

PD3 023

Modern NMR spectroscopy of proteins and peptides in solution and its relevance to drug design

E.R.P. Zuiderweg*, S.R. van Doren, A.V. Kurochkin, R.R. Neubig and A. Majumdar

*Biophysics Research Division and Department of Biological Chemistry, The University of Michigan,
553 North University, Ann Arbor, MI 48109, U.S.A.*

Received 6 August 1993

Accepted 5 October 1993

Key words: Three-dimensional structure; Isotopic labeling; Dynamics, Multidimensional NMR

SUMMARY

The knowledge of the three-dimensional (3D) structures and conformational dynamics of proteins and peptides is important for the understanding of biochemical and genetic data derived for these molecules. This understanding can ultimately be of help in drug design. We describe here the role of Nuclear Magnetic Resonance (NMR) spectroscopy in this process for three distinct situations: for small proteins, where relatively simple NMR methods can be used for full 3D structure determination; for larger proteins that require multinuclear multidimensional NMR but for which full 3D structures can still be obtained; and for small peptides that are studied in interaction with macromolecules (receptors) using specialized NMR techniques. A fourth situation, pertaining to large systems where only partial structural information can be obtained from NMR data, is briefly discussed. Molecules of interest to the biomedical field (C5a and stromelysin) are discussed as examples.

INTRODUCTION

The importance of nuclear magnetic resonance (NMR) spectroscopy for drug design arises from the fact that this technique has become a major tool for the determination of conformations of biological molecules in solution. Once conformations of (bound) drugs, their target molecules or their complexes are known, molecular design processes can take place to help improve many aspects of the quality of the drug. Here we will focus on the impact of NMR spectroscopy on the determination of protein and peptide conformations in solution and how the information obtained can affect the design process. It is quite customary to try to compare the conformational information obtained from NMR with that obtained from X-ray crystallography. We will not go into these questions, except that we will discuss an intriguing difference in crystal and solution structures of the anaphylatoxin C3a. Instead, let us ask what distinguishes NMR from other spectro-

*To whom correspondence should be addressed.

sopic techniques such as UV/VIS, infrared or Mössbauer. Why can one determine 3D structures of proteins with NMR but not with other spectroscopic methods? It is certainly not the basic physics: in NMR, as well as in other spectroscopies, transitions between energy levels are induced by matching the frequency of a photon between them (the transitions are called resonances in NMR); the fact that an external magnet is needed to create the difference in energy levels for NMR is not really relevant. As in other spectroscopies, the excited NMR states relax back to their ground state with some characteristic lifetime. What distinguishes NMR from most other spectroscopies is the timescale of such lifetimes. NMR relaxation times can be seconds in contrast to the characteristic nano- and picoseconds in optical spectroscopies, owing to the fact that the nuclei involved in NMR are physically very isolated from each other. The slow relaxation (long lifetime) of nonequilibrium states of the nuclear spin energy levels allows application of *coherent* spectroscopy. This means that it is technically feasible to pump a second photon (or many more) into the system after initial excitation, before it can relax back to its ground state. Thus, multiple-quantum excitations can be generated with relative ease in NMR. Transition energies can be passed between nuclei on resonant terms. The long lifetimes also enable application of NMR in many frequency dimensions, which in turn makes it possible to assign the resonances in the NMR spectrum to individual nuclei in the molecule under study. Energy transfers between many assigned resonances can then be studied. The efficiencies of these energy transfers are dependent on the structure and conformation of the molecule. In particular, the important energy transfer mechanism called the Nuclear Overhauser Effect (NOE) is only effective when nuclei are separated by no more than 5 Å in space. Thus, when an NOE can be measured between two resonances, it means that the pair of nuclei associated with these resonances are close in space. When a dense network of these NOEs between assigned resonances can be measured, the distance information can be transformed into a molecular model, generally using computational methods. Depending on the characteristics of the molecule under study and the conditions of the experiment, 3D structures can be determined with atomic precision for molecules up to at least 20 and possibly 30 kDa. An important feature is that these high-resolution NMR structures are (and have to be) determined in solution, where in most cases conditions can be chosen similar to native ones.

In order to give the general reader the opportunity to appreciate the possibilities and limitations of the modern NMR methods (which are often referred to as two-, three-, four- or multidimensional NMR, heteronuclear, multinuclear NMR, triple- and quadruple-resonance NMR, isotope-edited, isotope-directed NMR and all combinations thereof), we include a very basic NMR tutorial in the next section. A major part of this contribution consists of a description of an NMR study of the small inflammatory protein C5a and the impact of the obtained information on drug design. The following section illustrates the initial stages of an NMR study of the much larger human stromelysin for which the new multinuclear multidimensional NMR methods have to be used. The last section will describe how these methods come together in the study of smaller ligands in interaction with larger structures.

SOME NMR BASICS

NOE

A very important energy transfer mechanism is the Nuclear Overhauser Effect [1]. It is mediated by the through-space magnetic interaction between the nuclear magnets themselves. When the

molecule tumbles through solution, the interaction between these magnets, represented by ordinary dipoles, becomes modulated because it is dependent on their relative orientations. The radio-frequency electromagnetic fields generated by the time-modulated magnetic interaction will cause magnetic resonance. This 'internal' magnetic resonance will even out (relax) the nonequilibrium magnetic states initially generated by the NMR experiment. The rates of relaxation are inversely proportional to r_{ij}^6 , where r_{ij} is the distance between nuclei i and j in an NOE pair. Careful measurement of these rates (NOE build-up rates) yields distance information in the range 2–5 Å, with a precision of a few tenths of an Angstrom.

Two-dimensional NMR

NMR energies are small compared to thermal energies and the nuclear energy levels will be nearly equally populated at equilibrium. The small difference that does exist gives rise to a net macroscopic magnetic moment. At equilibrium, the magnetic moment is perfectly aligned with the magnetic field. NMR experiments are carried out by delivering radio-frequency (r.f.) pulses to the sample. This is equivalent to whipping a spinning top; the top will start to precess around the vertical axis. Analogously, the magnetic moment will precess around the direction of the external magnetic field. This precession gives rise to an oscillating electromagnetic field (Free Induction Decay or FID) that can be picked up by a wire coil antenna. The frequency of the FID is equal to the frequency of the NMR transition excited in the first place. When different nuclei are excited in the sample, the time-domain signal will be a superimposition of all the individual transitions with different frequencies. Fourier transformation of this composite FID gives the frequency spectrum (see row 1 of Fig. 1). In the 2D NMR experiment one does not wait until the FID has died out; because of the slow relaxation back to equilibrium as described above, one has time to issue a second pulse on the nonequilibrium state. The effect of this second pulse will depend on the moment it is delivered; just like the outcome of a second whiplash on the top's precession will depend on the position of the top when it is hit. Let us assume for simplicity that the second r.f. pulse does nothing but select for the amplitude of the signal as indicated in Fig. 1. If we record the effect of this second pulse as a function of the delay between the two pulses (this time is denoted t_1), we obtain FIDs (denoted t_2) that are amplitude-modulated as a function of t_1 . Fourier transformation of this series of FIDs gives rise to a series of amplitude-modulated spectra. The frequency of this amplitude modulation is of course given by the frequency present during t_1 ; this frequency can be measured by carrying out a second Fourier transformation over corresponding points of the amplitude-modulated spectra in the orthogonal direction [2]. The obtained 2D frequency-domain spectrum (f_1, f_2), which is generally represented as a contour diagram, is not very interesting when the frequencies in t_1 and t_2 are the same: the 2D spectrum is a diagonal as shown in Fig. 1. The 2D spectrum becomes informative when the frequencies in t_1 and t_2 are not equal. This can be invoked when, instead of a single pulse, a rest period flanked by two pulses is inserted in the experiment. During this rest time the NOE energy transfer becomes active. A customary description of the processes involved is that resonance energy residing on nucleus A with frequency f_A during t_1 is transferred to the spatially close nucleus B in the rest period. After the rest it will be measured with an FID (t_2) of frequency f_B . The 2D spectrum is now a correlation diagram: the off-diagonal peak at coordinates (f_A, f_B) correlates the two nuclei A and B, in this example by NOE. As a real example, consider the 2D NOE spectrum of C5a, a 75-amino acid protein of the complement system (Fig. 3). The off-diagonal

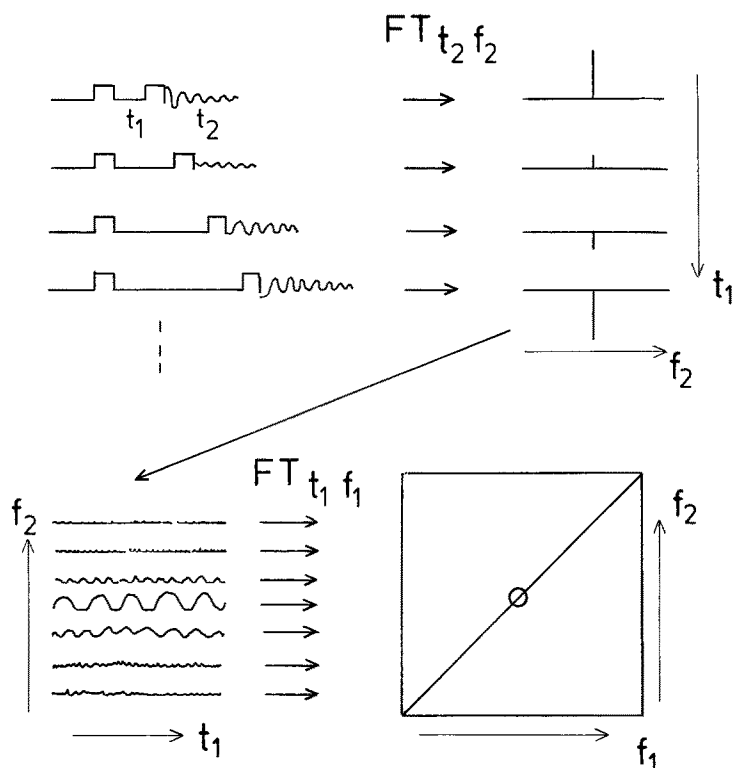


Fig. 1. Schematic representation of the recording and processing of a 2D NMR spectrum. Left top: the squares represent the r.f. pulses that demarcate the two time periods t_1 and t_2 . The squiggly lines are the FIDs; note that they are amplitude modulated as a function of t_1 . Right top: the result after Fourier transformation of this series of FIDs. The frequency-domain spectra (f_2) are amplitude modulated as a function of t_1 . Left bottom: a transposed representation of the same data. Corresponding points in f_2 give rise to t_1 'FIDs'. Right bottom: the final result of the 2D Fourier transformation: a peak is found on the diagonal.

peaks give the distance correlations between the nuclei in the sample. If the assignments of the resonances are known, these correlations can be interpreted in structural terms.

Coherence transfer through scalar couplings

The NOE is not the only energy-transfer process between transitions. The second, and equally important process is known as coherence transfer through scalar couplings or just as coherence transfer [2]. The scalar coupling is an interaction between nuclear spins, mediated by electrons in the chemical bonds joining the nuclei. By this process, nucleus A senses the state of nucleus B. Scalar couplings occur in principle between all magnetic nuclei in a molecule; however, the effect becomes unmeasurably small if more than three or four chemical bonds separate the nuclei. When three or four intervening bonds are involved in the mediation of the interaction, the scalar coupling depends on the dihedral angle described by the chemical bonds and becomes a conformational marker by itself. The dihedral angles obtained from measurement of the couplings are used as additional input for structure-determination algorithms.

It is impossible to describe the energy *transfer* by scalar coupling from one transition to the

other in terms other than quantum mechanics. In very general terms, however, one can state that when the energy levels of nucleus A are affected by the nuclear state of the scalarly coupled nucleus B, transitions of B (which affect the nuclear state of B) are sensed by transitions of A. Thus, transfer of transition 'information' between nuclei is possible via the scalar coupling. The first 2D experiment called COSY (for correlation spectroscopy — a much too generic name) just transfers coherence from one nucleus to the other over the scalar couplings [2]. When executed as a 2D experiment, cross peaks will thus indicate that the corresponding nuclei are part of the same molecule and separated by not more than three (or sometimes four) chemical bonds. The COSY-style experiments thus delineate the chemical structure of a component, whereas NOESY-style experiments characterize the conformation of the molecule.

Multidimensional NMR

The principle of 2D NMR is easily extended to three or more dimensions. First, we need mechanisms to transfer information between nuclei. For example, let information from the resonance associated with nucleus A be carried to nucleus B by NOE and from nucleus B to nucleus C by scalar coupling. If we now record the FID of C as a function of the time the magnetization resided on A *and* as a function of the time it resided on B, we obtain a 3D time-domain matrix. Fourier transformation of this 3D experiment will give 3D cross peaks in a frequency-domain spectrum. The existence of the cross peak will show that A is spatially close to B, which is in turn within three chemical bonds of C. The three frequency axes of the cross peak determine the frequencies of A, B and C, respectively. Practical examples of 3D experiments will be discussed in conjunction with the NMR studies of stromelysin.

A 3D experiment can still be depicted in our minds, but we have problems if dimensionality goes beyond that. However, molecules, quantum mechanics, NMR instrument and computers easily cope. Take a (not so hypothetical) 5D experiment in which magnetization information is passed by heteronuclear scalar couplings in a labeled peptide from $H^{\alpha}(i)$ to $C^{\alpha}(i)$ to $CO(i)$ to $N(i+1)$ to $H^N(i+1)$. Occurrence of cross peaks in the 5D matrix would indicate that the named nuclei are connected in the molecular framework. The five dimensions would give the five frequencies of the connected nuclei. If the assignment for one of them is known, this experiment would give the assignment for the other four. The number of achievable dimensions is practically limited by relaxation, sensitivity and resolution. For most larger molecules where multidimensional NMR is necessary for data analysis, the limit lies currently around four dimensions.

THREE-DIMENSIONAL STRUCTURE DETERMINATION OF C5a AND IDENTIFICATION OF THE C5a-RECEPTOR INTERACTION

Background

C5a is the N-terminal 74-amino acid fragment of the C5 complement protein, released upon cleavage by C3b. C5a is a primary mediator of inflammation and recruits leukocytes to the site of complement activation. Pathologic complement activation occurs in auto-immune diseases such as arthritis and asthma, resulting in tissue damage inflicted by the recruited lymphocytes (for a review, see Reference 3). It is therefore thought that a compound that competes with the binding of C5a to its leukocyte receptor in an antagonistic way may be a useful therapeutic against some of the symptoms of the mentioned diseases. The basic idea is to develop a low-molecular-weight

compound that mimics the receptor-interface site of C5a on the basis of the 3D structure of that site. No 3D structure was known for C5a; a 3.5 Å resolution crystal structure was known for the highly homologous C3a [4]; the receptor for C5a was unavailable. The approach was (i) to determine the 3D structure of C5a by NMR in solution; (ii) to use mutagenesis to identify residues that are involved in the C5a receptor binding; and (iii) to combine (i) and (ii) in a 3D template to be used for rational drug design.

NMR structure determination

With 75 amino acids, C5a is sufficiently small to be amenable to 2D proton NMR. The upper limit for such approaches lies around 10 kDa; above this molecular weight the relaxation processes become too fast to allow the through-bond proton COSY-type experiments which are crucial for 2D NMR data analysis [5]. For C5a, we used such experiments to correlate the resonances within the spin systems of amino acids. Most amino acids give rise to unique connectivity patterns in these experiments and could thus be recognized by type. The identified amino acid residues were placed in the primary structure with the aid of a NOESY experiment. We made use of the property that at least one of the distances $d\alpha N(i, i + 1)$ or $dNN(i, i \pm 1)$ between the alpha and amide protons (N) or between amide protons of sequential residues i and $i + 1$ is smaller than 3 Å and thus within range of NOE energy transfer [5]. Thus connections between the protons of sequentially adjacent residues were made, leading to the so-called sequence-specific assign-

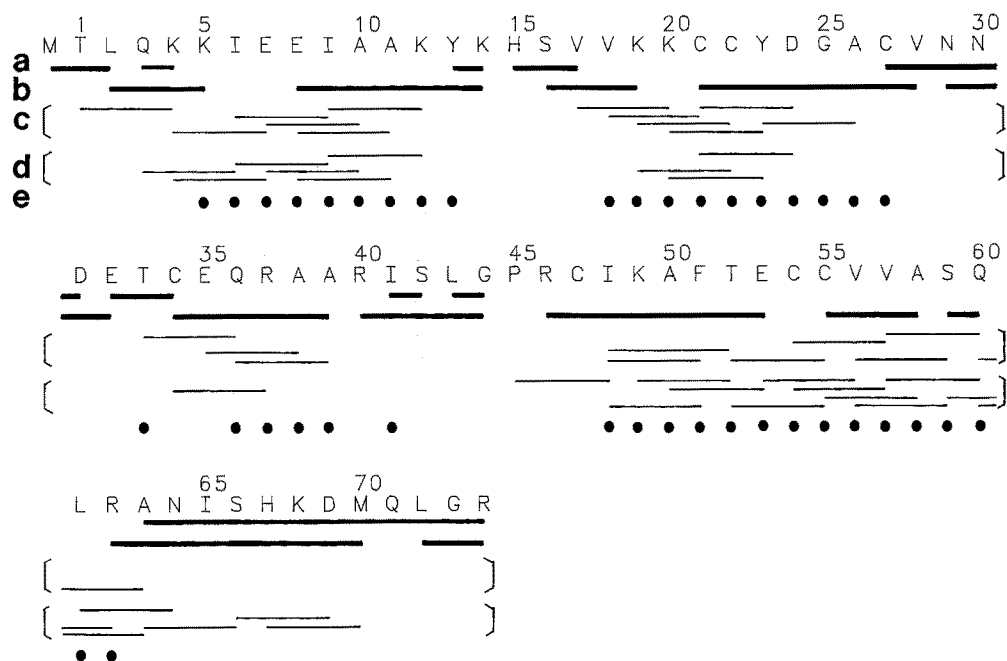


Fig. 2. Secondary structure determination of C5a in solution. Lines in lane a represent the presence of a strong cross peak $d\alpha N(i, i + 1)$ in the NOESY spectrum (indicative of extended structure) for the indicated residues. Lines that connect residues in lane b point to the presence of a strong cross peak $dNN(i, i \pm 1)$ (indicative of helical structure) in the NOESY spectrum. Lines in lane c represent strong cross peaks $d\alpha N(i, i + 3)$ and $d\alpha N(i, i + 4)$ and lines in lane d indicate $d\alpha\beta(i, i + 3)$ (all indicative of helical structure). Lane e identifies amide protons protected from exchange with the solvent.

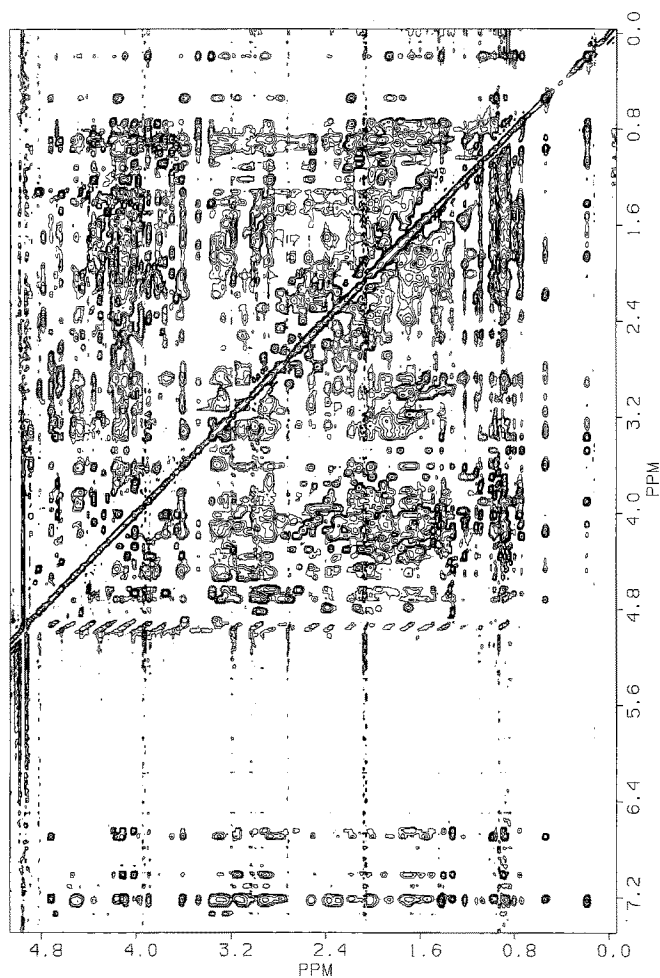


Fig. 3. Part of the NOESY spectrum of C5a.

ment [6]. Figure 2 shows the obtained connections for C5a. It is immediately apparent that many strong dNN NOEs were observed for this protein. Of course, this means that the corresponding distances in the protein are small; this is indicative of the presence of helices, where dNN is 2.8 Å. In extended structures the distance is 4.8 Å and generally beyond detection possibility by NOE. The $d\alpha N(i, i + 1)$ distance in helical regions is relatively long (3.5 Å) and thus gives rise to weaker NOEs in these areas. Once there are indications that a protein is helical, one undertakes a grid search for so-called medium-range connectivities that corroborate the identification of secondary structure [7]: the α -helical pitch of 3.6 residues per turn brings the H^α of residue i in close proximity to the H^N of residues $i + 3$ (3.5 Å) and $i + 4$ (4.2 Å). The corresponding NOEs, termed $d\alpha N(i, i + 3)$ and $d\alpha N(i, i + 4)$, are shown in Fig. 2, together with an NOE corresponding to another small distance in a helix ($d\alpha\beta(i, i + 3)$). The locations of four α -helices are clearly outlined by these NOEs. In order to further characterize the secondary structure, the solvent exchange rates of the labile amide protons were determined. Amide protons exchanging slowly are protected from solvent by hydrogen bonding, protein bulk or both [8]. In the case of small proteins such

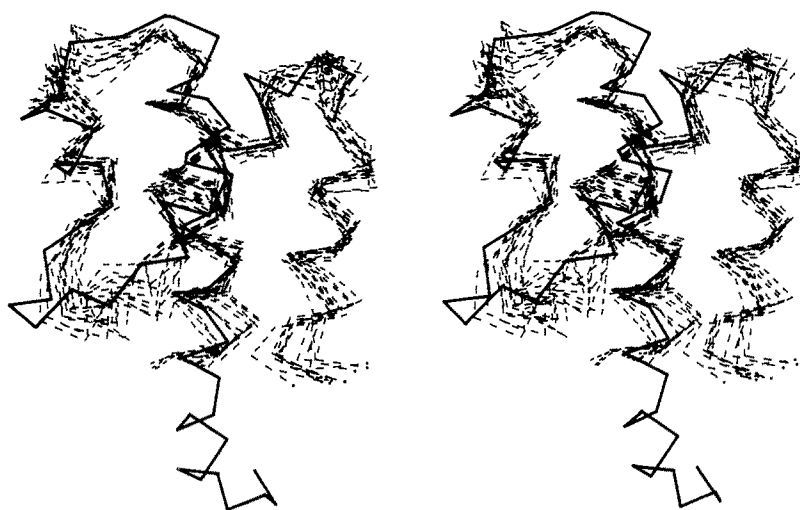


Fig. 4. Stereodiamgram of the 3D structure of C5a in solution as determined from NMR data, represented as an ensemble of 11 structures using dashed lines (C^α atoms only). The 3D structure of C3a in the crystal as determined from X-ray diffraction data is shown as a bold line.

as C5a, hydrogen bonding is the most likely cause of exchange protection. The slowly exchanging protons are indicated in Fig. 2 and align perfectly with the helical stretches. In parallel and antiparallel β -sheets, $d\alpha N$ distances are very short (2.2 Å), giving rise to very strong NOE cross peaks and dNN distances too long for appreciable NOEs. It is clear that C5a does not contain such stretches of sheet. For the C-terminal 10 residues one observes fairly strong $d\alpha N$ and dNN NOEs *simultaneously* and also some medium-range connectivities. These NOEs indicate that this area does not conform to a standard secondary structure where these NOEs are mutually exclusive; rather, the molecule is here in dynamic equilibrium between many states. This is corroborated by the fact that amide protons exchange fast in this region, showing that no stable hydrogen-bonding pattern is formed. Relaxation measurements revealed that the mobility of the C-terminal part of C5a becomes monotonically larger when proceeding from Ala⁶³ towards the end of the molecule; the last four residues move ten times as fast in solution as the core of the protein and behave as if they were part of a small peptide [9]. It is difficult to state the nature of the dynamic equilibrium; because of the r^{-6} distance dependence of NOEs, close distances are strongly over-emphasized in the average NOE. Thus, even when the majority of time is spent as random extended structure and only a small fraction of time in a helical conformation, dNN NOEs would accumulate. (This nongeometric averaging of NOE is utilized in transferred NOE experiments discussed below.)

The results of this investigation thus reveal that C5a consists of four α -helices and that the molecule has a partially disordered C-terminal decapeptide. In order to proceed with 3D structure determination [10], the NOE data were further searched for NOEs between protons that are remote in sequence, but still within 4 Å. For an α -helical protein, most of these so-called long-range NOEs are found between side-chain resonances and thus occur in a relatively crowded part of the NOESY spectrum (see Fig. 3). The NOEs were interpreted with extreme care: only NOEs that could be observed at a very short mixing time (50 ms) were included in the calculation. The

3D structure of C5a was determined using these long-range NOEs and the data determining the secondary structure (no hydrogen bonds in the first rounds). As no NOEs were identified positioning the C-terminal 11 residues of the molecule (corroborating its random structure), structure calculations were carried out for the fragment containing residues 1–63 only. The structure of C5a was computed from the NMR data with a dihedral-angle-driven distance geometry (DG) program [11,12] and 5 ps of restrained molecular dynamics (MD) [13–15]. The final ensemble of 11 structures [10] was compatible with the observed presences and absences of NOEs and with well-established model potential-energy functions. The ensemble is shown in Fig. 4, together with the crystal structure of C3a.

Structure evaluation, mutagenesis and receptor interaction

Comparison of the solution structure of C5a [10] with the crystal structure of C3a [4] (Fig. 4) reveals correspondences as well as some striking differences. The second and third helices (see Fig. 2) as well as the beginning of the large fourth helix correspond between the two structures; the topology of these elements is identical in both structures. However, no N-terminal helix is observed for C3a in the crystal, whereas a well-defined, well-docked helix is seen for C5a in solution. A well-ordered C-terminal fragment is observed for C3a in the crystal, whereas the conformation of C5a in solution is disordered in that region. Are the observed differences due to the difference in species, to the difference in conditions, or to both? This puzzle was (partially) solved by the determination of the secondary structure of C3a in solution [16]. We found that the C-terminus of C3a in solution is also disordered, just like that of C5a. Apparently, intermolecular contacts in the crystal stabilize a helical conformation for this region. The NMR data for C3a in solution also indicate that C3a, like C5a, has a helix at its N-terminus. However, the data indicate that this helix is less stable than its C5a counterpart (fraying towards the N-terminal side) and that the docking of this helix against the core of the molecule is unstable as well. Nevertheless, the helix exists in solution. It thus turns out that the differences observed between the crystal structure for C3a and the solution structure for C5a are mainly due to differences in conditions.

Once the structure was determined, functional information that was acquired over the years could be put in context. It was determined that a synthetic octapeptide corresponding to the last eight residues of C5a is active in receptor binding [17]. It was therefore very disappointing to find that the C-terminal part of C5a is unstructured in solution; no structural template for peptidomimetics could thus be generated. The NMR information *did* suggest that there was no particular reason to try to make peptides of a particular secondary structure. The NMR-derived noninformation in this region is also of relevance to drug design for C3a. Peptides corresponding to the C-terminal region of C3a account for 100% of the activity of the molecule and efforts have been directed to improvement of the models of these peptides by conformational restriction [18]. The NMR data clearly caution against the uncritical use of crystal coordinates for structural modeling of these peptides.

The C-terminus accounts for only part of the activity of C5a, since synthetic peptides based on the natural sequence have 10^{-4} M receptor-binding affinity while the affinity of C5a for the same receptor is in the subnanomolar range. It is remarkable, however, that these peptides are full agonists. Genetically constructed truncated C5a molecules comprising residues 1–62 (the 'core') also showed residual receptor binding and full agonism at 10^{-5} M. It therefore seems that the C-terminal part and the core of the molecule are both required for full affinity [19]. This is

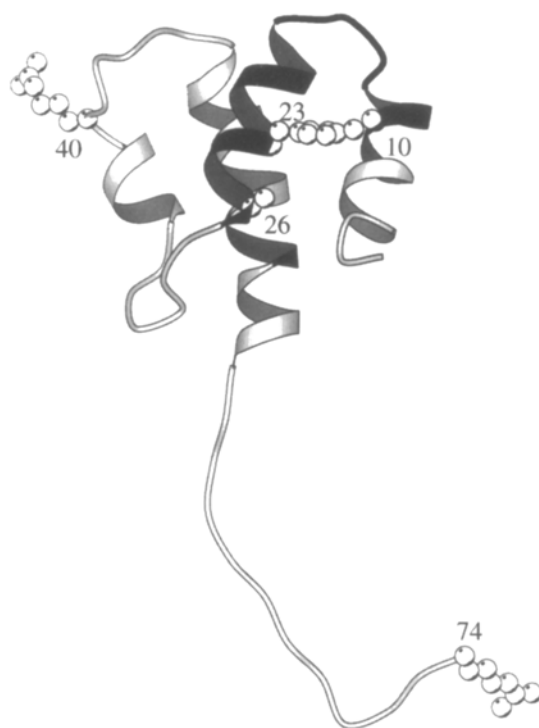


Fig. 5. Molscript [69] drawing of C5a in solution. Several residues of interest, as discussed in the text, are indicated. The conformation of the extended C-terminal part was chosen arbitrarily. The darker shading in the interface region between the first two helices corresponds to the area displaying large chemical-shift changes for the mutant Ala²⁶→Met (Fig. 6).

corroborated by results of mutagenesis and chemical modification. Step-by-step hydrolysis of the N-terminus of C5a resulted in a progressive loss of activity and mutations Ala²⁶→Met and Arg⁴⁰→Gly in the 'core' gave rise to a 10-fold loss in activity [19]. Inspection of the 3D structure of C5a shows that these 'hot spots' are spread over the entire molecule (Fig. 5) and that they are not adjacent to the C-terminus at all. In order to help understand these results, we used NMR to study some of the mutants. Mutants that did not show a change in activity were found to be structurally equivalent to wild-type C5a, as the 2D NOESY data for these molecules superimposed. Almost complete superimposition of NOESY spectra of wild-type and mutant Arg⁴⁰→Gly was also possible; however, minor structural perturbations were seen for the backbone of residue 40 itself and for its direct neighbors. Therefore, the loss of affinity in this mutant must indicate that Arg⁴⁰, or at the very least its direct surroundings, are involved in receptor binding. As expected, Arg⁴⁰ is a surface residue. The mutant Ala²⁶→Met is very different. Ala²⁶ is found on the interface between the first and second helix and is buried (Fig. 5). How can mutagenesis here cause activity loss? The answer is that more global conformational changes are induced by this amino acid change. Careful NMR studies indicated that the chemical shifts of residues in the entire interface between the first two helices were disturbed (Fig. 6), suggestive of extensive change in this area (indicated as darkened area in Fig. 5). The largest shift changes were found in the loop between the helices, roughly 15 Å distant from the mutagenesis site. As the figure shows,

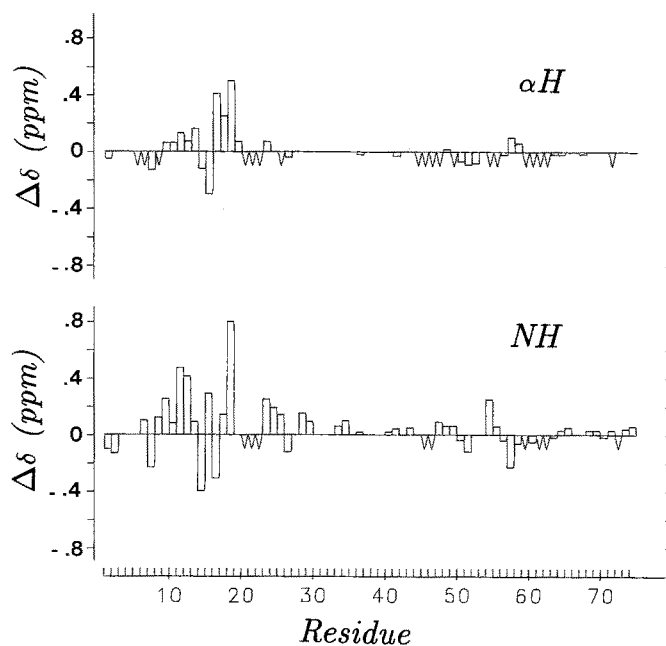


Fig. 6. The changes in chemical shift for H^α and H^N in the C5a mutant Ala²⁶→Met.

no other areas of significant shift changes were found; importantly, both regions already implied in activity, the C-terminus and the Arg⁴⁰ area, do not show any changes in NMR parameters. A key NOE defining the docking of the first helix to the remainder of the molecule, the NOE Ala¹⁰-Tyr²³ (residues shown in Fig. 5), was absent from the NOESY spectrum of the mutant. Taken together, these data strongly suggest that the docking of the first helix is affected and that possibly the helix has come loose in a time-shared manner. It is interesting that residue 26 in C3a is also a methionine and that the first helix in C3a seems loose as well. The combination of NMR data and mutagenesis thus showed that in addition to the C-terminus, Arg⁴⁰ and the N-terminus are important for receptor binding. Mutagenesis of solvent-exposed residues in the first helix, however, did not affect activity. It therefore appears that not the helix itself, but rather the loop region between the first two helices is involved and that the second receptor binding site in C5a is very extended and that it is noncontiguous. It is located at the top of the molecule in the view of Fig. 5. It must be emphasized that the identification of the turn region between the first two helices is based on what seems a dynamic rather than a static change in structure. It is interesting to speculate that this secondary site mimics the C-terminal site (or vice versa). It was found that the C-terminal residue Arg⁷⁴ is extremely important for activity in peptides derived from that area. It is thus possibly not coincidental that an arginine residue (Arg⁴⁰) is part of the 'top' site. We have obtained some preliminary mutagenesis data, showing that other residues in the top of the molecule are relevant for receptor binding, thus corroborating the notion of a secondary binding site at this location.

This information on the secondary site suggests the following strategy for drug design: potential drugs should be based on a scaffolding of extended structure to span the distance from Arg⁴⁰ to the turn between the first two helices, while the functional groups on this framework can be derived

from those found to be optimal in the C-terminus-derived peptides. The NMR work showed that the C-terminus of C5a is not well ordered in solution. One should study a peptide corresponding to this region when bound to the C5a receptor, using techniques such as transferred NOE or isotope editing (described below) to obtain useful information about this portion of the molecule.

In addition to information directly useful for the drug design process, the combined NMR and mutagenesis studies on C5a and C3a revealed that (i) solution and crystal structures can be different, sometimes in extremely important areas; (ii) mutagenesis of internal residues can give rise to widespread conformational changes that compound difficulties in interpretation of the biological consequences of these mutations; and (iii) these changes can be as subtle as changes in a conformational equilibrium in solution.

THE NMR STUDY OF THE STROMELYSIN CATALYTIC DOMAIN

Background

Human stromelysin belongs to a family of matrix metalloproteinases involved in the remodeling of the extracellular matrix. Stromelysin is a wide-spectrum zinc endoproteinase of MW 52 kDa that hydrolyzes collagens, gelatins, fibronectin, laminin and several cartilage proteoglycans. Stromelysin can participate in its own activation and in that of other matrix metalloproteinases such as collagenases and gelatinases (for reviews see References 20 and 21). It appears that elevated levels of matrix metalloproteinases in general and stromelysins in particular are associated with tumor metastasis and invasion, caused by degradation of the extracellular matrix [21–28]. Enhanced activity of stromelysin was also observed in tissues afflicted by rheumatoid arthritis [29] and osteoarthritis [30]. Stromelysin and the family of matrix metalloproteinase proteins as a whole are therefore important targets for antimetastasis and arthritis drug research.

No 3D structures are known for any of the matrix metalloproteinases to date, nor for any other protein that shares significant overall amino acid homology with these molecules. With its molecular weight of 52 kDa, stromelysin is beyond the current limits of NMR methodology for detailed studies. However, strong evidence supports the hypothesis that the N-terminal 19.5-kDa domain of the mature protein (Phe⁸³–Pro²⁵⁶) is necessary and sufficient for stromelysin catalytic activity [31–34].

The size of the stromelysin catalytic domain is within the limits of the new multidimensional multinuclear NMR methods. We describe here some aspects of the resonance assignment work currently being carried out. The ultimate goal is to determine the 3D structure in solution to help understand the architecture, catalytic mechanism, specificity, activation and *in vitro* inhibition of stromelysin. This knowledge will then be utilized to help in the design of synthetic inhibitors specific for this protein.

3D NMR of stromelysin

For molecules larger than 10 kDa, analysis of the structure with proton NMR as described for C5a becomes impossible. The reasons are threefold [35]: (i) Larger molecules display more resonances and therefore create spectral overlap problems. This point is already appreciated when inspecting the NOESY spectrum of C5a (Fig. 3): not much space is left for more cross peaks. (ii) With larger molecules, relaxation becomes more effective. The larger linewidths associated with this enhanced relaxation compound the overlap problem. (iii) Proton coherence transfer

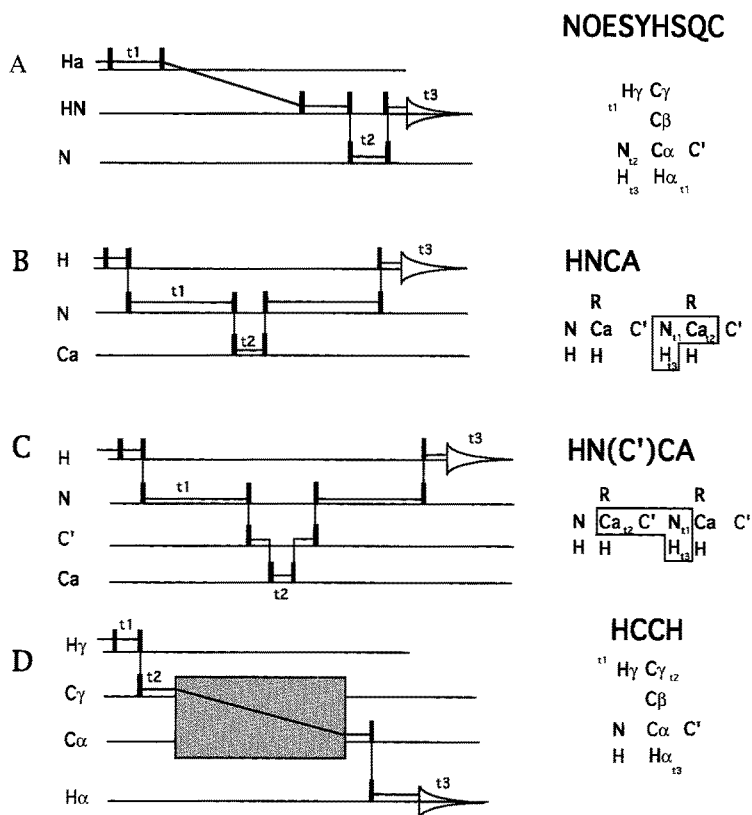


Fig. 7. Simplified pulse sequences for the four 3D NMR experiments discussed in the text. Black boxes show time demarcations; the bold lines indicate the flow of coherence. The coherence residences t_1 , t_2 and t_3 are also shown for the peptide fragments on the right. The hatched box in the HCCH experiment indicates coherence transfer by HOHAHA methods.

is slow as compared to the fast relaxation and becomes inefficient; for these molecules, experiments such as COSY display unacceptable low sensitivity.

The first problem is (partially) overcome by increasing the dimensionality of the experiments. It would seem that 3D proton experiments as discussed above would be good candidates to solve the problem. This is not the case, because such experiments are dependent on coherence transfer over ${}^1\text{H}$ - ${}^1\text{H}$ scalar couplings and will fail, except for extraordinary cases, on the basis of the third problem. The key development that enabled application of modern multidimensional NMR is the use of heteronuclear NMR techniques. Heteronuclear scalar couplings are generally large, thereby allowing efficient coherence transfer and thus allowing the construction of multidimensional experiments. A problem is that the naturally abundant carbon nucleus, ${}^{12}\text{C}$, does not have a magnetic moment and that the naturally most abundant nitrogen isotope, ${}^{14}\text{N}$, has spin 1 and displays very broad resonance lines. These nuclei need to be replaced by the stable isotopes ${}^{13}\text{C}$ and ${}^{15}\text{N}$, which are both high-resolution NMR nuclei. For proteins this can be accomplished relatively easily by expressing the cloned protein in *E. coli*, growing on a medium containing ${}^{15}\text{NH}_4\text{Cl}$ and glucose ${}^{13}\text{C}_6$.

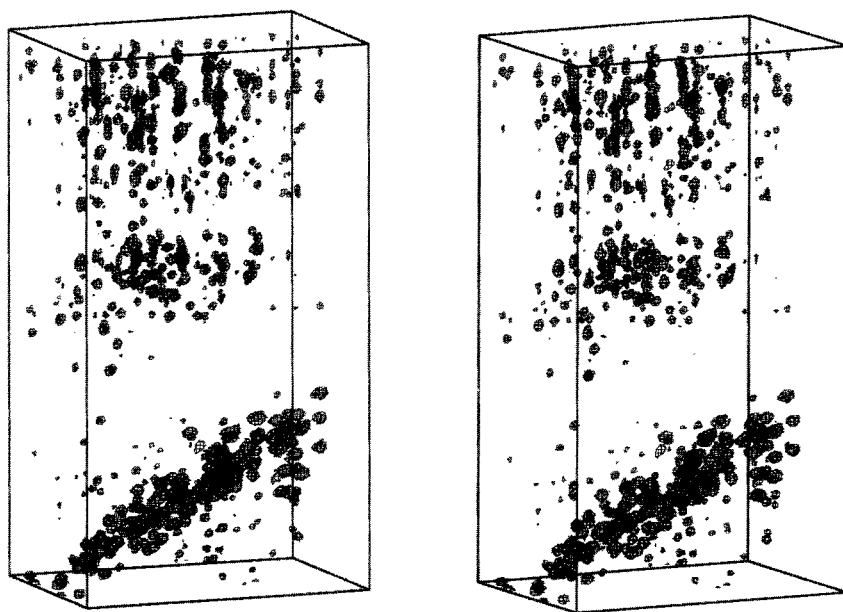


Fig. 8. Stereodialogram of the 3D ^{15}N -resolved NOESY-HSQC spectrum of the stromelysin catalytic domain. The width and height of the box represent proton dimensions; the depth is the ^{15}N -dimension.

The first heteronuclear 3D NMR experiment [36] was the 3D ^{15}N -resolved NOESY experiment. Even though it was carried out on a labeled tripeptide, it was recognized that the experiment is the key to the analysis of larger molecules by NMR. The principle of the 3D ^{15}N -resolved NOESY spectrum is straightforward when referring to Fig. 7, where we have drawn simplified representations of several key heteronuclear 3D NMR experiments. We trace the coherence path of a $d\alpha\text{N}(i,i+1)$ NOE, an NOE between the H^α of residue i and the H^N of residue $i+1$, in the 3D NOESY-HSQC (HSQC means Heteronuclear Single Quantum Correlation) experiment (Fig. 7A). Coherence starts on the H^α (t_1), and is transferred by NOE to the $\text{H}^\text{N}(i+1)$. Before the frequency of the H^N is measured, the coherence is transferred to the attached ^{15}N nucleus, the frequency of which is sampled (t_2). The coherence is then brought back to the H^N where its frequency is finally measured (t_3). This rather complicated pathway is followed for reasons of sensitivity. A similar path is followed by the NOE between the H^α and its own H^N ; coherence starting on $\text{H}^\alpha(i+1)$ will be transferred to $\text{H}^\text{N}(i+1)$ as well. As a result, all protons reporting by NOE to a specific H^N will be labeled with the amide nitrogen frequency of that amide proton. One thus obtains all these NOEs on a 'skewer' at frequency coordinates $(f_2, f_3) = (^{15}\text{N}, \text{H}^\text{N})$ in 3D space. An ensemble of skewers is seen on a particular plane with a particular ^{15}N frequency. The experiment is extremely powerful and works well on even the largest molecules. The reason is that the ^{15}N - H^N scalar coupling is very large (90 Hz) so that coherence can be quickly transferred without much relaxation loss. The ^{15}N resonances themselves are very narrow (around 8 Hz for proteins of 20 kDa) and relaxation is limited when the ^{15}N frequency is sampled. As a result, the ^{15}N -resolved NOESY-HSQC experiment has a sensitivity close to that of the parent NOESY experiment for proteins up to 20 kDa [36–38]. Figure 8 shows a stereoplot of the 3D NOESY-HSQC experiment as carried out for the stromelysin catalytic domain. Data in such spectra are

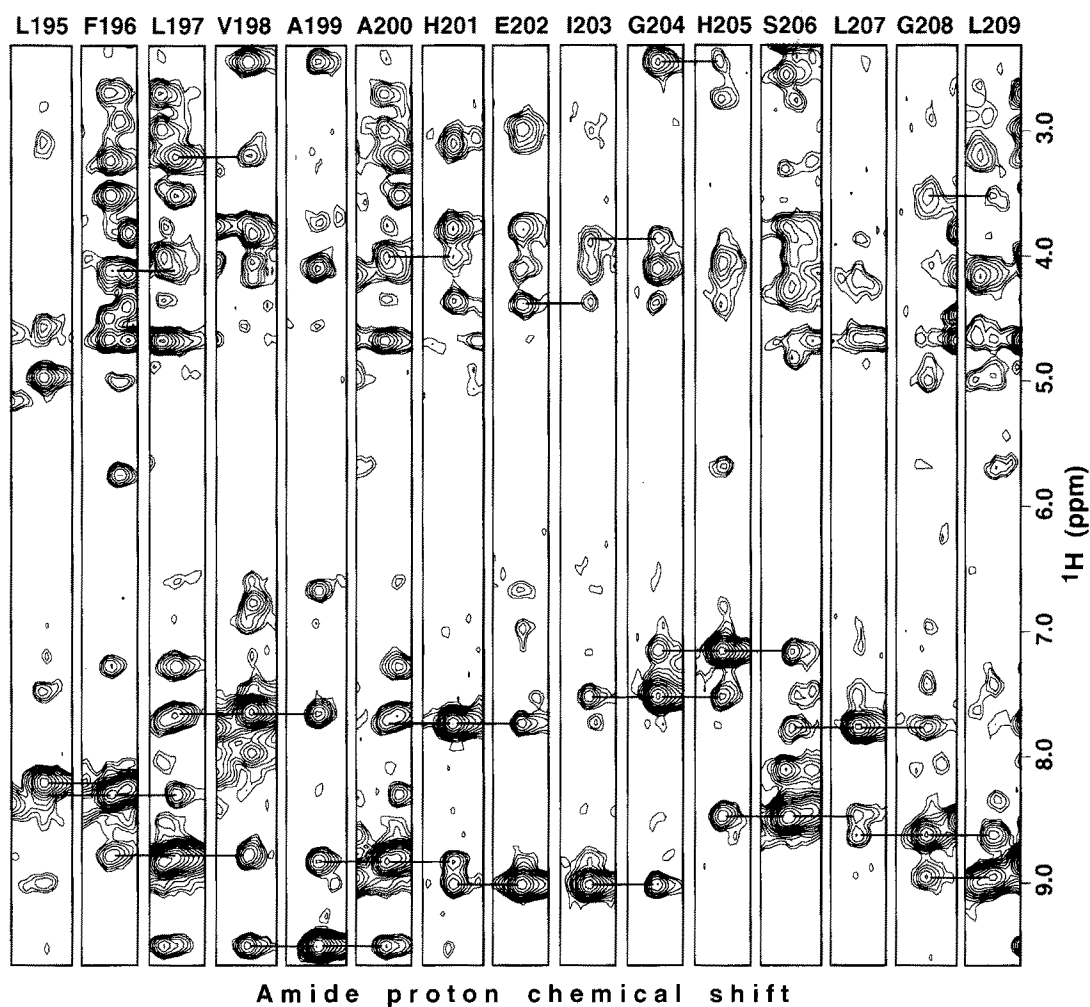


Fig. 9. The alignment of the ^1H -amide 'skewers' at the corresponding ^{15}N frequency from the 3D ^{15}N -resolved NOESY-HSQC spectrum. The data were collected with a 1.5-mM sample of ^{15}N -labeled stromelysin catalytic domain. The sequential alignment of the residues forming the active-site helix (residues Leu¹⁹⁵ through Leu²⁰⁹) is shown. Sequential $d\text{NN}(i, i \pm 1)$ and $d\alpha\text{N}(i, i + 1)$ NOEs are indicated with thin lines.

analyzed by taking 2D cross sections perpendicular to the ^{15}N axis. The resulting planes look like ^1H - ^1H NOESY spectra of small molecules and are virtually devoid of overlap. However, skewers belonging to adjacent residues will generally not lie in the same plane. The entire 3D matrix must thus be scanned for corresponding skewers. Corresponding skewers can be 'pasted' on a single plane and compared on the computer screen [39]. A plot of a sequence of adjacent skewers is shown in Fig. 9. The NOEs $d\text{NN}(i, i \pm 1)$ and $d\alpha\text{N}(i, i + 1)$ as drawn confirm that these skewers indeed belong to amino acid residues adjacent in sequence.

A similar 3D ^{15}N 'editing', 'resolving' or 'separating' technique can be applied to through-bond ^1H - ^1H correlation (HOHAHA) experiments [35,39]. Thus in a HOHAHA spectrum, skewers in 3D space contain the more-or-less complete spin systems of the amino acid residues. A combined

analysis of NOESY-HSQC and HOHAHA-HSQC will yield the sequential assignment: spin-system identification cross peaks obtained from HOHAHA-HSQC are aligned per corresponding skewer to identify the intraresidue cross peaks in NOESY-HSQC. The remaining cross peaks in the NOE skewers contain the sequential cross peaks from which the assignment can be obtained. This (semiconventional) method of combined analysis works well only for some larger proteins which have exceptionally small linewidths and contain much β -structure (because of sensitivity limitations of the HOHAHA experiment itself). For stromelysin, the HOHAHA-HSQC experiment turned out to be of insufficient sensitivity for practical use.

Structure determination of larger proteins using ^{13}C -resolved 3D and 4D NMR

Before describing the relatively involved alternative methods for the assignment of the resonances of larger proteins such as stromelysin, we want to sketch the entire process of structure determination of larger molecules. Assuming that resonance assignments can be obtained, one wants to proceed by identifying NOEs for structure calculation. Many of these NOEs will come from side-chain–side-chain interactions that resonate in the most crowded part of the 2D NOESY spectrum (see Fig. 10A). ^{15}N -editing will not help here, since the protons are not ^{15}N -bound. By labeling the molecule with ^{13}C , the NOESY data can be pulled apart with respect to the ^{13}C chemical shifts (Fig. 10B). Considerable simplification is achieved in a fashion completely analogous to that described for ^{15}N -resolved NOESY spectra [40]. However, it was found that the resolution is still insufficient in three dimensions and the experiment was therefore extended to a 4D doubly ^{13}C -resolved NOESY experiment [41–43]. It was found that the 4D NOESY spectrum is essential for structure determination of larger proteins. The consequence is that the ^{13}C resonances have to be assigned as well. Once that process is completed, structure determination can proceed on the basis of the identified NOEs as previously described (distance geometry/molecular dynamics/simulated annealing [11–15]).

Through-bond main-chain-directed assignments for stromelysin

These methods are referred to as triple-resonance assignment methods and can only be carried out for proteins that are uniformly and 100% labeled in ^{13}C and ^{15}N . The majority of these new techniques were pioneered and developed in the laboratory of Bax [44]. The objective in triple-resonance experiments is to transfer magnetization over the backbone atoms and to measure the frequency of the nuclei involved in the transfer. Coherence transfer from nucleus A to nucleus B is achieved by COSY-type mechanisms involving heteronuclear scalar couplings. Most triple-resonance experiments involve the HN unit from the peptide linkage as anchor. In the HNCA experiment (Reference 45, Fig. 7B), coherence starts at the amide proton and is transferred to the ^{15}N nucleus, where the frequency is sampled (t_1). The coherence is further transferred to the C^α nucleus and that frequency is sampled. The coherence is now pulsed back over the same pathway to the same amide proton for which the FID is detected (t_3). The experiment is carried out in this complicated way for reasons of sensitivity, related with magnitudes of the scalar couplings, the relaxation characteristics of the different nuclei involved and exploitation of Boltzmann polarization transfer. After 3D Fourier transformation one obtains one 3D cross peak for every amino acid residue (except proline) whose three frequency coordinates are the amide nitrogen, the α -carbon and amide proton in f_1 , f_2 and f_3 , respectively. In the related 3D HN(CO)CA experiment (Reference 46, Fig. 7C) (which takes the coherence in a different direction and uses one additional

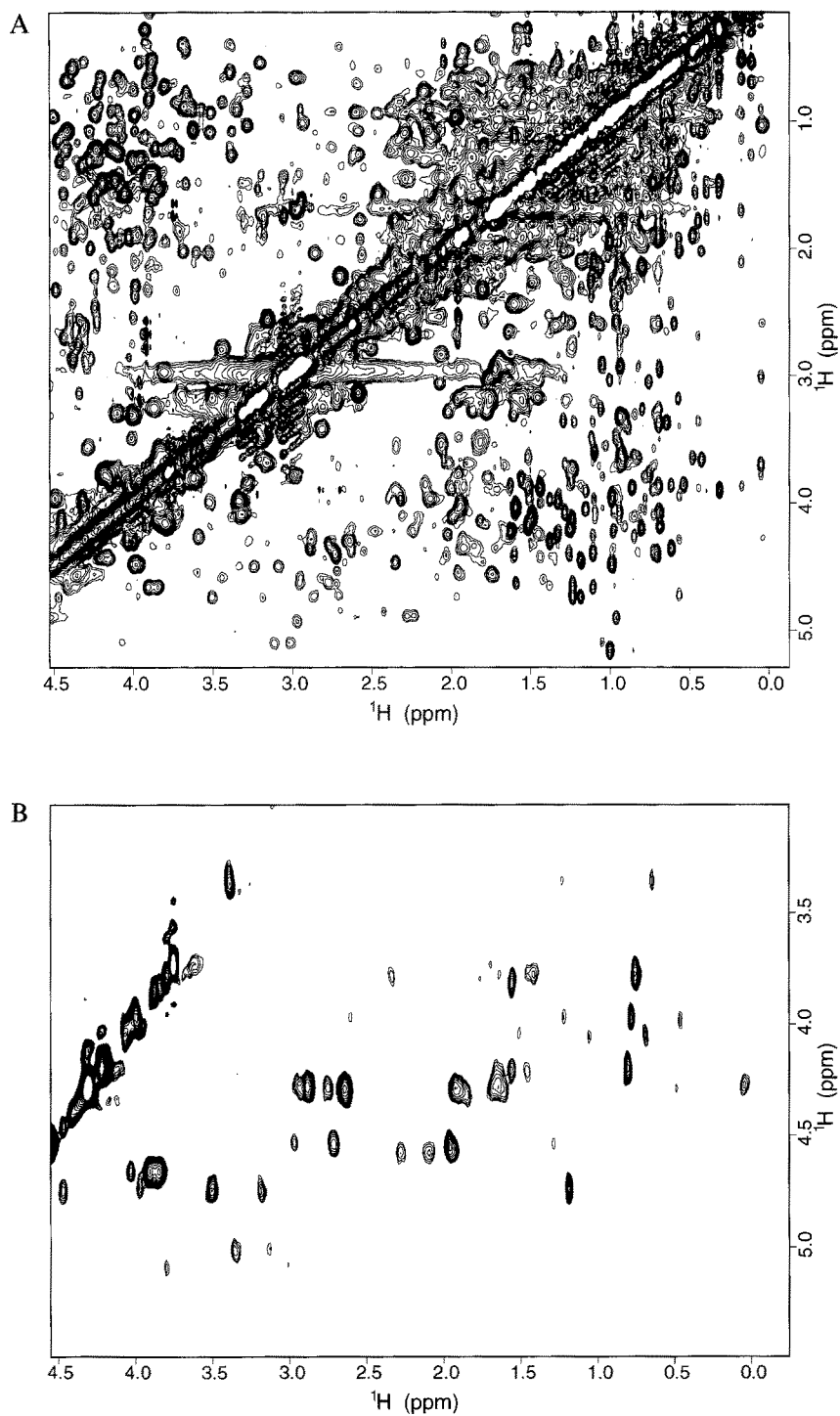


Fig. 10. The 3D ^{13}C -resolved HSQC-NOESY experiment of a sample of ^{13}C -labeled T4-lysozyme (19.5 kDa). (A) A projection of the data. (B) A ^1H - ^1H plane through the data showing the achieved simplification.

step), the frequency coordinates are the amide nitrogen of residue i , the α -carbon of residue $i - 1$ and the amide proton of residue i in f_1 , f_2 and f_3 , respectively.

Accordingly, the 2D H^N shift-correlation map is correlated with the C^α chemical shifts of the same and the previous amino acid residues in the 3D HNCA and HN(CO)CA experiments, respectively. Thus, theoretically, all H, N and C^α resonances in the protein can be interconnected with these two experiments and be assigned. In practice, however, this is not possible due to overlap in the chemical shifts of the C^α resonances. In fact, the linkage from amino acid to amino acid is only a 1D connection and much overlap occurs. Therefore, additional pathways need to be exploited as well. A combination of 3D HN(CA)HA [47] and HA(CACO)NH [48] worked well for stromelysin to give another 1D linkage between the amino acids but now over H^α . Thus, for a given amino acid linkage, as long as both H^α and C^α are not simultaneously degenerate with another H^α and C^α pair, the connectivity can be derived from the four experiments in conjunction. Figure 11 shows a double 3D connectivity path using the two pairs of 3D experiments described above for the stromelysin catalytic domain. For every peptide linkage a computer search routine identified a C^α -(CO)- H^N - C^α 'tee' formed from the 3D HNCA and HN(CO)CA experiments as shown in the upper row, or an H^α (C^α CO) H^N (C^α) H^α tee from 3D HN(CA)HA and HA(CACO)NH as shown in the lower row. The tees are connected with corresponding C^α resonances for the top row and corresponding H^α resonances for the bottom row. The requirement is, of course, that both pathways delineate the same amino acid connections as indicated.

The connectivity pathways generally do not reach from one side of the protein to the other, because of locally unfavorable relaxation conditions and/or breaks due to the presence of proline residues. Therefore, a method is needed to establish the positions of the fragments in the protein. A particularly elegant way is found by using residue-specific labeling [49]. For stromelysin, we have used the protein with only phenylalanine, alanine or leucine ^{15}N -labeled. Proton- ^{15}N chemical-shift correlation maps of these constructs thus gave the coordinates of the ^{15}N - 1H shifts for these residues. Subsequently this information is used to assign amino acid type identification to two of the three dimensions of some of the cross peaks of the triple-resonance experiments. Using this very limited information, all triple-resonance-connected peptide fragments could be placed in the sequence [50].

NMR assignments can already help in understanding protein structure: Possibilities for large proteins

The NMR (main-chain) assignment procedure is an integral part of a full structure determination process. Nevertheless, the obtained assignments can already be used to aid in the understanding of the protein structure and function before the structure determination is complete. First, it is only a small step to scan for the secondary-structure NOEs once all H^α and H^N protons have been assigned [5]. For larger proteins the process is essentially the same as that illustrated for C5a. The availability of experimentally determined secondary structure is of great importance for homology modeling and for theoretical protein-structure calculations: the positions of deletions and insertions, generally between the elements of secondary structure, will become apparent, helping in the alignment of homologous structures. The availability of the secondary structure may also reveal homology previously gone undetected, thus allowing modeling efforts to start early [50]. Secondly, the assigned spectra may function as a monitor for structural alterations upon mutagenesis of chemical modifications; generally one will use the ^{15}N - 1H correlation map

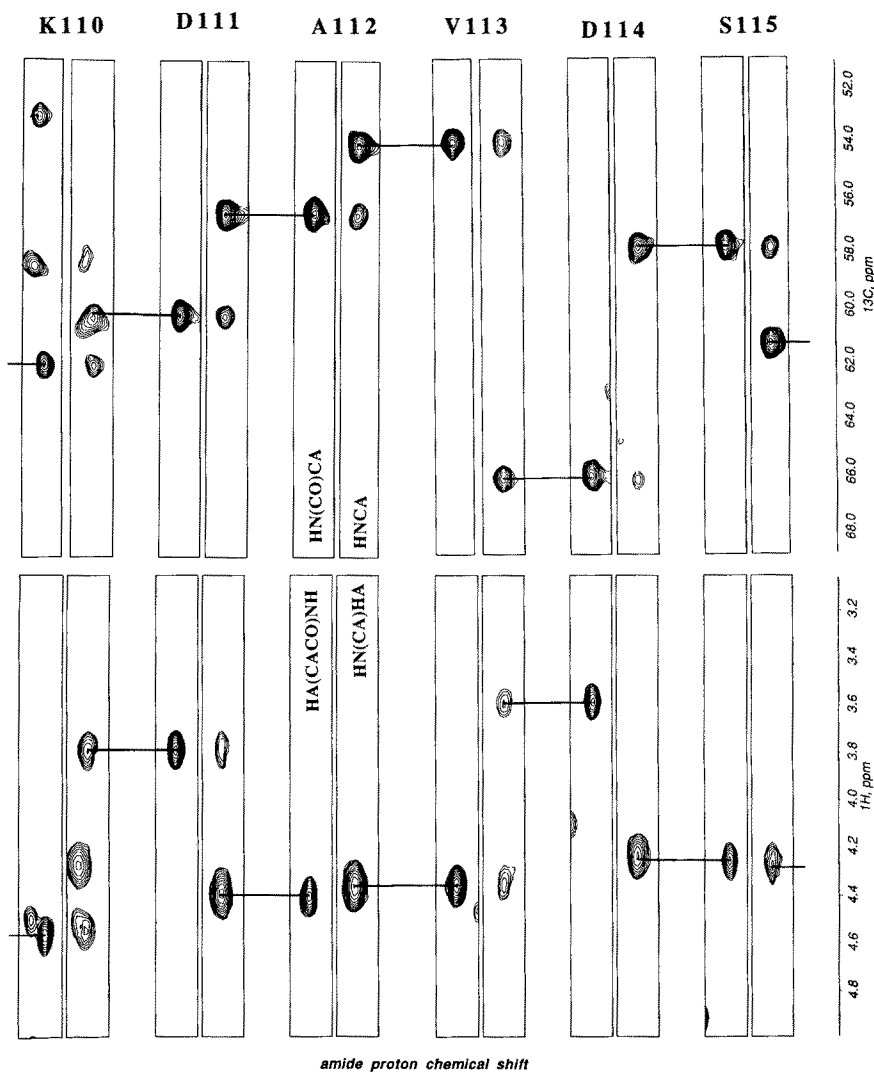


Fig. 11. Alignment of the results from the 3D HNCA, HN(CO)CA, HN(CA)HA and HA(CACO)NH experiments for the assignment of several of the main-chain resonances of the stromelysin catalytic domain. In the upper half of the figure alternating (^1H , $^{13}\text{C}^\alpha$) planes of HNCA and HN(CO)CA spectra at particular ^{15}N frequencies are shown as exemplified for Ala¹¹². The ^{15}N planes and the H^N shift ranges were taken at the frequencies corresponding to the residues shown at the top of the figure. The HNCA/HN(CO)CA pairs identify the transpeptide $\text{C}^\alpha(i-1)\text{N}(i)\text{H}(i)\text{C}^\alpha(i)$ tees. The tees are connected sequentially as indicated. The lower half of the figure shows a similar construction for the $\text{H}^\alpha(i-1)\text{N}(i)\text{H}(i)\text{H}^\alpha(i)$ tees and their connections.

for these purposes. These maps can also be used to monitor intermolecular interactions, i.e., shift changes in resonances upon binding of ligands can be used to suggest locations of the interface of interaction. Finally, the availability of assignments is sufficient to obtain dynamic information from heteronuclear relaxation measurements [51] and interpret the results on a structural level when a structural model can be postulated.

All of the above uses of main-chain assignments also apply to very large systems for which full structure determination may not (yet) be possible because of problems in obtaining side-chain assignments (see next paragraph). Three-dimensional models on the basis of secondary structure may suggest spatial proximities between residues that could be tested using specific isotopic labels. An interactive process between modeling and experimentation can develop in this way. Modern NMR spectroscopy can thus help in understanding the structure and function of these large molecules. Modeling based on the obtained results will then possibly contribute to drug design processes.

Three-dimensional structure determination of larger proteins: Side-chain assignments

As argued above for the case of C5a, side-chain assignments are necessary for experimental structure determination of any detail. Again, heteronuclear NMR methods are used for this process. The one-bond couplings between ^{13}C nuclei are between 34 and 55 Hz. ^{13}C lines are approximately 20–30 Hz wide for proteins in the 20 kDa range. This means that through-bond information, necessary for the side-chain assignments, can be moved over the ^{13}C – ^{13}C bonds. These so-called HCCH experiments (Fig. 7D) are carried out [52–54] by starting magnetization on e.g. a $^1\text{H}^\gamma$ nucleus and measuring its frequency (t_1), then transferring it to the attached $^{13}\text{C}^\gamma$ and measuring its frequency (t_2). From there, the coherence is pushed by ^{13}C – ^{13}C transfer (HOHAHA) over the ^{13}C side-chain skeleton to e.g. $^{13}\text{C}^\alpha$. At the end of the trip the coherence is transferred out to the attached H^α proton where the FID is measured (t_3). As a result, one obtains ^1H – ^1H transfer edited by ^{13}C frequencies. The experiment serves three purposes: (i) it makes through-bond correlation possible; (ii) resolution is obtained because the experiment is carried out in three or four dimensions; and (iii) spin-system assignments are obtained for the side-chain proton and ^{13}C resonances simultaneously. It is relatively straightforward to interpret the HCCH data after the assignment of the main-chain resonances. Note that we have reversed the order of proceeding in assignment here: in the conventional 2D or semiconventional ^{15}N -resolved 3D methods, one starts with side-chain identifications and ends with linking them up using NOEs; here we first link the main chain and then connect up the side chains. New developments in this area focus on the integration of main-chain and side-chain assignments. In principle, magnetization may start on the peripheral protons of the side chain, transferred to the carbon frame of the side chain, transferred to the C^α resonance by ^{13}C – ^{13}C COSY-like transfer, followed by transfer to the other main-chain atoms. Most new experiments correlate the side-chain carbon or proton chemical shifts or both of residue i with the amide $^{15}\text{N}/^1\text{H}$ pair of residue $i + 1$ [55,56].

Once side-chain resonance assignments have been obtained, 3D structure determination proceeds by identifying the NOE cross peaks in 3D and 4D ^{13}C -resolved NOESY spectra. The obtained distance restraints are used in the structure-determination programs as usual. Using these methods, several interesting structures and complexes of larger molecules have been determined: calmodulin [57], interleukin 1β [58] and the complex of cyclosporin A/cyclophilin [59]. From there, the path to drug design will follow the same lines as described for C5a; first, biological data will be placed into a structural context. Mutagenesis experiments can be suggested and monitored by NMR. For instance, with stromelysin we anticipate being able to identify and place in 3D space the essential active-site residues. On this basis, a lead inhibitor can be suggested and its complex with the protein studied, after which one hopes to enter an iterative cycle of structure determination and drug improvement.

PROBLEMS WITH NMR OF SMALL LIGANDS

It is clear from the descriptions above that NMR analysis of larger proteins becomes much more involved than the analysis of smaller proteins. One would thus assume that conformational analysis, and interactive drug design, must be very easy for small ligands such as (unrestrained) peptides. It turns out, however, that this is not true at all: although the techniques involved are relatively simple (^1H - ^1H COSY-type experiments and NOESY or ROESY experiments work well), the interpretation of the conformational data is not. The problem is that small, nonglobular molecules are often in dynamical equilibrium between different conformations. An example was described above for the C-terminal peptide of C5a; simultaneous appearance of $d\alpha\text{N}$ and $d\text{NN}$ NOEs is not compatible with a single structure. Currently, it is possible to characterize the dynamic state better by measuring many heteronuclear three-bond couplings [60] in a labeled version of the peptide, allowing the formulation of a dynamic model compatible with scalar and NOE information. But even if one would have determined this equilibrium, necessarily involving very small free-energy differences between the substates, one would not know which substate would be relevant to receptor binding. In principle one *never* knows this: it is not very likely that one will recognize the presence of a minor conformation (say 10%) in an ensemble from the NMR data, especially when that conformation is in rapid dynamic equilibrium with the main conformation. However, the free-energy difference of 1.4 kcal/mol between these states may be easily overcome by the free energy of binding; the bound state is then different from the perceived state in solution. Even the NMR analysis of constrained molecules may lead to wrong conclusions as was shown by Fesik et al. [61] in their studies of cyclosporin A.

Clearly, the only way of knowing the conformation of a small ligand when bound to a larger molecule is to determine its structure when it is in that state. Now the small-molecule NMR becomes a large-molecule NMR problem again. A straightforward way is to determine the structure of the entire complex, but the size of most of these complexes is beyond the upper limits of even modern NMR methods. In the following sections, the two different approaches that cope with these problems, transferred NOE and isotope editing, will be discussed.

Transferred NOE

The transferred NOE techniques have been analyzed and popularized mainly by Clore and Gronenborn [62,63]. The transferred NOE occurs in equilibrium mixtures where a small ligand is in excess over a larger molecule. As an example, we describe a study of the binding of an α_2 -adrenergic receptor-derived peptide to G_o -protein (Neubig and Zuiderweg, unpublished results). The peptide, named 'Q', is modeled after the carboxy-terminal portion of the putative third cytoplasmic loop of the receptor. This region of the receptor is involved in coupling to the G_o -protein. Free in solution, the peptide shows hardly any NOE intensities (Fig. 12), indicating the lack of spatial proximities in the peptide, an NOE-ineffective correlation time, or both. When only 2% G_o -protein is present, clear NOEs are observed between the different residues of the peptide spectrum (Fig. 12). The explanation for the results is that NOEs build up between spatially close protons in the peptide when the peptide is transiently bound to the G-protein. When the exchange kinetics is faster than the disappearance of the NOE by relaxation processes in the free state, NOEs will build up according to the proximities in the bound conformation every time the peptide binds. NOEs will also build up according to the proximities in the free conforma-

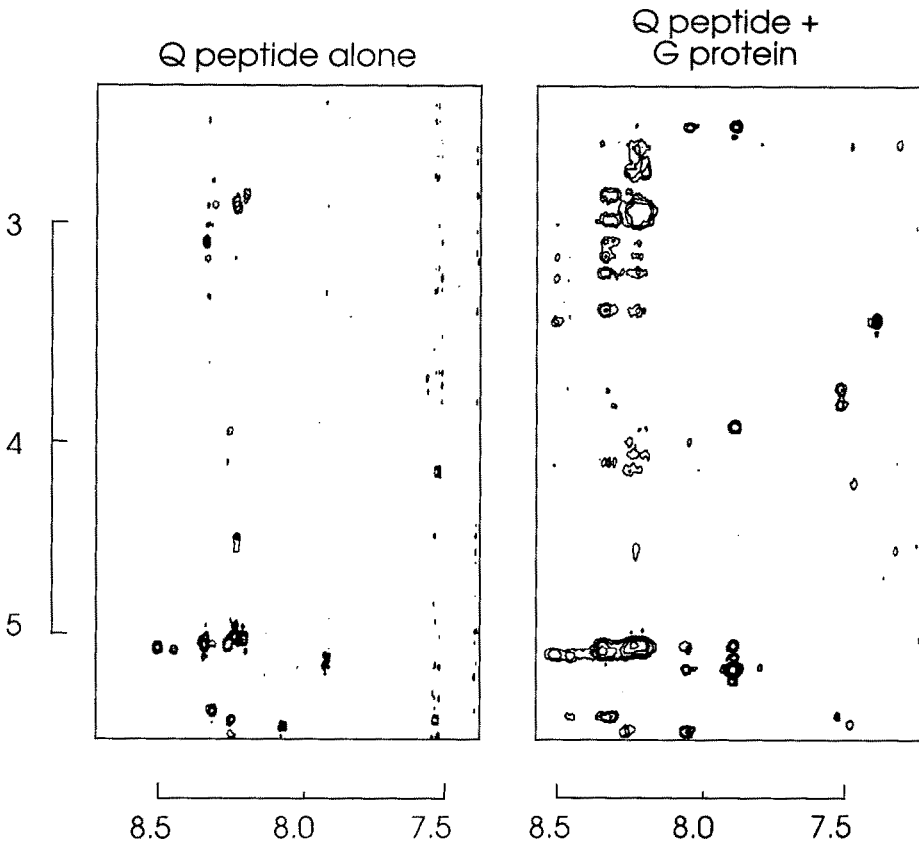


Fig. 12. Transferred NOE spectra of the Q-peptide of the α -adrenergic receptor. Left: Q-peptide alone. Right: Q-peptide in exchange with G_o -protein.

tion every time the peptide is free. The resulting NOE is thus a time-average between bound and free states. Since the free state does not display much NOE by itself, the observed NOEs must be due to proximities in the bound state, even though this occurs only 2% of the time. The elegance of the transferred-NOE experiment is that information on the bound state can be obtained from a relatively easily analyzed NMR spectrum that is closely related to the free-state spectrum. The spectrum will generally be a fast-exchange chemical-shift average between free and bound shifts (but it can also be slow exchange on the chemical-shift scale as long as the exchange rates are faster than the relaxation). An additional advantage of the transferred-NOE method is that only relatively small quantities of the receptor are needed. However, a major problem with the transferred-NOE experiment is related to this very same advantage: the excess peptide may bind nonspecifically to many low-affinity sites on the receptor and the observed transferred NOEs may thus very well reflect these nonspecifically bound states. It is thus of extreme importance to carry out the proper control experiments to rule out such occurrences. A good literature example is that given by Behling et al. [64], who studied the conformation of acetylcholine when bound to the nicotinic receptor using transferred NOE. These authors had at their disposal α -bungarotoxin, which is known to compete for the same binding site on the receptor as acetylcholine. Upon

addition of the α -bungarotoxin to the sample containing the acetylcholine–receptor mixture, the transferred NOEs disappeared, proving that they were due to proximities in the specific interaction state. We are currently in the process of identifying such controls for the Q-peptide G-protein. Irrespective of the nature of the binding in our example, the data in Fig. 12 clearly show strong $d\alpha N$ and negligible to weak dNN connections, suggesting that the peptide is in an extended conformation in this binding mode.

Isotope editing

Another approach that studies more directly the bound conformation involves the use of isotopes. Much of this approach has been developed in the laboratory of Fesik [65]. In the very early days of NMR, isotope editing was already carried out by, for instance, using ^{13}C - or ^{19}F -labeled ligands and observing shift and/or relaxation changes of these nuclei upon binding. Nowadays, one is generally not content with only shift information; in order to obtain conformational information on the bound state, one wants to detect ^1H – ^1H NOEs between ligand protons in the bound state, and possibly also between protons of the ligand and the receptor in order to characterize the interface between the molecules.

A straightforward but expensive way to obtain such information is to fully deuterate the larger molecule; the only proton signals remaining are those from the bound ligand and NOESY spectra of the ligand only can be obtained. This method was carried out to study mellitin when bound to perdeuterated calmodulin [66]. The advantage of this technique is that simple proton NMR methods *can* be used. Thus, NOESY structure determination of the bound ligand can be performed. The obtained information may be very valuable for drug design purposes, as the ligand is studied in its relevant state. The disadvantage of the perdeuteration approach is that simple proton NMR methods *must* be used. The powerful heteronuclear 3D and 4D methods are excluded and severe restrictions exist on the size of molecules that can be handled. The other drawback is that it is difficult to generate large quantities of perdeuterated proteins, since growth of *E. coli* in the toxic D_2O medium is generally quite modest.

A second approach using deuterium for isotopic editing involves labeling of the ligand. Two experiments are recorded: (i) a NOESY experiment using a normal, proton-labeled ligand bound to a normal, proton-labeled receptor and (ii) a NOESY experiment using deuterated ligand, also bound to its proton-labeled receptor [67]. The 2D spectra are subtracted and the difference spectrum shows only NOEs between ligand protons and between ligand protons and receptor protons. Again, structural information about the relevant bound state is obtained. In addition, NOEs between ligand and receptor are identified, showing which ligand groups are directly involved in binding. This information is obviously invaluable for drug design processes, even when the receptor groups involved in this process cannot be identified (easily). The deuterated-ligand approach has the same advantage/disadvantage as the deuterated-receptor approach described above, in that proton NMR can and must be used, therefore limiting the application to smaller complexes only. Nevertheless, quite useful information could be obtained for the bound state of a peptidic inhibitor bound to pepsin [67]. A possible advantage over the perdeuterated-macromolecule approach is that it is relatively easy to prepare the deuterated ligand if it is a natural peptide (using solid-state synthesis with deuterated amino acids).

The third and most powerful use of isotopes is isotope-edited NMR using ^{15}N - and/or ^{13}C -labeled ligand. This method was actually a predecessor of the heteronuclear 3D experiments.

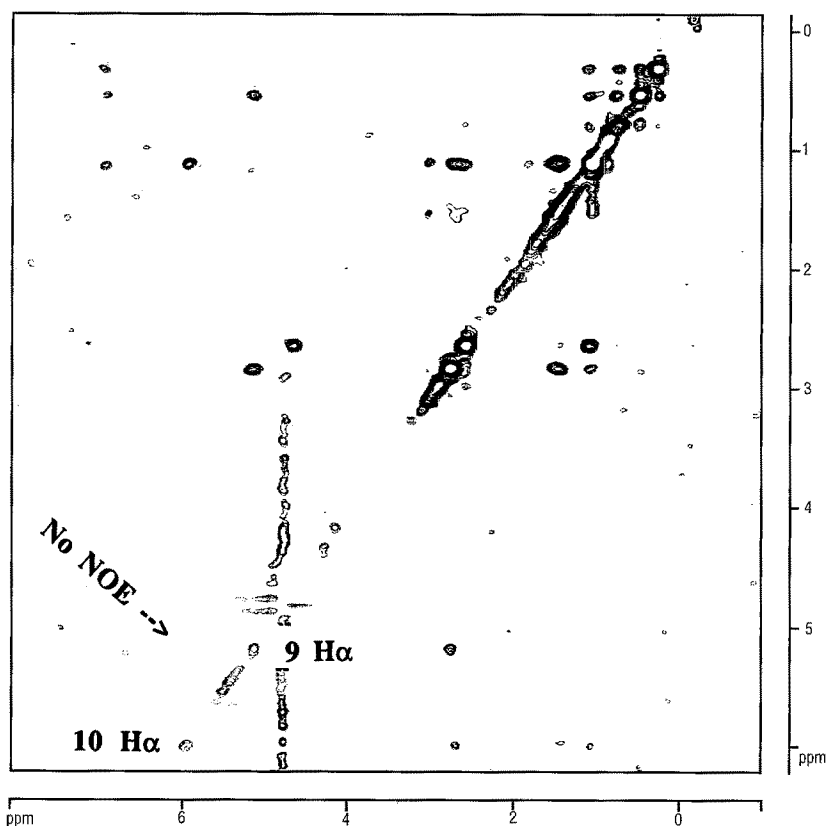


Fig. 13. ^{13}C -edited NOESY spectrum of 9,10 ^{13}C -labeled cyclosporin A bound to unlabeled cyclophilin [56]. The data were obtained using a sample kindly provided by Dr. S.W. Fesik (Abbott Laboratories).

Basically, the same technique as the NOESY-HSQC experiment can be used to edit or filter the ^1H subspectrum of an isotopically labeled ligand out of the ^1H spectrum of the complex. In this method one alternates the phase of the r.f. pulses, affecting the heteronuclei on every other scan [65,68]. Protons bound to these heteronuclei through one-bond scalar coupling will sense these phase alterations and can be specifically detected by difference spectroscopy since signals of protons bound to nuclei not affected by these pulses are subtracted out. An example is shown in Fig. 13, where the proton NMR spectrum of partially ^{13}C -labeled cyclosporin A (CsA) bound to cyclophilin was made visible using ^{13}C -isotope edited 2D NOE [61]. This type of isotopic editing is very powerful, as it can also be executed as heteronuclear 3D and 4D experiments. Therefore, larger systems can be analyzed with these techniques than possible with the perdeuteration methods. A very important discovery with respect to the structure of cyclosporin was made with the help of the spectrum in Fig. 13, which illustrates the importance of studying bound ligands. The crystal structure of CsA [70] as well as the NMR structure of free CsA in solution [70] (chloroform) show the presence of a cis peptide bond between residues 9 and 10. In the NMR data, this conformation was evidenced by a strong NOE cross peak between the α -protons of residues 9 and 10. However, in the NOESY spectrum of CsA bound to cyclophilin (in aqueous

solvent) no NOE cross peak is observed between the H^α protons of these residues, as indicated in the figure. This is very strong evidence that this peptide bond is trans in the complex, thus contrasting with the conformation as obtained from free CsA in chloroform solution and in the crystal. (Recent NMR investigations of a more soluble CsA analog indicate that a trans peptide bond may prevail for this related free molecule in solution when dissolved in H₂O [71].)

It is clear that information of this type is indispensable for successful drug design. Fesik's group went on to determine the full 3D structure of fully ¹³C-labeled CsA while bound to cyclophilin, using (isotopic-filtered) heteronuclear 3D and 4D methods [59]. The NMR approach for the study of this small molecule has here become identical to that used for larger molecules.

CONCLUSIONS

NMR spectroscopy of proteins and peptides can aid in the drug design process in several different ways. For small proteins (MW < 10 kDa), full 3D structures can be determined with relative ease. The 3D structure is used to help understand biochemical and genetic experiments on the protein; ultimately the relevant part of the 3D structure can serve as a template for the design of drugs. For larger proteins (10 < MW < 30 kDa) the NMR assignment process is more difficult, involving many multidimensional NMR experiments using isotopically labeled proteins. Nevertheless, the structure determination methods are all laid out for this class of molecules and are basically not different from those used for the smaller molecules. The drug design process can still be based on the full structure and possibly of the complex of the drug and the protein. For even larger molecules, NMR studies may not proceed beyond the assignment and secondary-structure stage. However, the obtained information may be used to aid in computer modeling procedures, to monitor binding studies and to gather dynamic information. Drug design will thus indirectly follow from the NMR data through modeling. The conformation of many small peptidic ligands is difficult to characterize in solution, because of dynamic averaging processes. Therefore, these ligands are best studied when interacting or complexed with the macromolecular receptor. Several NMR techniques that were developed to obtain structural information in these heterogeneous systems were discussed. Direct valuable information on the conformation of the bound drug can be obtained with these methods.

The presence of dynamic equilibria between different conformations can be detected by NMR, although the exact characterization of the states involved is not straightforward. This dynamic NMR information shows that the conformation of a protein in a crystal can be biased. It also reveals that site-directed mutagenesis can induce subtle conformational changes that are dynamic rather than static in nature.

ACKNOWLEDGEMENTS

E.Z. thanks the many former colleagues at Abbott Laboratories who were involved in the C5a work: Drs. D. Nettlesheim, W. Mandrecki, K. Mollison, J. Henkin, J. Greer and G. Carter. E.Z. also thanks Drs. S. Fesik, E. Olejniczak, R. Gampe and A. Petros for the many years of collaboration in the development of heteronuclear 3D and 4D NMR. We thank E. Fauman for help with Fig. 5. Some of the research described here is sponsored through NSF grant MCB-9218573. The Program in Protein Structure and Design at the University of Michigan is acknowledged for

additional financial aid to the stromelysin project. R.N. acknowledges NIH support (grant HL46417).

REFERENCES

- 1 Noggle, J.H. and Schirmer, R.E., *The Nuclear Overhauser Effect: Chemical Applications*, Academic Press, London, 1971.
- 2 Ernst, R.R., Bodenhausen, G. and Aue, W.P., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
- 3 Hugli, T.E., *Crit. Rev. Immunol.*, 1 (1981) 132.
- 4 Huber, R., Scholtze, H., Paques, E.P. and Deisenhofer, J., *Hoppe-Seyler's Z. Physiol. Chem.*, 250 (1980) 1389.
- 5 Wüthrich, K., *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY, 1986.
- 6 Zuiderweg, E.R.P., Mollison, K.W., Henkin, J. and Carter, G.W., *Biochemistry*, 27 (1988) 3568.
- 7 Zuiderweg, E.R.P., Kaptein, R. and Wüthrich, K., *Proc. Natl. Acad. Sci. USA*, 80 (1983) 5837.
- 8 Wagner, G., *Q. Rev. Biophys.*, 16 (1983) 1.
- 9 Zuiderweg, E.R.P., Nettesheim, D.G., Fesik, S.W., Olejniczak, E.T., Mandrecki, W., Mollison, K.W., Greer, J. and Carter, G.W., In Live, D., Armitage, I. and Patel, D. (Eds.) *Frontiers of NMR in Molecular Biology*, A.R. Liss, Inc., New York, NY, 1990, pp. 75-87.
- 10 Zuiderweg, E.R.P., Nettesheim, D.G., Mollison, K.W. and Carter, G.W., *Biochemistry*, 28 (1989) 172.
- 11 Braun, W. and Gö, N., *J. Mol. Biol.*, 186 (1985) 611.
- 12 Guntert, P. and Wüthrich, K., *J. Biomol. NMR*, 1 (1991) 447.
- 13 Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F., *J. Mol. Biol.*, 182 (1985) 179.
- 14 Clore, G.M., Brünger, A.T., Karplus, M. and Gronenborn, A.M., *J. Mol. Biol.*, 191 (1986) 523.
- 15 Brünger, A.T., *XPLOR Molecular Dynamics*, Yale University, New Haven, CT, 1991.
- 16 Nettesheim, D.G., Edalji, R.P., Mollison, K.W., Greer, J. and Zuiderweg, E.R.P., *Proc. Natl. Acad. Sci. USA*, 85 (1988) 5036.
- 17 Edalji, R.P., Mollison, K.W., Zuiderweg, E.R.P., Fey, T.A., Krause, R.A., Conway, R.G., Miller, L., Lane, B., Henkin, J., Greer, J. and Carter, G.W., *Fed. Proc.*, 46 (1987) 980.
- 18 Lu, Z.-X., Fok, K.-F., Erickson, B.W. and Hugli, T.E., *J. Biol. Chem.*, 259 (1984) 7367.
- 19 Mollison, K.W., Mandrecki, W., Zuiderweg, E.R.P., Fayer, L., Fey, T.A., Krause, R.A., Conway, R.G., Miller, L., Edalji, R.P., Shallcross, M.A., Lane, B., Fox, J.L. and Carter, G.W., *Proc. Natl. Acad. Sci. USA*, 86 (1989) 292.
- 20 Woessner Jr., J.F., *FASEB J.*, 5 (1991) 2145.
- 21 Matrisian, L.M., *BioEssays*, 14 (1992) 455.
- 22 Basset, P., Bellocq, J.P., Wolf, C., Stoll, I., Hutin, P., Limacher, J.M., Podhajces, O.Y., Chenard, M.P., Rio, M.C. and Chambon, P., *Nature*, 348 (1990) 69.
- 23 Clavel, C., Polette, M., Doco, M., Benninger, I. and Berembaut, P., *Bull. Cancer*, 79 (1992) 261.
- 24 Stetler-Stevenson, W.G., *Cancer Metastasis Rev.*, 2 (1990) 289.
- 25 Polette, M., Clavel, C., Muller, D., Abecassis, J., Binnering, I. and Birembaut, P., *Invasion Metastasis*, 11 (1991) 76.
- 26 Kusakawa, J., Sasaguri, Y., Shima, I., Kameyama, T. and Morimatsu, M., *J. Oral Pathol. Med.*, 21 (1992) 221.
- 27 Sreenath, T., Matrisian, L.H., Stetler-Stevenson, W., Gattoni-Celli, S. and Pozzatti, R.O., *Cancer Res.*, 52 (1992) 4942.
- 28 Engel, G., Popowicz, P., Norling, H., Svensson, C., Auer, G., Akusjarvi, G. and Linder, S., *Int. J. Cancer*, 51 (1992) 761.
- 29 Walakovits, L.A., Moore, V.L., Bhardwaj, N., Gallick, G.S. and Lark, M.W., *Arthritis Rheum.*, 35 (1992) 35.
- 30 Okada, Y., Shinmei, M., Tanaka, O., Naka, K., Kimura, A., Nakanishi, I., Bayliss, M.T., Iwata, K. and Nagase, H., *Lab. Invest.*, 66 (1992) 680.
- 31 Nagase, H., Enghild, J.J., Suzuki, K. and Salvesen, G., *Biochemistry*, 29 (1990) 5783.
- 32 Clark, I.M. and Cawston, T.E., *Biochem. J.*, 263 (1989) 201.
- 33 Ye, Q.-Z., Johnson, L.L., Hupe, D.J. and Baragi, V., *Biochemistry*, 31 (1992) 11231.
- 34 Marcy, A.I., Eiberger, L.L., Harrison, R., Chan, H.K., Hutchinson, N.I., Hagmann, W.K., Cameron, P.M., Boulton, D.A. and Hermes, J.D., *Biochemistry*, 30 (1991) 6476.
- 35 Fesik, S.W. and Zuiderweg, E.R.P., *Q. Rev. Biophys.*, 23 (1990) 97.
- 36 Fesik, S.W. and Zuiderweg, E.R.P., *J. Magn. Reson.*, 78 (1988) 588.

- 37 Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M., *Biochemistry*, 28 (1989) 6150.
- 38 Zuiderweg, E.R.P. and Fesik, S.W., *Biochemistry*, 28 (1989) 2387.
- 39 Clore, G.M. and Gronenborn, A.M., *Prog. NMR Spectrosc.*, 23 (1991) 43.
- 40 Zuiderweg, E.R.P., McIntosh, L.P., Dahlquist, F.W. and Fesik, S.W., *J. Magn. Reson.*, 86 (1990) 210.
- 41 Kay, L.E., Clore, G.M., Bax, A. and Gronenborn, A.M., *Science*, 249 (1990) 411.
- 42 Clore, G.M., Kay, L.E., Bax, A. and Gronenborn, A.M., *Biochemistry*, 30 (1991) 12.
- 43 Zuiderweg, E.R.P., Petros, A.M., Fesik, S.W. and Olejniczak, E.T., *J. Am. Chem. Soc.*, 113 (1991) 370.
- 44 Ikura, M., Kay, L.E. and Bax, A., *Biochemistry*, 29 (1990) 4659.
- 45 Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A.M. and Bax, A., *Biochemistry*, 31 (1992) 8180.
- 46 Grzesiek, S. and Bax, A., *J. Magn. Reson.*, 96 (1992) 432.
- 47 Clubb, R.T., Thanabal, V., Osborne, C. and Wagner, G., *Biochemistry*, 30 (1991) 7718.
- 48 Boucher, W., Laue, E.D., Campbell-Burk, S. and Domaille, P.J., *J. Am. Chem. Soc.*, 114 (1992) 2262.
- 49 Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W., *Methods Enzymol.*, 177 (1989) 44.
- 50 Van Doren, S.R., Kurochkin, A.V., Ye, Q.-Z., Johnson, L.L., Hupe, D.J. and Zuiderweg, E.R.P., *Biochemistry*, 1993, in press.
- 51 Peng, J.W. and Wagner, G., *J. Magn. Reson.*, 98 (1992) 308.
- 52 Fesik, S.W., Eaton, H.L., Olejniczak, E.T., Zuiderweg, E.R.P., McIntosh, L.P. and Dahlquist, F.W., *J. Am. Chem. Soc.*, 112 (1990) 886.
- 53 Bax, A., Clore, G.M., Driscoll, P.C., Gronenborn, A.M., Ikura, M. and Kay, L.E., *J. Magn. Reson.*, 87 (1990) 620.
- 54 Majumdar, A., Wang, H., Morshauer, R. and Zuiderweg, E.R.P., *J. Biomol. NMR*, 3 (1993) 387.
- 55 Grzesiek, S. and Bax, A., *J. Am. Chem. Soc.*, 114 (1992) 6291.
- 56 Grzesiek, S., Anglister, A. and Bax, A., *J. Magn. Reson.*, B101 (1993) 114.
- 57 Ikura, M., Clore, G.M., Gronenborn, A., Zhu, G., Klee, C.B. and Bax, A., *Science*, 256 (1992) 632.
- 58 Clore, G.M., Wingfield, P.T. and Gronenborn, A.M., *Biochemistry*, 30 (1991) 2315.
- 59 Theriault, Y., Logan, T.M., Meadows, R., Yu, L., Olejniczak, E.T., Holzman, T.F., Simmer, R.L. and Fesik, S.W., *Nature*, 361 (1993) 88.
- 60 Montelione, G.T. and Wagner, G., *J. Magn. Reson.*, 87 (1990) 183.
- 61 Fesik, S.W., Gampe Jr., R.T., Holzman, T.F., Egan, D.A., Edalji, R., Luly, J.R., Simmer, R., Helfrich, R., Kishore, V. and Rich, D.H., *Science*, 250 (1990) 1406.
- 62 Clore, G.M. and Gronenborn, A.M., *J. Magn. Reson.*, 53 (1983) 423.
- 63 Gronenborn, A.M. and Clore, G.M., *Biochem. Pharmacol.*, 40 (1990) 115.
- 64 Behling, R.W., Yamana, T., Navon, G. and Jelinsky, L.W., *Proc. Natl. Acad. Sci. USA*, 85 (1988) 6721.
- 65 Fesik, S.W., *Biochemistry*, 27 (1988) 8297.
- 66 Seeholzer, S.H., Cohn, M., Putkey, J.A., Means, A.R. and Crespi, H.L., *Proc. Natl. Acad. Sci. USA*, 83 (1986) 3634.
- 67 Fesik, S.W., Zuiderweg, E.R.P., Olejniczak, E.T. and Gampe Jr., R.T., *Biochem. Pharmacol.*, 40 (1990) 161.
- 68 Otting, G., Senn, H., Wagner, G. and Wüthrich, K., *J. Magn. Reson.*, 70 (1986) 500.
- 69 Kraulis, P.J., *J. Appl. Crystallogr.*, 24 (1991) 946.
- 70 Loosli, H.R., Kessler, H., Oschkinat, H., Weber, H.P., Petcher, T.J. and Widmer, A., *Helv. Chim. Acta*, 68 (1985) 682.
- 71 Fesik, S.W., 206th American Chemical Society National Meeting, Chicago, 1993.