Solid-State NMR Studies of the Dynamics and Structure of Mouse Keratin Intermediate Filaments

James W. Mack,[‡,§] Dennis A. Torchia,[*,‡] and Peter M. Steinert[∥]

Bone Research Branch, National Institute of Dental Research, and Dermatology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

Received September 23, 1987; Revised Manuscript Received February 8, 1988

ABSTRACT: The molecular dynamics and structural organization of mouse epidermal keratin intermediate filaments (IF) have been studied via solid-state nuclear magnetic resonance (NMR) experiments performed on IF labeled both in vivo and in vitro with isotopically enriched amino acids. As a probe of the organization of the peripheral glycine-rich end domains of the IF, carbon-13 NMR experiments have been performed on subfilamentous forms (prekeratin) and on IF reassembled in vitro that had been labeled with either [1-^{13}C]glycine or [2-^{13}C]glycine, as more than 90% of the glycines of the keratins are located in the end domains. Although cross-labeling to seryl residues was observed, the proportion of serine located in the end domains is nearly the same as that for glycine. Measurements of carbon relaxation times, nuclear Overhauser enhancements, and signal intensities show that the motions of the peptide backbone in the end domains are effectively isotropic, with average correlation times distributed over the range of 0.2–20 ns. These results indicate that the end domains of IF are remarkably flexible and have little or no structural order. To probe the structural organization of the coiled-coil rod domains of the IF, separate samples of native keratin IF, raised in primary tissue culture, were labeled with L-[1-^{13}C]leucine, L-[$^{2}H_{10}$]leucine, or L-[2,3,3-$^{2}H_3$]leucine, as greater than 90% of the leucyl residues of the keratin IF types studied are located in the coiled coils which form the central core of IF. Carbon-13 spectra of L-[1-^{13}C]leucyl-labeled IF display only a rather small level of intensity at the carbonyl position, a fact initially ascribed to the ability to detect only the small fraction of leucyl residues found in the flexible end domains and not the great majority which are resident in a perhaps much more rigid core. Deuterium NMR experiments performed on IF labeled with deuteriated leucines indeed reveal a marked degree of peptide backbone rigidity within the coiled coils, confirming the initial conclusions of the carbon-13 data. In addition, analysis of the deuterium powder-pattern line shapes shows that the leucyl side chains have a substantial degree of conformational flexibility. Simulations of the line shapes of L-[$^{2}H_{10}$]leucyl-labeled IF indicate that the side-chain dynamics are well approximated by a model involving transitions between two rotameric states that differ from each other by rotations of 120° about both the C^{α}–C^{β} and C^{β}–C^{γ} bonds. These data, demonstrating relative peptide backbone rigidity yet side-chain flexibility, are interpreted to mean that the coiled coils of these keratin IF are not tightly packed together but rather form a somewhat looser structure which permits a significant degree of side-chain mobility. This structural model is consistent with the flexuous nature of IF when visualized both in intact tissue and by negative staining in vitro.

Intermediate filaments (IF)[1] are ubiquitous constituents of the cytoskeletons of most differentiated types of eukaryotic cells (Steinert et al., 1985a; Osborn & Weber, 1986; Fraser et al., 1987). Those of epithelial cells (the keratins) are perhaps the most complex since about 30 different protein chains are now known to exist in a given mammalian species. Other IF proteins seem to be less complex: there are about three to four different neurofilament chains in neuronal tissues, three to four different lamin chains in the lamina complex of the karyoskeleton of nucleated mammalian cells, and only one known protein chain for each of glial fibrillary acidic protein, desmin, and vimentin IF.

Despite wide differences in their patterns of expression, chemistry, and probable functions, all IF-forming protein chains share remarkable similarities in structure. They all consist of a central rod domain of precisely conserved secondary structure, which is flanked by end domains of highly variable size and amino acid sequence. The central rod domain consists of four discrete α-helical segments that possess a characteristic seven-residue repeat which favors the formation of a two-chain coiled-coil molecule and are interspersed with linker sequences that cannot form coiled coils. Analyses of the sequences, however, reveal at least five distinct sequence types among IF proteins: type I for acidic keratins; type II for neutral–basic keratins; type III for vimentin, desmin, and glial fibrillary acidic protein; type IV for neurofilaments; and type V for lamins. The next level of IF structural hierarchy is the two-chain coiled-coil molecule, which is formed by two chains aligned in parallel and in exact axial register and is stabilized by hydrophobic interactions of aliphatic side chains along the axis of the coiled coil. Higher orders of IF structure are not yet resolved. However, it is thought that a pair of molecules

‡ National Institute of Dental Research.

§ Supported by an NIH/NRSA postdoctoral fellowship (Grant F32 DE05433).

∥ National Cancer Institute.

[1] Abbreviations: IF, intermediate filaments; k*n* (*n* as integer), keratin chain *n*; EDTA, ethylenediaminetetraacetic acid; SDS, sodium dodecyl sulfate; TMS, tetramethylsilane; 1-^{13}C, designates carbon-13 label at peptide carbonyl carbon; 2-^{13}C, designates carbon-13 label at peptide α-carbon position; NOE, nuclear Overhauser enhancement; NMR, nuclear magnetic resonance; ppm, parts per million.

forms a four-chain building block, several of which associate to form an intact IF at least 10–15 nm in diameter. Both native in vivo IF and especially in vitro reassembled IF are known to be polymorphic; that is, they have varying masses-per-unit-length and contain 4–12 (most commonly 8) four-chain units in cross section (Steven et al., 1985; Steinert & Parry, 1985; Fraser et al., 1986). All of these data point to the conclusion that the conserved rod domains of the IF chains drive higher orders of IF structure by forming the basic core of IF and thus are responsible for the curvilinear morphological similarity of all IF, including lamins (Parry et al., 1987).

On the other hand, a number of biochemical, morphological, electron microscopic, and predictive modeling studies indicate that large portions of the end domains of the IF chains protrude from this α-helical core to form a less dense periphery about the IF (Steinert et al., 1985a; Steven et al., 1985; Fraser et al., 1987; Osborn & Weber, 1986). This paradigm is teleologically attractive because the known variability of these end domain sequences could thus adequately explain the chemical and functional diversity of the IF. However, attempts to systematically analyze the structure and organization of these end domains have been less successful. Nevertheless, the available sequence data suggest a complex subdomainal organization. For example, in the keratins, for which the largest body of sequence information exists, there are variable V1 and V2 subdomains, often highly enriched in glycines and/or hydroxy amino acids, flanked by basic E1 and E2 end subdomains. Type II keratins only possess additionally homologous sequences H1 and H2 immediately adjacent to the rod domain (Steinert et al., 1985a,b).

In order to further study the structures in both the central α-helical core of IF and their end domains, we have initiated nuclear magnetic resonance (NMR) studies of mouse epidermal keratin IF. Modern pulsed NMR spectroscopy is an established method for investigating molecular dynamics in solids (Spiess, 1978; Mehring, 1983). The techniques are well suited to the study of biologial macromolecules, provided that isotopic labeling is employed (Griffin, 1981; Opella, 1982; Kinery et al., 1984; Torchia, 1984). Herein we report observations of mouse keratin IF labeled with ^{13}C and ^{2}H.

Terminally differentiated mouse epidermis synthesizes IF containing almost exclusively keratin 1 and keratin 10 chains [using the nomenclature system of Heid et al. (1986)]. From the known amino acid sequences of these two protein chains (Steinert et al., 1983, 1985b), it is found that their V1 and V2 subdomains are very glycine- (ca. 60%) and serine- (ca. 30%) rich. Further, in k1 93% of the glycyl residues and 86% of the seryl residues are located in the end domains, while in k10 the glycyl and seryl percentages of the end domains are 92 and 90, respectively. Similarly, 90–93% of the total number of leucines are confined to the rod domains of these chains. Cultured mouse epidermal cells express IF containing mostly keratin 5, 6, 14, and 16 chains (Roop et al., 1986). Although sequence data on these chains is incomplete [keratin 6 (Steinert et al., 1984); keratin 14 (Rentrop et al., 1986); others (Roop and Steinert, unpublished data)], the distributions of glycines in the end domains (ca. 90%) and leucines in the rod domains (ca. 90%) are similar. On the basis of these data, we have labeled keratins 1 and 10 in vivo with [1-^{13}C]glycine, [2-^{13}C]glycine, and [1-^{13}C]leucine. Because of the low and variable levels of incorporation attained with in vivo labeling, we also employed cell culture techniques to produce native keratin IF containing keratin chains 5, 6, 14, and 16 labeled with L-[1-^{13}C]leucine, L-[$^{2}H_{10}$]leucine, or L-[2,3,3-$^{2}H_3$]leucine. In this way, we have been able to study the end domain (glycine rich) and rod domain (leucine rich) organization of keratin IF via solid-state nuclear magnetic resonance.

Experimental Procedures

Isolation of in Vivo Keratin IF from Mouse Epidermis. Female BALB/c mice, estimated to be 18 days pregnant, were supplied ad libitum with water containing 5% glucose and 7–10 mg/mL [1-^{13}C]- or [2-^{13}C]glycine or DL-[1-^{13}C]leucine. The newborn pups were sacrificed at 2 days of age and skinned, and the epidermis was separated by being heated in phosphate-buffered saline at 60 °C for 30 s. Keratins were extracted from the tissue with vigorous homogenization in 0.1 M sodium citrate, pH 2.6. This dissolves the keratin IF of the living cell layers only into a native subfilamentous form termed prekeratin. This has a molecular mass of 500–1000 kDa with dimensions of about 4 × 100 nm. It thus consists of several four-chain complexes associated laterally and longitudinally (Matoltsy, 1965; Skerrow, 1974; Steinert, 1975). The prekeratin was then purified by three cycles of isoelectric precipitation at pH 7 with 1 M Tris base and redissolution in the citrate buffer (Matoltsy, 1965; Skerrow, 1974; Steinert, 1975). As judged by polyacrylamide gel electrophoresis in the presence of SDS, the resulting protein consisted of keratins 1 and 10 and smaller amounts of keratins 5 and 14. Other abundant soluble proteins, including filaggrin and its precursors, were quantitatively removed by this process. For NMR studies, an isoelectric precipitate of prekeratin, neutralized to pH 7, was utilized. In some experiments, this preparation was dissolved in a buffer of 8 M urea, 50 mM Tris-HCl (pH 7.6), 1 mM EDTA, and 1 mM dithiothreitol, and the keratin IF were reassembled in vitro by dialysis into 5 mM Tris-HCl, pH 7.6, 1 mM EDTA, and 1 mM dithiothreitol (Steinert et al., 1976). The filaments were harvested by centrifugation at 100000g for 30 min. This clear keratin IF pellet was then used directly for NMR experiments.

Isolation of Keratin IF from Cultured Mouse Epidermal Cells. Newborn BALB/c mouse epidermal cells (20 × 10^6 per 150-mm dish) were cultured in medium 199 prepared without L-leucine and containing 8% fetal calf serum and low calcium ion concentration (about 0.1 mM). Added to this medium was DL-[1-^{13}C]leucine (MSD Isotopes), DL-[$^{2}H_{10}$]-leucine (MSD Isotopes), or L-[2,3,3-$^{2}H_3$]leucine (Prochem Isotopes), at a concentration of 200 mg/L (4 times the normal leucine concentration within the medium). Cells were cultured exactly as described elsewhere (Hennings et al., 1980) for 3 days, by which time confluent growth had been achieved. Cytoskeletal extracts of the cells were then prepared as described (Steinert et al., 1982) with a buffer of 1.0 M KCl–1% Triton X-100 in phosphate-buffered saline and protease inhibitors to yield undenatured native keratin IF samples. By polyacrylamide gel electrophoresis in the presence of SDS, these cytoskeletons were found to consist almost entirely (ca. 90%) of keratins 5, 6, 14, and 16. Intact cytoskeletal pellets were then used directly for NMR experiments. Before use, the keratin cytoskeletal preparations were washed thoroughly in phosphate-buffered saline reconstituted from deuterium-depleted water (Aldrich).

Several milligrams of each protein sample was hydrolyzed in 6 N HCl (Pierce) for 24 h at 112 °C, and high-resolution NMR spectra were obtained for checking the level of incorporation of the label and for scrambling of the label.

NMR Experimental Procedures. The 62.98-MHz carbon-13 NMR spectra were obtained on a home-built spectrometer (5.9-T Oxford magnet) described elsewhere (Sarkar et al., 1983). A home-built double-resonance probe with a 5 mm i.d. solenoidal coil was used. The carbon 90° pulse length

was ca. 4.5 μs for this probe. An auxilliary logic box (designed and constructed by R. G. Tschudin) allowed for either low-level broad-banded decoupling with WALTZ-16 (Shaka et al., 1983) modulation (ca. 0.8-kHz rotating frame field and 3 kHz of effective bandwidth) or high-level coherent decoupling (ca. 60-kHz rotating frame field). Signals were digitized by a NIC-2090 digital oscilloscope interfaced to the spectrometer. Homogeneity of carbon signals for this probe arrangement was typically ca. 50 Hz on an ethylene glycol sample.

The 25.15-MHz carbon-13 spectra (1.9-T Oxford magnet) were obtained on a spectrometer and probe (5 mm) of similar design to those used at 5.9 T. Rotating frame field strenghs were about equal to those used at 62.98 MHz.

Carbon-13 spin–lattice relaxation times were measured with the saturation-recovery method. Transverse relaxation times were measured with a 90°–t–180°-detect Hahn spin-echo pulse sequence. Nuclear Overhauser enhancements (NOE) were determined by comparing spectral intensities observed with and without proton saturation.

Deuterium spectra at 38.45 and 76.77 MHz were obtained on the 5.9-T instrument and on a modified NIC-500, respectively. Homemade probes with 5 mm i.d. solenoidal coils were used, and 90° pulse lengths were 2.2–2.5 μs. For some experiments at the higher field, a 6 mm i.d. coil was used, with a 90° pulse length of ca. 3.2 μs. Quadrupole echo spectra were obtained with the standard ($90_{\pm x}$–t–90_y–t–detect) sequence (Davis et al., 1976) with t set to either 30 or 40 μs. Inversion-recovery spectra were obtained by preceeding the quadrupole echo sequence with a 180°–τ– sequence. Quadrature signals were digitized at a rate of 500 ns, without prior analog filtering. Exponential apodization was applied to the time-averaged signals resulting in 1.0–2.5 kHz of Lorentzian line broadening.

Simulations of the ^2H line shapes were performed on a VAX 750 using a two-site program developed at NIH and a multisite program developed at the National Magnet Laboratory (Wittebort et al., 1987). Both programs correct the line shape for the effects of finite pulse-power levels (Bloom et al., 1980; Hiyama et al. 1986) and short anisotropic T_2 values (Speiss & Sillescu, 1981).

Results and Discussion

Structural Organization of Glycine-Rich End Domains. Initial experiments were performed on a pellet of native keratin IF, obtained as a subfilamentous form (prekeratin) from intact mouse epidermis, that had been labeled with [1-^{13}C]glycine (carbonyl label). The pellet wet weight was ca. 21 mg, with a protein content of ca. 10% by weight. A 62.98-MHz carbon spectrum of this sample obtained with WALTZ-16 (low-level) decoupling is shown in Figure 1A. The carbonyl peak at ca. 173 ppm is essentially Lorentzian with a full width at half-height of about 5 ppm and rises well above the upfield aliphatic peaks. Comparison of this spectrum with a spectrum of IF in natural abundance (Figure 1B) shows that only the carbonyl signal is significantly enhanced beyond the natural abundance level.

IF samples were reassembled in vitro from these same labeled native prekeratin preparations to compare the behavior of the glycine-rich domains in the two structural forms. The 62.98-MHz WALTZ-decoupled spectrum of reassembled IF, Figure 1B, is essentially the same as the spectrum of the native subfilamentous IF, Figure 2A. A high-level decoupled spectrum of the reassembled IF reveals a slight increase in carbonyl peak amplitude and sharpening of the line (Figure 2B).

The observation of narrow, isotropic glycine carbonyl signals in spectra obtained with WALTZ decoupling strongly suggests

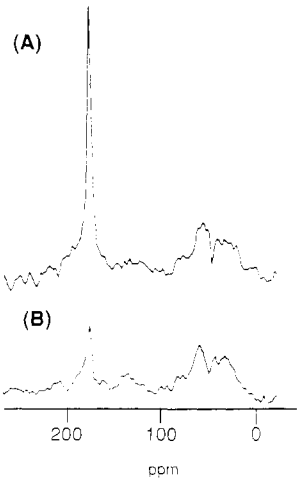

FIGURE 1: Carbon-13 spectra at 25 °C of keratin IF, obtained as prekeratin from intact mouse epidermis, at 62.98 MHz under WALTZ-modulated proton decoupling. (A) Filaments labeled with [1-^{13}C]glycine. (B) Unlabeled filaments. Each spectrum represents 12 600 scans, and ca. 200 Hz of Lorentzian broadening is used. Each spectrum was obtained with a recycle time of 5 s and with NOE.

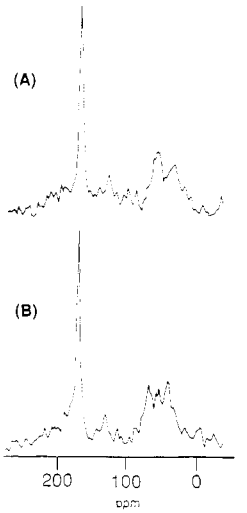

FIGURE 2: Carbon-13 spectra at 25 °C of IF, reassembled in vitro from native prekeratin, labeled with [1-^{13}C]glycine acquired at 62.98 MHz. (A) Spectrum obtained under conditions of low-level WALTZ-16 decoupling with Overhauser enhancement. (B) Spectrum obtained with high-level coherent decoupling with NOE. About 200 Hz of line broadening was applied. Each spectrum is the average of 12 600 acquisitions. Recycle times of 5 s were used.

that, as a consequence of molecular motion, there is a high degree of intrinsic dipolar decoupling of the glycine carbonyl carbons from neighboring protons. This follows because WALTZ decoupling is insufficient to average the short-range carbon–proton dipolar couplings of glycyl residues whose motions are highly restricted. The large static chemical shift anisotropy of the glycyl carbonyl carbon (Stark et al., 1983) is also completely averaged by molecular motion. Therefore, the glycyl residues, which are almost exclusively located in the amino- and carboxyl-terminal end domains of the keratin chains, undergo a high degree of segmental motion which is independent of the state of assembly of the IF.

The small effect of high-power decoupling on the glycyl carbonyl signal, Figure 2, is further evidence that molecular motion alone decouples the glycyl carbonyl carbons from protons. In view of the fact that high-power decoupling sharpens signals of immobile carbons (Mehring, 1983), one might ask why signals arising from carbonyl and aliphatic carbons in the IF core domains are not evident in the high-

Table I: Experimental Spin–Lattice Relaxation Times for the Carbonyl Carbon of Glycine in Labeled Keratin IF[a]

frequency (MHz)	T_1[b]	T_2[c]	NOE[d]
25.15	1.14		2.2
62.98	1.81	17.2	

[a] Measurements were made at 62.98 and 25.15 MHz at 25 °C. The Overhauser enhancement at 25.15 MHz and the transverse decay time at 62.98 MHz are also given. [b] In seconds; uncertainty 10%. [c] In milliseconds; uncertainty 20%. [d] $1 + \eta$; uncertainty 25%.

power decoupled spectrum, Figure 2B. Motionally restricted carbonyl carbons have large line widths and are difficult to detect even when high-power decoupling is applied because of their large static chemical shift anisotropies, ca. 150 ppm. In contrast, immobile aliphatic carbons, having shift anisotropies in the 20–60 ppm range, are more readily observed when high-power decoupling is applied. Unfortunately, the low signal to noise ratio of the aliphatic signals in Figure 2—a consequence of limited amount of IF available—precludes determining what effect high-power decoupling has upon the aliphatic signal intensity. We will return to this topic when discussing the spectra in Figure 3, which were obtained on larger amounts of IF.

The measured relaxation times, T_1 and T_2, of the labeled glycyl carbonyl carbons are listed in Table I. The ^1H–^{13}C dipole interaction and chemical shift anisotropy (Abragam, 1961) contribute to carbonyl carbon spin relaxation. Using the carbonyl carbon shift anisotropy parameters measured by Stark et al. (1983) and assuming that the carbonyl carbon dipolar relaxation rate is $^1/_{15}$ that of a protonated carbon (Lyerla & Torchia, 1975), we calculated the shift anisotropy and dipolar contributions to the carbonyl relaxation times using a log χ^2 distribution of a correlation times model (Schaefer, 1973). The ratio of the observed carbonyl T_1 values at 62.98 and 25.14 MHz is 1.58. The log χ^2 calculations using p = 11 and b = 1000 (Schaefer, 1973; Lyerla & Torchia, 1975; Flemming et al., 1980) predict that this T_1 ratio occurs at an average correlation time, $\bar{\tau}$, of 2.1 ns. A similar result, $\bar{\tau}$ = 2.5 ns, is obtained from the ratio of the carbonyl T_1 and T_2 measured at 62.98 MHz. The carbonyl carbon NOE value predicted at 25.14 MHz by the log χ^2 model, assuming that $\bar{\tau}$ = 2 ns, is 1.85, a value which agrees with the observed NOE, Table I, to within the experimental uncertainty.

Although we are able to analyze the relaxation data assuming that all residues in the terminal domains have the same average correlation time, some heterogeneity in the dynamical behavior of residues in these domains might be expected because of factors such as heterogeneity in the amino acid sequence and restrictions on flexibility imposed by the IF core on portions of the IF terminal domains. We will return to the subject of dynamical heterogeneity when analyzing the spectra of the α-carbon-labeled IF.

We have thus far assumed that the carbonyl label did not scramble into other amino acids. However, metabolic pathways are known that readily interconvert glycine and serine. We checked for scrambling of the glycine carbonyl label by comparing a carbon spectrum of hydrolyzed IF in natural abundance with a spectrum of hydrolyzed IF labeled with [1-^{13}C]glycine. This comparison revealed the presence of signals of both [1-^{13}C]serine, 172.9 ppm, and [1-^{13}C]glycine, 173.3 ppm, having relative signal areas that agreed with the theoretical value of 4:7 calculated from the protein composition (Steinert et al., 1983). This result shows that the label had equilibrated between serine and glycine before incorporation into the IF. Although both seryl and glycyl residues contribute to the labeled carbonyl carbon signal observed in Figures 1 and 2, our preceeding discussion of mobility in the IF terminal domains still applies because, as mentioned above, most of the serines (86% in k1 and 90% in k10) like most of the glycines (93% in k1 and 92% in k10) are in the IF terminal domains (Steinert et al., 1983, 1985b).

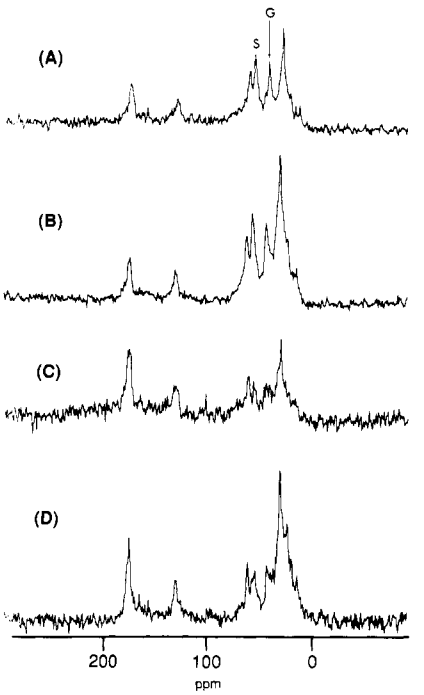

FIGURE 3: Carbon-13 spectra at 25 °C and 62.98 MHz of isolated prekeratin IF. (A) A WALTZ-16-decoupled spectrum of a sample incorporating [2-^{13}C]glycine. The α-carbon resonance is at ca. 42 ppm. S and G mark the locations of the serine and glycine α-carbon peaks. (B) A spectrum of the same sample taken under high-level coherent decoupling. (C) Spectrum of a sample incorporating L-[1-^{13}C] leucine, WALTZ decoupling. (D) A high-level decoupled spectrum of the same sample as in (C). The upfield aliphatic regions of spectra C and D serve essentially as natural abundance comparisons for the analogous regions of spectra A and B of the α-labeled sample. Lorentzian broadening of 20 Hz is used in each spectrum. Each spectrum is the result of ca. 16 000 scans.

Because of their nearly equal chemical shifts, the glycyl and seryl carbonyl carbon signals are not resolved in Figures 1 and 2. However, the glycyl and seryl α-carbon shifts differ by 12–15 ppm (Howarth & Lilley, 1978), and the carbon spectra of IF labeled with [2-^{13}C]glycine, Figure 3A,B, have signals at 42 and 57 ppm which are assigned to glycyl and seryl α-carbons, respectively. The signal at 61 ppm is assigned to the methylene carbon of Tris.

Although sharp α-carbon signals of seryl and glycyl residues are evident in Figure 3A,B, the intensity of these signals is only slightly above the natural abundance signal intensity, Figure 3C,D, in the aliphatic region of the spectrum. This observation contrasts with what was found for the [1-^{13}C]glycine-labeled IF and suggests that the sharp signals do not represent all of the labeled α-carbons. This hypothesis was checked by analyzing a carbon spectrum of a hydrolyzed labeled IF sample which showed that both the seryl and glycyl residues were enriched sixfold with carbon-13. This level of enrichment together with the fact that there are ca. 1.2 times as many glycyl α-carbons in the terminal domains as protonated aromatic carbons implies that the area of the sharp glycyl α-carbon signal in Figure 3A,B should be about 7 times greater than the area of aromatic carbon signal centered at 130 ppm. This is clearly not the case, and on the basis of the signal areas observed, we estimate that only 15–20% of the glycyl residues contribute to the sharp glycyl α-carbon signal. Similar con-

siderations show that 35–50% of the seryl residues are represented by the sharp seryl α-carbon signal.

Further analysis, employing the log χ^2 model, shows that these intensity anomalies are quite reasonably explained as the product of dynamical heterogeneity of the end domain residues. The analysis, discussed in the supplementary material (see paragraph at end of paper regarding supplementary material) shows that if average correlation times span the range of 0.2–20 ns, then large line widths and concomitant reduced intensity levels are predicted for the glycyl and seryl α-carbons.

The use of high-power decoupling increases the signal area in the aliphatic region by ca. 50%, Figure 3B, as compared with the signal area obtained with WALTZ-16 decoupling, Figure 3A. We ascribe the signal increase to contributions from motionally restricted carbons in the IF core domain whose resonances are too broad to detect in the absence of high-power decoupling.

The collection of data gathered for both the carbonyl- and α-carbon-labeled samples is consistent with a model of effectively isotropic motions of the glycine and serine residues of the IF terminal domains. The average correlation times of these motions are in the 0.2–20-ns range. Accordingly, it can be concluded that there is a high degree of flexibility of the pepide backbone of the end domain sequences, and little or no ordered structure is likely to exist.

Structural Organization of the Leucine-Rich Rod Domains. (*A*) *Carbon-13 Labeling.* In the interest of comparing the mobility of the coiled-coil core regions of the IF with that of the glycine/serine-rich end domains, we labeled native keratin IF, obtained as prekeratin, in vivo with L-[1-^{13}C]leucine. The IF spectra recorded with WALTZ and high-level decoupling, Figure 3C,D, have carbonyl signal intensities that are only ca. 30% larger than natural abundance signals Figure 3A,B. This observation suggests that most of the labeled leucyl carbonyl signal is not detected.

A carbon spectrum of the [1-^{13}C]leucine-labeled IF following hydrolysis shows that the leucine carbonyl carbon is ca. 15-fold enriched. However, the observed carbonyl signal intensity enhancement in the labeled IF spectra, ca. 30%, agrees with that calculated (25–30%) by assuming that only the five leucyl residues in the IF terminal domains have narrow lines and are observed in Figure 3C,D. The remaining 46 leucyl residues per chain in the core domain evidently have very broad lines as a consequence of their very limited mobility. We estimate that the residual shift anisotropy of the core leucyl carbonyl carbons exceeds 100 ppm in view of the fact that we were not able to observe their signals employing high-power decoupling. In contrast, an aliphatic carbon signal increase of ca. 50% occurs when high-power decoupling is used, cf. Figure 3C,D, because of contributions of decoupled aliphatic IF core carbons which have shift anisotropies of only 20–40 ppm (Schaefer & Stejskal, 1979).

These data indicate that the majority of the leucyl residues must occupy regions of relatively low peptide backbone flexibility. The observed leucyl carbonyl signal is due mainly to a minor component of much greater flexibility. This conclusion is consistent with the distribution of leucyl residues in these IF preparations: ca. 90% are in the rod domains, and ca. 10% are in the glycine/serine-rich end domains (Steinert et al., 1983, 1985b), which we have shown are highly mobile.

(*B*) L-[$^2H_{10}$]*Leucine Label.* In order to study dynamics of leucyl side chains in the IF core domain, further experiments were carried out with deuterium NMR. Keratin IF, isolated as cytoskeletal extracts labeled in cell culture with L-[$^2H_{10}$]-leucine, were prepared. Tissue cultures were used to produce the IF so as to maximize ^{2}H uptake.

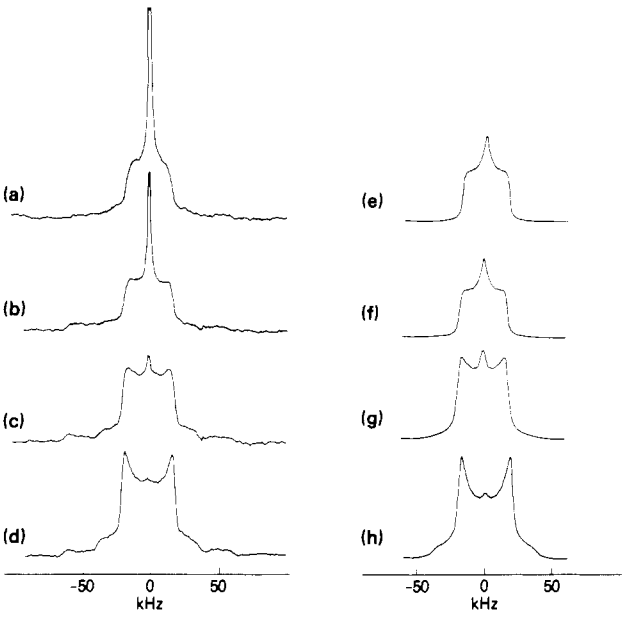

FIGURE 4: Comparison of 76.77-MHz ^{2}H spectra of L-[$^2H_{10}$]-leucine-labeled IF with line shapes generated by a discrete two-site jump model. Spectra at temperatures (a) 26, (b) 5, (c) −23, and (d) −45 °C, each an average of 25 000 scans. Simulations with two-stie half-angle of 54.75° between two equally populated equilibrium sites with "fast" and "slow" jump rates (k_f and k_s) assigned to two different classes of "molecules". (e) $k_f = 1 \times 10^6$, $k_s = 1.65 \times 10^4$; (f) $k_f = 5 \times 10^5$, $k_s = 1.65 \times 10^4$; (g) $k_f = 6.5 \times 10^4$, $k_s = 5.73 \times 10^3$; (h) $k_f = k_s = 3.8 \times 10^3$. A 1:1 ratio of slowly and rapidly jumping components is assumed in the simulations.

A series of ^{2}H spectra recorded in the 26 to −45 °C temperature range, Figure 4a–d, consists mainly of powder line shapes rather than isotropic resonances. This result contrasts sharply with the observed glycyl/seryl end domain carbon resonances and implies a much lower degree of mobility for most of the leucyl residues.

At 5 °C and above, an isotropic component, comprising ca. 20% of the signal area, is observed. Some of this component may arise from the deuterons remaining in ^{2}H-depleted buffer, but the primary source is the deuterons in the ca. 10% of leucyl residues in the IF end domains. Like the glycyl residues, the leucyl residues would generate an isotropic resonance, and computer simulations (to be discussed) show that the isotropic component is well fitted assuming that 10% of the leucyl residues in the IF are isotropically mobile. This result shows that the remaining leucyl residues are in domains of substantial rigidity, in accord with the ^{13}C data. They furthermore provide strong evidence that the core domains of the IF obtained from tissue culture were in the organized native state and were not denatured.

(*C*) *Simulation of Line Shapes of* L-[$^2H_{10}$]*Leucyl-Labeled Samples.* The $\eta \approx 1$ leucine methyl patterns observed in Figure 5a,b are similar to line shapes observed in spectra of collagen fibrils (Batchelder et al., 1982) and of bacteriophage fd coat protein (Colnago et al., 1987) labeled with L-[$^2H_{10}$]-leucine. Following these investigators, we analyze our IF spectra assuming a dynamic isomerization between two predominant side-chain conformations that differ from each other by rotations of 120° about the C^α–C^β and the C^β–C^γ bonds (Benedetti, 1977; Janin et al., 1978). The net effect of these rotations is to change the orientation of each C^γ–C^δ bond (a methyl threefold axis) by 109.5°. If the two conformations have equal populations, an $\eta = 1$ ^{2}H powder pattern is obtained when isomerization is rapid (Soda & Chiba, 1969).

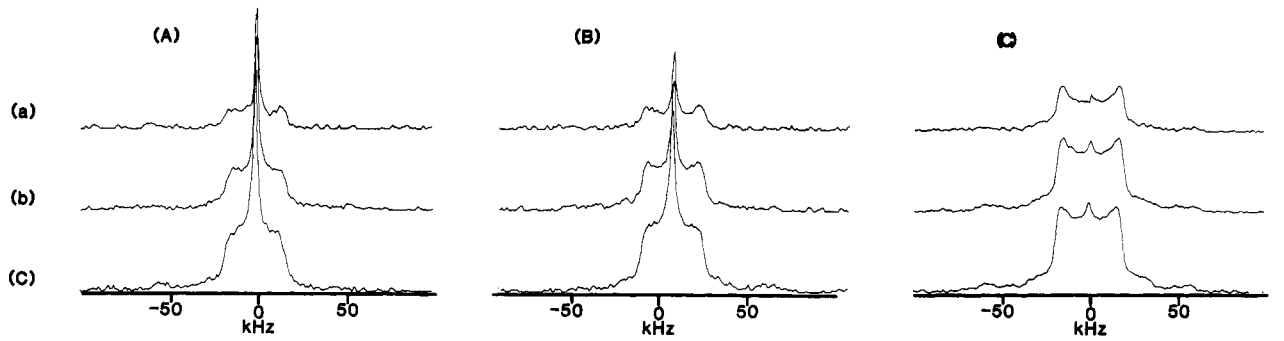

FIGURE 5: A series of spectra of keratin IF at 76.77 MHz, produced in primary tissue culture, labeled with L-[$^2H_{10}$]leucine at temperatures of (A) 26, (B) 5, and (C) −23 °C, as a function of τ, the time interval between the 90° pulses of the quadrupole echo sequence. (a) τ = 100 μs, (b) τ = 50 μs, and (c) τ = 30 μs. Each spectrum is an average of 25 000 scans.

Table II: Powder-Averaged Deuterium Spin–Lattice Relaxation Times (T_1) at 76.77 MHz (and at 38.45 MHz at 26 °C Only) and Values of the ω_2 Principle Frequency for the Methyl Deuterium Resonances of Keratin IF Labeled with L-[$^2H_{10}$]Leucine as a Function of Temperature

temp (°C)	T_1 (ms)	ω_2 (kHz)
26	35.05	16.5
	28.68[a]	
5		17.8
−23	19.29	19.0
−45	11.58	19.7

[a] Value measured at 38.45 MHz.

Calculations using the above model provided good fits of line shapes observed at 5 and 26 °C, but discrepancies between the calculated and observed relative intensities were large; i.e., the ratio of calculated signal intensity at 26 °C to that at −45 °C is 3 times greater than is observed. Furthermore, the line shape at −23 °C could not be satisfactorily simulated.

Because the −23 °C spectrum appears to be a superposition of two line shapes, one having $\eta \approx 0$ and the other having $\eta \approx 1$, line shapes were calculated assuming that there were two classes of leucyl side chains, distinguished by their rates of isomerization, Figure 4e–h. The calculated relative intensities are in much better agreement with experimental values than are those generated assuming a single isomerization rate. Although the intensities calculated at 26 and 5 °C had to be reduced by ca. 45% in order to match the data, some reduction in observed signal intensity is expected at the higher temperatures because of small amplitude motions, dielectric losses, and reduced Boltzmann factors. The appearance of the calculated line shapes is rather insensitive to variations of the slow vs fast component proportions over the range from 1:1 to 2:1. The relative integrated intensities of the simulated spectra are in closest agreement with the experimental values when the ratio is 2:1, however.

Steinert et al. (1984) have shown that for k6 73% of the leucyl residues in the rod domain occupy either a or d positions in the heptad repeat sequence. We therefore tentatively assign the slowly isomerizing leucyl residues to a and d positions in the heptad repeat and the rapidly isomerizing leucyl residues to the remaining positions in the heptad or to leucyl residues in the linker sequences.

Molecular motions, whose rates approximate the quadrupolar coupling constant, reduce ^2H T_2 values and cause significant intensity losses in ^2H powder-pattern line shapes (Speiss & Sillescu, 1981). As a consequence of short ^2H T_2 values, the line-shape components calculated assuming slow jump rates at 5 and 26 °C lose about 60% of their signal intensities. In contrast, the components calculated assuming the fast rates lose only ca. 10% of their signal intensities.

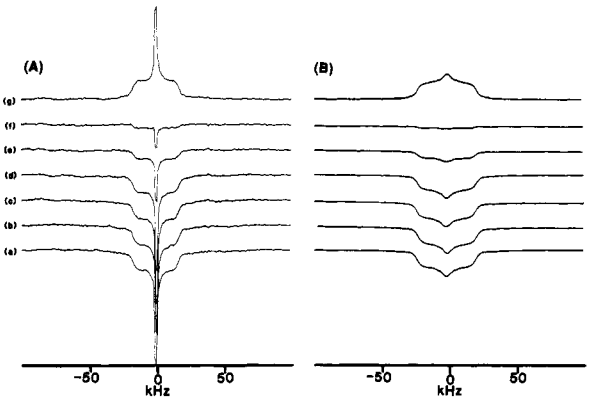

FIGURE 6: Comparison of actual and simulated inversion-recovery profiles at 76.77 MHz for keratin IF, produced in primary tissue culture, labeled with L-[$^2H_{10}$]leucine. (A) Actual data and (B) calculated profile generated by the six-site discreet hopping model. Recovery times are (a) 0.75, (b) 1.0, (c) 2.0, (d) 3.0, (e) 12.0, (f) 20.0, and (g) 100.0 ms. The experimental spectra are averages of 25 000 scans each. The simulated spectra assume a threefold methyl rotation rate of 4.0×10^9 s^{-1} and a side-chain isomerization rate of 1.0×10^6 s^{-1}.

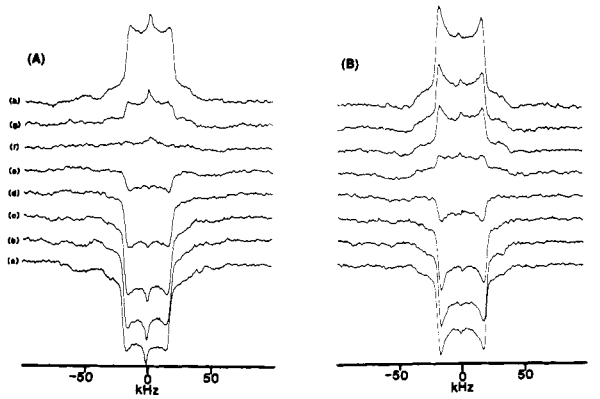

FIGURE 7: Inversion-recovery profiles of IF labeled with L-[$^2H_{10}$]leucine at 76.77 MHz and temperatures of (A) −23 and (B) −45 °C. The recovery times are (a) 0.75, (b) 1.0, (c) 2.0, (d) 5.0, (e) 10.0, (f) 15.0, (g) 20.0, and (h) 50.0 ms. Each spectrum is an average of 25 000 scans.

The reduction of signal intensity with increasing τ is clearly evident in Figure 5. Particularly noteworthy is the anisotropic nature of the signal reduction at 26 and 5 °C, in agreement with that calculated for two-site jumping. This model predicts that the portions of the powder pattern in the vicinity of the three principal frequencies decay the most slowly (Speiss & Sillescu, 1981). This prediction is indeed verified, indicating that the two-site model is a good approximation of leucyl side-chain dynamics.

Further information about leucyl side-chain dynamics is provided by analysis of spin–lattice relaxation times, T_1, Figures 6 and 7 and Table II. The field independence of the T_1 values measured at 26 °C, Table II, shows that the spin–lattice relaxation is the consequence of motions having correlation times, τ_c, on the fast side of the T_1 minimum; i.e. $\tau_c < 10^{-9}$ s. This conclusion is confirmed by the observation that the methyl T_1 values decrease as temperature decreases. Therefore, a rapid motion, such as three-site methyl jumps (Batchelder et al., 1983), rather than the two-site side-chain isomerization, causes the spin–lattice relaxation.

A six-site jump model, which combines the slow two-site isomerization and the fast three-site methyl jumps, can account for both the ^2H line shapes and relaxation data, Figure 6. Details of the calculations performed with the six-site model are provided in the supplementary material. The simulation of the 26 °C inversion-recovery line shape, Figure 6B, obtained with this model was done assuming a single isomerization rate, 10^6 s^{-1}, for all leucyl side chains. This simulation and others (not shown) confirm two things: that methyl rotation is the principal factor determining the powder-averaged T_1 and that the composite effects of the two types of motion are responsible for the suppression of the relatively prominent T_1 anisotropy that is produced by methyl rotation alone (Batchelder et al., 1983).

(*D*) L-[*2,3,3-*$^2H_3$]*Leucine Label.* We prepared this sample in order to determine whether the large-amplitude side-chain motions evident at the methyl positions are indeed present at positions closer to the peptide backbone. While these positions are labeled in the [$^2H_{10}$]leucine-labeled IF, their signals are obscured by the strong signals of the methyl deuterons. The 76.77-MHz spectra of this sample imply the continuance at the β-position of the same large-amplitude motion found at the terminating methyl positions of the the leucyl side chain. The room temperature line shape manifests sizable distortion due to T_2 anisotropy. At −23 °C these distortions are no longer very prominent as the side-chain motion is largely quenched at this point. Further exposition on these experiments, including spectra at 24 and −23 °C, is given in the supplementary material.

Comparisons with Other Fibrous Proteins. The degree of mobility observed in the glycine/serine-rich terminal domains of the keratin IF, as manifested in the carbon spectra of samples labeled with [1-^{13}C]glycine, Figures 1 and 2, is very much like what has been observed for the extensible regions of elastin. In the low-power decoupled spectrum of chick aortic elastin, the valyl carbonyl line width, 4 ppm (Fleming et al., 1980; Torchia et al., 1983), is comparable to the value observed for [1-^{13}C]glycyl-labeled keratin IF. In both cases the observation of narrow isotropic carbonyl signals points to a high degree of flexibility of the peptide backbone that effectively averages the carbonyl carbon/proton dipolar couplings and the carbonyl shift anisotropy.

The ^{13}C NMR spectra observed for collagen fibers contrast sharply with spectra observed for the end domains of keratin IF and the extensible domains of elastin. The glycyl carbonyl resonances of various collagens are broad powder patterns having chemical shift anisotropies ranging from 109 to 140 ppm (Sarkar et al., 1983). Only upon unfolding of the helix structure is the line width of the glycyl carbonyl comparable to the value observed in keratin IF, Figure 1A,B (Torchia & Piez, 1973; Jelinski & Torchia, 1979).

The data described in this paper show that the peptide backbone of the leucine-rich coiled-coil domains which form the core of keratin IF is also severely restricted in mobility, similar to collagen but in dramatic contrast to the glycine/serine-rich end domains of the IF. This conclusion is supported by the observations that (a) only a small percentage (ca. 10%) of the [1-^{13}C]leucyl signal is observed in the NMR spectra of the IF, Figure 3C,D, and (b) the motion of the leucyl side chain is restricted to methyl rotation plus a two-site isomerization.

Sequence and structural data have firmly established that coiled-coil α-helices form by the interaction of neighboring protein chains which are composed of characteristic seven-residue (heptad) repeat units, in which the first, a, and fourth, d, positions are preferentially occupied by apolar amino acids. The two-chain molecule of keratin and of other such coiled-coil proteins is thus stabilized primarily by hydrophobic interactions of the interfacing apolar residues which are buried along the axis of the coiled coil. Higher orders of IF structure, involving the lateral and longitudinal packing of the coiled coils of the individual molecules, are thought to be dominated by ionic interactions involving residues at the other heptad positions. In this way, a substantially rodlike structure is thus considered to form in IF.

The data presented in this paper show that the peptide backbone in the vicinity of the leucine-labeled sites of the rod domains executes restricted motions similar to those observed in fibrous collagen. Additionally, our analysis indicates that the leucyl side chains of the rod domains execute large-amplitude motions similar to those of leucyl side chains in collagen. This implies that the hydrophobic interface stabilizing the two protein chains of the coiled-coil unit in IF is not rigid, but presumably adopts a somewhat open and flexible arrangement. Other types of experimental data exist in support of this view. First, various type III and IV IF chains possess a single cysteine residue in their rod domains in the f heptad position, which can be catalytically oxidized to disulfide bonds within the two-chain coiled-coil molecule (Quinlan et al., 1982; Carden & Eagles, 1983). Since amino acids in the f heptad position are directed radially outward from the coiled coil, this means that the coiled coil must be locally quite flexible in order to permit alignment of the cysteine residues of neighboring chains to form the disulfide bond (Parry et al., 1985). Second, detailed structural studies of the two-chain coiled coil of tropomyosin have revealed a significant degree of axial flexibility of the coiled coil (Phillips et al., 1980).

Taken together, the degree of long-range axial flexibility of the peptide backbone of the rod domains in IF, deduced by analogy with collagen data, and the mobility of the apolar side chains in the rod domains indicate that IF are not long rigid structures but rather possess a marked degree of axial freedom. These NMR results therefore provide a direct structural and molecular explanation for the morphological data which have established that IF, visualized in vivo and reassembled in vitro, are flexible, extensible structures that display considerable mobility within cells under a wide variety of conditions. It is well established that the epidermal keratin specific IF associated protein filaggrin interacts with keratin IF, an interaction presumably involving either the end domains or the rod domains of the IF. We would expect that the extent of conformational flexibility of either or of both the end and rod domains may change significantly upon the formation of macrofibrils with filaggrin. We are currently exploring this possibility with additional NMR experiments.

In summary, we have used solid-state NMR spectroscopy to establish that the glycine/serine-rich end domains of epidermal keratin IF have little or no structural order. Additionally, our data have revealed that that the central core of

the IF structure, composed principally of the α-helical rod domains, is relatively rigid in comparison to the end domains but that the notable degree of mobility found for the apolar (leucyl) side chains is consistent with the macroscopically flexible nature of IF.

ACKNOWLEDGMENTS

We thank Dr. S. Yuspa for providing us with murine epidermal cell cultures and Dr. R. G. Griffin for providing us with a copy of the multiple-site NMR line-shape simulation program developed in his laboratory. We thank Dr. A. C. Steven for numerous discussions during the course of this work and for previewing the manuscript. The expert technical support provided by R. G. Tschudin has been invaluable.

SUPPLEMENTARY MATERIAL AVAILABLE

Further discussion of some of the considerations that went into the analysis of the glycyl/seryl α-carbon intensities, a section describing in more detail the specifics of the six-state discrete jumping model used in analyzing the leucyl side-chain dynamics, and a more detailed discussion of the interpretation of the experiments done on the L-[2,3,3-$^2H_3$]leucine-labeled filaments (9 pages). Ordering information is given on any current masthead page.

REFERENCES

Abragam, A. (1961) *Principles of Nuclear Magnetism*, Oxford University Press, London.

Barnes, R. G. (1974) *Adv. Nucl. Quadrupole Reson. 1*, 335–357.

Batchelder, L. S., Sullivan, C. E., Jelinski, L. W., & Torchia, D. A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 386–389.

Batchelder, L. S., Niu, C. H., & Torchia, D. A. (1983) *J. Am. Chem. Soc. 105*, 2228–2231.

Benedetti, E. (1977) in *Proceedings of the Fifth American Peptide Symposium* (Goodman, M., & Meienhofer, J., Eds.) pp 257–273, Wiley, New York.

Bloom, M., Davis, J. H., & Valic, M. I. (1980) *Can. J. Phys. 58*, 1510–1517.

Burnett, L. J., & Muller, B. H. (1971) *J. Chem. Phys. 55*, 5829–5831.

Colnago, L. A., Valentine, K. G., & Opella, S. J. (1987) *Biochemistry 26*, 847–854.

Davis, J. H., Jeffrey, K. R., Bloom, M., Valic, M. I., & Higgs, T. P. (1976) *Chem. Phys. Lett. 42*, 390–396.

Engel, A., Eichner, R., & Parry, D. A. D. (1985) *J. Ultrastruct. Res. 90*, 323–335.

Fleming, W. W., Sullivan, C. E., & Torchia, D. A. (1980) *Biopolymers 19*, 597–617.

Fraser, R. D. B., MacRae, T. P., Parry, D. A. D., & Suzuki, K. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1179–1183.

Fraser, R. D. B., Steinert, P. M., & Steven A. C. (1987) *Trends Biochem. Sci. (Pers. Ed.) 12*, 43–45.

Griffin, R. G. (1981) *Methods Enzymol. 72*, 108–174.

Hartmann, S. R., & Hahn, E. L. (1962) *Phys. Rev. 128*, 2042–2053.

Heid, H. W., Werner, E., & Franke, W. W. (1986) *Differentiation (Berlin) 32*, 101–119.

Hennings, H., Michael, D., Cheng, C., Steinert, P. M., Holbrook, K. A., & Yuspa, S. H. (1980) *Cell (Cambridge, Mass.) 19*, 245–254.

Hentschel, R., & Speiss, H. W. (1979) *J. Magn. Reson. 35*, 157–162.

Hiyama, Y., Silverton, J. V., Torchia, D. A., Gerig, J. T., & Hammond, S. J. (1986) *J. Am. Chem. Soc. 108*, 2715–2723.

Howarth, O. W., & Lilley, D. M. J. (1978) *Prog. Nucl. Magn. Reson. Spectrosc. 12*, 1–40.

Janin, J., Wodak, S., Levitt, M., & Maigrot, B. (1978) *J. Mol. Biol. 125*, 357–386.

Jelinksi, L. W., & Torchia, D. A. (1979) *J. Mol. Biol. 133*, 45–65.

Keniry, M. A., Gutowsky, H. S., & Oldfield, E. (1984) *Nature (London) 307*, 383–386.

Lyerla, J. R., & Torchia, D. A. (1975) *Biochemistry 14*, 5175–5183.

Matoltsy, A. G. (1965) in *Biology of the Skin and Hair Growth* (Lyne, A. G., & Short, B. F., Eds.) pp 291–312, American Elsevier, New York.

Mehring, M. (1983) *NMR: Basic Princ. Prog., 2nd Ed., 11*, 1–340.

Opella, S. J. (1982) *Annu. Rev. Phys. Chem. 33*, 533–562.

Osborn, M., & Weber, K. (1986) *Trends Biochem. Sci. (Pers. Ed.) 11*, 469–472.

Parry, D. A. D., Conway, J., & Steinert, P. M. (1987) *Int. J. Biol. Macromol. 9*, 137–145.

Phillips, G. N., Fillers, J. P., & Cohen, C. (1980) *Biophys. J. 32*, 485–502.

Pines, A., Gibby, M. G., & Waugh, J. S. (1973) *J. Chem. Phys. 59*, 569–590.

Rentrop, M., Knapp, B., Winter, H., & Schweizer, J. (1986) *J. Cell Biol. 103*, 2583–2591.

Roop, D. R., Toftgard, R., Kronenberg, M. S., Clark, J. H., & Yuspa, S. H. (1984) in *Molecular Biology of the Cytoskeleton* (Borisy, G. G., Cleveland, & Murphy, D. B., Eds.) pp 409–414, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Roop, D. R., Huitfeld, H., & Yuspa, S. H. (1987) *Differentiation (Berlin) 35*, 43–150.

Sarkar, S. K., Sullivan, C. E., & Torchia, D. A. (1983) *J. Biol. Chem. 258*, 9762–9767.

Sarkar, S. K., Sullivan, C. E., & Torchia, D. A. (1985) *Biochemistry 24*, 2348–2354.

Schaefer, J. (1973) *Macromolecules 6*, 882.

Schaefer, J., & Stejskal, E. O. (1979) in *Topics in Carbon-13 NMR Spectroscopy* (Levy, G. C., Ed.) pp 283–324, Wiley, New York.

Shaka, A. J., Keeler, J., & Freeman, R. (1983) *J. Magn. Reson. 53*, 313–340.

Skerrow, D. (1974) *Biochem. Biophys. Res. Commun. 59*, 1311–1316.

Soda, G., & Chiba, T. (1969) *J. Chem. Phys. 50*, 439–455.

Speiss, H. W. (1978) *NMR: Basic Princ. Prog. 15*, 55–214.

Speiss, H. W., & Sillescu, H. (1981) *J. Magn. Reson. 42*, 381–389.

Stark, R. E., Jelinski, L. W., Rubin, D. J., Torchia, D. A., & Griffin, R. G. (1983) *J. Magn. Reson. 55*, 266–273.

Steinert, P. M. (1975) *Biochem. J. 149*, 39–48.

Steinert, P. M., & Parry, D. A. D. (1985) *Annu. Rev. Cell. Biol. 1*, 41–65.

Steinert, P. M., Idler, W. W., & Zimmerman, S. B. (1976) *J. Mol. Biol. 108*, 547–567.

Steinert, P. M., Aynardi, M., Zackroff, R. V., & Goldman, R. D. (1982) in *Methods and Perspectives in Cell Biology* (Wilson, A., Ed.) Vol. 24, Part A, pp 399–419, Academic, New York.

Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L., & Steven, A. C. (1983) *Nature (London) 302*, 794–800.

Steinert, P. M., Parry, D. A. D., Raccosin, E. L., Idler, W. W., Steven, A. C., Trus, B. L., & Roop, D. R. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5709–5713.

Steinert, P. M., Steven, A. C., & Roop, D. R. (1985a) *Cell (Cambridge, Mass.) 42*, 411-420.
Steinert, P. M., Parry, D. A. D., Idler, W. W., Johnson, L. D., Steven, A. C., & Roop, D. R. (1985b) *J. Biol. Chem. 260*, 7142-7149.
Steven, A. C., Hainfeld, J. F., Trus, B. L., Wall, J. S., & Steinert, P. M. (1983) *J. Cell Biol. 97*, 1939-1944.
Steven, A. C., Trus, B. L., Hainfeld, J. F., Wall, J. S., & Steinert, P. M. (1985) *Ann. N.Y. Acad. Sci. 455*, 371-379.
Torchia, D. A. (1984) *Annu. Rev. Biophys. Bioeng. 13*, 125-144.
Torchia, D. A., & Piez, K. A. (1973) *J. Mol. Biol. 76*, 419-424.
Torchia, D. A., Batchelder, L. S., Fleming, W. W., Jelinski, L. W., Sarkar, S. K., & Sullivan, C. E. (1983) *Ciba Found. Symp. 93*, 98-111.
Wittebort, R. J., Olejniczak, E. T., & Griffin, R. G. (1987) *J. Chem. Phys. 86*, 5411-5420.

Tyrosine Hydrogen-Bonding and Environmental Effects in Proteins Probed by Ultraviolet Resonance Raman Spectroscopy†

Peter G. Hildebrandt,‡ Robert A. Copeland,§ and Thomas G. Spiro*
Department of Chemistry, Princeton University, Princeton, New Jersey 08544
Jacek Otlewski‖ and Michael Laskowski, Jr.
Department of Chemistry, Purdue University, West Lafayette, Indiana 47907
Franklyn G. Prendergast
Department of Biochemistry and Molecular Biology, Mayo Foundation, Rochester, Minnesota 55901
Received December 30, 1987; Revised Manuscript Received March 14, 1988

ABSTRACT: Ultraviolet resonance Raman spectra with 229-nm excitation are reported for aqueous tyrosine and for ovomucoid third domain proteins from chicken [OMCHI3(-)] and from chachalaca [OMCHA(-)], as well as α_1-, α_2-, and β-purothionin. At this excitation wavelength interference from phenylalanine is minimized, and it is possible to determine the frequencies of the Tyr ring modes ν_{8a} and ν_{8b}. The ν_{8b} frequency decreases with the degree of Tyr H-bond donation, reaching a limiting value for deprotonated tyrosine. This spectroscopic indicator of H-bond strength was calibrated by using the model compound *p*-cresol in H-bond acceptor solutions for which the enthalpy of H-bond formation can be obtained from the literature. With this calibration it is possible to estimate Tyr H-bond enthalpies in proteins for which Tyr is a H-bond donor; values of 13.7, 9.6, and 11.2 kcal/mol were found for OMCHA3(-) and for α_1- (or α_2-) and β-purothionin, respectively. The intensity of the 1176-cm^{-1} ν_{9a} band of Tyr excited at 229 nm and also the intensity ratio of the Tyr 830/850-cm^{-1} Fermi doublet excited at 200 nm both correlate strongly with the estimated H-bond enthalpies, but large deviations are seen for the purothionins, reflecting a special environment for the Tyr residue of these proteins, which is believed to be constrained in a hydrophobic pocket. The molar intensity of the strong $\sim$1000-cm^{-1} ν_{12} band of phenylalanine in aqueous solution is about half the value observed in most proteins. Addition of ethylene glycol to aqueous phenylalanine increases the intensity, which attains a value similar to those seen in proteins. Protein environmental effects on UVRR intensities for aromatics are expected to be common.

The importance of hydrogen bonding (H-bonding)[1] in protein structure and reactivity is well established (Creighton, 1983). Amide backbone H-bonding is the key structural element determining secondary structures of proteins. The H-bonding of individual amino acid side chains to other residues and to substrates plays important roles in enzyme mechanisms (Cantor & Schimmel, 1980a; Walsh, 1979). Among the amino acid side chains that can participate in H-bonding, the phenolic hydroxide of tyrosine (Tyr) can serve as either proton donor or acceptor, or both simultaneously (Baker & Hubbard, 1984). For this reason biochemists have long sought spectroscopic probes of tyrosine H-bonding in proteins.

The most widely used method for studying Tyr phenolic H-bonding is ultraviolet (UV) (difference) absorption spectroscopy (Demchenko, 1986; Cantor & Schimmel, 1980b). However, this approach requires one to differentiate the effect of Tyr hydrogen bonding from all the possible interactions of the chromophore in the protein matrix. Fluorescence spectroscopy (Demchenko, 1986) may also be employed but is

† This work was supported at Princeton by NIH Grant GM 25158 to T.G.S. and at Purdue by NIH Grant GM 10831 to M.L.
* Author to whom correspondence should be addressed.
‡ Present address: Max-Planck-Institut für biophysikalische Chemie, Abteilung Spektroskopie, D-3400 Göttingen, FRG.
§ Present address: Arthur Amos Noyes Laboratory of Chemical Physics, 127-72, California Institute of Technology, Pasadena, CA 91125.
‖ Present address: Institute of Biochemistry, University of Wroclaw, Tamka 2, 50-137 Wroclaw, Poland.

[1] Abbreviations: UVRR, ultraviolet resonance Raman; H-bond, hydrogen bond; Tyr, tyrosine; Trp, tryptophan; Phe, phenylalanine; OHMCHA3(-), chachalaca ovomucoid third domain carbohydrate free; OMCHI3(-), chicken ovomucoid third domain carbohydrate free; Erb, erabutoxin B; Pα_1, Pα_2,Pβ, purothionin α_1, α_2, β; S/N, signal to noise ratio; Asp, aspartate.

0006-2960/88/0427-5426$01.50/0