

# Characterization of the linker peptide of the single-chain $F_v$ fragment of an antibody by NMR spectroscopy

Christian Freund<sup>a</sup>, Alfred Ross<sup>a</sup>, Birgith Guth<sup>b</sup>, Andreas Plückthun<sup>b</sup> and Tad A. Holak<sup>a</sup>

Department of <sup>a</sup>Structural Research and <sup>b</sup>Protein Engineering, Max-Planck-Institut für Biochemie, D-8033 Martinsried, Germany

Received 25 January 1993; revised version received 16 February 1993

A comparison of the single-chain  $F_v$  fragment of the antibody McPC603 (sc $F_v$ ) with its corresponding unlinked  $F_v$  fragment has been carried out with <sup>15</sup>N-edited NMR spectroscopy. The two  $F_v$  fragments adopt the same structure, indicating that the linker does not perturb the folding of the domains. This also directly demonstrates that folding in vivo ( $F_v$  fragment) and in vitro (sc $F_v$  fragment) leads to the same structure. The main differences in the spectra of the uniformly <sup>15</sup>N-labeled sc $F_v$  and  $F_v$  fragments are due to signals of Gly and Ser from the linker peptide of the sc $F_v$  fragment. The linker peptide has been mapped with NMR spectra of <sup>15</sup>N-glycine- and <sup>15</sup>N-glycine/<sup>15</sup>N-serine-labeled sc $F_v$  fragments. The <sup>15</sup>N  $T_2$  relaxation data indicate that the linker peptide is more flexible than the rest of the molecule.

Nuclear magnetic resonance; Antibody; Single chain  $F_v$

## 1. INTRODUCTION

$F_v$  fragments are the smallest functional units of antibody molecules that still contain the complete antigen binding site. They are heterodimers, comprising the variable light ( $V_L$ ) and the variable heavy ( $V_H$ ) chain of the whole antibody. To obtain functional  $F_v$  or  $F_{ab}$  fragments in *E. coli* [1,2], both chains must be co-secreted to the periplasm, where they fold and associate. Alternatively, the two chains of the  $F_v$  fragment can be linked by a peptide linker to give a single-chain  $F_v$  fragment (sc $F_v$ ) that can be functionally secreted to the periplasm [3] or obtained as inclusion bodies in the cytoplasm [4,5]. In the latter case, the protein must be solubilized and refolded in vitro.

In the present paper the  $F_v$  fragment of the antibody McPC603 is compared to the corresponding sc $F_v$  fragment,  $V_H$ -(Gly<sub>4</sub>Ser)<sub>3</sub>- $V_L$ , with the aid of the <sup>15</sup>N-edited 2D NMR spectroscopy. The linker sequence has been mapped by preparing <sup>15</sup>N-glycine- and <sup>15</sup>N-glycine/<sup>15</sup>N-serine-labeled sc $F_v$  fragments. A comparison of the HMQC-spectra of these samples with spectra of the uniformly <sup>15</sup>N-labeled  $F_v$  and sc $F_v$  fragments has allowed the assignment of the residues belonging to the linker sequence of the sc $F_v$  fragment.

## 2. MATERIALS AND METHODS

### 2.1. Expression and purification of the <sup>15</sup>N-labeled $F_v$ fragment of McPC603

The expression vector used was pTACKAN which is a derivative

of the vector pASK30 [6], in which the *lac* promoter is substituted by the *tac* promoter and a kanamycin resistance is introduced. The *E. coli* strain MRE600 [7] harboring this plasmid was grown in minimal 1\*A medium [8], supplemented with <sup>15</sup>N-ammonium chloride at 25°C. Induction with 1 mM IPTG at OD<sub>550</sub> = 1.8 and further growth for 3 h in a 50 l fermenter resulted in 220 g cell paste. After adding 440 ml of BBS (0.2 M sodium borate, 0.16 M NaCl, pH 8.0), the cells were lysed by passing the suspension three times through a Gaulin press at a pressure of 250 bar. The lysate was then centrifuged at 15,000 × *g* for 30 min, and the supernatant was directly applied to a phosphorylcholine affinity column [9]. Bound protein was eluted with a solution of 5 mM PC in BBS. 20 mg of uniformly <sup>15</sup>N-labeled  $F_v$  fragment was obtained.

### 2.2. Expression and in vitro folding of the sc $F_v$ fragments

The expression vector used in this case was pTFT74 (Freund et al., unpublished results), in which the sc $F_v$  fragment is under control of the T7 promoter. The 5'-untranslated region between the conserved 23 bp of the promoter and the start codon of the sc $F_v$  fragment are almost identical to the wild-type *gene 10* sequence of phage T7, with the only difference that a *lac* operator sequence insertion is present in the construct. The terminator sequence of *gene 10* was introduced just behind the coding sequence for the sc $F_v$  fragment. The host strain used for expression was *E. coli* BL21(DE3) [10], which carries the T7-RNA polymerase in the chromosome under the control of the *lacUV5* promoter.

A 10 litre shaking culture (37°C) using the same medium as above was used for the uniformly labeled sample and a 1\*A medium supplemented with all amino acids [11] (only glycine being <sup>15</sup>N labeled) was used in the case of the glycine-labeled sample. All amino acids were supplemented except serine in the glycine/serine-labeled sample (only glycine being <sup>15</sup>N labeled). Cultures were induced at an OD<sub>550</sub> = 0.4 and growth was allowed for another 4 h. The cells were centrifuged at 5,000 × *g* for 10 min, resuspended in 100 ml of 10 mM Tris, 2 mM MgCl<sub>2</sub>, and lysed by a two-fold passage through a French Pressure Cell at 1.1 kbar. The lysate was treated first with DNase (ca. 10,000 u) and RNase (ca. 500 u) for 30 min at 37°C, and then 0.5 M EDTA (20 ml) and Triton X-100 (1 ml) were added, and the solution was stored for 30 min in an ice bath. After centrifugation at 12,000 × *g* for 10 min, the pellet was washed with 0.5 M GdnCl, 0.1 M Tris, 20 mM EDTA, pH 6.8, once and then two more times with 0.1 M Tris, 5 mM EDTA, pH 6.8.

Correspondence address: T.A. Holak or A. Plückthun, Max-Planck-Institut für Biochemie, D-8033 Martinsried, Germany. Fax: (49) (89) 8578 3777.

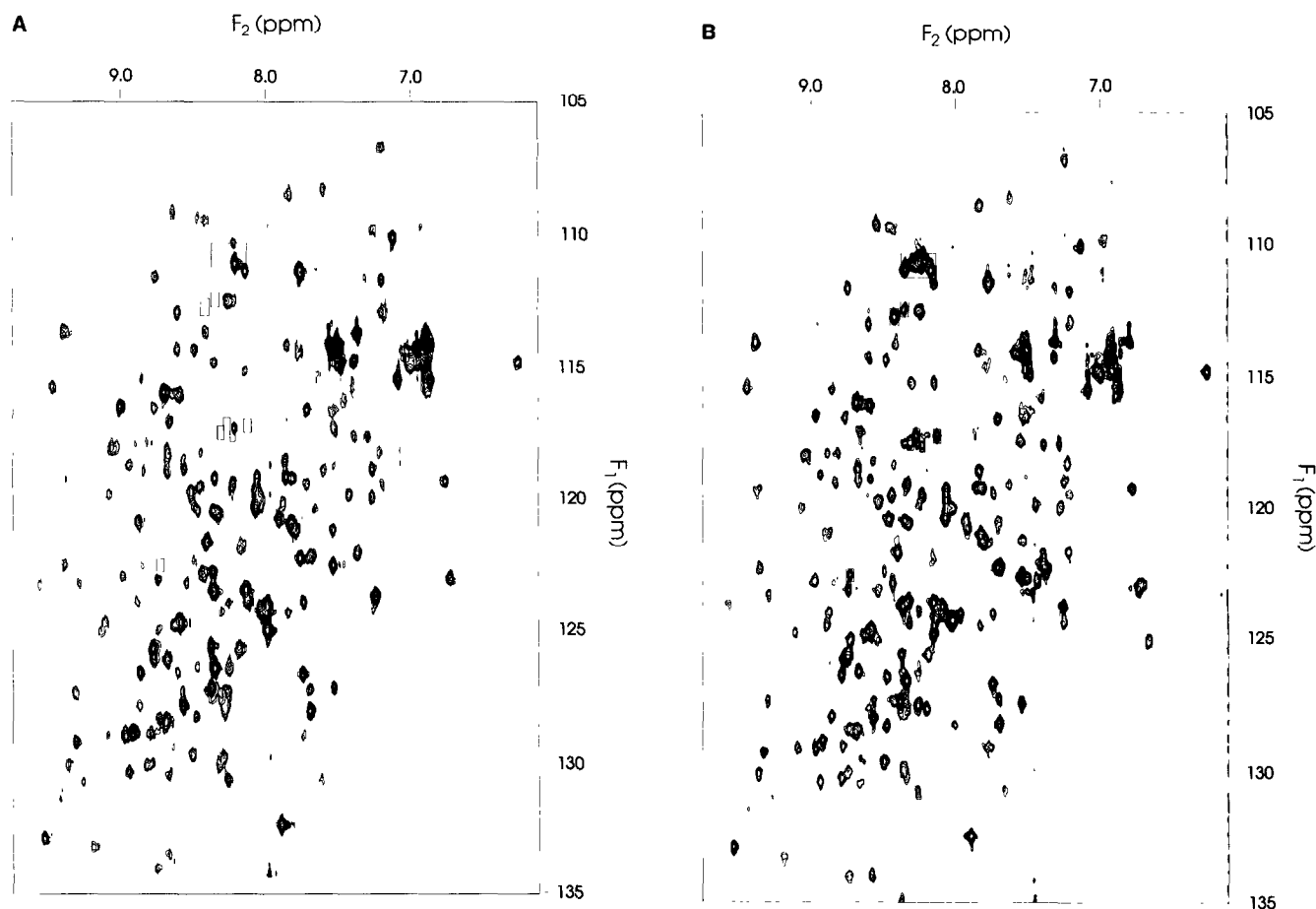


Fig. 1. The HMQC spectra of (A)  $^{15}\text{N}$ -uniformly-labeled  $\text{F}_v$ , and (B)  $^{15}\text{N}$ -uniformly-labeled  $\text{scF}_v$  fragment. In the regions marked by rectangles there are additional peaks in the  $\text{scF}_v$  fragment spectrum compared to the  $\text{F}_v$  fragment spectrum. These peaks are due to the two C-terminal serines in  $\text{V}_H$  and residues of the peptide linker  $(\text{Gly}_4\text{Ser})_3$  in the  $\text{scF}_v$  fragment.

The pellet was then solubilized in 50 ml of 5.5 M GdnCl, 0.2 M Tris, 2 mM EDTA, 0.14 M DTE, pH 9.5 and stirred for 2 h under argon at room temperature. After centrifugation at  $20,000 \times g$  for 30 min, the supernatant was dialyzed three times against 1 l of the same buffer without DTE. Renaturation was performed by diluting the solubilise into 5 l of renaturation buffer (0.8 M arginine, 0.2 M Tris, 2 mM EDTA, 1 mM reduced glutathione, 0.2 mM oxidized glutathione, 0.2 mM of benzamidine hydrochloride, 0.2 mM  $\epsilon$ -amino caproic acid, 0.2 mM phosphorylcholine, pH 9.5), allowing the solution to stand at  $10^\circ\text{C}$  for 100 h. The protein was then concentrated using an Amicon RA2000 cell, resulting in a final volume of 300 ml solution, which was dialyzed against 5 l of BBS (0.2 M sodium borate, 0.4 M NaCl) four times. The functional protein was further purified by affinity chromatography on phosphorylcholine as described above. The yield of a functional labeled protein was typically 1 mg per liter *E. coli* and  $\text{OD}_{550}$ .

### 2.3. NMR spectroscopy

Protein solutions were concentrated by centrifugation in Centricon-10 tubes at  $2,200 \times g$ . NMR samples contained 1 mM protein, 1 mM phosphorylcholine, 1 mM EDTA, and 30 mM potassium phosphate, pH 6.0 in 90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$ . The pulse sequence used for measuring HMQC spectra is described in [12]; the optimized delay for the coherence transfer was 4.8 ms. Presaturation on the water frequency was performed for 1.3 s. During acquisition, decoupling of  $^{15}\text{N}$  frequencies was achieved using the GARP pulse sequence [13] with 1 kHz field strength. The size of the FID matrix was 2K real  $\times$  128 complex points, and TPPI was used to obtain phase-sensitive spectra. In  $\text{F}_1$  the

carrier frequency was set to 119.85 ppm, and in  $\text{F}_2$  the carrier frequency was switched to the middle of the NH region (8 ppm) after presaturation. The  $\text{T}_2$  measurement for the  $^{15}\text{N}$ -glycine-labeled sample was carried out according to the sequence described in [14], which includes a modified CPMG sequence to remove relaxation caused by chemical-shift anisotropy. No incrementation of  $t_1$  was performed; the relaxation delay was incremented by 7.52 ms from one experiment to the next one, thereby achieving twenty 1D spectra. Only the signals that were not overlapped in the 1D spectra were used to calculate  $\text{T}_2$  relaxation times. The other parameters used were the same as for the HMQC spectra. All spectra were recorded at 300 K on a Bruker AMX-600 spectrometer.

## 3. RESULTS AND DISCUSSION

Fig. 1A and B show the HMQC spectra of the uniformly  $^{15}\text{N}$ -labeled  $\text{F}_v$  and  $\text{scF}_v$  fragments, respectively. The spectra of the two samples are remarkably similar and show only minor shifts for a few peaks except in the regions highlighted with rectangles in Fig. 1. The similarity suggests that there are no significant structural changes between the  $\text{scF}_v$  and the unlinked  $\text{F}_v$  fragment. This finding is in agreement with our earlier measurements of antigen binding that showed no significant differences between the  $\text{scF}_v$  and  $\text{F}_v$  fragments in their

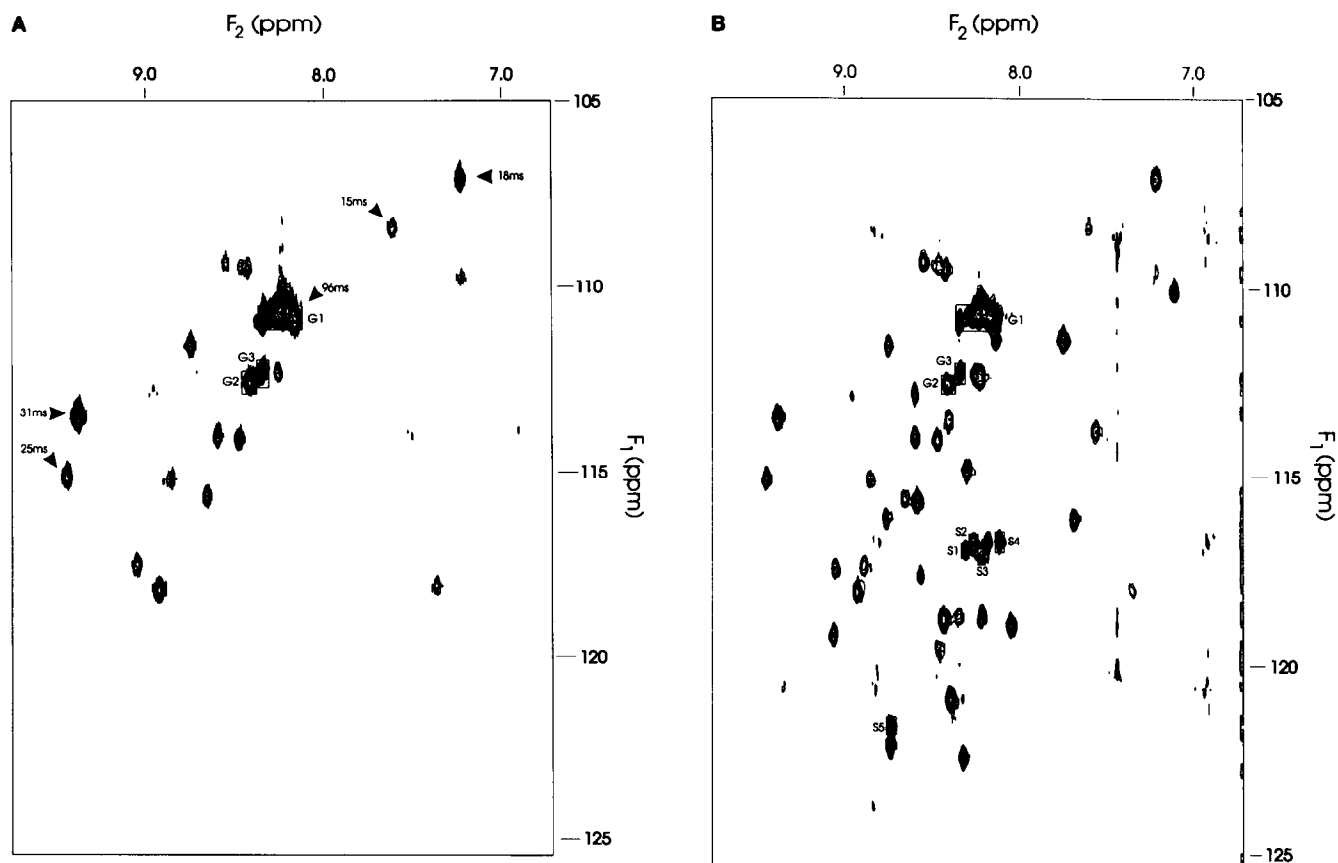


Fig. 2. The HMQC spectra of (A)  $^{15}\text{N}$ -glycine-labeled and (B)  $^{15}\text{N}$ -glycine/serine-labeled  $\text{scF}_v$  fragment. The peaks marked with rectangles are those present only in the  $\text{scF}_v$  and not in the  $\text{F}_v$  fragment (see Fig. 1). In (A) the  $T_2$  relaxation times are indicated for peaks for which  $T_2$  relaxation data could be fitted to the exponential with a correlation coefficient larger than 95%.

binding to antigen [3]. The peaks marked in Fig. 1 are only present in the  $\text{scF}_v$  fragment and should therefore be due to the peptide linker. As expected, these additional peaks are also present in spectra of the selectively labeled  $\text{scF}_v$  (Fig. 2). In order to distinguish glycine from serine residues, the spectrum of  $\text{scF}_v$  with only glycine residues labeled was compared to the spectrum of the glycine/serine-labeled  $\text{scF}_v$  (Fig. 2). There are five peaks in Fig. 2B assigned to serines of the  $\text{scF}_v$  fragment. Only three serines are present in the linker ( $\text{Gly}_4\text{Ser}_3$ ). The two other signals are from the two sequential serines that are at the C-terminus of the  $\text{V}_H$  domain in  $\text{F}_v$ , and are directly preceding the linker in  $\text{scF}_v$ , thus changing their chemical environment upon connecting the linker.

In the case of the glycine residues of the linker peptide, severe overlap prevents their individual assignments, and all but two residues (G2 and G3) appear in the bulk peak G1 (Figs. 1 and 2). In order to determine the number of glycine residues in this peak, we carefully integrated the G1 area and compared the integration with the average value for the two single glycine peaks of the linker (peaks G2 and G3 in Fig. 2A). Since there are two peaks from the  $\text{F}_v$  in the G1 region (Fig. 1A),

which overlap with the bulk glycine peak of the  $\text{scF}_v$ , and assuming that these peaks are also glycines, the anticipated number of 10 glycine residues from  $\text{scF}_v$  in peak G1 was obtained.

The average volume for a single glycine cross peak of the linker peptide is about two times as large as the average volume of a single glycine residue within the domains of the  $\text{scF}_v$  fragment. This is consistent with the observation of strong linker peptide residue signals as compared to the intra-domain residue signals in other  $T_2$  relaxation time-dependent spectra of the  $\text{scF}_v$  fragment ( $^{15}\text{N}$ -edited 3D-TOCSY, HNCA-experiment, data not shown). In order to quantify these observations, the  $^{15}\text{N}$   $T_2$  relaxation times of the linker glycines and four intra-domain glycines were measured. It can be seen from Fig. 2A that the transversal relaxation times for the intra-domain peaks are between 15 and 31 ms and that the  $T_2$  for the linker glycines is three times longer than the longest intra-domain  $T_2$  value (96 ms vs. 31 ms). Although there may be some differences in the dynamical behavior of the glycine residues within the domains, the most dramatic difference is observed for the linker glycines, as compared to the intra-domain glycines. The long  $T_2$  relaxation time of the linker pep-

tide glycines clearly indicates an enhanced flexibility for this part of the molecule. The narrow chemical shift dispersion of the peptide linker residues suggests an almost identical chemical environment for these residues. Also, the chemical shifts of the linker residues are very close to random coil values [15]. This is further evidence that the linker region is unstructured.

In conclusion, the NMR data indicate the linker to be a passive entity with an enhanced flexibility constrained only by the distance between the two domains. The NMR analysis also shows that the same folding state is reached by the F<sub>v</sub> and scF<sub>v</sub> fragments, even though they were obtained by different expression systems. The secretion, which can be carried out for either F<sub>v</sub> or scF<sub>v</sub> fragments does not yield as much protein, but involves fewer steps. The cytoplasmic expression gives a large amount of insoluble scF<sub>v</sub> fragment, but the in vitro folding takes only 20% of the protein to the native state, even after a wide search of suitable conditions. Aggregation and formation of molecules with incorrect disulfide bonds seem to be the major problems during the refolding procedure [16,17], and may also limit the in vivo production of native antibody fragments [18].

*Acknowledgements* This work was supported by a research grant from the Bundesministerium für Forschung und Technologie.

## REFERENCES

- [1] Skerra, A. and Plückthun, A. (1988) *Science* 240, 1038–1041.
- [2] Better, M., Chang, C.P., Robinson, R.R. and Horwitz, A.H. (1988) *Science* 240, 1042–1043.
- [3] Glockshuber, R., Malia, M., Pfitzinger, I. and Plückthun, A. (1990) *Biochemistry* 29, 1362–1367.
- [4] Bird, R.E., Hardman, K.D., Jacobson, J.W., Johnson, S., Kaufman, B.M., Lee, S., Lee, T., Pope, S.H., Riordan, G.S. and Whitlow, M. (1988) *Science* 242, 423–426.
- [5] Huston, S.J., Mudgett-Hunter, M., Tai, M.S., Novotny, J., Margolies, M.N., Ridge, R.J., Brucoleri, R.E., Haber, E., Crea, R. and Oppermann, H. (1988) *Proc. Natl. Acad. Sci. USA* 85, 5879–5883.
- [6] Skerra, A., Pfitzinger, I. and Plückthun, A. (1991) *Biotechnology* 9, 273–278.
- [7] Wade, H.E. and Robinson, H.K. (1966) *Biochem. J.* 101, 467–479.
- [8] Miller, J.H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- [9] Chesebro, B. and Metzger, H. (1972) *Biochemistry* 11, 766–771.
- [10] Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Methods Enzymol.* 185, 60–89.
- [11] Senn, H., Eugster, A., Otting, G., Suter, F. and Wüthrich, K. (1987) *Eur. Biophys. J.* 14, 301–306.
- [12] Bax, A., Griffey, R.H. and Hawkins, B.L. (1983) *J. Magn. Reson.* 55, 301–315.
- [13] Shaka, A.J., Baker, P.B. and Freeman, R. (1985) *J. Magn. Reson.* 64, 547–551.
- [14] Kay, L.E., Nicholson, L.K., Delaglio, F., Bax, A. and Torchia, D.A. (1992) *J. Magn. Reson.* 97, 359–375.
- [15] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.* 222, 311–333.
- [16] Buchner, J. and Rudolph, R. (1991) *Biotechnology* 9, 157–163.
- [17] Buchner, J., Pastan, I. and Brinkmann, U. (1992) *Anal. Biochem.* 205, 263–270.
- [18] Knappik, A., Krebber, C. and Plückthun, A. (1993) *Biotechnology* 11, 77–83.