Production of recombinant isotopically labelled peptide by fusion to an insoluble partner protein: generation of integrin αvβ6 binding peptides for NMR

Jane L. Wagstaff, Mark J. Howard and Richard A. Williamson

The integrin αvβ6 is up-regulated in several cancers and has clinical potential for both tumour imaging and therapy. Peptide ligands have been developed which show good binding specificity for αvβ6 and provide an opportunity to study the interaction in more detail by NMR. Such studies ideally require ^{15}N and ^{13}C labelled peptides, and recombinant expression within *E. coli* provides a cost effective way of generating isotopically labelled proteins and peptides. In this study we have used an insoluble fusion partner (ketosteroid isomerase) to produce high yields of recombinant peptide. The insoluble nature of the fusion allowed simple product recovery by cell lysis and centrifugation, and thorough washing of the insoluble pellet to remove contaminating proteins avoided the need for nickel-affinity chromatography in denaturing conditions which is the standard procedure. The protocol described here is convenient to scale-up and requires only one chromatography step (reverse-phase HPLC) which is comparable to solid-phase synthesis.

Introduction

The epithelial-specific integrin αvβ6 is expressed at low or undetectable amounts in normal human adult tissue but up-regulated during tissue remodelling processes, *e.g.* wound healing, development and inflammation,[1] and in more severe pathologies such as chronic skin wounds[2] and cancer.[3] αvβ6 binds with high-affinity to protein ligands containing an RGD (Arg-Gly-Asp) motif. Known naturally occurring ligands include fibronectin, tenascin, the latency associated peptides (LAP) of TGFβ1[4] and TGFβ3,[5] and the VP1 coat protein of foot-and-mouth disease virus (FMDV).[6] The amino acid consensus sequence RGDLxxL/I has been shown to be a key determinant of the binding specificity of αvβ6[7] and experiments with a 15-mer peptide containing this motif were shown to prevent infection of baby hamster cells by FMDV.[8]

Peptide antagonists for αvβ6 have the potential for clinical use in both the treatment and imaging of several cancers including pancreatic and colon cancer. Structural analysis of 20-mer peptides from the VP1 coat protein of FMDV and the latency associated peptide of TGFβ1 revealed that secondary structure is also important for binding potency, in particular, an α-helix immediately proximal to the RGD hairpin that brings the two Leu residues in the consensus sequence together on the same face of the helix.[9] The potency of different peptides was also found to be related to the length and propensity of helix formation, and furthermore, NMR experiments revealed that the two Leu residues are in close contact with the integrin when the peptide is bound which suggests that the helix is present in the bound conformation of the peptide.[9]

To enable further NMR-based studies into the structure, dynamics and binding interactions of peptide ligands for αvβ6, it was desirable to generate ^{15}N and ^{13}C isotopically labelled peptides. Such an approach allows the use of more sophisticated multi-dimensional NMR and nucleus-selective editing techniques to provide more detailed information than possible with homonuclear ^{1}H NMR alone.[10]

The production of peptides for NMR analysis has typically been carried out by solid-phase synthesis. This method is convenient and relatively inexpensive, but costs increase significantly for isotopically labelled peptides (^{2}H, ^{13}C, ^{15}N). A cheaper alternative is to use recombinant expression in *E. coli*. This method allows large quantities of peptide to be generated with higher fidelity than chemical synthesis, and the host can be grown on relatively inexpensive media containing ^{2}H, ^{13}C and ^{15}N as sole hydrogen, carbon and nitrogen sources.

Several expression systems have been tried for the recombinant production of peptides in *E. coli*. Early work revealed that peptides need to be expressed with a fusion partner or in tandem constructs to protect the product from intracellular degradation.[11] Several fusion systems are commercially available, the most popular being fusions with maltose binding protein, glutathione S-transferase, small ubiquitin-like modifier and thioredoxin. These fusion partners are highly expressed, soluble and provide a convenient method of affinity purification. The highest peptide yields have been obtained from those fusion partners with the lowest molecular weight, and good results have been obtained from the B1 domain of protein G which is only 6 kDa in size (12–50 mg l^{-1} purified peptide[12]). Peptides can be cleaved from the soluble fusion using restriction proteinases, *e.g.* thrombin, enterokinase or factor Xa.

An alternative to using a system that generates a soluble fusion protein on intracellular expression in *E. coli* is to use a

Protein Science Group, School of Biosciences, University of Kent, Canterbury, CT2 7NJ, UK. E-mail: r.a.williamson@kent.ac.uk, m.j.howard@kent.ac.uk

fusion partner that is highly insoluble. Such a system has a number of benefits. Insoluble proteins often have the highest expression yields,[13,14] and as the recombinant product is sequestered into inclusion bodies, the peptide is further protected from degradation. The formation of inclusion bodies allows the recombinant product to be recovered easily by simple cell lysis followed by centrifugation. The sequestration also allows the expression of toxic peptides at high level.[15,16] Several systems have been described,[17–21] of which one (ketosteroid isomerase; 125 residues) is commercially available (pET31b, Novagen).

In this study, four 20-mer peptide ligands for $\alpha v \beta 6$ were expressed as C-terminal fusions with ketosteroid isomerase (KSI) in E. coli.[17] The pET31b vector allows the insertion of tandem copies of the DNA peptide-coding sequence to increase final peptide yield, and constructs with 1, 2 or 3 peptide repeats were characterised. The vector codes for a C-terminal His$_6$-tag to allow nickel-affinity purification of the fusion protein in denaturing conditions. The peptides are then released from the fusion partner, His$_6$ tag and each other by cyanogen bromide (CNBr) cleavage at methionine residues that separate each peptide sequence. In this work we have shown that the standard nickel-affinity purification step in denaturing conditions used by the majority of workers can be replaced by an inclusion body washing procedure to remove contaminating E. coli proteins that is both more convenient and less costly. Final peptide yields of between 5 and 15 mg l^{-1} of culture were obtained from minimal media with ^{15}N and/or ^{13}C incorporation.

Materials and methods

Design and construction of expression vector

HPLC purified coding and non-coding DNA sequences for four 20-mer RGD-containing peptides were purchased from MWG (London, UK) with 5$'$ phosphate groups. The sequences were optimised for E. coli codon usage and designed to give ATG overhangs at both 3$'$ ends when annealed (the ATG nucleotides code for the Met residues between each peptide). The amino acid sequences of the 4 peptides were: A20fmdv2 (NAVPNLRGDLQVLAQKVART), A20lap2 (GFTTGRRGDLATIHGLNRPF), DBD1 (EKCPNLRGD-LQVLAQKVCRT) and TGFβ3 (NEDDHQRGDLGRLKK-QKDHH).

The coding and non coding oligonucleotides (1.0 pmol μl^{-1}) for each peptide were annealed by heating to 100 °C or 121 °C in TES buffer (20 mM Tris-HCl, 2 mM EDTA, 100 mM NaCl, pH 8.0) for 5 min in an insulated water bath or bench autoclave before allowing to cool slowly over several hours to room temperature. Samples of the annealed DNA were ligated with pET31b(+) previously cleaved with the restriction enzyme AlwNI to generate compatible ends and then dephosphorylated. E. coli (DH5α) were transformed with the ligated DNA and ampicillin-resistant colonies screened for the correct incorporation of one or more DNA inserts by DNA restriction analysis (using XhoI and XbaI enzymes), PCR (using primers that anneal at the T7 terminator sequence (GCTAGTTATTGCTCAGCGG) and within the KSI coding

sequence (GGCAAGGTGGTGAGCATC)) and DNA sequencing of the inserted region (Lark Technologies, UK).

Protein expression

E. coli (strain BL21 (DE3); Novagen) were transformed with pET31b containing 1, 2 or 3 peptide repeats and grown at 37 °C in Luria Bertani (LB) medium or M9 minimal medium containing ^{15}N ammonium sulphate (0.6 g l^{-1}) or ^{13}C glucose (2 g l^{-1}) as sole nitrogen or carbon sources. Cultures were grown at 37 °C with shaking at 200 rpm until the cell density as measured by absorbance at 600 nm reached 0.6. At this point, IPTG (1.0 mM) was added to induce recombinant protein expression and the culture incubated under the same conditions for a further 3 h.

Inclusion body isolation and washing

After the induction period (final A$_{600}$ = 1.2–1.4), E. coli were harvested from 400 ml cultures by centrifugation (6000g, 15 min, 4 °C) and resuspended in 10 ml of lysis buffer (50 mM Tris-HCl, 2 mM EDTA, pH 8). Cells were lysed by the addition of lysozyme (0.1 mg ml^{-1}) and Triton X-100 (0.1% v/v) as previously described.[22,23] The genomic DNA released on cell lysis was removed by incubation with DNase I (0.02 mg ml^{-1}) and MgCl$_2$ (10 mM) for 20 min at room temperature. To ensure complete cell breakage, lysis mixtures were also cooled on ice and ultrasonicated for 2 min with alternate 10 s pulse and wait periods.

Insoluble protein was collected by centrifugation (12 000g, 10 min at 4 °C) and resuspended in 2.5 ml of wash buffer (50 mM Tris-HCl, 10 mM EDTA, 0.5% (w/v) Triton X-100, pH 8.0). The suspension was transferred to a hand-held ground-glass homogenizer to allow maximal disaggregation of the inclusion body pellet. The inclusion bodies were recovered from the wash buffer by centrifugation (12 000g, 10 min at 4 °C) and the washing step repeated. The procedure was finally repeated twice with water to remove detergent from the final preparation.

Cyanogen bromide cleavage and peptide recovery

Washed inclusion bodies were dissolved in 5 ml of formic acid (80% (v/v)) to which CNBr (0.2 g) was added and the mixture incubated for 18–24 h at room temperature in the dark. Non-volatile material was recovered from the digest by lyophilization after first diluting the mixture 5-fold with water. This material was then resuspended in 2.5 ml of 50 mM sodium phosphate buffer containing 100 mM NaCl and the pH adjusted to 7.5–8.0 by the addition of NaOH. The suspension was stirred for 16 h at room temperature in the dark to allow solubilisation of the released peptides. The soluble fraction was collected by centrifugation.

Modifications to the procedure were made for the A20lap2 and DBD1 peptides. CNBr cleavage of A20lap2 was carried out in 80% (v/v) trifluoroacetic acid (TFA) as an alternative to formic acid[24] to avoid formylation of Thr residues in the sequence. The DBD1 peptide was designed to be crosslinked by an intramolecular disulfide bond and solubilisation of this peptide was carried out in 50 mM Tris-HCl, 100 mM NaCl (pH 8.0) containing dithiothreitol (DTT) at a 2-fold

molar excess over peptide concentration to ensure the cysteine residues remained reduced during the solubilisation period. After separation of the KSI carrier by centrifugation, the DBD1 peptide was oxidized by the addition of a 4-fold molar excess of oxidized glutathione before purification by reverse-phase HPLC.

Reverse-phase HPLC purification

Resolubilised peptides were purified to homogeneity by HPLC on a preparative C18 reverse-phase column (5 mm × 150 mm, Vidac) at room temperature. Peptides were eluted with a linear gradient of acetonitrile (containing 0.045% TFA) in water (containing 0.05% TFA). Absorbance was measured at 220 nm and peptide fractions were collected manually to minimize cross-contamination. Peptides were recovered from the HPLC fractions by lyophilisation and their masses determined by electrospray mass spectrometry.

Peptide yield quantification

Estimates for peptide yield were made by comparison of peak areas seen on reverse-phase HPLC with known amounts of an A20fmdv2 peptide standard. The standard peptide was produced recombinantly as a KSI-fusion and purified by reverse-phase HPLC as described above. The concentration of the standard was determined by amino acid analysis (Alta Biosciences, Birmingham, UK).

Nuclear magnetic resonance (NMR)

$^{15}N/^{1}H$ HSQC spectra were acquired on 1.0 mM peptide samples in 20 mM sodium phosphate buffer (pH 6.5) containing 100 mM NaCl and 10% (w/v) $^{2}H_2O$ at 10 °C. Spectra were collected on a 600 MHz Varian UnityINOVA spectrometer equipped with a 5 mm HCN z-pulse field gradient probe as previously described.[25] Solvent-suppression to reduce the water signal was carried out by the WATERGATE method.[26] Data was processed using NMRPipe[27] and analysed using the CCPN Analysis package.[28]

Results and discussion

Ligation of the *Alw*NI cut pET31b vector with the annealed double-stranded DNA coding for the 20-mer peptides typically produced constructs with 1, 2 or 3 tandem inserted sequences as determined by DNA restriction analysis and PCR. The likelihood of a particular colony containing a construct with 1, 2 or 3 inserts was found to be 53%, 17% and 4%, respectively. DNA sequencing revealed high error rates within the annealed region of the DNA inserts, requiring constructs from up to 3 separate clones to be sequenced to ensure that a correct coding sequence was found. Error rates were generally lower when peptides were annealed by slow cooling from 121 °C, rather than 100 °C as initially used. Small scale preliminary work (data not shown) with peptide A20fmdv2 revealed that final peptide yields were highest from constructs containing 2 tandem repeats, and hence two-repeat constructs were selected for all further production work. This finding is consistent with other published work using peptides of similar length.[12,29,30]
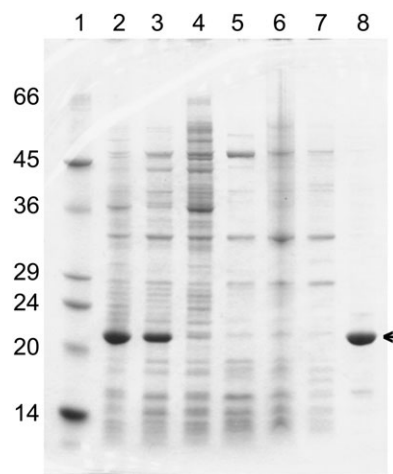

Fig. 1 Expression and purification of a KSI fusion containing two sequence repeats for the peptide A20fmdv2 analysed by SDS-PAGE. Lane 1, molecular weight marker proteins (kDa); lane 2, total *E. coli* proteins after IPTG induction of KSI-A20fmdv2 expression; lane 3, insoluble protein fraction separated by centrifugation after cell lysis; lane 4, soluble protein fraction after cell lysis; lanes 5 and 6, wash supernatants collected from washing the insoluble fraction (inclusion bodies) with Triton X-100/EDTA; lane 7, wash supernatant with water; lane 8, washed inclusion bodies. Significant purification of the recombinant product (indicated by arrow) was achieved by inclusion body washing. The purity of the recombinant product in lane 8 was estimated by gel densitometry to be >85%.

Constructs of interest were expressed in *E. coli* strain BL21 (DE3) grown in LB medium for unlabelled peptides, or minimal medium containing ^{15}N ammonium sulphate and/or ^{13}C glucose as sole nitrogen and carbon sources for isotopically labelled peptides. The KSI–peptide fusion was found to be entirely insoluble as expected (Fig. 1). Insoluble proteins sequestered into inclusion bodies are conveniently purified from the majority of host cell proteins by simple cell lysis and centrifugation. Such methods typically give purities of 30–50%. However, further purification can be easily achieved by washing the inclusion bodies in detergent or low concentrations of denaturants. Washing KSI–peptide fusions in 0.1% Triton X-100 and EDTA gave inclusion body preparations that were typically >70% recombinant product (Fig. 1). This simple washing step avoided the need for more costly nickel affinity chromatography in denaturing conditions which is the standard and often-used purification approach. Fig. 1 shows SDS-PAGE analysis of samples taken from the various stages of cell lysis and inclusion body washing for KSI-A20fmdv2.

Washed inclusion bodies were solubilised directly in 80% formic acid to allow cleavage with CNBr to release the peptides from the KSI carrier and the C-terminal His$_6$ tag. Efficient cleavage was confirmed by SDS-PAGE analysis of samples taken after CNBr treatment and lyophilisation (Fig. 2). Peptides were recovered by resolubilisation of the lyophilised material in phosphate buffer and then purified to homogeneity by preparative reverse-phase HPLC. A typical reverse-phase HPLC chromatogram is shown in Fig. 3. Chemical cleavage by CNBr generates peptides with an authentic N-terminus, but the methionine residue at the C-terminus of the peptide where cleavage occurs is converted
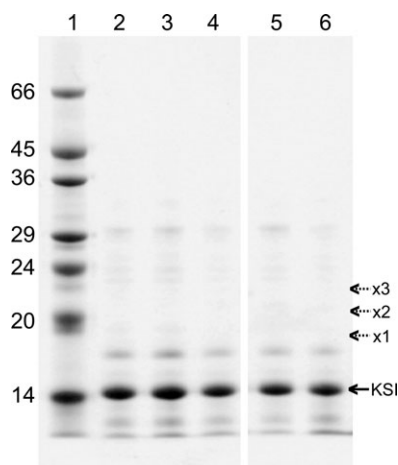
Fig. 2 SDS-PAGE analysis of the remaining insoluble fraction of KSI-A20fmdv2 with different numbers of peptide repeats after CNBr cleavage, lyophilisation and peptide extraction. Lanes 1, molecular weight marker proteins (kDa); lanes 2 and 3, KSI-A20fmdv2 with ×1 repeat; lane 4, KSI-A20fmdv2 with ×2 repeat; lanes 5 and 6, KSI-A20fmdv2 with ×3 repeat. The size of the major protein band (shown by unbroken arrow) is consistent with KSI only hence showing the high efficiency of the CNBr cleavage. The expected migration positions of the different tandem repeat species if chemical cleavage had not taken place are indicated by the broken arrows.
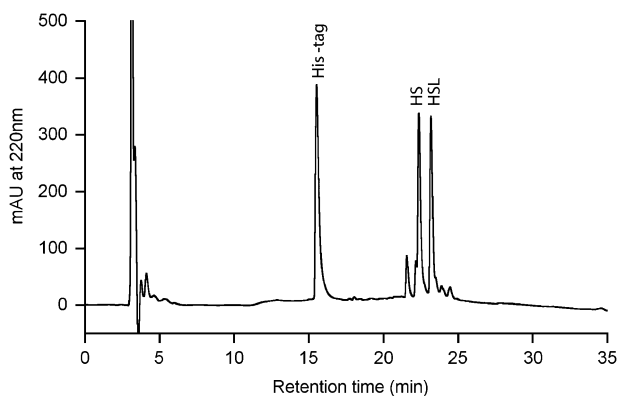

Fig. 3 Reverse-phase HPLC analysis of the A20fmdv2 peptide (NAVPNLRGDLQVLAQKVART) solubilised from the KSI-fusion after CNBr-cleavage and lyophilisation. HS denotes the homoserine derivative, HSL the homoserine lactone. The C-terminal His$_6$-peptide (LLEHHHHHH) is also released from the insoluble KSI carrier.

into a mixture of products (homoserine and homoserine lactone). These products can be resolved by reverse-phase HPLC (Fig. 3), or the mixture converted to a single product either by hydrolysis of the lactone to homoserine by treatment with NaOH, or lactonisation of the homoserine to the lactone by treatment with TFA.[15,31] The proportion of homoserine to homoserine lactone was found to be variable, but typically a 1 : 1 ratio was seen. Although not utilised in this work, the lactone provides a convenient attachment point for specific modification using amine-derivatives of reagents such as biotin or fluorescein.[15,17]

Chemical cleavage with cyanogen bromide is inexpensive and highly efficient, but the approach prevents the use of methionine residues within the peptide sequence. Isoleucine is often chosen as a conservative replacement for methionine, but if methionine is an absolute requirement, then cysteine or tryptophan could be considered as alternative residues for which specific chemical cleavage protocols are available.[17,29] The specificity of the *Alw*N1 restriction enzyme used for cloning in the pET31b vector allows it to be engineered to accept any 3-base 3′ overhang on the insert DNA and hence the opportunity to change the residue at the cleavage site.

Expression and purification yields for the four αvβ6-binding peptides produced as KSI-fusions are shown in Table 1. Expression yields were typically in excess of 80 mg l^{-1} in LB medium and 60 mg l^{-1} in minimal medium, whilst purified peptide yields averaged 18 mg l^{-1} from LB medium and 9 mg l^{-1} from minimal medium. These results are comparable with other data reported in the literature[15,17,29,32,33] and higher than those typically seen with some commonly used soluble fusion partners (*e.g.* glutathione S-transferase[34]). Indeed, we have previously attempted to express ^{15}N-labelled A20lap (a peptide sequence identical to A20lap2 except for a Met to Leu change) as a fusion with glutathione S-transferase and only achieved yields of 0.23 mg l^{-1} (unpublished data). Peptide recovery expressed as a percentage of the theoretical maximum (Table 1) was variable between different experiments and preparations, although SDS-PAGE analysis revealed that peptide cleavage from the carrier protein was essentially complete in all cases (Fig. 2). Significant error in estimating these values is likely to come from the determination of fusion protein expression yield. Due to the insoluble nature of this material, quantities were estimated by comparison of

Table 1 Production of recombinant peptides as insoluble KSI-fusions

Peptide (21 aa)	No. tandem repeats	KSI-peptide length (aa)	%Peptide/KSI fusion (Da/Da)	Yield of KSI fusion[a]/mg l^{-1}		Yield of pure peptide[b,c]/mg l^{-1}	
				LB medium	Minimal medium	LB medium	Minimal medium
A20fmdv2	1	155	13.4	160	—	16 (75)	—
	2	176	23.7	180	130	19 (45)	7.0 (23)
A20lap2	2	176	23.9	86	70	30 (146)[e]	15 (90)
DBD1[d]	2	176	24.5	80	65	7.4 (38)	5.3 (33)
TGFβ3	2	176	25.7	—	70	—	10 (56)

[a] Yields estimated by comparison of coomassie-blue stained protein bands on SDS-PAGE to known standards. [b] Peptide concentrations determined by peak area on reverse-phase HPLC compared to a standard (A20fmdv2) accurately quantified by amino acid analysis. Yields include both the homoserine and homoserine lactone derivative and are for the peptide after preparative reverse-phase HPLC. [c] Values in parentheses are % theoretical maximum yield. [d] CNBr cleavage carried out in 80% TFA to avoid formylation of Thr residues within the sequence. [e] Value of % theoretical maximum yield higher than expected due to suspected error in estimating fusion protein yield where uniform sampling of suspended inclusion preparations is required.

band-staining intensities on SDS-PAGE to known amounts of a standard. Accurate sampling of an inclusion body suspension can be difficult, and band staining comparisons are also prone to error. Peptide recoveries were not on average different to those reported by others,[32] and recoveries could be improved slightly (5–10%) by repeat buffer extractions.

Mass spectrometry analysis of the purified peptides confirmed their authenticity (all experimentally determined masses were within 0.5 Da of the expected mass) and incorporation of ^{15}N and ^{13}C nuclei was typically >98%. NMR analysis also confirmed the chemical authenticity of the peptides, and in the case of A20fmdv2, allowed direct

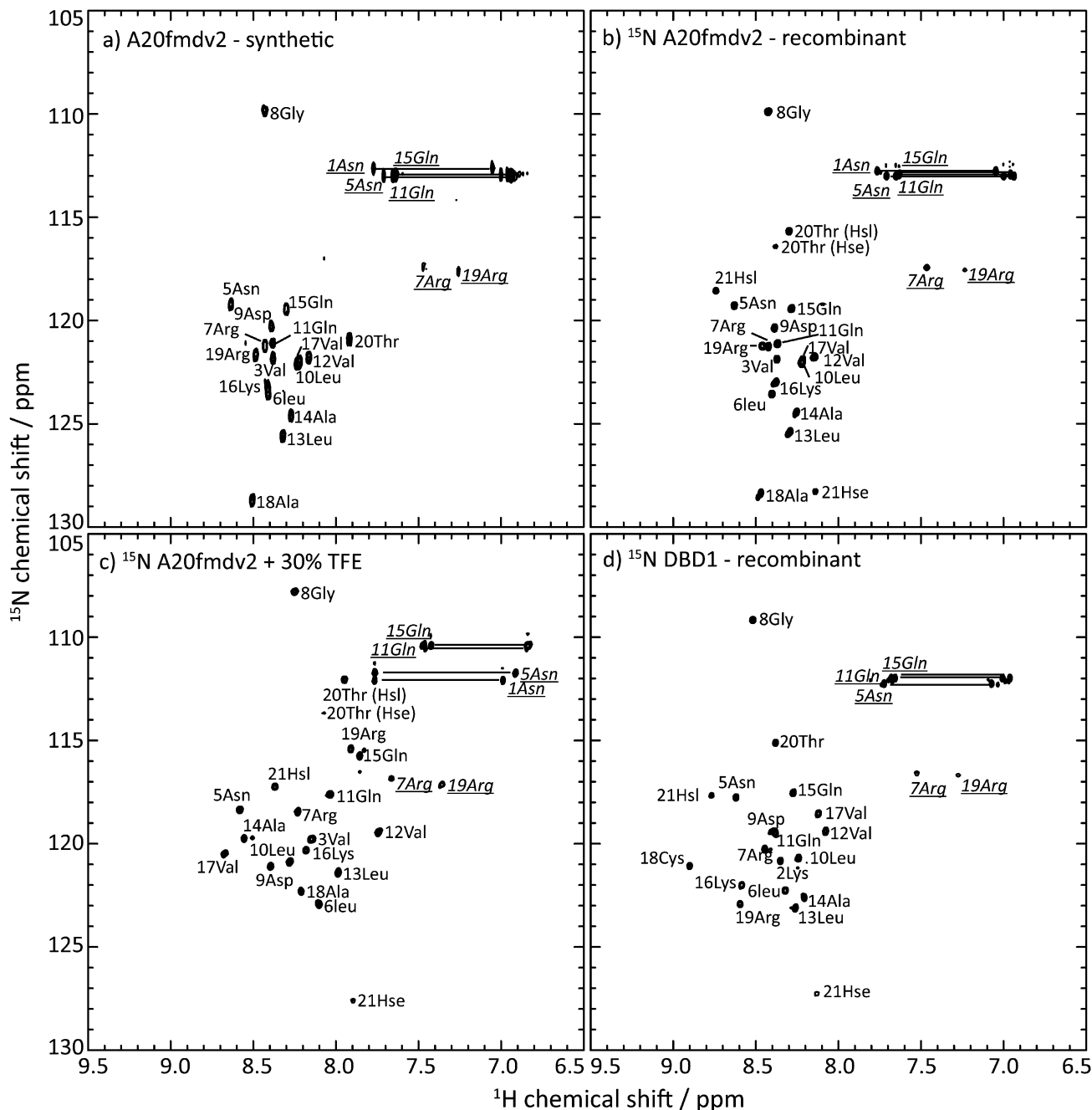**Fig. 4** $^{15}N/^1H$ HSQC spectra of (a) chemically synthesized A20fmdv2 (data collected from naturally abundant ^{15}N isotope (0.37%)), (b) recombinant ^{15}N-labelled A20fmdv2, (c) recombinant ^{15}N-labelled A20fmdv2 in the presence of 30% (v/v) trifluoroethanol, and (d) recombinant ^{15}N-labelled DBD1. Peaks are labelled in each spectrum with the residue number and residue type; labels in regular text are backbone NH resonances, those in italics and underlined are for NH or NH_2 groups in sidechains. Assignments were made from recombinant peptide using standard ^{15}N-edited NOESY and TOCSY experiments. The recombinant samples were enriched in the homoserine lactone derivative (Hsl) by separation on reverse-phase HPLC, although a small amount of the homoserine derivative was also present (Hse). The chemical shift of the penultimate amino acid in A20fmdv2 (b and c) is also affected by the nature of the homoserine derivative and these two species are labeled 20Thr (Hse) and 20Thr (Hsl).

comparison with chemically synthesised material (Fig. 4a and b). The ^{15}N/^{1}H HSQC spectra shown in Fig. 4 show a signal for each backbone amide group (NH), and a pair of peaks for each sidechain NH$_2$ group (Asn and Gln). The recombinant material includes an additional C-terminal homoserine or homoserine lactone residue. The presence of the homoserine residue changes the environment of the preceding residues; hence the peak for the backbone NH of Thr20 is shifted from 121 to 116 ppm in the ^{15}N dimension, and a chemical shift change is also seen for Arg19, but to a much smaller extent. If the presence of a C-terminal homoserine or homoserine lactone residue is undesirable, then a stop codon can be added immediately at the end of the peptide. This approach would prevent the use of tandem constructs, but good yields of recombinant peptide are still obtained even with a single unit (see Table 1). This method would also avoid the production of the His$_6$ tag peptide from which the target peptide must be purified.

The A20fmdv2 peptide has been shown to be essentially unstructured in aqueous solution at pH 6.5.[9] This finding is consistent with the ^{1}H chemical shifts seen for the backbone NH groups (Fig. 4b) that all fall between 8.1 and 8.6 ppm and are typical for a random coil conformation.[35] The addition of the structure-enhancing co-solvent trifluoroethanol (TFE) to A20fmdv2 (Fig. 4c) caused increased dispersion of the signals in the ^{1}H dimension suggesting the formation of secondary structure. NMR distance constraints (NOEs) collected in the presence of TFE revealed a backbone helical conformation immediately proximal to the RGD motif (residues Leu[10] to Val[17]), whereas these NOE signals were absent from NOESY spectra collected in the absence of TFE.[9] The peptide DBD1 (Fig. 4d) was designed as a variant of the A20fmdv2 sequence with an intramolecular disulphide bond between the N- and C-terminal regions to constrain the peptide into a "loop"-like structure. This arrangement stabilises the formation of the helix after the RGD motif in aqueous conditions which is observed in the ^{15}N/^{1}H HSQC spectrum by increased ^{1}H resonance dispersion (Fig. 4d) and confirmed by further NMR structural analysis (unpublished data).

In summary, the KSI fusion approach provides an efficient method of generating recombinant peptides. The method described here avoids the need for nickel affinity purification in denaturing conditions and employs a single column purification step (preparative reverse-phase HPLC) no different than would be routinely used after solid-phase peptide synthesis. Final yields are high and typically several fold greater than could be obtained from other recombinant methods. Importantly, the protocol allows simple scale-up and is particularly useful for generating isotopically labelled peptides for biological NMR studies. The use of an insoluble fusion partner for recombinant peptide production has not been as widely adopted as we feel it deserves.

Acknowledgements

References

1 J. M. Breuss, J. Gallo, H. M. Delisser, I. V. Klimanskaya, H. G. Folkesson, J. F. Pittet, S. L. Nishimura, K. Aldape, D. V. Landers, W. Carpenter, N. Gillett, D. Sheppard, M. A. Matthay, S. M. Albelda, R. H. Kramer and R. Pytela, *J. Cell Sci.*, 1995, **108**, 2241–2251.

2 L. Hakkinen, L. Koivisto, H. Gardner, U. Saarialho-Kere, J. M. Carroll, M. Lakso, H. Rauvala, M. Laato, J. Heino and H. Larjava, *Am. J. Pathol.*, 2004, **164**, 229–242.

3 G. J. Thomas, M. L. Nystrom and J. F. Marshall, *J. Oral Pathol. Med.*, 2006, **35**, 1–10.

4 J. S. Munger, X. Z. Huang, H. Kawakatsu, M. J. D. Griffiths, S. L. Dalton, J. F. Wu, J. F. Pittet, N. Kaminski, C. Garat, M. A. Matthay, D. B. Rifkin and D. Sheppard, *Cell (Cambridge, Mass.)*, 1999, **96**, 319–328.

5 J. P. Annes, D. B. Rifkin and J. S. Munger, *FEBS Lett.*, 2002, **511**, 65–68.

6 T. Jackson, D. Sheppard, M. Denyer, W. Blakemore and A. M. Q. King, *J. Virol.*, 2000, **74**, 4949–4956.

7 S. Kraft, B. Diefenbach, R. Mehta, A. Jonczyk, G. A. Luckenbach and S. L. Goodman, *J. Biol. Chem.*, 1999, **274**, 1979–1985.

8 M. G. Mateu, M. L. Valero, D. Andreu and E. Domingo, *J. Biol. Chem.*, 1996, **271**, 12814–12819.

9 D. DiCara, C. Rapisarda, J. L. Sutcliffe, S. M. Violette, P. H. Weinreb, I. R. Hart, M. J. Howard and J. F. Marshall, *J. Biol. Chem.*, 2007, **282**, 9657–9665.

10 J. L. Wagstaff, S. Vallath, J. F. Marshall, R. A. Williamson and M. J. Howard, *Chem. Commun.*, 2010, DOI: 10.1039/C0CC01846E.

11 S.-H. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 4627–4631.

12 B. W. Koenig, W. Rogowski and J. M. Louis, *J. Biomol. NMR*, 2003, **26**, 193–202.

13 H. L. Lilie, E. Schwatz and R. Rudolph, *Curr. Opin. Biotechnol.*, 1998, **9**, 497–501.

14 E. De Bernardez Clark, *Curr. Opin. Biotechnol.*, 2001, **12**, 202–207.

15 A. Majerla, J. Kidric and R. Jerala, *J. Biomol. NMR*, 2000, **18**, 145–151.

16 I. Čipáková, J. Gašperík and E. Hostinozá, *Protein Expression Purif.*, 2006, **45**, 269–274.

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

18 M. J. Osbourne, Z. Su, V. Sridaran and F. Ni, *J. Biomol. NMR*, 2003, **26**, 317–326.

19 D. H. Jones, E. H. Ball, S. Sharpe, K. R. Barber and C. W. Grant, *Biochem. J.*, 2000, **39**, 1870–1878.

20 M. Sharon, M. Gorlash, R. Levy, Y. Hayek and J. Anglister, *Protein Expression Purif.*, 2002, **24**, 374–383.

21 I. T. Yonemoto, M. R. Wood, W. E. Balch and J. W. Kelly, *Protein Sci.*, 2009, **18**, 1978–1986.

22 F. Marston, *Biochem. J.*, 1986, **240**, 1–12.

23 R. A. Williamson, D. Natalia, C. K. Gee, G. Murphy, M. D. Carr and R. B. Freedman, *Eur. J. Biochem.*, 1996, **241**, 476–483.

24 D. R. Goodlett, F. B. Armstrong, R. J. Creech and R. B. van Breeman, *Anal. Biochem.*, 1990, **186**, 116–120.

25 H. I. Alanen, R. A. Williamson, M. J. Howard, F. S. Hatahet, K. E. H. Salo, A. Kauppila, S. Kellokumpu and L. W. Ruddock, *J. Biol. Chem.*, 2006, **281**, 33727–33738.

26 M. Piotto, V. Saudek and V. Sklenar, *J. Biomol. NMR*, 1992, **2**, 661–665.

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

28 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins*, 2005, **59**, 687–696.

29 B. M. Hartmann, E. Kaar, R. J. Falconer, B. Zeng and A. P. J. Middleberg, *J. Biotechnol.*, 2008, **135**, 85–91.

30 J. M. Riley, A. Aggeli, R. J. Koopmans and M. J. McPherson, *Biotechnol. Bioeng.*, 2009, **103**, 241–251.

31 R. E. Offord, *Biochem. J.*, 1972, **129**, 499–501.

32 S. Sharpe, W.-M. Yau and R. Tycko, *Protein Expression Purif.*, 2005, **42**, 200–210.

33 T.-J. Park, J.-S. Kim, S.-S. Choi and Y. Kim, *Protein Expression Purif.*, 2009, **65**, 23–29.

34 P. T. F. Williamson, J. F. Roth, T. Haddingham and A. Watts, *Protein Expression Purif.*, 2000, **19**, 271–275.

35 A. E. Meekhof and S. M. V. Freund, *J. Mol. Biol.*, 1999, **286**, 579–592.