Journal of Cellular Biochemistry 30:291–309 (1986)
Protein Structure, Folding, and Design 135–153

One- and Two-Dimensional NMR Spectral Analysis of the Consequences of Single Amino Acid Replacements in Proteins

John L. Markley, David H. Croll, R. Krishnamoorthi, Gilberto Ortiz-Polo, William M. Westler, W.C. Bogard, Jr., and M. Laskowski, Jr.

Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706 (J.L.M., R.K., W.M.W.) and Department of Chemistry, Purdue University, West Lafayette, Indiana 47907 (D.H.C., G.O.-P., W.C.B., M.L.)

The traditional approach of using homologous sequences to elucidate the role of specific amino acid residues in protein structure and function becomes more meaningful as the number of differences is minimized, with the limit being alteration of a single residue. For small proteins in solution, NMR spectroscopy offers a means of obtaining detailed information about each residue and its response to a given change in the protein sequence. Extraction of this information has been aided by recent progress in spectrometer technology (higher magnetic fields, more sensitive signal detection, more sophisticated computers) and experimental strategies (new NMR pulse sequences including multiple-quantum and two-dimensional NMR methods). The set of avian ovomucoid third domains, which consists of the third domain proper plus a short leader (connecting peptide) and has a maximum of 56 amino acid residues, offers an attractive system for developing experimental methods for investigating sequence-structure and struc-

Abbreviations used: OMXXX3, unglycosylated ovomucoid third domain isolated following limited proteolysis of native ovomucoid by staphylococcal proteinase (unless noted otherwise) from the following species of birds: OMTKY3, turkey (*Meleagris gallopavo*); OMIPF3, [Indian] peafowl (*Pavo cristatus*); OMSVP3, silver pheasant (*Lophura nycthemera*); OMGMQ3, Gambel's quail (*Lophortyx gambelii*); OMJPQ3, Japanese quail (*Coturnix coturnix japonica*); OMCHI3, chicken (*Gallus gallus*); OMFTD3, fulvous tree duck (*Dendrocygna bicolor*). COSY, homonuclear two-dimensional correlated spectroscopy; NOESY, two-dimensional nuclear Overhauser effect spectroscopy. pH*, uncorrected glass electrode pH meter reading of a sample solution containing ^2H_2O made following calibration of the instrument with standard ^1H_2O buffers; 2DFT, two-dimensional Fourier transform; FID, free induction decay; TSP, 3-(trimethylsilyl)propionic acid-2,2,3,3-^2H (sodium salt); DSS, 3-(trimethylsilyl)-1-propanesulfonic acid (sodium salt); TMS, tetramethylsilane.

David H. Croll is now at Biophysics Institute, Boston University Medical Center, Boston, MA 02118.

W.C. Bogard, Jr., is now at Centocor, 244 Great Valley Parkway, Malvern, PA 19355.

Received June 10, 1985; revised and accepted September 17, 1985.

ture-function relationships in proteins. Our NMR results provide examples of sequence effects on pK_a' values, average conformation, and internal motion of amino acid side chains.

Key words: NMR spectroscopy, ovomucoid, ovomucoid third domain, protein, hydrogen ion dissociation constant, histidine, tyrosine

Developments in the broad field of biotechnology, including improved methods of peptide synthesis, efficient gene cloning procedures, and site-selective mutagenesis, make possible the synthesis of small proteins of essentially any desired sequence. These advances have revived interest in several classical areas of protein chemistry, both because biotechnology provides exciting new tools for studies of protein structure-function relationships and because additional fundamental knowledge of proteins is needed for the full practical exploitation of the new technology. In general, it is easier now to produce a new protein (at least in limited quantity) than to determine its structure and functional properties. Recent studies have demonstrated the prospects for engineering proteins with altered properties. For a given functional goal, the basic questions are (1) what sequence to create and (2) how to evaluate the protein that has been made. Two approaches to protein design currently are being explored. In the first, slight alterations are made in the sequence of a known protein in order to change, for example, its activity, specificity, or stability. In the second, a structural framework is synthesized that has been designed to have desired functional properties on the basis of principles deduced from the catalog of known protein X-ray structures; an active site may then be appended to this structure. Progress with both approaches will depend on the development of efficient methods for determining the conformational and functional properties of the designed proteins. These include biochemical or biological assays along with X-ray crystallography and other techniques of physical biochemistry.

What role can NMR spectroscopy be expected to play in this endeavor? Since NMR chemical shifts are highly sensitive to molecular configuration, the spectra can be used to determine if a protein has a unique conformation and how similar this structure is to that of other protein molecules [1]. It is sometimes desirable to determine whether two proteins are identical, for example, whether an engineered protein is identical to the original one produced in situ by the gene that was cloned. In cases of altered conformation, the nuclear Overhauser effect can be used to deduce the proximity of groups, and this information may serve to determine a rough structure for the altered protein [2]. NMR spectroscopy is also ideally suited for determining information about the chemical properties of protein groups and their bonding interactions. Determination of microscopic pK_a' values of functional groups in small proteins is routinely carried out by NMR [1,3] and, in favorable cases, can be performed with proteins of M_r 60,000 or higher [1,4]. Other chemical properties amenable to study by NMR are solvent accessibilities, hydrogen bonding, and participation of particular groups in ligand binding [1]. Kinetic data over a wide range of rates can be obtained by an analysis of NMR parameters. Spectral data can be used to measure the rates and activation energies of a variety of dynamic molecular properties, including hydrogen exchange rates at individual sites, on and off rate constants for ligand binding, rates of conformational changes affecting protein structure, and rates of internal motions of side chains (for example, flip rates of aromatic rings) [5].

What are the desiderata of a prototype protein system for NMR investigations of the effects of amino acid replacements? Small size is important, because it minimizes the effort required to resolve and assign NMR signals and limits the number of residue positions that need to be changed in order to evaluate the structural and functional roles of various parts of the molecule. Owing to intrinsic limitations in NMR sensitivity, protein solubility and stability are important factors along with the ability to make large amounts of the parent protein and its variants. An X-ray structure of at least one member of the protein family is essential for providing detailed geometrical information. Finally, one wishes to relate the spectral information to known physical biochemical information about the functional properties of the parent and variant molecules. With these ideas in mind, we have chosen two protein systems for our NMR investigations. One is the group of staphylococcal nucleases that David Shortle has produced by mutagenesis [6], and the other is the set of avian ovomucoid third domains that have been isolated and sequenced by Michael Laskowski, Jr., and coworkers [7]. The first NMR study of a single amino acid replacement in a protein involved two naturally occurring variants of staphylococcal nuclease [8]. We shall concentrate here, however, on the results with the simpler (smaller) protein family, the ovomucoid third domains.

Ovomucoid is, after ovalbumin, the most abundant protein in avian egg white. The ovomucoid molecule, a glycoprotein of about 186 amino acid residues [9], contains three, tandem, homologous domains, which evolved as the consequence of two gene duplications with the subsequent loss of one domain [10]. Although we have carried out NMR studies of the whole protein and each isolated domain [11], most of our NMR work [12–14] has been with the fragment of 56 or fewer amino acid residues ($M_r \sim 6{,}000$) that contains the third domain proper plus a short leader (connecting peptide). For simplicity this fragment will be referred to as the *ovomucoid third domain* (OMXXX3). Michael Laskowski's group at Purdue has sequenced over 100 ovomucoid third domains [15]. If the third domain proper is considered, 42 pairs of these sequences differ by only one substitution. Sequences of several ovomucoid third domains that we have studied by NMR spectroscopy are compared in Figure 1. These sequences are prepared by limited proteolysis of the whole ovomucoid. The third domain so produced is partially glycosylated ($\sim 50\%$) at asparagine 45, and the oligosaccharide-containing and unglycosylated species can be separated by gel exclusion chromatography [16]. For simplicity, the work we report here concerns the unglycosylated protein. Numerous species of birds are raised commercially, and from the whites of their eggs one can easily prepare 1 gm or more of the isolated third domain.

Ovomucoid domains function as efficient serine proteinase inhibitors with K_{assoc} values as high as 7×10^{11} [7]. Comparisons of ovomucoid third domains from closely related species of birds have revealed radical changes in their specificity for various proteinases resulting from sequence changes [7]. The most important residue for specificity is the reactive site P1 residue (residue 18 in Figure 1), with other residues in the contact region between the proteinase and inhibitor of secondary importance and noncontact residues of tertiary importance if any at all [17]. Of considerable interest is the question whether it will be possible to devise a sequence-to-reactivity algorithm for this set of protein proteinase inhibitors in which the contributions of single substitutions are treated in linear fashion [18]. The results to date indicate that this strategy works remarkably well and suggest that a given amino acid residue replacement generally causes only minor perturbations at remote sites.

```
                1         10        20        30        40        50
OMTKY3  LAAVSVDCSEYPKPACTLEYRPLCGSDNKTYGNKCNFCNAVVESNGTLTLSHFGKC
OMIPF3  LAAVSVDCSEYPKPACTLEHRPLCGSDNKTYGNKCNFCNAVVESNGTLTLSHFGKC
OMSVP3  LAAVSVDCSEYPKPACTMEYRPLCGSDNKTYGNKCNFCNAVVESNGTLTLSHFGKC
OMGMQ3  FAAVSVDCSEYPKPACTLEYRPLCGSDNKTYANKCNFCNAVVESNGTLTLSHFGKC
OMJPQ3  LAAVSVDCSEYPKPACPKDYRPVCGSDNKTYSNKCNFCNAVVESNGTLTLNHFGKC
OMCHI3  LAAVSVDCSEYPKPDCTAEDRPLCGSDNKTYGNKCNFCNAVVESNGTLTLSHFGKC
OMFTD3  VAT--VDCSDYPKPACTLEYMPLCGSDNKTYGNKCNFCNAVVDSNGTLTLSHFGKC
```

Fig. 1. Sequences of seven ovomucoid third domains that have been studied by NMR spectroscopy: OMTKY3, turkey [9]; OMIPF3, peafowl [7]; OMSVP3, silver pheasant [9]; OMGMQ3, Gambel's quail [9]; OMJPQ3, Japanese quail [9]; OMCHI3, chicken [9]; and OMFTD3, fulvous tree duck [15]. The ovomucoid third domain isolated from Japanese quail eggs shows a Ser/Gly polymorphism at position 32 that is generally found to be about 70%:30% when OMJPQ3 is isolated from a large number of eggs [16]. The sequence shown here is that of the predominant (Ser) form; and should be labeled more properly as OMJPQ3S. The numbering system we are employing starts with the residue following the staphylococcal proteinase cleavage site [9]. Residues that differ from the corresponding ones in OMTKY3 are underscored.

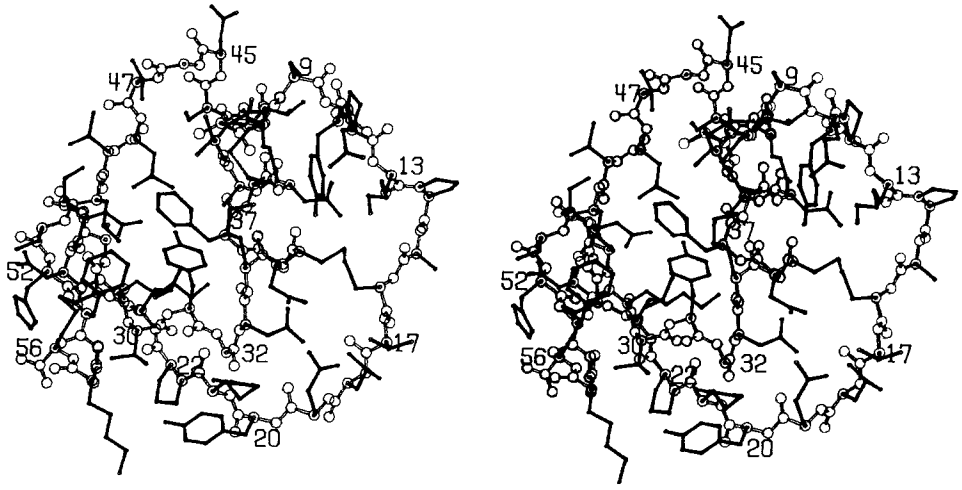

Fig. 2. Three-dimensional structure of turkey ovomucoid third domain (OMTKY3) derived from the X-ray crystallographic structure of the complex between OMTKY3 and *Streptomyces griseus* proteinase B [20].

These ideas are supported by a recent comparison of the X-ray structures of ovomucoid third domains from Japanese quail (OMJPQ3), turkey (OMTKY3), and silver pheasant (OMSVP3) [19]. The structure of OMTKY3 was determined from crystals of its complex with *Streptomyces griseus* proteinase B [20], whereas the OMJPQ3 [21] and OMSVP3 [19] structures were from crystals of the proteins alone. The three-dimensional structure of OMTKY3 is shown in Figure 2. The differences between the structures of OMKTY3 and OMSVP3, which have the same sequence except for a $Leu^{18} \rightarrow Met^{18}$ replacement at the P1 reactive site residue, could be explained in terms of slight conformational rearrangements of OMTKY3 caused by

its interaction with the proteinase [19]. The OMSVP3 and OMJPQ3 structures differ in the amino terminal region (probably as the result of intermolecular interactions in the crystal), in the region around residue 17 (as the result of a $Pro^{17} \rightarrow Thr^{17}$ substitution), and in the region between residues 44 and 48 (a perturbation by a remote substitution) [19].

Here we summarize our one- and two-dimensional NMR studies of seven ovomucoid third domains (Fig. 1). We show that residue replacements provide an effective method of making primary spectral assignments that can then be extended to the rest of the molecule by two-dimensional NMR methods. The results show that the overall structure of these proteins is conserved but that amino acid replacements can lead to local changes in structure and dynamics.

MATERIALS AND METHODS

Proteins

Eggs were obtained from bird farms or local suppliers. Staphylococcal proteinase (*S aureus* V8) was purchased from Miles Laboratories, Inc. (Elkhart, IN). NMR glassware was from Wilmad Co. The ^2H_2O (99.8% isotope) was from Bio-Rad Laboratories (Richmond, CA).

Crude ovomucoid was isolated from avian egg whites by the method of Lineweaver and Murray [22] as modified by Bogard et al [16] and purified by gel filtration through Sephadex G-75. The third domains were produced, unless noted otherwise, by limited proteolysis of native ovomucoid with staphylococcal proteinase and purified by chromatographic methods [16] to obtain the nonglycosylated OMXXX3 component. The identity and purity of OMXXX3 samples were verified by amino acid analysis. The OMXXX3 concentration in solutions used for NMR spectroscopy depended on the NMR experiment and is given in the figure legends. The pH of each solution was adjusted to the desired value with 1 M KO^2H or 1 M ^2HCl.

NMR Spectroscopy

NMR spectra were obtained with Nicolet NT-200, NT-360, or NT-470 spectrometers as described previously [11–13,23]. Details of unpublished NMR procedures used are given in the figure legends. Unless otherwise noted, all ^1H chemical shifts are referenced to TSP; 1,4-dioxane (assumed to be 67.8 ppm from TMS) was used as an internal reference for ^{13}C chemical shifts, which are reported relative to TMS.

RESULTS AND DISCUSSION

One-Dimensional (1D) NMR Spectroscopy

A general comparison of the ovomucoid third domains can be made on the basis of comparisons of 1D ^1H NMR spectra. Figure 3 presents spectra of ovomucoid third domains from silver pheasant, turkey, and Gambel's quail. OMTKY3 and OMSVP3 differ by $Leu^{18} \rightarrow Met^{18}$, and OMTKY3 and OMGMQ3 differ by $^1Leu \rightarrow {}^1Phe$ and $Gly^{32} \rightarrow Ala^{32}$. Difference spectra, generated by subtracting spectra of members of these two protein pairs, show the expected intensity in the spectral regions of the replaced residues; however, they exhibit minor changes in the signals from other residues. To interpret these differences, it is necessary to assign their origin to

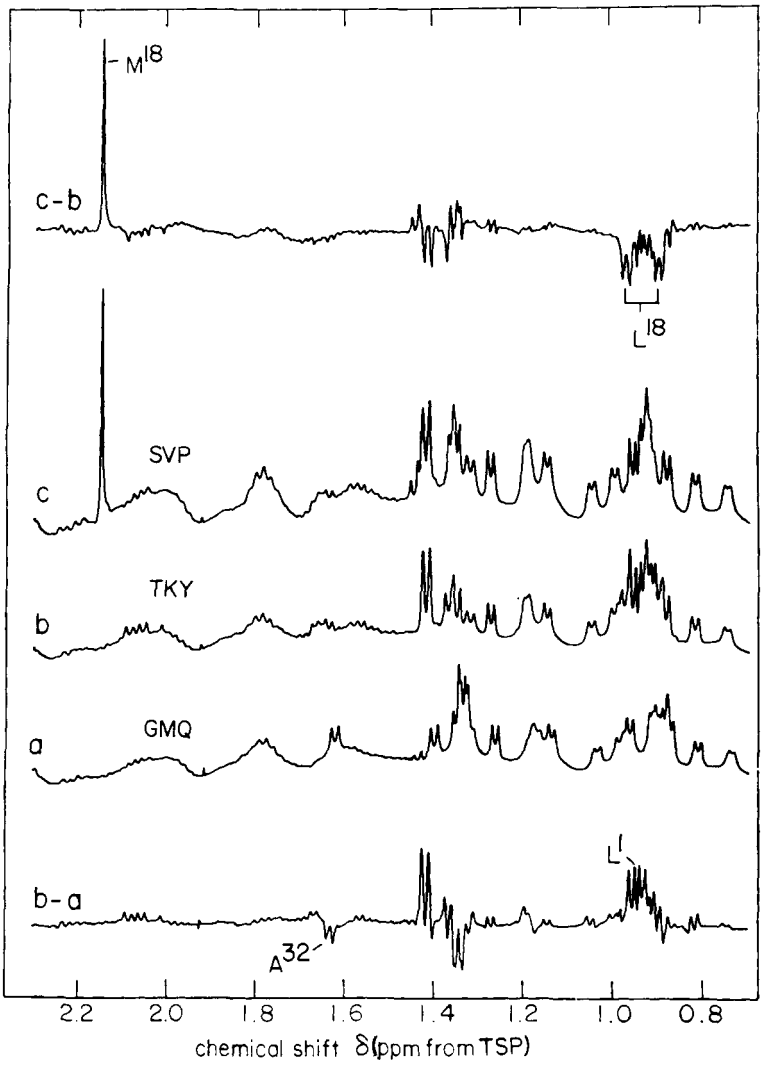

Fig. 3. Part of the aliphatic spectral region of the 470 MHz ^1H NMR spectrum of three ovomucoid third domains: a, Gambel's quail (OMGMQ3); b, turkey (OMTKY3); and c, silver pheasant (OMSVP3). The difference spectra, shown at the top and bottom, were generated by digitally subtracting the spectra of the individual proteins. OMGMQ3 and OMTKY3 differ by substitutions at two positions: $Phe^1 \rightarrow Leu^1$ and $Ala^{32} \rightarrow Gly^{32}$. OMSVP3 and OMTKY3 differ by a single substitution: $Met^{18} \rightarrow Leu^{18}$ (Fig. 1). The major peaks in the difference spectra originate from residues at the position(s) substituted. Sample conditions were 4 mM protein in 0.2 M KCl in ^2H_2O, pH* 8.0. The NMR probe temperature was 25°C. The spectra were collected by using a standard single-pulse experiment. A 90° pulse was used to excite the spins, and 512 free induction decays were collected into 16K of data memory. The spectral width was 5,000 Hz, and the data were collected using quadrature detection. The acquisition time was 1.64 sec, and the recycle delay was 4 sec.

particular nuclei and explain the changes in terms of altered protein structure. The former is easier to accomplish than the latter, so that we frequently are in the position of being able to deduce that a change occurs at a particular residue while being vague about the nature of that change.

We have detailed a number of cases in which an amino acid substitution leads to changes in the chemical shifts of neighboring residues. For example, among the four variants whose ^1H NMR spectra are compared in Figure 4, apart from OMGMQ3, which has Ala^{32} in place of the Gly^{32} present in the other three proteins, none of the sequence differences involve alanines or threonines—residues whose methyl resonances appear in the spectral region shown. The substitution $Tyr^{20} \rightarrow His^{20}$ (OMTKY3 vs OMIPF3) leads to upfield shifts in the Thr^{30} and Thr^{17} methyl peaks. The side chains of residues 20 and 30 are adjacent to one another, but residue 17 is somewhat remote, and its chemical shift change must mirror a change in the conformation of the reactive site loop. A similar change in the chemical shift of the methyl of Thr^{17} accompanies the $Leu^{18} \rightarrow Met^{18}$ substitution (OMTKY3 vs OMSVP3). OMGMQ3, which differs from OMTKY3 by two substitutions, $Leu^1 \rightarrow Phe^1$ and

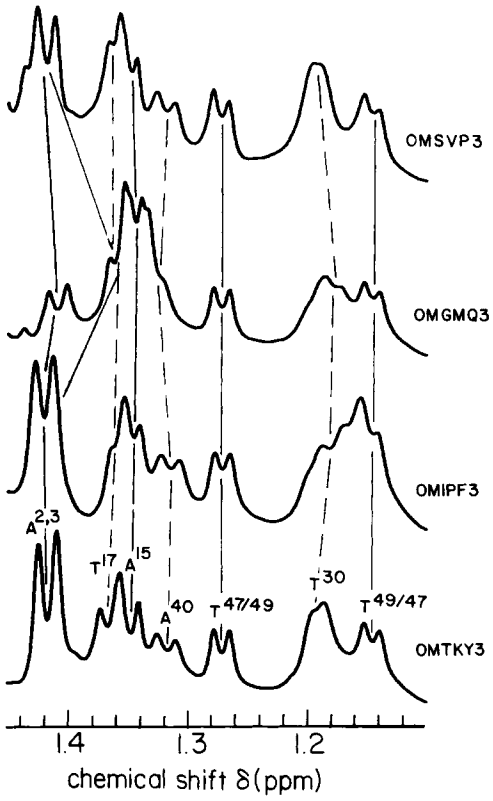

Fig. 4. Examples of spectral differences arising from perturbation of residues neighboring sites of amino acid substitutions in ovomucoid third domains. Shown is the methyl region of 470 MHz ^1H NMR spectra of four variant proteins. Sample conditions were 4 mM protein in 0.2 M KCl in ^2H_2O, pH* 8.0. The NMR probe temperature was 25°C. Spectral acquisition was as described in Figure 3.

$Gly^{32} \rightarrow Ala^{32}$, shows changes in the chemical shifts of peaks assigned to the methyls of Ala^{2}, Ala^{3}, Ala^{40}, and Thr^{17}.

Figure 5 compares the aromatic region of ^{1}H NMR spectra of OMTKY3 and OMIPF3. Recall that these proteins differ by Tyr $^{20} \rightarrow His^{20}$. The difference spectrum clearly shows a positive AA'XX' pattern (pair of doublets) from the ring protons of Tyr^{20} (OMTKY3) and two negative singlets from the ring protons of His^{20} (OMIPF3). The difference spectrum (Fig. 5c) contains further changes that can be explained by a perturbation of the chemical shift of the ring protons of His^{52}. This chemical shift change results from slight differences in the pH of the two samples; these peaks are coincident in the ^{1}H NMR spectrum (not shown) of a 1:1 mixture of the two proteins obtained at this pH. The histidine pK_a' values are obtained from ^{1}H NMR titration curves (Fig. 6). The pK_a' of His^{52} is 7.53 ± 0.01* in OMTKY3 and 7.64 ± 0.01 in OMIPF3. The inflection with a pH_{mid} around 2.7 ± 0.1 in the titration curve of His^{52} has been attributed to the proton disssociation of the carboxyl group of the C-terminal Cys^{56} [11]. The results show that the pK_a' of the carboxyl terminus is the same,

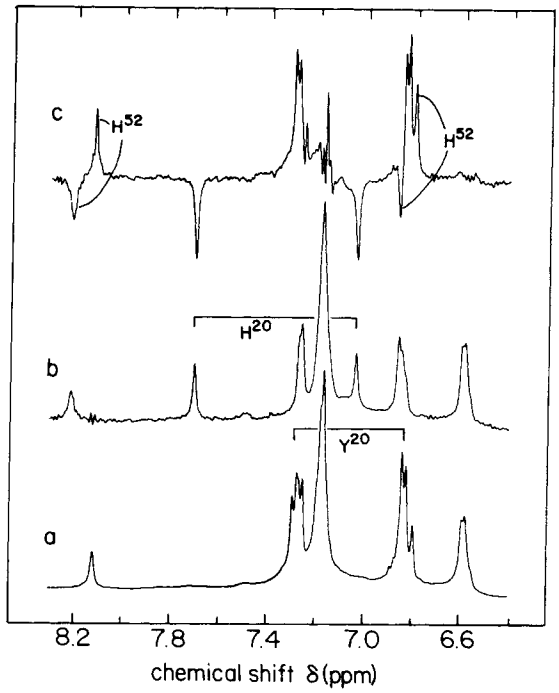

Fig. 5. Comparison of aromatic spectral region of 470 MHz ^{1}H NMR spectra of two ovomucoid third domains: a, turkey (OMTKY3); and b, peafowl (OMIPF3), which differ by the substitution $Tyr^{20} \rightarrow His^{20}$. The difference spectrum (c) reveals shifts in the positions of the ring protons of His^{52}, a residue common to both OMTKY3 and OMIPF3. Sample conditions were 4 mM protein in 0.2 M KCl in $^{2}H_2O$, pH* 8.0, 25°C. Spectral acquisition was as described in Figure 3.

*The errors given represent 1 SD from the nonlinear least squares computer fit to a theoretical titration curve. They do not include other errors, such as those associated with temperature fluctuations or the measurements of pH and chemical shifts. The pK_a' values probably are accurate only to 0.1 pH unit.

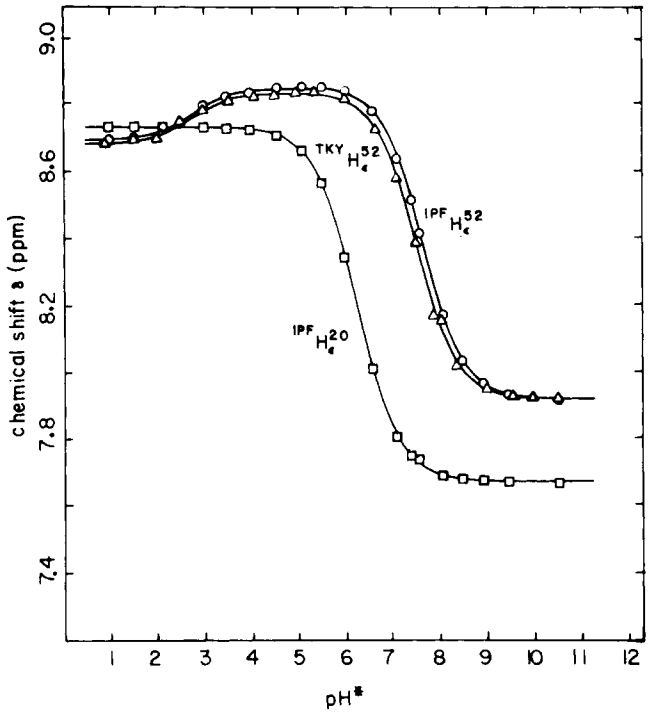

Fig. 6. ^{1}H NMR pH titration curves of the histidine residues of turkey (OMTKY3) and peafowl (OMIPF3) ovomucoid third domains. The pK_a' of His^{52} is the same, within experimental error in the two proteins and is not sensitive to the substitution $Tyr^{20} \rightarrow His^{20}$. In that residues 20 and 52 are remote from one another in the three dimensional structure of the molecule (Fig. 2), the pK_a' shift must have an indirect origin. Experimental conditions are given in the caption to Figure 5.

within experimental error, in both proteins. The additional histidine in OMIPF3 has a pK_a' value of 6.24 $\pm$ 0.01.

We have used ^{13}C NMR to compare the tyrosine titration curves in another pair of proteins differing by a single substitution, OMTKY3 and OMSVP3. The results (Fig. 7) show that the substitution, $Leu^{18} \rightarrow Met^{18}$, at the P1 position does not lead to significant changes in the pK_a' values of Tyr^{11} (9.76 $\pm$ 0.02 in OMTKY3, 9.74 $\pm$ 0.02 in OMSVP3) and Tyr^{20} (10.82 $\pm$ 0.04 in OMTKY3, 10.88 $\pm$ 0.10 in OMSVP3). These results are consistent with the very close agreement between the three-dimensional structures of the two proteins, r.m.s. deviation of only 0.036 nm for all main-chain atoms except those of residues 1–9 [19]. Tyr^{31} has an abnormal pK_a' value, which is well above 11.5. Its NMR signal is perturbed with a pH_{mid} of 2.55 $\pm$ 0.04 in OMTKY3 (Fig. 7), which is attributed to titration of Asp^{27} [23] in agreement with the interpretation of the pH dependence of the optical spectrum [24]. The X-ray structures of OMJPQ3 [21], OMTKY3 [20], and OMSVP3 [19] each show a hydrogen bond between the hydroxyl of Tyr^{31} and the carboxylate of Asp^{27}. One thus predicts that Asp^{27} will have a depressed pK_a' and that Tyr^{31} will have an elevated pK_a', as is confirmed by the NMR results (Fig. 7).

^{13}C NMR is less responsive than ^{1}HNMR to tyrosine side-chain dynamics. Studies of the temperature dependence of the ^{1}H NMR spectrum of OMTKY3 have

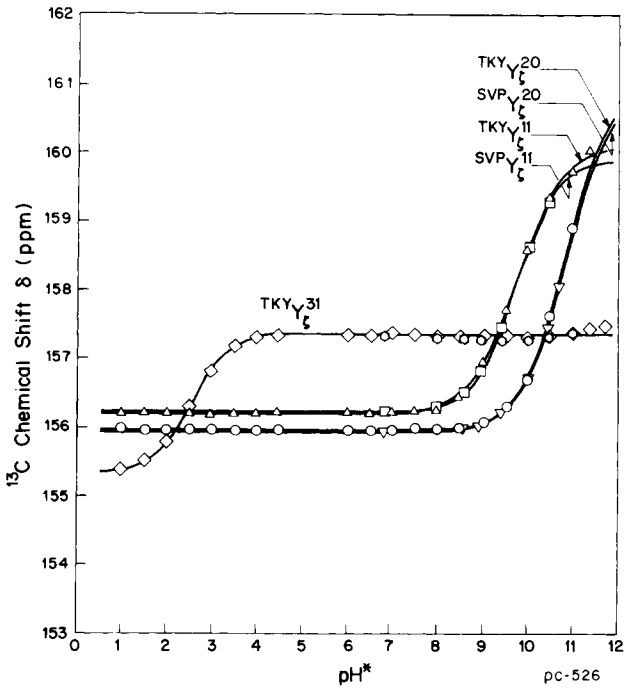

Fig. 7. ^{13}C NMR pH titration curves of the tyrosine residues of OMTKY3 and OMSVP3 based on the chemical shifts of the ring C_ζ. The titration curves reveal that the substitution $Leu^{18} \rightarrow Met^{18}$ has no effect on the pK_a' values of Tyr^{11} and Tyr^{20}. No spectra of OMSVP3 were obtained below pH* 6.85. The samples contained 15 mM protein in 0.2 M KCl in ^2H_2O. The probe temperature was 25°C.

shown that Tyr^{31} flips slowly on the NMR time scale at low temperatures but rapidly at high temperatures (above 30°C). At 25°C, the activation energy for ring flipping is about 13 kcal mol^{-1}. The ring flip rate is also pH-dependent [25]. The line shape of the signal from Tyr^{11} also depends on pH but, more interestingly, on sequence changes. This is illustrated qualitatively by the ^1H NMR titration curves (Fig. 8) obtained for this residue in OMJPQ3, OMSVP3, and OMCHI3. Whereas the signals from Tyr^{11} can be followed between pH 10 and 12 in OMSVP3 and OMCHI3, the corresponding signals in OMJPQ3 are broad and can be resolved only at extremely high pH. This suggests that one of the amino acid replacements that makes OMJPQ3 differ from the others (see Fig. 1) is responsible. It has been suggested [23] that the critical substitution is $Pro^{17} \rightarrow Thr^{17}$, which leads to a loosening of the loop that surrounds Tyr^{11} and allows it to rotate more freely. This interpretation is supported by the recent comparison of the three-dimensional structures of OMJPQ3 and OMSVP3, which suggests that this substitution at residue 17 leads to a minor rearrangement in that loop [19].

Homonuclear Two-Dimensional NMR Spectroscopy

Although considerable information about the structure and properties of ovomucoid third domains can be obtained from 1D NMR spectra, particularly if spectra of single substitution variants are compared, two-dimensional NMR spectroscopy [26] provides a valuable means of spectral simplification and useful approaches to

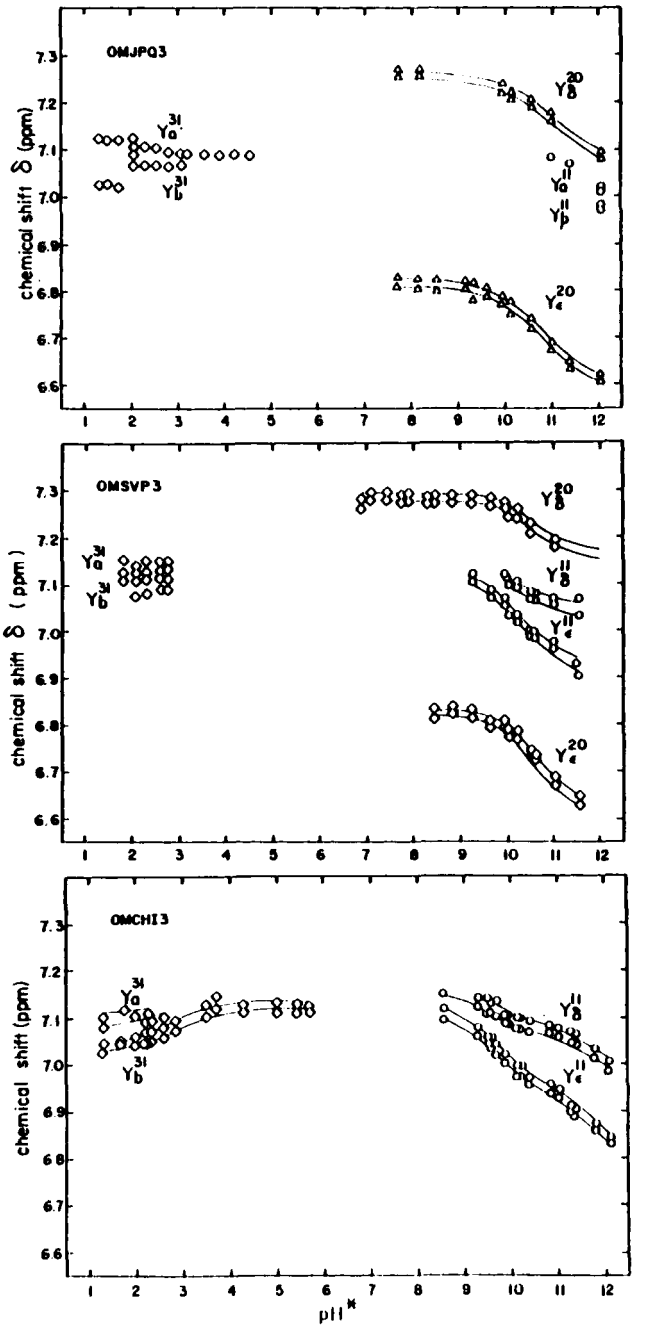

Fig. 8. Titration curves of the tyrosine ring proton resonances of isolated ovomucoid third domains from chicken (OMCHI3), Japanese quail (OMJPQ3), and silver pheasant (OMSVP3). Note that the chicken and Japanese quail ovomucoid third domains used in this study were prepared following limited proteolyis with thermolysin and lack the amino terminal tripeptide LAA shown in Figure 1. The solid lines are computer-generated theoretical titration curves. The data for OMCHI3 and OMJPQ3 were obtained at 360 MHz and the data for OMSVP3 at 470 MHz. The samples contained ~ 2 mM protein and 0.2 M KCl dissolved in ^2H_2O; the NMR probe temperature was 25°C. The chemical shift scale is referenced to DSS.

extending assignments to other parts of the molecule. In addition, two-dimensional nuclear Overhauser effect spectroscopy (NOESY) [27] affords the most efficient means of identifying resonances from pairs of protons that are close to one another in space (eg, within ~ 2–5 Å).

We have used the sequential NMR assignment method developed by Wüthrich and coworkers [28] to assign over 70% of the backbone ^{1}H NMR peaks of OMTKY3. According to this method, the homonuclear (^{1}H) two-dimensional correlated spectrum (COSY) is compared to the NOESY plot. Both 2D spectra are obtained in ^1H_2O solution so that the peptide N–H resonance can be observed. The COSY plot contains off-diagonal peaks that establish the connectivity (via spin-spin coupling) of the C_α–*H* and N–*H* atoms of a given residue. The NOESY plot contains off-diagonal peaks that establish the connectivity (via a dipole-dipole cross relaxation mechanism) between the N–*H* of a given residue and a C–*H* or N–*H* of the preceeding residue in the sequence. By putting this information together, one can trace the connectivity of peaks along the peptide backbone, as is illustrated in Figure 9, which shows how assignments in the reactive site region have been made. The NOE connectivity shown is that between the N–*H* of residue i and the C_α–*H* of residue i–1. One must have a primary assignment to start this spiral of extended assignments. We have obtained such starting points by comparing COSY plots of closely related variants. An example is shown in Figure 10, which compares the C_α–*H*—N–*H* connectivity region of COSY plots from OMFTD3 (fulvous tree duck) and OMTKY3. In the reactive site region, this pair of proteins differs by $Met^{21} \rightarrow Arg^{21}$, and the presence of a peak in the region expected for an arginine residue in the COSY plot of OMTKY3 but not that of OMFTD3 serves to provide the C_α–*H* and N–*H* chemical shifts for Arg^{21}.

Figure 11 shows 1D ^{1}H NMR spectra of four ovomucoid third domains that illustrate sequence-dependent changes in the chemical shift of the C_α–*H* peak of Tyr^{20}. We have interpreted these changes in terms of a steric interaction between the phenyl ring of Tyr^{20} and the side chain at position 32. Evidence for such an interaction comes from the geometry of that region as established by X-ray results [20] and NOESY data (Fig. 12). The chemical shift of the C_α–*H* of Tyr^{20} is abnormally to low field because of the magnetic anisotropy (ring current effect) of its phenyl ring. Introduction of a bulky group at the side-chain position of residue 32 (eg, $Gly^{32} \rightarrow Ala^{32}$ as in OMTKY3 vs OMGMQ3) alters the *mean* tilt angle of the Tyr^{20} ring (the ring executes 180° flips rapidly on the NMR time scale). The chemical shift of the Tyr^{20} C_α–*H* of OMTKY3 is constant in the pH range 5–8 (Fig. 13a). A similar interaction between the C_α–*H* and the imidazole ring of His^{20} is observed in OMIPF3 and explains the abnormal C_α–*H* chemical shift (Fig. 12); moreover, the chemical shift of the C_α–*H* is pH-dependent, with a pH_{mid} that matches the pK_a' determined for His^{20} (Fig. 13b).

Heteronuclear Two-Dimensional NMR Spectroscopy

We have demonstrated recently that heteronuclear 2D NMR techniques can be applied to a small protein such as ovomucoid third domain without recourse to isotopic labeling. The experiments require larger amounts of material than does homonuclear 2D NMR spectroscopy, but, with the higher sensitivity afforded by newer NMR equipment, they should be quite feasible. We have used the $^{13}C\{^1H\}$ correlated experiment (observe carbon, indirectly observe proton) [29] to cross-assign the resonances of directly bonded ^1H–^{13}C pairs. Figure 14A shows the methyl region of the 2D $^{13}C\{^1H\}$ correlated spectrum of OMTKY3 [13]. The heteronuclear 2D FT

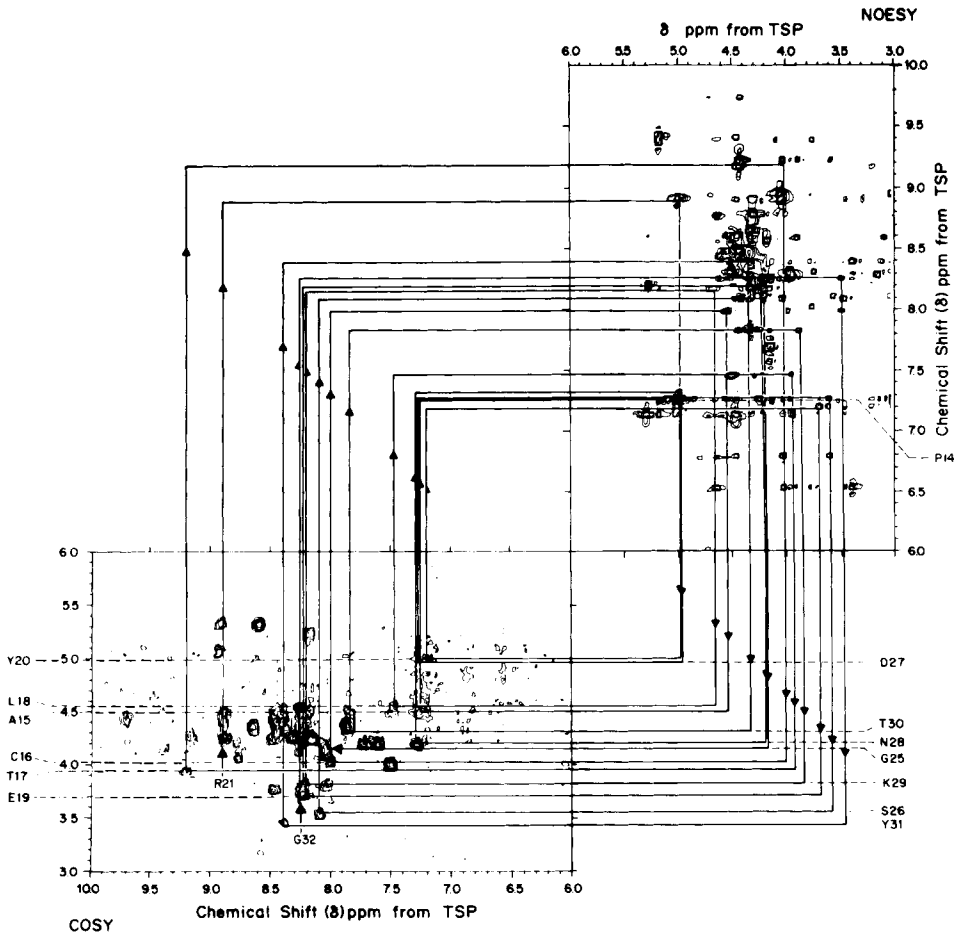

Fig. 9. Spiral of sequential assignments of backbone ^{1}H NMR peaks of turkey ovomucoid third domain (OMTKY3) obtained by the method of Wüthrich et al [28]. The sequential assignments start from the N-*H* of Arg^{21} and extend to the N-*H* of Ala^{15}. The assignment of the N-*H* of Arg^{21} is shown in Figure 10. Assignments of the N-*H* peaks of Tyr^{20} and Ala^{15} were confirmed by spectra of variants with substitutions at these positions. The C_α-*H* of Pro^{14} was identified in the NOESY plot. A second stretch of sequential assignments is indicated from Gly^{32} to Gly^{25}. The COSY data were obtained by using a $90°$-τ-$60°_\phi$ pulse sequence; the phase of ϕ of the second pulse was changed from x to y; the phase of the receiver alternated by 180° between successive accumulations. The delay, τ, was incremented in steps of 100 μsec, 512 such blocks were collected; each block contained the sum of 52 free induction decays (FIDs) accumulated into 1,024 data points. The ^{1}H transmitter was positioned at 10.08 ppm from TSP. The sweep width in both dimensions was equivalent to 5,000 Hz. The delay between successive accumulations was 3 sec. The FIDs were apodized by multiplication with a sine function and zero-filled once in both dimensions, before Fourier transformation, in order to improve the resolution of cross-peakis. The NOESY data were collected by using a $90°_\phi$-τ-$90°$-τ_{mix}-$90°_\phi$ pulse sequence. The phases of the first and third pulses were changed from x to $-$x in successive accumulations. The variable delay, τ, was incremented in steps of 98 μsec, which was equal to the dwell time for digitization, and yielded a sweep width in each dimension of 5,102 Hz. The mixing time, τ_{mix}, was fixed at 100 msec. A recycling time of 3 sec was used; 512 blocks of data were collected; 76 FIDs were accumulated into 1,024 points for each block. The ^{1}H transmitter was positioned at 10.25 ppm from TSP. The FIDs were multiplied by a line-broadening factor to suppress spin-correlated cross-peaks that occur as artifacts. Both the COSY and NOESY data were obtained with a Nicolet NT-470 NMR spectrometer with a Nicolet 1180 computer and 293B pulse programmer. The water signal was irradiated with the decoupler only during the relaxation delay. The sample contained 16 mM protein in 0.2 M KCl in H_2O, pH 6.1. The NMR probe temperature was 25°C.

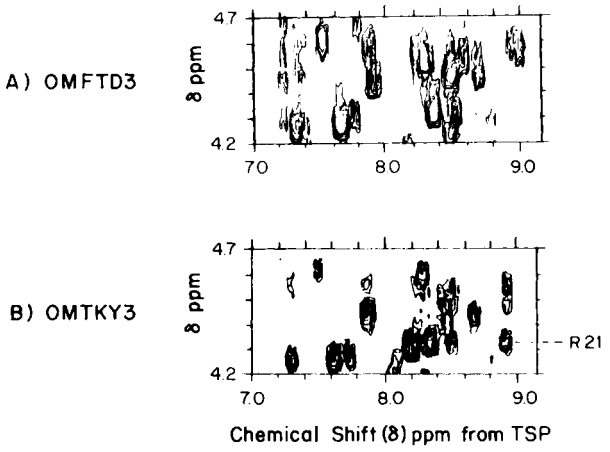

Fig. 10. Assignment of the N-*H* of Arg^{21} by comparison of COSY plots of ovomucoid third domains from turkey (OMTKY3), which has Arg^{21}, and fulvous tree duck (OMFTD3), which has Met^{21}. The COSY data were obtained as described in Figure 9. The samples contained ~ 10 mM protein in 0.2 M KCl in H_2O, pH 6.1. The NMR probe temperature was 25°C.

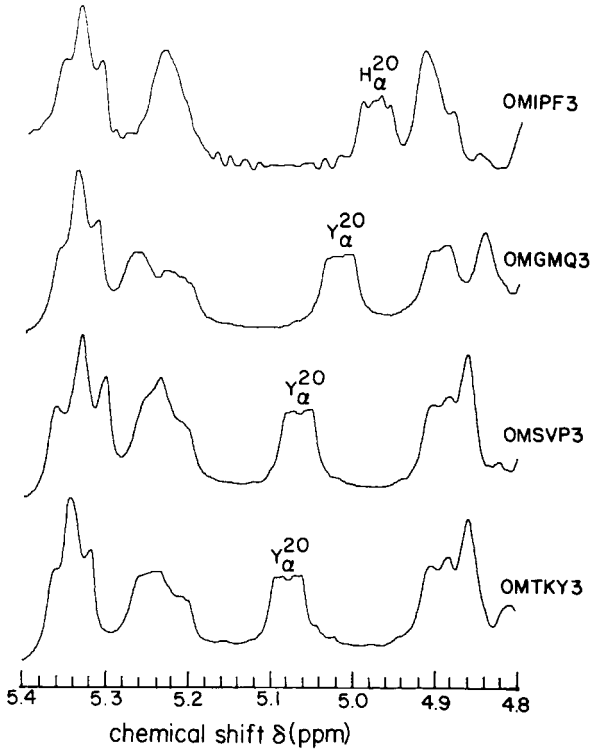

Fig. 11. ^1H NMR spectra (470 MHz) of four ovomucoid third domains illustrating the sequence-dependent chemical shift of the C_α-*H* of Tyr^{20}. The top spectrum shows the position of C_α-*H* of His^{20} in OMIPF3. The samples contained 4 mM protein in 0.2 M KCl in ^2H_2O, pH* 8.0. The NMR probe temperature was 25°C.

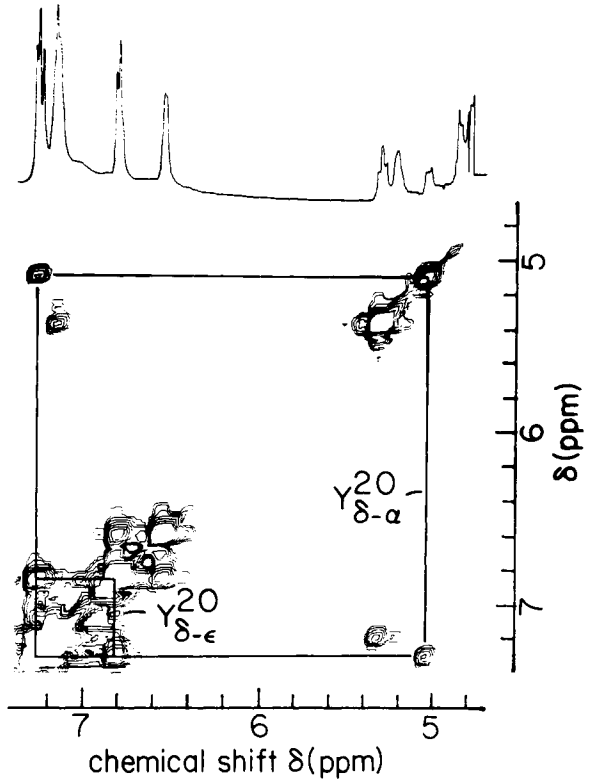

Fig. 12. NOESY plot (470 MHz) showing evidence for the interaction between the C_α-*H* and ring C_δ-*H* of Tyr^{20} in turkey ovomucoid third domain. The mixing time used was 100 msec. For reference, the NOE cross-peak from the adjacent tyrosine ring protons is indicated. The 1D 470 MHz ^1H NMR spectrum of the protein is shown at top. The sample contained 10 mM OMTKY3 dissolved in 0.2 mM KCl in ^2H_2O, pH* 4.6. The probe temperature was 25°C. The NOESY data were collected by using a $90°_\phi$-τ-$90°$-τ_{mix}-$90°_\phi$ pulse sequence. The phases of the first and third pulses were changed from x to −x in successive data acquisitions to remove axial peaks. The mixing time, τ_{mix}, was fixed at 100 msec. The spectral width in each dimension was 5,000 Hz, and the carrier frequency was placed at the low-field end of the spectrum. Each of 512 blocks of 1,024 data points was the sum of 16 transients. The acquisition time was 51.3 msec, and the recycle time was 6 sec. Data processing was as described in Figure 9.

method allows the resolution of numerous methyl peaks that overlap in the 1D ^1H NMR spectrum. Several of the cross-peak assignments shown were based on known proton assignments. Others have been obtained through comparison of $^{13}C\{^1H\}$ correlated spectra of variant ovomucoid third domains (eg, Fig. 14b and c).

Resonances from ^1H–^{15}N groups can be cross-assigned by the method of Bax and coworkers [30] based on connectivities established via multiple quantum (ie, forbidden) transitions. We have used this method to assign resonances in the ^{15}N NMR spectrum of OMTKY3 [31] on the basis of ^1H NMR assignments obtained previously by the sequential assignment method described above.

CONCLUSIONS

NMR spectroscopy can provide a wealth of information about the effects of single-site replacements in proteins. Extraction of this information requires that

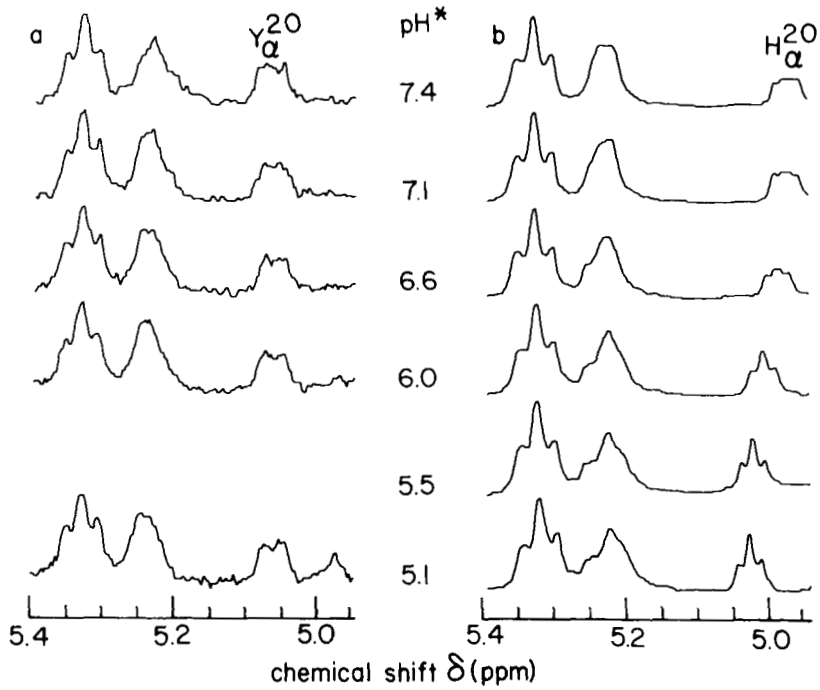

Fig. 13. ^{1}H NMR spectra (470 MHz) illustrating the pH dependence of the C_α-*H* of residue 20 in turkey ovomucoid third domain, which has Tyr^{20} (a), and peafowl ovomucoid third domain, which has His^{20} (b). The samples contained 4 mM protein in 0.2 mM KCl in ^2H_2O. The pH* value is indicated. Probe temperature was 25°C.

individual peaks be resolved and assigned and that the observed site-specific changes in NMR parameters be interpreted in structural or functional terms. 2D NMR methods greatly assist in the tasks of resolution and assignment. Room for improvement remains in the theoretical interpretation of NMR chemical shift values. Changes in chemical shifts with pH can be interpreted more readily, and NMR provides the method of choice for investigating proton dissociations in proteins. NOESY data appear to be promising as a means of establishing the nature of changes in local structure.

The agreement between the X-ray and NMR data on ovomucoid third domains is satisfying. In general, interactions predicted by assigned peaks in NOESY spectra correspond with the X-ray results. NMR spectra may be useful in establishing the solution structure of the amino terminal region of the molecule; the conformation of this region from X-ray data changes from one variant to another as the result of different crystal packing forces.

A general finding of the NMR and X-ray studies of this series of proteins is that the effects of substitutions are generally localized. Such conservation of structure in spite of modification has been observed in a number of the systems. The strategy proposed for obtaining a sequence-to-function algorithm [18] relies on the independence of substitutions at different sites; thus this property of the system should make such predictions more reliable. The independence of different sites also simplifies the

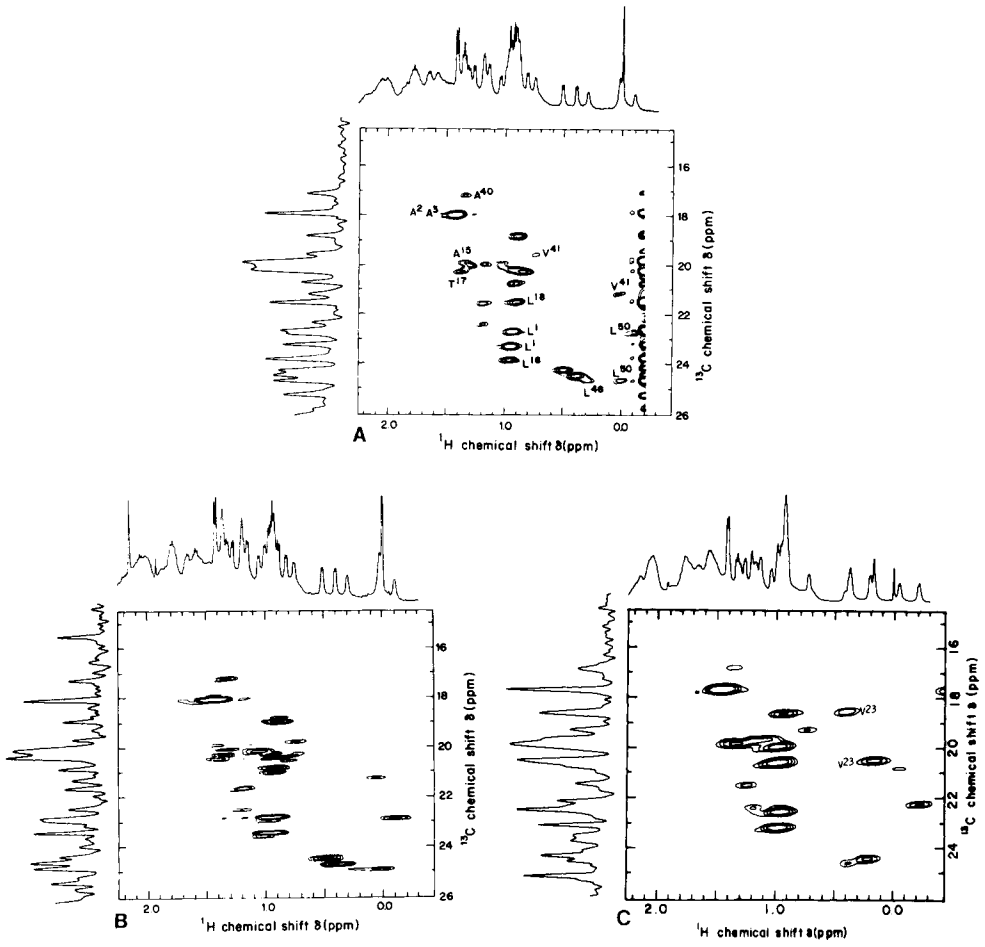

Fig. 14. Natural abundance 2D Fourier transform $^{13}C\{^1H\}$ chemical shift correlated spectra, obtained with a Nicolet NT-200 spectrometer, of the methyl region of ovomucoid third domain from turkey (A), silver pheasant (B), and Japanese quail (C). Along the left side of each contour plot is the corresponding 50.3 MHz ^{13}C NMR spectrum, and along the top is the corresponding 470 MHz ^1H NMR spectrum. The samples were dissolved in 2.5 ml of 0.2 M KCl in ^2H_2O to give final protein concentrations of 12 mM, OMTKY3; 15 mM, OMSVP3; 14 mM, OMJPQ3. The pH* of each sample was 8.0, and the probe temperature was approximately 23°C. Details of the experimental procedure used were as given in reference 13 except that 32 blocks (rather than 64 blocks) were collected for spectra of OMSVP3 (B) and OMJPQ3 (C). Contour peaks are labeled in A to show the ^1H–^{13}C cross-assignments for the methyl groups of several amino acid residues in the proteins.The expected cross peak in B from the S–CH_3 of Met[18] of OMSVP3 is not observed in the figure, but it shows up if a longer experimental delay time is used to prevent saturation. Comparison of contour plots A and B confirms the assignment of cross-peaks to the methyls of Leu[18] in OMTKY3 (OMSVP3 has Met[18]). Comparison of contour plots A and C reveals the presence of additional cross-peaks in the spectrum of OMJPQ3 that are assigned to the methyls of Val[23] (OMTKY has Leu[23]) and confirms the assignment of the methyl of Thr[17] in spectrum A of OMTKY3 (OMJPQ has Pro[17]).

NMR assignment strategy, since difference spectra of single-site variants will reveal peaks only from the replaced residues. In fact, we have found it possible to use pairs of ovomucoid third domains differing by three or four replacements for assignment purposes; simultaneous assignments to all the replaced residues speeds up the process. In that all the OMXXX3 variants studied to date represent naturally occurring proteins, the conservation of structure may be the result of selective pressure to produce stable and efficient proteinase inhibitors.

As was mentioned above, the X-ray and NMR results do point to minor exceptions to this basic independence of sites, and it will be interesting to find out whether more such examples will emerge from studies of additional variants. Modification of some of the internal residues that are constant in all known OMXXX3 sequences can be expected to lead to global structural arrangements. Such studies must await advances in expression of the cloned gene or chemical synthesis of ovomucoid third domain peptides. In this connection, a single residue substitution in staphylococcal nuclease ($Phe^{76} \rightarrow Val^{76}$) should be mentioned; this replacement causes a 10°C decrease in the thermal stability of the protein [6] and a radical change in the ^1H NMR spectrum [32]. Here the changes in the NMR spectrum are so extensive that it might be necessary to assign the spectrum of each variant independently. Other substitutions in the same molecule, however, for example, $His^{124} \rightarrow Leu^{124}$ [8], produce localized perturbations.

ACKNOWLEDGMENTS

These studies were supported by NIH grants GM19907 and RR01077 (J.L.M.) and GM10831 (M.L.). The authors thank Carlton Schlicher and Bob Santini for able assistance with NMR instrumentation and Jim Cook for amino acid analyses. NMR spectra were obtained at the Purdue University Biochemical Magnetic Resonance Laboratory, which is supported by the Biomedical Research Technology Program Branch of the Division of Research Resources, National Institutes of Health.

REFERENCES

1. Markley JL, Ulrich EL: Annu Rev Biophys Bioeng 13:493, 1984.
2. Braun W, Bösch C, Brown L, Gō N, Wüthrich K: Biochim Biophys Acta 557:135, 1981.
3. Markley JL: Acc Chem Res 8:70, 1975.
4. Rhyu GI, Ray WJ Jr, Markley JL: Biochemistry 24:4746, 1985.
5. Wagner G: Quart Rev Biophys 16:1, 1983.
6. Shortle D: J Cell Biochem (in press).
7. Empie M, Laskowski M Jr: Biochemistry 21:2274, 1982.
8. Jardetzky O, Markley JL: Il Farmaco Ed Sci 25:894, 1970.
9. Kato I, Kohr WJ, Laskowski M Jr: In Magnusson S, Ottensen M, Foltmann B, Danø K, Neurath H (eds): "Regulatory Proteolytic Enzymes and Their Inhibitors." Oxford, Pergamon, 1978, vol 47, pp. 197–206.
10. Laskowski M Jr, Kato I: Annu Rev Biochem 49:593, 1980.
11. Ogino T, Croll DH, Kato I, Markley JL: Biochemistry 21:3452, 1982.
12. Markley JL, Westler WM, Chan T-M, Kojiro CL, Ulrich EL: Fed Proc 43:2648, 1984.
13. Westler WM, Ortiz-Polo G, Markley JL: J Magnet Reson 58:354, 1984.
14. Wu X, Westler WM, Markley JL: J Magnet Reson 59:524, 1984.
15. Kato I, Ardelt W, Cook J, Denton A, Empie MW, Kohr WJ, Park SJ, Parks K, Schatzley BL, Schoenberger OL, Tashiro M, Vichot G, Wieczorek A, Wieczorek M, Whatley HE, Laskowski M Jr: (Submitted).

16. Bogard WC Jr, Kato I, Laskowski M Jr: J Biol Chem 255, 1980.
17. Laskowski M Jr, Empie MW, Kato I, Kohr WJ, Ardelt W, Bogard WC Jr, Weber E, Papamokos E, Bode W, Huber R: In Eggerer H, Huber R (eds): "Structural and Functional Aspects of Enzyme Catalysis." Berlin, Heidelberg: Springer Verlag, 1981, Vol 32, pp 136–152.
18. Laskowski M Jr: Biochem Pharmacol 29:2089, 1980.
19. Bode W, Epp O, Huber R, Laskowski M Jr, Ardelt W: Eur J Biochem 147:387, 1985.
20. Fujinaga M, Read RJ, Sielecki A, Ardelt W, Laskowski M Jr, James MNG: Proc Natl Acad Sci USA 79:4868, 1982.
21. Papamokos E, Weber E, Bode W, Huber R, Empie MW, Kato I, Laskowski M Jr: J Mol Biol 158:515, 1982.
22. Lineweaver H, Murray CW: J Biol Chem 171:565, 1947.
23. Croll DH: PhD thesis, Purdue University, 1982.
24. March CJ: PhD thesis, Purdue University, 1980.
25. Westler WM, Markley JL: (Unpublished data).
26. Bax A: "Two Dimensional Nuclear Magnetic Resonance in Liquids." Amsterdam: Reidel, 1982.
27. Kumar A, Ernst RR, Wüthrich K: Biochem Biophys Res Commun 95:1, 1980.
28. Wüthrich K, Wider G, Wagner G, Braun W: J Mol Biol 155:311, 1982.
29. Maudsley AA, Ernst RR: Chem Phys Lett 50:368, 1977.
30. Bax A, Griffey RH, Hawkins BL: J Magnet Reson 55:301, 1983.
31. Ortiz-Polo G, Krishnamoorthi R, Markley JL, Live DH, Davis DG, Cowburn D: J Magnet Reson (in press).
32. Ulrich EL, Shortle D, Markley JL: (Unpublished data).