9

Human Chorionic Gonadotropin: Progress in Determining Its Tertiary Structure

JOYCE W. LUSTBADER, CHUAN WANG, XIAOLU ZHANG, STEVEN BIRKEN, HAO WU, JONATHAN MILES BROWN, DAVID YARMUSH, SUSAN POLLAK, AND ROBERT E. CANFIELD

In humans, maintenance of pregnancy requires extending the lifetime of the corpus luteum, which produces essential steroids. The signal for this maintenance is provided by *human chorionic gonadotropin* (hCG), a dimeric glycoprotein hormone produced by the trophoblast cells of the early embryo. The absence of hCG would result in rapid termination of early pregnancy.

Solution of the three-dimensional structure of hCG, the hormone of pregnancy, could have a profound impact on society. Uncontrolled population growth is a major obstacle to improving the quality of life in many nations. One potential route to controlling population growth is the use of antagonists to hCG, since pregnancy cannot be sustained in its absence. Furthermore, solution of the structure of hCG would also permit a reasonable prediction of the structure of follicle stimulating hormone, the homologous pituitary hormone controlling ovulation.

Understanding the action of hCG may also be important in alleviating infertility due to short or poor luteal phase, the condition in which hCG secretion is insufficient to maintain adequate formation of progestational compounds. Development of hCG agonists may improve therapeutic prospects for patients with this form of infertility. A rational approach to the design of both agonists and antagonists of hCG presupposes knowledge of the hormone's three-dimensional structure.

For more than twenty years laboratory groups have been attempting to determine the three-dimensional structure of the glycoprotein hormones, the family to which hCG belongs. Knowledge of the tertiary structure of the hormones would aid in understanding their structure-function re-

lationships. Detailed knowledge of a given glycoprotein hormone's three-dimensional structure would allow the chemist to identify a specific amino acid's contribution to the function of the hormone and then to change those properties through site-specific mutagenesis. The chemist could also, for example, "build a better hormone" by synthesizing molecules that had higher affinities for the receptor than the native hormone. Alternatively, a synthetic peptide or small molecule could be developed that would block the binding of the hormone to the receptor. To make such compounds requires an intimate knowledge of the hormone's surface, and that understanding cannot be derived from knowledge of the primary amino acid structure of the protein. With regard to hCG in particular, two groups, one of which is ours, obtained large diffractable crystals of hCG, and yet little new structural information has been presented since those initial reports (1, 2).

This chapter describes the progress in efforts to determine the structure of hCG. hCG is one of four glycoprotein hormones, and the only one secreted by the trophoblast. The other three glycoproteins secreted by the pituitary are *follicle stimulating hormone* (FSH), *thyroid stimulating hormone* (TSH), and *luteinizing hormone* (LH). The hormones are dimeric in structure, and within a species they share an identical α-subunit, with the biological specificity of the hormone conferred by the β-subunit; the primary action of hCG is to signal the corpus luteum of pregnancy to maintain steroid production. The intact hormone's subunits are held together by noncovalent interactions. Dissociation of the dimer into its monomeric subunits can be achieved by adding chaotropic reagents or acid pH. Both subunits possess a large number of disulfide bonds, and this greatly restricts the number of different conformations the hormone can assume. The degree of homology in the sequences in the α-subunit from different species can be determined by aligning the positions of the cysteines; the α has 10 cysteines (all conserved in species that have been protein-sequenced to date), resulting in 5 disulfide bridges in the folded subunit. Similarly the β-subunits can be aligned by their 12 cysteines, resulting in 6 disulfide bridges. Since more than 70% of the β-subunit associated with any given glycoprotein hormone is homologous with the β-subunits of the other members of the family, one could reasonably extrapolate from knowledge of the structure of one subunit to the structures of the other subunits. Since CG and its most homologous analog LH, both bind to the same receptor, indicating that their conformations are similar, either one could have been chosen for the studies to be described.

A three-dimensional structure of hCG can be generated in one of three ways: (i) model building, (ii) crystallography, and (iii) *nuclear magnetic resonance* (NMR) imaging. The first technique is an indirect method that capitalizes on other known structures that have significant sequence

homologies to the protein under study. The latter two methods rely on direct physical data.

Model Building

Obstacles to hCG model building are threefold: (i) No sequence homology to any known structure has been identified, (ii) little in the way of interatomic distance data is available, and (iii) the disulfide bridge pairings are still unresolved. Therefore, traditional model building, primarily based on known homologous structures, cannot be accomplished until the structure of at least one member of the glycoprotein hormones is solved.

Unfortunately, a method to derive the three-dimensional structure of a protein from its primary amino acid sequence has not yet been determined, although testable hypothetical models can be generated from the many computer-based modeling programs. Such models are probably incorrect in their detail, but they aid the investigator in visualizing the surface topology of the protein. Molecular biology experiments aimed at defining epitopes, biologic efficacy, and dimeric contacts have greatly increased our knowledge of the surface residues, providing additional inputs to a modeling program. Because of the considerable degree of homology between the α- and β-subunits from different species, initial modeling paradigms relied heavily on the placement of conserved regions. Regions of conserved sequences probably exist internalized in the center of the molecule and may not participate directly in the function of the hormone, but may aid in the proper folding of the molecule and, consequently, be crucial in relation to function. Those positions in the sequences with variable amino acids are probably located on the surface and are subject to a greater degree of evolutionary change.

For the studies to be discussed in this chapter, the modeling program PAKGGRAF was used. A brief summary of the modeling studies is presented here, while an extensive report appears elsewhere (3). The precursor of PAKGGRAF was originally developed by Levinthal more than thirty years ago (4). It has been continually revised over the ensuing years. Its advantage over other modeling programs is that it performs both torsional and Cartesian minimization operations and can alternate between the two interactively. This allows the researcher to start with unrealistic coordinates and refine the structure by moving atoms over several angstroms. A number of models have been presented by other groups (5, 6). These models, although informative, do not assign coordinates to atoms and therefore cannot be stringently compared. The model presented in this paper relies heavily on the Mise and Bahl assignments for the disulfide pairings (7, 8). Other experimental data incorporated into this model were derived from the reactivities of glycoprotein

hormones of various species with a selected panel of monoclonal antibodies (3). These comparisons allowed for the placement of certain amino acids on the surface of the hormone.

α-Subunit

The α-subunit appears to have a structure with three dominant loops. As has been indicated by *circular dichroism* (CD) spectra analyses, the α-subunit is predominantly β sheet with little α helix. Figure 9.1 depicts the α-subunit indicating the loops: loop 1, residues 7–31; loop 2, residues 32–59; and loop 3, residues 60–84. The amino terminal of the first loop is exposed in the dimer and consequently can be cleaved by enzymes. The remainder of the first loop is hidden in the dimer. Residues 27–30 of this loop are likely to be the nucleating site for the folding of this subunit. The greatest divergence in sequences between species appears in the first part of loop 1; this observation is consistent with the previously reported data indicating that this part of the molecule is not involved in the subunit interface. The second loop is more strongly conserved, except at four positions, 45, 48, 54, and 57. The data suggest that the region preceding Asn 52 is involved in subunit association; that Tyr 37 is not iodinated in the dimer; that glycosylation of Thr 39 prevents subunit association; and that Lys 45 can be cross-linked to Asn 111 on β. Evidence from site-specific mutagenesis studies indicates that a mutation or a deletion of residues 38, 39, or 40 reduces or destroys the ability of α to combine with β. The amino terminal of the third loop is also involved in the subunit

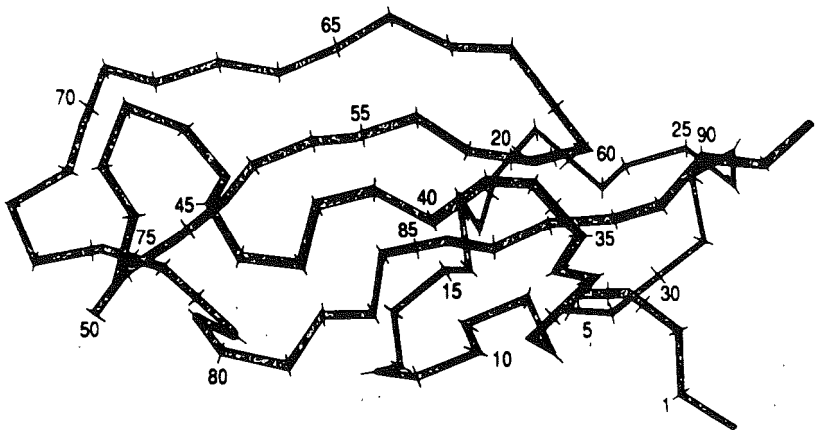

FIGURE 9.1. The α-carbon backbone of the α-subunit of hCG with every fifth residue labeled. Reprinted with permission from Lustbader, Yarmush, Birken, Puett, and Canfield, © The Endocrine Society, 1993.

interface (Fig. 9.2). The carboxyterminal region (73–92) of this loop is antigenic and may play a role in receptor binding.

β-Subunit

Our model includes residues 1–112 of the β-subunit of hCG only, omitting the carboxyterminal peptide, which includes residues 113–145. The β-subunit of our proposed model contains five antiparallel strands of β sheet, with turns proximal to residues 20, 38, 58, 70, 83, 90, and 96 (Fig. 9.3). The CD spectrum of hCGβ depicts a preponderance of β structure with only limited, if any, helicity (3). Puett and Birken demonstrated that reduced, carboxymethylated hCGβ and some tryptic fragments become helical in aqueous solution and that the helix index increases in the presence of trifluoroethanol (9).

The initial portion of the first cysteine loop (9–90) projects away from the subunit interface, influenced probably by the large bulky carbohydrate group at Asn 13. Additionally, residues 8–10 have been described as part

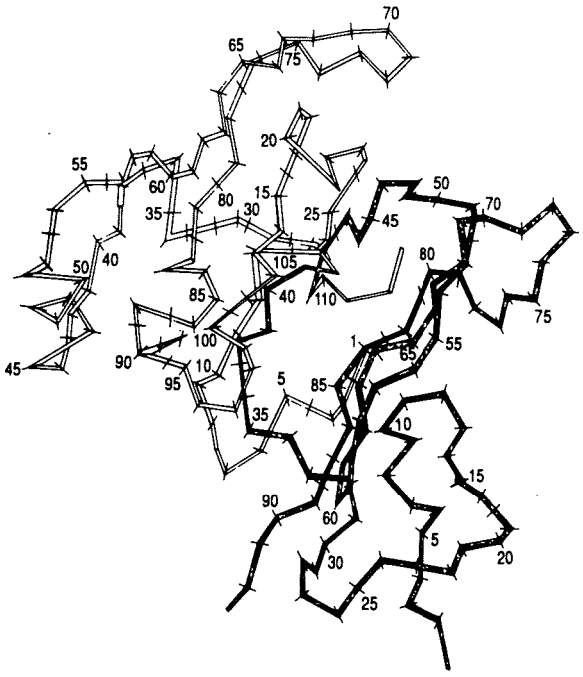

FIGURE 9.2. The α-carbon backbone placement of the α-subunit (shaded) of hCG relative to the β-subunit (residues 1–112) (outlined) with every fifth residue labeled. Reprinted with permission from Lustbader, Yarmush, Birken, Puett, and Canfield, © The Endocrine Society, 1993.

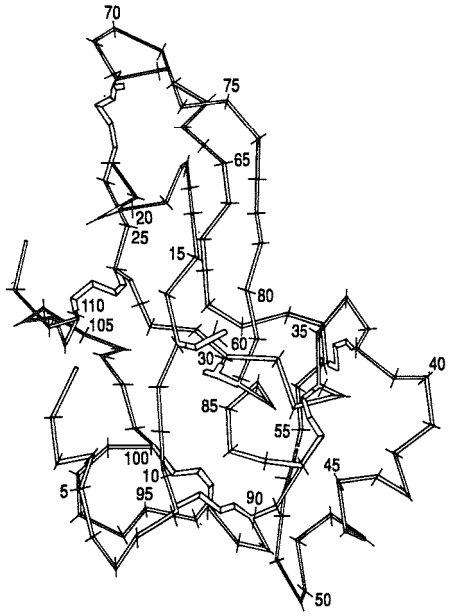

FIGURE 9.3. The α-carbon backbone of the β-subunit of hCG with every fifth residue labeled. Reprinted with permission from Lustbader, Yarmush, Birken, Puett, and Canfield, © The Endocrine Society, 1993.

of the B204 epitope. An area of high sequence conservation is proximal to Glu 19. A mutation of Lys 20 decreases receptor binding significantly, indicating that this region has a role in the receptor interface. Because of carbohydrate at Asn 30, the region between Cys 26 and Cys 34 is likely to project away from any contact. The CAGY region is highly conserved among species and is therefore likely to be involved in receptor or subunit interaction. The conserved region around Asp 111 is likely to be located near the subunit contact since this residue can be cross-linked to the α-subunit (Fig. 9.2).

X-Ray Crystallography

Crystallography is so far the most powerful technique for determining structures of proteins of moderate molecular weight. Since the solution of the hemoglobin structure, developments in methodology and computers have brought about an exponential increase in the number of solved three-dimensional protein structures. The first step toward a structure determination is the purification and crystallization of the target protein.

The history of the successful crystallizations of hCG is a story of international cooperation. Although our laboratory has worked with hCG for over two decades and has prepared all of the reference preparations distributed by the NIH, we had never seriously pursued crystal growth in our own laboratory. We lacked a crystallographer and had been discouraged in such attempts by the numerous anecdotal tales from experts in the field who had spent a great deal of time in fruitless efforts to obtain crystals. These stories had discouraged many investigators from such efforts, and it was widely accepted that hCG and the other glycoprotein hormones could not be crystallized into useful, diffractable crystals by current technology and that the solution of the structures of these hormones would probably not be by crystal diffraction.

In 1982 a crystallographer joined the group (J.W.L.) to pursue a postdoctoral project of bacterial expression of hCG under the tutelage of James Roberts, who was then at Columbia University. Since bacteria do not glycosylate their proteins, this was an avenue of obtaining a naturally carbohydrate-free form of hCG to be used for crystallographic studies. It was known that carbohydrate interfered with crystal growth, and this seemed to be a reasonable study. Unfortunately, we could never obtain intact hCG from the bacteria and only observed some denatured material. The sugar moieties must somehow assist in folding of the molecule. It was at this time that we embarked on a collaboration with Integrated Genetics, who had also tried bacterial expression with hCG and had the same result, but had gone on to express intact hCG in Cos cells. This was the first demonstration of an expression of a biologically active dimeric hormone in recombinant cells (10). These studies also led to an expression system in C127 cells that could produce significant quantities of this recombinant hCG that could be used for crystallographic experiments (11).

Late in 1988 our lab was visited by Francis Morgan from St. Vincent's Institute of Medical Research in Melbourne, Australia. Dr. Morgan was a long-time friend of the laboratory, having spent several years working with us (R.E.C.) in the early 70s on the primary structure of hCG and, in fact, was the postdoctoral mentor of one of us (S.B.). Dr. Morgan indicated that there was some successful crystallization work with hCG by a member of his laboratory, Dr. Neil Isaacs, and a student, D. Harris. Recognizing this as a very major accomplishment, we sought and received approval from the NIH to send Dr. Isaacs several hundred milligrams of the most recently purified hCG reference preparation to facilitate his progress.

In early 1989 Dr. Isaacs sent us a preprint of his manuscript describing the hCG crystallization, which soon appeared in the *Journal of Biological Chemistry*. This manuscript seemed quite remarkable since the hCG employed was the Organon material, which is only 1/3 hCG by weight. This crude material was simply treated with *hydrogen fluoride* (HF) to

remove most of the sugar groups and then gel filtered. The hCG crystals appeared to have grown from material that may have been 75% or less of HF-treated hCG in the mother liquor. No characterization of the content of the crystals was presented. It was not clear that these crystals were composed of hCG. However, future studies proved that they were pure hCG. Fortuitously, at that time, the crystallographer (J.W.L.) in our laboratory was able to pursue crystallization work in our own facility. We proceeded to repeat Neil Isaacs's studies, but with purified hCG in our own laboratories. Nick Pileggi in the University's Protein Core laboratory was quite adept at the HF deglycosylation procedure, and we tailored as careful a treatment scheme as possible. We also made modifications in an attempt to treat the protein in as gentle a manner as possible after HF deglycosylation, such as adding a high-strength neutral Tris buffer instead of sodium hydroxide to neutralize residual HF. We also did a very careful size selection of the molecules for the crystallization work by running an SDS-gel profile on each fraction from a gel-permeation column separation of the HF-treated hCG. Crystals of hCG were successfully grown and completely characterized.

Through the efforts of Mary Ann Gawinowicz, it was found that the HF-treated hCG seemed to accumulate peptide bond cleavages during the long period of crystal growth in acid pH. Since examination of the SDS-gel patterns of HF-treated hCG showed that the heterogeneity of the molecules had increased, not decreased, after HF treatment and that the total carbohydrate content was only reduced by about half, we reasoned that simple removal of the negatively charged sugars may provide the only necessary pretreatment needed to allow the formation of orderly crystals. This proved to be the case, and neuraminidase-treated hCG crystals diffracted better than HF-treated materials. They were found to undergo less peptide bond cleavage during crystal growth, indicating less damage from the HF-treatment. In the light of Neil Isaacs's findings that hCG crystallized from impure mixtures, we also investigated whether the crystals selected particular species of hCG molecules and found that this was not the case and that the crystals reflected the content of the mother liquor. In fact, the crystals seemed to be even more heterogeneous than the mother liquor because of accumulated additional peptide bond cleavages over time.

We have crystallized urinary HF-hCG, urinary asialo-hCG, and r-HF-hCG (Fig. 9.4) using the hanging drop vapor diffusion method (2) with a range of ammonium sulfate at pH 4.2. Whereas large crystals of HF-hCG tended to form within 4 to 6 weeks, the neuraminidase-treated hCG required 8 to 12 weeks to grow (2). The crystals are morphologically identical and appear as stout bipyramids whose faces are usually somewhat rounded. These crystals typically reach a size of $0.4 \times 0.4 \times 0.6$ mm. The HF-treated hCG crystals diffract to a limit spacing of 2.8 Å, whereas the asialo-hCG crystals diffract to a slightly better resolution

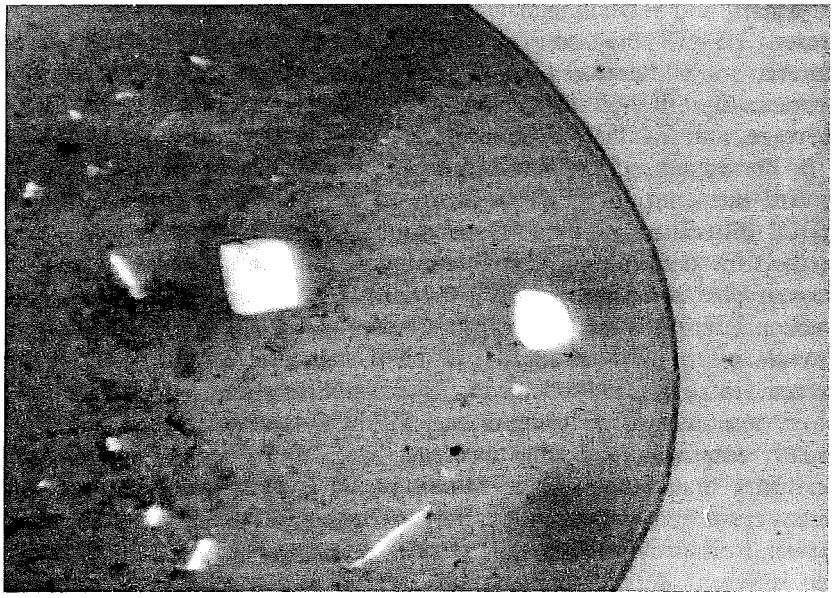

FIGURE 9.4. Hydrogen-fluoride treated recombinant hCG (HF-r-hCG) crystals. The crystals appear as stout bipyramids.

of 2.6 Å. Crystals from neuraminidase-treated hCG, however, have not been regularly reproducible. Each of these crystals possesses the prerequisites for a successful structural determination at intermediate resolution.

Diffraction patterns from these crystals could be indexed in a hexagonal lattice that belongs to space group $P6_122$, or its enantiomorph $P6_522$ (2). Unit cell dimensions are roughly a = b = 88 Å and c = 177 Å, with slight variation among the three forms of hCG. The solvent content of the crystal is estimated to be 59%, assuming one protein molecule per crystallographic asymmetric unit. Most protein crystals have solvent contents of between 27% and 65%, with 43% occurring most frequently (12). The patterns of diffracted intensities from the HF-hCG crystals and the asialo-hCG crystals are not visually distinguishable, suggesting that additional sugars removed by the HF-treatment might be rather disordered in the neuraminidase-treated hCG crystals (2).

Chemical characterization of the various forms of hCG showed that HF treatment increased size heterogeneity of the β-subunit, which was augmented even further by the crystallization process (2). Although the starting HF-hCG contained 1%–3% of β-subunit beginning at Leu 45, N-terminal sequencing revealed the presence of peptide bond cleavages at the acid-labile Asp-Pro (112–113) position and to a lesser degree at Asp-Ser (117–118) in the HF-hCG crystals (2). Although the resolution limits

of urinary and recombinant hCG crystals appear to be similar, the expressed HF-hCG had considerably less nicking than the urinary source material and is therefore a better starting material for crystallization. The neuraminidase-treated hCG starting material or crystal had the least cleavages and was the most homogeneous of the three forms of hCG (2).

In the meantime, we have continued to collaborate with Dr. Isaacs toward the solution of the structure of urinary forms of hCG by the crystal diffraction method. Conventionally, the most commonly used method for determining a nonhomologous protein structure is *multiple isomorphous replacement* (MIR) (13). In this method, differences in diffraction intensities due to incorporation of heavy atoms into the crystals provide information about the detailed structural arrangement of the protein. In collaboration with Dr. Isaacs and his group, we are searching heavy atom derivatives using urinary HF-hCG crystals. Such searches are usually empirical, and some proteins turn out to be more difficult to derivatize than others. Heavy atom-labeled crystals may also have different crystal parameters, since they are nonisomorphous with the native crystal. Both problems necessitate trials of various heavy atom compounds and different derivatizing methods. We (J.W.L.) have provided over 50 high-quality crystals to Dr. Isaacs for the heavy atom replacement procedure. During this period of time we learned of the importance of very careful HF treatment if crystals were to be obtained. Dr. Isaacs informed us of problems that he had with a number of batches of the purified hCG (sent to him earlier), which did not crystallize well presumably due to the harshness of the HF treatment or problems in the selection of the proper fractions to pool for crystallization. We also ran into problems in the capability of different batches of neuraminidase-treated hCG to crystallize. This may have been due to protease contamination. In any event, the effort to solve the structure of urinary hCG remains in the hands of Dr. Isaacs by collaborative agreement.

We have pursued a parallel, but different, approach, since we do not know if it will prove possible to obtain suitable heavy atom replacements for the urinary HF-treated hCG crystals. This work is being pursued collaboratively with Wayne Hendrickson and is dependent on use of recombinant expressed hCG. These studies generally require incorporation of selenomethionine into hCG and are described below.

We obtained high levels of expression of *recombinant hCG* (rhCG) from *Chinese hamster ovary* (CHO) cells subcloned from cells provided by Irving Boime. We have also formed a collaboration with Ares Advanced Technology (a division of Serono) for high-level expression of hCG. Serono had acquired the glycoprotein hormone section of Integrated Genetics, which first expressed hCG and with whom we had worked earlier. A method developed by Dr. Wayne Hendrickson, *multiwavelength anomalous diffraction* (MAD) (14, 15), has resulted in many successful applications (16–20). With the presence of strong diffractors, in this

case selenium atoms introduced by substituting methionines in hCG with selenomethionines, diffraction intensities from a single-type selenomethionine-containing crystal would vary between two sets of otherwise equivalent reflections (Bijvoet difference[1]) and wavelength (dispersive difference[2]). This phenomenon is a result of the so-called anomalous scattering, which is inappreciable for lighter atoms in proteins at typically achievable wavelengths. These variations provide phasing information of the structure. Since only one type of crystal is required, there is perfect isomorphism and no derivatization is required. MAD experiments require a wavelength-tunable synchrotron radiation source. The strong synchrotron radiation should provide a more accurate measurement of diffraction data, and hence, generally small signals from different wavelengths could be extracted. There are a total of four methionines in the primary sequence of hCG—that is, one methionine in every 59 residues. Assuming 100% of the methionines are substituted by selenomethionines, the phasing power is 7.8% from Bijvoet difference and 4.7% from dispersive difference. Previous studies have shown that signals of this magnitude should be ample for a successful structure determination by MAD phasing (21). Theoretically, phasing information from MIR and MAD experiments are complementary to each other and thus can be combined in the structure solution.

The difficulty associated with MAD is obtaining sufficient quantities of hCG with a high incorporation level of selenomethionine. Although some *Escherichia coli* strains are able to grow for 100 generations (22) in selenomethionine, the slight difference in chemistry and metabolism between selenomethionine and methionine makes it toxic to cells, especially mammalian cells as a result of their greater complexity. The starting cell culture must grow in methionine-containing medium since the cells do not tolerate selenomethionine at this stage (21). The expression of selenomethionine protein therefore mandates careful arrangement of timing and concentration levels of the selenomethionine for the preincubation and incubation period to replace the native methionine. We have recently successfully expressed and purified sufficient quantities of selenomethionine HF-rhCG. Amino acid analysis and mass spectroscopy showed complete selenomethionine substitution within the limits of the sensitivity of our analysis.

Small crystals have been obtained from these selenomethionine HF-hCG preparations using the same crystallization condition that has produced native crystals. In many cases, selenomethionine proteins crystal-

[1] Bijvoet difference is the difference in structure factor amplitudes of Friedel pairs (structure factor hkl and hkl).
[2] Dispersive difference is the difference in structure factor amplitudes at different wavelengths.

lize isomorphously under conditions similar to those in which the native materials crystallize (21). Small adjustments to the crystallization conditions are often necessary to allow for possible differences in solubility and hydrophobicity. Here, the selenomethionine rhCG appears to differ from the wild-type rhCG to a somewhat greater extent, and this may necessitate a search for completely new crystallization conditions. In collaboration with Dr. Wayne Hendrickson and coworkers, we have recently initiated a joint effort toward this end. Once diffraction quality crystals are obtained, we should be able to determine the three-dimensional structure of hCG. Such a structure would also facilitate further structural studies on individual subunits of hCG and on the interactions of hCG with its receptor and antibodies.

Nuclear Magnetic Resonance

Although X-ray crystallography is still the technique of choice for large proteins, both NMR (23, 24) and X-ray crystallography can be used to determine the *three-dimensional* (3D) structure of small- to medium-sized proteins. NMR is particularly attractive for intact hCG, given the difficulties of obtaining intact hCG crystals (1, 2). In addition, NMR allows the determination of the structure of intact hCG in solution, with its full complement of carbohydrate, which most closely resembles its physiologic condition. In the past, structures of many small proteins (less than approximately 10 kd) have been successfully determined by using the *two-dimensional* (2D) proton NMR method (24–27). For larger proteins (up to approximately 40 kd), however, it is necessary to apply the multidimensional heteronuclear NMR method (28–32).

Intact hCG is too large to be studied by the 2D proton NMR method. The problems are threefold: First, the large molecular weight (38.6 kd) severely broadens proton NMR resonances, preventing the investigator from obtaining high-quality NMR spectra. Second, the very large number of proton resonances leads to extremely complex NMR spectra for analysis. Third, the presence of a large amount of carbohydrate (comprising 1/3 of the molecular weight) causes further signal overlaps in NMR spectra, which also makes a complete analysis of NMR spectra difficult.

In order to probe the structure of hCG, we have applied the 2D proton NMR method to the α-subunit, which is common to all members of the glycoprotein hormone family. To facilitate proton resonance assignments, the α-subunit was treated with HF to partially remove carbohydrate. In this case, it was found necessary to treat the α-subunit under harsher conditions with HF in order to remove enough carbohydrate to permit gathering some interpretable NMR data. In one case, the α-subunit was treated with HF twice, and a pool was made of the smallest molecular weight alpha species. NMR studies of the HF-treated α-subunit were

carried out in our laboratories, as well as those of Dinshaw Patel. To obtain structural information, the majority of proton resonances, at least those of backbone protons, must be first identified. Therefore, the study starts with proton resonance assignments. This usually involves two steps (24). Amino acid spin systems[3] can be identified according to their characteristic covalent connectivity[4] as observed in various correlation spectra. This is accomplished by relying mainly on three-bond proton-proton J-couplings[5] (24). The order of individual amino acids along the peptide chain is then determined by observation of *nuclear Overhauser effects* (NOE)[6] (33) that result in NOE spectra from protons that are close in space (24).

To identify amino acids, a series of homonuclear correlation spectra of the HF-treated α-subunit, including *correlation* (COSY)[7], *relayed correlation* (RELAYED-COSY), *total correlation* (TOCSY), and *double-quantum correlation* (DQ-COSY) experiments, were recorded in both H_2O and D_2O. Figure 9.5 shows the NH-CαH fingerprint region of the correlation spectrum of the α-subunit of hCG in H_2O. In this spectral region, each amino acid shows a cross-peak at the chemical shifts of its amide and α protons, except that each proline shows none and each glycine shows two. The same region of the relayed-correlation spectrum (Fig. 9.6) shows not only the direct NH-CαH J-coupling correlation, but also the indirect NH-CβH J-coupling correlation. With careful analysis of the data, the NH, CαH, and/or CβH protons of all 85 amino acids other than proline are identified. Among these 85 amino acids, all alanines, glycines, threonines, and most valines are identified by their characteristic patterns in the correlation spectra. The assignments are summarized in Table 9.1. At present, sequential proton resonance assignments have not been completed because the extensive overlap of the proton resonances precludes reliable identification of some signals. As shown in Table 9.1, approximately 62% of the amide protons and more than 90% of the CαH and CβH protons possess indistinct chemical shifts from one another. Degeneracy of the proton resonances causes great difficulties in identification of sequential NOE connectivities.

---

[3] Spin systems refers to the type and number of equivalent nuclear spins in a molecule and their relationships.

[4] Covalent connectivity indicates a covalent bond connection between atoms in a molecule.

[5] J-coupling arises from a specific interaction of one nucleus with another that produces a resonance splitting in an ordinary spectra.

[6] Nuclear Overhauser effect refers to the change in intensity of a nuclear spin when another spin is perturbed by spin saturation or spin inversion (through-space experiments).

[7] COSY, RELAYED COSY, TOCSY, and DQ-COSY refer to types of homonuclear 2D NMR experiments (through-bond experiments).

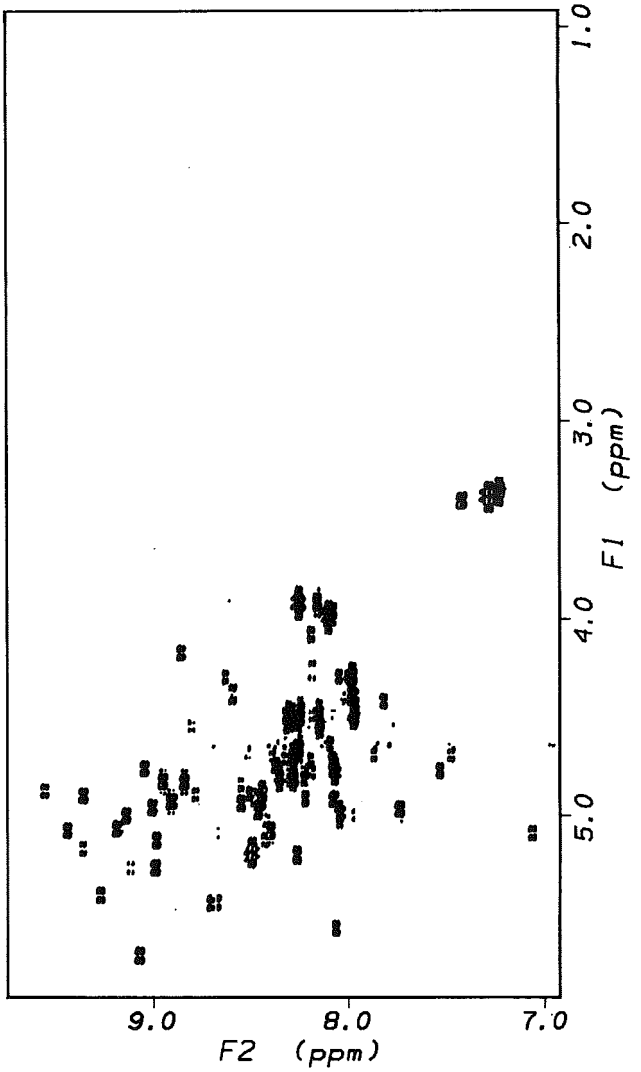

FIGURE 9.5. The NH-CαH fingerprint region of COSY spectrum of HF-treated α-hCG at pH 3.5. The spectrum is recorded on a Bruker AMX500 spectrometer at 500 MHz frequency and 60.0 °C. Each cross peak identifies the NH and CαH of individual amino acids.

Multidimensional triple-resonance (^{1}H, ^{13}C, and ^{15}N) NMR has proved to be a powerful tool for determining the structures of larger proteins (28–39). As in the 2D proton NMR method (24), proton resonance assignments are crucial in structure elucidation. This new technique provides significant improvement in spectral resolution by increasing the

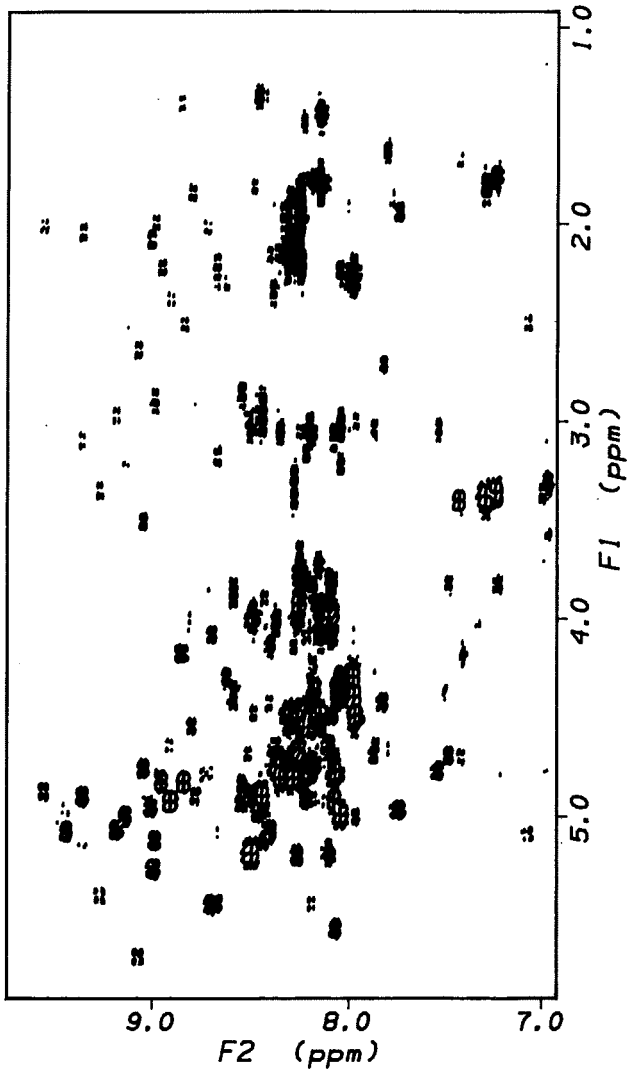

FIGURE 9.6. The NH-CαH-CβH region of relayed-COSY spectrum of HF-treated α-hCG at pH 3.5 and 60.0 °C. This spectrum is used to identify the NH-CαH-CβH fragments of individual amino acids.

spectral dimensionality to spread proton resonances according to the chemical shifts of the directly attached ^{13}C or ^{15}N. In addition, this method provides high sensitivity because it utilizes large one-bond heteronuclear J-couplings and is thus much less sensitive to the broad line width of proton resonances. The applications of multidimensional triple-resonance NMR have been successfully demonstrated by deter-

TABLE 9.1. Proton chemical shifts (ppm) of HF-deglycosylated α-hCG[a].

Residue[b]	NH	CαH	CβH	CγH	
1	9.54	4.87	2.01		
2	9.43	5.07	3.65/3.52		
3	9.35	5.16	3.09/2.62		
4	9.35	4.89	2.04		
5	9.27	5.40	3.35/2.48		
6	9.18	5.06	2.97		
7	9.13	4.99	3.19		
8	9.11	5.25	2.55		
9	9.07	5.70	2.64/3.18		
10	9.04	4.76	3.50		
11	8.99	4.95	2.08		
12	8.98	5.26	2.88		
13	8.98	5.12	1.99		
14	8.95	4.82	2.22	0.90/1.03	Val
15	8.91	4.63	2.13		
16	8.90	4.92	2.38		
17	8.85	4.17	1.39		Ala
18	8.84	4.82	2.51	0.98/1.02	Val
19	8.80	4.54	1.84		
20	8.78	4.89	c		
21	8.72	4.78	2.01		
22	8.70	5.46	4.08		
23	8.68	4.62	2.19/2.30		
24	8.67	5.44	3.18		
25	8.65	5.08	c		
26	8.62	4.29	2.29		
27	8.55	4.83	c		
28	8.55	4.94	2.85		
29	8.52	4.68	2.99		
30	8.51	4.89	3.02		
31	8.49	5.18	3.99		
32	8.48	4.47	1.81	1.09/1.05	Val
33	8.46	4.97	1.35		Ala
34	8.44	4.89	2.92/3.02		
35	8.43	4.92	3.89		
36	8.42	5.12	0.75		Ala
37	8.42	5.01	4.14	1.28	Thr
38	8.41	4.42	2.15		
39	8.40	5.07	4.12	1.25	Thr
40	8.39	4.10	2.39	1.19	Val
41	8.37	4.74	4.00		
42	8.35	4.80	3.05		
43	8.30	4.04	2.40		
44	8.28	4.47	4.14	1.31	Thr
45	8.26	5.19	3.76		
46	8.25	3.91	2.27/2.16		
47	8.22	4.81	3.05		
48	8.21	4.91	3.16		
49	8.19	4.71	c		
50	8.19	4.07	5.44		

TABLE 9.1. *Continued.*

51	8.18	4.49	1.78		
52	8.16	3.91	3.72		
53	8.15	4.48	1.44		Ala
54	8.15	4.55	1.79		
55	8.10	3.99	5.18		
56	8.09	4.64	3.79		
57	8.08	4.48	3.05		
58	8.08	4.71	4.47	1.29	Thr
59	8.07	4.77	4.36	1.43	Thr
60	8.07	4.91	4.10	1.21	Thr
61	8.06	5.57	3.37/3.01		
62	8.05	4.29	2.23	1.08/1.10	Val
63	8.04	4.99	3.23/3.01		
64	7.98	4.07	c		
65	7.98	4.40	2.29	1.08	Val
66	7.98	4.29	2.25		
67	7.97	4.53	4.38	1.37	Thr
68	7.97	4.48	4.38	1.36	Thr
69	7.97	5.00	3.00		
70	7.87	4.69	3.03		
71	7.82	4.42	2.71		
72	7.80	4.67	1.63		Ala
73	7.78	4.52	1.88		
74	7.74	4.96	1.94		
75	7.53	4.77	3.03		
76	7.47	4.69	3.82		
77	7.42	3.39	1.67/1.87		
78	7.41	4.70	c		
79	7.28	3.38	1.81		
80	7.23	3.35	1.75		
81	7.07	5.09	2.06		
82	8.80	4.31/3.99			Gly
83	8.59	4.38/3.86			Gly
84	8.18	4.25/3.81			Gly
85	8.07	4.77/3.93			Gly

[a] Proton chemical shifts of HF deglycosylated α-hCG at pH 3.5 and 60 °C.
[b] The ordering of residues is purely arbitrary.
[c] Not identified due to signal overlap.
Source: Reprinted with permission from Lustbader, Yarmush, Birken, Puett, and Canfield (3). © The Endocrine Society, 1993.

mination of the structures of several medium-sized (about 20 kd) ^{13}C,^{15}N-labeled proteins (31, 32).

Determination of the entire structure of intact hCG presents a challenge for today's NMR technology because of its large molecular weight, as well as its large carbohydrate content. Multidimensional triple-resonance NMR will be applied to intact hCG. This approach requires the hormone

to be uniformly enriched with ^{13}C and ^{15}N isotopes. Therefore, we are pursuing expression of uniformly ^{13}C,^{15}N-labeled hCG in mammalian cells.

In 1992 we were searching for more feasible methods for performing NMR studies on hCG. It was obvious that chemical deglycosylation, even the harsh double HF treatment that we had performed with the α-subunit, was inadequate because of residual sugar content. Incorporation of the NMR active isotopes, ^{13}C and ^{15}N, was the only feasible response. The problem was that whereas ^{13}C and ^{15}N can easily be incorporated into bacterial proteins by growing the bacteria on ^{13}C glucose and ^{15}N ammonium salts, mammalian cells require a medium containing all the amino acids preformed. For the production of ^{13}C- and ^{15}N-labeled mammalian proteins, these amino acids must all be present and all universally labeled with ^{13}C and ^{15}N. Unfortunately, not all amino acids are commercially available in ^{13}C- and ^{15}N-labeled form, and moreover, some cannot be prepared from published procedures.

Fortuitously, we saw an advertisement from Martek that stated that dual-labeled amino acids would soon be available. Martek was contacted at an apparently critical time in their efforts to produce such dual-labeled materials, and one key member of their research team, Jonathan Miles Brown, was already active in producing NMR-labeled materials for use by the general research community. He was at the point of designing universally labeled mammalian cell growth media and needed a collaborator to help in assessing various formulations of these media. We entered into such a collaboration using an ongoing expression system for hCG that had been refined to produce hCG for use in the crystallization work. Several members of our group visited Martek and found an innovative company producing products from algae. ^{13}C- and ^{15}N-labeled proteins were produced by supplying the algae with ^{13}C- and ^{15}N-labeled carbon dioxide and nitrogen salts, respectively. Jonathan Miles Brown had developed a scheme to hydrolyze the algal proteins in a fashion that preserves all the labeled amino acids, remarkably, including the amides, and that produces them in high purity. Much exploration of mixtures of concentrations of amino acids was necessary, but in a period of months, a suitable media formulation was developed that would sustain mammalian cells that were producing NMR-labeled hCG. This represents a great advance in the potential to solve structures of the mammalian proteins, especially glycoproteins, which cannot be correctly expressed with sugars intact in any cells other than mammalian.

In our upcoming studies, the α- and β-subunits will first be dissociated and their structures determined separately. This enables the analysis of a smaller molecular weight molecule, increasing the probability of high-quality NMR spectra and facilitating data analysis. A variety of triple-resonance experiments have been developed for resonance assignments of polypeptides (28–30, 34–43). These experiments are also essential for

glycoproteins. Alternatively, it will be neceessary for us to develop more effective resonance assignment procedures for the carbohydrate moieties of the glycoprotein. To determine the entire structure of hCG, two intact hCG molecules will be prepared, each consisting of a ^{13}C,^{15}N-labeled subunit and its unlabeled complementary subunit. This approach will facilitate resonance assignments, as well as measurements of structural information. Complexity of heteronuclear NMR spectra of such heterodimeric hormones is the same as these of their subunits alone because the heteronuclear NMR permits the selective observation of those protons bonded to ^{13}C and/or ^{15}N instead of the protons bonded to ^{12}C or ^{14}N, or vice versa. However, the greater molecular weight of intact hCG will severely broaden NMR resonances, especially those carbons bonded to protons, such as α carbons. For intact hCG, optimization experiments are needed to overcome this problem. Generally, this problem is a major obstacle to the application of NMR technology applied to larger molecules, including the glycoprotein hormones.

The application of computer modeling, X-ray crystallography, and NMR methods is bringing about encouraging progress toward the solution of the three-dimensional structure of hCG. Modeling has yielded a representation of the structure with defined atomic coordinates that can be tested against experimental data. However, the chief benefit of modeling will come after the hCG structure is solved by one or both of the methods described in this chapter. The structures of the other homologous glycoprotein hormones will then be predicted by modeling techniques. Most importantly, hormone analogs will be precisely designed from the known geometry of the hormone surface. Agonists and antagonists can then be constructed by the techniques of molecular biology and tested in vitro and in vivo. From these studies, small molecules may be designed that will serve as effective regulators of human fertility and permit a simple and humane control of human population growth.

Acknowledgment. This work was supported by NIH Grant HD-15454.

References

1. Harris DC, Machin KJ, Evin GM, Morgan FJ, Isaacs NW. Preliminary X-ray diffraction analysis of human chorionic gonadotropin. J Biol Chem 1989;264: 6705–6.
2. Lustbader JW, Birken S, Pileggi NF, et al. Crystallization and characterization of human chorionic gonadotropin in chemically deglycosylated and enzymatically desialylated states. Biochem 1989;28:9239–43.
3. Lustbader JW, Yarmush DL, Birken S, Puett D, Canfield RE. The application of chemical studies of human chorionic gonadotropin to visualize its three-dimensional structure. Endocr Rev 1993;14:1–21.

4. Katz L, Levinthal C. Interactive computer graphics and representation of complex biological structures. Annu Rev Biophys Bioeng 1972;1:465–504.
5. Moyle WR, Matzuk MM, Campbell RK, et al. Localization of residues that confer antibody binding specificity using human chorionic gonadotropin/luteinizing hormone β subunit chimeras and mutants. J Biol Chem 1990;265:8511–8.
6. Willey KP, Leidenberger F. Functionally distinct agonist and receptor-binding regions in human chorionic gonadotropin. J Biol Chem 1989;264:19716–29.
7. Mise T, Bahl OP. Assignment of disulfide bonds in the beta subunit of human chorionic gonadotropin. J Biol Chem 1981;256:6587–92.
8. Mise T, Bahl OP. Assignment of disulfide bonds in the alpha subunit of human chorionic gonadotropin. J Biol Chem 1980;255:8516–22.
9. Puett D, Birken S. Helix formation in reduced S-carboxymethylated human choriogonadotropin β subunit and tryptic peptides. J Protein Chem 1989;8:779–94.
10. Reddy VB, Beck AK, Garramone AJ, Vellucci V, Lustbader J, Bernstine EG. Expression of human choriogonadotropin in monkey cells using a single simian virus 40 vector. Proc Natl Acad Sci USA 1985;82:3644–8.
11. Lustbader J, Birken S, Pollak S, et al. Characterization of the expression products of recombinant human choriogonadotropin and subunits. J Biol Chem 1987;262:14204–12.
12. Matthews BW. Solvent content of protein crystals. J Mol Biol 1968;33:491–7.
13. Watenpaugh KD. Overview of phasing by isomorphous replacement. Methods Enzymol 1985;115:3–14.
14. Hendrickson WA. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 1991;254:51–8.
15. Hendrickson WA, Smith JL, Sheriff S. Direct phase determination based on anomalous scattering. Methods Enzymol 1985;115:41–54.
16. Leahy DJ, Hendrickson WA, Aukhil I, Erickson HP. Structure of a fibronectin type III domain from Tenascin phased by MAD analysis of the selenomethionyl protein. Science 1992;258:987–91.
17. Weis WI, Kahn R, Fourme R, Drickamer K, Hendrickson WA. Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. Science 1991;254:1608–15.
18. Yang W, Hendrickson WAS, Crouch RJ, Satow Y. Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. Science 1990;249:1398–405.
19. Graves BJ, Hatada MH, Hendrickson WA, Miller JK, Madison VS, Satow Y. Structure of interleukin 1 A at 2.7 A resolution. Biochemistry 1990;29:2679–84.
20. Guss JM, Merritt EA, Phizackerley RP, et al. Phase determination by multiple-wavelength x-ray diffraction: crystal structure of a basic "blue" copper protein from cucumbers. Science 1988;241:806–11.
21. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J 1990;9:1665–72.
22. Cowie DB, Cohen GN. Biochim Biophys Acta 1957;26:252–61.

23. Ernst RR, Bodenhausen G, Wokaun A. Principles of nuclear magnetic resonance in one and two dimensions. Oxford: Oxford Science Publications, 1986.
24. Wuthrich K. NMR of proteins and nucleic acids. New York: Wiley, 1986.
25. Markley JL. Two-dimensional nuclear magnetic resonance spectroscopy of proteins: an overview. Methods Enzymol 1989;176:12–34.
26. Clore GM, Gronenborn AM. Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. CRC Crit Rev Biochem Mol Biol 1989;24:479–564.
27. Wagner G. NMR investigations of protein structure. Prog NMR Spectroscopy 1990;22:101–39.
28. Kay L, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 1990;89: 496–514.
29. Ikura M, Kay M, Bax A. Novel approach for sequential assignment of ^{1}H, ^{13}C and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy: application to calmodulin. Biochemistry 1990;29:4659–67.
30. Montelione G, Wagner G. Conformation-independent sequential NMR connections in isotope-enriched polypeptides by ^{1}H-^{13}C-^{15}N triple-resonance experiments. J Magn Reson 1990;87:183–8.
31. Ikura M, Clore GM, Gronenborn AM, Zhu G, Klee G, Bax A. Solution structure of a Calmodulin-target peptide complex by multidimensional NMR. Science 1992;256:632–7.
32. Clore GM, Wingfield P, Gronenborn AM. High-resolution three-dimensional structure of Interleukin β in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy. Biochemistry 1991;30:2315–23.
33. Noggle JH, Schirmer RE. The nuclear Overhauser effect: chemical applications. New York: Academic Press, 1971.
34. Boucher W, Laue E, Campbell S, Domaille P. Four-dimensional heteronuclear triple-resonance NMR methods for the assignment of backbone nuclei in proteins. J Am Chem Soc 1992;114:2262–4.
35. Kay L, Wittekind M, McCoy M, Friedrichs M, Miller L. 4D NMR triple-resonance experiments for assignment of protein backbone nuclei using shared constant-time evolution periods. J Magn Reson 1992;98:443–50.
36. Olejniczak ET, Xu RX, Petros AM, Fesik SW. Optimized constant-time 4D HNCαHα and HN(CO)CαHα experiments: applications to the backbone assignments of the FKBP/Asomycin complex. J Magn Reson 1992;100:444–50.
37. Grzesiek S, Bax AJ. Improved 3D triple-resonance NMR techniques applied to a 31 kD protein. Magn Reson 1992;96:432–40.
38. Palmer AG III, Fairbrother WJ, Cavangh J, Wrught PE, Rance M. Improved resolution in three-dimensional constant-time triple resonance NMR spectroscopy of proteins. J Biomol NMR 1992;2:103–8.
39. Farmer BT II, Venters RA, Spicer LD, Wittekind MG, Mueller L. Refocused optimized HNCα: increased sensitivity and resolution in large macromolecules. J Biomol NMR 1992;2:195–8.
40. Clubb RT, Thanabal V, Wagner G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ^{1}H, ^{15}N and ^{13}C: chemical shifts in ^{15}N-^{13}C-labeled proteins. J Magn Reson 1992;97:213–7.

41. Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M, Kay L. Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of ^{13}C-labeled proteins. J Magn Reson 1990;87:620–7.
42. Bax A, Clore GM, Gronenborn AM. ^{1}H-^{1}H correlation via isotopic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. J Magn Reson 1990;88:425–31.
43. Fesik SW, Eaton HL, Olejniczak ET, Zuiderweg ERP, McIntosh LP, Dahlquist FW. Proton-proton correlation via carbon-carbon couplings: a three-dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with carbon-13. J Am Chem Soc 1990;112:886–8.