Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6705–6710, June 1998
Biochemistry

Expressed protein ligation: A general method for protein engineering

TOM W. MUIR*†‡, DOLAN SONDHI†§, AND PHILIP A. COLE†‡§

Laboratories of *Synthetic Protein Chemistry and §Bioorganic Chemistry, Rockefeller University, 1230 York Avenue, New York, NY 10021

Communicated by Samuel J. Danishefsky, Memorial Sloan–Kettering Cancer Center, New York, NY, April 1, 1998 (received for review January 27, 1998)

ABSTRACT **A protein semisynthesis method—expressed protein ligation—is described that involves the chemoselective addition of a peptide to a recombinant protein. This method was used to ligate a phosphotyrosine peptide to the C terminus of the protein tyrosine kinase C-terminal Src kinase (Csk). By intercepting a thioester generated in the recombinant protein with an N-terminal cysteine containing synthetic peptide, near quantitative chemical ligation of the peptide to the protein was achieved. The semisynthetic tail-phosphorylated Csk showed evidence of an intramolecular phosphotyrosine-Src homology 2 interaction and an unexpected increase in catalytic phosphoryl transfer efficiency toward a physiologically relevant substrate compared with the non-tail-phosphorylated control. This work illustrates that expressed protein ligation is a simple and powerful new method in protein engineering to introduce sequences of unnatural amino acids, posttranslational modifications, and biophysical probes into proteins of any size.**

The ability to alter protein structure and function by introducing unnatural amino acids has great potential to enhance our understanding of proteins, generate new tools for biomedical research, and create novel therapeutic agents (1–6). For example, site-specific incorporation of posttranslational modifications such as phosphoamino acids could lead to molecules having altered structural and functional properties that would be useful tools in understanding complex processes such as signal transduction. Equally, the introduction of small molecules like fluorescent tags, spin resonance probes, or cross-linking agents into proteins at well-defined positions could lead to new biochemical insights in countless biological systems. Although powerful, each of the currently existing techniques to introduce unnatural molecules into proteins has associated with it certain synthetic or practical limitations that have limited their widespread application (1–6).

In so-called "native chemical ligation" an N-terminal cysteine-containing peptide is chemically ligated to a second peptide possessing an $^{\alpha}$thioester group with the resultant formation of a native peptide bond at the ligation junction (5). Until now, peptide $^{\alpha}$thioesters for use in native chemical ligation have been generated solely through chemical synthesis, a technically demanding process that imposes certain size constraints on the technique. Thus, native chemical ligation has proven very useful for the total synthesis of small proteins and protein domains, but has not been extended to the synthesis of proteins beyond ≈15 kDa. One way to overcome this size limitation would be to chemically ligate small synthetic sequences to much larger recombinant protein fragments. Although methods for the generation of recombinant proteins possessing N-terminal cysteine residues have been reported (7), there is currently no way of producing recombinant protein $^{\alpha}$thioesters for use in native chemical ligation. Thus, native chemical ligation has not been completely interfaced with recombinant protein semisynthesis. The current challenge was therefore to devise a way of generating the necessary protein $^{\alpha}$thioesters by using a straightforward and widely available recombinant expression method.

Protein splicing, the process in which a protein undergoes an intramolecular rearrangement resulting in the extrusion of an internal sequence (intein) and the joining of the lateral sequences (exteins), has been shown to involve the intermediacy of a thioester (8, 9). A mutant version of the splicing protein has been demonstrated to be defective in completion of the splicing reaction but still capable of thioester intermediate formation (8, 9). The commercially available pCYB vectors for *Escherichia coli* protein expression result in the generation of $^{\alpha}$thioesters where a protein of interest can be expressed in frame fused with an intein-chitin binding domain (CBD) sequence (9). In the standard experiment, the protein of interest is cleaved from the intein-CBD with DTT or 2-mercaptoethanol by a transthioesterification reaction while the chimera is bound to a chitin column. We suspected that it might be possible to intercept the $^{\alpha}$thioester with a synthetic peptide rather than DTT, thereby generating a semisynthetic protein containing a chemoselectively ligated synthetic peptide sequence.

To investigate the feasibility of this ligation process we decided to undertake the synthesis of a pair of semisynthetic C-terminal Src kinase (Csk) proteins. This target system was chosen in part because it satisfied two principal objectives of this initial study, which were: (*i*) to demonstrate that intein-fusion constructs could be used as a source of large recombinant protein $^{\alpha}$thioesters for use in native chemical ligation, and (*ii*) to synthesize challenging molecules that could not have been generated any other way.

Csk is a 50-kDa protein that catalyzes the phosphorylation of a highly conserved tyrosine within the C-terminal tail of Src family members (10–17). This modification results in an intramolecular interaction between the Src homology 2 (SH2) domain and the C-terminal phosphotyrosine within Src, an association that leads to a significant conformational change and catalytic repression *in vivo* (Fig. 1*A*) (10–15). Csk is similar in architectural layout to Src family members and is 40–50% identical in amino acid sequence to this protein family (16, 17), however, it lacks a C-terminal tyrosine-containing tail and is not known to be regulated through phosphorylation (Fig. 1) (10, 11). We were therefore interested in the possibility of adding a phosphotyrosine tail to Csk and determining the

Abbreviations: Csk, C-terminal Src kinase; Csk^{PEP}, Csk-[CONH]-peptide (the ligation production of Csk and unphosphorylated peptide); Csk^{pPEP}, Csk-[CONH]-phosphopeptide (the ligation product of Csk and phosphopeptide); CBD, chitin binding domain; SH2, Src homology 2.

†T.W.M., D.S., and P.A.C. contributed equally to this work.
‡To whom reprint requests should be addressed. e-mail: cole@rockvax.rockefeller.edu and muirt@rockvax.rockefeller.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

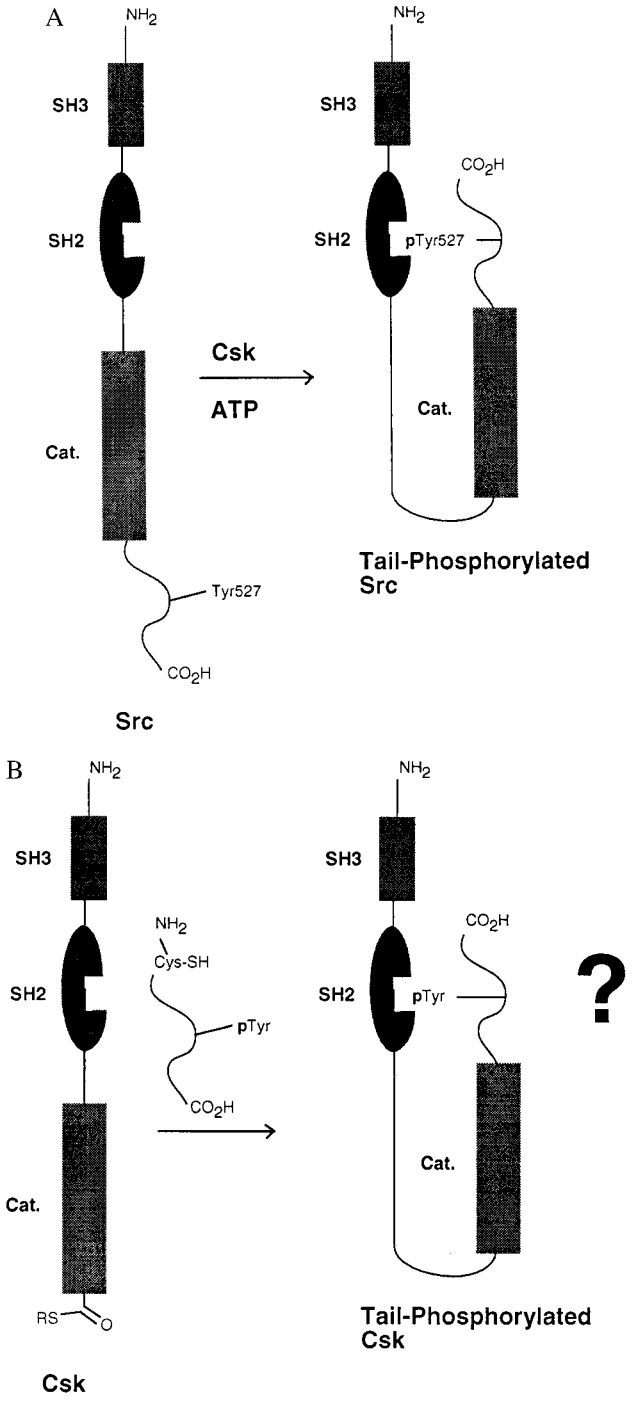

FIG. 1. Phosphotyrosine tails in Src and Csk. (*A*) Phosphorylation of the Src tail on tyrosine-527 is catalyzed by Csk. This phosphorylation results in a conformational change involving an intramolecular interaction between the Src tail and the SH2 domain. (*B*) Csk is highly homologous to Src but lacks a C-terminal tyrosine-containing tail. Proposed ligation of a phosphotyrosine tail might lead to a conformational change like that found in Src. Csk is not a Src family member and in addition to its lack of a C-terminal tail, Csk also lacks two key Src features: an activating tyrosine-containing loop, as well as an N-terminal myristoylation site (16, 17).

potential effects of this modification on conformation (Fig. 1*B*) and catalytic behavior.

Although extending the Csk C terminus could in principle be carried out using recombinant methods (18), there would be no way of directing specific phosphorylation of a single engineered tyrosine (wild-type Csk has 20 tyrosine residues) by chemical or enzymatic methods. Bacterial recombinant protein expression of wild-type human Csk is complex because it requires chaperonins for significant soluble expression (18). In fact, attempts at simply adding a 9-aa tyrosine-containing tail to wild-type Csk led to very poor protein expression using these established methods (P.A.C., unpublished data). Furthermore, none of the existing protein engineering techniques appeared to be suitable for the generation of this large and complex protein. For example, the current size limit of total chemical synthesis by native chemical ligation is 100 ± 50 aa, putting the 450-aa Csk well beyond the range of this approach. Similarly, the presence of 10 cysteine residues in wild-type Csk prohibits a site-selective cysteine-modification strategy. Thus, producing the desired semisynthetic Csk protein(s) in sufficient quantities for enzymatic and biochemical studies was a significant chemical challenge.

MATERIALS AND METHODS

Peptide Synthesis. Phosphorylated and unphosphorylated peptides were manually synthesized by fluorenylmethoxycarbonyl and *tert*-butoxycarbonyl solid phase peptide synthesis, respectively. Phosphotyrosine was introduced during fluorenylmethoxycarbonyl chain assembly in the phosphate unprotected form. Orthogonal protection of the ε-NH_2 group of the C-terminal Lys residue with either fluorenylmethoxycarbonyl (*tert*-butoxycarbonyl strategy) or dimethyldioxocyclohexylidene (fluorenylmethoxycarbonyl strategy) allowed direct attachment of fluorescein (activated as an *N*-hydroxysuccinimide ester) before the final cleavage step. After cleavage, peptides were purified to homogeneity by HPLC and characterized by electrospray MS.

Expressed Protein Ligation. See Fig. 2 for the overall strategy. Primers containing an *Nde*I site (upstream) and a *Sma*I site (downstream) were used to PCR amplify full-length wild-type human *csk* DNA for in frame insertion upstream of the intein/CBD encoding sequence in the vector pCYB2 (New England Biolabs). The resultant plasmid pCYB2-CSK, which was free of mutations in the Csk coding region based on DNA sequencing, was then cotransformed into *E. coli* DH5α with the GroES and GroEL expression plasmid pREP4-groESL using dual selection with ampicillin and kanamycin and cells grown and lysed as described (19). Chitin resin (1 ml) in a disposable plastic column was washed with 20 ml of equilibration buffer (25 mM Na·Hepes, pH 7.0/250 mM NaCl/1 mM Na·EDTA/0.1% Triton X-100). Cell lysate (10 ml, made 0.1% in Triton X-100) was passed through the column at a flow rate of 0.5 ml/min and the flow-through was reapplied at a similar rate. The column was then washed with 30 ml of equilibration buffer and treated with 2 ml 2% vol/vol thiophenol in equilibration buffer (minus Triton X-100) at 1 ml/min followed immediately by 1 ml of 2 mM peptide + 2% thiophenol + equilibration buffer (minus Triton X-100). After 24 h standing at 25°C, the column was eluted with equilibration buffer and the desired product appeared in the initial 2.5 ml; it was dialyzed (Dispodialyzer, 25 kDa cutoff, Fisher) against 25 mM Na·Hepes (pH 7.7), 2 mM DTT, 500 mM NaCl at 4°C for 4–5 days to remove unligated peptide. Protein was estimated to be ≈20% pure by SDS/PAGE with the principal contaminants shown to be GroEL (60%) and DnaK (20%) (see Fig. 3). Quantification of the semisynthetic Csk proteins was based on the intensity of Coomassie blue stained bands on SDS/PAGE compared with known amounts of wild-type Csk and relative amounts of Csk semisynthetic proteins further confirmed by quantitative fluorescence imaging (Storm, Molecular Dynamics). Expressed protein ligation reactions were carried out at least three times and gave reproducible yields. Purification of the proteins by phosphotyrosine affinity chromatography (see below) allowed chaperonin removal (Fig. 3) and showed that chaperonins did not effect kinase activity of either Csk-

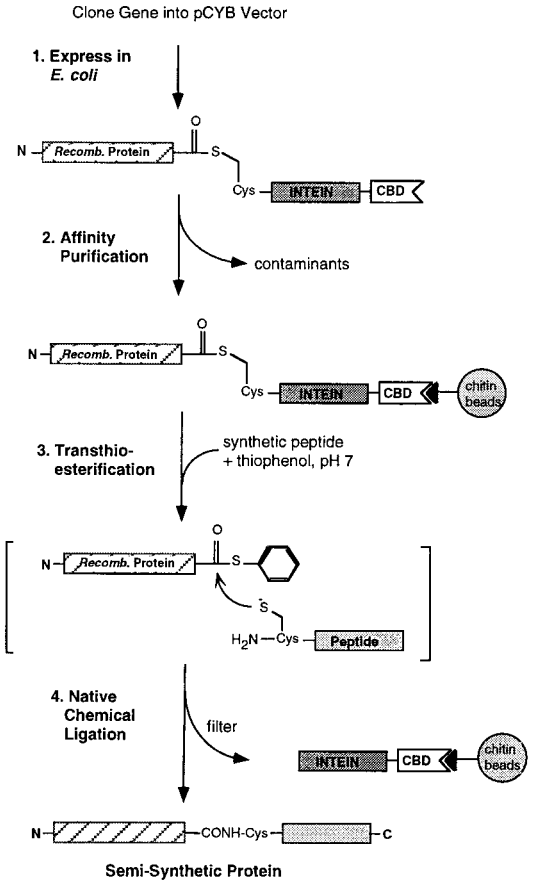

FIG. 2. The principle of expressed protein ligation. In the first step, the gene or gene fragment is cloned into the commercially available PCYB2-IMPACT vector (New England Biolabs) by using the *Nde*I and *Sma*I restriction sites. This cloning strategy results in the addition of a Pro-Gly appended to the native C terminus of the protein of interest. The presence of a C-terminal glycine has been shown to accelerate native chemical ligation (24) and thus reduces the chance of side reactions. After expression and affinity purification of the fusion protein by binding to the chitin resin, the chemical ligation step is initiated by incubating the resin-bound protein with thiophenol and synthetic peptide in buffer. This results in the *in situ* generation of a highly reactive phenyl $^{\alpha}$thioester derivative of the protein that then rapidly ligates with the synthetic peptide to afford the desired semisynthetic protein.

[CONH]-peptide (Csk^{PEP}) or Csk-[CONH]-phosphopeptide (Csk^{pPEP}). Complete characterization of the proteins by MS was carried out after reverse-phase HPLC purification (Fig. 3*C*).

Kinase Assays with Poly(Glu,Tyr). These were performed as described (20) where transfer of ^{32}P from [γ-^{32}P]ATP to poly(Glu,Tyr) was monitored. Autophosphorylation of the semisynthetic Csk ligation proteins was shown to be insignificant. Kinetic parameters were as follows: Csk^{PEP}, K_m of ATP = 31 ± 2 μM, K_m of poly(Glu,Tyr) = 19 ± 3 μg/ml, k_{cat} = 17 ± 1 min^{-1}; Csk^{pPEP}, K_m of ATP = 34 ± 11 μM, K_m of poly(Glu,Tyr) = 30 ± 3 μg/ml, k_{cat} = 19 ± 1 min^{-1}; wild-type Csk (20), K_m of ATP = 12 ± 1 μM, K_m of poly(Glu,Tyr) = 48 ± 2 μg/ml, k_{cat} = 40 ± 5 min^{-1}.

Phosphotyrosine Affinity Chromatography. This was performed in a manner analogous to procedures described (19, 21). A preequilibrated 8 × 8 mm column of *O*-phospho-L-tyrosine-agarose resin was eluted over 5 min with 1 ml of a semisynthetic protein solution (≈1 μM) and eluted at 0.5 ml/min with eight column volumes of low salt buffer (25 mM Hepes, pH 7.7/2 mM DTT) followed by eight column volumes of high salt buffer (25 mM Hepes, pH 7.7/2 mM DTT/500 mM NaCl). The material eluted under high salt conditions was quantitated by fluorescence imaging on SDS/PAGE. The amount bound in the case of Csk^{PEP} was 50 ± 10% and that for Csk^{pPEP} was 10 ± 4%. Binding studies gave similar results on two separate occasions.

Gel Filtration. Gel filtration of semisynthetic proteins was carried out on a Superdex-75 column (Pharmacia) in 20 mM Tris·acetate (pH 8.0) at 0.5 ml/min at room temperature using the proteins ribonuclease A (13,700 Da), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), and BSA (66 kDa) to generate a standard curve. Detection of the standards and wild-type Csk was done by monitoring UV absorbance at 280 nm and for the semisynthetic Csk proteins was done by monitoring fluorescence emission at 520 nm. The concentration of the semisynthetic Csk proteins during column loading was ≈1 μM. The calculated molecular masses were: wild-type Csk (50 kDa), Csk^{PEP} (54 kDa), and Csk^{pPEP} (54 kDa) with an estimated SE ± 10%.

RESULTS AND DISCUSSION

An intein-CBD expression plasmid containing full-length wild-type human Csk DNA was generated and coexpressed in *E. coli* along with GroES and GroEL. The soluble fraction was passed over chitin resin and the resin was washed and then treated with 50 mM DTT containing buffer overnight. This led to generation of full-length Csk protein. Kinase assay of the Csk generated in this manner showed that it was fully active. Treatment of the resin bound Csk-intein-chitin fusion with mercaptoacetic acid and cysteine also afforded comparable quantities of Csk (whereas treatment with *N*-acetylcysteine, for reasons that are unclear, gave no detectable yield of Csk).

Given these results, we designed and synthesized the tyrosine phosphorylated and unphosphorylated forms of the peptide NH_2-Cys-Glu-Asp-Asn-Glu-Tyr-Thr-Ala-Arg-Glu-aminocaproate-Lys-ε-[fluorescein]-CO_2H. Although not homologous to the Src tail, the sequence is derived from the highly conserved activating autophosphorylation site of Src family kinases (10, 11) and is of a similar length to the Src tail. In its tyrosine phosphorylated form, this sequence has been shown to bind specifically to the SH2 domain of Csk (21–23). In the projected semisynthetic protein, the number of amino acids between the highly conserved phosphotyrosine-interacting FLVRES motif of the Csk SH2 domain and the engineered phosphotyrosine in the tail is identical to the analogous spacing in the Src family member Lck. The N-terminal cysteine residue was included in the peptide to facilitate native chemical ligation. Incorporation of a C-terminal fluorescent tag via a flexible linker was envisaged to serve as a sensitive marker of successful ligation and as a probe for further biochemical studies.

Initial efforts to react the unphosphorylated peptide with the resin-bound Csk-intein-CBD fusion protein without added thiol cofactors were unsuccessful, and coaddition of mercaptoacetic acid led to cleavage of Csk from the fusion protein, without any detectable ligation. However, inclusion of 2% thiophenol in the reaction buffer led to extremely efficient (>90%) ligation of the synthetic peptide to the recombinant protein as evidenced by the production of a highly fluorescent 52-kDa protein band on SDS/PAGE (Figs. 2–4). As illustrated in Fig. 2, we hypothesize a two-step, one-pot process involving an initial transthioesterification event followed by immediate native chemical ligation. The initial transthioesterification step is critical as it alleviates any steric hindrance present around the fusion protein thioester, and creates a reaction sink involving the formation of a highly reactive phenyl $^{\alpha}$thioester derivative of the recombinant Csk protein (24). The corresponding mercaptoacetic acid thioester would be expected to be much less reactive than the phenyl thioester thus accounting for the results in our preliminary studies.

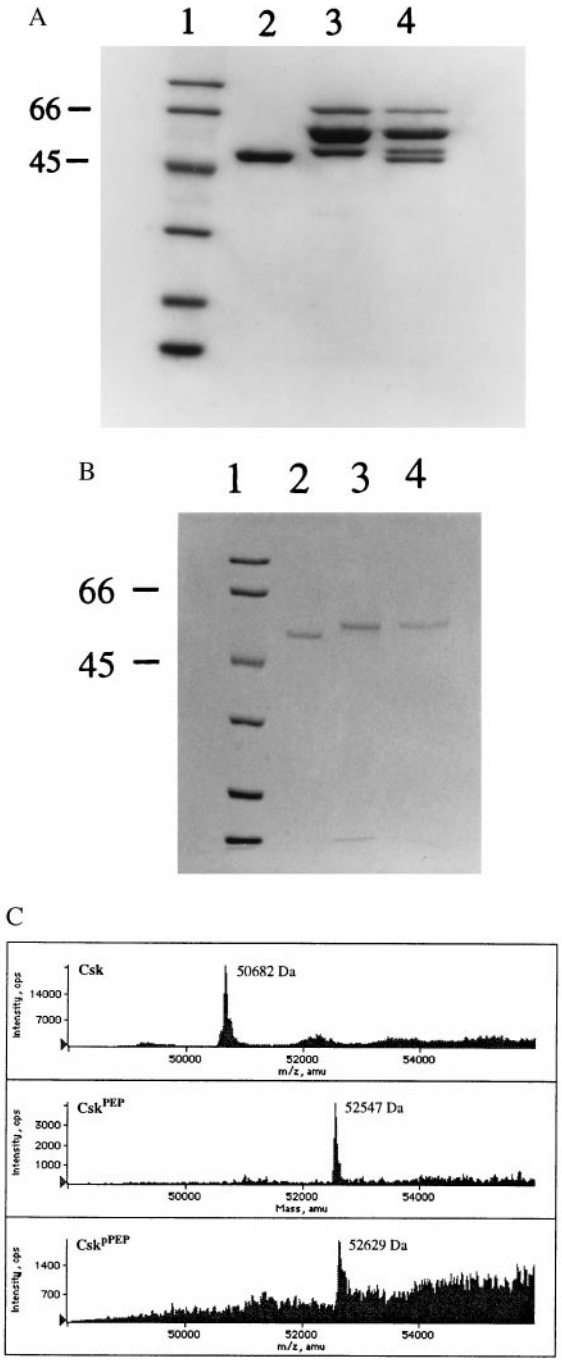

FIG. 3. Characterization of semisynthetic proteins. (*A*) A Coomassie-stained 10% SDS/PAGE gel of Csk^{PEP} crude reaction product mixture. Lane 1: molecular mass markers from the top: 97, 66, 45, 31, and 21.5 kDa. Lane 2: Wild-type Csk. Lane 3: Csk^{PEP} crude ligation product mixture; a combination of N-terminal sequencing and electrospray MS indicated that the bands at 56 kDa and 69 kDa were GroEL and DnaK, respectively. Lane 4: Comixture of wild-type Csk and Csk^{PEP} crude ligation product mixture. (*B*) A Coomassie-stained 10% SDS/PAGE gel of the purified semisynthetic Csk proteins. Lane 1: Molecular mass markers as listed in *A*. Lane 2: Wild-type Csk. Lane 3: Csk^{pPEP} purified by reverse-phase HPLC. Lane 4: Csk^{PEP} purified by phosphotyrosine affinity chromatography. (*C*) Characterization of semisynthetic proteins by electrospray MS. (*Top*) Full-length wild-type Csk, expected mass = 50,705 Da (average isotope). (*Middle*) Csk^{PEP}, expected mass = 52,540 Da (average isotope). (*Bottom*) Csk^{pPEP}, expected mass = 52,619 Da (average isotope). Each sample was isolated by reverse-phase HPLC and mass analyzed by using a Perkin–Elmer-Sciex (Thornhill, ON, Canada) API-100 mass spectrometer. Predicted masses were calculated by using the program MACBIOMASS (S. Vemuri and T. Lee, City of Hope, Duarte, CA). Note, the ligated Csk products were engineered to have the sequence Pro-Gly added to their C termini, and Edman sequencing indicated that the N-terminal methionine had been removed from the Csk expressed in the pCYB2 vector. That the ligation products contained only one N terminus (i.e., from Csk) combined with the MS data provides unambiguous characterization of the semisynthetic proteins.

In each of the two ligation reactions, the crude product mixtures were nearly free of unligated material as exemplified by lanes 3 and 4 of Fig. 3*A*, but both preparations were contaminated with GroEL and DnaK. Although yields were not optimized, an estimated 0.5 mg of ligation product per 2.5 l bacterial cell culture was produced. Ligation conditions were not disruptive to Csk protein folding because both semisynthetic constructs were catalytically active (see *Materials and Methods*). Further characterization of Csk^{PEP} and Csk^{pPEP} after chromatographic purification to remove the chaperonins demonstrates the slightly higher apparent molecular weights of the ligation products compared with wild-type Csk on SDS/PAGE (see Fig. 3*B*). The enzymologic activity of the purified material was identical to the chaperone-containing mixture. Electrospray MS of the purified semisynthetic proteins gave molecular masses in good agreement with the predicted values (Fig. 3*C*). Interestingly, Edman sequencing revealed that the N-terminal methionine residue in the protein produced as an intein-CBD fusion was completely removed (no such cleavage is observed when Csk is expressed in the standard form).

The Csk ligation products were subjected to extensive dialysis to remove unreacted peptide. After dialysis, affinity purification over a phosphotyrosine column (19, 21) was attempted for both Csk^{PEP} and Csk^{pPEP} to assess potential conformational differences. While about 50% of Csk^{PEP} bound to the phosphotyrosine resin under low salt conditions, comparable to wild-type Csk, only ≈10% Csk^{pPEP} bound to the phosphotyrosine resin under similar conditions. It is presumed that the SH2 domain of Csk^{pPEP} is less available for affinity column interaction because it is prebound to the phosphotyrosine sequence of Csk^{pPEP}. Similar behavior has been reported with Src family members (10, 21, 25).

In the case of Src family members, the interaction between the phosphotyrosine tail and the SH2 domain has been shown to be intramolecular (12, 13, 15). Nondenaturing PAGE (6%) with fluorescence imaging showed that Csk^{pPEP} had a slightly faster migration time compared with Csk^{PEP}, consistent with Csk^{pPEP} being monomeric with an increased electrostatic effect (data not shown). A nondenaturing PAGE of the Src family member Lck in its tail-phosphorylated and unphosphorylated form showed very similar behavior (25). Gel filtration further suggested that both the semisynthetic Csk proteins were monomeric (see *Materials and Methods*), evidence that the proposed interaction between the phosphotyrosine tail and SH2 domain in Csk^{pPEP} is intramolecular.

Limited proteolysis studies with subtilisin further suggested a conformational difference between Csk^{PEP} and Csk^{pPEP}, with Csk^{pPEP} showing a reproducibly slower proteolytic degradation rate as demonstrated by the persistent fluorescent bands on SDS/PAGE in Fig. 4. Varying the time and protease concentration failed to reveal the build-up of a fluorescent intermediate with molecular mass of >14 kDa for proteolysis of Csk^{PEP} suggesting that the fluorescent peptide tail was cleaved rapidly in this semisynthetic protein. The ≈38 kDa fragment produced in the proteolysis of Csk^{pPEP} clearly contains an intact C terminus because of its fluorescence, and is ≈7 kDa larger than the primary site of wild-type Csk cleavage under similar conditions, with the cleavage site in the latter at the SH2 domain-catalytic domain junction (data not shown). The position of cleavage in Csk^{pPEP} is nearer to the N-terminal edge of the SH2 domain. Interestingly, tail-phosphorylated and unphosphorylated forms of Src show distinct proteolytic

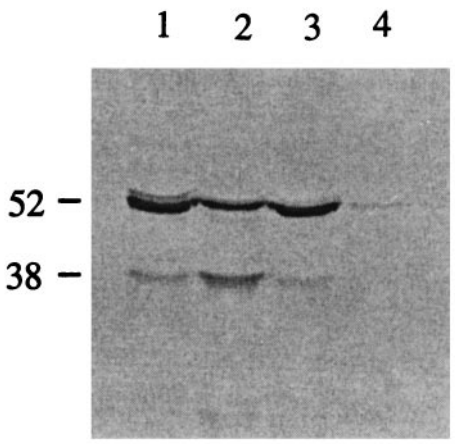

FIG. 4. Fluorescence imaging of an SDS/PAGE showing the results of proteolytic digestions of Csk^{PEP} and Csk^{pPEP} with subtilisin. Lane 1: Csk^{pPEP} − subtilisin; lane 2: Csk^{pPEP} + subtilisin; lane 3: Csk^{PEP} − subtilisin; lane 4: Csk^{PEP} + subtilisin. Reactions conditions: Csk^{PEP} and Csk^{pPEP} (1 μg) in 20 μl buffer (20 mM Tris·acetate, pH 8.0/10% glycerol/2 mM DTT) treated with subtilisin Carlsberg (12.5 ng) for 30 min at 4°C. Fluorescence imaging was done on a Storm instrument (Molecular Dynamics). Proteolysis reactions were performed in parallel under identical conditions with different protein preparations on four different occasions and all gave comparable results to those shown here.

degradation patterns, comparable to those of the semisynthetic Csk proteins (10, 26). In the case of Src, the overall proteolysis rate is reduced for the tail-phosphorylated form and the C-terminal tail region is particularly resistant to proteolysis when phosphorylated compared with the unphosphorylated form (26). It is important to emphasize that the fluorescein probe greatly simplified the analysis of this proteolysis data and that, in general, the ability to label the C terminus of proteins with fluorescent markers using expressed protein ligation should be of significant value in protein mapping and footprinting experiments.

In sum, the phosphotyrosine affinity, nondenaturing PAGE, gel filtration, and proteolysis results support the proposition that appending a phosphotyrosine tail to Csk results in a new conformation involving an intramolecular interaction between the SH2 domain and the tail phosphotyrosine. Such a conformational switch could lead to new biological activities in cell signal transduction. As a first stage toward investigating this possibility, the catalytic behavior of the semisynthetic Csk proteins was investigated using both peptide and protein substrates.

The chemical mechanism of peptide and protein substrate phosphorylation by Csk appears to be similar but the basis for selectivity is different for the two substrate classes (25). Specifically, the interaction of Csk with its physiologic substrates appears to require multidomain structural interactions, so that it could be expected that a perturbed Csk conformation might well lead to altered catalytic effects with a Src family substrate. This is indeed observed to be the case, but in contrast to expectations based on the Src model, Csk^{pPEP} is significantly more active than Csk^{PEP} in phosphorylating the physiologically relevant substrate Lck (Fig. 5 *A* and *B*). Both phosphorylation reactions showed linear activity as a function of time and enzyme concentration (Fig. 5 *A* and *B*). At low Lck substrate concentration, the rate of phosphorylation by the Csk^{pPEP} is ≈5-fold *higher* compared with Csk^{PEP}. This is likely to be primarily a specific Lck-binding difference to the two semisynthetic Csk proteins because the difference in the rates becomes less at higher Lck concentration (see Fig. 5*C*) and because the kinetics with the peptide substrate poly(Glu,Tyr) were so similar for Csk^{PEP} and Csk^{pPEP} (see *Materials and Methods*).

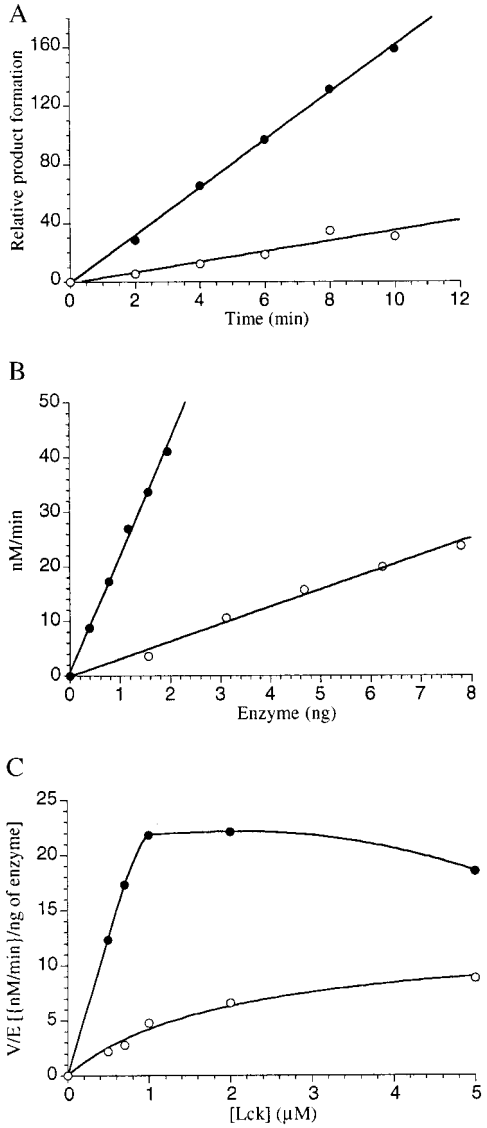

FIG. 5. Phosphorylation of Lck catalyzed by semisynthetic Csk proteins. (*A*) Time course of Csk^{PEP} and Csk^{pPEP}-catalyzed phosphorylation of Lck. ●, Csk^{pPEP}; ○, Csk^{PEP}. Relative product formation is corrected for the amount of Csk^{PEP} and Csk^{pPEP} utilized. [Lck] = 1 μM. (*B*) Kinase activity vs. Csk^{PEP} and Csk^{pPEP} concentration with the substrate Lck. ●, Csk^{pPEP}; ○, Csk^{PEP}. [Lck] = 1 μM. Reaction time = 2 min. (*C*) Rate of phosphorylation of Lck catalyzed by Csk^{PEP} and Csk^{pPEP} vs. Lck concentration. ●, Csk^{pPEP}; ○, Csk^{PEP}. The data for Csk^{pPEP} could not be strictly fit to standard Michaelis-Menten kinetics because of apparent substrate inhibition. For Csk^{pPEP} the K_m (apparent) (Lck) estimated as the substrate concentration necessary for 50% maximal velocity is 0.4 ± 0.1 μM. The data for Csk^{PEP} was fit to the Michaelis-Menten equation and gave K_m (apparent) (Lck) = 2 ± 0.5 μM. The Lck K_m with wild-type Csk enzyme under these conditions is 3–5 μM (25). All assays were performed as described (25) on a 15 μl scale at 30°C by using catalytically impaired Lck (N-terminal 63 aa deleted, containing a K273R mutation) at fixed and near-saturating ATP concentration [10 μM, ≈3 × K_m (apparent)] and optimal $MnCl_2$ concentration (5 mM) under initial conditions (<10% turnover of the limiting substrate). All assays were carried out at least two times and duplicate points generally agreed within 10%. Kinase activity was shown to be identical for the phosphotyrosine-affinity purified and unpurified semisynthetic proteins by using Lck as substrate. Kinase specificity for the 505-tyrosine of Lck was confirmed by demonstrating that 505-phosphorylated Lck was not detectably phosphorylated by either semisynthetic protein using conditions described (25).

It should also be noted that the repression of Src kinase activity in its tail phosphorylated form involves a complex

network of intramolecular interactions within the Src protein (13, 15, 27), and Csk appears to lack the ability to recapitulate this inhibited state.‖ Although the structural basis for increased kinase activity of tail-phosphorylated Csk on a physiologic substrate is not yet understood,** these results allow the prediction that conformational engineering using expressed protein ligation should give rise to interesting effects in signal transduction.

In this paper, we have introduced a method in protein engineering. Our approach ''expressed protein ligation'' is very simple, involving a single chemical step, and effectively unites the fields of synthetic peptide chemistry and recombinant protein biotechnology. In doing so, it allows proteins to be subjected to the same level of systematic chemical investigation previously restricted to the study of small bioactive peptides. In addition to the work described in this report, we have now successfully applied expressed protein ligation to two other polypeptide systems, including the entire sigma subunit of bacterial RNA polymerase (28). Furthermore, N-terminal methionine deletion that has been observed in the several cases using the intein expression vector creates the possibility for ligation of peptides or proteins to an N-terminal cysteine (placed at the second codon of the recombinant protein). This may further expand the utility of expressed protein ligation as a means of inserting unnatural or isotopically labeled amino acids in the middle as well as at the ends of recombinant proteins.

‖Repression of Src largely requires an unphosphorylated activating loop in addition to tail phosphorylation. A major feature of Src repression appears to be that tail phosphorylation results in a conformational change that prevents autophosphorylation of the activating loop (14, 27). Because Csk does not have an activating loop it would have been impossible to predict the results on Csk's catalytic activity resulting from the conformational change due to tail phosphorylation. However, these results implicitly argue that one or more structural features that lead to catalytic repression in tail-phosphorylated Src is/are absent in tail-phosphorylated Csk.

**The simple model where the increased catalytic efficiency results from an interaction of the Lck-SH2 domain with the phosphotyrosine tail of Csk^{pPEP} appears unlikely because the addition of the peptide NH_2-Cys-Glu-Asp-Asn-Glu-phosTyr-Thr-Ala-Arg-Glu-aminocaproate-Lys-ε-[fluorescein]-CO_2H (250 μM) had minimal effect on either Csk^{pPEP} or Csk^{PEP} catalyzed Lck phosphorylation.

We are grateful to G. Cotton, I. Moarefi, K. Severinov, J. Kuriyan, W. Xu, M. Eck, S. Burley, D. Cowburn, R. Alani, M.-Q. Xu, and F. Perler for advice and assistance, and to the Damon Runyon Scholars Award Program (P.A.C), the Irving A. Hansen Memorial Foundation (P.A.C.), the National Institutes of Health (T.W.M. and P.A.C.), and the Pew Scholars Program in the Biomedical Sciences (T.W.M.) for financial support.

1. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. (1989) *Science* **244,** 182–188.
2. Wallace, C. J. A. (1995) *Curr. Opin. Biotechnol.* **6,** 403–410.
3. Witte, K., Sears, P., Martin, R. & Wong, C.-H. (1997) *J. Am. Chem. Soc.* **119,** 2114–2118.
4. Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J. & Wells, J. A. (1994) *Science* **266,** 243–247.
5. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. (1994) *Science* **266,** 776–779.
6. Muir, T. W., Dawson, P. E. & Kent, S. B. H. (1997) *Methods Enzymol.* **289,** 266–298.
7. Erlanson, D. A., Chytil, M. & Verdine, G. L. (1996) *Chem. Biol.* **3,** 981–991.
8. Xu, M.-Q. & Perler, F. B. (1996) *EMBO J.* **15,** 5146–5153.
9. Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., *et al.* (1997) *Gene* **192,** 271–281.
10. Brown, M. T. & Cooper, J. A. (1996) *Biochim. Biophys. Acta* **1287,** 121–149.
11. Superti-Furga, G. & Courtneidge, S. A. (1995) *BioEssays* **17,** 321–330.
12. Boerner, R. J., Kassel, D. B., Barker, S. C., Ellis, B., DeLacy, P. & Knight, W. B. (1996) *Biochemistry* **35,** 9519–9525.
13. Sicheri, F., Moarefi, I. & Kuriyan, J. (1997) *Nature (London)* **385,** 602–609.
14. Moarefi, I., LaFevre-Bernt, M., Sicheri, F., Huse, M., Lee, C.-H., Kuriyan, J. & Miller, W. T. (1997) *Nature (London)* **385,** 650–653.
15. Xu, W., Harrison, S. C. & Eck, M. J. (1997) *Nature (London)* **385,** 595–602.
16. Nada, S., Okada, M., MacAuley, A., Cooper, J. A. & Nakagawa, H. (1991) *Nature (London)* **351,** 69–72.
17. Partanen, J., Armstrong, E., Bergman, M., Makela, T. P., Hirvonen, H., Huebner, K. & Alitalo, K. (1991) *Oncogene* **6,** 2013–2018.
18. Cole, P. A. (1996) *Structure (London)* **4,** 239–242.
19. Grace, M. R., Walsh, C. T. & Cole, P. A. (1997) *Biochemistry* **36,** 1874–1881.
20. Cole, P. A., Burn, P., Takacs, B. & Walsh, C. T. (1994) *J. Biol. Chem.* **269,** 30880–30887.
21. Koegel, M., Kypta, R. M., Bergman, M., Alitalo, K. & Courtneidge, S. A. (1994) *Biochem. J.* **302,** 737–744.
22. Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., *et al.* (1994) *Mol. Cell. Biol.* **14,** 2777–2785.
23. Bougeret, C., Delaunay, T., Romero, F., Jullien, P., Sabe, H., Hanafusa, H., Benarous, R. & Fischer, S. (1996) *J. Biol. Chem.* **271,** 7465–7472.
24. Dawson, P. E., Churchill, M. J., Ghadiri, M. R. & Kent, S. B. H. (1997) *J. Am. Chem. Soc.* **119**, 4325–4329.
25. Sondhi, D., Xu, W., Songyang, Z., Eck, M. J. & Cole, P. A. (1998) *Biochemistry* **37,** 165–172.
26. MacAuley, A. & Cooper, J. A. (1989) *Mol. Cell. Biol.* **9,** 2648–2656.
27. Hardwick, J. S. & Sefton, B. M. (1997) *J. Biol. Chem.* **272,** 25429–25432.
28. Severinov, K. & Muir, T. W. (1998) *J. Biol. Chem.* **273,** in press.