesion and functional assays. Bioconjug Chem 12: 346–353.

15. Reineke U, Volkmer-Engert R, Schneider-Mergener J (2001) Applications of peptide arrays prepared by the SPOT-technology. Curr Opin Biotechnol 12: 59–64.

16. Diks SH, Kok K, O'Toole T, Hommes DW, van DP, et al. (2004) Kinome profiling for studying lipopolysaccharide signal transduction in human peripheral blood mononuclear cells. J Biol Chem 279: 49206–49213.

17. Diella F, Cameron S, Gemund C, Linding R, Via A, et al. (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. BMC Bioinformatics 5: 79.

18. Lowenberg M, Tuynman J, Bilderbeek J, Gaber T, Buttgereit F, et al. (2005) Rapid immunosuppressive effects of glucocorticoids mediated through Lck and Fyn. Blood 106: 1703–1710.

19. Diks SH, Parikh K, van der SM, Joore J, Ritsema T, et al. (2007) Evidence for a minimal eukaryotic phosphoproteome? PLoS One 2: e777.

20. Parikh K, Poppema S, Peppelenbosch MP, Visser L (2009) Extracellular ligation-dependent CD45RB enzymatic activity negatively regulates lipid raft signal transduction. Blood 113: 594–603.

21. Parikh K, Peppelenbosch MP, Ritsema T (2009) Kinome profiling using peptide arrays in eukaryotic cells. Methods Mol Biol 527: 269–80, x.

22. van Baal JW, Diks SH, Wanders RJ, Rygiel AM, Milano F, et al. (2006) Comparison of kinome profiles of Barrett's esophagus with normal squamous esophagus and normal gastric cardia. Cancer Res 66: 11605–11612.

23. Warne PH, Viciana PR, Downward J (1993) Direct interaction of Ras and the amino-terminal region of Raf-1 in vitro. Nature 364: 352–355.

24. Moodie SA, Willumsen BM, Weber MJ, Wolfman A (1993) Complexes of Ras.GTP with Raf-1 and mitogen-activated protein kinase kinase. Science 260: 1658–1661.

25. Koide H, Satoh T, Nakafuku M, Kaziro Y (1993) GTP-dependent association of Raf-1 with Ha-Ras: identification of Raf as a target downstream of Ras in mammalian cells. Proc Natl Acad Sci U S A 90: 8683–8686.

26. Vojtek AB, Hollenberg SM, Cooper JA (1993) Mammalian Ras interacts directly with the serine/threonine kinase Raf. Cell 74: 205–214.

27. Rapp UR, Goldsborough MD, Mark GE, Bonner TI, Groffen J, et al. (1983) Structure and biological activity of v-raf, a unique oncogene transduced by a retrovirus. Proc Natl Acad Sci U S A 80: 4218–4222.

28. Jansen HW, Lurz R, Bister K, Bonner TI, Mark GE, et al. (1984) Homologous cell-derived oncogenes in avian carcinoma virus MH2 and murine sarcoma virus 3611. Nature 307: 281–284.

29. Moelling K, Heimann B, Beimling P, Rapp UR, Sander T (1984) Serine- and threonine-specific protein kinase activities of purified gag-mil and gag-raf proteins. Nature 312: 558–561.

30. Lowenberg M, Verhaar A, van den BB, ten KF, van DS, et al. (2005) Specific inhibition of c-Raf activity by semapimod induces clinical remission in severe Crohn's disease. J Immunol 175: 2293–2300.

31. Ballester A, Velasco A, Tobena R, Alemany S (1998) Cot kinase activates tumor necrosis factor-alpha gene expression in a cyclosporin A-resistant manner. J Biol Chem 273: 14099–14106.

32. Tsatsanis C, Patriotis C, Bear SE, Tsichlis PN (1998) The Tpl-2 protooncoprotein activates the nuclear factor of activated T cells and induces interleukin 2 expression in T cell lines. Proc Natl Acad Sci U S A 95: 3827–3832.

33. Ballester A, Tobena R, Lisbona C, Calvo V, Alemany S (1997) Cot kinase regulation of IL-2 production in Jurkat T cells. J Immunol 159: 1613–1618.

34. Belich MP, Salmeron A, Johnston LH, Ley SC (1999) TPL-2 kinase regulates the proteolysis of the NF-kappaB-inhibitory protein NF-kappaB1 p105. Nature 397: 363–368.

35. Lin X, Cunningham ET Jr, Mu Y, Geleziunas R, Greene WC (1999) The protooncogene Cot kinase participates in CD3/CD28 induction of NF-kappaB acting through the NF-kappaB-inducing kinase and IkappaB kinases. Immunity 10: 271–280.

36. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, et al. (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res 13: 2363–2371.

37. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16: 10881–10890.

38. Kyriakis JM, App H, Zhang XF, Banerjee P, Brautigan DL, et al. (1992) Raf-1 activates MAP kinase-kinase. Nature 358: 417–421.

39. Dent P, Haser W, Haystead TA, Vincent LA, Roberts TM, et al. (1992) Activation of mitogen-activated protein kinase kinase by v-Raf in NIH 3T3 cells and in vitro. Science 257: 1404–1407.

40. Howe LR, Leevers SJ, Gomez N, Nakielny S, Cohen P, et al. (1992) Activation of the MAP kinase pathway by the protein kinase raf. Cell 71: 335–342.

41. Lin X, Mu Y, Cunningham ET Jr, Marcu KB, Geleziunas R, et al. (1998) Molecular determinants of NF-kappaB-inducing kinase action. Mol Cell Biol 18: 5899–5907.

42. Nishikawa K, Toker A, Johannes FJ, Songyang Z, Cantley LC (1997) Determination of the specific substrate sequence motifs of protein kinase C isozymes. J Biol Chem 272: 952–960.

43. Obata T, Yaffe MB, Leparc GG, Piro ET, Maegawa H, et al. (2000) Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. J Biol Chem 275: 36108–36115.

44. Songyang Z, Lu KP, Kwon YT, Tsai LH, Filhol O, et al. (1996) A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. Mol Cell Biol 16: 6486–6493.

45. Hardie DG, Carling D, Carlson M (1998) The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell? Annu Rev Biochem 67: 821–855.

46. Songyang Z, Shoelson SE, McGlade J, Olivier P, Pawson T, et al. (1994) Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. Mol Cell Biol 14: 2777–2785.

47. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, et al. (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform 3: 265–274.

48. Hasegawa H, Nakai M, Tanimukai S, Taniguchi T, Terashima A, et al. (2001) Microglial signaling by amyloid beta protein through mitogen-activated protein kinase mediating phosphorylation of MARCKS. Neuroreport 12: 2567–2571.

49. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.