Journal of Biomolecular NMR, 5 (1995) 48–58
ESCOM

J-Bio NMR 263

Secondary structure determination by NMR spectroscopy of an immunoglobulin-like domain from the giant muscle protein titin

Mark Pfuhl*, Matthias Gautel, Anastasia S. Politou, Catherine Joseph and Annalisa Pastore

European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

Received 15 November 1994
Accepted 11 January 1995

Keywords: Muscles; Titin ig-like domain; Secondary structure

Summary

We present the complete ^{15}N and ^{1}H NMR assignment and the secondary structure of an immunoglobulin-like domain from the giant muscle protein titin. The assignment was obtained using homonuclear and ^{15}N heteronuclear 2D and 3D experiments. The complementarity of 3D TOCSY-NOESY and 3D ^{15}N NOESY-HSQC experiments, using WATERGATE for water suppression, allowed an efficient assignment of otherwise ambiguous cross peaks and was helpful in overcoming poor TOCSY transfer for some amino acids. The secondary structure is derived from specific NOEs between backbone α- and amide protons, secondary chemical shifts of α-protons and chemical exchange for the backbone amide protons. It consists of eight β-strands, forming two β-sheets with four strands each, similar to the classical β-sandwich of the immunoglobulin superfamily, as previously predicted by sequence analysis. Two of the β-strands are connected by type II β-turns; the first β-strand forms a β-bulge. The whole topology is very similar to the only *intra*cellular immunoglobulin-like domain for which a structure has been determined so far, i.e., telokin.

Introduction

Titin is a giant (MW ≈ 3 MDa) modular protein specific for striated muscle of vertebrates (Maruyama et al., 1977; Wang and Ramirez-Mitchell, 1979), recently reviewed by Maruyama (1994). It has a string-like shape (Nave et al., 1989) and extends from the M- to the Z-line of the sarcomere (Fürst et al., 1988). Titin has specific functions, varying along its sequence. I-band titin contributes to the elastic properties of muscle; in the A-band titin seems to be a major organizer of the thick filament via interactions with myosin and other accessory proteins (e.g. C-protein) (Fürst et al., 1992). The region located in the M-line interacts extensively with M-line proteins, forming the M-line meshwork (Vinkemeier et al., 1993). The recent completion of the total titin cDNA sequence determination (S. Labeit, personal communication) has provided the basis for an understanding of titin functions at the molecular level. With a few exceptions, e.g., insertions at the C-terminus (Gautel et al., 1993), the whole titin molecule consists of small modules of ≈100 amino acids in size. These have been suggested to belong to either fibronectin III (type I) or immunoglobulin (type II) superfamilies. Sequence periodicity and the different arrangement of modules in varying regions of the molecule have been suggested to parallel their different functional roles. The most obvious example is the correlation of the periodicity of titin superrepeats in the A-band (II-I-I-II-I-I-I-II-I-I-I) with the C-protein binding sites (Labeit et al., 1992).

A next step in the general understanding of titin can be provided by three-dimensional structure determination of single domains. Based on the observed sequence similarity, it may be sufficient to determine the structures of a few reference modules that are representative of each of the different regions. Extracting key features in terms of folding determinants and individual functions may then lead to the reconstruction of the sequence–structure–function relationships of the whole molecule.

We have recently started the characterization of type

*To whom correspondence should be addressed.

II modules from the A-band and the M-line (Politou et al., 1994a,b). It was demonstrated by CD and NMR spectroscopy that single domains are indeed folded independently and display predominantly β-sheet conformation. Using thermal and urea unfolding experiments, a broad range of stabilities was detected, leading to the hypothesis of a strongly conserved structural framework which would allow a high degree of adaptation towards individual, specific functions. From the domains characterized, we have selected one of the type II modules located in the M-line (referred to here as M5, but named M11 by Politou et al. (1994b)). M5, the sequence of which is highly representative of the whole titin subclass, shows excellent spectral dispersion and high stability. This module has been studied by NMR spectroscopy, with the final aim of obtaining sufficient data for high-resolution structure determination. As a prerequisite, however, the assignment of all resonances and their correlations is necessary to allow the identification of structure elements prior to the calculation of the three-dimensional structure. In this paper we present the complete assignment of the proton and nitrogen spectra of this domain. The secondary structure as determined from NMR parameters is compared to that of telokin, the only other structure known so far of an *intra*cellular immunoglobulin-like domain (Holden et al., 1992).

Materials and Methods

Protein expression and purification

The cDNAs coding for the titin type II domain M5 from the M-line region of the molecule were isolated by means of the polymerase chain reaction (Saiki et al., 1985) using the human AB5 clone as template (M5: AC X69490, bp 11365–11637) (Gautel et al., 1993). The DNA fragments obtained were subcloned into the pET8c vector (Studier et al., 1990) and fused N-terminally with an oligonucleotide linker encoding a His^6 tag sequence, where two additional serines were introduced as linkers. After induction of transformed BL21 [DE3] pLysS cells (Studier and Moffat, 1991) with 0.2 mM IPTG for 4 h, the harvested cell pellet was treated with 5 μg/ml lysozyme on ice and sonicated in 50 mM sodium phosphate, pH 8.0, 150 mM NaCl and 0.1% Triton X-100. After centrifugation at 25 000 g and washing, insoluble inclusion bodies were dissolved in 8 M urea, 50 mM potassium phosphate, pH 8.0, and 20 mM β-mercaptoethanol. The urea extract was clarified by centrifugation at 25 000 g and the soluble supernatant was fractionated by metal chelate affinity chromatography on Ni^{2+}–NTA agarose (Qiagen), essentially as described previously (LeGrice and Grüninger-Leitch, 1990). Pure fractions from imidazole gradients were pooled, diluted at 100 μg/ml and dialyzed against several buffer changes of 20 mM Tris/acetate, pH 8.0, 1 mM EDTA, and 0.5 mM PMSF. After dialysis, insoluble protein was pelleted at 25 000 g and the supernatant was passed over 50 ml of DEAE-Sephacel equilibrated in the dialysis buffer. Unbound fractions contained pure titin domains, as judged by SDS-PAGE (Lämmli, 1970) and mass spectrometry. They were subsequently concentrated by vacuum dialysis against the assay buffer (20 mM deuterated sodium acetate, pH 4.5). To prepare ^{15}N-labeled protein, BL21 [DE3] pLysS cells containing the M5 expression plasmid were grown at 37 °C at 200 rpm in a 2000 ml Eppendorf flask from single colonies for 12 h in 500 ml M9 minimal medium, supplemented with 0.5% glucose, 1 μg/ml Biotin, 1 μg/ml Thiamine, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 0.5 μg/ml $FeCl_3$, 0.05 μg/ml $ZnCl_2$, 0.01 μg/ml $CuCl_2$, 0.01 μg/ml $CoCl_2$, 0.03 μg/ml Na_3BO_3, 2 ng/ml $MoNa_2O_4.2\ H_2O$, 1.6 μg/ml $MnCl_2.6\ H_2O$ and 30 μg/ml ampicilin. For expression, the preculture was diluted fourfold in minimal medium in which the nitrogen source had been completely exchanged against $^{15}NH_4Cl$. Protein expression was induced by addition of 100 μM IPTG to the medium. After further growth at 37 °C and 200 rpm, cells were harvested by centrifugation in a Beckman preparative centrifuge at 4000 rpm.

NMR spectroscopy

All NMR spectra were recorded on 1–2 mM samples in 20 mM deuterated acetate or phosphate buffer at pH values between 4.5 and 5.5. ^{15}N-edited spectra were recorded on a 2 mM ^{15}N-labeled sample at pH 4.5. Spectra were recorded at temperatures between 290 and 310 K. All spectra were measured on a Bruker AMX-600 spectrometer. Water suppression was accomplished using WATERGATE (Piotto et al., 1992) in both homo- and heteronuclear spectra (Sklenář et al., 1993). To record spectra in D_2O, 0.5 ml of a 2 mM sample in 20 mM phosphate was lyophilized and redissolved in the same volume of D_2O. Exchange of amide protons was further analyzed by recording ^{15}N HSQC WATERGATE experiments with and without additional presaturation of the water resonance (Li and Montelione, 1994). The extent of exchange was determined by the ratio of the peak intensities in the two spectra. To avoid additional saturation of the water resonance by the WATERGATE gradients, long relaxation delays were used (5 s) (Stonehouse et al., 1994). The NOE/exchange cross peaks of amide protons with water were taken from the 3D ^{15}N NOESY-HSQC spectrum. TOCSY spectra (Ernst et al., 1987) were recorded using the MLEV17 (Bax and Davis, 1985) and TOWNY (Kadkhodaei et al., 1993) mixing sequences for the 3D and 2D experiments, respectively. To suppress coherence transfer via dipolar coupling, delays of 18 μs were introduced in the MLEV17 sequence (Griesinger et al., 1988). The spectral widths in all experiments were 14.1 and 50 ppm for ^1H and ^{15}N, respectively. Chemical shifts are given relative to DSS for proton and to NH_4Cl for

nitrogen resonances. Spectra were processed on a Silicon Graphics 4D 480 VGX computer using UXNMR/P (BRUKER, Karlsruhe) for the 2D spectra and NMR3D (kindly provided by Dr. C. Cieslar, MPI, Martinsried) for the 3D spectra. The FIDs in the 2D NOESY and TOWNY spectra were weighted with a Gaussian function with a line broadening of 10 Hz and a weighting factor of 0.1 in F2, and with a $\pi/4$ shifted sine bell in F1 prior to Fourier transformation. The FIDs in the 3D spectra were multiplied with Gaussian functions in all dimensions. In the acquisition dimension, the parameters were identical to those in the 2D spectra; in the indirect dimensions the line broadening was 15 Hz and the weighting factor 0.15. The F1 dimension of the 2D spectra was extended by zero-filling, usually from 512 to 2048 data points. In the 3D spectra, both indirect dimensions (F1 and F2) were extended to their final size by linear prediction. All spectra were baseline corrected by a polynomial. The residual water signal in the 3D TOCSY-NOESY spectrum was further reduced by means of the Karhunen–Loeve transformation in NMR3D (Mitschang et al., 1991). Efficient reduction was reached by subtracting the three largest vectors. All spectra were analyzed and plotted using AURELIA (BRUKER, Karlsruhe).

Results

Identification of amino acid spin systems

For a number of residues, the TOCSY transfer from the amide protons was very weak. From the α-protons, however, almost all residues gave a complete transfer to the side chain. A 3D TOCSY-NOESY spectrum proved essential to resolve the α-proton chemical shifts via NOEs between α- and amide protons. F1-F2 planes (TOCSY) were then analyzed at different amide proton chemical shifts in F3. Even when the intraresidual amide to α-proton NOE connectivities were too weak to provide a clear TOCSY pattern, the NOE of the α-proton to the amide proton of the sequentially following residue was usually sufficient to observe an appreciable TOCSY pattern. In this way, the identification of amino acid spin systems was directly combined with the sequential assignment.

Aromatic residues Phe[57] and Trp[33] were assigned directly by the clear J-coupling patterns in 2D TOCSY and COSY spectra. The other phenylalanine and tyrosine residues could be distinguished by the differences in ^{13}C chemical shifts for their δ- and ε-carbons. His[46] was distinguished from the histidines of the His-tag on the ground that it is the only histidine having intra- as well as interresidual NOEs. The aromatic ring systems were connected to their AMX part via NOEs.

Aliphatic residues Valine spin systems were readily identified by the transfer from the α-protons to the side-chain resonances in TOCSY spectra. The connection to their amide protons had to be achieved via 2D NOESY spectra, because of the poor 2D TOCSY transfer from amide to α-protons in the valines having high field-shifted α-resonances (Val[28], Val[62], Val[39]). Leucines and isoleucines had some stronger connectivities of α- and amide protons in 2D TOCSY spectra, but often lacked a complete side-chain pattern. By the combined use of TOCSY spectra recorded at different mixing times and COSY spectra, complete identification and distinction of leucine and isoleucine residues was possible.

Small residues Of the 12 threonines, 10 were immediately identified from TOCSY spectra. Only Thr[23] and Thr[42] gave poor transfer from their amide protons and were identified by analyzing connections to the α-protons. The serines were identified mainly from ^1H α-proton to side-chain J-coupling patterns, with the exception of Ser[65] and Ser[77], which were discovered only in ^{15}N-edited spectra. Glycine residues were identified by the observation of cross peaks between non-degenerate α-protons in COSY and short mixing time TOCSY spectra that were absent in long mixing time TOCSY spectra. Ala[17] and Ala[83] were identified by very strong ^1H amide–β-proton cross peaks in TOCSY spectra, while the amide proton of Ala[44] gave only poor connectivities. The overlap of the α-amide cross peak of Ala[64] with those of Phe[19] and Thr[49] could only be resolved in ^{15}N-resolved spectra.

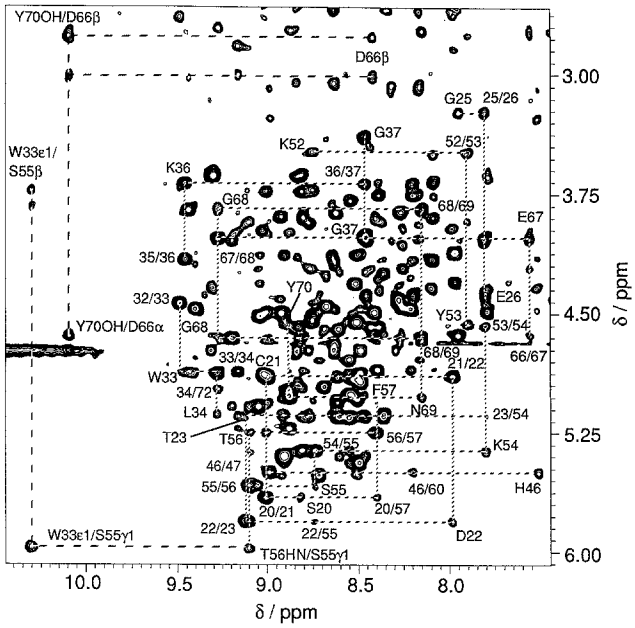

Fig. 1. Fingerprint region of a 2D NOESY spectrum recorded with a mixing time of 100 ms at T = 300 K of a 2.0 mM sample in 20 mM deuterated acetate, pH 5.5. Stretches of sequentially assigned parts are indicated by dotted lines. The intraresidual NOEs are labeled with residue name and number, the sequential α-amide as well as a few interstrand NOEs are labeled with the sequence numbers of the contributing amino acids. NOEs (weak) are labeled in the same way, e.g. 20/57. NOEs involving the two stable hydroxyl protons (on the left side and at the bottom) are indicated by dashed lines and labeled explicitly.

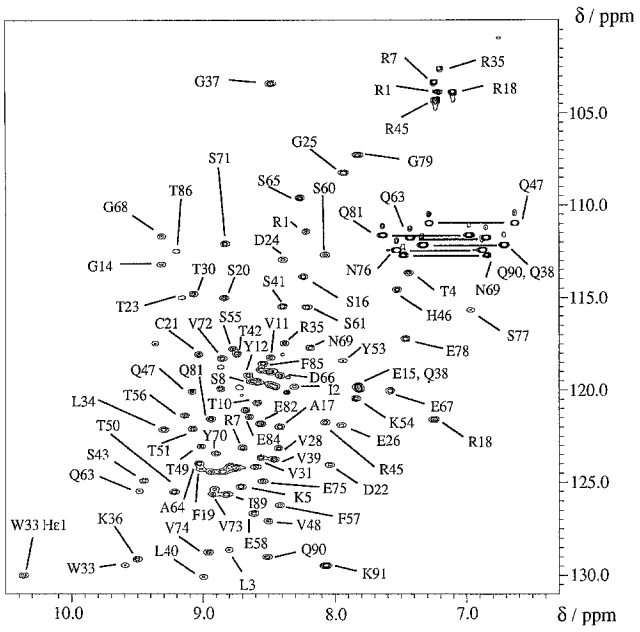

Fig. 2. ^{15}N HSQC experiment at T = 300 K. Experimental conditions were as described in Fig. 1. Arginine N^{ε} resonances are at the top right. Below the arginines are the side-chain amides of asparagines and glutamines.

Charged residues Lysine and arginine residues gave good TOCSY patterns, both from α- and amide protons. Complete side-chain assignments were possible because of the large spread in α-proton frequencies for the lysines. Almost all arginine resonances could be assigned supported by the α-δ correlations present in the TOCSY spectra.

Asn and Gln With the exception of Asn^{76}, all glutamines and asparagines gave good TOCSY transfer from the backbone amides into the side chain. They were distinguished from glutamates and aspartates via NOEs to the side-chain amides. The latter were confirmed in ^{15}N-resolved NOESY spectra.

Identification of exchangeable side-chain protons Two AMX spin systems were identified with extra TOCSY connections at 5.99 and 1.33 ppm, respectively, which are inconsistent with any standard residue. Both extra resonances disappeared in D_2O and were later assigned as the hydroxyl proton of Ser^{55} (5.99 ppm) and the sulfhydryl proton of Cys^{21} (1.31 ppm), consistent with the sequential assignment obtained so far. Another unusual resonance, giving no connections in TOCSY spectra but displaying very strong NOEs to Tyr^{70}, was found at 10.1 ppm. The possibility for this proton to be bound to a nitrogen was excluded by analysis of the ^{15}N HSQC experiment (see Figs. 1 and 2). The resonance was assigned to the hydroxyl proton of Tyr^{70}. All side-chain amide groups of asparagine and glutamine were assigned based on comparison of 2D TOCSY and NOESY spectra recorded in H_2O and D_2O, to distinguish them from aromatic resonances. The results were checked against ^{15}N and ^{13}C HSQC spectra. All arginine ε-resonances were assigned based on a com-

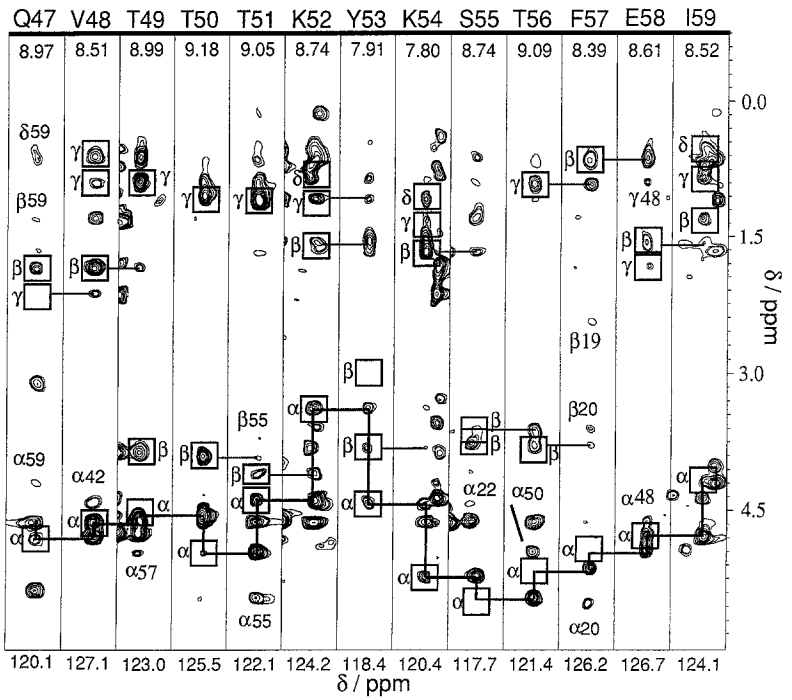

Fig. 3. Strips along F1, taken from a 3D ^{15}N NOESY-HSQC experiment. Experimental conditions were as described in Fig. 1. The F2 (^{15}N) coordinate of each strip is given at the bottom, the F3 (^{1}H) coordinate is given at the top. Intraresidual NOEs are indicated by boxes with the corresponding labels. Sequential backbone–backbone NOEs are indicated by thick lines, sequential side chain–backbone NOEs by thin lines. Non-sequential, medium- and long-range NOEs are labeled explicitly. The figure refers to the antiparallel β-strand formed by strands D (Gln^{47}-Thr^{51}) and E (Lys^{54}-Ile^{59}) and connected by a β-turn (Lys^{52} and Tyr^{53}).

bination of ^{15}N HSQC and 3D ^{15}N NOESY-HSQC spectra. No lysine ζ group was assigned.

Peptide bond isomerism for prolines

The spin systems of the three prolines were identified in TOCSY spectra with 80 ms mixing time, recorded at T = 310 K. Only Pro[6] could be identified as a trans residue from the presence of typical NOESY peaks between its δ-protons and the α-proton of Arg[5] (Wüthrich, 1986). Similar conclusions for the other two prolines could not be drawn, due to partial overlap of their spin systems with the water (Pro[27]) or with the resonances of the preceding residue (Val[28]/Pro[29]). Their conformation could be identified conclusively by a ^{13}C HSQC experiment at natural abundance. The values obtained for both Pro[6] and Pro[29] indicate trans peptide bonds, as their β- and γ-carbons resonate below 30 ppm and above 25 ppm, respectively. In contrast, a β- and a γ-carbon at 32 and 23 ppm, respectively, indicate that Pro[27] has a cis peptide bond (Wüthrich, 1976). In addition, NOEs between the α-protons of Pro[27] and Glu[26] were identified in a spectrum measured in D_2O.

Sequential assignments

Sequential assignment was started using the standard strategy (Wüthrich, 1986), i.e., intraresidue peaks in the fingerprint region were connected via NOEs from the α-proton to the amide proton of the following amino acid. The first stretches of sequence identified in the 2D spectra were the regions Glu[67]-Tyr[70], as also shown in Fig. 6 in Politou et al. (1994b), and Val[48]-Ile[59] (see Fig. 3). Other, shorter stretches, especially from β-strand regions, could also be identified because of the well-resolved low-field resonances of their α-protons. Examples are the stretches Cys[21]-Thr[23] and Glu[82]-Glu[84], as well as loop regions such as Arg[35]-Gly[37] or Gly[25]-Glu[26] (see Fig. 1). Extending these fragments was difficult using only 2D spectra, because of extensive ambiguity in the connection to the α-proton of the previous residue or the amide proton of the following residue. Ambiguities were greatly resolved by using NOESY spectra acquired at different temperatures and 3D TOCSY-NOESY and 3D ^{15}N NOESY-HSQC spectra. The particular advantage of the 3D TOCSY-NOESY is that it enables extraction of sequential information by using the transfer to the side-chain protons for dispersion in F1-F2 in case of weak TOCSY α-amide connections that render a fingerprint peak essentially invisible.

However, this experiment resolved only a few amide resonance ambiguities, because the WATERGATE left only the most high- and low-field shifted α protons in F3

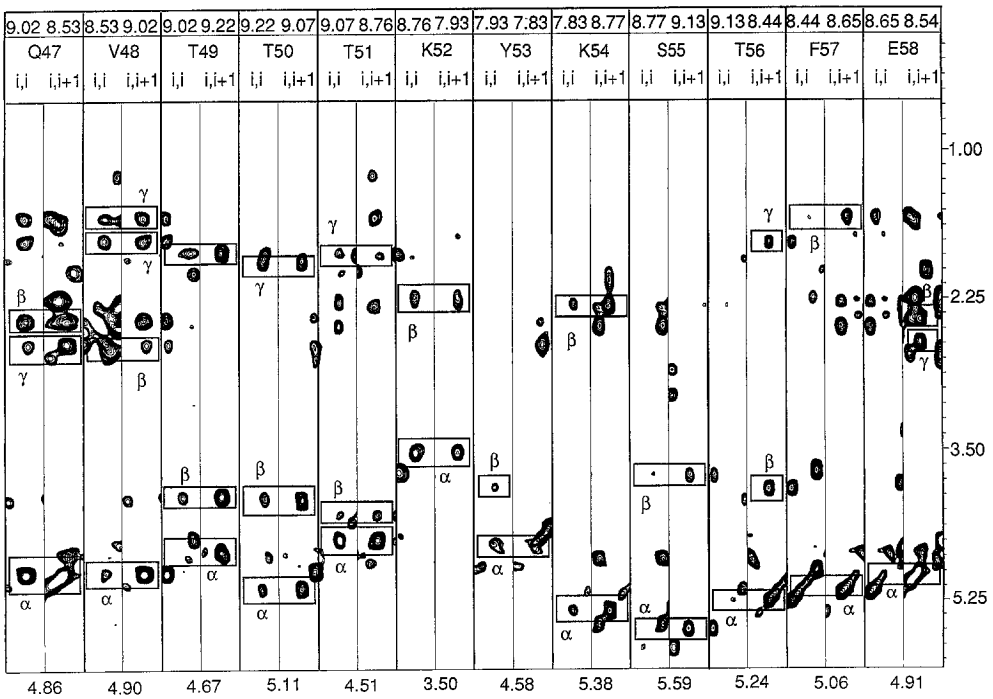

Fig. 4. Selected F1-F2 (TOCSY) strips from the 3D TOCSY-NOESY. Experimental conditions were as described in Fig. 1. Proton resonances ranging from α to side-chain resonances are aligned along F1 (vertical) in each strip making TOCSY connections to the α-proton in F2. Each TOCSY pattern is shown two times: once via a NOE to the intraresidual amide proton in F3 (i,i) and once to the amide proton of the sequentially following residue in F3 (i,i+1). The frequency at which each strip was taken in F2 (α-proton) is shown at the bottom; the frequency at which the strip was taken in F3 (amide proton) is shown at the top. For some residues no TOCSY pattern is visible for the intraresidual connection (e.g. Thr[56] or Glu[58]) because of weak intraresidual NOEs between α- and amide protons typical for β-sheets. Sometimes there are more peaks in a trace than expected due to overlap. Comparison of two neighbouring strips, however, indicates which peaks are part of the spin systems in question. This is emphasized by boxes connecting the 'correct' peaks that belong to one spin system, in particular for Gln[47] and Thr[51].

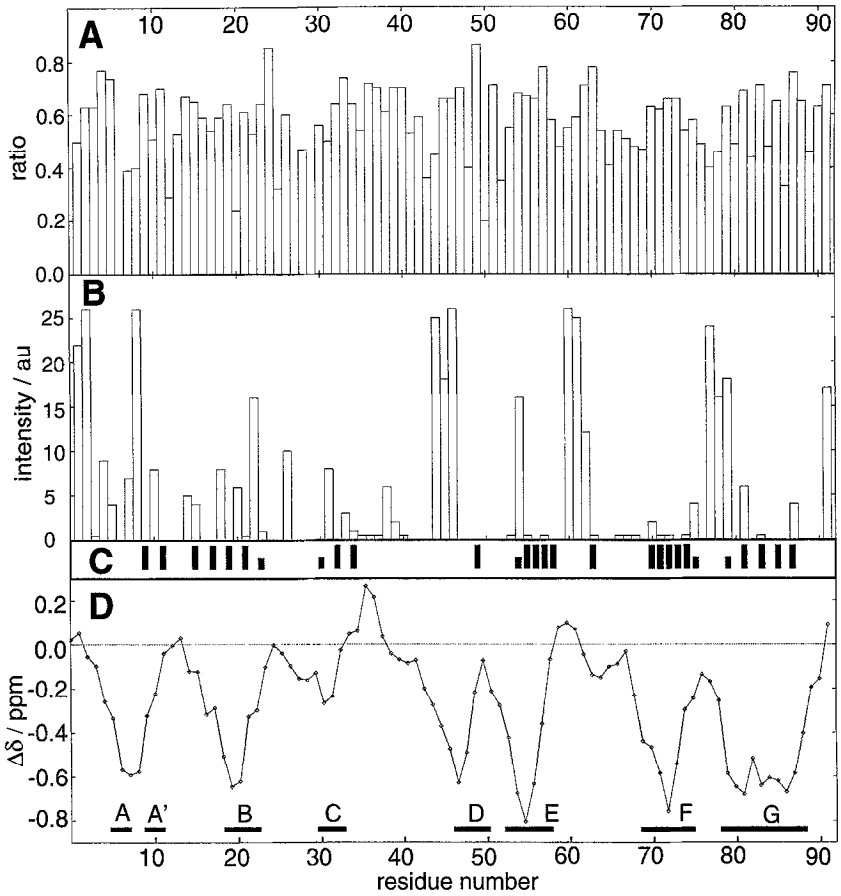

Fig. 5. Exchange of amide protons and secondary chemical shifts. The results from three different experiments are combined. The ratio of peak intensities in ^{15}N HSQC experiments with and without presaturation of the water resonance for 1 s is shown in panel A. Saturation of the water resonance is transferred to the amide protons via chemical exchange, so that smaller peaks indicate faster exchange. Panel B shows intensities of amide proton water NOE/exchange peaks taken from the 3D ^{15}N NOESY-HSQC spectrum. The water is completely undisturbed until immediately before the start of acquisition (F3), such that NOEs from water (in F1) to amide protons (in F2) are observable. Data are shown only for residues that do not display intra- or interresidual NOEs to α-protons close to the water resonance. The black bars in panel C indicate amide protons that are stable for longer than 30 days (large) or 8 h (small) in D_2O solution. Secondary chemical shifts of α-protons at T = 300 K are shown in panel D. Values are obtained by subtracting the measured values from the random coil chemical shifts (Wüthrich, 1986). Averages over two consecutive values are shown to render the pattern more evident. The presence of β-conformation is clearly indicated by stretches of residues having consistently negative values. The position of β-strands according to the characteristic NOE pattern is indicated by black bars at the bottom of panel D.

and these were well resolved anyway. NOEs with ambiguous amide resonances were resolved instead by the ^{15}N dispersion in the 3D ^{15}N NOESY-HSQC. By combining the information provided by the two 3D spectra, the sequential assignment became straightforward.

Examples of F1-F2 (NOESY) slices in the 3D ^{15}N NOESY-HSQC spectrum are shown in Fig. 3 and F1-F2 (TOCSY) slices in the 3D TOCSY-NOESY spectrum are shown in Fig. 4. Intraresidual NOEs in the slices of the 3D ^{15}N NOESY-HSQC spectrum were identified by superimposing with the corresponding 3D ^{15}N TOCSY-HSQC spectrum. NOEs collected from F1-F2 slices in the 3D ^{15}N NOESY-HSQC were assigned by analysis of a F1-F2 plane in the 3D TOCSY-NOESY spectrum, taken at the same value in F3 as the slice in the 3D ^{15}N TOCSY-HSQC.

The search for sequential partners in the 3D spectra was guided by the candidate frequencies obtained from the 2D spectra. In this way, the assignments of all but one strand were completed. Also, some loops could be connected to the strands, e.g., residues Arg^{35}-Gly^{37} were connected to the strand Thr^{30}-Leu^{34} based on NOEs from the 3D ^{15}N TOCSY-HSQC. The loops Asp^{66}-Ser^{60}, Gly^{25}-Pro^{29}, Glu^{13}-Ala^{17}, the N-terminus and the parallel strand Ser^{8}-Tyr^{12} were assigned only on the basis of the 3D spectra. Sequential connectivities from the prolines to the previous residue were obtained from the 3D TOCSY-NOESY spectrum, with the exception of Pro^{29} (see above). The connection to the sequentially following residues was taken from the 3D ^{15}N NOESY-HSQC for all three prolines. The complete sequential assignment is shown in Table 1.

Identification of elements of secondary structure

Elements of secondary structure are usually identified by characteristic chemical shifts of α-protons (Pastore and

TABLE 1
^{1}H AND ^{15}N ASSIGNMENTS (ppm) FOR THE M5 MODULE OF TITIN

Residue	N	H^N	H^α	H^β	Others
Arg[1]	112.8	8.25	4.46	1.60	H^γ 0.64; H^δ 2.76; H^ε 7.22; N^ε 96.2
Ile[2]	119.6	8.24	3.93	1.51	$H^{\gamma 1}$ 1.44; $H^{\gamma 2}$ 0.67; $H^{\delta 1}$ 0.80
Leu[3]	128.6	8.77	4.21	1.47 / 1.34	H^γ 1.16; H^δ 0.75, 0.69
Thr[4]	113.6	7.38	4.46	3.97	H^γ 1.19
Lys[5]	125.1	8.69	4.27	1.94	H^γ 1.70; H^δ 1.09; H^ε 3.07
Pro[6]	–		4.92	0.94	H^γ 1.76 / 1.69; H^δ 3.41
Arg[7]	123.1	8.71	4.59	1.95	H^γ 1.78; H^δ 3.26; H^ε 7.24; N^ε 96.7
Ser[8]	119.6	8.57	5.14	4.25 / 3.87	–
Met[9]	118.2	8.46	5.01	1.96 / 1.91	H^γ 2.42 / 2.32
Thr[10]	120.7	8.54	5.43	3.81	H^γ 0.91
Val[11]	118.5	8.49	4.72	2.41	H^γ 0.79 / 0.58
Tyr[12]	119.2	8.63	4.73	2.59 / 3.06	H^δ 7.22; H^ε 7.22
Glu[13]	119.1	8.84	3.66	2.02 / 1.97	H^γ 2.27 / 2.16
Gly[14]	113.2	9.29	4.27 / 3.64	–	–
Glu[15]	119.8	7.78	4.46	2.04 / 2.14	H^γ 2.25
Ser[16]	113.8	8.23	4.87	3.76 / 3.69	–
Ala[17]	121.9	8.41	4.31	1.05	–
Arg[18]	121.6	7.21	4.49	2.05 / 1.80	H^γ 1.26; H^δ 3.00; H^ε 7.10; N^ε 96.2
Phe[19]	123.9	9.03	4.66	2.45	H^δ 6.83; H^ε 7.03; H^ζ 6.98
Ser[20]	115.0	8.82	5.65	3.72	–
Cys[21]	118.1	9.09	4.91	2.71 / 2.51	H^γ 1.33
Asp[22]	124.0	7.98	5.81	2.87 / 2.57	–
Thr[23]	114.9	9.11	5.16	3.88	H^γ 0.94
Asp[24]	115.5	8.37	4.64	2.56 / 2.32	–
Gly[25]	108.2	7.95	4.05 / 3.24	–	–
Glu[26]	121.9	7.79	4.36	1.97 / 2.04	H^γ 2.16
Pro[27]	–	–	4.66	1.95	H^γ 1.84; H^δ 3.76 / 3.62
Val[28]	123.3	8.43	3.74	1.77	H^γ 0.85, 0.58
Pro[29]	–	–	4.51	2.04	H^γ 1.86; H^δ 3.58
Thr[30]	114.8	9.03	4.31	4.14	H^γ 1.01
Val[31]	123.6	8.51	4.54	1.82	H^γ 0.76 / 0.25
Thr[32]	124.2	8.77	4.45	3.74	H^γ 1.19
Trp[33]	129.4	9.49	4.86	2.68 / 2.57	$H^{\zeta 2}$ 7.06; $H^{\varepsilon 3}$ 7.35; $H^{\delta 1}$ 7.19; $H^{\zeta 3}$ 6.50; $H^{\eta 2}$ 6.38; $H^{\varepsilon 1}$ 10.4
Leu[34]	122.1	9.27	5.13	1.58	H^γ 0.97; H^δ 0.58 / 0.49
Arg[35]	117.4	8.36	4.18	1.58 / 1.46	H^γ 0.96; H^δ 3.13; H^ε 7.20; N^ε 97.4
Lys[36]	129.1	9.45	3.67	1.34	H^γ 1.96; H^δ 1.65; H^ε 2.90
Gly[37]	103.2	8.48	4.03 / 3.41	–	–
Gln[38]	119.8	7.78	4.46	2.30 / 2.23	H^γ 2.43; H^ε 7.14 / 6.82; N^ε 112.2
Val[39]	123.7	8.39	3.91	1.89	H^γ 0.96 / 0.63
Leu[40]	130.1	8.91	4.43	1.71	H^γ 1.49; H^δ 0.61 / 0.29
Ser[41]	115.3	8.25	4.71	3.88	–
Thr[42]	119.8	8.62	4.47	4.20	H^γ 1.40
Ser[43]	124.8	9.43	4.73	4.17	–
Ala[44]	121.1	8.65	4.16	1.42	–
Arg[45]	121.7	8.09	4.36	1.90	H^γ 1.55; H^δ 3.02; H^ε 7.23; N^ε 95.7
His[46]	114.6	7.52	5.49	3.28 / 3.10	H^δ 8.27; H^ε 7.09
Gln[47]	120.1	8.97	4.86	1.98	H^γ 2.26; H^ε 7.25 / 6.61; N^ε 110.9
Val[48]	127.1	8.51	4.89	1.99	H^γ 1.01 / 0.74
Thr[49]	123.0	8.99	4.67	4.00	H^γ 1.14
Thr[50]	125.5	9.18	5.11	4.05	H^γ 1.27
Thr[51]	122.1	9.05	4.51	4.22	H^γ 1.18
Lys[52]	124.2	8.74	3.50	1.78 / 1.67	H^γ 1.18; H^δ 0.98; H^ε 2.89
Tyr[53]	118.4	7.91	4.58	3.93 / 2.80	H^δ 7.24; H^ε 6.96
Lys[54]	120.4	7.80	5.38	1.79	H^γ 1.47; H^δ 1.17; H^ε 2.92
Ser[55]	117.7	8.74	5.59	3.83 / 3.73	H^γ 5.97
Thr[56]	121.4	9.09	5.24	3.93	H^γ 1.05
Phe[57]	126.2	8.39	5.06	0.75	H^δ 6.44; H^ε 6.71; H^ζ 6.57
Glu[58]	126.7	8.61	4.91	1.67 / 1.73	H^γ 1.94
Ile[59]	124.1	8.52	4.40	1.50	$H^{\gamma 1}$ 0.83; $H^{\gamma 2}$ 0.97; $H^{\delta 1}$ 0.61
Ser[60]	112.7	8.19	4.23	3.76 / 3.69	–

TABLE 1
(continued)

Residue	N	H^N	H^α	H^β	Others
Ser^{61}	115.4	8.08	3.91	3.67 / 3.51	–
Val^{62}	124.4	8.77	3.87	2.00	H^γ 0.83 / 0.93
Gln^{63}	125.5	9.46	4.88	1.96	H^γ 2.31 / 2.24; H^ε 7.40 / 6.80; N^ε 111.7
Ala^{64}	123.9	9.02	4.67	1.44	–
Ser^{65}	109.4	8.17	4.79	3.75	–
Asp^{66}	119.2	8.42	4.64	3.03 / 2.82	–
Glu^{67}	120.0	7.56	4.04	2.16	H^γ 2.72 / 2.39
Gly^{68}	111.7	9.29	4.63 / 3.83	–	–
Asn^{69}	117.7	8.14	5.01	2.72 / 2.50	H^δ 7.45 / 6.81; N^δ 112.7
Tyr^{70}	123.4	8.87	4.61	2.73 / 2.45	H^δ 6.75; H^ε 6.65; H^η 10.08
Ser^{71}	112.1	8.86	5.38	3.74 / 3.63	–
Val^{72}	118.3	8.79	4.97	1.17	H^γ 0.46 / –0.42
Val^{73}	125.6	8.90	4.57	1.70	H^γ 0.72
Val^{74}	128.7	8.92	5.04	1.72	H^γ 0.75
Glu^{75}	124.9	8.51	5.15	2.02	H^γ 2.23
Asn^{76}	120.1	8.41	4.56	3.22 / 2.71	H^δ 7.49 / 6.84; N^δ 112.4
Ser^{77}	115.6	6.97	4.06	3.78 / 3.73	–
Glu^{78}	117.2	7.42	4.42	2.01	H^γ 2.29
Gly^{79}	107.2	7.98	4.04 / 4.51	–	–
Lys^{80}	118.9	8.45	5.39	1.66	H^γ 1.18; H^δ 1.42; H^ε 2.91
Gln^{81}	121.5	8.92	4.61	1.89 / 1.75	H^γ 2.41; H^ε 7.61 / 6.94; N^ε 111.6
Glu^{82}	121.8	8.50	5.50	1.88 / 1.80	H^γ 2.06
Ala^{83}	124.0	8.72	4.79	1.39	–
Glu^{84}	121.4	8.55	5.14	2.05 / 1.94	H^γ 2.26
Phe^{85}	118.8	8.55	4.86	3.05 / 2.86	H^δ 6.85; H^ε 6.54
Thr^{86}	112.5	9.15	5.22	4.01	H^γ 1.14
Leu^{87}	125.3	8.93	5.42	2.25	H^γ 1.42; H^δ 0.91 / 0.65
Thr^{88}	124.4	8.93	4.86	4.18	H^γ 1.23
Ile^{89}	125.8	8.83	4.96	1.75	$H^{\gamma 1}$ 1.60; $H^{\gamma 2}$ 0.73; $H^{\delta 1}$ 0.62
Gln^{90}	128.9	8.49	4.14	1.84	H^γ 1.41; H^ε 7.28 / 6.67; N^ε 112.2
Lys^{91}	121.8	7.99	4.32	1.68	H^γ 1.84; H^δ 1.41; H^ε 3.02

^{15}N resonance frequencies are given for main-chain as well as for side-chain nitrogens, with the exception of lysine ζ nitrogens. Experimental conditions: T = 300 K, 20 mM deuterated acetate, pH 4.8.

Saudek, 1990; Wishart et al., 1991), protection of amide protons from chemical exchange and, most important, by the detection of specific patterns of short-, medium- and long-range NOEs (Wüthrich, 1986). The secondary chemical shifts of the α-protons are shown in Fig. 5. It is possible to distinguish seven regions with consistently negative secondary chemical shifts (typical of β-sheet structures) of at least –0.15 ppm, ranging between residues 3–12, 16–24, 29–33, 44–50, 50–59, 68–77 and 78–89. The maximum values reached in all these minima are around –0.6 ppm, with the exception of region 29–33, which reaches only values of slightly more than –0.2.

The protection of amide protons was measured by exchange in D_2O, solvent saturation transfer and by semi-quantitative analysis of exchange/NOE intensities in the 3D ^{15}N NOESY-HSQC experiment. These experiments are mostly complementary. While D_2O exchange and the NOE/exchange experiments give more appreciable variations than the saturation transfer experiment, they may present limitations for residues with α-protons that resonate too close to the water signal. Currently, this problem can be circumvented only by using doubly $^{15}N/^{13}C$-labeled samples (Grzesiek and Bax, 1993). The saturation transfer experiment does not suffer from this problem. In addition, only two 2D experiments are needed that profit from the usually good dispersion in the ^{15}N HSQC experiment, so that values for essentially all residues were obtained. The results of all three experiments performed on M5 are in perfect agreement (Fig. 5), as is most evident for residues 70–81. The existence of β-strands is immediately obvious by the alternating pattern of protected and unprotected amide protons. Residues Met^9 and Val^{11} are protected, while residues Ser^8, Thr^{10} and Tyr^{12} are unprotected, suggesting a β-strand facing its partner strand with the amides of Met^9 and Val^{11}. Similar results are observed for residues 15–21, 31–35 and 80–86. All of these regions are external strands. Two regions with continuous stretches of protected amides between 55–58 and 70–75 are more indicative of internal strands, buried in the core of the β-sheet.

The NOEs indicative of β-sheets are displayed in Fig. 6. By analysis of 2D (see Fig. 1) and 3D spectra (see Fig.

3) a complete set of NOEs was collected, defining a total of eight β-strands. With the aid of interstrand NOEs, these can be grouped into two β-sheets, each containing four strands. All strands are antiparallel, with the exception of the pair G-A'. A closer view on the general topology is obtained by analyzing the list of NOEs, including also side-chain protons, as shown in Fig. 7. The regions of parallel and antiparallel sheets are clearly evident. Three other distinct elements of secondary structure were identified besides the β-strands, i.e., two type I β-turns connecting strands E and D as well as strands G and F according to specific short-range NOEs (Wüthrich et al., 1984) and a β-bulge (Richardson et al., 1978) near the N-terminus, characterised by a strong sequential amide–amide NOE between Leu^3 and Thr^4 and the 'inversion' of the pattern of interstrand NOEs between Leu^3 and Thr^4.

The definition of the β-strands by the NOEs accounts almost perfectly for the secondary chemical shifts, as well as for the protected amide protons. The only exception is the β-strand Thr^{30}-Arg^{35} that displays much weaker secondary chemical shifts than all other strands. However, Trp^{33} may compensate for the low-field shift expected for the α-protons in a β-sheet. Among the protected amide protons are some that are not part of a β-strand, as defined by the NOEs. Glu^{15} and Ala^{17} follow the alternating pattern of protected amides that continues with Phe^{19} and Cys^{21} and the weaker protected Thr^{23}, without exhibiting the typical NOEs. The protection should be caused by the amides pointing into the core of the molecule, thus being almost unaccessible from the outside. This assumption is confirmed by the fact that Ile^{59}, being the last residue on the opposite strand, still shows the characteristic NOEs but has its amide proton pointing to the outside of the molecule. The same holds for the following residues (Ser^{60}-Val^{62}) in this strand, which exhibit consistently fast exchange of their amide protons in all three experiments. All amide protons that are presumably involved in hydrogen bonds between strands A and B exchange relatively fast in D_2O experiments (see Fig. 5). This is not observed for the other external strands (C, D and A'). However, we expect a higher flexibility of the N-terminus, so that the A strand may be formed for most of the time but with a high rate of 'opening and closing'. This hypothesis is supported by stability data, showing that the M5 domain is stabilized when N-terminally extended by six residues (Politou et al., 1994a) (the melting temperature increases

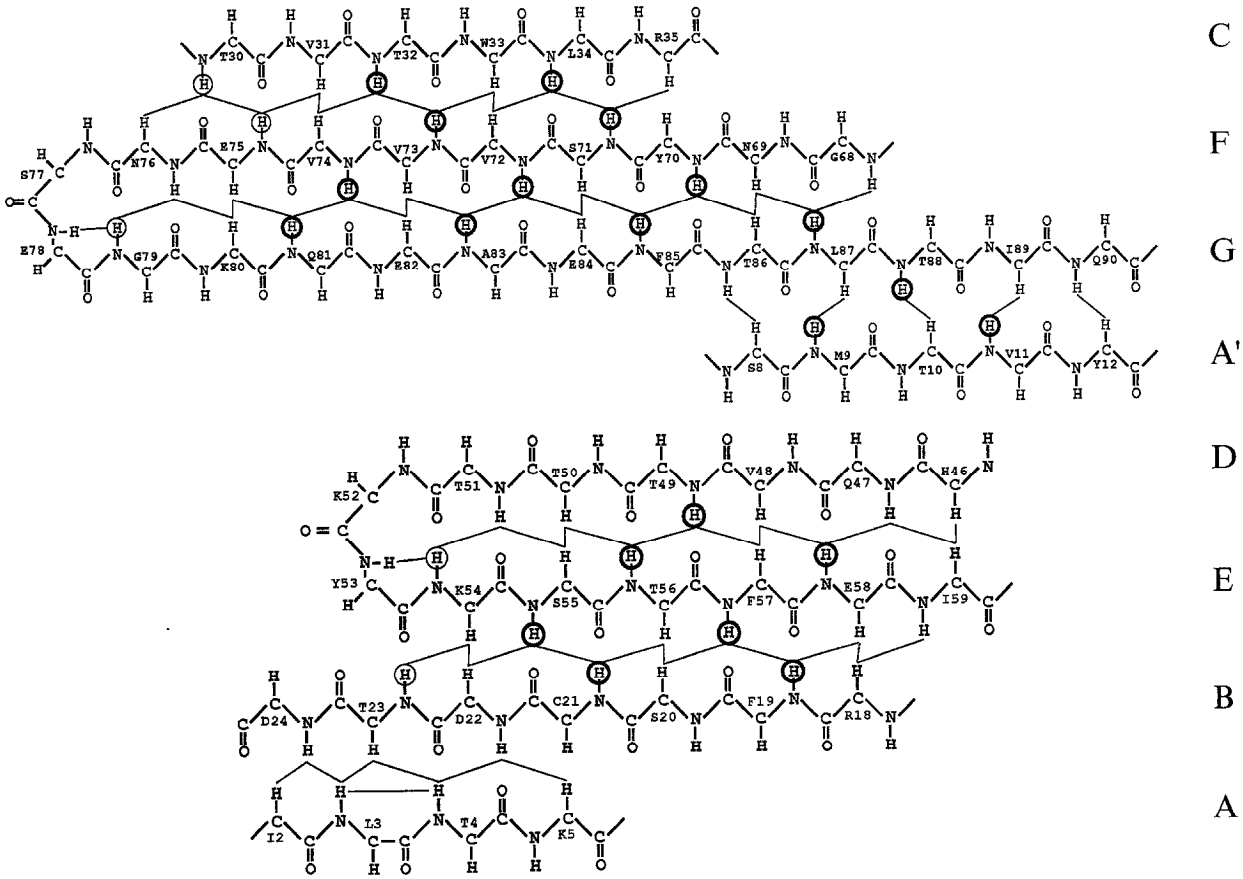

Fig. 6. Definition of the β-strands according to interstrand backbone–backbone NOEs. NOEs that are indicative of secondary structure are labeled by thin lines. Amide protons stable for more than 30 days in D_2O are labeled with thick circles; amide protons stable for more than 8 h are labeled with thin circles. The rest of the amide protons exchange within less than 8 h. A β-bulge is clearly defined for residues 2 to 4. Two type II β-turns were identified, connecting strands E and D as well as F and G. Nomenclature conventions were taken from Harpaz and Chothia (1994).

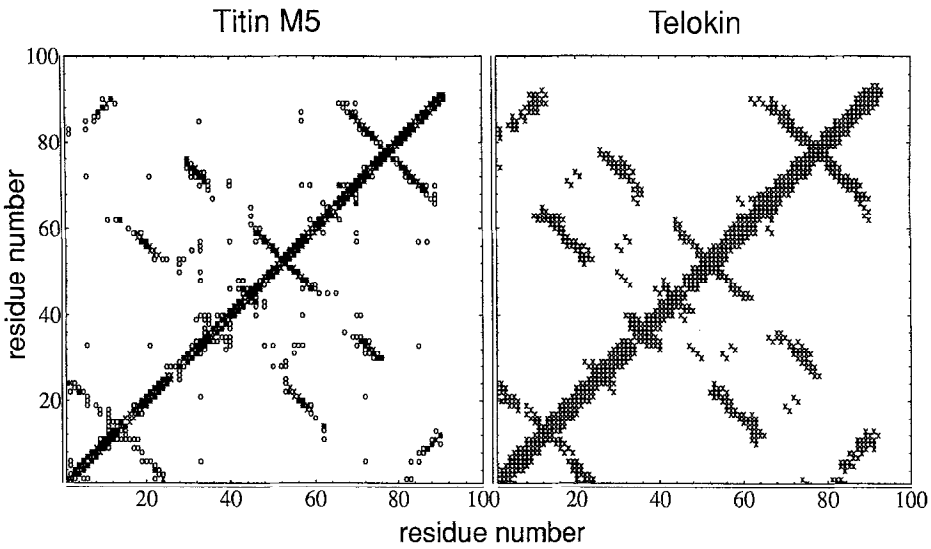

Fig. 7. Left: NOEs extracted from a NOESY spectrum with a mixing time of 100 ms, measured at T = 300 K. Crosses indicate distances < 3.0 Å, circles indicate distances > 3.0 Å. Right: distances calculated from the structure of telokin (Brookhaven entry 1tlk) using XPLOR by averaging over all atoms of every residue with a cutoff of 9 Å. Only the part that is homologous to the M5 domain is shown. Matrix plots were generated with MATHEMATICA.

from 50.5 to 59.3 °C; the free energy of folding goes from 3.61 to 5.47 kcal mol^{-1}). The improved stability against thermally and urea-induced unfolding is not restricted to the M5 module, but appears to be a general property of titin type II modules. The additional residues could stabilize the interface of the A and B strands by making interactions to the G strand as well as to the loop connecting strands B and C. More direct support for additional flexibility of the A strand comes from variations in the three experiments, performed to analyze amide proton exchange. Although vanished within less than 8 h in D_2O, the amides of Leu^3 and Thr^4 show rather high intensities in the solvent transfer experiment and weak or no NOE/exchange peaks to water in the 3D ^{15}N NOESY-HSQC experiment. A similar behaviour is observed for the residues in strand B: Asp^{22} and Asp^{24} exchange fast in D_2O, but have a higher degree of protection in the other two experiments.

Conclusions

Essentially complete sequential assignment was obtained for the immunoglobulin-like module M5 from titin by the combined use of homo- and ^{15}N heteronuclear 2D and 3D spectra. Analysis of amide proton chemical exchange data, using a combination of three independent methods together with secondary chemical shifts and medium- and long-range NOEs, established a well-defined secondary structure. The protein contains β-strands, connected by turns and bulges. The whole structure is divided into two β-sheets, forming the β-sandwich that is well established for the structure of immunoglobulins. Each β-sheet contains four strands. The sheet comprising strands A, B, E and D is arranged in an antiparallel fashion (see Fig. 6). The other sheet, containing strands A', G, F and C, is antiparallel with the exception of the C-terminal part of strand G, which is arranged in a fashion parallel to strand A'. The strands F and G, as well as D and E are connected by type II β-turns. The central part of strand A' forms a β-bulge.

From the overall fold as defined above, it can be concluded that the M5 module is a member of the immunoglobulin superfamily, as previously predicted on the basis of sequence alignments. This family has been the subject of numerous investigations concerning sequence–structure relationships. The ever increasing amount of sequences and structures from extra- and more recently also intracellular proteins has resulted in the classification of subfamilies (Bork et al., 1994; Harpaz and Chothia, 1994). The only other structure solved so far for an intracellular immunoglobulin-like protein from eukaryotes, telokin (Holden et al., 1992), was assumed to be an intermediate between the well-established V and C subfamilies. For this reason, a new subfamily called the I-set was proposed (Harpaz and Chothia, 1994). Although the sequence identity of the M5 module to telokin is rather low, i.e. 25%, the positions of the β-strands determined from NMR parameters in M5 match very well. This is evident when comparing the list of NOEs for M5 with a matrix of distances from the structure of telokin in Fig. 7. Both patterns match very well when the different lengths of the N-termini of the two proteins are taken into account. It should be noted in addition that most of the residues assumed to be determinants for the folding of the core are well conserved in the sequences among the titin type II modules and telokin. The most appreciable difference

is the lack of any data confirming the existence of strand C' in M5. The sequence alignment, however, shows an extremely low homology in this region between M5 and telokin, but also among all the type II domains of titin. It seems therefore unlikely that the C' strand is a key feature of the I-set. To derive more conclusive arguments, the structure determinations of M5 and other domains from different regions of the molecule have to be awaited.

Acknowledgements

The authors wish to thank Toby Gibson and Siegfried Labeit for stimulating discussions on titin, Vladimir Sklenář for help with WATERGATE, Lorenz Mitschang for help with gradients, Thorsten Dieckmann for help with 3D processing, Andrew Atkinson for valuable suggestions and David Grindrod for computer support. This work was supported by a grant from the Human Frontiers Science Organisation.

References

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **65**, 533–535.

Bork, P., Holm, L. and Sander, C. (1994) *J. Mol. Biol.*, **242**, 309–320.

Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, International Series of Monographs on Chemistry, Vol. 14, Oxford Science Publications, Oxford.

Fürst, D.O., Osborn, M. and Weber, K. (1988) *J. Cell Biol.*, **106**, 1563–1572.

Fürst, D.O., Vinkemeier, U. and Weber, K. (1992) *J. Cell Sci.*, **102**, 769–778.

Gautel, M., Leonhard, K. and Labeit, S. (1993) *EMBO J.*, **10**, 3827–3834.

Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 627–638.

Harpaz, Y. and Chothia, C. (1994) *J. Mol. Biol.*, **238**, 528–539.

Holden, H.M., Ito, M., Hartshorne, D.J. and Rayment, I. (1992) *J. Mol. Biol.*, **227**, 840–851.

Kadkhodaei, M., Hwang, T.-L., Tang, J. and Shaka, A.J. (1993) *J. Magn. Reson. Ser. A*, **105**, 104–107.

Labeit, S., Gautel, M., Lakey, A. and Trinick, J. (1992) *EMBO J.*, **11**, 1711–1716.

Lämmli, U.K. (1970) *Nature*, **227**, 680–685.

LeGrice, S.F.J. and Grüninger-Leitch, F. (1990) *Eur. J. Biochem.*, **87**, 307–314.

Li, Y.-C. and Montelione, G.T. (1994) *J. Magn. Reson.*, **105**, 45–51.

Maruyama, K., Matsubara, S., Natori, R., Nonomura, Y., Kimura, S., Ohashi, K., Murakami, F., Handa, S. and Eguchi, G. (1977) *J. Biochem.*, **82**, 317–337.

Maruyama, K. (1994) *Biophys. Chem.*, **50**, 73–85.

Mitschang, L., Cieslar, C., Holak, T.A. and Oschkinat, H. (1991) *J. Magn. Reson.*, **92**, 208–217.

Nave, R., Fürst, D.O. and Weber, K. (1989) *J. Cell Biol.*, **109**, 2177–2187.

Pastore, A. and Saudek, V. (1990) *J. Magn. Reson.*, **90**, 165–176.

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

Politou, A.S., Gautel, M., Joseph, C. and Pastore, A. (1994a) *FEBS Lett.*, **352**, 27–31.

Politou, A.S., Gautel, M., Pfuhl, M., Labeit, S. and Pastore, A. (1994b) *Biochemistry*, **33**, 4730–4737.

Richardson, J.S., Getzoff, E.D. and Richardson, D.C. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 2574–2578.

Saiki, R.K., Scharf, S.J., Faloona, F., Mullis, G.T. and Ehrlich, H.A. (1985) *Science*, **230**, 1350–1354.

Sklenář, V., Piotto, M., Leppik, R. and Saudek, V. (1993) *J. Magn. Reson.*, **102**, 241–245.

Stonehouse, J., Shaw, G.L., Keeler, J. and Laue, E.D. (1994) *J. Magn. Reson.*, **107**, 178–184.

Studier, F.W., Rosenberg, A.H. and Dubendorf, J.W. (1990) *Methods Enzymol.*, **185**, 62–89.

Studier, F.W. and Moffat, B.A. (1991) *J. Mol. Biol.*, **189**, 113–130.

Vinkemeier, U., Obermann, W., Weber, K. and Fürst, D.O. (1993) *J. Cell Sci.*, **106**, 319–330.

Wang, K. and Ramirez-Mitchell, R. (1979) *J. Cell Biol.*, **83**, 389a.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.

Wüthrich, K. (1976) *NMR of Amino Acids and Nucleotides*, North Holland, Amsterdam.

Wüthrich, K., Billeter, M. and Braun, W. (1984) *J. Mol. Biol.*, **180**, 715–740.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.