New Vectors for High Level Expression of Recombinant Proteins in Bacteria

David J. Hakes and Jack E. Dixon

Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109-0606

A system has been developed for synthesis and rapid purification of recombinant polypeptides expressed in frame with glutathione S-transferase (D. B. Smith and K. S. Johnson, 1988, *Gene* 67, 31–40). Expressed fusion proteins are purified from bacterial extracts by glutathione–agarose affinity chromatography. A thrombin protease cleavage site allowed for proteolysis of the fusion protein. We reported the construction of the vector pGEX-KG (K. Guan and J. E. Dixon, 1991, *Anal. Biochem.* 192, 262–267) which has a glycine-rich "kinker" immediately after the thrombin cleavage site. This kinker dramatically improved the thrombin cleavage efficiency of several fusion proteins. One potential drawback of expressing proteins in this vector is that, following proteolytic cleavage, unrelated amino acids from the vector remain at the amino terminus of the released protein. These extensions could affect enzymatic activity or protein structure. We have constructed two new vectors, pGEX-KT and pGEX-KN, which have the glycine kinker placed N-terminal to the thrombin cleavage site in order to minimize the unrelated amino acids associated with the cleaved protein. The change in location of the kinker had no effect on the increased thrombin cleavage efficiency. A strategy combining the kinker in the vector pGEX-KN with polymerase chain reaction has also been developed to express fusion proteins which when cleaved with thrombin released a protein having no amino terminal extensions of any kind. © 1992 Academic Press, Inc.

The ability to produce foreign polypeptides in *Escherichia coli* as fusion proteins with various carrier proteins has facilitated large scale purification (1–3). Glutathione S-transferase (GST)[1] fusion proteins can be

expressed using the vector pGEX-2T (3). This vector introduced a thrombin cleavage site (4) C-terminal to the GST protein. The fusion protein can be readily purified using glutathione–agarose chromatography. Following cleavage of the fusion protein with thrombin, the glutathione S-transferase can be removed from the protein of interest by a second round of chromatography on the glutathione–agarose column. We modified the pGEX-2T vector, constructing the vector pGEX-KG (5) in which a glycine "kinker" (PGISGGGGG) was introduced after the thrombin cleavage site of pGEX-2T. The presence of this kinker resulted in improved thrombin cleavage of an expressed GST-rat tyrosine phosphatase (PTPU323) fusion protein (6). The released PTPU323 peptide contained several unrelated amino acids from the fusion protein at its N-terminal, but this did not affect the activity of the protein. With other expressed proteins, however, these N-terminal extensions could alter their structure or affect their activity. We have designed two new plasmids which utilize the increased efficiency of thrombin cleavage associated with the glycine kinker while minimizing N-terminal amino acid extensions following thrombin cleavage.

MATERIALS AND METHODS

Materials

Vector pGEX-1 was from Pharmacia LKB. *E. coli* strain TG1 [K12, Δ(lac–pro), supE, thi, hsdD5/F'traD36, proA+B+, lacI^q, lacZ ΔM15] was from Amersham. Glutathione, glutathione agarose, and human thrombin were purchased from Sigma. Restriction enzymes were obtained from U.S. Biochemicals or New England Biolabs and the sequencing kit was from USB. Problot membrane was obtained from Applied Biosystems. Oligonucleotides were synthesized at the Laboratory for Macromolecular Structure, Purdue University.

Construction of Plasmids

pGEX-KT. Two complementary synthetic oligonucleotides (5'-GATCTGTCTGGTGGTGGTGGTGGT-

[1] Abbreviations used: GST, glutathione S-transferase; PCR, polymerase chain reaction; PTP, protein tyrosine phosphatase; IPTG, isopropyl β-D-thiogalactoside; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

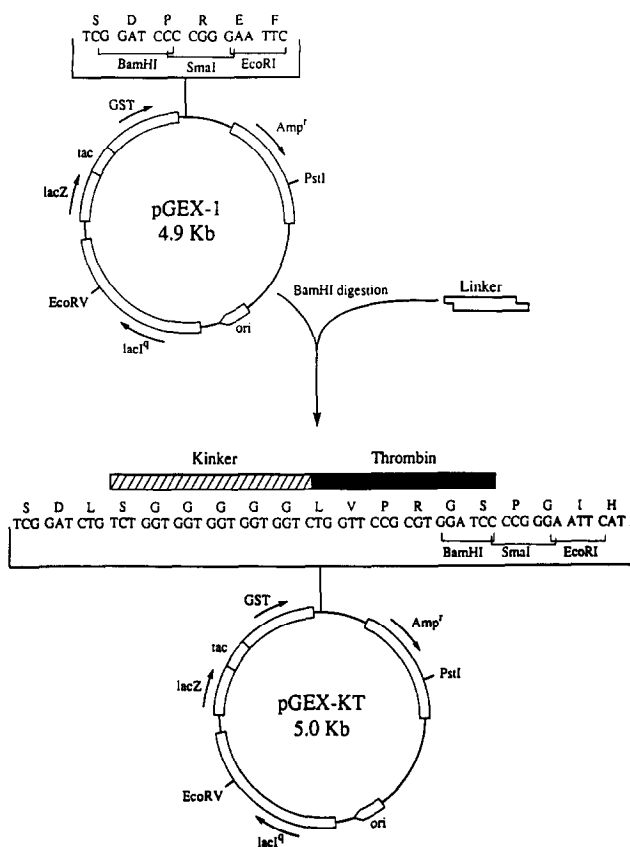FIG. 1. Construction of pGEX-KT. A synthetic oligonucleotide linker (⌐⌐) was inserted into the BamHI site of pGEX-1 to construct pGEX-KT. The sequence of the linker is shown. The amino acids coded for by the linker are shown above the nucleotide sequence. The polyglycine kinker (▨) and thrombin cleavage site (■) are indicated. The unique restriction sequences are shown below the nucleotide sequence. The genes for β-lactamase (AMP^r), the lac repressor (lacI^q), a portion of β-galactosidase (LacZ), and glutathione S-transferase (GST), as well as the origin of replication (ori) and the hybrid trp-lac promoter (tac) are shown.

CTGGTTCCGCGTG and 5'-GATCCACGCGGAACC-AGACCACCACCACCACCAGACA) were mixed in equal amounts and heated to 90°C for 10 min before slowly cooling to room temperature over 1 h. The annealed oligonucleotides were ligated into BamHI-digested pGEX-1 (Fig. 1A). The resultant plasmid, pGEX-KT, was transformed into the E. coli strain TG1. The orientation of the oligonucleotide linker was determined by double-stranded sequencing (7).

Two synthetic oligonucleotides (5'-TTTGGATCC-CGCGCCCAATTCGAGCAGATC and 5'-TTTGAA-TTCCTAGCCATTGTGAGGCTCCAA) were used in the polymerase chain reaction (PCR) with pKG-PTPU323 (5) as a template. The resulting fragment, coding for a portion of a rat brain tyrosine phosphatase (PTP) missing the hydrophobic region near the C-terminus, was digested with BamHI and EcoRI and ligated into BamHI- and EcoRI-digested pGEX-KT.

pGEX-KN.   Two complementary synthetic oligonucleotides (5'-GATCTGGGTGGTGGTGGCGGCCGCG and 5'-GATCCGCGGCCGCCACCACCACCCA) were prepared as above and ligated into BamHI-digested pGEX-1 (Fig. 4). The plasmid, termed pGEX-KN, was transformed into the E. coli strain TG1. The orientation of the linker was determined by double-stranded sequencing (7).

Two synthetic oligonucleotides (5'-TTTGCGGCC-GCCTGGTTCCGCGTATGGAAATGGAGAAGGA-ATTCGAGCA and 5'-TTTGAATTCCTAGCCATTG-TGAGGCTCCAA) were used in the PCR with pKG-PTPU323 (5) as a template. The resultant fragment, coding for a rat PTP missing the hydrophobic region near the C-terminal, but containing the initiator methionine, was digested with NotI and EcoRI and ligated into BamHI- and EcoRI-digested pGEX-KN.

Two synthetic oligonucleotides (5'-TTTGCGGCC GCCTGGTTCCGCGTATGCGCTATTACGAAAAAA-TTGATGG and 5'-TTTGAATTCTCATGCGCCTTC-TCCCTGTAC) were used in the PCR with a λ-gt10 library as a template. The resulting fragment, encompassing the entire coding sequence of the λ-phosphatase (8–9), was digested with NotI and EcoRI and ligated into NotI- and EcoRI-digested pGEX-KN.

Two synthetic oligonucleotides (5'-TTTGCGGCC GCCTGGTTCCGCGTATGAACTTATCATTAAG-CGATCTTC) and (5'-TTTGTCGGATCCTGAATT-CGAATAAATATTTACATTAGC) were used in PCR with a YOP51 cDNA clone as a template. The resulting fragment, encompassing the entire coding sequence of the Yersinia phosphatase (10) was digested with NotI and EcoRI and ligated into NotI- and EcoRI-digested pGEX-KN.

Purification of Fusion Protein

The expression of fusion proteins was performed as previously described (5). A 10-ml overnight culture was inoculated into 1 liter of LB media containing 100 μg/ml ampicillin. The culture was grown with shaking until reaching an OD$_{600}$ of 1. IPTG was then added to 0.2 mM final concentration and the culture was grown for an additional 3 h. The cells were centrifuged and the pellet was resuspended in PBST buffer (150 mM NaCl, 16 mM Na$_2$HPO$_4$, 4 mM NaH$_2$PO$_4$, pH 7.3, 1% Triton) + 0.1% β-mercaptoethanol. The cells were lysed by passage through a French press and the lysate was centrifuged for 10 min at 10,000g. The supernatant was mixed with 3 ml of 50% glutathione–agarose beads and incubated at 4°C for 30 min. The beads were washed three times with 10 ml PBST, followed by two washes with 50 mM Tris · HCl (pH 8.0). The fusion protein was eluted with 3
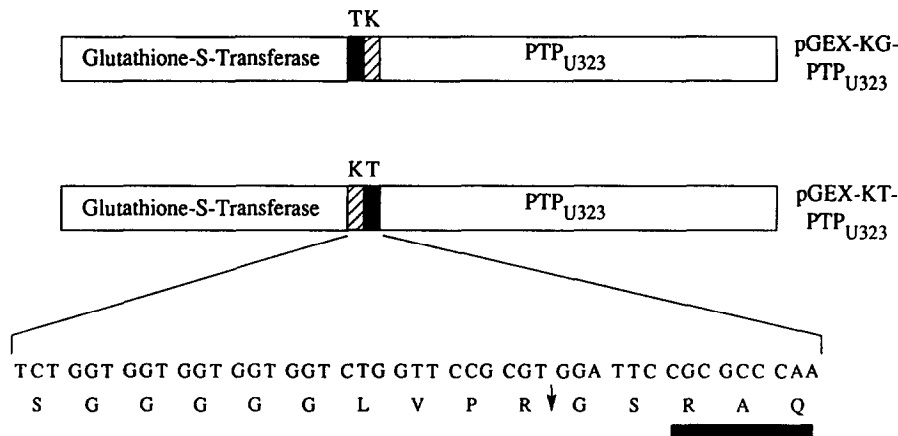
FIG. 2. Representation of pGEX-KG-PTPU323 and pGEX-KT-PTPU323. Synthetic oligonucleotides were used with PCR to subclone PTPU323 into the BamHI and EcoRI sites of pGEX-KT. The thrombin cleavage site (■) and the glycine kinker (▨) are shown in the schematic representation of the two fusion proteins. The nucleotide and amino acid sequences are shown below the pGEX-KT-PTPU323 figure. The site of thrombin cleavage is indicated by the arrow. The beginning of the coding sequence of PTPU323 is indicated by the dark underline.

ml of 50 mM Tris · HCl (pH 8.0) containing 5 mM glutathione.

Thrombin Cleavage

Fusion protein was incubated with 1:50 or 1:500 (w/w) thrombin in thrombin cleavage buffer [50 mM Tris · HCl (pH 8.0), 100 mM NaCl, 2.5 mM CaCl$_2$, 0.1% β-mercaptoethanol] for various times at room temperature. Sequencing of the thrombin-cleaved PTPU323 and Yersinia PP proteins was done using an Applied Biosystems Model 470A gas-phase sequencer using 0.5–
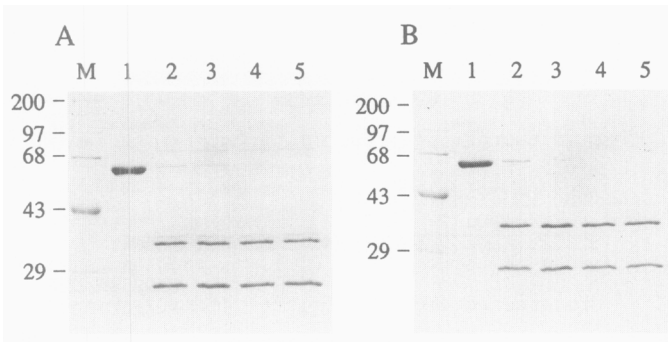

FIG. 3. Thrombin cleavage of the fusion proteins. The molecular weight of both fusion proteins is 63 kDa. Following thrombin cleavage, GST and PTPU323 are 26 and 37 kDa, respectively. Fusion proteins were incubated at room temperature for various times (see Materials and Methods). (A) Thrombin cleavage of KG-PTPU323. The thrombin to fusion protein ratio is 1:500. Lane M, molecular weights markers; lane 1, KG-PTPU323, no thrombin; lane 2, 5 min; lane 3, 10 min; lane 4, 20 min; lane 5, 40 min. (B) Thrombin cleavage of KT-PTPU323. The thrombin ratio is 1:500. Lane M, molecular weight markers; lane 1, KT-PTPU323, no thrombin; lane 2, 5 min; lane 3, 10 min; lane 4, 20 min; lane 5, 40 min.
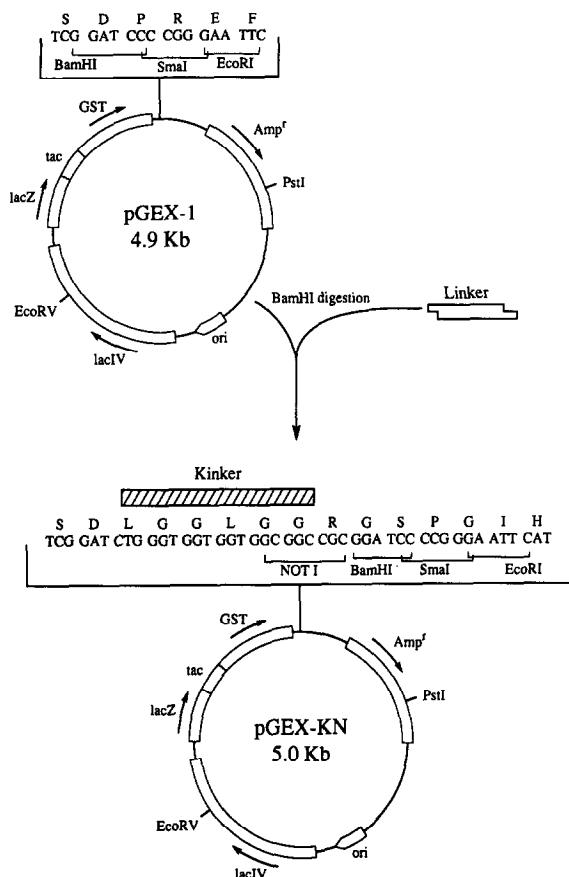
1 nmol of protein transferred to Problot membranes according to manufacturers instructions.

RESULTS AND DISCUSSION

Insertion of the Glycine Kinker N-Terminal to the Thrombin Cleavage Site

To enhance thrombin cleavage of pGEX-2T, we previously inserted a glycine kinker prior to the cleavage site (5). The resulting vector, pGEX-KG, produced a fusion protein which was readily cleaved by thrombin, but the resulting protein had several glycine residues at its amino terminus. The usefulness of this vector is reflected in the fact that we have had over 300 requests for the pGEX-KG plasmid. In order to eliminate the amino terminal extension, we have now constructed a plasmid where the glycine residues were placed prior to the thrombin cleavage site. The resulting protein, following thrombin cleavage of the fusion protein, does not contain the extraneous glycine residues. The synthetic oligonucleotides shown in Fig. 1 were inserted into the BamHI site of pGEX-1. This linker placed the amino acid sequence LSGGGGG immediately preceding the thrombin cleavage site found in pGEX-2T, while maintaining the restriction sites found in the original pGEX-1 vector. This altered vector was termed pGEX-KT (Fig. 1).

The increased efficiency of thrombin cleavage with pGEX-KG was demonstrated using a fusion protein produced from a protein tyrosine phosphatase (PTPU323) (5). We utilized the same PTP for our studies using pGEX-KT. The cDNA encoding the first 323 amino acids of the phosphatase was subcloned into pGEX-KT to produce the plasmid pGEX-KT-PTPU323 (Fig. 2; Materials and Methods). The major

FIG. 4. Construction of pGEX-KN. A synthetic oligonucleotide linker (⌐⌐ₗₗ) was inserted into the BamHI site of pGEX-1 to construct pGEX-KN. The sequence of the linker is shown. The amino acids coded for by the linker are shown above the nucleotide sequence. The polyglycine kinker (▧) is indicated. The unique restriction sites are shown under the nucleotide sequence. The genes for β-lactamase (AMP^r), the lac repressor (lacI^q), a portion of the β-galactosidase (LacZ), and glutathione S-transferase (GST) as well as the origin of replication (ori) and the hybrid trp-lac promoter (tac) are shown.

difference between the two plasmids is in the placement of the glycine kinker with respect to the thrombin cleavage site (Fig. 2). Both 63-kDa fusion proteins were expressed to high levels in E. coli and purified by glutathione–agarose affinity chromatography (3).

The efficiency of thrombin cleavage of KT-PTPU323 and KG-PTPU323 was compared in Fig. 3. Uncleaved, purified KG-PTPU323 and KT-PTPU323 fusion proteins are shown in lanes 1 of Figs. 3A and 3B, respectively. Only minor protein contaminants are visible in each lane. Lanes 2–5 of Figs. 3A and 3B represent the extent of thrombin cleavage following 5, 10, 20, and 40 min of incubation at room temperature with a 1:500 ratio of enzyme to fusion protein. The rates of cleavage of KT-PTPU323 and KG-PTPU323 are virtually identical, with complete cleavage of the fusion protein after 10 min (lane 3, Figs. 3A and 3B). These results suggest that

the glycine kinker increases the efficiency of cleavage of the GST-PTPU323 fusion protein by thrombin when placed on either side of the thrombin cleavage site.

The N-terminal sequence of the PTPU323 peptide released following thrombin cleavage of KT-PTPU323 was determined. The proteins were subjected to electrophoresis on a 10% SDS–PAGE (11) gel and transferred to Problot membranes. Protein (0.5–1 nmol) was sequenced using an Applied Biosystems Model 470A gas-phase sequencer. The sequence (GSRAQFEQID-KAGNWA) contained the predicted Gly–Ser extension N-terminal to the PTP sequence. These results suggest that the pGEX-KT vector will prove very useful in expressing and purifying large amounts of protein containing minimal amino terminal modifications.
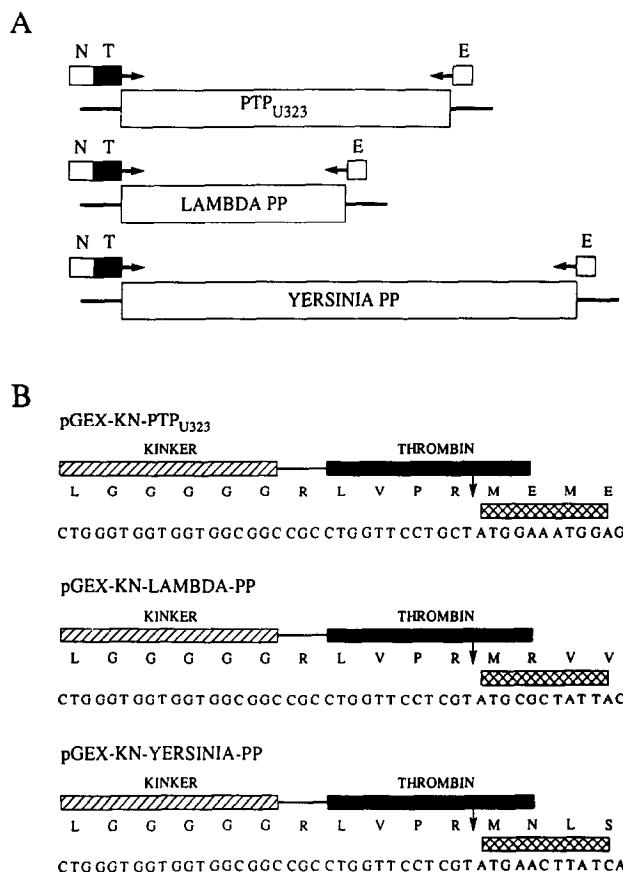

FIG. 5. Construction of KN fusion proteins. (A) Strategy for PCR of PTPU323, the λ-phosphatase (8,9), and the Yersinia phosphatase (10). The coding sequences used as templates for PCR are shown. The synthetic oligonucleotides are indicated by □→. The NotI (N) and EcoRI (E) sites as well as the partial thrombin cleavage sites (T, ■) are shown above the boxes. (B) The nucleotide and amino acid sequences of the junctions of the KN-PTPU323, KN-Lambda-PP, and KN-Yersinia-PP fusion proteins are shown. The glycine kinkers (▧) and the putative thrombin cleavage sites (■) are indicated. The site of potential thrombin cleavage is shown with an arrow (↓). The sequence of the mature portions of each protein (▩) is shown.
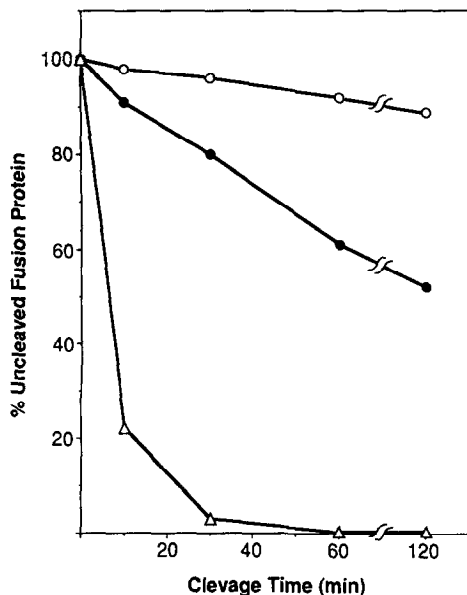
FIG. 6. Thrombin cleavage of KN fusion proteins. The percentage of uncleaved pGEX-KN-PTPU323 (O), pGEX-KN-Lambda-PP (●) and pGEX-KN-Yersinia-PP (△), following digestion with thrombin (1:50, enzyme:substrate) for various periods of time at room temperature was estimated by densitometric scans of SDS polyacrylamide gels.

Expressed Proteins Lacking N-Terminal Extensions

It would be advantageous to be able to produce large quantities of protein which contain the exact N-terminus found in the native protein. We designed a pGEX vector which could be used in conjunction with the PCR (and thrombin cleavage) to produce proteins free of any extraneous N-terminal amino acids. The optimum cleavage site for thrombin has been proposed to have the structure P4–P3–Pro–Arg–P1'–P2'; where P1' and P2' are nonacidic amino acids and P3 and P4 are hydrophobic amino acids (4). The thrombin cleavage sites found in pGEX-2T, pGEX-KG, and pGEX-KT (LVPRSG) conform to this consensus sequence. Many native proteins contain an N-terminal methionine and this residue at position P1' would fit into the consensus cleavage sequence.

The synthetic oligonucleotide shown in Fig. 4 was inserted into the BamHI site of pGEX-1. This linker placed the amino acid sequence LGGGGGR along with a NotI restriction site immediately before the restriction sites found in the original pGEX-1 vector. This modified vector was termed pGEX-KN.

The strategy for subcloning fusion proteins in pGEX-KN is shown in Fig. 5A. Synthetic oligonucleotides were used along with PCR to insert coding sequences between the NotI and EcoRI restriction sites. The 5' oligonucleotide also contains sequences coding for the first four amino acids of a thrombin recognition sequence

(LVPR) (Fig. 5A). The two amino acids following the site of thrombin cleavage (C-terminal to the R residue) correspond to the exact amino terminus of the anticipated fusion protein (Fig. 5A). NotI was chosen as the 5' restriction site because its 8-bp recognition sequence would rarely occur in a typical DNA coding region.

Proteins with an acidic, a basic, or a neutral amino acid in the P2' position of the thrombin cleavage site were expressed as fusion proteins and subsequently subjected to enzymatic cleavage. The three proteins are all phosphatases. An acidic site at P2' is present in PTPU323, a protein phosphatase present in λ (8,9) incorporates a basic residue at P2' and a bacterial phosphatase from Yersinia (10) contains a neutral amino acid at this site. The structure of the junctions of KN-PTPU323, KN-Lambda-PP, and KN-Yersinia-PP are shown in Fig. 5B. KN-PTPU323 encodes a 63-kDa fusion protein, KN-Lambda-PP codes for a 51-kDa fusion protein, and KN-Yersinia-PP encodes a 77-kDa fusion protein. All three fusion proteins contain the glycine kinker N-terminal to a putative thrombin cleavage site. All three fusion proteins were expressed to high levels (10–30 mg/liter of bacteria) in E. coli and purified by glutathione–agarose chromatography.

The efficiency of thrombin cleavage of KN-PTPU323, KN-Lambda-PP, and KN-Yersinia-PP were compared in Fig. 6. There are major differences in the rates of cleavage of the three fusion proteins. KN-PTPU323 (open circles in Fig. 6) showed very low levels of cleavage even following 2 h of incubation with a 1:50, thrombin:substrate ratio. This result is consistent with the low levels of cleavage predicted with an acidic residue in the P2' position of the thrombin cleavage consensus site (4). KN-Lambda-PP (closed circles in Fig. 6) showed significant levels of cleavage. The uncleaved fusion protein represented approximately half of the protein forms present after a 2-h incubation. KN-Yersinia-PP (open triangles in Fig. 6) showed essentially complete cleavage following a 30-min incubation. Interestingly, the released 51-kDa Yersinia PP contained an internal cryptic thrombin cleavage site producing 34- and 17-kDa proteins (data not shown). The rates of thrombin cleavage at the cryptic site and the site located at the fusion protein juncture were virtually identical. The two resultant peptides were sequenced and the 17-kDa peptide was found to contain the correct predicted N-terminal sequence of the Yersinia-PP (data not shown). These results suggest that a Met in the P1' position of the recognition site necessitates an uncharged amino acid in the P2' position for efficient thrombin cleavage.

In conclusion, two vectors have been modified to produce large concentrations of recombinant protein which can be rapidly purified by affinity chromatography. The

vector, pGEX-KT, incorporates the glycine kinker prior to the thrombin cleavage site and this affords rapid and complete cleavage of the fusion protein with only two additional amino acids at the C-terminus (Gly–Ser). The second vector, pGEX-KN, can be used to express a protein containing the initiator methionine if the second residue in the sequence is uncharged. Thrombin cleavage is not as rapid as that seen with the pGEX-KT vector, but quantitative cleavage can be obtained with thrombin:substrate ratios of 1:50.

ACKNOWLEDGMENT

REFERENCES

1. Shuman, H. A., Silhavy, T. J., and Beckwith, J. R. (1980) *J. Biol. Chem.* **255**, 68–174.
2. Bedoulle, H., and Duplay, P. (1988) *Eur. J. Biochem.* **171**, 541–549.
3. Smith, D. B., and Johnson, K. S. (1988) *Gene* **67**, 31–40.
4. Chang, J.-Y. (1985) *Eur. J. Biochem.* **151**, 217–224.
5. Guan, K., and Dixon, J. E. (1991) *Anal. Biochem.* **192**, 262–267.
6. Guan, K. L., Haun, R. S., Watson, S. J., Geahlen, R. L., and Dixon, J. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1501–1505.
7. Sanger, F., Nicklen, S., and Colson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5453–5467.
8. Cohen, P. T., Collins, J. F., Coulson, A. F., Berndt, N., and da Cruz e Silva, O. B. (1988) *Gene* **69**, 131–134.
9. Cohen, P. T., and Cohen, P. (1989) *Biochem. J.* **260**, 931–934.
10. Guan, K. L., and Dixon, J. E. (1990) *Science* **249**, 553–556.
11. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.