Journal of Biomolecular NMR, **26:** 193–202, 2003.
KLUWER/ESCOM

A rapid method to attain isotope labeled small soluble peptides for NMR studies

Bernd W. Koenig[a,*], Marco Rogowski[b] & John M. Louis[c]
[a]*Structural Biology Institute, IBI-2, Research Centre Jülich, D-52425 Jülich and Institute of Physical Biology, Heinrich-Heine-Universität, D-40225 Düsseldorf, Germany;* [b]*Department of Structural Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland;* [c]*Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, MD 20892-0520, U.S.A.*

Received 14 October 2002; Accepted 17 March 2003

Key words: E. coli, fusion protein, isotope labeling, NMR, recombinant peptide

Abstract

A widely applicable strategy is presented for efficient and rapid production of small water soluble peptides expressed as fusion proteins with the immunoglobulin-binding domain of streptococcal protein G. A simple extraction and purification scheme that includes a protease cleavage step to release the target peptide is described. The yield of authentic target peptide exceeds 10 mg per liter of culture. Production of U-^{13}C, ^{15}N and highly deuterated U-^{13}C, ^{15}N isotope labeled peptide is demonstrated for the 11 residue S2 peptide, corresponding to the C-terminus of the α-subunit of transducin, and the coiled coil trimerization domain from cartilage matrix protein (CMPcc), respectively. Heteronuclear two-dimensional NMR spectra are used for initial peptide characterization.

Abbreviations: CMPcc – 43 residue coiled coil domain of chicken cartilage matrix protein plus four additional non-native N terminal residues; S2 – 11 residue peptide analog of the C-terminus of α-subunit of bovine transducin; GB1 – immunoglobulin-binding domain of streptococcal protein G; GEV-CMPcc – vector for expression of the fusion protein GB1 linked to CMPcc; GEV-S2 – vector for expression of the fusion protein GB1 linked to S2; CNBr – cyanogen bromide; Fxa – factor Xa protease; ESMS – electrospray mass spectrometry.

Introduction

Small peptides are important in numerous biological processes like peptide hormones in signal transduction, antimicrobial peptides in host defense, or peptide inhibitors in enzyme regulation. Biological action of these peptides involves complex formation. Characterization of the bound form of the peptides is key to a detailed understanding of such processes. Moreover, noncovalent peptide-protein complexes can serve as models for understanding protein-protein interaction where the binding domain of one protein is represented by a peptide, e.g., in studies aiming at structure and dynamics of antibody-antigen complexes (Tugarinov et al., 1999).

NMR is a powerful method for characterizing the structure, dynamics, and interactions of biological macromolecules. However, to make use of the full potential of NMR methods, biomolecules labeled with NMR active stable isotopes, in particular with ^{15}N, ^{13}C, and ^{2}H, are required. Multidimensional, multinuclear NMR experiments reduce spectral overlap, greatly simplify resonance and NOE assignments, and allow for spectral editing or filtering schemes. The latter is particularly important when studying complexes by allowing exclusive retention of NMR signals of selected molecules or moieties based on appropriate isotope labeling strategies (Walters et al., 2001).

The bound conformation of small ligands that weakly bind to complexes far too large for liquid state

*To whom correspondence should be addressed. E-mail: b.koenig@fz-juelich.de

NMR methodology, e.g., membrane anchored receptors, may still be accessible with transferred NOE (Clore and Gronenborn, 1982; Ni, 1994) or transferred dipolar coupling measurements (Koenig et al., 2002) on the free form of the ligand. ^{13}C and/or ^{15}N labeled peptides are required to measure heteronuclear transferred dipolar couplings. Interpretation of two-dimensional NOESY spectra of transiently binding small peptides often suffer from peak overlap due to low dispersion of ^{1}H spectra of the unfolded peptide. The overlap problem can be overcome by the use of three-dimensional heteronuclear pulse schemes if isotope labeled peptide is available. Transverse and longitudinal relaxation times of heteronuclei (^{13}C, ^{15}N) and steady state heteronuclear NOEs (^{1}H-^{15}N or ^{1}H-^{13}C) reflect motional properties of the backbone and side chains of polypeptides on the pico- to millisecond time scales (Kay, 1998). Probing the dynamic properties of peptides does greatly benefit from availability of isotope labeled material (Alexandrescu et al., 1998; Campbell et al., 2000).

Genetic engineering followed by efficient expression of isotope labeled biomolecules in milligram amounts has been indispensable for recent successes in biological NMR applications. Much effort has been devoted to recombinant expression and refolding of medium size proteins (Gardner and Kay, 1998). However, recombinant production of small peptides in sizable quantities still presents a formidable challenge. Although, small peptides can be produced chemically by solid-state synthesis (Merrifield, 1995), this process becomes prohibitively expensive even for uniformly labeled short peptides.

Direct expression of short peptides in bacterial hosts in many instances is susceptible to rapid proteolytic degradation. However, gene fusions have been successfully engineered coding for hybrid proteins with the target peptides being part of a larger protein (Itakura et al., 1977; Goeddel et al., 1979), which are sufficiently stable and well characterized for recombinant production. Proteins that express well and allow rapid purification by affinity chromatography are attractive fusion partners.

Numerous expression vectors to produce fusion proteins are commercially available that allow easy in-frame insertion of a target DNA sequence. Examples include fusions with maltose binding protein (di Guan et al., 1988), glutathione S-transferase (Smith and Johnson, 1988), ketosteroid isomerase (KSI) (Kuliopulos and Walsh, 1994), and thioredoxin (LaVallie et al., 1993), etc. However, design of fusion proteins for production of small peptides requires balancing several aspects. The final yield of purified peptide depends on the expression level and stability of the fusion protein, but also on the relative size of the target peptide to fusion partner. A small molecular weight of the fusion partner enhances the yield of the target peptide and is particularly desirable if expensive isotope enriched growth media are used. Fusion proteins containing multiple repeats of the peptide sequence of interest might further increase peptide yield (Kuliopulos and Walsh, 1994; Jonasson et al., 1998). Provisions must be made to allow release of the intact peptide from the fusion protein by either chemical or enzymatic means. Sequence modifications at the N- and/or C-termini, resulting from the strategy chosen to release the target peptide, may affect the properties of short peptides and should be avoided. The solubility, localization and purification scheme of the fusion peptide is often influenced by the properties of the fusion tag. For production of peptides that are toxic to the host and for highly hydrophobic peptides like membrane spanning segments of membrane proteins, a fusion partner with a bias towards formation of inclusion bodies is desirable. A fusion tag that accumulates to high levels in the cytoplasm or that directs the fusion into the periplasmic space might be more appropriate for production of soluble peptides.

In spite of a wealth of information, expression of fusion proteins is a complex process and does not always follow simple predictions. It might be necessary to test several strategies with different fusion tags in search for a highly efficient protocol that works for a given peptide. It is unlikely that a universal fusion tag and purification strategy will emerge that would allow production of any given peptide. Instead there is a strong need for several different systems that can be tried and perhaps modified in the quest for a suitable expression scheme to attain the target peptide of interest.

Here we describe the highly efficient recombinant production and purification of isotope labeled soluble peptides that are expressed as C-terminal fusions to the immunoglobulin binding domain of streptococcal protein G (GB1 domain) in *E. coli*. The peptides were inserted as gene fusions into the GEV2 vector, a pET based system originally designed for high yield production and rapid NMR characterization of isotope labeled fusions of the GB1 domain with proteins of interest (Huth et al., 1997). A specific protease cleavage site is introduced immediately prior to the peptide sequence allowing easy recovery of peptide molecules

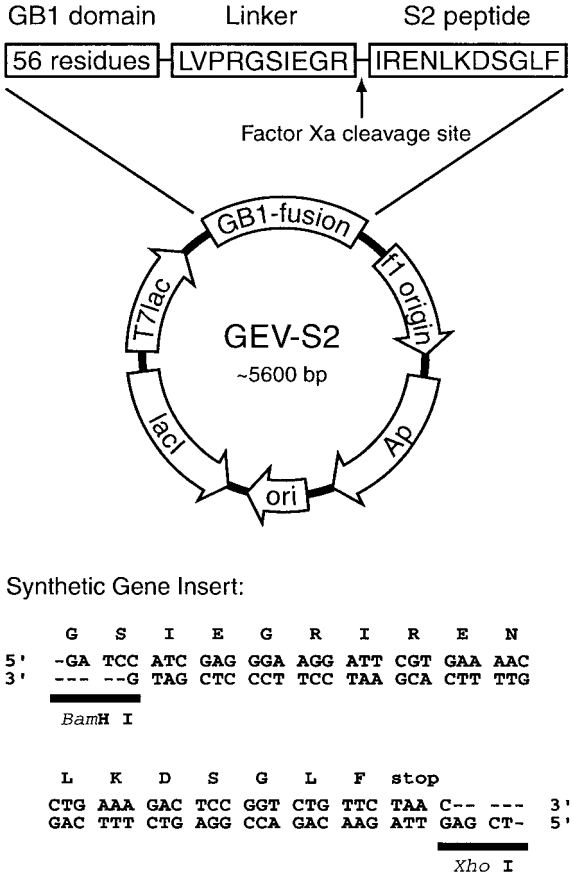

Figure 1. An illustration of the expression vector GEV-S2 containing the synthetic S2 gene fragment cloned between the *Bam*HI and *Xho*I restriction sites of a pET21-based vector GEV2 (Huth et al., 1997). Amino acid sequence of the linker region and the target S2 peptide are shown and the factor Xa cleavage site is indicated by an arrow.

exactly matching the desired sequence without any modifications of the N-terminus.

Materials and methods

Construction of GEV-S2 and GEV-CMPcc expression vectors

The oligonucleotide coding for a factor Xa (Fxa) cleavage site immediately followed by the S2 peptide and a stop codon and the complementary strands were synthesized and purified (Midland Certified Reagent Co., Midland, TX). The nucleotide sequence was biased for optimal codon usage in *E. coli* (www.kazusa.or.jp/codon). The oligonucleotides were designed to provide 5′ *Bam*HI and 3′ *Xho*I cohesive

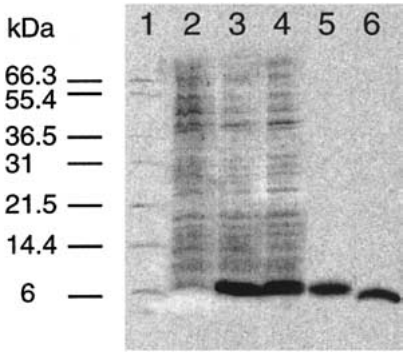

Figure 2. Steps in the purification of the S2 peptide analyzed by 20% SDS-PAGE. Lane 1, protein standard Mark 12 (Novex, Carlsbad, CA); lanes 2 to 4, supernatant of the cell lysate of *E. coli* BL21 transformed with GEV-S2 plasmid harvested prior to induction (lane 2) and 3 hours post-induction (lanes 3 and 4); lane 5, GB1-S2 fusion protein after size-exclusion chromatography; lane 6, Fxa digest of GB1-S2 fusion protein (note that only the GB1 domain containing the linker region is visible).

ends for cloning into the GEV2 vector (Figure 1) (Huth et al., 1997). 2.4 nmol of each oligonucleotide were annealed in 40 µl of buffer (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA) at 95 °C for 5 min followed by cooling to 25 °C over 55 min prior to ligation to the vector.

The gene coding for the 43 residue CMPcc (Argraves et al., 1987), a stop codon, and four non-native residues (GSHM) flanking the N-terminus of CMPcc was PCR amplified from a previously engineered vector (Wiltscheck et al., 1997). The primers were designed to provide a 5′ *Bam*HI site followed by a Fxa cleavage site and a 3′ *Xho*I site in the CMPcc insert for cloning into the GEV2 vector (Huth et al. 1997).

Expression and purification of labeled fusion proteins

E. coli BL21-Gold (DE3) (Stratagene, La Jolla, CA) were transformed with the GEV-S2 plasmid. Cells were grown in a modified minimal medium for uniform ^{15}N and/or ^{13}C labeling with $^{15}NH_4Cl$ and/or [$^{13}C_6$] glucose as the sole nitrogen and carbon sources, respectively. The medium contained 1.2 g l^{-1} NH_4Cl, 0.5% glucose, 0.02% yeast extract, 100 mg l^{-1} carbenicillin, 0.1 mg l^{-1} biotin, 2 mg l^{-1} thiamin, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 1 ml trace elements per liter. 100 ml of trace elements solution contained per liter 600 mg $CaCl_2 \cdot 2H_2O$; 600 mg $FeSO_4 \cdot 7H_2O$; 500 mg EDTA; 115 mg $MnCl_2 \cdot 4H_2O$; 80 mg $CoCl_2 \cdot 6H_2O$; 70 mg $ZnSO_4 \cdot 7H_2O$; 30 mg $CuCl_2 \cdot 2H_2O$; 25 mg $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$; and 2 mg H_3BO_3.

Cells were grown in 50 ml of medium at 37 °C overnight. Cultures were diluted 20 fold with medium and grown to late exponential phase at 37 °C. Expression was induced at $OD_{600} = 0.7$ with 2 mM isopropyl β-thiogalactosidase (IPTG) for 3 h. Cells were harvested by centrifugation at $6000 \times g$ for 20 min at 4 °C. Pelleted cells were kept at −70 °C for at least 1 h, thawed and resuspended in 1 × PBS, and then heated for 5 min at 80 °C. The cell suspension was chilled on ice for 10 min and spun at $20000 \times g$ for 30 min at 4 °C. The supernatant containing the fusion protein was extensively dialyzed against deionized water using a 3000 MWCO Slide-A-Lyzer cassette (Pierce, Rockford, IL), filtered (0.22 μm), and concentrated using Centriprep YM-3 concentrators (Millipore, Bedford, MA).

The GB1-S2 fusion protein was purified by size-exclusion chromatography on a Superdex-75 column (2.6×60 cm; Amersham Biosciences, Piscataway, NJ) in Factor Xa buffer (20 mm Tris-HCl, pH 8, 100 mM NaCl without 2 mM $CaCl_2$). Peak fractions of the fusion protein were pooled and concentrated to about 20 mg protein per ml using Centriprep YM-3 devices. An apparent molecular weight of about 6 kDa was estimated for the GB1-S2 fusion protein by SDS-PAGE (Figure 2). However, electrospray mass spectrometry (ESMS) confirmed the expected molecular weight of 8990 g mol^{-1} of the doubly labeled fusion protein.

CMPcc was expressed in *E. coli* BL21(DE3) (Novagen, Madison, WI). Cells were grown in 1 l of M9 medium adapted for growth in deuterium oxide (>97% ^2H) (Vanatalu et al., 1993). Fed-batch fermentation was performed in a 2.5-l bioreactor (Infors, Basel, Switzerland) equipped with on-line measurement of oxygen. Total consumption of ^{13}C – glucose and $^{15}NH_4Cl$ per liter of fermentation medium was 3.75 g and 1.5 g, respectively. Gene expression was induced at $OD_{600} = 2.4$ with 1.5 mM IPTG for 4 h (final $OD_{600} = 3.8$).

The GB1-CMPcc fusion was isolated by affinity chromatography at room temperature using a 20 ml column packed with IgG Sepharose 6 Fast Flow (Amersham Biosciences, Piscataway, NJ) following an established protocol for isolation of GB1 fusions (Huth et al., 1997). Fractions containing the fusion protein were pooled, lyophilized, and resuspended in Factor Xa buffer containing 2 mM $CaCl_2$.

Enzymatic digestion and peptide purification

The fusion protein was subjected to Factor Xa cleavage (New England Biolabs, Beverly, MA) at 20 °C using a protein : enzyme weight ratio of 1000:1. The digest was stopped after 16 h by freezing the sample at −20 °C.

The cleaved S2 peptide was separated from the GB1 domain, monitored at 215 nm, on a Superdex 30 column (2.6×60 cm; Amersham Biosciences, Piscataway, NJ) in 20 mM sodium phosphate buffer, pH 6.7, 0.020 mM EDTA, and 0.01% NaN_3. S2 peptide containing fractions were pooled and concentrated to about 3 mg ml^{-1} by lyophilization. The peptide was desalted by applying the sample on the same Superdex 30 column equilibrated in deionized water. The pure S2 peptide was lyophilized. The final yield was ∼ 15 mg U-^{13}C, ^{15}N labeled S2 peptide per liter of minimal medium. ESMS confirmed the expected molecular weight of 1362 g mol^{-1} of the doubly labeled peptide.

Following enzymatic cleavage, CMPcc was reduced with 300 mM DTT for 30 minutes at 80 °C and centrifuged for 10 min at $5000 \times g$ at 22 °C. The supernatant was adjusted between pH 4 to 5 with 10% trifluoroacetic acid (TFA) and applied to a Vydac C_4 reverse-phase HPLC column (2.2×25 cm; GraceVydac, Hesperia, CA). Protein was eluted with a linear acetonitrile gradient (0 to 90%) in water containing 0.1% TFA. The fraction with reduced CMPcc was lyophilized and resuspended in 50 mM Tris-HCl buffer, pH 8.0, and 0.5 M NaCl. Oxidation with a glutathione redox system (5 mM oxidized and 2 mM reduced glutathione) for 5 days resulted in trimeric CMPcc. The buffer was exchanged to 100 mM Tris-HCl, pH 7.1 and 150 mM NaCl and the protein was concentrated using Centricon YM-3 devices. The final yield was ∼ 48 mg CMPcc peptide per liter of deuterated M9 medium. Note that the target peptide in this instance contains four additional residues at the N-terminus that are not part of the natural CMPcc sequence, nevertheless we refer to this 47 residue peptide as CMPcc. These four non-native residues were added to create a peptide that exactly matches the sequence of the 47 residue CMPcc peptide used in a previous NMR study (Wiltscheck et al., 1997).

NMR characterization

Two-dimensional NMR spectra on S2 and CMPcc were acquired on Bruker DMX600 and DRX800 spectrometers, respectively, using actively shielded x,

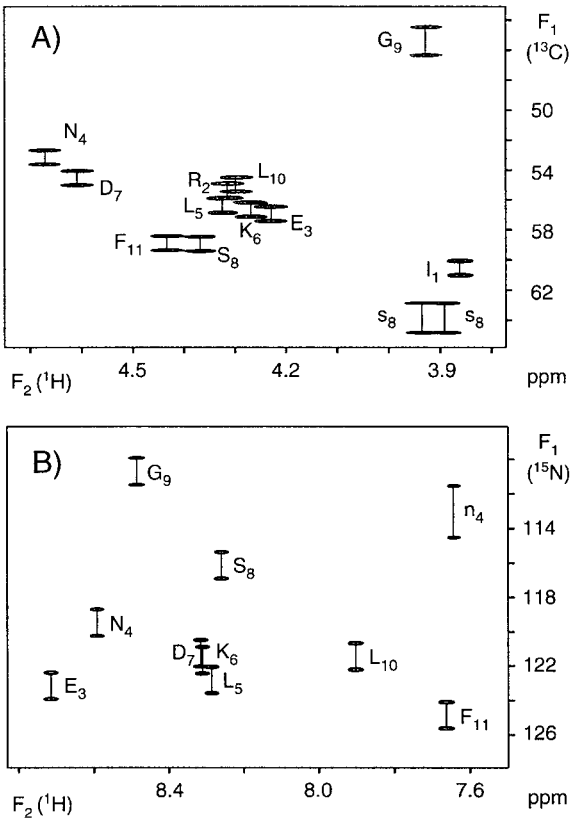

Figure 3. 600 MHz HSQC spectra without ^{1}H decoupling in the t_1 dimension for the U-^{13}C, ^{15}N doubly labeled S2 peptide (2.5 mM) at 20 °C. A) Cα-Hα region of ^{1}H-^{13}C CT-HSQC. Cβ-Hβ signals of Ser8 are indicated by lowercase letters. Cα-Hα signals of Gly9 are negative. B) Backbone amide region of ^{1}H-^{15}N HSQC. Signal of the side chain amino group of Asn4 is indicated by lowercase letter.

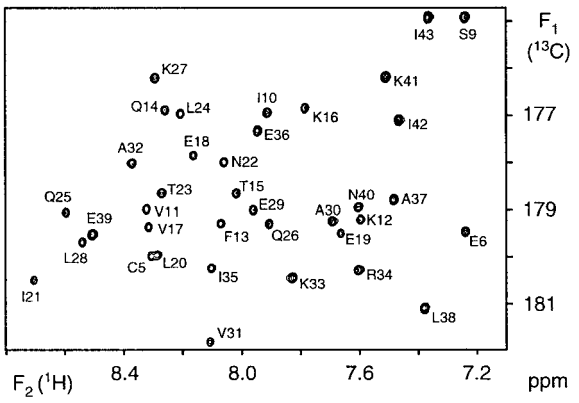

Figure 4. Two-dimensional ^{1}H-^{13}C projection of a HNCO-TROSY spectrum of oxidized trimeric CMPcc (highly deuterated, U-^{13}C, ^{15}N) recorded at 800 MHz ^{1}H frequency at 50 °C. H^{N}(i)-C′(i−1) correlations are labeled for 37 of the 46 non-proline residues of CMPcc. The residues of the N-terminal tag (GSHM) and residues 1 to 3, 7 and 8 of the native CMPcc sequence are not observed due to fast solvent exchange of the amide protons and/or high mobility of the N-terminus of the peptide.

y, z gradient triple resonance probes. Spectra were processed with the NmrPipe software package (Delaglio et al., 1995).

U-^{13}C, ^{15}N labeled S2 peptide (2.5 mM) was studied in 10 mM Hepes, 20 mM KCl buffer (90% H_2O/10% D_2O) at pH 6.6 and T = 20 °C (Figure 3). A ^{1}H-^{13}C CT-HSQC (Santoro and King, 1992; Vuister and Bax, 1992) without proton decoupling in the t_1 dimension was recorded with 106 × 512 complex data points, and total acquisition times of 28 ms (t_1) and 68 ms (t_2). Quadrature in the t_1 dimension was achieved using the Rance-Kay recipe (Kay et al., 1992). Apodization was performed with 63°-shifted sine bell (t_1) and squared sine bell (t_2) window functions. The data matrix was zero-filled to 512 × 2,048 points and Fourier transformed. A ^{1}H-^{15}N HSQC without proton decoupling in the t_1 dimension was recorded with 65 × 1,024 complex data points, resulting in total acquisition times of 120 ms (t_1) and 114 ms (t_2). Quadrature in the t_1 dimension was achieved in the States-TPPI manner. Data were apodized with 72°-shifted sine bell (t_1) and squared sine bell (t_2) window functions, prior to zero filling (256 × 2,048 points) and Fourier transformation.

Trimeric, highly deuterated U-^{13}C, ^{15}N labeled CMPcc (1.5 mM) was studied in 100 mM Tris-HCl, 150 mM NaCl buffer (90% H_2O/10% D_2O) at pH 7.1 and T = 50 °C (Figure 4). A two-dimensional ^{1}H-^{13}C HNCO-TROSY (Pervushin et al., 1997; Yang and Kay, 1999) was recorded with 35 × 1,024 complex data points, with acquisition times of 23.8 ms (t_1) and 85 ms (t_2). Quadrature in the indirect dimension was achieved in the States manner. Data were apodized with 75°-shifted sine bell (t_1) and squared sine bell (t_2) window functions and zero filled to 128 × 2,048 points.

Results and discussion

Genetically engineered gene fusions coding for short soluble target peptides linked to the GB1 domain of streptococcal protein G enable extraordinarily efficient, time- and cost-effective recombinant production of a wide variety of isotopically labeled peptides for NMR studies. Six different soluble peptides have been produced as fusions with the GB1 domain in our laboratories. This manuscript describes plasmid design and purification of the S2 and CMPcc peptides. One of the other four water-soluble peptides produced by following closely the same strategy is a 22-mer pep-

tide that corresponds to the 20 N-terminal residues of RNAse A flanked by an N-terminal Gly-Ser dipeptide, termed S-peptide. Recombinant production of the GB1-S-peptide fusion protein resulted in about tenfold higher yield of S-peptide (22 mg per liter of minimal medium) as compared to a previously reported system, where the S-peptide was expressed as part of a fusion with a 6 kDa carrier protein in *E. coli* (Alexandrescu et al., 1998). The remaining three peptides represent 31, 33 and 37 residue-long overlapping fragments of the extracellular domain of the CD4 glycoprotein of human T cells. The final yield of CD4 peptides was about 4 mg per liter of complex medium, $\sim$ 3- and 11-fold less than that obtained for the S2 and CMPcc peptides, respectively. Most likely, the 3-fold lower glucose concentration used in the medium for the expression of the GB1-CD4 fusions contributes in part to the lower yields (M. Rogowski and S. Grzesiek, unpublished result).

Peptides containing about fifty residues or more are often sufficiently resistant against proteolytic degradation in the host cell and allow direct expression in *E. coli*. However, even production of such peptides as part of a fusion protein with GB1 can significantly increase the yield of peptide. For example, the amount of the 47 residue peptide CMPcc obtained per liter of growth medium is roughly threefold higher if peptide production is based on expression of the GEV-CMPcc vector as described in this manuscript in comparison to direct expression of CMPcc using the pET-15b based vector described by Wiltscheck et al. (1997) (M. Rogowski, unpublished data).

The GB1 domain of streptococcal protein G exhibits an extraordinary stable three-dimensional structure and an extreme thermal stability with a melting temperature of 87 °C (Gronenborn et al., 1991). GB1 on its own expresses at very high levels in *E. coli* and nearly the same level of expression and accumulation pattern during induction is observed for several other GB1 fusions, where GB1 is linked to proteins of moderate size (Huth et al., 1997). Similarly, high level expression of GB1 fusions with short peptides is observed under conditions originally optimized for expression of GB1. The three-dimensional structure of GB1 is largely preserved when fused to small target proteins as indicated by similar chemical shift values (Huth et al., 1997). We speculate that the extremely stable fold of the GB1 domain may in part play a role in the stability of the soluble target peptides when fused to the GB1 domain. However, to our knowledge the exact mechanism by which the target peptide is protected from degradation in the host cell remains unknown.

Factor Xa protease, based on its specificity and the design of the fusion protein, is one among a few proteases that permits the release of the target protein without additional N-terminal residues. This method of releasing the target polypeptide using Fxa together with the high-level expression as a GB1 fusion protein nearly constitutes a perfect system for rapid production of unlabeled and isotope labeled soluble peptides. Fxa is known to hydrolyze the peptide bond following the tetrapeptide sequence Ile-Glu-Gly-Arg (Nagai and Thogersen, 1987). Cleavage of the GB1-peptide fusions presented here with commercially available Fxa proved to be very efficient even at a protein to enzyme weight ratio of 1000:1, well below the manufacturer's recommendation. Digests were basically complete after 4 hours. Apparently, the cleavage site on the folded fusion protein is readily accessible for the enzyme. However, despite the high specificity of Fxa, the enzyme does occasionally attack secondary cleavage sites (Nagai and Thogersen, 1987). A much-reduced rate of cleavage was observed after the sequence IIGR (residues 84–87 of the mature HIV-1 protease; J.M. Louis, unpublished data) and after RSGR and IDGR sequences within the Rep catalytic domain of TYLC (Campos-Olivas et al., 2002). Thus, caution should be exerted when the target peptide or protein also bears the dipeptide sequence GR. Enterokinase with a specificity Asp-Asp-Asp-Asp-Leu, where the cleavage occurs after the C-terminal Leu residue is a good alternative to Fxa in such cases.

Two alternative protocols are available for easy purification of the fusion protein, using either IgG affinity or size exclusion chromatography. Both methods allow close to 100% recovery of the fusion protein in a simple one-step procedure. Another advantage of GB1-peptide fusions is the moderately small size of the GB1-domain and linker region (total of 66 amino acids) resulting in significantly higher yields of the target peptide. This point becomes particularly important if bacteria are grown in expensive isotope labeled substrates.

The obvious advantages of short peptides uniformly or selectively labeled with NMR active isotopes in many biomolecular NMR applications has triggered an intensive search for cost-effective and practical strategies to produce such peptides in milligram amounts. Over the last few years several fusion protein based strategies have been proposed. Soluble peptides have been obtained from fusion proteins that

Table 1. Strategies for stable recombinant production of short peptides (< 50 residues) as fusion proteins

Peptide # of residues	Fusion partner # of residues	Enzyme or chemical cleavage	Extra residues/ modification	Accumulation site	Peptide yield [mg l^{-1}]		Reference
					Luria–Bertani medium	Minimal medium	
α-Factor[a] 13	Bacterial ketosteroid isomerase + His-tag 131	CNBr	HS[b]	Inclusion bodies	56[a]		Kuliopulos and Walsh, 1994
ACTH(1–24) peptide 24	*E. coli* thioredoxin + His-tag + linker ~120	Enterokinase	None	Periplasm	10	6	Uegaki et al., 1996
PAK pilin peptide 17	*E. coli* OmpA + *de novo* polypeptide 82	CNBr	HS[b]	Periplasm		~2	Tripet et al., 1996 Campbell et al., 1997
Bradykinin 9	Glutathione S-transferase + linker 230	Plasma kallikrein	None	Cytoplasm		~0.3	Ottleben et al., 1997
S-peptide 20	Polypeptide + His-tag ~50	Thrombin	GS[c]	Cytoplasm		1.5	Alexandrescu et al., 1998
C-peptide[a] 31	HSA binding region of Streptococcal protein G ~210	Trypsin and Carboxy-peptidase B	None	Cytoplasm	23[a]		Jonasson et al., 1998
Mastoparan-X 14	Yeast ubiquitin + His-tag 86	Ubiquitin hydrolase	None	Inclusion bodies and cytoplasm		0.6	Kohno et al., 1998
LF12 11	Bacterial ketosteroid isomerase + His-tag 131	CNBr	HS[b]	Inclusion bodies	6		Majerle et al., 2000
Neurotensin 13	Glutathione S-transferase + linker 218	CNBr	None	Cytoplasm	1.8		Williamson et al., 2000
Neu peptide 43	Anthranilate synthase + His-tag 323	CNBr	$(His)_6$ and HS[b]	Inclusion bodies		5–10	Jones et al., 2000
Vpu(2–37) 36	Part of the *E. coli* trpΔLE 1413 polypeptide ~105	CNBr	None	Inclusion bodies		0.7	Opella et al., 2001
V3 peptide 23	RNA binding domain of hnRNCP1 + linker 113	CNBr	None	Inclusion bodies		6–13	Sharon et al., 2002
S2 peptide 11	GB1 domain of Streptococcal protein G + linker 66	Factor Xa	None	Cytoplasm		15	This manuscript
CMPcc 47	GB1 domain of Streptococcal protein G + linker 66	Factor Xa	None	Cytoplasm		48	This manuscript

[a]Tandem repeats of the peptide linked to the fusion partner.
[b]C-terminal homoserine (lactone) as a result of CNBr cleavage.
[c]Residues Gly-Ser are retained at the N-terminus of the target peptide based on thrombin specificity.

accumulate in the cytoplasm (Ottleben et al., 1997; Jonasson et al., 1998; Alexandrescu et al., 1998; Williamson et al., 2000), in the periplasmic space (Uegaki et al., 1996; Tripet et al., 1996), or in inclusion bodies (Kuliopulos and Walsh, 1994).

In the case of membrane spanning peptides or peptides that are toxic to the host, fusion partners that tend to direct the fusion protein into inclusion bodies seem to be the preferred choice (Jones et al., 2000; Majerle et al., 2000; Opella et al., 2001). Proteins in inclusion bodies are largely protected from degradation by cellular proteases. Protein recovery and purification are possible under denaturing conditions either using high concentrations of guanidine hydrochloride or urea. In some cases it may still be possible to isolate the target peptide by exchanging the protein, subsequent to its initial purification under denaturing conditions, into a suitable buffer at a low protein concentrations for protease cleavage. On the other hand, chemical cleavage with cyanogen bromide, frequently used for release of target peptide after solubilization of inclusion bodies, indiscriminately severs all methionyl peptide bonds in the fusion protein and converts the methionines at the C-termini of the resulting fragments into homoserine (lactone) (Gross, 1967). Depending on the design of the fusion protein this might add an unnatural homoserine (lactone) at the C-terminus of the target peptide.

Strategies for the expression of small peptides differ greatly with respect to the carrier protein, accumulation site, cleavage strategy, and most importantly yield of purified peptide per liter of medium. Table 1 summarizes published fusion strategies used for recombinant production of peptides smaller than 50 residues in length in *E. coli*. The GB1 domain is one of the smallest carrier proteins producing a high peptide yield. For soluble peptides, the cloning, expression and purification steps using the GB1 fusion system prove to be one of the simplest procedures available. Fusion proteins containing repeats of the target peptide are expected to give a higher yield than the GB1 fusion system. However, such systems involve a more complex purification scheme and the resulting peptides carry some modifications (Kuliopulos and Walsh, 1994; Jonasson et al., 1998). For hydrophobic peptides, the GB1 fusion strategy may not be the system of choice. For example, a gene fusion encoding GB1 linked to a hydrophobic peptide corresponding to the M2 proton channel of the influenza A virus (Lamb et al., 1985) failed to give significant expression as compared to a fusion with the KSI domain (J.J. Chou and J.M. Louis, unpublished result). The somewhat lower yield for the three water-soluble CD4 peptides also indicates that the expression level of GB1-peptide fusions depends on the sequence composition of the target peptide. The strategies listed in Table 1 may help to choose an appropriate system for the production of a given target peptide or to inspire newer methods.

The extremely slow rotational diffusion of membranous particles prohibits direct characterization of membrane proteins by liquid state NMR. However, the bound structure of a ligand that interacts only transiently with a membrane protein can be obtained from high resolution NMR experiments on the free form of the ligand. For example, residual dipolar couplings can be used to characterize the membrane protein bound ligand (Koenig et al., 2002; Koenig, 2002). In case of small peptides moving freely in solution, internuclear dipolar couplings average to zero due to fast isotropic rotational diffusion. Transient binding of a peptide to a practically immobilized membrane protein that is aligned with respect to the magnetic field can transfer a small degree of alignment to the peptide (Koenig et al., 2000). In a partially aligned peptide, dipolar couplings between pairs of nuclei no longer average to zero. Most importantly, these residual dipolar couplings carry precise information on the structure and orientation of the bound ligand. Provided the peptide alignment is sufficiently weak, the high-resolution character of the NMR spectrum is preserved and the largest, i.e., mainly one-bond, dipolar couplings can be measured as changes of the corresponding J splittings (Tjandra and Bax, 1997; Tjandra et al., 1997). Use of residual dipolar couplings in molecular dynamics based calculation of peptide structure is feasible only, if a large number of dipolar couplings is available. Uniform labeling of the peptide with ^{13}C and ^{15}N is a key requirement for rapid detection of a sufficiently large set of one-bond heteronuclear ^{13}C-^{1}H and ^{15}N-^{1}H dipolar couplings.

In case of the short S2 peptide, heteronuclear ^{1}J splittings are conveniently extracted from the vertical traces of two-dimensional ^{1}H coupled HSQC spectra. All expected ^{1}H-^{13}C correlations were identified in the ^{1}H-^{13}C CT-HSQC spectrum of the U-^{13}C, ^{15}N doubly labeled S2 peptide. Peaks were assigned based on previously determined proton chemical shifts (deposited at the BMRB database, accession code 5376). Isotope labeling with ^{13}C allows recording of the complete CT-HSQC with two scans in as little as 9 min. The Cα-Hα region is presented in Figure 3A and shows only very little peak overlap. ^{1}H-^{15}N correlations of nine

backbone amides and of the side chain NH_2 of Asn4 are observed in the ^{1}H-^{15}N HSQC of isotope labeled S2 peptide in Figure 3B. The backbone amine protons of Ile1 and the backbone amide proton of Arg2 are not detected due to fast chemical exchange with the solvent.

The spectra of free S2 peptide shown in Figure 3 were recorded prior to photo activation of the peptide-binding meta II state of rhodopsin. Light exposure of rhodopsin results in distinct changes of the ^{1}J splittings and allows to extract a total of 38 residual dipolar couplings which characterize the rhodopsin bound S2 peptide (Koenig et al., 2002).

Availability of partially deuterated U-^{13}C, ^{15}N labeled biomolecules (Gardner and Kay, 1998) as well as a new class of multidimensional NMR experiments that are based on transverse relaxation optimized spectroscopy (TROSY) (Pervushin et al., 1997) have significantly increased the size of proteins and protein complexes amenable to solution NMR (Kay, 2001). Triply labeled peptides are needed to take full advantage of the TROSY methodology for investigation of small peptides in the context of noncovalently bound large complexes. As demonstrated here for CMPcc, a fusion of a small soluble peptide with GB1 is an attractive system for efficient recombinant production of such triply labeled peptides. A two-dimensional ^{1}H-^{13}C HNCO-TROSY of trimeric CMPcc is shown in Figure 4. High mobility of the N-terminus of the peptide is reflected by the absence of H^N(i)-C′(i−1) correlations for several N-terminal residues in agreement with previous NMR experiments on CMPcc (Wiltscheck et al., 1997).

Conclusions

GB1-peptide fusions allow efficient recombinant production of small soluble peptides in *E. coli*. A simple purification scheme involving a specific protease cleavage step to yield a high amount of the target peptide characterizes the system. Finally, GB1-peptide fusions provide an affordable way to make isotope labeled peptide for NMR studies.

Acknowledgements

We thank Angela Gronenborn, James Chou, and Stephan Grzesiek, for stimulating discussions. Sonja Hess is gratefully acknowledged for performing ESMS analysis of the S2 fusion protein and peptide, and Victor Jaravine for NMR experiments on CMPcc.

References

Alexandrescu, A.T., Rathgeb-Szabo, K., Rumpel, K., Jahnke, W., Schulthess, T. and Kammerer, R.A. (1998) *Protein Sci.*, **7**, 389–402.

Argraves, W.S., Deak, F., Sparks, K.J., Kiss, I. and Goetinck, P. F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 464–468.

Campbell, A.P., Bautista, D.L., Tripet, B., Wong, W.Y., Irvin, R.T., Hodges, R.S. and Sykes, B.D. (1997) *Biochemistry*, **36**, 12791–12801.

Campbell, A.P., Spyracopoulos, L., Irvin, R.T. and Sykes, B.D. (2000) *J. Biomol. NMR*, **17**, 239–255.

Campos-Olivas R., Louis J.M., Clerot D., Gronenborn B. and Gronenborn A.M. (2002). *Proc. Natl. Acad. Sci. USA*, **99**, 10310–10315.

Clore, G.M. and Gronenborn, A.M. (1982) *J. Magn. Reson.*, **48**, 402–417.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

di Guan, C., Li, P., Riggs, P.D. and Inouye, H. (1988) *Gene*, **67**, 21–30.

Gardner, K.H. and Kay, L.E. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.

Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A.D. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 106–110.

Gronenborn, A.M., Filpula, D.R., Essig, N.Z., Achari, A., Whitlow, M., Wingfield, P.T. and Clore, G.M. (1991) *Science*, **253**, 657–661.

Gross, E. (1967) *Meth. Enzymol.*, **11**, 238–255.

Huth, J.R., Bewley, C.A., Jackson, B.M., Hinnebusch, A.G., Clore, G.M. and Gronenborn, A.M. (1997) *Protein Sci.*, **6**, 2359–2364.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F. and Boyer, H.W. (1977) *Science*, **198**, 1056–1063.

Jonasson, P., Nygren, P.A., Johansson, B.L., Wahren, J., Uhlen, M. and Stahl, S. (1998) *Gene*, **210**, 203–210.

Jones, D.H., Ball, E.H., Sharpe, S., Barber, K.R. and Grant, C. W. (2000) *Biochemistry*, **39**, 1870–1878.

Kay, L.E. (1998) *Nat. Struct. Biol*., **5** (Suppl.), 513–517.

Kay, L.E. (2001) *Meth. Enzymol.*, **339**, 174–203.

Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.

Koenig, B.W. (2002) *ChemBioChem*, **3**, 975–980.

Koenig, B.W., Kontaxis, G., Mitchell, D.C., Louis, J.M., Litman, B.J. and Bax, A. (2002) *J. Mol. Biol.*, **322**, 441–461.

Koenig, B.W., Mitchell, D.C., König, S., Grzesiek, S., Litman, B.J. and Bax, A. (2000) *J. Biomol. NMR*, **16**, 121–125.

Kohno, T., Kusunoki, H., Sato, K. and Wakamatsu, K. (1998) *J. Biomol. NMR*, **12**, 109–121.

Kuliopulos, A. and Walsh, C.T. (1994) *J. Am. Chem. Soc.*, **116**, 4599–4607.

Lamb R.A., Zebedee S.L. and Richardson C.D. (1985) *Cell*, **40**, 627–633.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F. and McCoy, J.M. (1993) *Biotechnology (NY)*, **11**, 187–193.

Majerle, A., Kidric, J. and Jerala, R. (2000) *J. Biomol. NMR*, **18**, 145–151.

Merrifield, B. (1995) In *Peptides: Synthesis, Structure, and Applications*, Gutte, B. (Ed.), Academic Press, San Diego, pp. 94–170.
Nagai, K. and Thogersen, H.C. (1987) *Meth. Enzymol.*, **153**, 461–481.
Ni, F. (1994) *Prog. NMR Spectrosc.*, **26**, 517–606.
Opella, S.J., Ma, C. and Marassi, F.M. (2001) *Meth. Enzymol.*, **339**, 285–313.
Ottleben, H., Haasemann, M., Ramachandran, R., Gorlach, M., Muller-Esterl, W. and Brown, L.R. (1997) *Eur. J. Biochem.*, **244**, 471–478.
Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371.
Santoro, J. and King, G.C. (1992) *J. Magn. Reson.*, **97**, 202–207.
Sharon, M., Gorlach, M., Levy, R., Hayek, Y. and Anglister, J. (2002) *Protein Expr. Purif.*, **24**, 374–383.
Smith, D.B. and Johnson, K.S. (1988) *Gene*, **67**, 31–40.
Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.
Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.
Tripet, B., Yu, L., Bautista, D.L., Wong, W.Y., Irvin, R.T. and Hodges, R.S. (1996) *Protein Eng.*, **9**, 1029–1042.
Tugarinov, V., Zvi, A., Levy, R. and Anglister, J. (1999) *Nat. Struct. Biol.*, **6**, 331–335.
Uegaki, K., Nemoto, N., Shimizu, M., Wada, T., Kyogoku, Y. and Kobayashi, Y. (1996) *FEBS Lett.*, **379**, 47–50.
Vanatalu, K., Paalme, T., Vilu, R., Burkhardt, N., Junemann, R., May, R., Ruhl, M., Wadzack, J. and Nierhaus, K.H. (1993) *Eur. J. Biochem.*, **216**, 315–321.
Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson.*, **98**, 428–435.
Walters, K.J., Ferentz, A.E., Hare, B.J., Hidalgo, P., Jasanoff, A., Matsuo, H., and Wagner, G. (2001) *Meth. Enzymol.*, **339**, 238–258.
Williamson, P.T., Roth, J.F., Haddingham, T. and Watts, A. (2000) *Protein Expr. Purif.*, **19**, 271–275.
Wiltscheck, R., Kammerer, R.A., Dames, S.A., Schulthess, T., Blommers, M.J., Engel, J. and Alexandrescu, A.T. (1997) *Protein Sci.*, **6**, 1734–1745.
Yang, D. and Kay, L.E. (1999) *J. Biomol. NMR*, **13**, 3–10.