Positional Scanning Peptide Libraries for Kinase Substrate Specificity Determinations: Straightforward and Reproducible Synthesis Using Pentafluorophenyl Esters

Thomas Ljungdahl,[†] Jenny Veide-Vilg,[‡] Fredrik Wallner,[†] Markus J. Tamás,[‡] and Morten Grøtli*,[†]

Department of Chemistry, Medicinal Chemistry, and Department of Cell & Molecular Biology, University of Gothenburg, Göteborg, Sweden

Received June 1, 2010

An efficient method to synthesize positional scanning synthetic combinatorial libraries (PS-SCLs) for studying the specificity of protein kinases is presented. Isokinetic ratios for pentafluorophenyl esters were determined iteratively using a new approach incorporating high performance liquid chromatography (HPLC) quantification and statistical experimental design. In the development process a large amount of work was put in to find efficient ways of screening for new isokinetic mixtures and to optimize the process of PS-SCL synthesis. The newly developed methods for the screening of isokinetic mixtures could be used for the screening of other interesting mixtures, but more importantly, the isokinetic ratios determined for the preactivated pentafluorophenyl esters were incorporated into a new efficient protocol. This straightforward protocol allows for a convenient synthesis of high quality PS-SCLs regardless of previous experience in solid phase synthesis.

Introduction

Synthetic combinatorial libraries (SCLs) are collections of large numbers of synthetic compounds, in which all possible combinations of the building blocks used are represented.[1] These are used to identify new drugs and to probe biological activity. While early SCLs required extensive deconvolution and additional iterative syntheses to find the combination giving the highest biological activity,[2] positional scanning techniques provide information on the substituent responsible for activity at each position of a compound in a single assay. The so-called positional scanning synthetic combinatorial library (PS-SCL) is a collection of compounds all of which share the same core structure and that are organized in a manner that uses well-defined mixtures to exponentially reduce the number of samples that need to be tested. This type of library was pioneered by Houghten et al. and has given researchers the possibility to efficiently search upward of a trillion compounds in a matter of weeks.[1,3−5] PS-SCLs have found applications in areas such as cancer drug development, vaccine development, and endoprotease studies.[6−13] Recently, a groundbreaking and time saving approach to establishing the optimal motifs for phosphorylation by protein kinases was presented by Hutti et al.[14] They adapted the concept of PS-SCLs to kinase assays using ^{32}P radiolabeled ATP and radiometry for detection and quantification. Determination of which protein kinases are responsible for phosphorylating which protein substrates in the cell is critical for mapping signal transduction cascades that regulate cell growth, division, differentiation, and death. The substrate specificity is partly determined by the catalytic domain of protein kinases, which recognizes phosphorylation motifs consisting of specific short linear sequences of amino acids that flank the residue to be phosphorylated.[15] Hutti et al. used a large library of peptide mixtures where one of the 20 naturally occurring proteogenic amino acids was fixed at each of nine positions surrounding a central serine/threonine residue, with the remaining positions degenerate (Figure 1).

In the degenerated positions, 17 of the 20 amino acids were incorporated competitively in such a way that all residues were equally represented. Cysteine, serine, and threonine were excluded; cysteine to avoid the formation of cystine bridges while serine and threonine were excluded to reduce background noise. This library is now commercially available but is very expensive, and the lack of publications describing its synthesis prevents others from doing their own

Peptide sequences

Y-A-Z-X-X-X-X-S/T-X-X-X-X-A-G-K-K(LC-Biotin)-NH_2
Y-A-X-Z-X-X-X-S/T-X-X-X-X-A-G-K-K(LC-Biotin)-NH_2
Y-A-X-X-Z-X-X-S/T-X-X-X-X-A-G-K-K(LC-Biotin)-NH_2
Y-A-X-X-X-Z-X-S/T-X-X-X-X-A-G-K-K(LC-Biotin)-NH_2
Y-A-X-X-X-X-Z-S/T-X-X-X-X-A-G-K-K(LC-Biotin)-NH_2
Y-A-X-X-X-X-X-S/T-Z-X-X-X-A-G-K-K(LC-Biotin)-NH_2
Y-A-X-X-X-X-X-S/T-X-Z-X-X-A-G-K-K(LC-Biotin)-NH_2
Y-A-X-X-X-X-X-S/T-X-X-Z-X-A-G-K-K(LC-Biotin)-NH_2
Y-A-X-X-X-X-X-S/T-X-X-X-Z-A-G-K-K(LC-Biotin)-NH_2

Figure 1. Peptide library sequences used by Hutti et al. for kinase substrate specificity studies. X = degenerate mixture of the 17 natural amino acids excluding Cys, Thr, and Ser. Z = fixed position varied between the 20 natural amino acids plus phosphothreonine and phosphotyrosine. LC = "long chain" version with an additional amino-hexanoic acid spacer between the biotin and lysine side chain. S/T = equimolar mixture of serine and threonine.[14]

* To whom correspondence should be addressed. E-mail: grotli@chem.gu.se.
[†] Department of Chemistry, Medicinal Chemistry.
[‡] Department of Cell & Molecular Biology.

10.1021/cc100095y
Published on Web 07/07/2010

modifications and expanding this ingenious method to other areas of interest. There are older publications describing similar syntheses, but none of them describe the composition of the all-important "isokinetic mixture" for a serine and threonine free scrambled position.[1,4] The critical information needed to synthesize a tailor-made PS-SCL is the composition of the so-called isokinetic mixture, that is, the ratio of amino acid reagents needed for a close to equimolar incorporation in each of the degenerated positions.[9] We therefore set out to find the isokinetic mixture ratios for preactivated amino acid esters and from this develop a straightforward protocol that would not require an excessive and rigorous control of reaction parameters. The use of preactivated amino acid derivatives reduces the amount of manual labor and improves reproducibility. The project thus entails method development for efficient screening of isokinetic mixtures, iterative determination of isokinetic mixtures, and library synthesis and evaluation.

Results and Discussion

The first task was the identification of suitable active esters and is covered in the first section. Directly following this is a section covering the specific issues and considerations of the iterative screening of isokinetic mixtures, in terms of number of components in the mixtures, scalability, and so forth. In the third section the stability of the pentafluorophenyl esters (Pfp-esters) is studied in detail, and suitable residues are selected based on these results. The development of a peptide model system and high performance liquid chromatography (HPLC) protocols to study and quantify the incorporation of the amino acids in the model peptide is covered in the fourth section. The fifth section deals with issues related to the design of the library, such as sequence orders and tags. The iterative process of finding the isokinetic mixture and the results from each iterative round are described in the sixth section. The details of the final protocol and the synthesis of the library are detailed in the seventh section. The use of this library to screen the specificity of the bovine protein kinase A, PKA, is covered in the eighth section. Comparison of the results from the PKA study served as a quality control as these results could be compared with similar studies in the literature. The work flow of the Results and Discussion section is illustrated in Figure 2.

1. Active Esters. We decided to use the stable, nonhygroscopic, and commercially available pentafluorophenyl esters as the reactive amino acid derivatives in the isokinetic mixture.[16] Using preactivated esters simplify synthesis by virtue of drastically reducing the number of manual steps needed for each coupling.

In the case of in situ activation of Fmoc-protected amino acids with free carboxylic acids, three compounds must be measured and mixed immediately prior to coupling; the amino acid derivatives, an activating agent (e.g., HBTU, TBTU, DCC, DIC, etc.) and, when using the so-called onium coupling reagents,[17] a tertiary amine. This approach, besides being labor intensive when performed on a library scale, suffers from additional drawbacks when it comes to mixture-based synthesis. The problems associated with in situ activation can be illustrated using HBTU as an example.

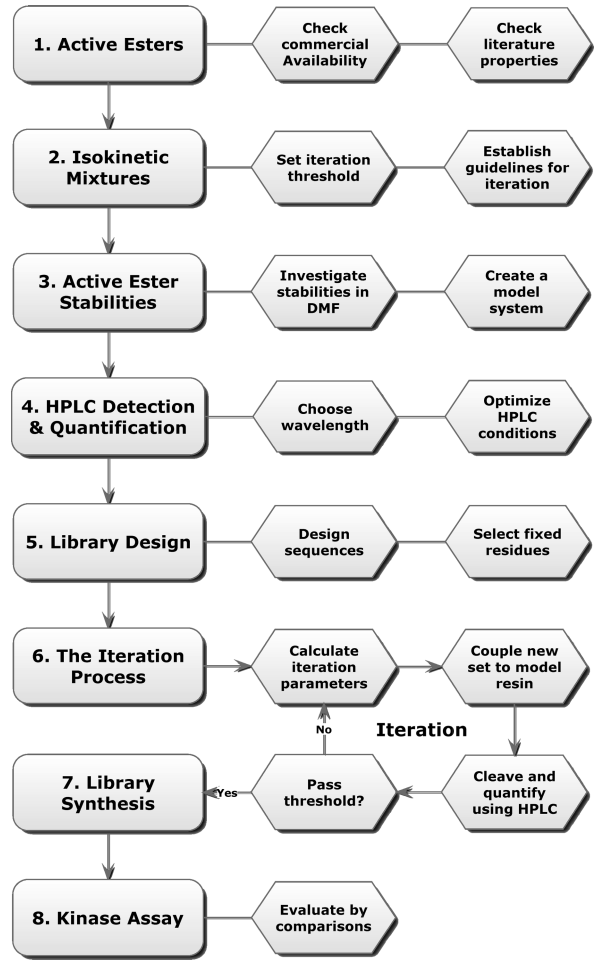

Figure 2. Schematic work flow showing the process from the choice of active esters all the way to the evaluation of the finished library. The numbered rounded blocks represent the different parts of the Results and Discussion section and the hexagonal shapes show different issues, approaches, and considerations in each section.

When using HBTU to couple an isokinetic mixture to the resin it is imperative to use precisely 1 equiv of the coupling reagent as the use of more HBTU than amino acids can cause the reagent to react with the N-terminus of the growing chain, forming an undesirable guanidino derivative.[17] Although not a major problem for standard couplings were an excess reagent is used, this could potentially shift the important resin loading:amino acid ratio and thus have a negative effect in mixture-based synthesis. There are a plethora of other coupling reagents available with different advantages and disadvantages, but the simplicity and high level of control over equivalents offered by the use of Pfp-esters make them the reagents of choice.

2. Isokinetic Mixtures. The isokinetic mixture is a nonequimolar mixture of reactive amino acid derivatives that, upon coupling to a resin-bound peptide, gives an equimolar incorporation of the constituent amino acid derivatives. This approach rests heavily on the findings by Ragnarsson et al.[18,19] and by Ostrech et al.[20] that the constitution of the isokinetic mixture is approximately independent of the resin-bound amino acid. They showed that the rate of coupling was dependent on the amino acid to be coupled, but the relative rate of coupling was approximately independent of the resin-bound amino acid. In earlier determinations of

isokinetic ratios a time-consuming and in some respects inaccurate method of iterating was used to find the correct mixture, that is, amino acid analysis after hydrolysis.[20] This method does not, by definition, distinguish between glutamine-glutamic acid (Gln-Glu) and asparagine-aspartic acid (Asn-Asp) as the amide side chains are hydrolyzed to their carboxylic acid counterpart during analysis.

To reduce the time and effort needed for the screening process we developed a rapid HPLC-based approach. This also has the additional benefit of distinguishing between Gln-Glu and Asn-Asp. A screening method was designed where a mixture of Pfp-amino acid derivatives was coupled to a resin bound dipeptide (Ala-Tyr), and the extent of incorporation of each was quantified by HPLC. Alanine was chosen as it is neither highly hydrophilic nor lipophilic which is of importance for the HPLC separation where the properties of the amino acids coupled onto the pre-synthesized Ala-Tyr resin should dominate. For example a polar amino acid like aspartic acid in the middle position instead of Ala (i.e., an Asp-Tyr resin) would cause all model tripeptides to elute much faster and would thus make separation difficult. The tyrosine residue is necessary for UV-quantification.

Acharya et al. used only binary mixtures in their work on carboxylic acids, and they were able to extend their results to a ternary mixture with reasonable accuracy (they classified a mixture as equimolar if the composition of each compound in the mixture fell within ±10%).[21] For the purpose of establishing the optimal motifs for phosphorylation by protein kinases it should be possible to regard ≤ ± 15% as equimolar. This statement is based on the fact that the commercially available library is reported to have a deviation from equimolarity exceeding ±50%.[14] We therefore started our iteration process using quaternary or quinary mixtures that could be manually adjusted iteratively to approach equimolarity. However, closer to equimolarity than ±20% was difficult to achieve by this approach. Furthermore, to test how far the results could be extrapolated, the ratios obtained after initial iterations were recalculated for a larger, combined, set of amino acids. When this set was run large differences in the product composition could be seen. It was soon clear that the iterations had to be performed on the full set of amino acids and even small variations of one of the components in such a mixture had a profound impact on the ratio of the rest. Interestingly, the largest impact was on the bulky amino acids that were present in largest excess (Val, Ile, His(Trt)).

During the work on tertiary and quinary mixtures we discovered that the active ester derived from arginine (Fmoc-Arg(Pbf)-OPfp) was not sufficiently stable in solution as it had a tendency to cyclize to a δ-lactame, as previously reported by Frank[22] (Scheme 1). The very basic guanidine moiety is not sufficiently deactivated by the electron withdrawing 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl to suppress intramolecular reactions.

3. Active Ester Stabilities. The finding that Fmoc-Arg(Pbf)-OPfp was not as stable as we had anticipated prompted us to further investigate how stable these Pfp-esters actually were in dimethylformamide (DMF) solution. There are examples in the literature of the stability of the Pfp-esters,[23,24] but we decided to check the stability in DMF, both in the time frame of the reaction (∼1 h) and in the time frame of stock solution storage (∼1 week). For the short-term stability two experiments were set up where three equally large aliquots of an isokinetic mixture consisting of four Pfp-esters were coupled to identical amounts of resin, but where two of the aliquots were diluted to half of the concentration of the first. The reactions were allowed to go to completion, and the idea was that the lower concentration should increase the time the active esters were present in solution before coupling. Any degradation in the time frame of the reaction should thus change the ratios of the incorporated amino acids. The results showed that the average variation for a quarternary or quinary mixture of Pfp-esters was ±1% when the same number of equivalents was used regardless of concentration. This shows conclusively that the Pfp-esters are sufficiently stable in the time frame of a coupling reaction. The results from the experiments to check long-term stability in DMF directly by HPLC proved difficult to analyze. When diluting aged 0.15 M stock solutions in DMF to around 1 mM for injection with DMF the diluted sample was degraded by trace impurities (probably amines) in the solvent. The consumption of the active esters stopped when the trace impurities had been consumed, but interpretation was difficult. However, it was possible to identify Pfp-esters that seemed to show a lower degree of stability than the rest. Fmoc-Arg(Pbf)-OPfp was clearly highly unstable. Two other Pfp-esters, Fmoc-Gln-OPfp and Fmoc-Gly-OPfp, seemed to display low stability. However, the trityl protected analogue of glutamine, Fmoc-Gln(Trt)-OPfp, seemed sufficiently stable. To effectively study the stability of the different Pfp-esters and to take intermolecular interactions into account, indirect experiments were performed. In these experiments groups of active esters were mixed and coupled to a small amount of resin. The same stock solution was repeatedly coupled to an identical amount of the model resin, cleaved and analyzed by HPLC. Couplings were made directly, after 1 day, 2 days, 5 days, and after 7 days. Comparison of the percentage area from the chromatograms after cleavage should reveal any instability. The arginine derivative was excluded because of its obvious instability, and Fmoc-Gln-Opfp was replaced by Fmoc-Gln(Trt)-OPfp. A representative example is given in

Scheme 1. Intramolecular Ring Closure of Arginine Competing with Desired Coupling Reactions

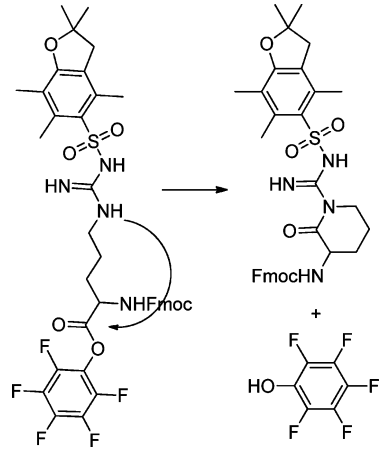

Table 1. Outcome of the Repeated Couplings Using Aged DMF Stock Solutions for the Asp, Lys, Ser, Ala, and Phe Derivatives[a]

days	Asp	Lys	Ser	Ala	Phe
0	18.1	20.3	24.1	20.2	17.4
1	18.3	20.6	23.8	20.0	17.2
2	17.9	20.8	24.1	19.9	17.3
5	17.7	20.9	24.3	19.9	17.2
7	17.7	20.6	24.2	20.5	17.1
coefficient of variation (%)	1.5	1.2	0.8	1.2	0.6

[a] The results for each amino acid are listed as percentage of total area in the HPLC chromatogram at 280 nm. Please note that the percentage areas do not directly correlate to the level of incorporation.

Table 1. In this group, Fmoc-Asp(O*t*Bu)-OPfp, Fmoc-Lys(Boc)-OPfp, Fmoc-Ser(*t*Bu)-OPfp, Fmoc-Ala-OPfp, and Fmoc-Phe-OPfp were included.

From these numbers it is obvious that the Pfp-esters included in the group are completely stable in DMF for at least a week at room temperature. The coefficient of variation of around 1% is equal to the margin of error for the experiments with quinary mixtures. From the rest of the experiments all the remaining Pfp-esters of interest were found to be stable for at least 2 days at room temperature. After 2 days the degradation of the highly reactive glycine derivative started influence the reaction outcome. Two days at room temperature is, however, more than sufficient for the purpose of handling and coupling.

However, the total time for a full library synthesis is in the order of a week, and to guarantee high quality results the stock solutions need to be stored colder in between couplings. A set of 15 selected Pfp-esters was mixed, divided, and stored in two different freezers (−18 and −80 °C). After 1, 3, and 6 days, aliquots were taken out and coupled to equally sized portions of the model resin. Comparison of the chromatograms revealed that the solution was fairly stable at −18 with a coefficient of variation of 6% and a maximum deviation of 20% after 6 days. However, much better results were obtained when storing at −80 °C where the stock solution was frozen solid. In these couplings no shift in product ratio was detected that was larger than the margin of error of the experimental setup (±1%). For maximum fidelity in library synthesis the stock solution should be divided into aliquots that are frozen in a −80 °C freezer and thawed immediately prior to use.

Finally, the quaternary and quinary stock solutions were aged for 6 months at room temperature and analyzed again. These studies supported the previous findings in that the only species completely consumed were glycine, arginine, and glutamine (Gln without side-chain protection). Interestingly, in one of the mixtures where the reactivities were closely matched (Gln, Glu, Tyr, Phe) the variation of the remaining components was less than ±18% even though one of the components (untritylated Gln) was completely consumed. In other mixtures where one of the components (in this case His(Trt)) were much less reactive than the rest, ratios far from equimolarity resulted when one of the reactive minor components was consumed (Gly). From these results it is safe to say that the assumption proposed by Acharya et al. that the isokinetic ratio determined for a binary mixture could be extrapolated to a ternary does not apply for quaternary or quinary mixtures and certainly not for a mixture of 15 components.

For the subsequent experiments a sufficiently stable mixture of 15 amino acids had thus been found. As mentioned above Ser, Thr, Cys, and Met were excluded from the isokinetic mixture. Arginine was also excluded as competing ring closing reactions would seriously compromise the reproducibility of the couplings. The scrambled positions in the peptide sequences of the library are included to provide enough sequences showing biological activity in a mixture for a detectable output signal to be generated. Even if arginine is excluded from this mixture it is likely that the presence of other similar amino acids like lysine, should provide a sufficiently active mixture. The strong signals observed in our kinase assays using the finished library confirmed this assumption (vide infra). Worth noting is that although no one has previously taken arginine instability into consideration in the context of isokinetic mixtures, the problem is not associated with the use of Pfp-esters specifically but is likely to exist with any active esters formed from arginine with a reactivity equal to or greater than the Pfp-ester derivative. On the contrary, with highly reactive intermediates formed by different coupling reagents, the cyclization to the δ-lactames should be substantial during the time frame of a peptide coupling. This is normally not a problem because couplings are generally made with an excess of reagent. However, in statistical mixture-based couplings even minute changes will have a major impact on the product ratio, and the inclusion of unstable substrates will infer a large dependence on the exact procedure used with respect to pre-mixing times, time of handling, reaction time (as a function of concentration and temperature), and so forth.

Our finding that the iterations had to be performed on the full set of 15 amino acids and the inherent difficulty in manually adjusting concentrations iteratively to find an equimolar mixture led us to search for a computer based approach. We decided to use statistical experimental design, also known as Design of Experiments (DOE), to help us in the search for the isokinetic mixture. In statistical experimental design multiple variables are changed at the same time according to a design that has been calculated to maximize the information content of the experiments.[25] To avoid too lengthy iterations when searching for the optimal amino acid ratios, constrained D-optimal mixture designs were used to plan the experiments.[26] The synthetic and analytical equipment used gave the possibility to perform and evaluate approximately 30 mixtures in parallel, and the designs were constructed as linear models in 15 factors to maximize the information gained by each iteration. The factors were the ratio of the 15 different amino acids used. The designs were then calculated using the statistical data analysis software Modde (Umetrics AB) as explained in the experimental section. New limits were set manually following each iterative round, starting from the calculated optimum for the previous iteration and accounting for experimental outliers and the results from earlier iterations. Experimental outliers represent artifacts caused by, for example, disturbing superimposition of a byproduct integral with a product integral in the HPLC chromatogram, an impure batch of

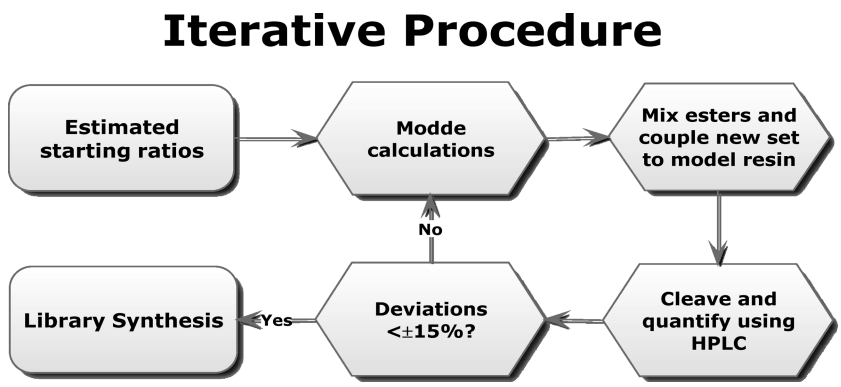

Figure 3. Workflow of the computer based iterative procedure.

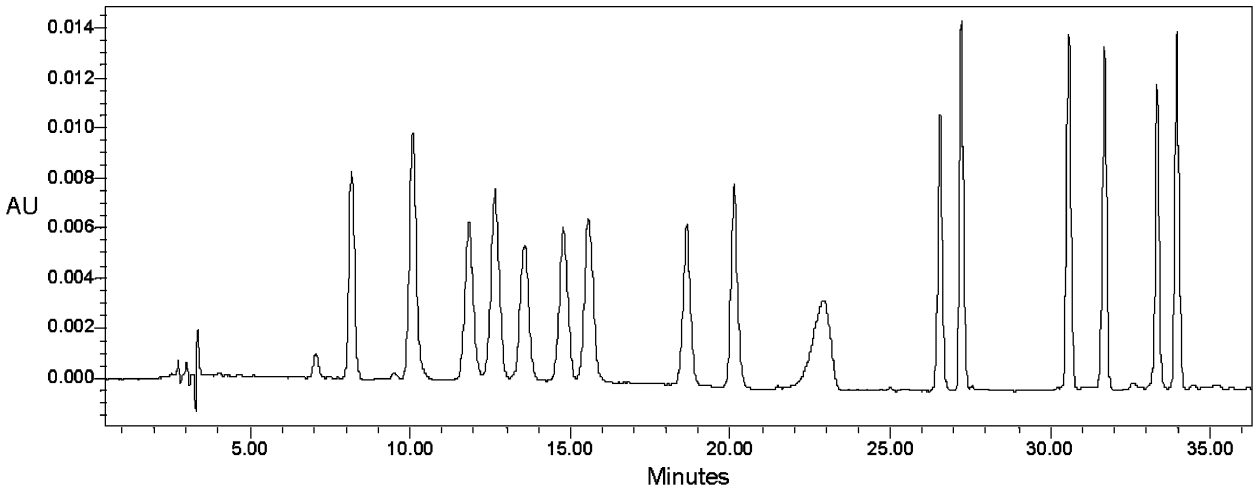

Figure 4. Chromatogram of 16 model tripeptides. The system used was A = 100 mM NH_4OAc pH 6.9, B = MeOH; 1.0 mL/min; isocratic 16% B for 10 min, a linear gradient to 20% B from 10 to 20 min, a linear gradient to 50% B from 20 to 30 min, and finally isocratic 50% B from 30 to 36 min.

Fmoc-His(Trt)-OPfp, and the poor stability of Fmoc-Arg-(Pbf)-OPfp. The work flow of the iterative procedure is illustrated in Figure 3.

4. HPLC Detection and Quantification. HPLC parameters were optimized for the specific tripeptides used in the determinations of isokinetic ratios. A large number of buffers, organic modifiers, columns, and gradients were tried. The optimal conditions turned out to be MeOH and aqueous ammonium acetate (100 mM, pH 6.9) on a C18-aq column with initial isocratic elution at 16.5% followed by a linear gradient to up to 50% MeOH. This system always produced baseline separation of the 16 tripeptides (arginine was included in iteration 1) used in the model reaction (Figure 4). The broadened peak corresponds to the Ac-Pro-Ala-Tyr-NH_2 tripeptide. This peak showed broadening in all different systems tried, but it was so well separated from the other peaks that it did not pose a problem.

The relative incorporation of each amino acid into the tripeptides was determined using UV-detection at 270 nm. This particular wavelength was identified by collecting and comparing the absorption spectra for all possible tripeptides in buffer and in MeOH/buffer at the particular MeOH concentration in which they elute. The standard deviation and maximum deviation for the entire group of tripeptides was plotted against wavelength (Figure 5). All tripeptide samples used in the UV measurements were meticulously prepared in triplicates, and three spectra were recorded for every sample. The average value of these nine spectra was used to represent each tripeptide in the subsequent calculations. Enough high quality data was thus available to allow for calculation of compensation factors for all tripeptides for any wavelength in the 250−290 nm span.

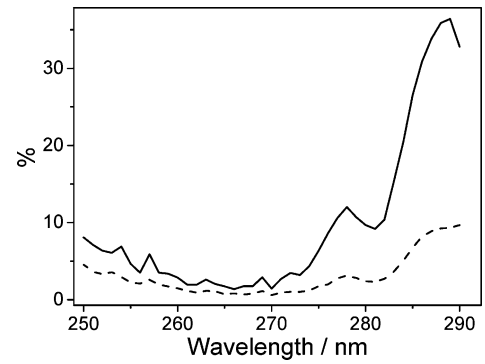

Figure 5. Absolute value of the maximum deviation (solid) and the standard deviation (dashed) of the difference between the absorption in pure buffer and MeOH/buffer for the entire set of tripeptides.

However, as it turned out, the difference between the absorption in buffer versus MeOH/buffer was very low at 270 nm. On closer examination of the differences at 270 nm it appeared that they showed a stochastic distribution around zero and thus reflected the margin of error of the experimental setup rather than an actual difference. Furthermore, at this wavelength only Tyr, Phe, and Trp absorb, and

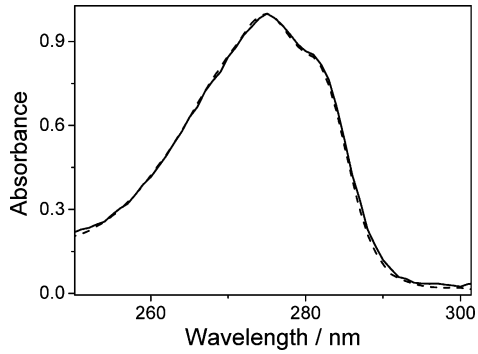

Figure 6. Overlaid normalized absorption spectra of the Ac-Tyr-Ala-Tyr-NH_2 (solid) and Ac-Ala-Ala-Tyr-NH_2 (dashed) tripeptides.

only three correction factors had to be determined. The correction factors should compensate for solvatochromic effects and any additional absorption by the amino acid in the third position of the tripeptide. The correction factor for the tripeptide with two tyrosine moieties (Ac-Tyr-Ala-Tyr-NH_2) was 0.5. That enough separation between the chromophores of the first and third amino acid residue was achieved is nicely illustrated by an overlay of the normalized spectra of Ac-Tyr-Ala-Tyr-NH_2 and Ac-Ala-Ala-Tyr-NH_2 where essentially identical spectra are obtained (Figure 6).

In the case of Ac-Phe-Ala-Tyr-NH_2 the influence of solvatochromic effects and additional absorption from the phenyl moiety was corrected for using the UV-spectra. First a constant describing the difference in absorption between 270 nm in MeOH/buffer and 280 nm in buffer was calculated (eq 1). At 280 nm the absorption of Phe is negligible. The part of the absorption at 270 nm originating from the Phe residue could thus be eliminated by multiplying the first coefficient with a second one describing the relationship between the absorption at 280 and 270 nm in buffer for a tyrosine containing tripeptide (eqs 2 and 3). The ratio of 0.98 from eq 2 is an average from the tripeptides not containing any additional chromophore, but is exceptionally stable for all the peptides so the value for any of the tripeptides could have been used.

$$K_1(270\text{ nm Tyr} + \text{Phe MeOH/buffer} \rightarrow 280\text{ nm Tyr} + \text{Phe buffer}) = 0.98 \quad (1)$$

$$K_2(280\text{ nm Tyr} + \text{Phe buffer} \rightarrow 270\text{ nm Tyr buffer}) = 0.98 \quad (2)$$

$$K_{(TOT)} + K_1 \times K_2 = 0.96 \quad (3)$$

For the Trp containing tripeptide three different constants had to be calculated and multiplied (eq 7). The first constant was calculated in the same manner as for the Phe containing tripeptide (eq 4). The second constant uses equations from the literature to calculate the part of the absorption at 280 nm in buffer that originates from the Tyr moiety (eq 5).[27] The third constant is identical to the one used for the Phe containing tripeptide (eq 6).

$$K_1(270\text{ nm Tyr} + \text{Trp MeOH/buffer} \rightarrow 280\text{ nm Tyr} + \text{Trp buffer}) = 1.02 \quad (4)$$

$$K_2(280\text{ nm Tyr} + \text{Trp buffer} \rightarrow 280\text{ nm Tyr buffer}) = 0.18 \quad (5)$$

$$K_3(280\text{ nm Tyr buffer} \rightarrow 270\text{ nm Tyr buffer}) = 0.98 \quad (6)$$

$$K_{(TOT)} = K_1 \times K_2 \times K_3 = 0.18 \quad (7)$$

The final constants are multiplied by the integrals from the chromatograms collected at 270 nm to obtain the corrected amino acid distribution. For further elaboration on the subject and a complete set of spectra of the data used in the calculations see Supporting Information.

5. Library Design. The library synthesized using the mixture we derived from the screening was in many respects similar to the commercially available one (http://www.anaspec.com/products/product.asp?id=44006) with three significant modifications incorporated to make the synthesis more straightforward. The first modification involved the attachment of biotin moieties directly to the N-terminus instead of to a lysine side chain. The side chain attachment would require either a more complicated synthesis or the use of an expensive pre-biotinylated lysine building block. The second modification was the exclusion of arginine from

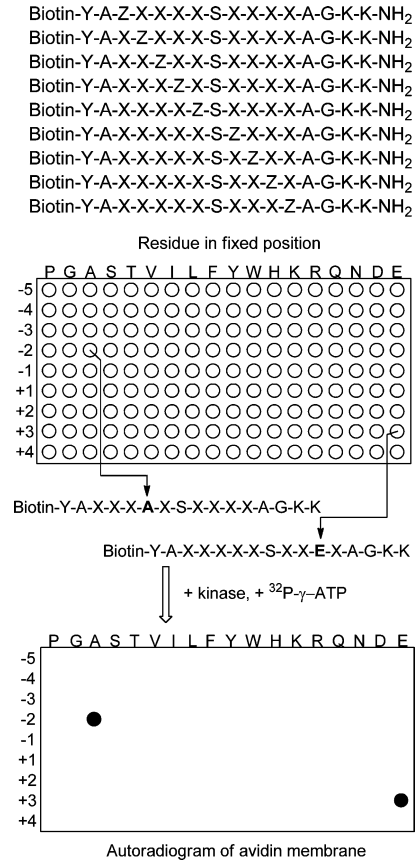

Figure 7. Sequences in our PS-SCL for kinase substrate specificity studies. The top part of the figure illustrates the layout of the entire library. The middle part shows how a sequence correlates to a specific well, and the lower part of the figure illustrates hits for alanine in the −2 position and glutamic acid in the +3 position after phosphorylation. X = degenerate mixture of the 15 natural amino acids excluding Met, Cys, Thr, Ser, and Arg. A fixed position is walked through the sequence and is varied between 18 natural amino acids (all except Met and Cys). The illustration is based on a similar figure in ref 15.

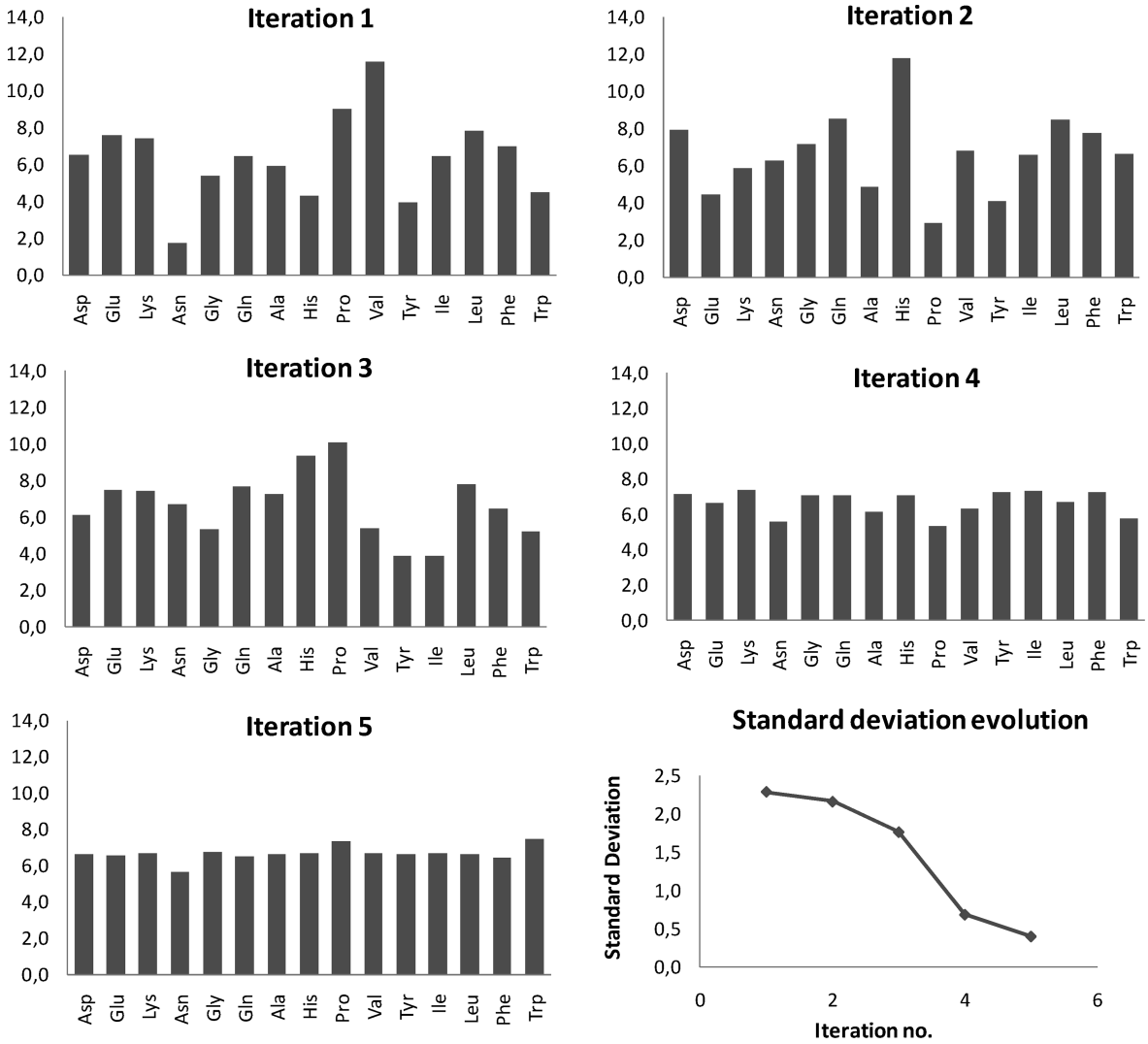

Figure 8. Level of incorporation in percent of the total amount of each of the 15 amino acids of the isokinetic mixture into the third position of the tripeptide for iterations 1−5 and the evolution of the standard deviation as a function of iteration number. The illustrations and the standard deviations are based on the best experiment from each iteration set. Arginine is omitted from the chart representing the first iteration for clarity.

the scrambled position because of its poor stability (vide supra). The third modification was the use of a serine as the phosphorylation motif instead of a serine/threonine mixture. All kinases able to phosphorylate threonine will also phosphorylate serine, so the inclusion of a mixture at this position would be redundant. The layout of the library and the principle of detecting hits are illustrated in Figure 7.

The Ala-Gly spacer between the scrambled positions and the Lys-Lys solubility tag was included in the original library for the purpose of Edman degradation. Even though we did not intend to perform Edman sequencing on our library, we kept the Ala-Gly spacer to allow for more productive comparisons with previously published data from studies using the commercial library.

6. Iteration Process. The iterations were performed in 22 or 30 parallel reactions using a Bohdan MiniBlock and a shaking table. If shaking was used, the variation between runs using the same stock solutions and identical amounts of model resin was around 1% but increased to 2−5% without shaking. Shaking is therefore preferable, but an acceptable result can be achieved without shaking as well. The starting point was taken from the initial experiments using quarternary and quinary mixtures recalculated to a full 16 amino acid mixture. Arginine was included in the first iteration, but was removed before the second. This gave the distribution shown in Figure 8 (raw data are available in the Supporting Information). An interesting property was discovered in the course of the iterations; residual trace quantities of piperidine from the Fmoc deprotection gave serious reproducibility problems with large deviations between runs using the same stock solution. The residual piperidine reacted preferentially with the most reactive Pfp-esters and thus changed the ratios. It turned out that five washes were sufficient to give high reproducibility if one minute shaking after each addition with DMF was introduced. Furthermore, our experiments showed that no gain in reproducibility was obtained if the number of washes was increased from 5 to 10 or 15. However, if shaking is inaccessible it could be advisible to extend the time of each wash with DMF or add more washes.

One week per iteration was normally required for design, preparation, execution, and analysis. After only five iterations we broke through the ±15% criterion (maximum deviation from the percentage representing equimolar concentration

Table 2. Isokinetic Ratios for Fmoc-Protected Pfp-Esters Determined by an Iterative Procedure

residue	ratio
Asp[a]	1.00
Gly	1.11
Phe	1.15
Ala	1.20
Tyr[b]	1.22
Glu[a]	1.35
Leu	1.39
Gln[c]	1.66
Trp[d]	1.74
Lys[d]	1.84
Pro	2.00
His[c]	2.49
Asn	5.05
Val	9.62
Ile	13.01

[a] O*t*Bu protected side chain. [b] *t*Bu protected side chain. [c] Trityl protected side chain. [d] Boc protected side chain.

of all species) which can be considered a very short time for such a highly complex mixture.

Furthermore, during the first three iterations substantial modifications were made on the way to correct for experimental problems such as arginine instability, byproduct peak superimposition in HPLC, and replacement of impure Fmoc-His(Trt)-OPfp. Without these complications the DOE approach would most likely have produced the finished mixture with even fewer iterations. After the fifth iteration when the threshold of $< \pm 15\%$ was reached, only three residues were outside $\pm 10\%$ from the desired 1/15 composition, the rest of the compounds were $< \pm 4\%$ with eight compounds $< \pm 1\%$. The average deviation was 3.3%. One or two more iterations would likely reduce the three $> \pm 10\%$ outliers to the lower numbers reached by the other residues, but the purities of the Pfp-esters are currently not high enough to motivate a lower threshold (for purities see Supporting Information). The method that we developed, however, could produce libraries with extremely narrow limit criteria if this is desired and high purity substrates are available.

7. Library Synthesis. The final protocol entails mixing of the commercially available Pfp-esters using the ratios shown in Table 2 so that a total concentration of 0.15 M is obtained, dividing this solution into the same number of aliquots as scrambled positions in the sequence, and storing the aliquots at −80 °C. The positions to be fixed in the library sequences are coupled to 0.15 M stock solutions of the pure Pfp-esters, and the scrambled positions are coupled to an aliquot of the isokinetic mixture thawed immediately prior to coupling.

The total amount of Pfp-esters in the solutions should correspond to 3 equiv compared to the resin loading. The resin used was a rink-amide resin and standard deprotection using 20% piperidine in DMF for 30 min was used for Fmoc removal. Five washes with DMF under shaking for 1 min each after deprotection was crucial for high fidelity results. Three washes with DMF under shaking for 1 min each was employed after coupling. Biotinylation was performed using standard methods with HBTU and DIPEA. Cleavage and deprotection was accomplished using three portions of TFA/H_2O/TIS (95:2.5:2.5).

8. Kinase Assay. The bovine protein kinase A (PKA) was used to test the performance of the newly synthesized library. A literature procedure using radiolabeled ATP was adapted to our library.[28] The signals were quantified using a Molecular Imager FX and Quantity One-4.6.8 software (BioRad) and normalized to the average value for all amino acids at a certain position in relation to the phosphorylated residue (Ser) (Figure 9, for radiographic image see Supporting Information Figure S1). At position −3, arginine gave a 12.5 times higher signal than average and at position −2, a 13.3 times higher signal. This is consistent with the known consensus sequence for PKA, which is R/K-R/K-X-S/T (ExPASy) (http://www.expasy.ch/cgi-bin/nicedoc.pl?PDOC00004). A lysine (K) at positions −2 and −3 also gave relatively high signals, 3.16 times the average, in our assay. Other pronounced values were those for arginine at position −4 and isoleucine at position +1 (3.89); this is also consistent with previous results by Hutti et al.[14] showing that a library synthesized by the new convenient method meets the required standards for kinase substrate specificity assays.

Conclusion

An efficient method to synthesize PS-SCLs for studying the specificity of protein kinases has been developed. The necessary isokinetic ratio was determined iteratively using a new HPLC based approach. A thorough study of a peptide model system revealed that 270 nm is a favorable wavelength to study product distribution as solvatochromic effects for tyrosine are negligible at this wavelength for different concentrations of MeOH ($\leq$50%) in buffer (100 mM NH_4OAc). With HPLC quantification and statistical experimental design the iterations quickly converged, and only five iterations, each consisting of 22 or 30 parallel reactions, were required to reach the predefined threshold of $\leq \pm 15\%$ deviation from equimolarity for any residue. The use of commercially available preactivated pentafluorophenyl esters of the amino acid derivatives reduced the number of operations during synthesis and increased reproducibility. Furthermore, activated esters of arginine showed low stability in solution, and their omission was deemed necessary for reproducibility. A few other Pfp-ester derivatives showed lower than expected stabilities, but careful choice of protective groups resolved these issues. It was also shown that stringent washing of the resin after deprotection of the Fmoc was crucial in mixture-based synthesis as traces of piperidine will disturb the ratio of incorporation. A library was synthesized using the new isokinetic ratio; this library was proven to be effective by comparison of the results from a PKA study with published results. The final protocol makes it convenient to synthesize high quality PS-SCLs regardless of previous experience in solid phase synthesis.

Experimental Section

Design of Experiments (DOE). D-optimal mixture designs were calculated and evaluated using Modde 8.0.2 (Umetrics AB). D-optimal mixture designs were calculated as linear models in 16 or 15 factors and three center points. For each experiment, five different designs were calculated with 27 ± 2 runs. The design with the highest G-efficiency

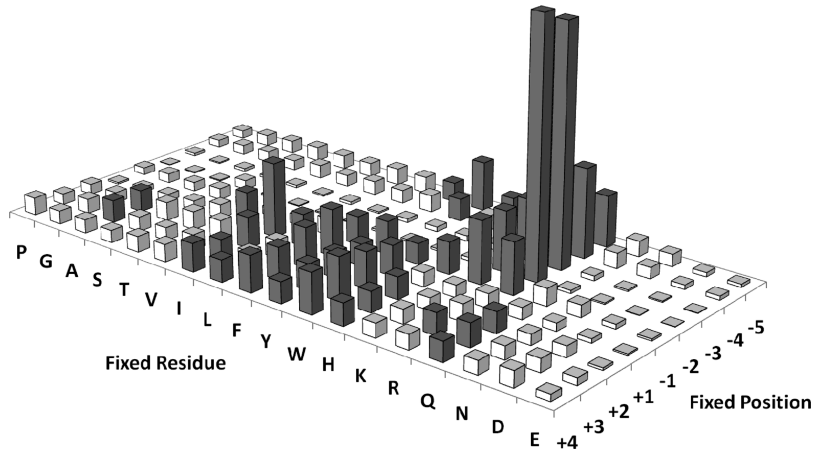

Figure 9. Sequence preference for phosphorylation by PKA. The bars represent the quantified phosphorylation of each peptide mixture, as determined by measurement of incorporation of radioactivity. Each phosphorylation value was normalized to the average value for all 18 amino acids at that position. Preferred sequences defined as having a value over 1 are shown as gray bars. Only the relative sizes of the bars are shown in the illustration. For specifics see Supporting Information.

was selected. The resulting designs can be found in the Supporting Information. After the experiments had been performed, multi-Y PLS models were calculated with the measured amino acid fractions as responses. The models received were of varying quality but were nonetheless used as such for optimizations of factors. The PLS models were used to optimize the factor levels through the optimizer in Modde. Amino acids that were considered to be erroneous in the experiment, for example, because of decay or underestimation in the previous optimization, were excluded from the optimization. For each amino acid, the response was to target 6.36 with a low limit of five and a high limit of seven. The optimizer was run repeatedly to give a result with acceptable responses with respect to the specific conditions in the experiment. The final conditions from the optimizer were converted to new factor limits manually. In that process, care was taken to adjust for problems in the last experiment and the experienced influence of a factor.

Iteration Process. For the iterations, reproducibility was a concern. The way to achieve this was to make 0.15 M stock solutions of all Pfp-esters using calibrated scales and diluting with DMF in 2 mL volumetric flasks. The flasks were silylated using Sigmacote to avoid droplet formation of the DMF. Some of the Pfp-esters were static and the use of an antistatic gun is advised for accurate weights. The exact amount in μL needed for each mixture was drawn by an Eppendorf Research pro fully automated pipet calibrated for DMF to compensate for capillary forces. A perfect correlation was obtained after the correction in the desired interval. The stock solution was thawed, vortexed to dissolve precipitated ester, and centrifuged before pipetting. Directly after the pipetting was finished from one stock solution, this was frozen on liquid nitrogen and the next one thawed. The stock solutions were then stored in a −80 °C freezer until the next iteration. The vials containing the mixtures to be coupled and analyzed were kept in a cooled metal block during pipetting to minimize degradation. The careful handling here of the solutions and mixtures is important as the iterations are much more sensitive than the actual library synthesis. The reason for this is that the amount of time needed for 15 × 30 careful additions of stock solution to the mixtures is significant (~6 h) and exceptionally high accuracy is needed for convergence in the iterations whereas small deviations can be tolerated in the library synthesis. After pipetting was complete, the mixtures were vortexed, centrifuged, and 176 μL quickly added to the deprotected resin (NH_2-Ala-Tyr-rink amide 0.37 mmol/g, 18 mg) and the shaking table started. The reaction was generally over in 1 h, but shaking was continued for 2 h, and then the reaction was for practical purposes allowed to sit without shaking overnight for cleaving and analysis the next day. Cleavage was achieved with TFA/H_2O/TIS 95:2.5:2.5 as can be considered the standard non-smelling cleavage cocktail for this protective group regimen ((Trp(Boc) was used). Evaporation by a stream of nitrogen gas using an aqueous sodium hydroxide bubbler to trap the TFA gave a solid gummy residue. This was treated with distilled ether, sonicated, and centrifuged in a total of three times to obtain the finished peptide mixture. This was dissolved in 50 μL of DMSO, and 5 μL of the resulting solution was drawn and diluted to 100 μL with H_2O and loaded into the HPLC for analysis.

Library Synthesis. The library was synthesized in parallel using a Bohdan MiniBlock and a shaking table. NH_2-Ala-Tyr-rink amide resin (18 mg, 0.37 mmol/g) was loaded into fritted polypropylene tubes using a 40 μL well dispenser plate. A DMF solution (150 μL, 0.15 M) of the appropriate Pfp-ester or isokinetic mixture was added using a dispensing pipet. This is just enough to cover the resin properly. Shaking was continued for 2 h, and the DMF removed by suction. Three DMF washes (300 μL) for each vessel were performed under shaking for 1 min. A solution of piperidine in DMF (300 μL, 20%) was added and shaking was started for 30 min. The piperidine solution was drained away, and five more DMF washes (300 μL) were performed with 1 min of shaking for each wash. The peptide chain was further extended by adding another aliquot of DMF solution (150 μL, 0.15 M) of the appropriate Pfp-ester or isokinetic mixture using a dispensing pipet. Once the correct length had been achieved the peptides were biotinylated by adding a biotinylation mixture (150 μL). This was prepared by mixing a solution of biotin in DMSO/DMF 1:1 (20.2 mL, 0.2 M), a solution of HBTU and HOBt (8.1 mL, 0.5 M in both), and

DIPEA (1.41 mL). Shaking was started overnight. Two washes with DMF/DMSO 1:1 (300 μL), two washes with DMF (300 μL), two washes with DCM (300 μL), and finally two washes with MeOH (300 μL) gave the finished resin-bound peptides. Drying using a stream of nitrogen for 2 h was followed by the addition of TFA/H_2O/TIS 95:2.5:2.5 (150 μL) and shaking for 1 h. This step was repeated two more times. Evaporation of the TFA using a stream of nitrogen and trapping the TFA using an aqueous sodium hydroxide bubbler gave a gummy residue. Trituration with distilled diethyl ether, sonication, centrifugation, and decantation was repeated three times to yield the finished peptides. The concentrations were adjusted using a UV−vis plate reader at 280 nm and Ac-Tyr-OMe as reference. The measurements were performed by dissolving the products in DMSO (100 μL) and taking an aliquot (5 μL). This was diluted 100 times, and 50 μL was added to a well on a 384 well Falcon UV-plate. The concentration was calculated and adjusted to 12.5 mM.

Kinase Assay. The kinase assay was performed mainly as described by Turk et al.[28] It was performed in 18 μL reactions on a 384-well polypropylene plate. Each peptide mix from the library was used as substrate in one well of the microtiter plate. The reactions were composed as follows: PKA (20 ng, SIGMA, P2645−400UN), ATP (100 μM), and γ[^{32}P]-ATP (0.25 μCi/μL in 16 μL kinase buffer). The kinase buffer was composed of Tris-HCl (50 mM; pH 7.4), $MgCl_2$ (20 mM), β-glycerophosphate (0.5 mM), and Tween-20 (0.1%). Peptide mixture (2 μL, 12.5 mM) was added as substrate to each reaction. The aqueous peptide mixtures (0.5 mM) were prepared by diluting the DMSO stock solutions (12.5 mM) in HEPES (pH 7.4, 20 mM).

The reaction plates were covered with adhesive tape and incubated in a water bath at 30 °C for 2 h. After incubation, the plates were chilled on ice, and 2 μL of each reaction were spotted out on a SAM^2 Biotin Capture Membrane (Promega, V7861) with a pin tool replicator (VP Scientific, VP 408FP6S2). The spots were allowed to incubate on the membrane for approximately 30 s, directly after which the membrane was immersed in the first wash solution. The membrane was washed twice in SDS/TBS (0.1%), twice in aqueous NaCl (2 M), three times in aqueous NaCl (2 M, containing 1% H_3PO_4), and twice in distilled water. After air-drying for 1 h, the membrane was wrapped in plastic and exposed to a phosphor storage screen for 16−17 h. The radioactive signal was then detected in a Molecular Imager FX (BioRad). To calculate the relative phosphorylation of the different peptide mixtures, signal strengths were quantified using Quantity One-4.6.8 (BioRad).

Acknowledgment. This project was funded by The Chemical Biology Platform at the University of Gothenburg, the Olle Engkvist Byggmästare Foundation, and the Signhild Engkvist Foundation.

Supporting Information Available. (1) Pfp-ester purities and lot numbers, (2) radiographic image from the PKA assay, (3) normalized numeric values from the PKA assay, (4) iteration designs and outcomes, (5) UV-measurements and reproducibility, (6) additional background for the wavelength choice, (7) raw data for Figure 5, and (8) UV-spectra for calculations of correction factors. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Dooley, C. T.; Houghten, R. A. Synthesis and Screening of Positional Scanning Combinatorial Libraries. In *Combinatorial Peptide Library Protocols*; Cabilly, S., Ed.; Humana Press: Totowa, NJ, 1998; Vol. 87, pp 13−24.
(2) Erb, E.; Janda, K. D.; Brenner, S. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11422–11426.
(3) Houghten, R. A.; Appel, J. R.; Blondelle, S. E.; Cuervo, J. H.; Dooley, C. T.; Pinilla, C. *BioTechniques* **1992**, *13*, 412–421.
(4) Dooley, C. T.; Houghten, R. A. *Life Sci.* **1993**, *52*, 1509–1517.
(5) Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. *BioTechniques* **1992**, *13*, 901–905.
(6) Appel, J. R.; Campbell, G. D.; Buencamino, J.; Houghten, R. A.; Pinilla, C. *J. Pept. Res.* **1998**, *52*, 346–355.
(7) Appel, J. R.; Johnson, J.; Narayanan, V. L.; Houghten, R. A. *Mol. Diversity* **1998**, *4*, 91–102.
(8) Houghten, R. A.; Appel, J. R.; Johnson, J.; Narayanan, V.; Pinilla, C.; Wilson, D.; Pines, T. *Ann. Oncol.* **1998**, *9*, 005.
(9) Houghten, R. A.; Pinilla, C.; Appel, J. R.; Blondelle, S. E.; Dooley, C. T.; Eichler, J.; Nefzi, A.; Ostresh, J. M. *J. Med. Chem.* **1999**, *42*, 3743–3778.
(10) Houghten, R. A.; Wilson, D. B.; Pinilla, C. *Drug Discovery Today* **2000**, *5*, 276–285.
(11) Houghten, R. A.; Pinilla, C.; Giulianotti, M. A.; Appel, J. R.; Dooley, C. T.; Nefzi, A.; Ostresh, J. M.; Yu, Y.; Maggiora, G. M.; Medina-Franco, J. L.; Brunner, D.; Schneider, J. *J. Comb. Chem.* **2008**, *10*, 3–19.
(12) Meldal, M.; Svendsen, I.; Breddam, K.; Auzanneau, F.-I. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3314–3318.
(13) Backes, B. J.; Harris, J. L.; Leonetti, F.; Craik, C. S.; Ellman, J. A. *Nat. Biotechnol.* **2000**, *18*, 187–193.
(14) Hutti, J. E.; Jarrell, E. T.; Chang, J. D.; Abbott, D. W.; Storz, P.; Toker, A.; Cantley, L. C.; Turk, B. E. *Nat. Methods* **2004**, *1*, 27–29.
(15) Yaffe, M. B. *Nat. Methods* **2004**, *1*, 13–14.
(16) Atherton, E.; Cameron, L. R.; Sheppard, R. C. *Tetrahedron* **1988**, *44*, 843–857.
(17) Albericio, F.; Bofill, J. M.; El-Faham, A.; Kates, S. A. *J. Org. Chem.* **1998**, *63*, 9678–9683.
(18) Ragnarsson, U.; Karlsson, S. M.; Sandberg, B. E. *Acta Chem. Scand.* **1971**, *25*, 1487–1489.
(19) Ragnarsson, U.; Karlsson, S. M.; Sandberg, B. E. *J. Org. Chem.* **1974**, *39*, 3837–3841.
(20) Ostresh, J. M.; Winkle, J. H.; Hamashin, V. T.; Houghten, R. A. *Biopolymers* **1994**, *34*, 1681–1689.
(21) Acharya, A. N.; Ostresh, J. M.; Houghten, R. A. *Biopolymers* **2002**, *65*, 32–39.
(22) Frank, R. *Tetrahedron* **1992**, *48*, 9217–9232.
(23) Sieber, P.; Riniker, B. *Tetrahedron Lett.* **1987**, *28*, 6031–6034.
(24) Ohfune, Y.; Kan, T.; Nakajima, T. *Tetrahedron* **1998**, *54*, 5207–5224.
(25) Antony, J. *Design of Experiments for Engineers and Scientists*; Butterworth-Heinemann: Oxford, 2003.
(26) Eriksson, L.; Johansson, E.; Wikström, C. *Chemom. Intell. Lab. Syst.* **1998**, *43*, 1–24.
(27) Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. *Protein Sci.* **1995**, *4*, 2411–2423.
(28) Turk, B. E.; Hutti, J. E.; Cantley, L. C. *Nat. Protoc.* **2006**, *1*, 375–379.

CC100095Y