Immunoglobulin-Type Domains of Titin: Same Fold, Different Stability?†

A. S. Politou,*,‡ M. Gautel, M. Pfuhl, S. Labeit, and A. Pastore

EMBL, Meyerhofstrasse 1, W69117 Heidelberg, Germany

Received December 2, 1993; Revised Manuscript Received January 27, 1994®

ABSTRACT: Titin is a 3-MDa protein thought to form a fibrous intracellular system in vertebrate striated muscle and to play an important role in sarcomere alignment during muscle contraction. It has also been implicated as a "molecular ruler", regulating the assembly and the precise length of the thick filaments [Whiting, A. J., Wardale, J., & Trinick, J. (1989) *J. Mol. Biol. 205*, 163–169]. Partial sequencing of titin-encoding cDNAs suggests that the protein is organized in a modular fashion, containing two classes of ~100-residue repeats [Labeit, S., Barlow, D. P., Gautel, M., Gibson, T., Holt, J., Hsieh, C. L., Francke, U., Leonard, K., Wardale, J., Whiting, A., & Trinick, J. (1990) *Nature 345*, 273–276]. These motifs, referred to as type I and type II modules, show sequence homology to the fibronectin III and immunoglobulin C2 superfamilies, respectively. Since the type II modules represent the most widely occurring motifs along the titin molecule, we expressed in *Escherichia coli* three domains of this type spanning different regions of the sarcomere (A-band and M-line) and studied their structure and stability. Using circular dichroism, nuclear magnetic resonance, and fluorescence spectroscopy, we showed that all the fragments examined are independently folded in solution and possess a β-sheet conformation. Furthermore, employing NMR analysis, we identified an overall folding pattern present in all modules and related to the Ig fold, as previously suggested by theoretical predictions. The stability of the modules over a range of conditions was investigated by measuring key thermodynamic parameters for both thermal and chemical denaturation and by monitoring amide proton exchange as a function of time. Despite the overall structural similarity, the stability of the modules seemed to differ; the motif corresponding to the M-line band was significantly more stable than the motifs corresponding to the A-band. Our data provide direct experimental evidence that the titin type II modules possess a β-sheet conformation and further suggest that similarly folded motifs located at different regions of the titin molecule may have distinct molecular properties and stability. As structural analysis of more titin domains is proceeding, these and related observations are expected to establish clear cut structure–function relationships and to unveil the exact cellular role of this protein.

Titin is the largest protein described to date (~3 MDa) and one of the few proteins specific to vertebrate striated muscle (Maruyama et al., 1984; Wang, 1985; Kurzban & Wang, 1988; Fürst et al., 1988). Related giant proteins are found in invertebrates (Benian et al., 1989; Ayme-Southgate et al., 1991). In the mature myofibril, titin is the third most abundant component of the sarcomere, after actin and myosin, comprising about 10% of its mass. Single titin molecules are "string-like" particles, over 1 μm in length, which span half the sarcomere, i.e., from M- to Z-line (Fürst et al., 1988; Nave et al., 1989). The part of the titin molecule located in the I-band appears to make elastic connections with the thick filaments and the Z-disk (Fürst et al., 1988; Horowits et al., 1989; Funatsu et al., 1990, 1993); the A-band region of titin contains a multiplicity of binding sites for myosin and C-protein (Labeit et al., 1992; Fürst et al., 1992); finally, in intimate association with other proteins, titin forms an integral part of the M-line (Vinkemeyer et al., 1993; Gautel et al., 1993). Because of these multiple interactions, titin is thought to play an important role in providing sarcomere alignment during muscle contraction and in regulating the assembly and the precise length of the thick filaments during myofibrillogenesis (Trinick et al., 1984; Whiting et al., 1989; Fulton & Isaacs, 1991; Isaacs et al., 1992; Gautel et al., 1993; Wang et al., 1993).

A detailed structural characterization of the titin molecule had until recently been precluded by its large size. However, the cloning and sequencing of cDNAs coding for an appreciable part of titin showed that this protein is organized in a typically modular fashion and that it contains two classes of repeated ~100-residue motifs (Labeit et al., 1990, 1992; Gautel et al., 1993). These modules, referred to as type I and type II, show sequence homology to the fibronectin III and immunoglobulin C2 superfamilies, respectively (Benian et al., 1989). The arrangement of the repeated sequences along the titin molecule shows considerable variation. In the elastic I-band region of the sarcomere there is an irregular alternation of type I and type II motifs including long stretches of exclusively class II motifs (S. Labeit, unpublished results). In the A-band region the two types of motifs form a very regular 11 domain super-repeat pattern (-II-I-I-II-I-I-I-II-I-I-I-) (Figure 1A). Finally, within the M-line, i.e., C-terminally, class II motifs are separated by nonrepetitive "linker" sequences (Figure 1B).

The modularity implied from the sequence data makes titin an ideal object for NMR[1] studies as the size of the motifs is well within reach of multidimensional NMR spectroscopy. Therefore, a structural study of the individual titin modules

† Supported by grants from the Deutsche Forschungsgemeinschaft (Ga 405/2-1, La 668/2-1), the European Community and the Human Frontiers Science Project.

* To whom correspondence should be addressed.

‡ Recipient of a BRIDGE Fellowship from EC.

® Abstract published in *Advance ACS Abstracts*, March 15, 1994.

[1] Abbreviations: NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; CD, circular dichroism; TOCSY, total correlation spectroscopy; TPPI, time-proportional phase incrementation; NOESY, nuclear Overhauser enhancement spectroscopy; DQF-COSY, double quantum filter correlated spectroscopy; HMQC, heteronuclear multiple quantum coherence; HSQC, heteronuclear single quantum coherence; ΔG, Gibbs free energy change; ΔS, entropy change; ΔH, enthalpy change; T_m, melting temperature; MLCK, myosin light chain kinase; NOE, nuclear Overhauser enhancement; UV, ultraviolet.

0006-2960/94/0433-4730$04.50/0 © 1994 American Chemical Society

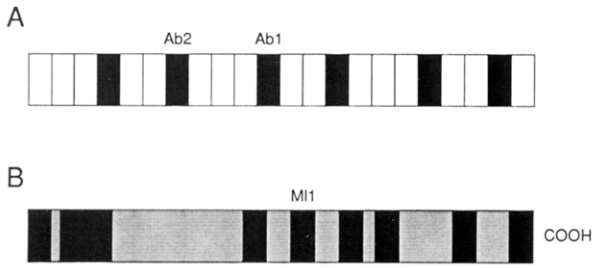


FIGURE 1: Arrangement of modules in the A-band (A) and in the M-line (B) part of the titin molecule. Black boxes indicate the type II and white ones the type I modules. Shaded boxes represent nonrepetitive "linker" sequences separating the type II modules in the M-line. Also indicated are the positions of the modules expressed and studied in the present work.

is both meaningful and feasible because it could eventually address important aspects of the function of titin, such as its elasticity, at the molecular level. Such an approach has been successfully applied to several modular proteins and is by now well established (Baron et al., 1990; Driscoll et al., 1991; Main et al., 1992).

We have undertaken a longterm structural analysis of representative titin modules in solution using a variety of techniques and mainly NMR spectroscopy. The general scope of this type of analysis and the initial data obtained with selected motifs made it obvious that a detailed study of the stability and conformational properties of individual domains is an absolute prerequisite before any rigorous 3D structure determination is attempted. At the present time we have concentrated on the structural analysis of type II modules, mainly for two reasons. First, these modules are present in all regions of the molecule in various combinations with modules of the same class or of class I or with nonrepetitive "linker" sequences. Second, they belong to the Ig superfamily, a diverse structural class which now includes intra- as well as extracellular proteins. The structures of these proteins have been extensively studied, allowing comparison among a large family of evolutionarily related proteins; however, the number of stability studies is rather limited (Buchner et al., 1991).

In this study, we provide experimental evidence showing that individual titin modules possess a predominantly β-sheet secondary structure. We also report the results of a stability survey for selected titin modules of class II from the A-band and the M-line which represents, to the best of our knowledge, the first of this type for single, isolated Ig modules.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification. cDNAs coding for various titin type II domains from the A-band and M-line regions of the molecule were isolated by PCR (Saiki et al., 1985) using the rabbit CE12 and human CH11 (Labeit et al., 1992) and human AB5 (Gautel et al., 1993) clones as templates (Ab2: EMBL data library AC X64698, bp 847–1137; Ml1: AC X69490, bp 11365–11637; Ab1: AC X64696, bp 14404–14685). 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, introducing an additional two serine residues in the linker. After induction of transformed BL21 [DE3] pLysE cells (Studier & Moffat, 1991) with 0.3 mM IPTG for 4 h, the harvested cell pellet was treated with lysozyme at 20 μg/mL and sonicated in 50 mM sodium phosphate, pH 8.0, 150 mM NaCl, and 0.1% Triton X-100. After centrifugation at 25000g and washing, the 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 25000g and the soluble supernatant fractionated by metal chelate affinity chromatography on Ni^{2+}–NTA agarose (Qiagen) essentially as described (LeGrice & Grueninger-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. Insoluble protein after dialysis was pelleted at 25000g and the supernatant passed over 50 mL of DEAE-Sephacel equilibrated in the dialysis buffer. Unbound fractions contained pure titin domains as judged by SDS–PAGE (Laemmli, 1970) and mass spectrometry and were concentrated by vacuum dialysis against the appropriate assay buffers.

Labeling. Uniformly ^{15}N-labeled samples were obtained from bacteria grown in M9 medium with $^{15}NH_4Cl$ as the sole nitrogen source. For the ^{13}C labeling, a mixture of ^{13}C-labeled saccharides and peptides from an acid hydrolysate was used.

Circular Dichroism–Thermal Unfolding. Circular dichroism (CD) spectra in the far ultraviolet were recorded on a Jasco J-710 spectropolarimeter, fitted with a thermostated cell holder and interfaced with a Neslab RTE-110 water bath. The instrument was calibrated with a 0.10% aqueous solution of d_{10}-camphorsulfonic acid. Quartz, thermostated cuvettes with 1- or 0.2-mm path lengths were used (Hellma). Spectra were typically recorded in 10 mM acetate buffer, pH 4.2, and were baseline corrected by subtraction of the appropriate buffer spectra. The combined absorbance of cell, sample, and solvent was kept less than 1 over the measured range. Thermal denaturation curves were obtained at a heating rate of 20 °C/h. Thermal denaturation curves of at least two different preparations were recorded for each module.

Fluorescence Spectroscopy–Urea Denaturation Studies. Urea denaturation of the titin domains was monitored by measuring the intrinsic fluorescence intensity of solutions containing typically 0.070 mg/mL of protein in 10 mM acetate buffer, pH 4.2, and urea in the concentration range of 0–8 M. At least two different samples of each module were used. Urea stock solutions (10 M) were prepared with "ultrapure" urea purchased from Schwarz/Mann Biotech and were used within 24 h. Fluorescence measurements were made in 1.0-cm quartz cuvettes thermostatted at 25 ± 0.1 °C either with an SLM-Aminco Bowman Series 2 or with an SLM 8000 spectrofluorimeter operating in ratio mode. The slit widths were 4 nm for both excitation and emission. For an excitation wavelength of 293 nm, the maximal change in fluorescence intensity between the folded and the unfolded form was obtained at emission wavelengths of 312 nm for Ab1, 314 nm for Ab2, and 315 nm for Ml1. The solutions were incubated at 25 °C for 10–12 h before the measurement. After the fluorescence measurements, the pH of four solutions near the midpoint of the transition was recorded on a PHM93 Radiometer pH meter after a double buffer adjustment, and the average was considered as the pH of denaturation. The resulting urea denaturation curves were analyzed, and the free energy of folding was determined by the linear extrapolation method (Pace et al., 1989) assuming a two-state mechanism of unfolding and by nonlinear regression analysis.

Nuclear Magnetic Resonance. The samples for NMR measurements typically contained 1–1.5 mM protein in 90% H_2O/10% D_2O and 10 mM deuterated acetate buffer, pH 4.2. Samples for the hydrogen exchange experiments were obtained by first lyophilizing the protein from its aqueous solutions and then redissolving it in 99.5% D_2O. All 2D NMR spectra were acquired on Bruker AMX-500 and AMX-600 spectrometers in the phase-sensitive mode (TPPI) either with preirradiation of the water resonance or with selective

```
          Ab1  A P D I D L D L E L   R K I I N I R A G G   S L R L F V . P I K   G R P T P E V K W G
          Ab2  . P T I V L D P T I   K D G L T I K A G D   T I V L N A I S I L   G K P L P K S S W S
          M11  . . . . R I L T K P   R S . M T V Y E G E   S A R F S C . D T D   G E P V P T V T W L
      Telokin  H V K P Y R T K T I   L D . M D V V E G S   A A R F D C . K V E   G Y P D P E V M W F

          Ab1  K V D G E I R D . .   . . A A I I D S T S   . S F T S L V L D N   V N R Y D S G K I T
          Ab2  K A G K D I R P S D   I T Q I T S T P T .   . . S S M L T I K Y   A T R K D A G E Y T
          M11  R K G Q V L S T S A   . . R H Q V T T T K   . Y K S T R E I S S   V Q A S D E G N Y S
      Telokin  K D D N P V K E S R   . . H F Q I D Y D E   E G N C S L T I S E   V C G D D D A K Y T

          Ab1  L T L E N S S G T K   S A F V T V R V L D   T
          Ab2  I T A T N P F G T K   V E H V K V T V L D   V
          M11  V V V E N S E G K Q   E A E F T L T I Q K
      Telokin  C K A V N S L G E A   T C T A E L L V E T   M
```

FIGURE 2: Alignment of all three titin modules against telokin.

excitation (WATERGATE pulse sequence; Piotto et al., 1992). Clean-TOCSY spectra (Griesinger et al., 1988) were measured using the MLEV-17 composite pulse cycle (Bax & Davis, 1985) and an optimized "cleaning" delay approximately 1.5 times longer than the low-power 90° pulse. Mixing times used were in the range 30–75 ms for the TOCSY and 50–200 ms for the NOESY spectra. Homonuclear 2D DQF-COSY, clean-TOCSY, and NOESY and heteronuclear ^1H–^{15}N HSQC, ^1H–^{15}N HSQC-TOCSY, and ^1H–^{15}N HSQC-NOESY as well as ^1H–^{13}C HMQC, ^1H–^{13}C HMQC-NOESY, and ^1H–^{13}C HMQC-TOCSY spectra (Bodenhausen & Ruben, 1980; Bax et al., 1990; Norwood et al., 1990) were recorded at 17, 27, and 35 °C, with 2048 data points in the acquisition domain and 512 data points in t_1. Data were processed on a Bruker X-32 data station using UXNMR software. Prior to Fourier transformation, the data were zero filled to 2048 points in the t_1 dimension and weighted with a Gaussian window in t_2 and a cosine window in t_1. A baseline correction was performed in both dimensions using a polynomial.

RESULTS

Selection of Module Boundaries. The precise boundaries of the modules were selected on the basis of an extensive sequence alignment of all type I and type II modules present in the A-band and M-line (70% of the whole titin sequence) (Higgins et al., 1994). The N-terminus of the A-band domains was selected around the well conserved proline. With the M-line domain two slightly different alignments of the N-terminus are possible, because of the lack of the starting proline common to most of the others. We chose the one that aligns the first hydrophobic residue (Ile) to the first Phe of telokin. The choice of the C-terminus was in all cases unambiguous, as it had to include the well conserved Hydrophobic–X–Hydrophobic (Figure 2).

Secondary Structure Deduced by Circular Dichroism. CD spectra of all three titin domains at room temperature (25 °C) are characteristic of a predominantly β-sheet secondary structure (Figure 3). Use of the method described by Chen et al. (1974) to estimate the secondary structure gave identical results for the three modules (52% in β-sheet and 48% in secondary structures other than β-sheet and helix). However, the error associated with the fitting process is large, so that the resulting percentages are only qualitatively significant.

CD spectra recorded in the pH range of 4–7 show the same features, indicating that there is no pH-dependent conformational change. A pH of 4.2 which is more favorable for NMR experiments was chosen for our further work, so that similar conditions could be used in all our studies.

Urea Denaturation. Fluorescence spectroscopy is ideally suited for monitoring of the unfolding, because in all three

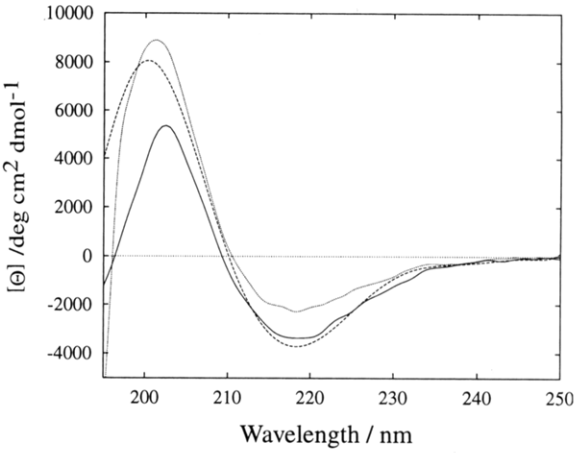


FIGURE 3: CD spectra of Ab1 (dotted line), Ab2 (solid line), and M11 (dashed line) at 25 °C, pH 4.2, in 10 mM acetate buffer.

cases the emission spectra of the denatured material differ significantly from the ones of the native protein both in intensity and in the maximum emission wavelength (21-, 28-, and 30-nm shift upon unfolding of M11, Ab1, and Ab2, respectively). The latter is highly indicative of a tryptophan well buried in the hydrophobic core of the protein (318 nm for Ab1 and 323 nm for Ab2 and M11 in their native forms).

The denaturation curves obtained by monitoring the intrinsic fluorescence of the modules at 312 nm for Ab1, 314 nm for Ab2, and 315 nm for the M11 domain are all characterized by the same sigmoidal shape (see Figure 4A for a representative curve). They can be divided into three regions: (a) the pretransition region that shows the effect of increasing urea concentration on the fluorescence intensity for the folded protein; (b) the transition region, which shows how the same property varies upon progression of unfolding; (c) the posttransition region, which shows the dependence of the fluorescence intensity on the denaturant concentration for the unfolded protein. An appreciable increase in the fluorescence intensity relative to the native state was observed at very low urea concentrations (<0.25 M) only in the case of the Ab1 module; a similar effect has been previously reported with other proteins and mainly in cases of guanidinium-induced unfolding (Pace et al., 1990). These points were not included in our analysis of the curve. There is also a slight linear dependence of fluorescence on urea concentration for urea concentrations higher than 7.5 M, in agreement with previous observations (Schmid et al., 1989; Pace et al., 1992).

A two-state folding mechanism was postulated to analyze the curves. Such an assumption is supported by the single-step shape of the unfolding curve (Figure 4A) and by its subsequent analysis (Figure 4B). On that basis ΔG can be

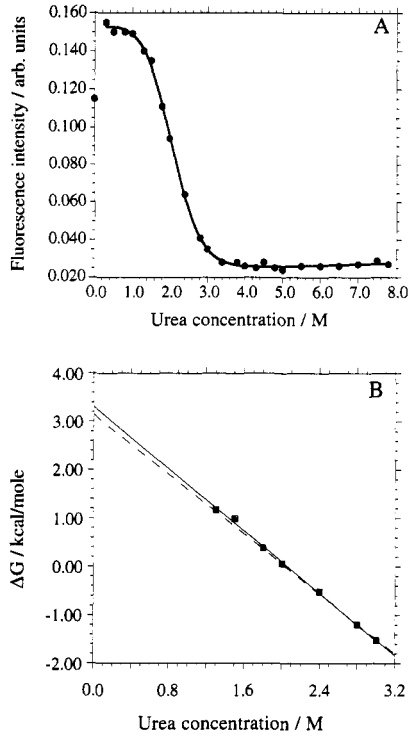


FIGURE 4: (A) Urea unfolding curve for the Ab1 module at pH 4.2, in 10 mM acetate buffer, 25 °C. The intrinsic fluorescence intensity was measured at 312 nm after excitation at 293 nm. The curve of best fit obtained from the nonlinear regression analysis is also shown. (B) ΔG as a function of the urea concentration for the transition region. ΔG was calculated from the data in panel A using eq 1 (points) and linearly extrapolated to zero according to eq 2 (solid line).

calculated as a function of urea concentration from the points in the transition region using

$$\Delta G = -RT \ln K = -RT \ln (f_u/f_f) = -RT \ln [(y_f - y)/(y - y_u)] \quad (1)$$

where R is the gas constant [1.987 calories/(deg·mol)], T is the absolute temperature, K is the equilibrium constant, f_f and f_u represent the fraction of protein present in the folded and unfolded conformation, respectively, y is the observed fluorescence intensity at the selected wavelength, and y_f and y_u are the values of the fluorescence intensities characteristic of the folded and unfolded conformation, respectively (Pace et al., 1989). Values of y_f and y_u were obtained by extrapolation of the pre- and posttransition baselines. A least-squares analysis was used to determine the equations for f_u and f_f in the transition region. In all three cases ΔG was found to vary linearly with urea concentration. Assuming that this linear dependence continues to zero concentration, the data were then fit to

$$\Delta G = \Delta G(H_2O) - m[\text{urea}] \quad (2)$$

where $\Delta G(H_2O)$ is the value of ΔG at 25 °C in the absence of the denaturant, known as conformational stability, and m is a measure of the steepness of the unfolding curves.

In all three cases, a very good fit was found between values of ΔG derived from the experimentally measured fluorescence intensity and those obtained by extrapolation in the transition region (Figure 4B). This lent additional support to our initial assumption of an one-step mechanism of unfolding.

It has been argued before (Santoro & Bolen, 1989) that this method of analysis underestimates the final error in the parameters determined (ΔG and m) because no error is assumed for the pre- and posttransition baselines. The use of a nonlinear regression analysis was suggested to fit the entire unfolding curve, such as that shown in Figure 4A, to

$$y = \{(y_f + m_f[U]) + (y_u + m_u[U])(\exp[-(\Delta G(H_2O))/RT - m[U]/RT)]\}/\{1 + \exp[-(\Delta G(H_2O)/RT - m[U]/RT)]\} \quad (3)$$

where m_f and m_u are the the slopes of the pre- and the posttransition lines, respectively, and [U] is the urea concentration. Using the nonlinear regression analysis program Kaleidagraph (Synergy Software, PCS Inc.), all six parameters of eq 3 were obtained with their standard errors.

Table 1: Thermodynamic Parameters Characterizing the Stability of the Three Titin Modules

module	T_m[a]	$[U]_{1/2}$[b] (M)	m[b] (cal mol^{-1} M^{-1})	$\Delta G(H_2O)$[b,c] (kcal mol^{-1})	ΔH_m[d] (kcal mol^{-1})
Ab1	43.9 (0.1)	2.06	1608 (38)	3.31 (0.15)	45.6 (1.6)
		2.03	1547 (78)	3.15 (0.21)	
Ab2	45.6 (0.3)	3.17	1313 (130)	4.18 (0.13)	61.3 (4.7)
		3.22	1463 (248)	4.71 (0.83)	
Ml1	52.5 (0.2)	4.37	985 (69)	4.31 (0.13)	68.5 (6.7)
		4.03	1445 (235)	5.80 (0.95)	

[a] Calculated from plots of ΔG vs T at $\Delta G = 0$. [b] For each module, the top line gives the results of a least-square analysis of plots of ΔG vs [urea] (eq 2), and the second line gives the results of the nonlinear regression analysis of the entire unfolding curve (eq 3). Errors are given in parentheses. [c] Conformational stability, at 25 °C, pH 4.2. [d] Obtained from the slope of ΔG vs T plots, ΔS_m, and T_m (= $T_m \Delta S_m$).

The results of the urea denaturation study are summarized in Table 1 and clearly show that the Ml1 module is the most stable of the three titin domains examined. For comparison, the results of both the linear least-squares fit and the nonlinear regression analysis are shown.

It should be emphasized at this point that more relevant in the present study is the relative stability of the modules and not the absolute values of $\Delta G(H_2O)$. As there is no single parameter to characterize (and no single method to calculate) differences in conformational stability between structurally related proteins, all relevant parameters are included in the table: (i) The midpoint of urea unfolding curve, $[\text{urea}]_{1/2}$, is the most "objective" parameter, in the sense that it can be determined quite accurately and reproducibly and is almost independent of the unfolding mechanism and the method of analysis used. (ii) m is not directly measured, and its value depends on the unfolding mechanism and the method of analysis; it is nevertheless useful as the only direct measure of the steepness of the unfolding curve. (iii) The values of $\Delta G(H_2O)$ *per se* also depend on the method of analysis used and include the error associated with m to an even larger extent, but they give an overall estimate of the stability of the protein and, as such, are quite useful.

The low values of the conformational stability could be partly due to the method of analysis used which is known to yield the lowest estimates of $\Delta G(H_2O)$ (Pace et al., 1989), but they could also reflect the nature of the proteins studied. It is reasonable to expect that modules cannot be as stable as an intact protein.

Thermal Denaturation Monitored by CD. Thermal denaturation of the three domains was monitored by following the change in the far-UV CD spectrum with increasing temperature. Since the maximal change in ellipticity between the folded and the unfolded conformation was obtained at 201–204 nm, this range of wavelengths was selected to monitor the thermal denaturation of the proteins. With increasing temperature there was a loss of the β-sheet CD pattern occurring at a different temperature for each domain (Table

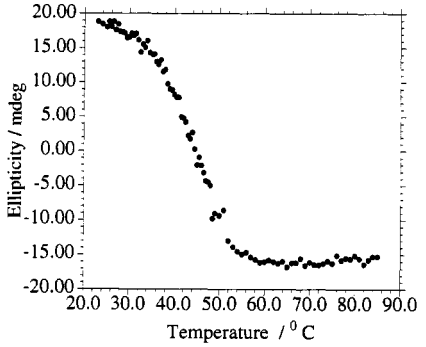


FIGURE 5: Thermal denaturation curve of Ab1 in 10 mM acetate buffer, pH 4.2, monitored by circular dichroism at 203 nm.

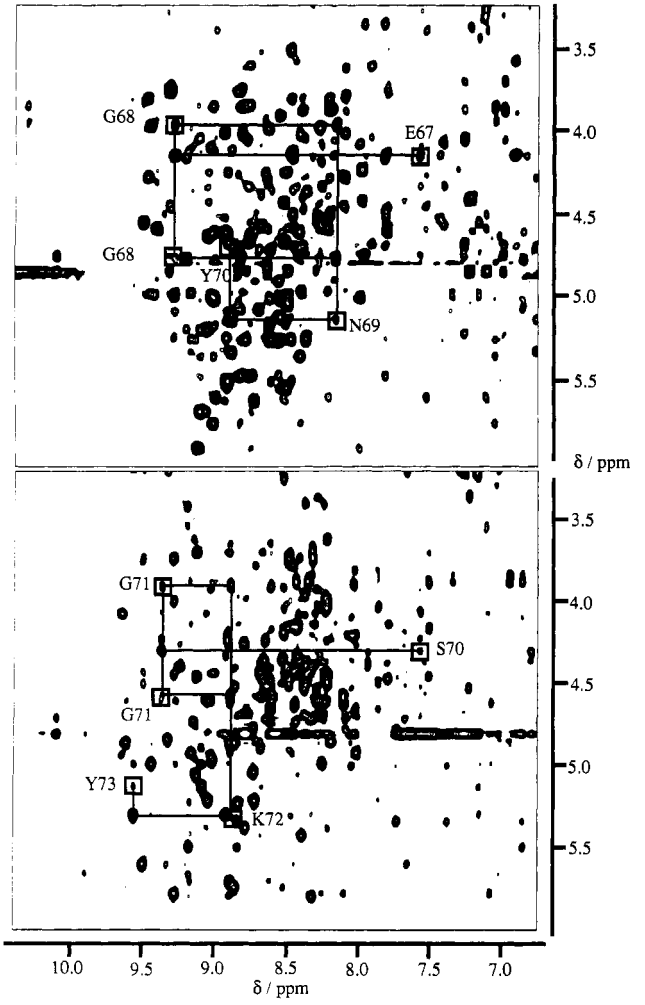


FIGURE 6: Fingerprint regions of NOESY spectra of Ml1 (top) and Ab1 (bottom) acquired under identical conditions (300 K, pH 4.2 in acetate buffer, mixing time = 120 ms). The sequential assignment pathways for a stretch of well-conserved residues are also shown.

1). Thermal denaturation curves (Figure 5) were analyzed in the standard manner, using a two-state approximation (Becktel & Schellman, 1987). Equations 1 and 2 corresponding to thermal denaturation were used to calculate the dependence of ΔG on temperature in the transition region, and subsequently the midpoint of thermal transition, T_m (where $\Delta G = 0$), and the enthalpy change, ΔH_m ($\Delta H_m = T_m \Delta S_m$, where ΔS_m is the slope of ΔG vs T), were determined. Calculation of ΔH_m values from the slope of van't Hoff plots yielded the same results. Data depicted in Table 1, which includes the T_m and ΔH_m values, show that the three titin domains differ in their thermal stability and that Ab1 is unstable relative to the other modules.

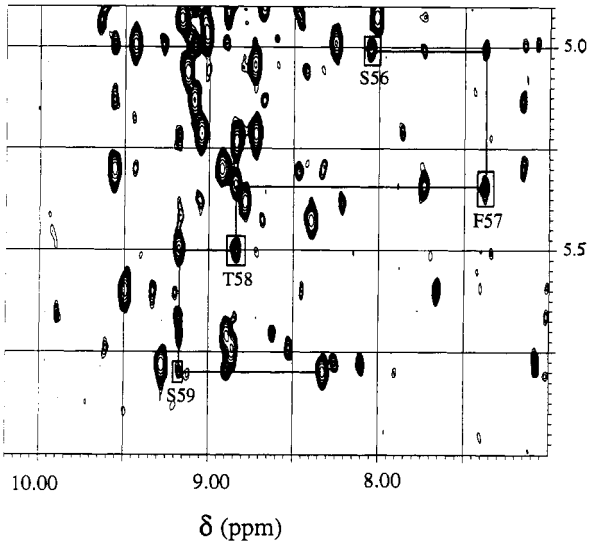


FIGURE 7: Part of the fingerprint region of a NOESY spectrum (mixing time = 120 ms) for Ab1 in acetate buffer, pH 4.2, recorded at 27 °C. The locations of the intraresidual amide proton–α-proton TOCSY cross-peaks are indicated with rectangular frames and the sequential assignment pathway with straight lines.

NMR Assignments. The NMR spectra of the three domains have a striking similarity and are all characteristic of a predominantly β-sheet protein (Figure 6). However, they differ markedly in the overall dispersion. Ml1 gives the best resolved spectra, while the NMR spectra of the Ab1 domain, although in general of a good quality, are characterized by regions of adequate dispersion mixed with regions of extensive overlap which make sequential assignment very difficult. For historical reasons, the NMR data presented here are those derived from the spectra of Ab1.

Initial spin system assignments were obtained from 2D homonuclear TOCSY and DQF-COSY spectra. Additional spin systems were identified in the 2D ^1H–^{15}N HSQC-TOCSY spectrum. Heteronuclear ^{13}C-based experiments proved invaluable; by using them the ambiguity in the aromatic region of the spectrum was completely resolved, and the previous assignments were confirmed.

Sequential assignment of the backbone ^1H and ^{15}N resonances (for the parts of the sequence assigned) was done in the conventional manner (Wüthrich, 1986) using mainly 2D ^1H–^1H NOESY, ^1H–^{15}N HSQC-NOESY, and ^1H–^{13}C HMQC-NOESY spectra to identify short-range through-space connectivities between the previously assigned spin systems. An example is given in Figure 7.

Use of any type of 3D experiment that could resolve the majority of the remaining ambiguities was precluded for the Ab1 module, because of degradation of the protein in the time required for the acquisition of a 3D spectrum. Therefore, we can presently be confident only about the assignment of 60% of the residues in Ab1. NMR work on the Ml1 domain, which shows better spectroscopic behavior, is in progress. A full account of this work will be published elsewhere.

NMR Secondary Structure. It has been widely accepted that secondary shifts, i.e., deviations of the NMR chemical shifts from their random coil values, contain valuable information regarding protein secondary structure (Pastore & Saudek, 1990; Spera & Bax, 1991; Wishart et al., 1991; Ikura et al., 1991). The $^1H_\alpha$ and $^{13}C_\alpha$ secondary shifts are believed to give the best correlation to the secondary structure (Wishart et al., 1991). Helices are characterized by positive $^{13}C_\alpha$ and negative $^1H_\alpha$ secondary shifts; the opposite is true for β-sheets. The magnitude of the deviations from the random

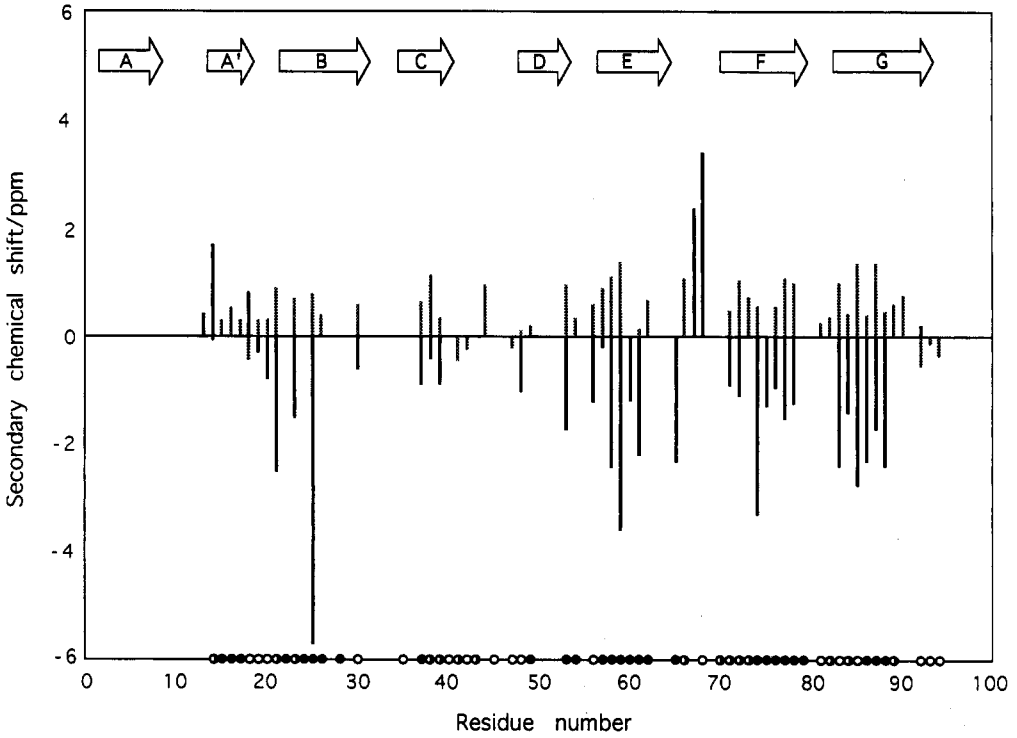


FIGURE 8: NMR secondary shifts and amide proton exchange. Plot of $^{13}C_\alpha$ (black bars) and $^1H_\alpha$ (shaded bars) secondary shifts vs residue number for the assigned part of Abl. Stretches of negative $^{13}C_\alpha$ (or positive $^1H_\alpha$) shifts indicate the location of β-strands. Qualitative amide proton–deuterium exchange data are also shown: (a) fast exchanging protons (open circles), i.e., those with lifetime less than 4 h; (b) medium (half-filled circles) with lifetime between 4 and 10 h; (c) slow exchanging protons (filled circles), i.e., those with lifetime between 10 and 26 h. No amide protons remained unexchanged after 26 h in D_2O. The putative location of the β-strands for the model used are also indicated and marked by the arrows on the top of the drawing. The residues for which no data are shown are the ones not assigned.

coil values is less pronounced for residues in β-sheets than for α-helix. However, these deviations are still characteristic of the secondary structure, and they can reliably be used for this purpose. In fact, there is evidence that this method of analysis can result, in certain cases, in a clearer definition of the secondary structure elements, such as ends of helices, than the pattern of sequential and medium-range NOEs considered alone (Shirakawa et al., 1993). In deducing the secondary shifts, we used the "coil" values given by Wishart et al. (1991), which represent average values from secondary structures other than β-sheets and α-helices (including random coil). Figure 6 shows secondary shifts for the $^1H_\alpha$ and $^{13}C_\alpha$ of the Abl residues assigned. There are clearly present stretches of secondary shifts of the same sign which identify the residues involved in β-strands (55% of the assigned residues).

Amide Proton Exchange. The Abl domain was lyophilized once from H_2O and then dissolved in D_2O. 2D TOCSY spectra were acquired at 17 °C starting immediately after dissolution; the acquisition time for each spectrum was 4 h. The majority of the cross-peaks for the amide protons (85%) were absent from the spectrum acquired 4 h following dissolution, while 26 h later there was none detectable. These data are summarized in Figure 8 and clearly show that exchange of most of the amide protons of Abl is rather rapid. This was not the case with Mll; an appreciable number of amide protons (12%) had not exchanged even after 1 month in D_2O.

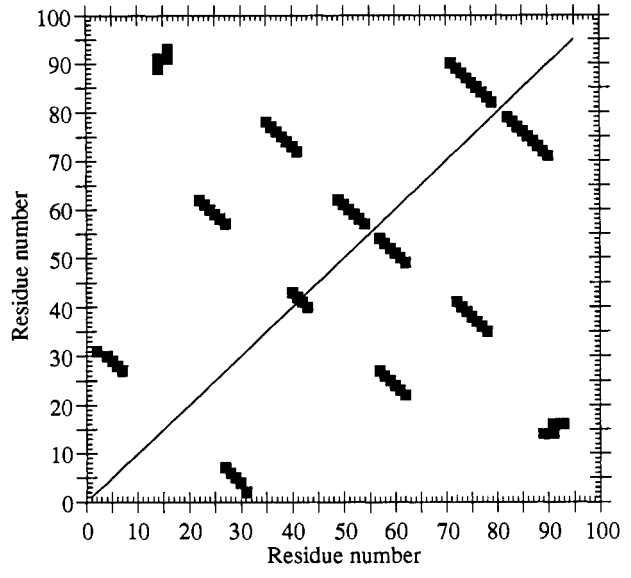
Modeling of the Tertiary Structure. The structure of type II modules was modeled after telokin according to the alignment shown in Figure 2 and suggested by Y. Harpaz and C. Chothia (personal communication). Telokin is the only member with known 3D structure (Holden et al., 1992) of the intracellular Ig subfamily, and it represents the C-terminal domain of MLCK; for this reason it is the most suitable one for the modeling of titin, despite the relatively low sequence homology. The pairwise identity between telokin and Abl, Ab2, and Mll is 26%, 25%, and 28%, respectively, while, if amino acid similarity is also considered, these numbers become 42%, 40%, and 45%, respectively. The overall fold of telokin is that of a β-sandwich of antiparallel β-sheets (Holden et al., 1992). In Figure 8 the regions where the β-strands are expected on the basis of this model are shown and compared with the experimentally determined locations of the β-strands. In the telokin structure 60% of the residues are located in β-strands, while roughly 55% of the Abl residues assigned showed secondary shifts characteristic for amino acids involved in β-strands.

In addition, a network of NH–NH and H_α–H_α NOE connectivities between nonsequential residues that could be confidently determined from our NMR data was compared with the corresponding interstrand contacts expected from the model. This is illustrated in Figure 9. NOE connectivities can de detected by NMR between protons that are less than 4 Å apart in space. Most of the expected NOEs that have not been experimentally detected correspond to residues for which the NMR assignment is not available.

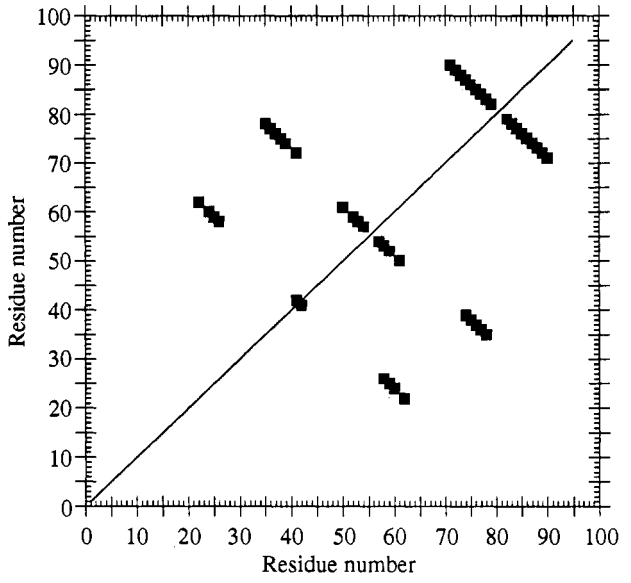
The ^1H, ^{13}C, and ^{15}N chemical shifts of the N-terminally attached His_6 sequence show in all three cases no appreciable dispersion and have values characteristic for random coil conformation. We can confidently conclude, therefore, that the His_6 tag does not participate in any way in the formation of secondary or tertiary structure.

DISCUSSION

The Ig fold, a stable β-sheet sandwich, has been elegantly described as "a stable platform upon which a diversity of sequences are displayed by varying the amino acids that are exposed on the external faces of the β sheets or on the loops of sequence connecting the β strands" (Williams et al., 1989). The wealth of sequences published in recent years has introduced even more variety in the family by including



(A)



(B)

FIGURE 9: Distance plot (A) derived from the model used and the experimentally determined connectivities (B) for the Ab1 module. In panel A all the interstrand NH–NH and H_α–H_α contacts for the model constructed after telokin are shown. In panel B the NH–NH and H_α–H_α NOE connectivities detected by NMR are indicated.

modules from intracellular proteins and forcing "the abandonment of conserved disulphide bond as the last invariant characteristic of an immunoglobulin-type domain" (Williams, 1987).

Secondary structure predictions and homology searches suggested that type II titin modules belong to the Ig superfamily (Labeit et al., 1990). The fluorescence, CD, and NMR spectra collected in this work provide direct experimental evidence in support of this hypothesis.

The wavelength of the fluorescence maximum shows that the unique tryptophan, the most conserved residue in all titin modules and in all members of the immunoglobulin superfamily (Lesk & Chothia, 1982; Williams & Barklay, 1988; Hsu & Steiner, 1992), is well buried in all three domains. This is consistent with the topology predicted.

The CD patterns obtained are typical of the so-called immunoglobulin fold, rich in β-sheets and in perfect agreement with those reported in the literature for "classical" immunoglobulins [V_{REI} fragment (Brahms & Brahms, 1980); F_c receptor (Gastinel et al., 1992); F_c fragments of IgG molecules (Hobbs et al., 1992); V_H and V_k Ab regions (Ito et al., 1992); mouse IgG (Perczel et al., 1992)] as well as for recent members of the family (single CD2 adhesion domain; Recny et al., 1990). This similarity is very meaningful, given that CD spectra of proteins of high β-pleated content, unlike those of the helical ones, show a large variation in shape and intensity (Perczel et al., 1992). CD spectra of the whole titin molecule have been reported in the past (Maruyama et al., 1986). The differences between those and the ones obtained in the present study could be attributed mainly to the different size of the fragments and the varying contribution of unfolded sequence stretches, as well as nonmodule sequences, in the different preparations of titin.

The NMR-derived plot of $^{13}C_\alpha$ and H_α secondary shifts vs residue number for Ab1 (Figure 8) also indicates the presence of β-sheet secondary structure and identifies the residues involved. There is a good agreement with the predicted location of β-strands in the sequence. Moreover, long-range connectivities (included in the contact map shown in Figure 9), representing interstrand backbone contacts, are identified in the NOESY spectra between protons from residues that are not found in the immediate proximity of each other along the primary sequence. These connectivities are characteristic of a β-sheet folding and agree very well with the fold expected on the basis of the modeling after the telokin structure.

Therefore, we can safely conclude that the three domains studied share a high degree of structural similarity, despite their overall low sequence homology. On the other hand, we should point out the difference in stability, as evidenced by the thermodynamic parameters shown in Table 1 and by the NMR-related spectroscopic behavior: while two modules appear to be quite stable and rigid, the other one, Ab1, was found much less stable and prone to degradation at high temperatures. The higher stability of the M-line domain relative to the others could perhaps be attributed to its position in the titin molecule. Within the M-line, class II motifs are separated by nonrepetitive "linker" sequences, so they could be structurally more "autonomous" than the ones in the A-band, where they are always closely flanked by type I domains (Figure 1). The M-line as a rigid anchoring plane for the thick filaments may also require a higher degree of stability of its constituting modules than along the myosin filaments, where a certain intrinsic elasticity has been demonstrated (Higuchi et al., 1992).

At this point, a word of caution should be added on the potential influence of the boundaries on the stability of isolated modules. Even in extracellular Ig domains, where the intron–exon boundaries are known, domains can be unstable when expressed singly. In the case of intracellular proteins, like titin, the situation is even less clear and that is the reason why our selection of boundaries was based on multiple alignment criteria (as described in detail under Results). We should add that the M-line domain, the one with the least defined boundaries and the shortest length, is the most stable, a fact that could further indicate the stabilizing effect of interdomain interactions.

Could the difference in stability derive from a different functional role? This question can only be answered by a detailed structural and biochemical analysis of motifs spanning a representative range of locations along the whole titin molecule; this analysis is included in the scope of our ongoing work. This type of work is complementary to investigations aiming at the characterization of the whole titin molecule. In a recent publication (Soteriou et al., 1993), CD and fluorescence data collected for the denaturation of the intact titin

molecule indicate a two-step mechanism of unfolding, with Gu-Cl midpoints at 0.1 M and 1.3 M Gu-Cl. The first transition has been interpreted as due to weak interactions between domains, while the second one indicates complete unfolding of titin. The average free energy of unfolding for each domain has subsequently been estimated at ~10 kcal/mol. While our data cannot, at this stage, lead to conclusions about domain interactions, they can yield a more direct estimate of thermodynamic parameters of isolated domains. Furthermore, large, filamentous molecules, such as titin, may collapse in solution producing a series of nonspecific interactions which might lead to artifactual estimates of their stability.

When the whole sequence of the titin molecule becomes available, a more extensive comparison between the structure and stability of modules from different regions will be possible and could explain important properties, such as the elasticity, at the molecular level.

ACKNOWLEDGMENTS

We are grateful to Catherine Joseph for technical assistance with the preparation of the proteins, to Toby Gibson for many helpful discussions, and to Arthur Lesk, Cyrus Chothia, and Spyros Georgatos for critical reading of the manuscript.

REFERENCES

Ayme-Southgate, A., Vigoreaux, Benian, G. M., & Pardue, M. L. (1991) *Proc. Natl. Acad. Sci U.S.A. 88*, 7973–7977.
Baron, M., Norman, D. G., Willis, A., & Campbell, I. D. (1990) *Nature 345*, 642–646.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 63*, 207–212.
Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., & Tschudin, R. (1990) *J. Magn. Reson. 86*, 304–318.
Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers 26*, 1859–1877.
Benian, G. M., Kiff, J. E., Neckelmann, N., Moerman, D. G., & Waterston, R. H. (1989) *Nature 342*, 45–50.
Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185–189.
Brahms, S., & Brahms, J. (1980) *J. Mol. Biol. 138*, 149–178.
Buchner, J., Lilie, H., Hinz, H. J., Jaenicke, R., Kiefhaber, T., & Rudolph, R. (1991) *Biochemistry 30*, 6922–6929.
Chen, Y. H., Yang, J. T., & Chau, K. H. (1974) *Biochemistry 13*, 3350–3359.
Driscoll, P. C., Cyster, J. G., Campbell, I. D., & Williams, A. F. (1991) *Nature 353*, 762–765.
Fürst, D. O., Osborn, M., Nave, R., & Weber, K. (1988) *J. Cell Biol. 106*, 1563–1572.
Fürst, D. O., Vinkemeier, U., & Weber, K. (1992) *J. Cell Sci. 102*, 769–778.
Fulton, A. B., & Isaacs, W. B. (1991) *BioEssays 13*, 157–161.
Funatsu, T., Higuchi. H., & Ishiwata S. (1990) *J. Cell Biol. 110*, 53–62.
Funatsu, T., Kono, E., & Tsukita, S. (1993) *J. Cell Biol. 120*, 711–724.
Gastinel, L. N., Simister, N. E., & Bjorkman, P. J. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 638–642.
Gautel M., Leonard K., & Labeit S. (1993) *EMBO J. 10*, 3827–3834.
Griesinger, G., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7878–7872.
Higgins, D. G., Labeit, S., & Gibson, T., (1994) *J. Mol. Evol.* (in press).
Higuchi, H., Suzuki, T., Kimura, S., Yoshioka, T., Maruyama, K., & Umazume, Y. (1992) *J. Muscle Res. Cell Motil. 13*, 285–294.
Hobbs, S. M., Jackson, L. E., & Hoadley, J. (1992) *Mol. Immunol. 29*, 949–956.
Holden, H. M., Ito, M., Hartshorne, D. J., & Rayment, I. (1992) *J. Mol. Biol. 227*, 840–851.
Horowits, R., Maruyama, K., & Podolsky, R. J. (1989) *J. Cell Biol. 109*, 2169–2176.
Hsu, E., & Steiner, L. A. (1992) *Curr. Opin. Struct. Biol. 2*, 422–431.
Ikura, M., Spera, S., & Bax, A. (1991) *Biochemistry 30*, 9216–9228.
Isaacs, W. B., Kim, I. S., Struve, A., & Fulton, A. B. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7496–7500.
Kurzban, G. P., & Wang, K. (1988) *Biochem. Biophys. Res. Commun. 155*, 1155–1161.
Labeit, S., Barlow, D. P., Gautel, M., Gibson, T., Holt, J., Hsieh, C. L., Francke, U., Leonard, K., Wardale, J., Whiting, A., & Trinick, J. (1990) *Nature 345*, 273–276.
Labeit, S., Gautel, M., Lackey, A., & Trinick, J. (1992) *EMBO J. 11*, 1711–1716.
Laemmli, U. K. (1970) *Nature 227*, 680–685.
LeGrice, S. F. J., & Grueninger-Leitch, F. (1990) *Eur. J. Biochem. 187*, 307–314.
Lesk, A. M., & Chothia, C. (1982) *J. Mol. Biol. 160*, 325–342.
Main, A. L., Harvey, T. S., Baron, M., Boyd, J., & Campbell, I. D. (1992) *Cell 71*, 671–678.
Maruyama, K., Kimura, S., Yoshidomi, H., Sawada, H., & Kikuchi, K. (1984) *J. Biochem. (Tokyo) 89*, 701–709.
Maruyama, K., Itoh, Y., & Arisaka, F. (1986) *FEBS Lett. 202*, 353–355.
Nave, R., Fürst, D. O., & Weber, K. (1989) *J. Cell Biol. 109*, 2177–2187.
Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N., & Campbell, I. D. (1990) *J. Magn. Reson. 87*, 488–501.
Pace, C. N., Shirley, B. A., & Thomson, J. A. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 311–330, IRL Press, Oxford.
Pace, C. N., Laurents, D. V., & Thomson, J. A. (1990) *Biochemistry 29*, 2564–2572.
Pace, C. N., Laurents, D. V., & Erickson, R. E. (1992) *Biochemistry 31*, 2728–2734.
Pastore, A., & Saudek, V. (1990) *J. Magn. Reson. 90*, 165–176.
Perczel, A., Park, K., & Fasman, G. D. (1992) *Proteins 13*, 57–69.
Piotto, M, Saudek, V., & Sklenar, V. (1992) *J. Biomol. NMR 2*, 661–664.
Recny, M. A., Neidhard, E. A., Sayre, P. H., Ciardelli, T. L., & Reinherz, E. L. (1990) *J. Biol. Chem. 265*, 8542–8549.
Saiki, R. K., Scharf, S. J., Faloona, F., Mullis, G. T., & Erlich, H. A. (1985) *Science 230*, 1350–1354.
Santoro, M. M., & Bolen, D. W. (1988) *Biochemistry 27*, 8063–8070.
Schmid, F. X. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 251–285, IRL Press, Oxford.
Shirakawa, M., Fairbrother, W. J., Serikawa, Y., Okhubo, T., Kyogoku, Y., & Wright, P. E. (1993) *Biochemistry 32*, 2144–2153.
Soteriou, A., Clarke, A., Martin, S., & Trinick J. (1993) *Proc. R. Soc. London B 254*, 83–86.
Spera, S., & Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490–5492.
Studier, F. W., & Moffat, B. A. (1991) *J. Mol. Biol. 189*, 113–130.
Studier, F. W., Rosenberg, A. H., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 62–89.
Trinick, J., Knight, P., & Whiting, A. J. (1984) *J. Mol. Biol. 180*, 331–356.
Vinkemeyer, U., Obermann, W., Weber, K., & Fürst, D. O. (1993) *J. Cell Sci. 106*, 319–330.
Wang, K. (1985) *Cell Muscle Motil. 6*, 315–369.
Whiting, A. J., Wardale, J., & Trinick, J. (1989) *J. Mol. Biol. 205*, 163–169.
Williams, A. F. (1987) *Immunol. Today 8*, 298–303.
Williams, A. F., & Barclay, A. N. (1988) *Annu. Rev. Immunol. 6*, 381–405.
Williams, A. F., Davis, S. J., He, Q., & Barclay, A. N. (1989) *Cold Spring Harbor Symp. Quant. Biol. 54*, 637–647.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311–333.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.