Protein Expression and Purification 67 (2009) 139–147

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Robust production of a peptide library using methodological synchronization

Kristen L. Huber, Kevin D. Olson, Jeanne A. Hardy *

Chemistry Department, University of Massachusetts at Amherst, Amherst, MA 01003, USA

ARTICLE INFO

Article history:
Received 20 February 2009
and in revised form 5 May 2009
Available online 18 May 2009

Keywords:
Peptide library
Fusion partner
Avian pancreatic polypeptide
Human rhinovirus 3C protease
Methodological synchronization
Isotopic labeling of peptides
Peptide expression

ABSTRACT

Peptide libraries have proven to be useful in applications such as substrate profiling, drug candidate screening and identifying protein–protein interaction partners. However, issues of fidelity, peptide length, and purity have been encountered when peptide libraries are chemically synthesized. Biochemically produced libraries, on the other hand, circumvent many of these issues due to the fidelity of the protein synthesis machinery. Using thioredoxin as an expression partner, a stably folded peptide scaffold (avian pancreatic polypeptide) and a compatible cleavage site for human rhinovirus 3C protease, we report a method that allows robust expression of a genetically encoded peptide library, which yields peptides of high purity. In addition, we report the use of methodological synchronization, an experimental design created for the production of a library, from initial cloning to peptide characterization, within a 5-week period of time. Total peptide yields ranged from 0.8% to 16%, which corresponds to 2–70 mg of pure peptide. Additionally, no correlation was observed between the ability to be expressed or overall yield of peptide-fusions and the intrinsic chemical characteristics of the peptides, indicating that this system can be used for a wide variety of peptide sequences with a range of chemical characteristics.

Introduction

Peptide libraries are commonly used in a variety of endeavors including identifying peptide–DNA interactions [1], as surrogates for screening protein–protein interactions [2–5], and as a basis for finding potential peptidic drug molecules. Peptide libraries have also been used for profiling substrate specificities of proteases [6–11], phosphatases [12–14], kinases [15–18], and other drug targets [19–22]. Thus production of high quality peptide libraries is of wide interest for many applications.

Typically, peptides for libraries have been produced via solid-phase synthesis, which involves sequential coupling of amine-protected amino acids to resin-bound amino acids. The resulting libraries generally contain peptides no longer than 15–20 amino acids, which can prove limiting in applications requiring longer peptides. The length constraints for chemically synthesized peptides are the result of coupling and deprotection efficiencies at each step, such that an exponential decrease in sequence fidelity is observed as a function of length. In addition, enantiomerization of nineteen of the 20 naturally occurring amino acids and the associated difficulties in purification of the isomers set practical limits to peptide lengths in library synthesis [23,24]. Sometimes peptide sequences are restricted due to steric clash of adjacent amino acids with bulky chemical catalysts or intermolecular aggregation which results in a low efficiency of the chemical coupling reaction [25–27]. Overall, chemically synthesized peptide libraries of longer lengths tend to have decreased sample quality, be costly and have increased production time.

Genetically encoded peptide libraries offer several advantages in library design. The resulting libraries can contain longer peptide lengths and significantly increased yields while avoiding the more common limitations of chemical synthesis. Biological synthesis of peptides of longer lengths can be particularly important for production of isotopically labeled peptides for use in heteronuclear nuclear magnetic resonance (NMR) spectroscopy. The principle advantages of biological production result from the fidelity of protein machinery, particularly because the ribosome is not limited in sequence or length of synthesized peptides and has an error rate of only 0.01% error per amino acid added [28]. Genetically encoded peptide libraries have not been widely used due to difficulties in the expression and purification of small peptides. This problem has been overcome by the use of protein-expression fusion partners such as glutathione-*S*-transferase [29], maltose binding protein [30], and thioredoxin [31] being the three most widely used fusions. While fusion proteins often promote production, this method can be time consuming and requires post expression protease processing and purification of the peptide from the protease. However, the addition of fusion partners has resulted in the improvement of the overall yields by enhancing solubility of the peptide of choice.

We report a method that allows robust expression of a genetically encoded peptide library that addresses the issues of purity, yield, length of the purified peptide, and batch-to-batch variability

* Corresponding author. Address: University of Massachusetts Amherst, 104 Lederle Graduate Research Tower, 710 North Pleasant St., Amherst, MA 01003, USA. Fax: +1 413 545 4490.
E-mail address: hardy@chem.umass.edu (J.A. Hardy).

1046-5928/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2009.05.006

which we have observed with chemically synthesized peptides, in addition to high cost and extended time of production. This genetically encoded system features thioredoxin as a fusion tag for the stably folded peptide scaffold avian pancreatic polypeptide (aPP). aPP is a 36 amino acid peptide that contains a hydrophobic core and hydrogen bond network between the α-helix and polyproline helix of the peptide [32,33] (Fig. 1A and B). These properties are responsible for its fold and stability. Previous work has demonstrated that the presence of the hydrophobic core and hydrogen bond network render aPP insensitive to mutations over much of the amino acid sequence [34,35] or to a C-terminal truncation [33,36].

A common difficulty with peptide expression is residual amino acid overhangs following the protease cleavage event. The native sequence of aPP allows inclusion of an N-terminal cleavage site for human rhinovirus 3C protease, a highly specific protease, without any unwanted amino acid additions to the peptides. Human rhinovirus 3C protease cleaves the sequence LEVLFQ-GP, generating a gly-pro overhang upon cleavage. In aPP the first two amino acids are gly-pro, so no residual amino acids are left upon cleavage. Combining this optimized expression system with methodological synchronization of site-directed mutagenesis, protein-expression and purification, a 20-member aPP variant library of peptides 27 amino acids in length was generated. This method allowed production of the desired library with overall improved yields, purity, and cost, all on a time frame that is comparable to synthetic peptide library methods on the same production scale.

Materials and methods

Cloning of recombinant peptide-fusion variants

In order to create the parent vector, pET32-peptide, for the peptide-fusion DNA variants, the aPP gene was amplified by PCR from the pJC20[35] vector (from Alana Schepartz and Doug Daniels). Using the forward primer 5′-GTACAA*CCATGG***CTGGAAGTGCTGTTTCAG**GGTCCGTCCCAGCCGACCTACCC-3′ that included the human rhinovirus 3C cleavage sequence (bold) and NcoI endonuclease restriction site (italics) and the reverse primer 5′-TCGAGC*CTCGAG***CTA**GTAACGGTGACGGGTAACAACGTTCAGG-3′ which encodes the XhoI restriction site (italics) and stop codon (bold), the desired gene was produced. This gene product was then ligated into the thioredoxin fusion tag-, 6× His purification tag-containing vector, pET32b (Novagen) via restriction sites NcoI and XhoI. Insertion was confirmed by sequence analysis (Genewiz, Inc.).

The newly constructed vector, pET32-peptide, was used for PCR-based site-directed mutagenesis via Quikchange® (Stratagene) to create the variable peptide sequences. Initial base constructs, B3 and B3II were created using two sets of primers. Mutational sites include Q4, T6, L17, I18, Y21, D23, Q25, L24, and Y27. Secondary base constructs B3DQ and B3NQ were produced from the B3 parent sequence by mutating sites D10, D11, and D16 while B3IQ and 3INQ were produced from the B3II sequence using two sets of primers. Remaining peptide sequences were produced from single step mutagenesis at position D11 utilizing one set of primers. The nucleotide sequences were optimized for expression in *Escherichia coli* and all constructs were verified via sequence analysis.

Expression and purification of recombinant peptide-fusion variants

DNA sequences encoding desired peptide sequences were transformed into *E. coli* strain BL21(DE3) competent cells (Novagen) for expression. These cells were inoculated into 2× YT media containing 100 μg/mL ampicillin (Sigma). Cells were grown at 37 °C until an OD_{600} of 0.6 was reached. Isopropyl β-D-1-thiogalactopyranoside (IPTG), (Anatrace) was then added to a final concentration of 1 mM. Cells were allowed to grow at 37 °C for an additional 3 h and then harvested by centrifugation at 5000 rpm (Sorvall SLC-4000 rotor) for 10 min.

Cells were resuspended in lysis buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 2 mM imidazole) and lysed using a microfluidizer system (Microfludics). The resulting solution was then centrifuged at 15,000 rpm (Sorvall SS-34 rotor) for 1 h at 4 °C to remove cellular debris. Lysates were passed over HiTrap™ Chelating HP columns

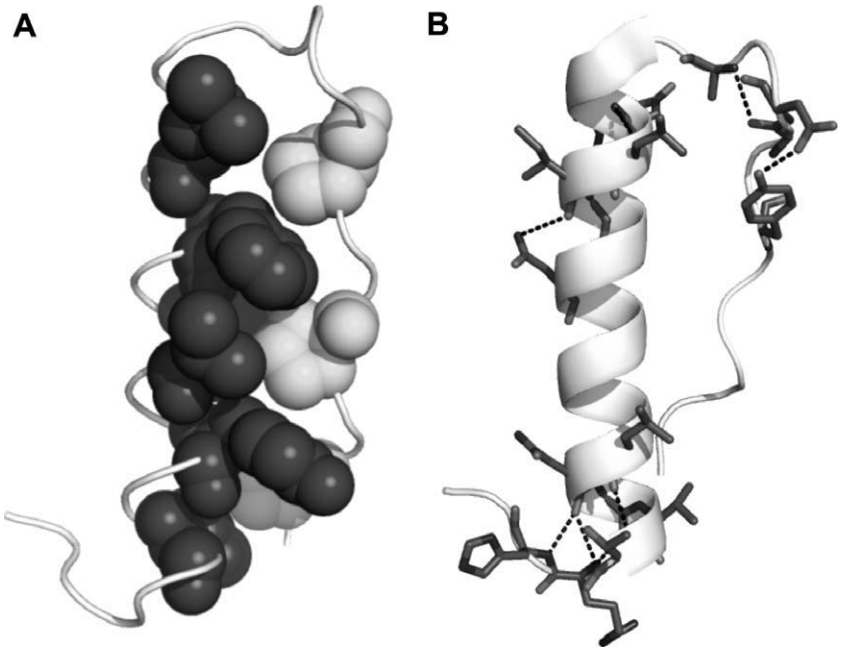

Fig. 1. aPP structure. (A) aPP backbone (lightest gray) has a hydrophobic core created by three prolines located on the poly-proline helix (medium gray) and various residues on the α-helix (dark gray). (B) aPP's internal hydrogen bond network, highlighted in dashed lines, is responsible for the impressive structural stabilization of a peptide that is just 36 amino acids long.

(GE Healthcare) charged with nickel. aPP, B3, B3DQ, B3DE, B3DK, B3DR, B3NE, B3NQ, B3NK, and B3II bound proteins were washed and eluted using a step gradient that consisted of a wash step using 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 10 mM imidazole and an elution step using 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 250 mM imidazole. Peptide-fusions B3, B3DK, B3DR, B3NQ, B3NE, B3NK, and B3II eluted proteins were subjected to ion-exchange chromatography to obtain additional purity. The previously eluted proteins were diluted 1:5 using buffer A (20 mM Tris, pH 8.0, 2 mM dithiothreitol (DTT)) and bound to Macro-Prep HighQ cartridge (Bio-Rad) at 5% buffer B (buffer B: 20 mM Tris, pH 8.0, 2 mM DTT, 1 M NaCl). Peptide-fusion proteins were eluted via a linear gradient that ranged from 50 to 350 mM NaCl over 240 min at a flow rate of 3 mL/min. Peptide-fusions B3NR, B3IQ, B3IK, B3IE, B3IR, B3IL, 3INQ, 3INE, 3INR, and 3INK were purified on a HiTrap™ Chelating HP columns (GE Healthcare) charged with nickel ions, using a gradient from 0 to 50 mM imidazole over 300 min. All peptide-fusions were analyzed by 16% SDS–PAGE gels that were stained with Coomassie brilliant blue (Sigma). Peptide-fusions were stored at −20 °C until required for cleavage and separation.

Expression and purification of human rhinovirus 3C protease

The human rhinovirus 3C protease gene in the pGEX vector (GE Healthcare) was transformed into *E. coli* strain BL21(DE3) competent cells (Novagen) for expression. These cells were inoculated into 2× YT media containing 100 μg/mL ampicillin (Sigma). Cells were grown at 37 °C until an OD_{600} of 0.6 was reached. IPTG was then added to a final concentration of 1 mM. Cells were allowed to grow at 20 °C for 18 h and then harvested by centrifugation at 5000 rpm (Sorvall SLC-4000 rotor) for 10 min.

Cells were resuspended in binding buffer (20 mM NaH_2PO_4, pH 7.4, 150 mM NaCl, 2 mM DTT) and lysed using a microfluidizer system (Microfludics). The resulting solution was then centrifuged at 15,000 rpm (Sorvall SS-34 rotor) for 1 h at 4 °C to remove cellular debris. Lysates were passed over GSTrap™ FF column (GE Healthcare) and washed with 50 mM Tris–HCl, 5 mM reduced glutathione, 2 mM DTT to remove loosely binding contaminants. The majority of the protease was eluted using 50 mM Tris–HCl, 15 mM reduced glutathione, 5 mM DTT followed by a second elution step of 50 mM Tris–HCl, 35 mM reduced glutathione, 5 mM DTT to release remaining protease and recharge the column.

The elution was diluted 1:5 using buffer A (20 mM Tris, pH 8.0, 2 mM DTT) and bound to Bio-Rad Macro-Prep HighQ cartridge at 2% buffer B (buffer B: 20 mM Tris, pH 8.0, 2 mM DTT, 1 M NaCl). Protease was eluted via a linear gradient from 20 to 100 mM NaCl over 50 min at a flow rate of 3 mL/min.

Peptide-fusion cleavage by human rhinovirus 3C protease and separation from fusion partners

Peptide-fusion and protease were incubated at 4 °C for a 5-h digestion period in a 1:25 protease to protein ratio. Reverse-phase HPLC was carried out on the cleaved peptide-fusion samples to separate the peptide from the protein components. Cleaved peptide-fusion samples were filtered through 0.22 μm syringe filter and loaded onto a Waters Sunfire Prep C18 column (10 × 50 mm, 5 μm, and 300 Å). Separation of the peptide from fusion protein and contaminants was carried out by reverse-phase HPLC on a Shimadzu HPLC system equipped with a SPD-20AV Prominence UV/Vis detector and LC-20AT Prominence liquid chromatograph. The mobile phase was 0.1% trifluoroacetic acid in water (buffer A), with an elutant of 0.1% trifluoroacetic acid in acetonitrile (buffer B). The column was developed with a three phase gradient consisting of a steep change in organic (5–30% buffer B over 2.5 min) for buffer component elution, a shallow step (30–45% buffer B over 15 min) for peptide elution, followed by a steep gradient (45–100% buffer B over 4 min) to regenerate the column and elute larger proteins such as thioredoxin and human rhinovirus 3C. Absorbance was monitored at 214 and 280 nm.

Mass spectrometry

To determine peptide molecular weight and degree of purity, peptide samples were analyzed using electrospray mass spectrometry. Lyophilized peptide samples were resuspended in water to a peptide concentration of 50 μM. Five milliliters of sample was analyzed using an Esquire-LC electrospray ion trap mass spectrometer (Bruker Daltonics, Inc.) set up with an ESI source and positive ion polarity. This system was equipped with an HP1100 HPLC system (Hewlett–Packard). Scanning was carried out between 200 and 2000 *m/z*, and the final spectra obtained were an average of 10 individual spectra. Equipment used was located at the Mass Spectrometry Center of the University of Massachusetts at Amherst.

Results

Construction of peptide-fusion expression vector and aPP variants

A peptide-expression system for the library of aPP-based peptides was generated as thioredoxin fusion proteins. A goal of this project was to streamline all steps in the process of cloning, expression and purification. In order to limit subsequent sub-cloning events, the parent vector (Fig. 2) was created through a single PCR reaction that resulted in the gene for a human rhinovirus 3C protease cleavage site (LEVLFQ) and the truncated aPP sequence (GPSQPTYPGDDAPVEDLIRFYNDLQQY). This gene product was then ligated into the thioredoxin fusion gene and 6× His purification tag containing vector, pET32b via restriction sites NcoI and XhoI. The resultant thioredoxin–peptide-fusion construct then served as the first base sequence for further mutagenesis.

aPP variant sequences were created by mutating codons for 6–12 of the amino acids in the truncated 27 amino acid peptide scaffold (Fig. 3A) via a Quikchange® (Stratagene) mutagenesis strategy. This limited library was designed as aPP mimics of natural protease-inhibiting peptides. Mutations included the desired variations in the original peptide sequences and took into consideration amino acids that are known to be structurally important for scaffold

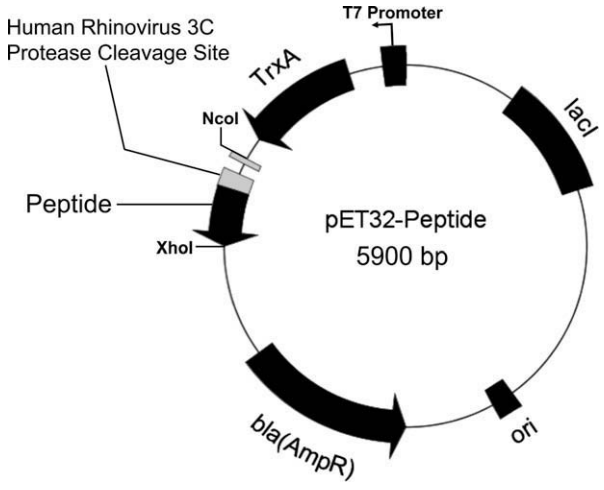

Fig. 2. Schematic representation of expression vector, pET32-peptide. The sequences encoding T7 promoter, a fusion partner thioredoxin (TrxA), 6× His tag, human rhinovirus 3C protease cleavable linker, peptide, and ampicillin resistance are shown.

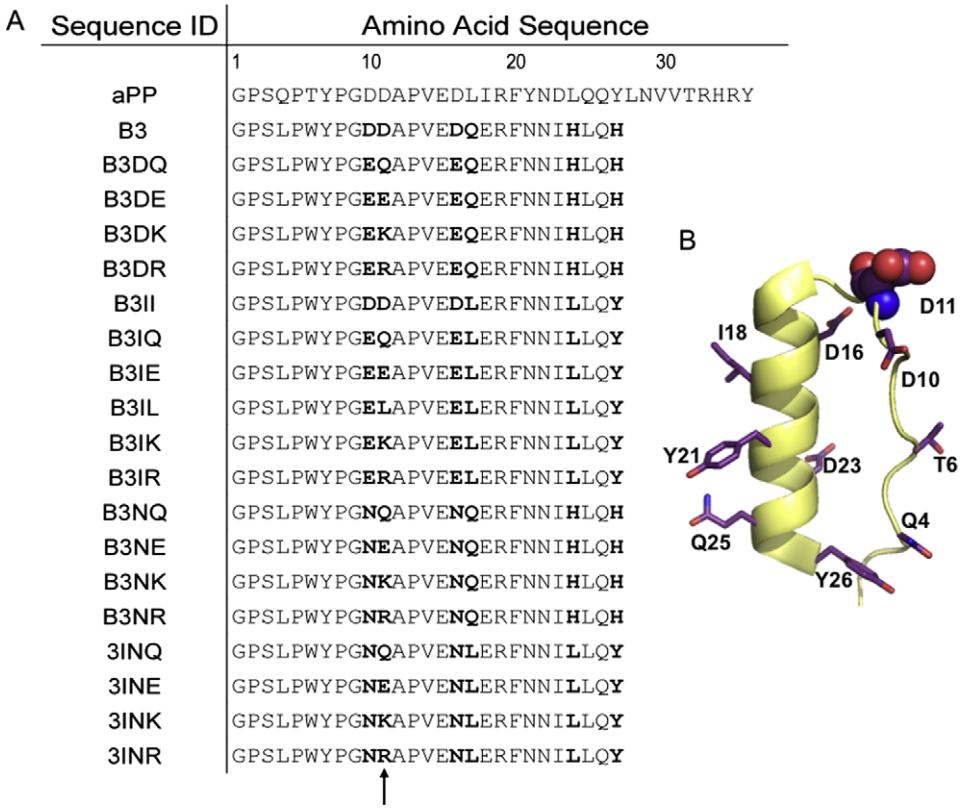

Fig. 3. Structure of peptide library. (A) Amino acid sequences of the 20-member peptide library. Highly interrogated residues, which vary relative to the starting B3 sequence, are bold. (B) Cartoon representation of truncated aPP peptide scaffold with sites of mutations shown in stick representation and highly interrogated site D11 represented in spheres.

stability. Sites of mutagenesis are throughout the truncated aPP sequence and are spread across the aPP structural elements (Fig. 3B). Mutational sites include Q4, T6, D10, and D11 on the polyproline helix and D16, D17, I18, Y21, D23, L24 Q25, and Y26 on the α-helix. In addition D11, was selected as a site to be extensively interrogated. By focusing our investigation in one region, we were able to produce many variants in a limited number of mutagenic steps. Moreover, all mutational oligonucleotides were designed to limit cost and allow rapid production of the new peptide encoding DNA constructs. These designs focused on producing an array of peptide sequences using the minimum number of rounds of mutagenesis. If mutagenesis had been performed by sequential means, 123 rounds of mutagenesis would have been necessary to build the complete library. In contrast, our method required only 29 round of mutagenesis. Mutations were quickly assessed by growing the transformed *E. coli* cells for 8 h and plasmid prepping the DNA in time for same day sequencing by an outsourced company (Genewiz Inc.). Upon sequence analysis, one out of two or more clones tested obtained the desired mutations. The success rate was directly related to the thoroughness of the DpnI digestion step during the Quikchange® protocol (data not shown).

Expression and purification of recombinant peptide library

Once the desired genetic mutations were verified, the vectors were transformed into the BL21(DE3) strain of *E. coli*, expressed and harvested. The resulting cell pellets were lysed and prepared for purification. Various avenues were tested to optimize the purification of the peptide-fusion from other bacterial proteins. This was done in order to minimize the potential additive loss of protein during each purification step. Therefore, nickel-affinity column chromatography using a step gradient alone or in combination with ion-exchange methods as well as nickel-affinity column chromatography using linear gradients were explored in order to obtain a peptide-fusion sample of adequate purity without sacrificing yield. For example, peptide B3NK was purified on a nickel-affinity column using a step gradient and ion-exchange methods while peptide B3NR was purified by a nickel-affinity column using a linear imidazole gradient only (Fig. 4). It is clear that a two-step purification yields protein of higher purity compared to nickel-affinity chromatography alone. However, since only a few contaminants are removed by the ion-exchange chromatography step in spite of a significant loss of sample, the single-step linear gradient using nickel-affinity chromatography became the preferred method of purification.

Once the peptide-fusions were purified to a satisfactory degree, cleavage of the peptide from its fusion partner and linkers was accomplished by the highly specific human rhinovirus 3C protease (Fig. 5). Various protease to peptide-fusion ratios were examined for optimal cleavage in minimal time. As a 5-h digestion with a 1:25 protease to protein ratio was sufficient for full cleavage of peptide from the fusion tag, this method emerged as the time-optimized protocol. In order to separate the liberated peptide from the other components in solution, reverse-phase HPLC was employed. By employing a three phase gradient which changes in organic phase from steep for buffer component elution, to shallow for peptide elution, to a second steep phase to regenerate the column and elute larger proteins such as thioredoxin and human rhinovirus 3C protease, the peptide was separated during a 21.5 min gradient. For example, cleaved peptides B3DE, B3II, and B3NR, all were successfully separated from the contaminating proteins and cleaved thioredoxin fusion partners via the three phase gradient (Fig. 6). Major peaks in the chromatogram correspond to buffer components (peak 1) and larger proteins such as the thioredoxin tag

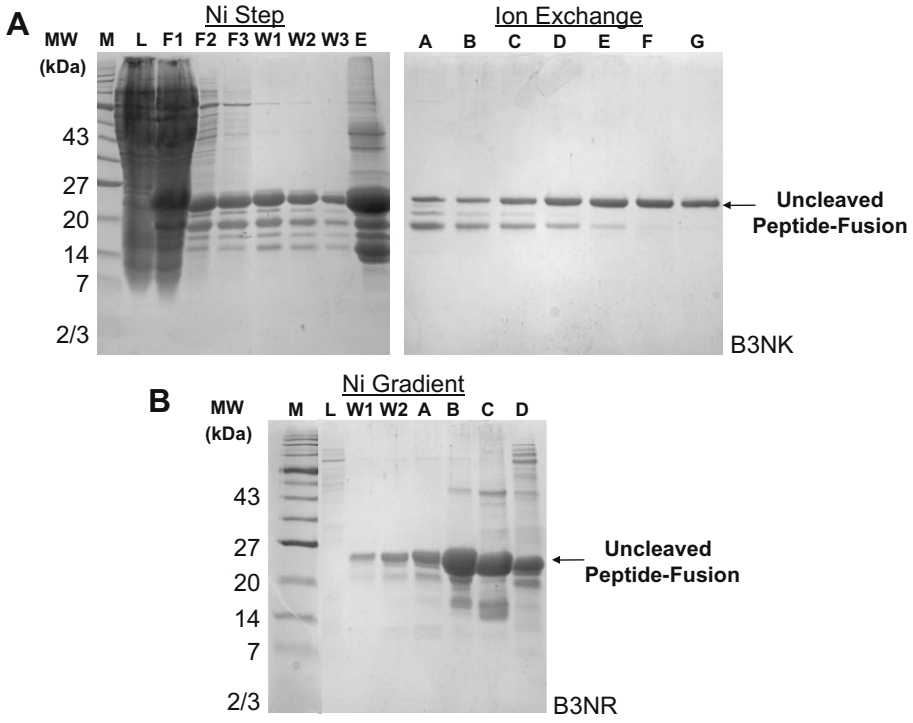

Fig. 4. Peptide-fusion purifications. (A) SDS–PAGE gels of peptide-fusion B3NK purified using nickel-affinity chromatography with a step gradient (load, L; flow through 1, F1; flow through 2, F2; flow through 3, F3; wash 1, W1; wash 2, W2; wash 3, W3; and elution, E) and ion-exchange (fractions A–G) are indicated. (B) SDS–PAGE gel of peptide-fusion B3NR purified using nickel-affinity chromatography and a linear imidazole elution gradient from 75 to 250 mM imidazole (load, L; wash 1, W1; wash 2, W2; and elution fractions A–C). Fraction D, which was eluted with 1 M imidazole, contained a mixture of tightly bound B3NK and contaminants. Uncleaved peptide-fusion MW = 24 kDa.

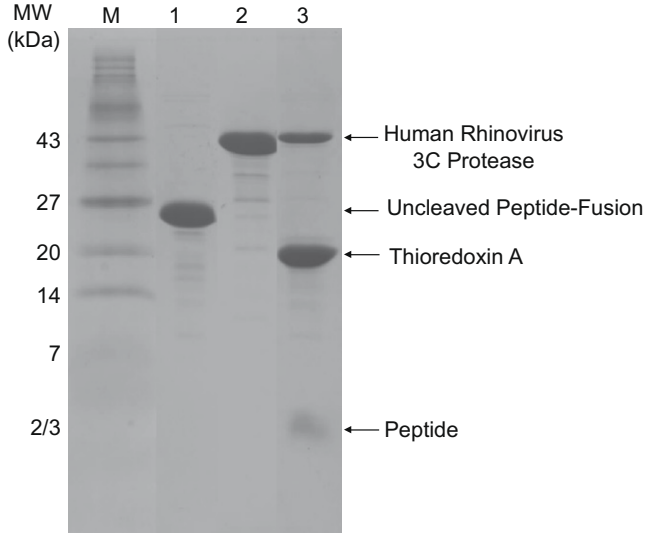

Fig. 5. Peptide–thioredoxin fusion protein cleavage. SDS–PAGE gel of the purified thioredoxin–peptide fusion (lane 1). Purified human rhinovirus 3C protease (lane 2) and cleavage reaction including human rhinovirus 3C protease, cleaved thioredoxin, and free aPP peptide (lane 3). Peptide MW = 3176 Da. Human rhinovirus 3C protease MW = 46,000 Da. Thioredoxin with tags MW = 27,912 Da.

and protease (peak 3) as verified by SDS–PAGE analysis. To verify the peptide identity and purity from the expected peptide samples (peak 2) on the reverse-phase HPLC chromatogram, electrospray ionization mass spectrometry (ESI-MS) was performed. The spectra indicate peptides are of the expected molecular weights (3119.1–3201.3 Da) (Fig. 7). In addition, the chromatogram for the liquid chromatography that is in line with the mass spectrometer showed only a single peak. Averaging of spectra from all regions of the peak resulted in uniform mass spectra indicating only one molecular species was present. These data suggest that the purified peptides are of extremely high purity.

In order to assess the overall utility of the peptide library construction system, the total amount of peptide fusion that was expressed in raw *E. coli* lysates is compared to the overall yield of each of the purification methods (Table 1). Total peptide yield ranged from 0.8% to 16% from the overall expression. A major contributor to low peptide yield is likely loss of protein at the step of column loading as the flow through during initial purification due to overloaded resin. In addition, other contributing losses are observed at the ionic exchange and reverse-phase HPLC portions of the purification. The ability to be expressed and the overall yield after purification for each peptide were compared to different peptide features including charge, isoelectric point, and amino acid characteristics. Peptide charge ranged from −0.06 to −4.06 and isoelectric point from 3.77 to 6.92. The number of acidic residues in the peptide sequences ranged from two to five while one to three basic residues, six to 12 polar residues and seven to nine hydrophobic residues were present in the various sequences. Favorably, no correlation between the expressibility of the peptide-fusion and intrinsic chemical characteristics of the peptides was found. Likewise, no correlation between peptidic character and overall yield of pure peptide was observed. This suggests that our system can be used for a wide variety of peptide sequences with a range of chemical characteristics.

Discussion

During chemical synthesis of peptides, sequence errors are a likely result due to the imperfect coupling efficiency at each step. For chemically synthesized peptides, the contaminants are extre-

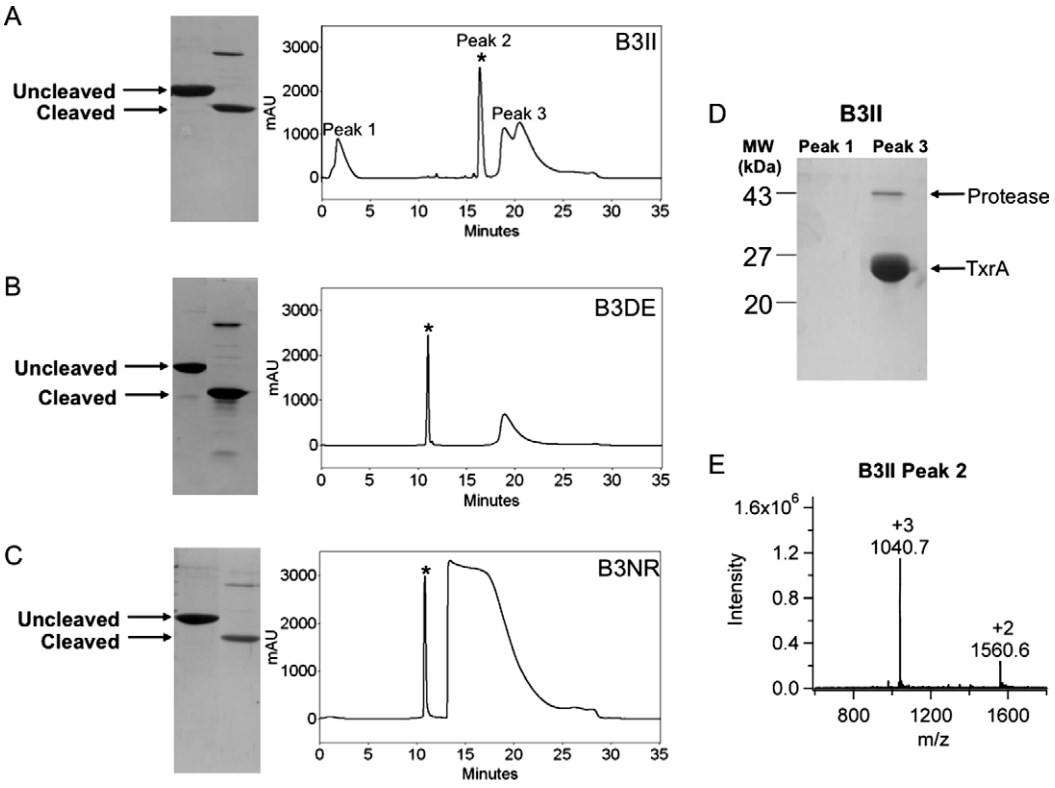

Fig. 6. HPLC purification chromatograms of cleaved peptide-fusion samples. (A–C) SDS–PAGE gel of uncleaved and cleaved peptide-fusion with corresponding HPLC separation chromatograms. Peak 1 consists of buffer components, peak 2 is eluted peptide indicated by * and peak 3 is cleaved thioredoxin and human rhinovirus 3C protease. (A) B3II, (B) B3DE, (C) B3NR, and (D) SDS–PAGE gel of peaks 1 and 3 from HPLC separation. (E) Mass spectra of HPLC peak 2, the purified peptide B3II, MW = 3119.1 Da.

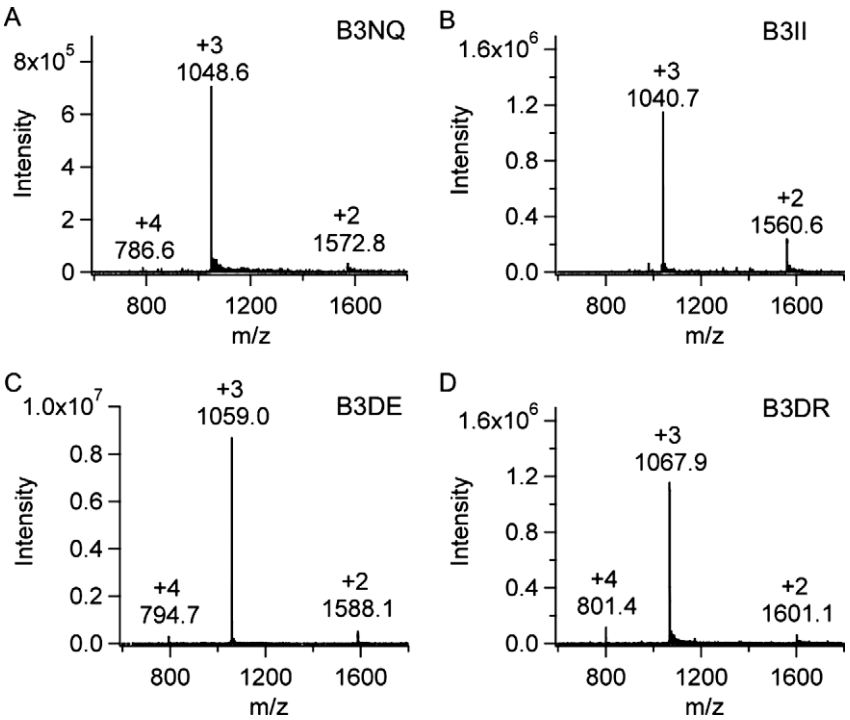

Fig. 7. Mass spectra of peptides. (A) B3NQ MW = 3143.2 Da, (B) B3II MW = 3119.1 Da, (C) B3DE MW = 3174.2 Da, and (D) B3DR MW = 3201.3. Spectra show double (+2) triply (+3), and quadruply (+4) charged ions.

mely similar to the desired product in size and chemical characteristics and are therefore difficult to separate based on chromatography. Peptides produced by biochemical synthesis by the *E. coli* ribosome are expected to be very homogeneous due to the ribosome's low error rate. In the system described here, any contaminants that remain after affinity purification and protease

Table 1
Peptide expression, yield, and chemical characteristics. Total peptide fusion expressed is compared to the overall yield of the peptide as a function of purification method used. Peptide characteristics such as charge, isoelectric point (IEP), the number of acidic, basic, polar, and non-polar amino acid residues are also tabulated.

Peptide	% Peptide in lysate	Purification method	Potential yield (mg)	Yield (%)	Charge	IEP	Acidic	Basic	Polar	Non-polar
aPP	26.1	Ni step	43	3.5	−2.15	4.75	5	3	12	9
B3	11.5	Ni/IE	2	1.0	−4.06	4.42	5	1	6	7
B3DQ	44.5	Ni step	N/A	N/A	−4.06	4.49	5	1	6	7
B3DE*	45.0	Ni step	50	5.0	−4.05	4.62	5	1	6	7
B3DK	45.9	Ni/IE	14	0.8	−2.06	5.40	4	2	6	7
B3DR	9.3	Ni/IE	5	1.4	−2.06	5.40	4	2	6	7
B3NQ	25.4	Ni/IE	11	1.1	−1.06	6.00	2	1	9	7
B3NE	20.9	Ni/IE	28	2.3	−2.06	5.33	3	1	8	7
B3NK	11.5	Ni/IE	8	1.0	−0.06	6.92	2	2	8	7
B3NR	24.5	Ni gradient	20	1.7	−0.06	6.92	2	2	8	7
B3II*	45.0	Ni/IE	70	5.8	−4.23	3.77	5	1	6	9
B3IQ	32.2	Ni gradient	31	3.9	−3.23	4.10	4	1	7	9
B3IE	35.9	Ni gradient	32	16.0	−4.23	3.98	5	1	6	9
B3IK	36.1	Ni gradient	25	2.5	−2.23	4.49	4	2	6	9
B3IR	37.6	Ni gradient	35	2.9	−2.23	4.49	4	2	6	9
B3IL	30.8	Ni gradient	40	4.0	−3.23	4.10	4	1	6	10
3INQ	27.0	Ni gradient	30	3.0	−1.24	4.53	2	1	9	9
3INE	36.6	Ni gradient	28	2.8	−2.23	4.25	3	1	8	9
3INK	19.2	Ni gradient	25	2.5	−0.24	6.14	2	2	8	9
3INR	29.7	Ni gradient	33	3.3	−0.24	6.14	2	2	8	9

Percent peptide in raw *E. coli* lystates was calculated using GeneTools (SynGene). *Indicates % peptide calculated was estimated by inspection of the stained SDS–PAGE gels. Ni, nickel-affinity chromatography; IE, ionic exchange chromatography; acidic, basic, polar and non-polar represent the number of amino acids in the 27-residue peptide that have the particular chemical characteristic.

cleavage of the biochemically synthesized peptide-fusions are unlikely to be chemically similar to the peptides. Our method for production of peptides uses reverse-phase HPLC purification as the final step. Because the non-peptide components are not chemically related to the peptides themselves, HPLC purification allows for simultaneous removal of the fusion partner, the protease, and any residual components of the buffer solution. This makes a simple, two-step purification possible. In addition, because of the high fidelity of biochemical synthesis, the homogeneity of our peptides is very impressive. This results in a library of extremely high purity.

All of the peptide-fusion proteins described here were purified using a six-histidine affinity tag. Following nickel-affinity chromatography using a step gradient purification the peptide-fusions were approximately 50% pure. We probed the effect of more extensively purifying the peptide-fusions to >95% purity by subsequent anion-exchange chromatography or to 80% purity by using a linear gradient of imidazole on the nickel-affinity column. Employing reverse-phase HPLC, we were able to robustly purify the cleaved peptide to homogeneity after either of these protocols. This finding significantly shortened our production time by obviating the ion-exchange purification step.

An additional advantage of the method described is that it is unnecessary for the peptide and the protease to be tagged with the same affinity tag, as is the case in some commercial protein-fusion purification systems. The method described here is compatible with any fusion partner and any protease. Since both the fusion partner and the protease are likely to be much larger than peptide products, it will be possible to separate the product from the contaminants in the final reverse-phase HPLC purification step. This fact is of particular importance in production of peptide libraries where additional amino acids at either the N- or C-termini can have significant effects on the overall properties of the peptide. Since any protease can be used, the overhanging amino acids from the protease cleavage site can be matched with the desired peptide sequence, or a protease that does not leave any overhang can be selected [37]. The fact that this method is not dependent on use of any particular affinity tags or on utilizing matching affinity tags also limits sub-cloning steps that are necessary to use the library production method.

This method of peptide production using genetically encoded peptides offers clear advantages in the length and fidelity of peptides that can be produced and the resultant purity of the final samples. In the field of chemical peptide synthesis, the concept of "difficult" sequences (e.g. hydrophobic sequences) exists. Conversely, the genetically encoded production method described here was insensitive to peptide sequence such that every sequence attempted worked the first time with no optimization of expression or purification procedure required. In order to make this method truly competitive with chemical synthesis, it is essential that libraries can be produced on the same general time scale as chemical synthesis. The scheduling strategy for the 20-member library we present here lays the foundation for production of much larger libraries using the same processes. To this end, we have time-optimized our expression, purification and characterization protocols. Using the protocols we report and have routinely executed, we have constructed a map for methodological synchronization (Fig. 8). By coordinating mutagenesis cycles, sequencing, transformation, expression, purification, and characterization, this peptide library can be produced rapidly. We generated this map reflecting the methods that were actually used, however, we have also performed sequencing reactions after an 8-h growth of cultures followed by mini-prepping of DNA for sequencing. If this were applied to each peptide construct, it would save one day from each segment of the mutagenesis portion of the plan. In addition, we have developed methodology so that Quikchange® mutagenized plasmids can be directly transformed into an expression strain of bacteria (such as BL21(DE3)) rather than into a cloning strain, which can save one step in the protocol. With these developments, the entire library could be generated even more rapidly. As presented, this methodological synchronization scheme allows construction of 20 peptide-expressing plasmids and purification of the resultant peptides to homogeneity in a 5-week period using one chromatography system, one HPLC and one liquid chromatography–mass spectrometry (LC–MS).[1]

[1] *Abbreviations used:* LC–MS, liquid chromatography–mass spectrometry; HPLC, high-performance liquid chromatography

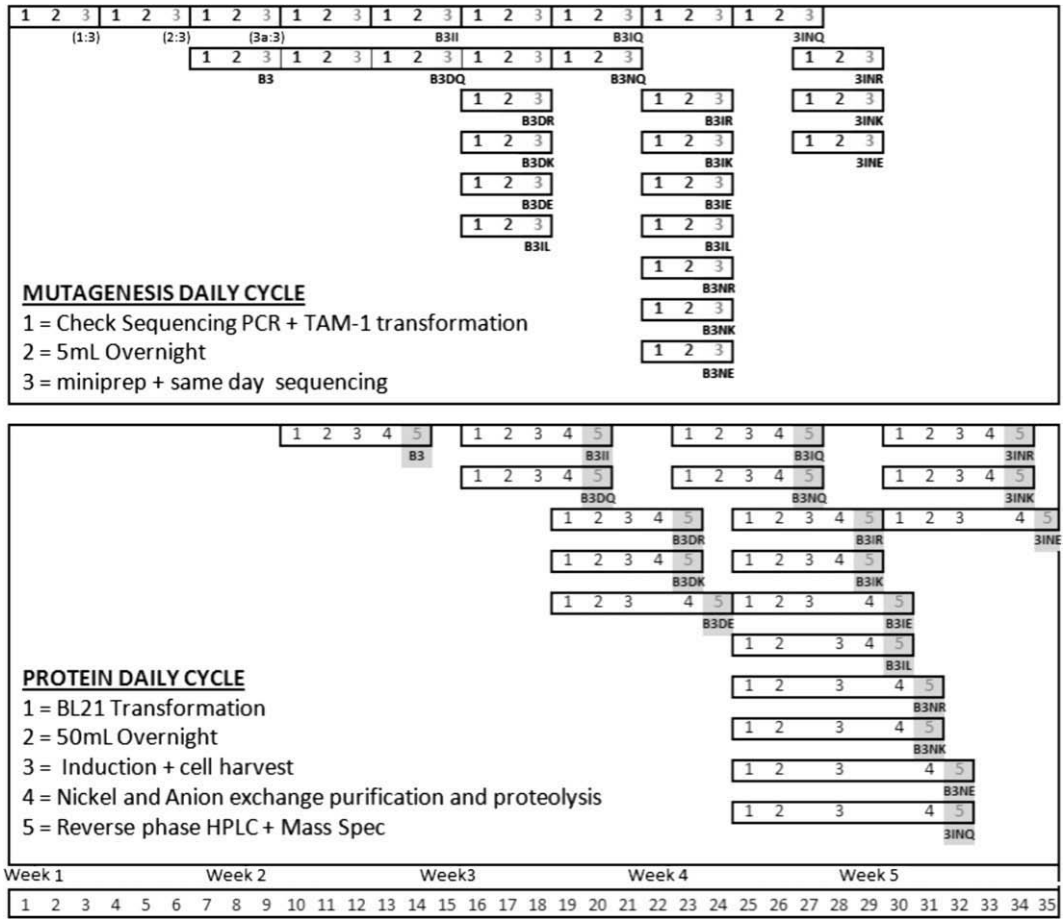

Fig. 8. Methodological Synchronization. An experimental map showing how synchronization of Quikchange® based mutagenesis and protein preparatory procedures allows the production of 20 peptides in a 5-week time frame. Mutagenesis daily cycle (top) outlines the number of days in which cloning steps such as Quikchange® mutagenesis, transformation, DNA purification and sequencing proceeds. 1:3, 2:3, and 3a:3 represent mutagenesis steps to create first peptide construct B3. Protein daily cycle (bottom) outlines the number of days in which expression, purification, and characterization for each peptide proceeds. The resultant DNA constructs are represented in light gray and purified peptides are represented in dark gray.

In this work, we have used no more than two liters of *E. coli* culture per peptide and have accepted losses at the affinity purification step. Nevertheless, the protocols discussed here could facilely be scaled up to produce even greater yields of pure peptide. For some of the peptide-fusions we have constructed, the yield of the expressed fusion protein as a fraction of total *E. coli* proteins is high enough (45% of total *E. coli* protein) that we can nearly envision skipping purification of the fusion protein altogether. By co-expressing the protease and the fusion protein it may be possible to perform a one-step purification by HPLC to yield large quantities of pure peptide.

Acknowledgments

This work was supported by NSF Chemical Innovation Center for Fueling the Future CHE-0739227, National Research Service Award T32 GM08515 from the National Institute of Health (K.L.H.), and NSF Award EEC-0649041 (K.D.O.). We thank Alana Schepartz and Doug Daniels for kindly providing the synthetic gene of avian pancreatic polypeptide (Yale University New Haven, CT, USA) and Dr. Sumana Ghosh for helpful discussions.

References

[1] R.L. Winston, J.M. Gottesfeld, Rapid identification of key amino-acid–DNA contacts through combinatorial peptide synthesis, Chemistry & Biology 7 (2000) 245–251.

[2] M. Rodriguez, S.S.C. Li, J.W. Harper, Z. Songyang, An oriented peptide array library (OPAL) strategy to study protein–protein interactions, Journal of Biological Chemistry 279 (2004) 8802.

[3] M.C. Sweeney, A.S. Wavreille, J. Park, J.P. Butchar, S. Tridandapani, D. Pei, Decoding protein–protein interactions through combinatorial chemistry: sequence specificity of SHP-1, SHP-2, and SHIP SH2 domains, Biochemistry 44 (2005) 14932–14947.

[4] S. Baltrusch, S. Lenzen, D.A. Okar, A.J. Lange, M. Tiedge, Characterization of glucokinase-binding protein epitopes by a phage-displayed peptide library identification of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase as a novel interaction partner, Journal of Biological Chemistry 276 (2001) 43915–43923.

[5] J.R.S. Newman, A.E. Keating, Comprehensive Identification of Human bZIP Interactions with Coiled-Coil Arrays, American Association for the Advancement of Science, 2003. pp. 2097–2101.

[6] Y. Choe, F. Leonetti, D.C. Greenbaum, F. Lecaille, M. Bogyo, D. Bromme, J.A. Ellman, C.S. Craik, Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities, Journal of Biological Chemistry 281 (2006) 12824.

[7] D.N. Gosalia, C.M. Salisbury, J.A. Ellman, S.L. Diamond, High throughput substrate specificity profiling of serine and cysteine proteases using solution-phase fluorogenic peptide microarrays*, Molecular & Cellular Proteomics 4 (2005) 626–636.

[8] D.N. Gosalia, C.M. Salisbury, D.J. Maly, J.A. Ellman, S.L. Diamond, Profiling serine protease substrate specificity with solution phase fluorogenic peptide microarrays, Proteomics 5 (2005) 1292–1298.

[9] J.L. Harris, B.J. Backes, F. Leonetti, S. Mahrus, J.A. Ellman, C.S. Craik, Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, Proceedings of the National Academy of Sciences of the United States of America 1 (2000) 7754–7759.

[10] B.E. Turk, L.L. Huang, E.T. Piro, L.C. Cantley, Determination of protease cleavage site motifs using mixture-based oriented peptide libraries, Nature Biotechnology 19 (2001) 661–667.

[11] D. Cuerrier, T. Moldoveanu, P.L. Davies, Determination of peptide substrate specificity for {micro}-calpain by a peptide library-based approach: the importance of primed side interactions, Journal of Biological Chemistry 280 (2005) 40632.
[12] M. Garaud, D. Pei, Substrate profiling of protein tyrosine phosphatase PTP1B by screening a combinatorial peptide library, Journal of the American Chemical Society 129 (2007) 5366–5367.
[13] G. Huyer, J. Kelly, J. Moffat, R. Zamboni, Z. Jia, M.J. Gresser, C. Ramachandran, Affinity selection from peptide libraries to determine substrate specificity of protein tyrosine phosphatases, Analytical Biochemistry 258 (1998) 19–30.
[14] S.W. Vetter, Z.Y. Zhang, Probing the phosphopeptide specificities of protein tyrosine phosphatases, SH2 and PTB domains with combinatorial library methods, Current Protein and Peptide Science 3 (2002) 365–397.
[15] J.E. Hutti, E.T. Jarrell, J.D. Chang, D.W. Abbott, P. Storz, A. Toker, L.C. Cantley, B.E. Turk, A rapid method for determining protein kinase phosphorylation specificity, Nature Methods 1 (2004) 27–29.
[16] A.S. Mah, A.E.H. Elia, G. Devgan, J. Ptacek, M. Schutkowski, M. Snyder, M.B. Yaffe, R.J. Deshaies, Substrate specificity analysis of protein kinase complex Dbf2-Mob1 by peptide library and proteome array screening, BMC Biochemistry 6 (2005) 22.
[17] B.E. Turke, J.E. Hutti, L.C. Cantley, Determining protein kinase substrate specificity by parallel solution-phase assay of large numbers of peptide substrates, Nature Protocols 1 (2006) 375–379.
[18] M.B. Yaffe, Study of substrate specificity of MAPKs using oriented peptide libraries, Methods in Molecular Biology 250 (2004) 237–250.
[19] V.V. Bartsevich, R.L. Juliano, Regulation of the MDR1 gene by transcriptional repressors selected using peptide combinatorial libraries, Molecular Pharmacology 58 (2000) 1–10.
[20] M. El-Mousawi, L. Tchistiakova, L. Yurchenko, G. Pietrzynski, M. Moreno, D. Stanimirovic, D. Ahmad, V. Alakhov, A vascular endothelial growth factor high affinity receptor 1-specific peptide with antiangiogenic activity identified using a phage display peptide library*, Journal of Biological Chemistry 278 (2003) 46681–46691.
[21] A. Jahnke, T. Wilmes, S. Adebahr, D. Pfeifer, T. Berg, M. Trepel, Leukemia targeting ligands isolated from phage display peptide libraries, Leukemia 21 (2007) 411.
[22] T. Obata, M.B. Yaffe, G.G. Leparc, E.T. Piro, H. Maegawa, A. Kashiwagi, R. Kikkawa, L.C. Cantley, Peptide and protein library screening defines optimal substrate motifs for AKT/PKB, Journal of Biological Chemistry 275 (2000) 36108–36115.
[23] N.L. Benoiton, Chemistry of Peptide Synthesis, Taylor & Francis, 2006.
[24] G.K.E. Scriba, Separation of Peptide Diastereomers, in: R.A. Meyers (Ed.), Encyclopedia of Analytical Chemistry: Instrumentation and Applications, John Wiley & Sons, Ltd., Chichester, 2000, pp. 5931–5946.
[25] J.P. Tam, Y.A. Lu, Coupling difficulty associated with interchain clustering and phase transition in solid phase peptide synthesis, Journal of the American Chemical Society 117 (1995) 12058–12063.
[26] S.B.H. Kent, Difficult sequences in stepwise peptide synthesis: common molecular origins in solution and solid phase, peptides: structure and function, in: C.M. Deber, V.J. Hruby, K.D. Kopple (Eds.), Proceedings of the 9th Annual American Peptide Symposium, Peirce Chem. Co., Rockford, IL, 1985, pp. 407–414.
[27] J.P. Tam, Enhancement of coupling efficiency in solid phase peptide synthesis by elevated temperature, peptides: structure and function, in: C.M. Deber, V.J. Hruby, K.D. Kopple (Eds.), Proceedings of the 9th Annual American Peptide Symposium, Peirce Chem. Co., Rockford, IL, 1985, pp. 423–425.
[28] R.F. Weaver, Molecular Biology, McGraw-Hill College, 2007.
[29] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase, Gene 67 (1988) 31–40.
[30] C. Guan, P. Li, P.D. Riggs, H. Inouye, Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein, Gene 67 (1988) 21–30.
[31] E.R. LaVallie, E.A. DiBlasio, S. Kovacic, K.L. Grant, P.F. Schendel, J.M. McCoy, A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm, Bio/Technology 11 (1993) 187–193.
[32] T.L. Blundell, J.E. Pitts, I.J. Tickle, S.P. Wood, C.W. Wu, X-ray analysis (1.4-angstrom resolution) of avian pancreatic polypeptide: small globular protein hormone, Proceedings of the National Academy of Sciences of the United States of America 78 (1981) 4175–4179.
[33] K. Tonan, Y. Kawata, K. Hamaguchi, Conformations of isolated fragments of pancreatic polypeptide, Biochemistry 29 (1990) 4424–4429.
[34] N.J. Zondlo, A. Schepartz, Highly specific DNA recognition by a designed miniature protein, Nature (London) 363 (1993) 38.
[35] J.W. Chin, R.M. Grotzfeld, M.A. Fabian, A. Schepartz, Methodology for optimizing functional miniature proteins based on avian pancreatic polypeptide using phage display, Bioorganic & Medicinal Chemistry Letters 11 (2001) 1501–1505.
[36] N. Shimba, A.M. Nomura, A.B. Marnett, C.S. Craik, Herpesvirus protease inhibition by dimer disruption, Journal of Virology 78 (2004) 6657–6665.
[37] B. Feeney, E.J. Soderblom, M.B. Goshe, A.C. Clark, Novel protein purification system utilizing an N-terminal fusion protein and a caspase-3 cleavable linker, Protein Expression and Purification 47 (2006) 311–318.