Journal of Biomolecular NMR, **17:** 337–347, 2000.
KLUWER/ESCOM

Expression of deuterium-isotope-labelled protein in the yeast *Pichia pastoris* for NMR studies

W.D. Morgan*, A. Kragt & J. Feeney
Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

Received 9 May 2000; Accepted 14 June 2000

Key words: deuterium isotope labelling, heterologous protein expression, heteronuclear NMR, methyl isotopomers, methylotrophic yeast, *Pichia pastoris*

Abstract

Deuterium isotope labelling is important for NMR studies of large proteins and complexes. Many eukaryotic proteins are difficult to express in bacteria, but can be efficiently produced in the methylotrophic yeast *Pichia pastoris*. In order to facilitate NMR studies of the malaria parasite merozoite surface protein-1 (MSP1) complex and its interactions with antibodies, we have investigated production of the MSP1-19 protein in *P. pastoris* grown in deuterated media. The resulting deuteration patterns were analyzed by NMR and mass spectrometry. We have compared growth characteristics and levels of heterologous protein expression in cells adapted to growth in deuterated media (95% D_2O), compared with expression in non-adapted cells. We have also compared the relative deuteration levels and the distribution pattern of residual protiation in protein from cells grown either in 95% D_2O medium with protiated methanol as carbon source, or in 95% D_2O medium containing deuterated methanol. A high level of uniform C_α deuteration was demonstrated, and the consequent reduction of backbone amide signal linewidths in [$^1H/^{15}N$]-correlation experiments was measured. Residual protiation at different positions in various amino acid residues, including the distribution of methyl isotopomers, was also investigated. The deuteration procedures examined here should facilitate economical expression of $^2H/^{13}C/^{15}N$-labelled protein samples for NMR studies of the structure and interactions of large proteins and protein complexes.

Abbreviations: BMGlc, buffered minimal medium with glucose; BMMe, buffered minimal medium with methanol; MSP1, malaria parasite merozoite surface protein; YNB, yeast nitrogen base; YPD, yeast extract peptone dextrose medium; OD_{600}, optical density at 600 nm; G418, Geneticin (aminoglycoside antibiotic); Fab, antigen-binding antibody fragment; scFv, single-chain variable-region antibody fragment.

Introduction

As part of a program for development of a malaria vaccine (Holder et al., 1999), we have recently determined the solution structure of the C-terminal fragment (MSP1-19) of the *Plasmodium falciparum* merozoite surface protein-1, an important candidate antigen (Morgan et al., 1999). The MSP1-19 C-terminal fragment, consisting of an EGF module pair (Figure 1), is derived from the MSP1-42 precursor (41.1 kDa) by proteolytic cleavage. This cleavage, designated secondary processing, occurs during or near the erythrocyte invasion event, and is an essential step for the invasion process (Blackman et al., 1991). Two fragments are produced, MSP1-33 (30.5 kDa) and MSP1-19 (10.6 kDa). The structure of the MSP1-33 and MSP1-42 species, as well as other components of the MSP1 complex, are of interest in addition to MSP1-19 for the development of blood-stage malaria vaccines that may inhibit proteolytic cleavage and invasion, and for targeting drugs against the invasion process. Furthermore, investigating the fine specificity

*To whom correspondence should be addressed. E-mail: wmorgan@nimr.mrc.ac.uk

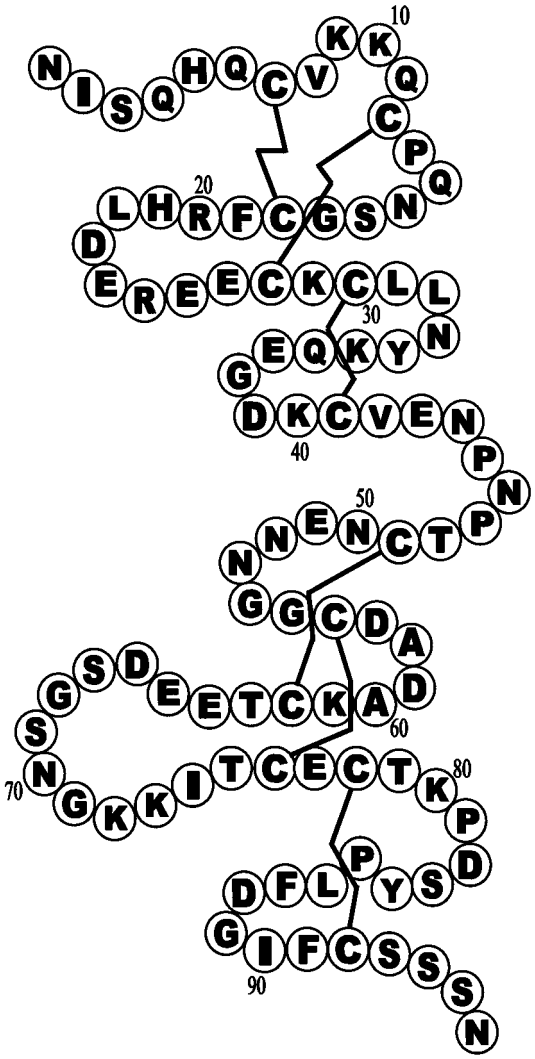

Figure 1. Schematic diagram of the *Plasmodium falciparum* MSP1-19 protein. The 96 amino acid merozoite surface protein-1 C-terminal fragment, consisting of an EGF module pair, is shown with the six disulphide bridges indicated. The fragment studied here also contains the vector-encoded N-terminal sequence YHHHHHH (Morgan et al., 1999). The PBD ID code for the NMR structure is 1cej.

of antibodies will be important for optimization of the candidate antigens. NMR studies of antibody–antigen complexes will be valuable for this work.

Stable isotope labelling is critical for successful multidimensional heteronuclear NMR studies of proteins. Labelling with deuterium (nuclear spin = 1) is an important factor, especially for studies of large proteins and macromolecular complexes. Deuterium labelling provides simplified spectra from the remaining ^{1}H nuclei and also has useful effects on relaxation properties of attached or adjacent atoms (^{1}H, ^{13}C, ^{15}N), and both factors contribute to more effective studies of large proteins (LeMaster, 1994; Markus et al., 1994; Arrowsmith and Wu, 1998). Extensive α-deuteration is particularly important for improving heteronuclear experiments involving the amide NH protons.

Bacterial systems have frequently been used to produce deuterium labelled proteins for NMR studies. (Gardner and Kay, 1998). However, many eukaryotic proteins are inefficiently expressed in bacteria, due to problems related to disulphide bond formation and folding. This is the case for the MSP1-19 protein, which had proved difficult to express in *E. coli*. In some cases, methylotrophic yeasts are suitable for high expression of eukaryotic proteins, and have been widely used in industrial biotechnology. One of these yeasts, *Pichia* (*Komagataella*) *pastoris*, is particularly attractive for this purpose, and yields of recombinant protein up to 12 g/L (30% of total protein synthesis) have been reported (Rosenfeld, 1999). In a few cases, proteins have been labelled with ^{15}N or ^{15}N/^{13}C for NMR studies using this organism, particularly as an alternative to inefficient bacterial systems (Laroche et al., 1994; White et al., 1994; McAlister et al., 1998; Morgan et al., 1999; Wood and Komives, 1999). We have used this approach to prepare the *P. falciparum* MSP1-19 protein, which was poorly expressed in *E. coli*, but efficiently expressed and labelled with ^{15}N/^{13}C in *P. pastoris* (Morgan et al., 1999).

Previous studies have described the growth characteristics of yeasts, including *P. pastoris*, in deuterated media (Katz and Crespi, 1966; Haon et al., 1993; Massou et al., 1999). In order to efficiently prepare isotopically labelled samples of MSP1 proteins, we have investigated ^{15}N/^{2}H and ^{13}C/^{2}H labelling procedures in *P. pastoris*. Here we report results on expression of the partially deuterated ^{15}N/^{13}C-labelled C-terminal fragment MSP1-19 in *P. pastoris*. The growth properties and expression levels of cells adapted to high D_2O concentration have been compared with those of nonadapted cells. The effect of partial deuteration by growth in media containing 95% D_2O, or with deuterated methanol as well as 95% D_2O, on the amide proton linewidth of [^{1}H/^{15}N]-HSQC spectra has been investigated. The distribution of residual protonation at different atom positions in the C-terminal fragment MSP1-19 for the various amino acid types has been established from examination of [^{1}H/^{13}C]-HSQC spectra. The development of protocols described here for deuteration of proteins in this system, and infor-

mation on the distribution pattern of deuterons, will be useful for obtaining suitably deuterated proteins for NMR studies in general, as well as for future studies of larger MSP1-components and of MSP1–antibody complexes.

Materials and methods

NMR spectroscopy

All spectra were acquired in 50 mM sodium phosphate, 50 mM NaCl, pH 6.5, 90% H_2O/10% D_2O at 25 °C, at concentrations of 0.07 to 0.28 mM, in 0.33 to 0.60 ml, with Varian Unity or Inova 600 MHz spectrometers. Samples were purified as described previously (Morgan et al., 1999).

The parameters used for the 2D-[^{1}H/^{15}N]-HSQC experiments were: F1 dimension: nucleus, ^{15}N; complex points, 180; spectral width, 2500 Hz; acquisition time, 72 ms; final data size, 512 points. F2 dimension: nucleus, ^{1}H; complex points, 1216; spectral width, 8000 Hz; acquisition time, 152 ms; final data size, 4096 points. Total time: 4.5 h. Solvent suppression was performed with the WATERGATE sequence (Piotto et al., 1992).

Separate experiments were performed for the aliphatic and aromatic regions in 2D-[^{1}H/^{13}C]-HSQC experiments. The carrier frequency was 39 ppm for the aliphatic region experiment and 119 ppm for the aromatic region experiment. The parameters used were: F1 dimension: nucleus, ^{13}C; complex points, 400 (aliphatic), 320 (aromatic); spectral width, 12000 Hz (aliphatic), 10000 Hz (aromatic); acquisition time, 33 ms (aliphatic), 32 ms (aromatic); 119 ppm (aromatic); final data size, 1024 points. F2 dimension: nucleus, ^{1}H; complex points, 1216; spectral width, 8000 Hz; acquisition time, 152 ms; final data size, 4096 points. Total time: 10 h (aliphatic), 8 h (aromatic). Pulsed field gradients were used for solvent suppression (Wider and Wüthrich, 1993). Deuterium decoupling was not used.

2D-HN(CO)CA-reference and -attenuated spectra for measurement of C_α-fractional deuteration were acquired as interleaved spectra using a 2D variation of the previously reported 3D pulse sequence (Yamazaki et al., 1994; Rosen et al., 1996). The parameters used were: F1 dimension: nucleus, ^{15}N; complex points, 24; spectral width, 2500 Hz; acquisition time, 9.6 ms; final data size, 1024 points. F2 dimension: nucleus, ^{1}H; complex points, 512; spectral width, 8000 Hz; acquisition time, 64 ms; final data size, 4096 points. Total time: 18 h. Solvent suppression was performed with the WATERGATE sequence.

The 2D-HSQC experiments were processed with Felix 98.0, using 90° shifted sinebell-squared window functions. ^{1}H linewidths were measured with Felix using peak optimization of 1D slices. The C_α-fractional deuteration experiment was processed with Felix using 90° shifted sinebell-squared window functions, and relative peak heights and volumes in the reference and attenuated spectra were measured. The C_α-fractional deuteration $f = I_{attenuated}/I_{reference}$ for non-glycine residues, and for glycine residues $I_{attenuated}/I_{reference} = f^2$.

Media

Buffered minimal media with either glucose (BMGlc) or methanol (BMMe) as carbon source were based on Invitrogen Corp. recipes. BMGlc medium contained 100 mM potassium phosphate, pH 6.0; yeast nitrogen base (0.34%, w/v) (DIFCO product number 0335-15: YNB without amino acids and without ammonium sulphate), ammonium sulphate (0.2% w/v); biotin (4×10–5%, w/v); Sigma antifoam 289 (0.01% v/v); D-glucose (0.5% w/v). BMMe medium contained 100 mM potassium phosphate, pH 6.0; yeast nitrogen base (0.34%, w/v), ammonium sulphate (0.2% w/v); biotin (4×10^{-5}%, w/v); Sigma antifoam 289 (0.01% v/v); methanol (0.5% v/v) (Romil, HPLC grade). The amount of carbon and nitrogen sources was adjusted for economical isotope labelling.

The following isotope labelled compounds were employed: D_2O, 99.8% (Norsk Hydro); $(^{15}NH_4)_2SO_4$ 99+ %, d_4-methanol 99.8% and ^{13}C-methanol 99.3% (Isotech); $^{13}C_6$-glucose, 99+%, and ^{13}C-d_4-methanol, ^{13}C 99%, D 99% (Cambridge Isotope Laboratories).

Adaptation of cells to deuterium oxide

Colonies of *Pichia (Komagataella) pastoris* strains SMD1168 (*his4 pep4*) and GS6 (SMD1168 strain transformed with *P. falciparum* MSP1-19/pPIC9K vector (His$^+$, G418^r, Mut$^+$)) were transferred from YPD agar plates into 1 ml liquid YPD media in 5 ml Falcon 2063 tubes and incubated in a shaker-incubator (30 °C, 250 rpm). After 24 h cells were transferred to fresh 10 ml YPD media in 50 ml conical tubes at 0.1 OD_{600}. After another 24 h cells were transferred to 2 ml BMGlc/H_2O in 5 ml tubes at 0.1 OD_{600} and cultured for ~72 h. Cells were subsequently transferred to 2 ml BMGlc/H_2O in 5 ml tubes at OD_{600} ~0.1. Another three identical transfers were done prior to the transfer of cells to fresh 10 ml BMGlc/H_2O

in 50 ml conical tubes at OD_{600} ~0.1. After 24 h cells were transferred to 2 ml BMGlc/25% D_2O in 5 ml tubes. The remainder of the cells were centrifuged for 5 min at 3000 × g and 4 °C, resuspended in 1 ml BMGlc/H_2O containing 15% glycerol and stored at −80 °C. For adaptation of cells to 25%, 50% and 70% D_2O, cells were cultured in BMGlc (containing the corresponding percentage of D_2O) in an identical manner to that used for growth of cells in BMGlc/H_2O. Adapted cells were stored in media containing the same percentage D_2O (25%, 50%, 70%) as was used for growth. For adaptation of cells to BMGlc media containing 90%, 95% and 99.8% D_2O cells were cultured to OD_{600} ~1.0 prior to transfer at OD_{600} ~0.1 in fresh media. For adaptation of cells to 90% and 95% D_2O and adaptation of cells to 99.8% D_2O transfers were done over two and three weeks, respectively to establish sufficient biomass (due to reduced growth rates). Furthermore, culture tubes were incubated at a 45 ° angle to increase aeration. The overall conditions were identical to those described for growth in BMGlc/H_2O.

Growth of nonadapted and adapted cells for growth comparison

Adapted (to growth in 95% D_2O media) and nonadapted GS6 cells (10 μL from frozen stocks) were transferred to 10 ml BMGlc/95% D_2O in 50 ml conical tubes and incubated at 30 °C with shaking at 250 rpm at a 45 ° angle. After ~72 h cells were transferred to 25 ml BMGlc/95% D_2O in 100 ml Erlenmeyer flasks at 0.1 OD_{600}. After ~24 h cells were transferred for the second time to 25 ml BMGlc/95% D_2O in 100 ml Erlenmeyer flasks at 0.1 OD_{600}. Incubation conditions were identical to those described above. Growth data (OD_{600}) was analysed using SigmaPlot 4.01. For comparison with growth in methanol medium, cells grown in BMGlc/95% D_2O were harvested, washed with BMMe/95% D_2O, and resuspended in BMMe at ~1 OD_{600} in 500 ml in 2L baffled Erlenmeyer flasks. Induced cultures were grown as described previously (Morgan et al., 1999).

Expression of isotopically labelled proteins

Biomass was generated in BMGlc media containing the appropriate isotopes, and expression was induced in BMMe media. Methanol was added daily at an amount equal to 0.5% v/v, for 2–4 days post-induction, until harvesting. Secreted MSP1-19 was deglycosylated and purified as described previously (Morgan et al., 1999). ^{2}H-labelled protein was produced with either 95% D_2O medium alone, or 95% D_2O with d_4-methanol. Starter cultures were 15 ml of BMGlc, and induction was done in 150 ml BMMe. ^{15}N/^{2}H-protein was produced with similar media, containing in addition ^{15}N ammonium sulphate. Starter cultures were 150 ml of BMGlc, and induction was done in 500 ml BMMe at ~1 OD_{600} initially. ^{13}C/^{2}H-labelled protein was produced with a culture grown in BMGlc/95% D_2O, $^{13}C_6$-glucose and induced with BMMe/95% D_2O, ^{13}C-methanol. A 100 ml starter culture was grown to an OD_{600} of 6, cells were harvested and resuspended in 100 ml inducing (BMMe) media. The final OD_{600} was 15.1. ^{15}N/^{13}C/^{2}H-labelled protein was produced in a similar fashion using ^{13}C-d_4-methanol during induction. Two 100 ml starter cultures were grown to an OD_{600} of 6.6 to 6.9, and each was induced in 100 ml labelled BMMe (containing 95% D_2O, $^{13}CD_3OD$, and $(^{15}NH_4)_2SO_4$), and grown to a final OD_{600} of 7.4 to 11.1.

Results

The *P. pastoris* strain SMD1168, transformed with the pPIC9K plasmid α-factor-MSP1-19 fusion construct, was first adapted to growth in increasing levels of D_2O, up to 99.8%, as described in the Methods section. This procedure was similar to that performed by Haon et al. (1993). Adapted cultures were stored frozen in 15% glycerol media containing 25%, 50%, 70%, 90%, 95%, and 99.8% D_2O. The culture adapted to 95% D_2O was used for most expression experiments. The media used were buffered minimal glucose (BMGlc) for generating initial biomass, and buffered minimal methanol (BMMe) for the induction phase, as previously described (Morgan et al., 1999). Expression from the *aox1* gene promoter of the vector is repressed in glucose medium, and induced by growth on methanol as the sole carbon source. The transformed strain is highly G418 resistant (indicating the presence of multiple copies of the vector), and has the Mut^+ phenotype (grows normally in methanol media). Under inducing conditions, His_6-tagged, partially glycosylated MSP1-19 is secreted into the culture medium, and the final yield of deglycosylated protein purified as described previously is up to 24 mg/L (Morgan et al., 1999).

The recombinant MSP1-expressing strain and the parent strain SMD1168 in buffered minimal glucose (BMGlc/H_2O) medium have similar growth characteristics (data not shown). The recombinant strain,

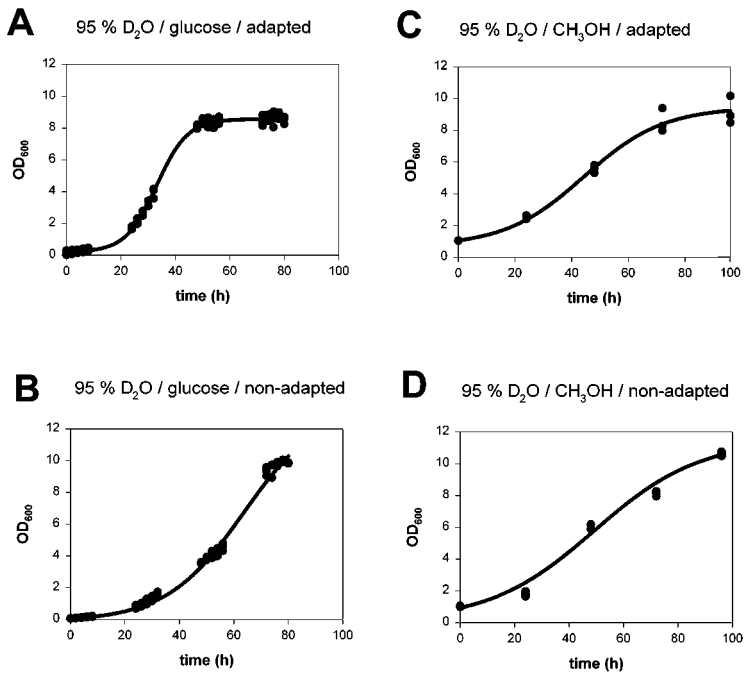

Figure 2. Growth comparison of D_2O-adapted and non-adapted *P. pastoris* cells. Optical density at 600 nm of cultures of *P. pastoris* transformed with the *P. falciparum* MSP1-19 expression vector. (A) Cells adapted to growth in 95% D_2O, in BMGlc/95% D_2O (glucose) medium. (B) Non-adapted cells, in BMGlc/95% D_2O (glucose) medium. (C) Cells adapted to growth in 95% D_2O, in BMMe/95% D_2O (methanol) medium. (D) Non-adapted cells, in BMMe/95% D_2O (methanol) medium.

adapted to growth in 95% D_2O medium, still grows normally when used to inoculate normal BMGlc/H_2O medium (data not shown). This is consistent with previous studies showing that the adaptation process involves somatic rather than genetic changes in the D_2O-adapted cells (Katz and Crespi, 1966). As expected from previous work with methylotrophic yeasts in deuterated media (Haon et al., 1993), the MSP1-transformed *P. pastoris* strain grew more slowly at high D_2O concentrations, with doubling times of 2.5 h (H_2O), 8 h (95% D_2O), and 10 h (99.8% D_2O). Growth data for 95% D_2O are shown in Figure 2A.

These growth curves compare the growth of adapted cells (adapted to 95% D_2O) with unadapted cells, in pre-induction (glucose) media and inducing (methanol) media. Adapted cells (Figure 2A) initially grew faster in BMGlc/95% D_2O than unadapted cells (Figure 2B). In contrast, the unadapted cells showed very slow initial growth, and did not reach stationary phase by the end of the experiment. However, there was little difference in growth of adapted and non-adapted cells in BMMe/95% D_2O (Figure 2C,D). This may indicate that different factors, such as oxygen concentration, influence the growth rate in BMMe/95% D_2O. It should also be noted that the non-adapted cells were grown in BMGlc/95% D_2O before induction. The growth rate in 95% D_2O for both adapted and non-adapted cells is relatively slow compared to that in normal media. Typical induction protocols in BMMe/H_2O gave vigorous growth from an initial value on inoculation of 1, reaching a final OD_{600} level of 18 at 4 days post-induction (Morgan et al., 1999). The relatively slow growth in 95% D_2O during the induction phase indicates that it may be useful to inoculate the methanol medium at a higher initial OD_{600} level. Initial values of 6–7 OD_{600} were used for preparation of $^{13}C/^{2}H$- and $^{15}N/^{13}C/^{2}H$-labelled protein (see Materials and methods).

The yield of protein secreted by the MSP1-transformed cells in BMMe/95% D_2O is shown in Figure 3. The medium was concentrated by ultrafiltration and then analyzed by SDS-PAGE. The average yield of protein from unadapted cells was 66% of the yield from the adapted cells, determined by image analysis of the gel bands.

The MSP1 C-terminal fragment from parallel 500 ml cultures in normal H_2O and 95% D_2O media (adapted cells) was purified to homogeneity as described previously (Morgan et al., 1999). The gel migrations of normal and partially deuterated protein

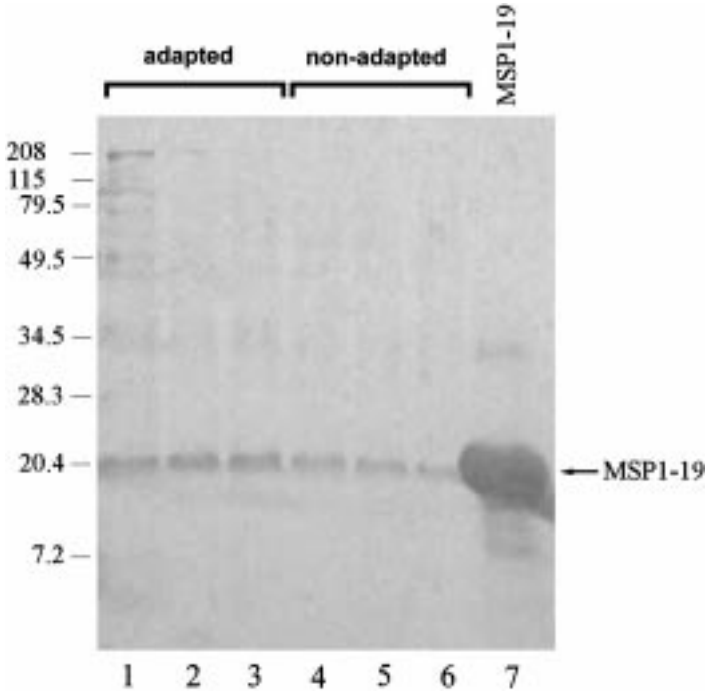

Figure 3. Comparative expression of MSP1-19 in adapted and non-adapted cells, in 95% D_2O medium. Trichloroacetic acid precipitates of supernatants containing secreted MSP1-19, analyzed by SDS-PAGE. Lanes 1–3: cultures of cells adapted to 95% D_2O (in triplicate). Lanes 4–6: cultures of non-adapted cells (in triplicate). Lane 7: purified MSP1-19 protein standard.

were identical. Although the yield is somewhat lower in deuterated media, the level is adequate for practical preparation of isotopically labelled proteins.

Samples of normal and partially deuterated (95% D_2O) MSP1 C-terminal fragment protein were dialyzed against deionized water, lyophilized, dissolved in 50% acetonitrile/0.3% formic acid, and analyzed by electrospray mass spectrometry. The mass of the partially deuterated protein was 12013.2 +/− 3.0 Da, compared to 11604.3 +/− 4.7 Da for the normal protein. This corresponds to an average substitution of deuterons for protons at 72% of the non-exchangeable positions in the protein. As shown below, the distribution of residual protonation is of course not uniform. This sample was produced using protonated carbon sources (glucose and methanol). In a further experiment, the additional deuteration obtained by using deuterated methanol (d-4, 99%) in addition to 95% D_2O was examined. In this experiment, ^{15}N ammonium sulphate was also included in the media for subsequent NMR work, and the mass for the ^{15}N/95% D_2O/CD_3OD product (12225.0 +/− 2.9 Da) is therefore compared to protein from ^{15}N/95% D_2O (12173.4 +/−2.5 Da). This indicates that an additional 9% fractional deuteration was obtained by including CD_3OD during the induction phase. One-dimensional ^1H NMR spectra demonstrated a high level of deuteration of aliphatic residues in samples from 95% D_2O/CH_3OH medium, which was increased somewhat in samples from 95% D_2O/CD_3OD medium (data not shown). The extent of deuteration was further quantified by the experiments discussed below.

The reduction of the amide proton linewidth in partially deuterated samples is of particular interest. This is very important for facilitating NMR studies of large proteins and protein complexes, and represents a major application of fractional deuteration. When samples grown in D_2O media are examined in H_2O-containing buffers, the majority of the amide groups readily exchange with the solvent to give NH groups. Deuteration at the C_α position causes a significant reduction in the proton linewidth of the amide signals. This effect can be seen in one-dimensional slices at the ^{15}N frequency of Gly71 through 2D [^1H/^{15}N]-HSQC data for normal and partially deuterated MSP1 C-terminal fragment samples (Figure 4). The proton linewidths for the Gly71 residue signal were ∼15.5 Hz (non-deuterated ^{15}N/^1H sample), ∼9.8 Hz (95% D_2O medium), and ∼9.3 Hz (95% D_2O medium with CD_3OD). Average proton linewidth values for

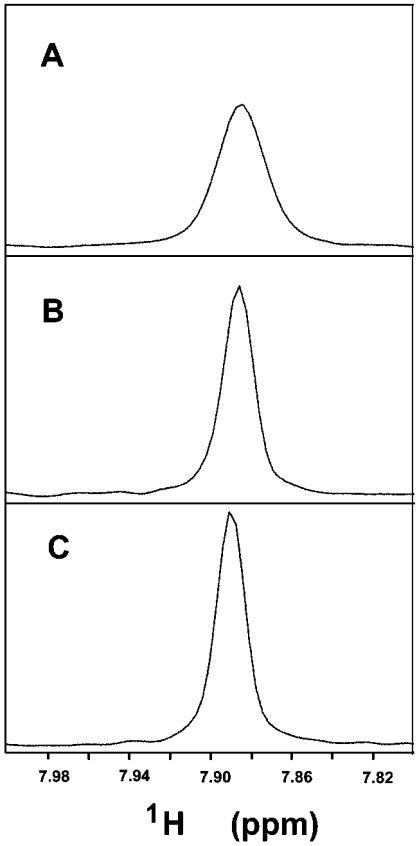

Figure 4. Improvement of ^{1}H-dimension linewidth by deuteration of non-exchangeable positions in ^{15}N/^{2}H-labelled MSP1-19 samples. 1D traces from 2D-[^{1}H/^{15}N]-HSQC spectra at the ^{15}N frequency of the Gly 71 residue. (A) ^{15}N-labelled (non-deuterated control) protein. (B) ^{15}N/^{2}H-labelled protein (grown in 95% D_2O with protiated methanol). (C) ^{15}N/^{2}H-labelled protein (grown in 95% D_2O with deuterated methanol (CD_3OD)).

residues with well-resolved signals that do not show exchange-line broadening were 17.1, 11.9, 11.3 +/− 1.5 Hz, respectively for the three sample types. These results indicate that fractional deuteration by growth in 95% D_2O causes an average linewidth reduction of ∼30%, with only a small additional reduction (3%) resulting from the inclusion of deuterated methanol.

Finally, ^{13}C/^{2}H double labelled samples were produced by growing the adapted MSP1-expressing cells first in BMGlc/95% D_2O medium containing protonated $^{13}C_6$-glucose, and then in BMMe/95% D_2O medium with protonated ^{13}C-methanol. A second sample was prepared by replacing $^{13}CH_3OH$ with $^{13}CD_3OD$ as the carbon source during the induction period. In this case, ^{15}N ammonium sulphate was also used. The yield of protein was ∼6 mg/L for these 0.1 to 0.2 L cultures.

The distribution of residual protonation in these samples was then analyzed by 2D ^{13}C-HSQC experiments. A previously prepared ^{15}N/^{13}C sample of MSP1-19 (Morgan et al., 1999) was used for the control experiment. Regions of the spectra containing the aliphatic or aromatic signals of the partially deuterated protein and the control proteins are shown in Figures 5–7.

The aliphatic region of the spectrum (Figure 5) shows that there is a very effective reduction of the H_α signals (> 85% for samples grown in 95% D_2O/CH_3OH and values near the noise level for samples grown in 95% D_2O/CD_3OD). This is consistent with the 1D ^{1}H data and observations of the amide ^{1}H linewidths described above. Furthermore, a modified HN(CO)CA experiment with a dephasing delay active on the $C_\alpha H$ species (but not $C_\alpha D$ species) showed that 90–95% deuteration was achieved with the sample grown in 95% D_2O/CD_3OD (data not shown). The efficient exchange at the C_α position presumably results from the H_α protons being derived from the solvent during transamination of α-keto-carboxylic acids. There is also a significant effect on many, but not all, H_β and H_γ signals. Residual signals from Ser H_β, Asp/Asn H_β, Pro H_δ, and Lys H_δ/H_ϵ atoms are relatively strong. The strength from the Lys side chain CH peaks is partly due to the clustering of most of the nine Lys residues near the same random coil chemical shifts.

Significant residual signals from different isotopomers of terminal methyl groups (Figure 6) of several residues were observed in samples grown in 95% D_2O/CH_3OH, including Ala H_β, Val H_γ, Leu H_δ, and Ile $H_{\gamma 2}$. These methyl groups are derived from pyruvate in common metabolic pathways (Rosen et al., 1996) (in general, as no specific information is available for *P. pastoris*). The signal intensities indicate that the CHD_2 form is the most populated, CH_2D also has a substantial population, but CH_3 is present only in trace amounts. There appears to be no preferential labelling of the diastereospecific methyl groups of Val and Leu residues. In contrast, the CH_2D peak is slightly more intense than the CHD_2 species for the Thr $H_{\gamma 2}$ and Ile $H_{\delta 1}$ methyl groups. These methyl groups are derived from a different pathway to that of the pyruvate-derived ones. The CD_3 species may also be present in all cases, but is not directly detectable by this experiment. The isotopomer chemical shifts were consistent with reported values of −0.02 ppm (^{1}H) and −0.3 ppm (^{13}C) per deuteron (Rosen et al., 1996). The CHD_2 form has a linewidth

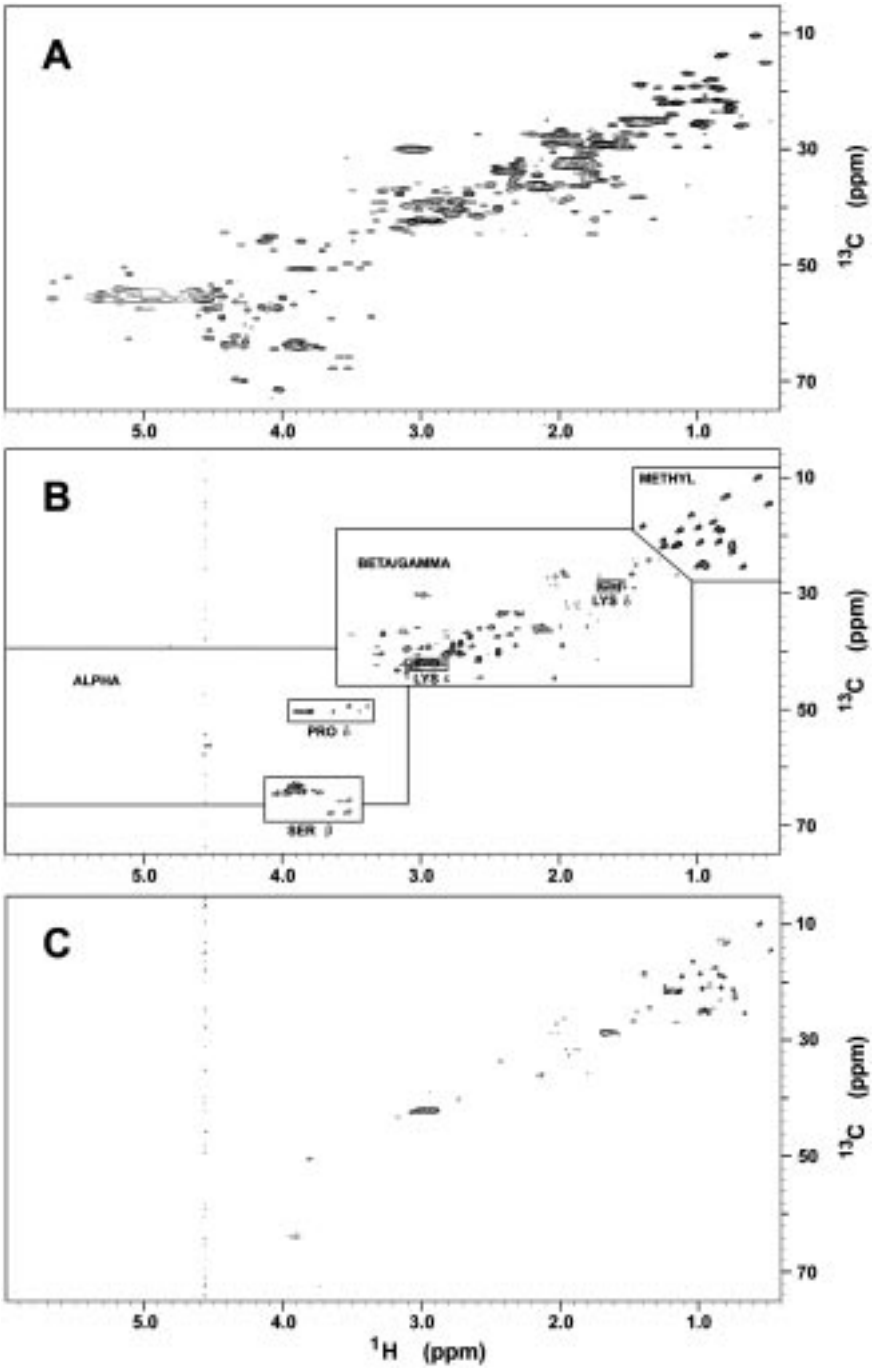

Figure 5. Distribution pattern of residual protiation in deuterated MSP1-19 samples. 2D-[^{1}H/^{13}C]-HSQC spectra of ^{13}C/^{2}H-labelled MSP1-19 protein (aliphatic region). (A) ^{15}N/^{13}C-labelled (non-deuterated control) protein. (B) ^{13}C/^{2}H-labelled protein (grown in 95% D_2O with protiated ^{13}C-methanol). (C) ^{15}N/^{13}C/^{2}H-labelled protein (grown in 95% D_2O with deuterated ^{13}C-methanol ($^{13}CD_3OD$)). Specific amino acid signals that remain relatively intense in deuterated samples (B,C) are indicated in B. The spectra are normalized to equivalent intensity for the methyl group signals.

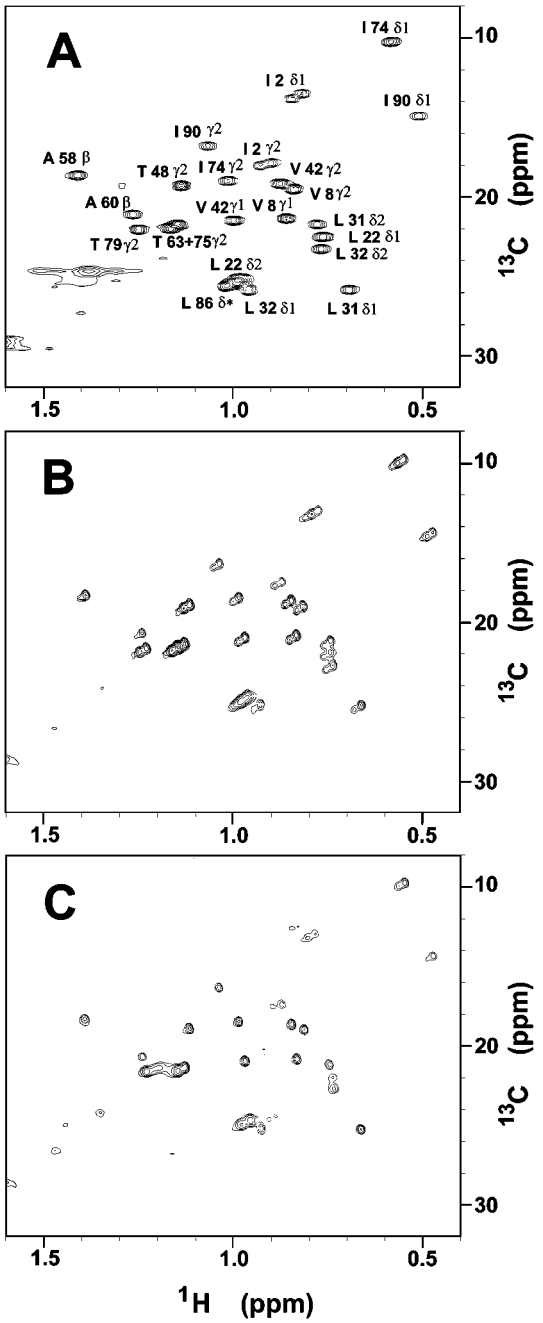

Figure 6. Analysis of methyl group isotopomers in deuterated MSP1-19 samples. This figure shows expansions of the methyl regions of the 2D-[^{1}H/^{13}C]-HSQC spectra presented in Figure 5. (A, B, C) Protein samples as described in the legend of Figure 5. The methyl group signals are shifted upfield in both ^{1}H and ^{13}C dimensions by deuterium isotope effects, depending on the number of deuterons present. Decreased linewidths are observed for the upfield CHD_2 species cross peaks.

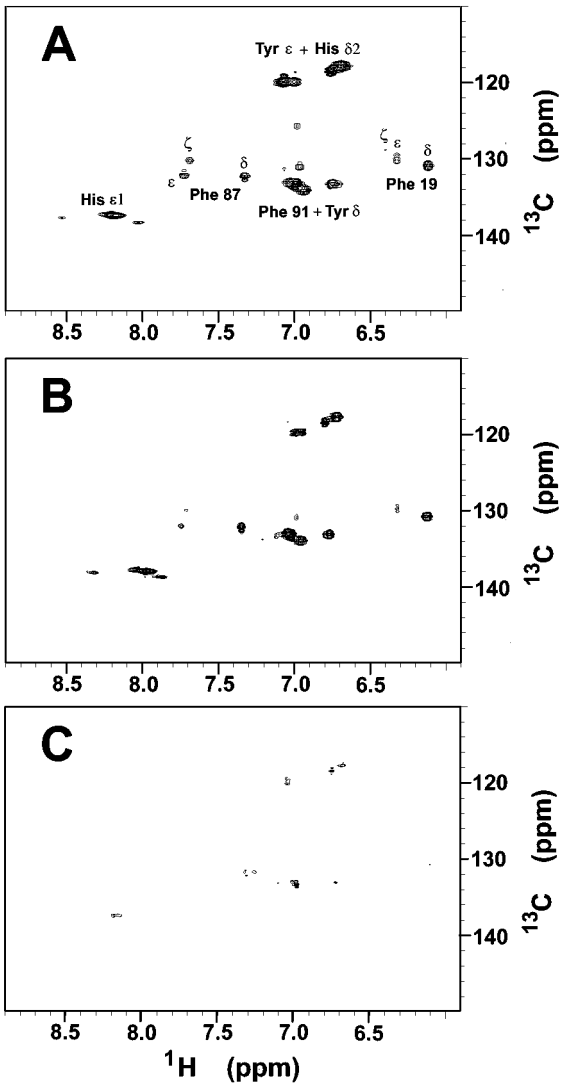

Figure 7. Residual protiation of aromatic amino acid side chains in deuterated MSP1-19 samples. 2D-[^{1}H/^{13}C]-HSQC spectra of ^{13}C/^{2}H-labelled MSP1-19 protein (aromatic region). (A, B, C) Protein samples as described in the legend of Figure 5. Intense signals from the N-terminal His_6 tag are present, as well as from the two His residues in the MSP1-19 sequence. There is also an N-terminal vector-encoded Tyr residue present.

of 10.4 Hz (Leu 31 δ1 peak), compared to 17.6 Hz for the nondeuterated sample. Deuterium decoupling was not used in this experiment, but will provide an additional linewidth reduction for these signals. For the sample grown in 95% D_2O/CD_3OD, the isotopomer distribution for the Thr $H_{\gamma 2}$ and Ile $H_{\delta 1}$ methyl groups is significantly shifted towards the singly protonated CHD_2 species, while for the pyruvate-derived methyl groups the CHD_2 form is clearly predominant, with a

trace of the CH_2D form present. The spectra have been normalized to approximately equal methyl group peak height; in reality the peak heights of the sample grown in 95% D_2O/CD_3OD are ∼30% of those for the 95% D_2O/CH_3OH case. In the aromatic region (Figure 7) for the sample grown in 95% D_2O/CH_3OH, relatively strong residual signals were seen, and could be assigned to Phe H_δ, H_ϵ, H_ζ, Tyr H_ϵ and His $H_{\epsilon 1}$ nuclei (His $H_{\delta 2}$ and Tyr H_δ signals overlap with other resonances). However, all of these aromatic signals were quite weak in the sample grown in 95% D_2O/CD_3OD (Figure 7C). Finally, no information was obtained on Trp and Met residues, since these are not present in the MSP1 C-terminal fragment.

Discussion

Conditions for partial deuteration of recombinant proteins expressed in the methylotrophic yeast *P. pastoris* have been investigated. Adaptation of the cells to growth at high D_2O concentration was found to improve the cell growth rate in glucose media and the yield of heterologous protein, but was not essential for expression. The overall substitution level was 72% of nonexchangeable positions for growth on protonated carbon sources in 95% D_2O media, with an additional 9% increase resulting from inclusion of deuterated methanol as carbon source during the induction period. The fractional deuteration level found here is consistent with values determined in another recent study, by mass spectrometry of amino acids (Massou et al., 1999). Our results suggest that deuterium isotope labelling can be economically achieved in the *Pichia* system in minimal media containing yeast nitrogen base and simple carbon and nitrogen sources, as previously used for $^{15}N/^{13}C$ labelling (Massou et al., 1999; Morgan et al., 1999). The time-consuming adaptation stage can probably be avoided in many cases, but may be important when expensive $^{13}C/^{2}H$-labelled carbon sources are used for producing highly deuterated protein. In general, deuterated carbon sources have less effect on cell growth than the severe effects seen at high concentrations of D_2O (Katz and Crespi, 1966; Haon et al., 1993). Therefore, it is possible that inclusion of deuterated glucose in the starting cultures before induction would be well tolerated and could increase the deuteration levels even further. Use of 99+% D_2O is also possible, but would probably significantly reduce the overall yield.

2D $[^1H/^{13}C]$-HSQC spectra showed that there is very effective and uniform deuteration at the C_α position with growth in 95% D_2O and protonated carbon sources. A fractional deuteration of $> 85\%$ was achieved with growth in 95% D_2O with protiated methanol. The 2D $[^1H/^{13}C]$-HSQC spectra, as well as results from an HN(CO)CA-type experiment used to measure fractional deuteration, demonstrated a level of 90–95% deuteration in samples grown in 95% D_2O media with deuterated methanol. $^{15}N/^{2}H$ samples produced in this way were found to have significantly improved backbone amide proton linewidths. The growth in 95% D_2O only will therefore be a suitable labelling system for preparation of the $^{15}N/^{2}H$-MSP1 C-terminal fragment for study of complexes with antibody Fab or scFv fragments by the conventional chemical shift perturbation methods. Use of deuterated methanol in addition to 95% D_2O had only a minor effect on the amide proton linewidth. This narrowing of the amide NH signals is consistent with the reduction in dipolar relaxation and scalar coupling effects at the NH proton following deuteration at the α-position. C_α relaxation will also be affected by deuteration at this position. The longer T_2 values of deuterated carbons will allow multidimensional heteronuclear experiments to be performed more efficiently with improved signal-to-noise ratios. A novel method for studying protein–protein interactions by NMR requires one component to be perdeuterated and labelled with ^{15}N (Takahashi et al., 2000). In this case, expression in *P. pastoris* using deuterated glucose, deuterated methanol, $(^{15}NH_4)_2SO_4$ and 99+% D_2O will be a very useful alternative to bacterial expression systems for certain proteins, such as the MSP-1 C-terminal fragment EGF module pair. Selective amino acid labelling in a deuterated background, by addition of specific protiated amino acids or substrates to growth media, has been widely used in bacteria (Rosen et al., 1996; Smith et al., 1996;). While this has not yet been reported in *P. pastoris*, this system is likely to be amenable to similar approaches.

The distribution of residual protonation in protein samples from *P. pastoris* grown in 95% D_2O media with either CH_3OH or CD_3OD as carbon source is quite interesting. With CH_3OH, in addition to efficient C_α position deuteration, many aliphatic H_β and H_γ signals were affected. Positions with significant residual protonation included methyl groups (Ala H_β, Thr $H_{\gamma 2}$, Val H_γ, Leu H_δ, Ile $H_{\gamma 2}$, Ile $H_{\delta 1}$) and aromatic side chains (Phe, Tyr, His). When CD_3OD is used as the carbon source, aromatic signals are very weak, and

methyl groups (as well as Lys H_δ/H_ϵ atoms) are the most intense signals.

Methyl groups in the *P. pastoris* sample grown in 95% D_2O/CH_3OH show a distribution of isotopomers, with two principal species, slightly different for the pyruvate-derived groups (Ala H_β, Val H_γ, Leu H_δ, and Ile $H_{\gamma 2}$) and for Thr γ_2 and Ile δ_1 methyl groups. Selective methyl protonation in *E. coli* grown in D_2O with protonated pyruvate also produces a distribution of methyl isotopomers, with CH_3 being the most abundant form (Rosen et al., 1996). However, in this system the Thr γ_2 and Ile δ_1 methyl groups are not protonated. With both yeast and bacteria expression systems, the methyl isotopomer heterogeneity may compromise resolution, although this can be relieved to some extent with purging pulses (Gardner et al., 1997) or multiplet editing (LeMaster, 1994). In this context the dominance of the CHD_2 species in methyl groups in the protein sample grown in 95% D_2O/CD_3OD may be very useful (although the peak intensity of the peaks is much lower than that in the 95% D_2O/CH_3OH sample). In addition, the CHD_2 peaks also have improved linewidths in the ^1H dimension relative to the CH_3 species.

The level of fractional deuteration at the C_α and C_β positions achieved here should be sufficient for optimising carbon relaxation in order to facilitate sequential assignment of larger molecules by triple resonance experiments, as well as to improve chemical shift perturbation experiments using $^1H/^{15}N$ correlation methods. Transverse relaxation-optimised spectroscopy (TROSY) facilitates signal detection in $^1H/^{15}N$ correlation experiments, and is very useful for studies of large molecules (Pervushin et al., 1997). However, deuterium replacement of non-exchangeable protons in protein samples can be applied simultaneously with the TROSY method for further improvement. The residual proton signals, particularly the sharp CHD_2 group peaks, may also be useful for experiments directed towards sequential assignment. As the level of residual protonation is rather low, it may be difficult to employ them in NOE experiments, for example for global fold determination (Rosen et al., 1996; Smith et al., 1996; Gardner et al., 1997). However, this possibility can be further investigated.

The *P. pastoris* system offers a combination of growth with relatively simple and inexpensive isotope labelled compounds, and the highest reported yields for any heterologous expression system (Rosenfeld, 1999). Thus, it is likely to prove very useful in future developments of biomolecular NMR spectrocopy.

Acknowledgements

We are grateful to Dr. K.S. Lilley for performing mass spectrometry, to Dr. T. Frenkiel for performing the modified HN(CO)CA experiment, and to Dr. G. Kelly for comments on the manuscript.

References

Arrowsmith, C.H. and Wu, Y.-S. (1998) *Prog. NMR Spectrosc.*, **32**, 277–286.

Blackman, M.J., Ling, I.R., Nicholls, S.C. and Holder, A.A. (1991) *Mol. Biochem. Parasitol.*, **49**, 29–34.

Gardner, K.H., Rosen, M.K. and Kay, L.E. (1997) *Biochemistry*, **36**, 1389–1401.

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

Haon, S., Auge, S., Tropis, M., Milon, A. and Lindley, N.D. (1993) *J. Label. Comp. Radiopharm.*, **33**, 1053–1063.

Holder, A.A., Guevara-Patiño, J.A., Uthaipibull, C., Syed, S.E.H., Ling, I.T., Scott-Finnigan, T. and Blackman, M.J. (1999) *Parasitologia*, **41**, 409–414.

Katz, J.J. and Crespi, H.L. (1966) *Science*, **151**, 1187–1194.

LaRoche, Y., Storme, V., De Meutter, J., Messens, J. and Lauwereys, M. (1994) *Bio/Technology*, **12**, 1119–1124.

LeMaster, D.M. (1994) *Prog. NMR Spectrosc.*, **26**, 371–419.

Markus, M.A., Dayie, K.T., Matsudaira, P. and Wagner, G. (1994) *J. Magn. Reson.*, **B105**, 192–195.

Massou, S., Puech, V., Talmont, F., Demange, P., Lindley, N.D., Tropis, M. and Milon, A. (1999) *J. Biomol. NMR*, **14**, 231–239.

McAlister, M.S.B., Davis, B., Pfuhl, M. and Driscoll, P.C. (1998) *Protein Eng.*, **11**, 847–853.

Morgan, W.D., Birdsall, B., Frenkiel, T.A., Gradwell, M.G., Burghaus, P.A., Syed, S.E.H., Uthaipibull, C., Holder, A.A. and Feeney, J. (1999) *J. Mol. Biol.*, **289**, 113–122.

Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) *Proc. Natl. Acad Sci. USA*, **94**, 12366–12371.

Piotto, M., Saudek, V. and Sklenár, V. (1992) *J. Biomol. NMR*, **2**, 661–665.

Rosen, M.K., Gardner, K.H., Willis, R.C., Parris, W.E., Pawson, T. and Kay, L.E. (1996) *J. Mol. Biol.*, **263**, 627–636.

Rosenfeld, S.A. (1999) *Methods Enzymol.*, **306**,154–169.

Smith, B.O., Ito, Y., Raine, A., Teichmann, S., Ben-Tovim, L., Nietlispach, D., Broadhurst, R.W., Terada, T., Kelly, M., Oschkinat, H., Shibata, T., Yokoyama, S. and Laue, E.D. (1996) *J. Biomol. NMR*, **8**, 360–368.

Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. and Shimada, I. (2000) *Nat. Struct. Biol.*, **7**, 220–223.

White, C.E., Kempi, N.M. and Komives, E.A. (1994) *Structure*, **2**, 1003–1005.

Wider, G. and Wüthrich, K. (1993) *J. Magn. Reson.*, **B102**, 239–241.

Wood, M.J. and Komives, E.A. (1999) *J. Biomol. NMR*, **13**, 149–159.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 11655–11666.