# 9.06 NMR – Small Molecules and Analysis of Complex Mixtures

**Arthur S. Edison**, University of Florida, Gainesville, FL, USA

**Frank C. Schroeder**, Cornell University, Ithaca, NY, USA





## 9.06.1 Introduction

Modern spectroscopic techniques have revolutionized compound identification and quantification. Only a few decades ago, identification of a structurally complex natural product would require multigram quantities of isolated material, which would then be subjected to series of derivatization and degradation experiments, aiming to deduce the unknown's structure from that of resulting derivatives or fragments that may represent known compounds. As a result of the tremendous advances in sensitivity and resolution of NMR spectroscopy over the past 30 years, identification of microgram quantities of new compounds has now become routine. For example, the structure of the polyketide antibiotic, erythromycin (**1**), was identified in 1957 only after extensive chemical and spectroscopic studies based on multigram amounts of isolated compound.[1–3] By the time its

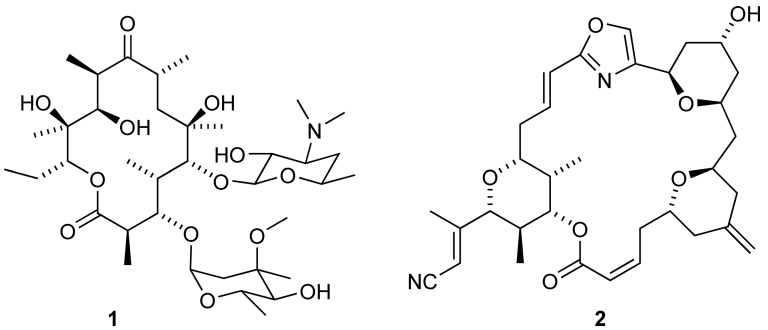


**Figure 1** Structures of the polyketide antibiotic erythromycin (**1**) and the marine polyketide hemi-phorboxazole A (**2**), which was recently identified based on a 16.5 μg (28 nmol) sample isolated from *Phorbas* sp.[7]

chemical structure was finally identified, erythromycin had already found extensive use in human medicine. Today, natural products of similar complexity, for example, hemi-phorboxazol (**2**), are routinely identified based on samples of 100 μg or less[4–7] (**Figure 1**). Of course, factors such as the structural complexity and novelty of the discovered compounds must be considered when making such comparisons. Whereas it may not be particularly challenging to design an analytical method that can reliably detect $10^{-9}$ mol of a known compound (e.g., a pesticide residue), determining the structure of an unknown natural product based on $10^{-9}$ mol of sample will likely present great difficulty.[8]

### 9.06.1.1 Sensitivity

A recent example illustrates how increases in sensitivity and the advent of multidimensional NMR spectroscopy have truly revolutionized organic structure determination. Identification of the first cardiotonic steroids from an invertebrate source in the late 1970s required the extraction of 28 000 *Photinus pyralis* fireflies. The crude extract was fractionated into five pure fractions, representing amounts from over 1 g down to 70 mg, which were then characterized by a combination of chemical and spectroscopic methods. Key structural information was afforded by one-dimensional $^{1}$H- and $^{13}$C-NMR spectroscopic analyses using a modest 250 MHz NMR spectrometer, resulting in identification of the bufadienolide (**3**).[9] Just over 25 years later, a similar analysis was carried out using a partially purified extract obtained from only 50 fireflies of the rare species, *Lucidota atra*. A 600 MHz spectrometer equipped with a microcoil probe was used,[10] allowing the characterization of 13 new bufadienolides (e.g., **4**) present in amounts ranging from 20 to 75 μg, corresponding to a decrease in sample requirement of roughly four orders of magnitude.[4] In another example, the disulfated steroid (**5**) was identified based on a sample of only 4 μg (6 nmol) and a 1.7 mm microprobe at 500 MHz. The steroid (**5**) functions as a 'sperm attracting and activating factor' (SAAF) in chemical signaling systems of the ascidian (sea squirt) *Ciona intestinalis* (**Figure 2**).[11]

Improvements in NMR spectroscopic sensitivity also benefit studies aimed at elucidation of the biological context of natural products. For example, NMR spectroscopic analysis of insect metabolites traditionally necessitated the pooling of material collected from multiple individuals, effectively eliminating NMR spectroscopy as a technique that could be used for the detailed analysis of metabolite dynamics in ecological studies. However, sensitivity increases derived from microsample NMR technology enabled Dossey *et al.*[12] to analyze metabolite mixtures within individual walking sticks, *Anisomorpha buprestoides*, permitting complete characterization of the iridioid anisomorphal (**6**) from a single insect specimen. Using the *A. buprestoides* secretion as a model system, subsequent work by Zhang *et al.*[13] demonstrated the application of covariance-based mixture analysis to automatically identify individual components in an unpurified sample. The ability to analyze individual specimens by both NMR spectroscopy and MS holds considerable promise for future biological studies.

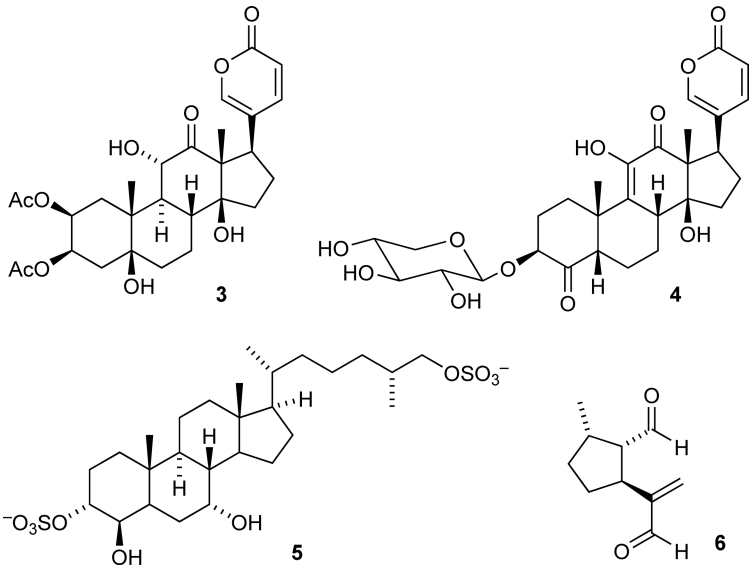


**Figure 2** Bufadienolides (**3** and **4**) from fireflies, SAAF (**5**) from the ascidian *Ciona intestinalis*, and anisomorphal (**6**) from walking sticks.

### 9.06.1.2 Mixtures

Traditionally, detailed NMR spectroscopic characterization of natural product samples was not initiated until largely pure samples of individual compounds had been obtained, usually through extensive chromatographic fractionation. However, the potential advantages of structure identification of individual components in mixtures have been widely recognized since at least the mid-1990s. Several techniques based on diffusion-ordered spectroscopy were developed to aid in this process including DOSY (diffusion-ordered spectroscopy)[14] and DECODES (diffusion-encoded spectroscopy).[15] Unfortunately, these methods often fail to resolve multiply overlapping signals and suffer from low dynamic range, which reduces their utility for structure determination in complex mixtures of organic small molecules. As a result, DOSY and related methods were primarily used to analyze mixtures of synthetic products[16–19] and never found widespread use in natural products research.

Using 2D NMR spectroscopy for the analysis of complex natural products mixtures recently regained momentum, as several studies demonstrated that using simple dqfCOSY (double-quantum-filtered correlation spectroscopy), TOCSY, HSQC, or HMBC spectra for complex mixtures offers exciting new perspectives for natural products research and chemical biology. Compared to mass spectrometric (MS) analyses of small molecule mixtures, such 2D NMR spectroscopic investigations offer the benefit of more detailed structural information, which is of particular relevance for the detection of unanticipated chemotypes. Recent examples include the identification of sulfated nucleosides, such as **7**, from spider venom,[20,21] the detection of ascarosides (e.g., **8**) as part of the mating signal in the nematode *Caenorhabditis elegans*,[22,23] and the identification of the highly unstable polyketide bacillaene (**9**) from *Bacillus subtilis*.[24] These studies show that using state-of-the-art NMR spectroscopy even minor components of complex small molecule mixtures can be characterized. Such NMR spectroscopic analyses of complex mixtures may not always permit complete structural assignments; however, additional results from mass spectroscopic analyses frequently allow proposing complete structures. As a result, the need for chromatographic separations is greatly reduced, which not only accelerates compound discovery, but also offers distinct advantages for the discovery of chemically unstable metabolites. It seems likely that the pervasive use of chromatography in natural products chemistry has skewed our knowledge of secondary metabolism, because sensitive compounds often do not survive extended exposure to solvents or chromatography media. In fact, the original motivation to explore the utility of high-resolution 2D NMR spectroscopy for the characterization of small molecule mixtures arose because alkaloids present in the poison gland secretion of *Myrmicaria* ants were found to be highly unstable for chromatographic isolation.[25,26] Myrmicarin

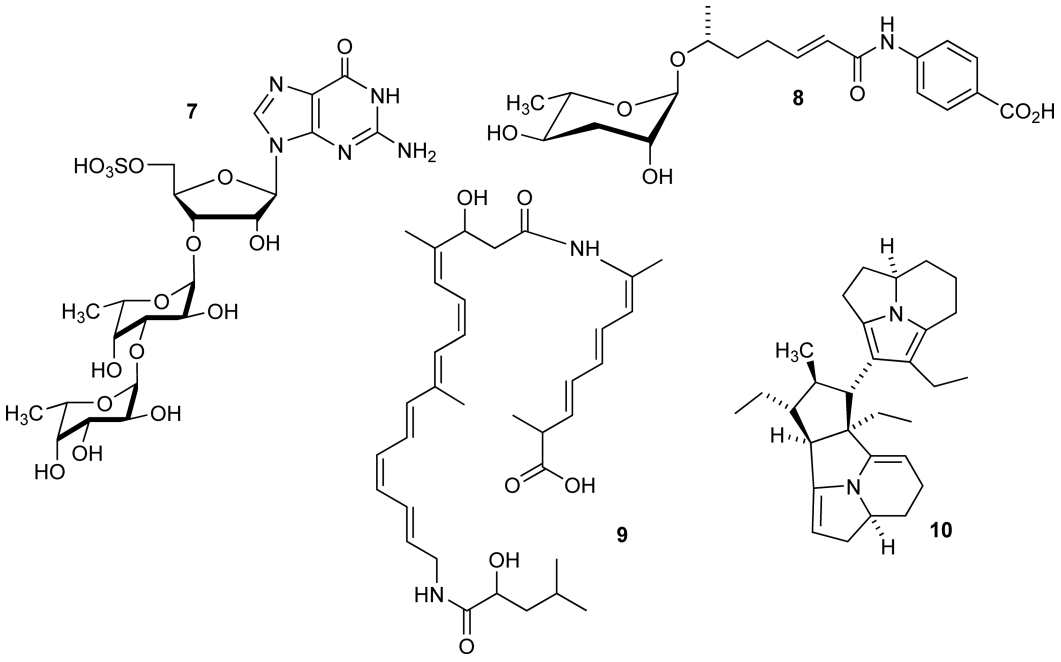


**Figure 3** Natural products identified from complex metabolite mixtures.

430A (**10**), the most unstable of the *Myrmicaria* alkaloids identified so far, thus represents one of the first members of a growing class of natural products that have never been isolated in pure form (**Figure 3**).

Advanced processing of spectroscopic data, taking advantage of statistical tools originally developed for metabolomics studies such as STOCSY,[27] SHY,[28] and others[29–32] could further enhance the utility of NMR spectroscopy and MS for analyzing natural products mixtures and correlating chemical information with biological data. However, to date there have been few reports on the application of metabolomics techniques to natural products research.[33]

In this chapter, we start with a brief overview of the standard methods currently used for the NMR-spectroscopic identification of natural products and other types of organic small molecules, which is followed by a section dedicated to NMR spectroscopic characterization of small molecule mixtures and a discussion of approaches to increase NMR spectroscopic sensitivity.

## 9.06.2 Routine NMR Spectroscopy for Natural Products Structure Elucidation

Strategies for NMR spectroscopic structure elucidation of organic compounds have been reviewed extensively.[34–37] In this section, we briefly describe a set of the most commonly useful 2D NMR spectra that is sufficient for most (though certainly not all) organic structure determination problems, and we comment on specific modifications of acquisition parameters that facilitate the analysis of mixtures.

Any NMR spectroscopic analysis of an organic sample will normally begin with the examination of a simple $^{1}$H-NMR spectrum, which serves to assess purity, concentration of minor components (if any), and overall complexity of the structures in the sample. Furthermore, the $^{1}$H spectrum provides an opportunity to examine line shape characteristics of the sample's components, and, if necessary, reevaluate solvent choice, sample concentration, or acquisition temperature. If large quantities of pure compound are available, 1D $^{13}$C-NMR spectra may also be useful. However, in most cases acquisition of a pair of ($^{1}$H,$^{13}$C)-HSQC and ($^{1}$H,$^{13}$C)-HMBC spectra will be a better use of spectrometer time, unless structural features are suspected that preclude full characterization by HSQC and HMBC. For example, compounds that feature quaternary carbon atoms that

cannot be detected by HMBC will often require acquisition of a 1D $^{13}C$ spectrum and, if possible, a 2D $^{13}C$-INADEQUATE (incredible natural abundance double quantum transfer experiment) spectrum.

Experienced natural products chemists may be able to recognize certain compound classes or characteristic structural features at this early stage of the analysis, and for known compounds such tentative structural assignments can often be confirmed through comparison with literature NMR data and additional mass spectrometric analyses. For unknown compounds, the next step in the structure elucidation process usually consists of acquisition of variants of four different types of 2D NMR spectra:

### 9.06.2.1 COSY and TOCSY

(1) ($^1H$,$^1H$)-COSY or TOCSY is used for characterization of the proton spin systems. A 'spin system' is represented by any group of protons that interact through scalar couplings, for example ethyl butanoate features two spin systems, one consisting of the five protons of the ethoxy group, and one consisting of the seven protons of the butanoyl group. There are many different COSY and TOCSY variants, differing in time requirement and type of ($^1H$,$^1H$)-coupling information provided. COSY spectra show crosspeaks only for directly coupled protons, whereas TOCSY spectra may show crosspeaks not only for protons directly coupled with each other, but also with other protons in the same spin system. For example, in COSY spectra, the alpha proton in the amino acid leucine will show crosspeaks to the adjacent beta-methylene protons, whereas in a TOCSY spectrum, depending on the experimental parameters, the alpha proton may have additional crosspeaks with protons of the gamma methine and the two methyl groups. TOCSY spectra are useful in situations where part of a spin system is obscured in the corresponding COSY spectrum, for example, due to an extensive overlap in the aliphatic region. TOCSY crosspeaks at the chemical shift of a nonobscured proton can often be used to reveal the obscured parts of the spin system. For example, TOCSY spectra are used extensively in the NMR-spectroscopic characterization of proteins where it is used to map complete amino acid spin systems onto the corresponding amide protons. TOCSY can also be useful for the analysis of natural products, especially for complex mixtures where overlap is often a problem. As described below, Bruschweiler's group has developed mixture analysis methods that are based on TOCSY spectra.[13,38,39] Acquisition parameters for TOCSY spectra can be roughly tuned to emphasize either COSY-type interactions with short mixing times or more complete spin system correlations with longer mixing times. Several mixing sequences are available on modern spectrometers, and some of the more popular are DIPSI-2[40] and MLEV-17.[41] Bax[42] has provided an excellent overview of TOCSY (also known as homonuclear Hartmann–Hahn or HOHAHA) that summarizes the basic principles and demonstrates applications.

For identifying new natural products, TOCSY spectra are often less straightforward to analyze than COSY spectra, because a TOCSY crosspeak does not necessarily indicate that two protons are coupled with each other – the presence of a crosspeak only shows that two protons are part of the same spin system. Furthermore, TOCSY spectra may be significantly more crowded than COSY spectra, and TOCSY crosspeak intensity is often difficult to correlate with structural properties. Finally, the fine structure of TOCSY crosspeaks is much less amenable to detailed analysis than dqfCOSY crosspeaks; as described below, the antiphase dqfCOSY crosspeaks contain information on proton multiplicities and scalar coupling constants, which TOCSY crosspeaks cannot provide.

Among variants of COSY, simple gradient COSY (gCOSY) spectra and phase-sensitive double-quantum-filtered COSY (dqfCOSY) spectra are the most useful for natural product analysis. gCOSY spectra can be acquired extremely fast – using a small molecule sample of >1 mg a decent spectrum can usually be acquired within 5–10 min. However, gCOSY spectra provide only very limited information. The fine structure of gCOSY crosspeaks is often poorly defined, which poses problems for differentiating signals of overlapping peaks and does not allow distinguishing crosspeaks that are due to large coupling constants from crosspeaks that are due to smaller couplings. This is of particular relevance for the analysis of complex spin systems where distinguishing between long-range couplings and stronger geminal or vicinal couplings is important, and where coupling constants may carry important information about relative configuration. In addition, artifacts are sometimes difficult to distinguish from 'real' crosspeaks in gCOSY spectra.

For these reasons, dqfCOSY spectra are usually a better choice for any compound or mixture sample that includes complex proton spin systems. If acquired using sufficiently long acquisition times (600 ms or more),

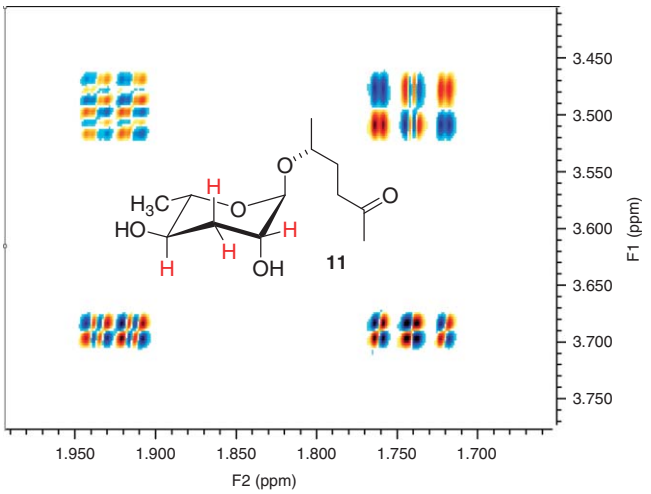


**Figure 4** Part of the dqfCOSY spectrum of the ascaroside (**11**), a component of the *Caenorhabditis elegans* dauer pheromone. The fine structure of the four shown crosspeaks permits accurate determination and assignments of the geminal and all vicinal coupling constants of the two methylene protons (red).[43]

dqfCOSY crosspeaks closely reflect the splitting patterns of corresponding multipletts in one-dimensional $^{1}H$ spectra. Based on their splitting patterns, dqfCOSY crosspeaks belonging to a specific proton can be easily recognized and grouped together, and as a result, overlapping signals can be clearly distinguished. Furthermore, the characteristic antisymmetric fine structure of each crosspeak not only allows for fairly accurate determination of coupling constant values, but also permits determining the coupling partner responsible for the coupling constant (**Figure 4**). Therefore, crosspeaks due to small coupling constants can be easily distinguished from crosspeaks due to large coupling constants. Another advantage resulting from the highly characteristic appearance of dqfCOSY crosspeaks is that artifacts can be recognized very easily. dqfCOSY spectra should always be acquired using pulse sequences that employ phase-cycling for coherence selection. Although gradient-selected versions of dqfCOSY are available, line shapes in these gradient versions are usually extremely poor.

DqfCOSY spectra usually provide sufficiently accurate values for coupling constants that are larger than twice the line width of the corresponding proton signals, for example, coupling constants larger than 2–4 Hz. However, for signals of protons that have several similar though not identical coupling constants, the interpretation of the dqfCOSY crosspeaks may present considerable difficulty. For analysis of such highly complex spin systems, or in situations where precise knowledge of small coupling constants is required, E.COSY ('exclusive' COSY) spectra are better suited.[44,45] E.COSY crosspeaks are less complex than dqfCOSY crosspeaks and can be used to obtain highly accurate values even for very small long-range coupling constants. For example, E.COSY was used to determine coupling constants in the ladybird beetle alkaloid psylloborine A (**12**) (**Figure 5**), which features mostly an aliphatic heptacyclic ring system. As evident from the dqfCOSY spectrum shown in **Figure 6**, this compound's spin systems are extremely complex and some dqfCOSY crosspeaks are

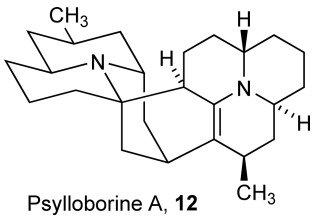


**Figure 5** Structure of the dimeric polyacetate alkaloid psylloborine from the ladybird beetle *Psyllobora vigintiduopunctata* (**12**).[46]

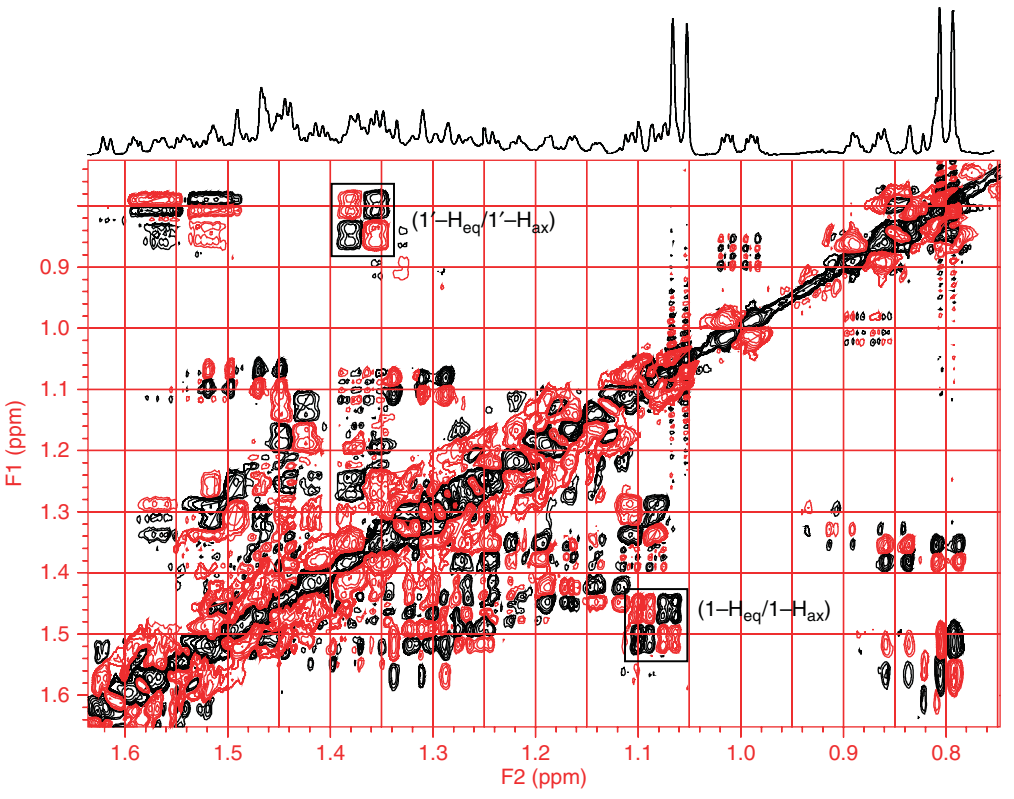


**Figure 6** The 75–1.64 ppm region of the $^{1}$H-NMR and the dqfCOSY spectrum of psylloborine A (**12**) ($C_6D_6$, 500 MHz). The vicinal coupling constants of the proton 1′ – $H_{eq}$ (1.37 ppm) cannot be directly extracted, due to poor resolution in F2 and overlap with other crosspeaks, for example, of the proton 8 – $H_{eq}$ at 1.36 ppm. The E.COSY signals corresponding to the crosspeaks (1 – $H_{eq}$/1 – $H_{ax}$) and (1′ – $H_{eq}$/1′ – $H_{ax}$) are shown in **Figure 7**.

difficult to interpret. However, the interpretation of corresponding E.COSY spectrum was straightforward (**Figure 7**).[46] It should be noted, however, signal to noise (S/N) of E.COSY is considerably lower than that of dqfCOSY, and that E.COSY requires careful calibration of pulse width in order to minimize artifacts.

### 9.06.2.2 HSQC and HMQC

($^{1}$H,$^{13}$C)-HSQC or HMQC serves to identify proton-bearing carbons and to associate these carbons with their attached protons. HSQC spectra feature generally better line shapes than HMQC and the commonly used multiplicity-edited HSQC versions offer the added benefit of distinguishing $CH_3$, $CH_2$, and CH groups. However, the HMQC pulse sequence is significantly shorter than the HSQC sequence, and therefore magnitude-mode HMQC spectra may have better S/N than magnitude-mode HSQC spectra. Use of adiabatic $^{13}$C- pulses in the HSQC sequence can significantly reduce this sensitivity disadvantage. For the analysis of complex mixtures with many overlapping proton and/or carbon signals, HSQC is much better suited than HMQC because of better line shapes. Both HMQC and HSQC are usually acquired with $^{1}$H-decoupling during acquisition and as a result feature one crosspeak per $^{1}$H-resonance that is located at a chemical shift value close to that of the $^{12}$C-attached protons. It should be noted, however, that the proton chemical shifts of HSQC or HMQC crosspeaks are not exactly identical to that of the main signals in the $^{1}$H [or ($^{1}$H,$^{1}$H)-COSY and ($^{1}$H,$^{13}$C)-HMBC] spectra. The latter represent $^{12}$C-bound protons (unless the sample is isotopically labeled), whereas the signals in HSQC/HMQC obviously represent $^{13}$C-bound protons, whose chemical shift values may differ slightly, and to a varying extent. Such small differences in chemical shift can make precise calibration of HMQC/HSQC spectra difficult and may present a problem for samples that feature

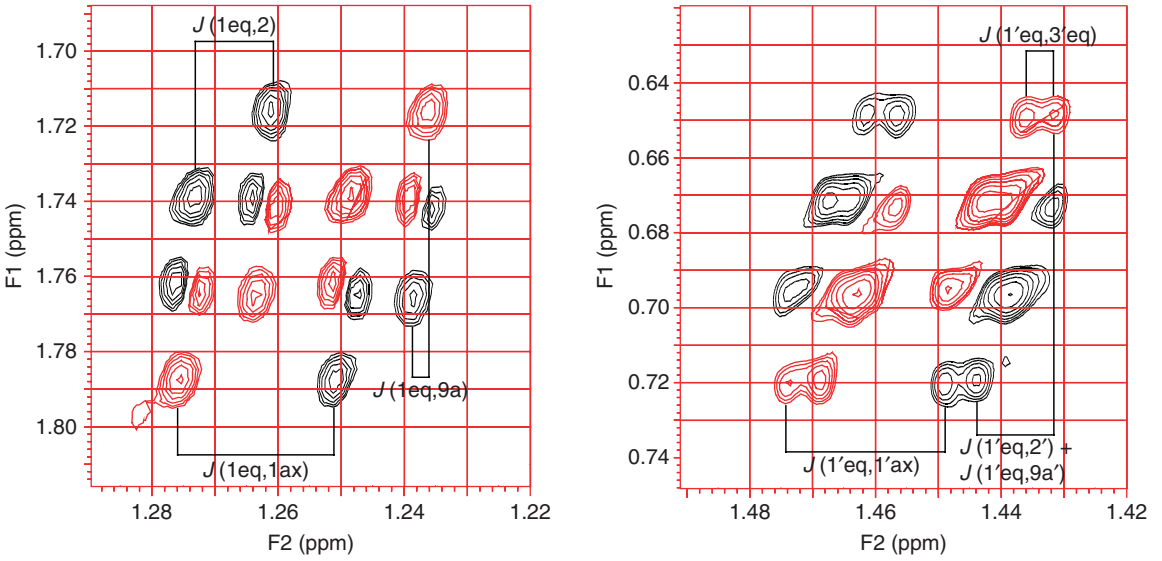


**Figure 7** Left: Crosspeak of the geminal pair $1-H_{eq}/1-H_{ax}$ in the E.COSY spectrum of psylloborine A (**12**) ($CD_2Cl_2$, 500 MHz). The passive vicinal couplings $J$(1eq,2) and $J$(1eq,9a) and the active coupling $J_{1eq,1ax}$ can be determined without interference from components of opposite phase. Right: E.COSY crosspeak of the geminal pair $1'-H_{eq}/1'-H_{ax}$. Although, in comparison to the corresponding dqfCOSY spectrum (**Figure 2**), the crosspeak is greatly simplified, the passive vicinal coupling constants $J$(1′eq,9a′) and $J$(1′eq,2′) cannot be directly extracted. Owing to partial overlap, only the sum $J$(1′eq,9a′) + $J$(1′eq,2′) can be determined. Since $J$(1′eq,9a′) is accessible from the E.COSY crosspeak of $9a'-H/1'-H_{ax}$ (not shown), $J$(1′eq,2′) can be calculated. The E.COSY crosspeak $1'-H_{eq}/1'-H_{ax}$ also allows one to determine the active geminal coupling $J$(1′eq,1′ax). In addition, a small four-bond coupling $J$(1′eq,3′eq) is revealed.[46]

many overlapping proton signals of very similar chemical shift, for example, complex mixtures. Acquisition of coupled ('nondecoupled') HSQC spectra is sometimes advantageous, for example, in cases where one-bond $^{1}H$–$^{13}C$ coupling constants are of interest, or in cases where decoupling would result in low-quality spectra due to sample heating (this can be an issue when using long acquisition times, or when using polar solvents, especially in the presence of salts).

### 9.06.2.3 HMBC

($^{1}H$,$^{13}C$)-HMBC provides correlations between protons and carbons that are two or three bonds apart from each other (though occasionally four-bond or even five-bond correlations may be observed). HMBC spectra are important for the detection of quaternary carbons and serve to link separate structural fragments obtained from analysis of COSY/TOCSY and HSQC/HMQC. However, interpretation of HMBC spectra often represents the most challenging step in the structure elucidation process, for several reasons. The intensity of crosspeaks in HMBC spectra is notoriously difficult to predict. Some two- or three-bond correlations may be extremely weak or not appear at all, and therefore the absence of an HMBC crosspeak cannot, *a priori*, be taken as evidence against a specific structural connection. Furthermore, there is no simple method for distinguishing two- and three-bond correlations. Additional difficulties may arise in situations where weak HMBC crosspeaks could represent either a three-bond or a four-bond (or rarely five-bond) correlation. Finally, because standard HMBC experiments are usually optimized for ($^{1}H$,$^{13}C$)-long range coupling constants of intermediate size, both very strong and very weak ($^{1}H$,$^{13}C$)-long-range couplings may give rise to weak crosspeaks in routine HMBC spectra, which can be source of considerable confusion. The latter concern can be addressed by acquiring two separate HMBC spectra using two different mixing delays, for example, 50 and 100 ms. It should be noted that when using long mixing delays, the acquisition time should be increased to at least twice the mixing delay, for example, for a delay of 100 ms, the acquisition time should be at least 200 ms, which is considerably above the default values in common HMBC parameter sets. Even when using standard parameters, HMBC signal

intensity is often somewhat lower than that of HSQC or HMQC spectra, primarily because of the relatively long mixing delay (40–120 ms) in the HMBC pulse sequence. Often $^{1}$H line shape is a good predictor of S/N in HMBC spectra: generally, narrow $^{1}$H line widths correlate with good S/N of corresponding signals in the HMBC spectra.

For most organic small molecules, acquisition of one-dimensional $^{13}$C spectra is not required when well-resolved HSQC and HMBC spectra are available. Exceptions include compounds that have quaternary carbons that simply do not show any HMBC correlations, for example, because there are no protons within two or three bonds of some carbons (see also Section 9.06.2.5). Another limitation of routine HMBC is that spectral resolution in the $^{13}$C-chemical shift dimension is limited. As is the case for HMQC (but not HSQC), HMBC crosspeaks are broadened in the $^{13}$C-chemical shift dimension by the ($^{1}$H,$^{1}$H)-coupling constants of the proton whose long-range ($^{1}$H,$^{13}$C)-coupling is observed. As a result, crosspeaks belonging to carbons with very similar chemical shifts can sometimes not be unambiguously assigned. However, the interfering ($^{1}$H,$^{1}$H)-couplings can be removed by using a constant-time variant of the standard HMBC experiment. Using band-selective, constant-time HMBC variants, spectra with extremely high resolution in the $^{13}$C-dimension can be easily obtained.[47]

### 9.06.2.4 NOESY and ROESY

($^{1}$H,$^{1}$H)-NOESY and ROESY provide information about spatial proximity of protons that are separated by up to about 5 Å, which can be used to determine relative configuration and, in some cases, conformation of organic small molecules. Other applications include the study of chemical exchange or investigations of the interaction of natural products with their protein targets. NOESY and ROESY spectra are similar, but the choice of which to use depends on the rate of tumbling of the molecule, which is roughly proportional to the molecular weight, but also depends on its polarity and on polarity and viscosity of the solvent. NOESY crosspeaks are opposite in sign for small and large molecules. For small molecules, the sign of NOESY crosspeaks is opposite to the sign of the diagonal, whereas for large molecules NOESY cross- and diagonal peaks have the same sign. Correspondingly, there is a range of molecules for which NOESY crosspeaks are close to zero,[48] and thus NOESY is not suitable. NOESY can generally be used for organic small molecules of molecular weights below 800 Da, unless the compound under investigation is very polar, or requires the use of a highly polar and/or viscous solvent, such as DMSO.

In ROESY spectra, the sign of the crosspeaks are always opposite to that of the diagonal peaks, and therefore ROESY is suitable for intermediate-sized molecules around 1 kDa and smaller molecules in viscous solvents or at low temperatures. Both NOESY and ROESY can also be used to investigate chemical exchange, and they show crosspeaks for nuclei that are in slow (relative to the chemical shift differences) exchange. Importantly, the sign of chemical exchange crosspeaks is the same as that of the diagonal, and so for small molecules both NOESY and ROESY can be used to distinguish chemical exchange crosspeaks from crosspeaks due to spatial proximity. In larger molecules, but sometimes also in small molecules, additional crosspeaks can occur through 'spin diffusion', a relay of magnetization along a chain of $^{1}$H's that are close together that leads to TOCSY-type crosspeaks. Mixing times for acquiring natural product NOESY or ROESY spectra should be set to 500–800 ms (NOESY) and 200–400 ms (ROESY).

### 9.06.2.5 Other Techniques and Current Limitations

Not all types of natural products can be sufficiently characterized using routine 2D NMR spectroscopy as described in the preceding section. For example, in highly unsaturated compounds some carbons may not be detected by HMBC simply because there are no protons within three bonds of these carbons. There are few ways to address this problem with NMR-spectroscopic means, and in some cases, chemical modification (e.g., hydrogenation) or degradation may be necessary in order to complete structural assignments. In a rare event that large amounts of the compound in question are available, ($^{13}$C,$^{13}$C)-correlations that provide direct evidence for carbon–carbon bonds such as ($^{13}$C,$^{13}$C)-INADEQUATE (or its $^{1}$H-detected cousin, ADEQUATE) can be useful.[49] However, due to the low natural abundance of $^{13}$C, sensitivity of INADEQUATE is extremely low, as only pairs of adjacent $^{13}$C atoms contribute to the signal. One of the

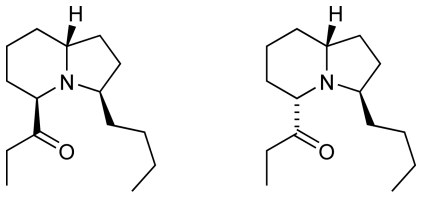


**Figure 8** Myrmicarin 237A (**13**) and 237B (**14**), which were identified using ($^{13}$C,$^{13}$C)-2D-INADEQUATE.[50]

very few examples for the use of INADEQUATE for natural product samples is presented by the identification of the indolizidine alkaloids myrmicarin 237A and 237B (**13** and **14**) (**Figure 8**).[50] These compounds equilibrate through keto–enol tautomerism and therefore had to be characterized as a 1:1 mixture of diastereomers, which resulted in extremely crowded COSY and HMBC spectra. A ($^{13}$C,$^{13}$C)-COSY-type INADEQUATE was then used to unambiguously distinguish between the $^{13}$C-resonances of the two diastereomers (**Figure 9**).

A different problem is posed by structures that include large numbers of NMR-inactive heteroatoms. In such cases, it may be impossible to assemble based on NMR spectroscopic data simply because there are too many possibilities for arranging the heteroatoms around the identified carbon- and hydrogen-based partial structures. For compounds that include nitrogen or phosphorus, $^{15}$N- and $^{31}$P-NMR spectroscopy can often supply important additional information.

As natural products chemists detect and investigate more and more complex structures, additional limitations of current NMR spectroscopic approaches have become apparent. Examples for compound classes that pose great difficulty for NMR spectroscopists include compounds with ill-defined conformations, natural products that occur as large families of structurally similar compounds, or oligomeric compounds whose

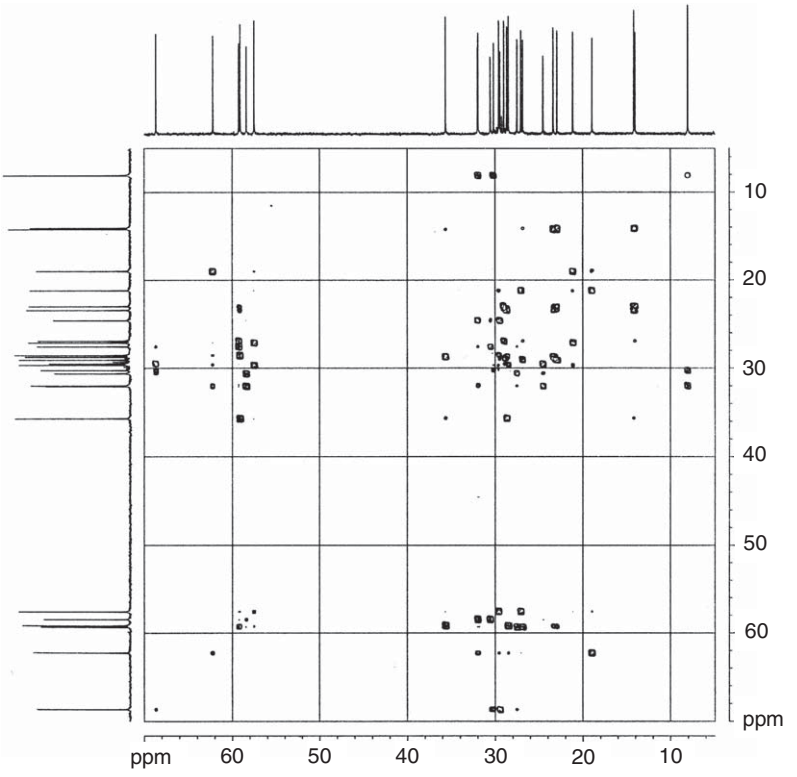


**Figure 9** COSY-like ($^{13}$C,$^{13}$C)-2D-INADEQUATE of a 220 mg sample of a 1:1 mixture of myrmicarin 237A (**13**) and 237B (**14**), acquired over 54 h.[50]

assembly follows an irregular scheme, such as complex glycosides or lipidated natural product derivatives. For the latter groups of compounds, some partial degradation and/or derivatization may still be required in order to enable NMR spectroscopic analysis.

## 9.06.3 Complex Mixtures

NMR spectroscopy evolved primarily as a tool for the characterization of pure compounds or simple well-defined mixtures (see Section 9.06.1.2), whereas strategies for the NMR spectroscopic identification of compounds from complex mixtures have been developed only recently. As several examples have shown, using NMR spectroscopy for the analysis of complex mixtures can open up new perspectives and may enable new lines of inquiry in both natural products chemistry and metabolomics. Before discussing some of these examples in greater detail, it is useful to consider what developments spawned the recent surge in applications of NMR spectroscopy to mixtures and what prevented earlier uses of NMR spectroscopy for this purpose.

Following the initial observation in the early 1950s that the resonance frequency of a nucleus is influenced by its chemical environment,[51] and that the fine structure of a resonance could be influenced by other nuclei through intervening chemical bonds,[51,52] chemists quickly seized upon the enormous potential of NMR spectroscopy for structure determination. As a result, NMR spectroscopy became one of the most important spectroscopic tools of organic chemists. Combining NMR spectroscopic with mass spectrometric analyses proved particularly useful, with MS providing information about molecular weight and atomic composition, and NMR spectroscopy contributing information about chemical environment and, importantly, connectivity and spatial configuration. Until about 1980, NMR spectroscopic structure elucidation was largely based on one-dimensional spectra, providing information about chemical shift (suggesting a specific chemical environment), relative signal intensity (indicating the number of a specific type of nuclei in the molecule), and multiplicity (suggesting connectivity between individual groups of nuclei in the molecule).[34,53] Extracting signal intensity and multiplicity information from one-dimensional NMR spectra depended crucially on sample purity, because signals of impurities could easily skew signal intensities or obstruct important splitting patterns. Furthermore, based on one-dimensional spectra it is often impossible to determine whether two signals represent nuclei that are part of the same molecule or whether they represent two (or more) separate structures. As a result, NMR spectroscopy was deemed largely unsuitable for the analysis of complex mixtures such as crude natural products extracts, and NMR spectroscopic analysis was usually initiated only after pure or almost pure samples of the compound(s) of interest had been obtained, generally as the endproduct of extensive chromatographic fractionation.

The eventual realization that NMR spectroscopy can nonetheless be applied most advantageously to the characterization of mixtures then depended on at least two separate developments. First, the advent of 2D-(and subsequently, multidimensional) NMR spectroscopy enabled much better access to connectivity information than could be obtained from the analysis of multiplets in one-dimensional spectra. 2D spectra such as COSY, HSQC, or HMBC yield correlations that correspond to connectivity through one or more chemical bonds.[34,54] Importantly, dispersion of signals along a second (or third) chemical shift dimension almost always removes any ambiguity resulting from overlap of signals in one dimension, and therefore enables recognition and identification of partial structures that may belong to several different compounds.[55]

However, even the advent of multidimensional spectroscopy alone did not yet suffice to make NMR spectroscopic analysis of complex mixtures broadly applicable. In the early days of 2D-spectroscopy, the capabilities of spectrometers and processing hardware limited resolution and dynamic range of the spectra severely. For example, processing of a very low-resolution COSY spectrum on a Bruker AC250P (250 MHz proton) spectrometer console in 1990 could take as much as 30 min. Moreover, early 2D spectra, especially the most useful inversely detected HSQC and HMBC, were prone to artifacts, making it nearly impossible to unambiguously discern signals representing minor components.

The advent of improved data acquisition systems and greatly increased computing power fundamentally changed the scope of multidimensional NMR spectroscopy, and today very low-artifact COSY, TOCSY, or HSQC spectra can be obtained whose resolution approximates that of one-dimensional spectra.[35,36] Along with increases in sensitivity and resolution derived from higher magnetic field strength and improved probe design

(see Section 9.06.4), these developments have set the stage for a broad exploration of the utility of NMR spectroscopy for characterizing complex mixtures. Sections 9.06.3.1 and 9.06.3.2 describe recent examples for using 2D NMR spectroscopy for the characterization of new natural products from complex biological extracts, whereas Section 9.06.3.3 describes a method for computational deconvolution of 2D-spectra of complex mixtures into one-dimensional subspectra that represent partial structures of individual components. Computational approaches have also been applied to ensembles of one-dimensional spectra of complex small molecule mixtures for the purpose of biomarker identification. Corresponding applications in metabolomics are discussed in Section 9.06.3.4.

### 9.06.3.1 NMR Spectroscopic Analysis of Complex Natural Products Mixtures

The idea to use 2D-NMR spectroscopy for a systematic characterization of crude or unfractionated natural products mixtures was first conceived in connection with research on the chemical ecology of arthropods.[20,21,25,26,46,56–58] During studies of the chemical composition of various arthropod secretions, several cases were encountered for which conventional analytical methodology based on fractionation of the secretions aiming at the isolation of individual components failed to identify the biologically active principles. It was concluded that the chromatography-based fractionation of these secretions had resulted in destruction or loss of the active components. As a consequence, the use of NMR spectroscopy for the characterization of native, entirely unfractionated materials was considered. As one of the first examples, the unfractionated defensive secretion of a Ladybird beetle pupa, *Epilachna borealis*, was subjected to 2D NMR spectroscopic analysis including dqfCOSY, NOESY, HSQC, and HMBC spectra, for which acquisition parameters were somewhat modified in order to obtain higher-resolution spectra.[58] This approach quickly resulted in the identification of a previously overlooked group of compounds that made up more than 50% of the secretion, a new family of macrocyclic lactone alkaloids, the polyazamacrolides, such as **15** and **16** (**Figure 10**).

These insect secretions presented a perfect starting point for exploring the utility of direct NMR spectroscopic analysis of crude mixtures, because the secretions consisted of mixtures of only one to three structurally

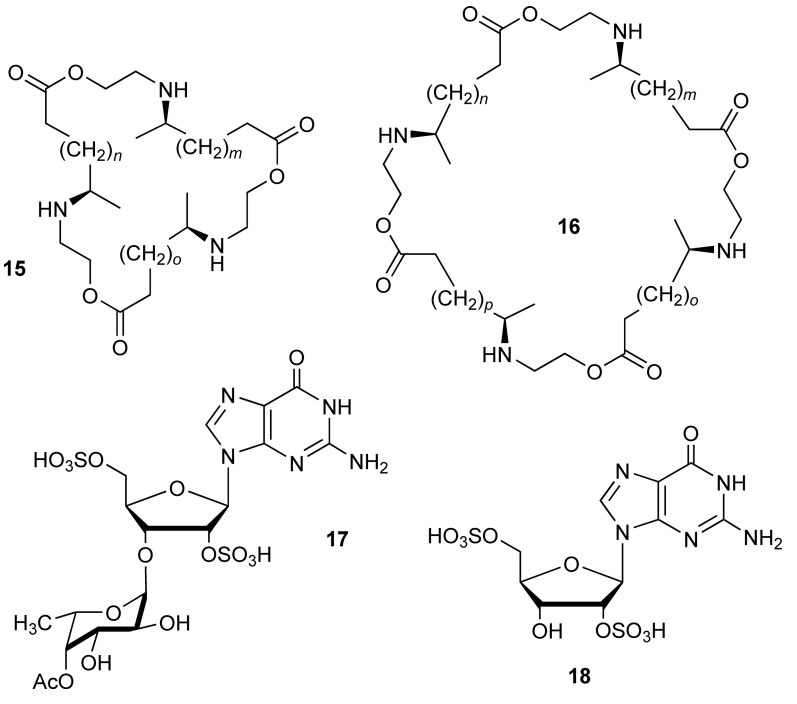


**Figure 10** Polyazamacrolides from pupae of the ladybird beetle *Epilachna borealis*[58] and sulfated nucleosides (**17**) and (**18**) identified from spider venom.[20,21,59]

distinct groups of small molecules. However, as recent analyses of crude spider venom have shown that even much more complicated mixtures of small molecules are amenable to NMR spectroscopic analysis.[20,21] NMR spectroscopic studies of crude spider venom were motivated by the earlier identification of a bis-sulfated nucleoside, HF-7 (**17**), a selective and potent kainate receptor antagonist, from the venom of the grass spider, *Hololena curta*.[59] The discovery of this entirely unexpected natural product suggested that spider venoms might harbor interesting new classes of neurotoxins. Moreover, it seemed unlikely that HF-7 is the only spider venom component of its kind. The question remained why sulfated nucleosides had previously escaped detection, even though spider venoms had been subject to intensive chemical scrutiny, which had led to identification of hundreds of proteins, peptides, acylated polyamines, and various small molecule neurotransmitters. Given this very high degree of complexity, it is not surprising that most previous studies of spider venom chemistry applied some form of chromatographic fractionation as a first step. Because sulfated nucleosides are somewhat susceptible to hydrolysis, it was suspected that sulfated nucleosides may have been overlooked in some earlier analyses as a result of decomposition during chromatographic fractionation. Building on experience gathered from characterization of the polyazamacrolides, it was thus attempted to characterize entire, unfractionated spider venom samples using 2D NMR spectra, including dqfCOSY, HMQC, and HMBC. This approach led to the identification of sulfated nucleosides such as **18** as important components in the venoms of several spider species, including previously well-studied species such as the hobo spider, *Tegenaria agrestis*, and the brown recluse spider, *Loxosceles recluse*.[20,21] Effectively, these 2D NMR spectroscopic analyses provided a largely undistorted and impartial view of spider venom composition, without any skewing of the results stemming from chromatographic separation.

It is important to note that such 'direct' NMR spectroscopic analyses of complex natural product mixtures may not always permit assigning complete structures. In many cases, a full or near-complete characterization will only be possible for a few major components, whereas more or less extensive partial structures will be obtained for minor components. However, any partial structures elucidated will provide important information that may be (1) used to search natural product databases for similar compounds, (2) combined with results from GC–MS or LC–MS analysis to develop better hypotheses about their structures, (3) used to develop a fractionation scheme tailored to the isolation of specific compounds of interest, and (4) used to design syntheses for the proposed structures.

Direct NMR spectroscopic analyses are particularly well-suited to examine natural product extracts for the presence of novel or unanticipated compounds. However, 2D spectra of natural product mixtures are often extremely complex, which limits the feasibility of using 2D NMR as a first-line tool for the characterization of large numbers of natural product extracts. In some cases, this concern can be addressed by focusing only on specific features in the spectra, for example, groups of crosspeaks that correlate with a certain biological activity or genotype. A method that facilitates recognition of 2D NMR signals relevant within a specific biological context, DANS ('differential analysis through 2D NMR spectroscopy'), is discussed in Section 9.06.3.2. Even if detailed spectral interpretation is not pursued, 2D spectra obtained for a crude natural products sample can be useful as a largely unbiased record of its original composition against which the results from any subsequent fractionation can be compared. Such comparisons can aid in recognizing artifacts or detecting loss of some components of the original mixture.

#### 9.06.3.1.1 Comparing NMR spectroscopic characterization of proteins and small molecule mixtures

In some regard, NMR spectra obtained from complex mixtures of small molecules resemble those of biological macromolecules such as large peptides, proteins, or oligonucleotides.[55] NMR spectra of both biological macromolecules and small molecule mixtures are similar in that they feature a very large number of overlapping signals, which through the use of a variety of two- or three-dimensional experiments can be assigned to individual substructures. In case of macromolecules such as proteins, these substructures may represent individual amino acids residues, whereas in the case of crude natural products extracts these substructures constitute fragments of the various secondary metabolites it contains. However, there are significant differences in the strategies for NMR spectroscopic analysis of crude natural products mixtures and biological macromolecules. Analysis of sets of NMR spectra from proteins or nucleic acids is primarily based on *template recognition*, and thus NMR spectroscopic analysis of biological macromolecules usually consists of sets of 2D and

3D experiments addressing specific structural features of these templates. For example, NMR spectroscopic analysis of proteins is based on a series of specialized NMR-pulse sequences designed to identify amino acid residues, the sequence of amino acids, and spatial proximity within the chain(s). Many macromolecular NMR pulse sequences are highly specific, such as the HNCO experiment to detect the repeating covalent structure of peptide bonds in proteins.[60] While analyzing crude natural products extracts, NMR spectroscopic experiments cannot be tailored in this way, because these crude mixtures usually contain a very large variety of components featuring highly diverse structures. Furthermore, the structural properties of these compounds will vary considerably between extracts, in a largely unpredictable manner. Therefore, NMR-based analysis of crude natural products extracts has to rely on experiments that focus on the most basic common features of organic molecules as frameworks of carbon and hydrogen. These experiments are primarily versions and combinations of ($^1$H,$^1$H)-COSY/TOCSY, ($^1$H,$^1$H)-NOESY/ROESY, and ($^1$H,$^{13}$C)-HSQC/HMQC/HMBC.[36]

When dealing with spectra of complex mixtures, signal overlap, especially in the proton dimensions, becomes a serious problem, which in some cases may necessitate some form of pre fractionation. Using high-field spectrometers can help increase spectral dispersion and thus reduce overlap, as the relative size of crosspeaks decreases approximately as $(1/F)^n$, with $F$ being the field strength of the magnet and $n$ the dimensionality of the experiment (neglecting line shape effects). Interference by overlap can be alleviated further by taking advantage of the much longer relaxation times of small molecules compared to those of macromolecules. The slower relaxation of small molecules allows for longer acquisition times especially for directly and indirectly detected proton magnetization, which results in correspondingly higher resolution of the spectra and permits detection of smaller scalar couplings. For example, the initial NMR spectroscopic characterization of crude natural products extracts in the studies discussed here was largely based on very high-resolution dqfCOSY spectra.[21,23,61]

#### 9.06.3.1.2 NMR spectroscopy versus mass-spectrometry-based approaches for characterizing crude mixtures

Traditionally, efforts to characterize unfractionated small-molecule mixtures have relied primarily on combinations of MS with HPLC or gas chromatography. As MS is extremely sensitive and typical GC–MS and LC–MS analyses are fast, can be automated easily, and thus can accommodate large numbers of samples, these techniques would seem extremely well suited for the purpose of characterizing libraries of unfractionated natural product extracts,[62] and in fact, various LC–MS-based approaches are being pursued to characterize fungal or bacterial metabolomes.[33] However, there are important drawbacks to the exclusively LC–MS-based approaches. One major disadvantage of using MS as the primary analytical tool is that most mass-spectrometric techniques are strongly biased toward the detection of a few specific compound classes.[63] For example, positive ion electrospray ionization MS is by orders of magnitude more sensitive for basic amines, amino acids, or peptides than for nonbasic polyketides or terpenoids. Alternative ionization techniques such as APCI, MALDI, and so on are biased in different ways and to varying degrees, and no single spectrometric approach is sufficient to provide an unbiased snapshot of a small molecule mixture of unknown composition. Regardless of the ionization technique chosen, the structural information available from mass spectrometric analyses is often insufficient for detailed structural assignments.[64] Although a few compound classes, notably peptides, can be characterized extremely well by MS[65,66] in most classes of small molecule metabolites, MS can provide only limited structural information beyond a tentative molecular formula. In the context of analyzing natural product extracts of diverse origins, this lack of structural information is particularly problematic, as decisions over further fractionation of an extract depend entirely on assumptions as to whether a specific extract is likely to contain new, interesting chemotypes or not. Therefore, the detailed structural information available through 2D NMR spectroscopy of small molecule mixtures can represent an invaluable addition to mass spectroscopic results. Of course, any NMR-based characterization of natural product mixtures normally will have to be complemented by HPLC–MS or GC–MS analyses.

A significant disadvantage of NMR-based approaches for the characterization of natural product mixtures is represented by the much lower sensitivity and dynamic range of NMR spectra compared to MS. Furthermore, the often high complexity of 2D NMR spectra obtained for mixtures can make their interpretation challenging.

The latter disadvantage could be overcome through the use of computational analysis, or through approaches based on graphical comparison of sets of 2D NMR spectra (DANS), as described in the following section.

### 9.06.3.2 Differential Analysis through 2D NMR Spectroscopy

One of the big remaining challenges in natural products chemistry is to develop better methods for connecting newly identified small molecule structures with their biological functions, including knowledge of the mechanisms regulating their biosynthesis and of their molecular targets. The traditional armamentarium of natural products chemistry appears ill-suited for this purpose, given the complexity of most organism's metabolomes and the scope of assigning functions to hundreds, if not thousands, of individual components, many of which represent previously undescribed chemical structures.[33] Efforts aimed at determining the structure of biologically relevant small molecules have traditionally relied on bioassay-guided fractionation, usually based on highly time-consuming multistep chromatographic fractionation schemes that require extensive biological assays at every stage in the process. As a result, approaches based on bioassay-guided fractionation often take years to tease out and identify the biologically active component(s), and the need for fractionation poses great difficulty in cases of synergism, that is, cases where more than one compound is required to elicit the monitored activity.[22,23] Importantly, for chemically unstable compounds chromatographic fractionation may be unsuitable all together.

Several recent studies have shown that 2D NMR analyses of natural product extracts can be highly effective for associating small molecules with specific biological properties, most significantly phenotype and genotype of the producing organism(s). These studies are based on differential analyses of 2D NMR spectra (DANS), a method for graphic comparison 2D NMR spectra representing different biological states, for example different phenotypes or genotypes.

#### *9.06.3.2.1 DANS for screening of a fungal extract library*

DANS was first used for the detection of differential expression of natural products in a small library of fungal extracts.[67] This library was derived from a *Tolypocladium cylindrosporum* strain that was cultured under a variety of 'stress' conditions, aiming to elicit the production of secondary metabolites from otherwise inactive biosynthetic pathways. The resulting unfractionated metabolite extracts were used to acquire dqfCOSY spectra with very high resolution in both dimensions. dqfCOSY was chosen for these studies because dqfCOSY crosspeaks feature highly regular fine structures and are thus particularly information rich. In addition, dqfCOSY spectra offer fairly good dynamic range, which often permits detailed characterization of spin systems representing even very minor components, as had been demonstrated with the examples described in Section 9.06.3.1. For differential analysis, dqfCOSY spectra corresponding to different extracts were superimposed onto each other, using a specific algorithm that suppressed signals common to all extracts, but highlighted signals unique to individual spectra. The algorithm chosen for this overlay allowed suppression of signals even in cases where compounds occurred at significantly different concentrations in different fungal extracts. As a result, only signals representing compounds whose expression was very strongly dependent on the culturing conditions were highlighted in the overlay. The DANS algorithm can be fine-tuned to reveal less severe differences as well, though it is not suitable for accurate quantitative measurements (**Figure 11**).

Application of DANS enabled fast screening of the *Tolypocladium* extract library for proton spin systems representing chemotypes that are produced only under specific conditions, and led to the identification of two new terpenoid indole alkaloids that are expressed under certain nutrient-deficient conditions, but do not get produced using standard culturing protocols. The structures of the two new indole alkaloids, TC-705A (**19**) and TC-705B (**20**), were proposed on the basis of NMR spectra obtained for the unfractionated extracts and subsequently confirmed through additional spectroscopic analyses of isolated samples.[67] In addition to TC-705A and TC-705B, differential expression of several known compounds was observed. These known compounds were identified based on comparison of NMR spectroscopic data obtained from DANS with literature data, in conjunction with results from additional mass spectrometric analyses.

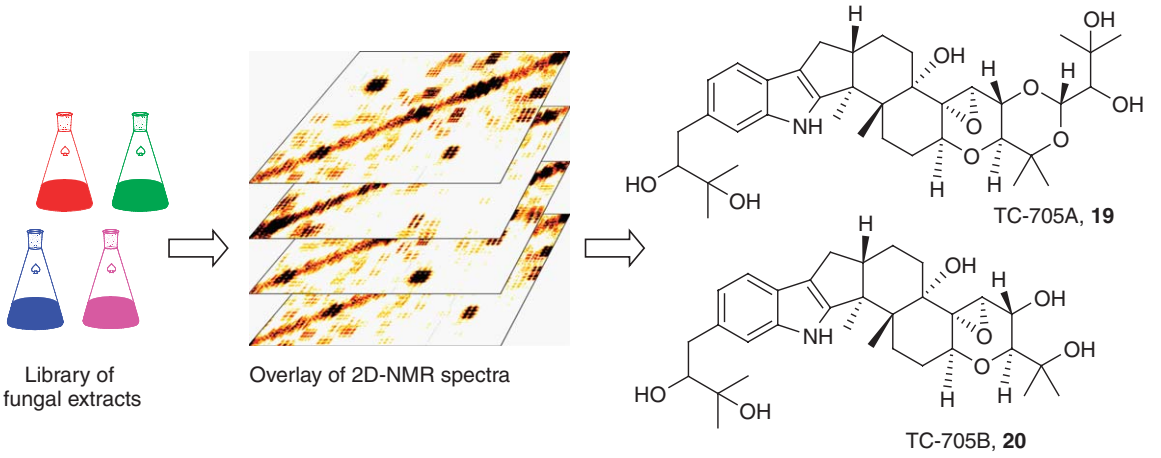


**Figure 11** Identification of new fungal natural products through DANS (schematic).[67]

#### 9.06.3.2.2 *DANS-based identification of bacillaene*

In a second example, DANS was used to determine the structure of the elusive product of the polyketide gene cluster pksX in *B. subtilis*.[24] The ~80 kb pksX gene cluster encodes an unusual hybrid polyketide/nonribosomal peptide synthase that had been linked to the production of the uncharacterized antibiotic bacillaene. Multiple copies of this synthase – each similar in size to the ribosome – assemble into a single organelle-like complex with a mass of tens to hundreds of megadaltons. The resource requirements of the assembled megacomplex suggest that bacillaene serves important biological functions. However, the unconventional domain organization of the PksX synthase and the presence of multiple enzymes that act in *trans* rather than in the standard assembly-line mode that is characteristic of polyketide and nonribosomal peptide biosynthesis precluded bioinformatic prediction of bacillaene's structure. Furthermore, isolation of bacillaene using traditional activity-based fractionation could not be accomplished due to the molecule's chemical instability.

Therefore, identification of bacillaene based on NMR spectra of largely unfractionated bacterial extracts was pursued. DANS-based comparison of a bacillaene-producing *B. subtilis* strain and a corresponding knockout strain clearly identified distinct proton spin systems present in the bacillaene producer but absent in the knockout. Acquisition of additional ($^{1}$H,$^{13}$C)-HMQC, ($^{1}$H,$^{13}$C)-, and ($^{1}$H,$^{15}$N)-HMBC, and ROESY spectra for the bacillaene-producing strain subsequently permitted full identification of the two main products of PksX, bacillaene (**9**) and dihydrobacillaene (**21**), along with several double-bond stereoisomers. The biosynthesis of bacillaene by the PksX synthase was subsequently investigated by Moldenhauer *et al.*[68] Small molecules like bacillaene, which link genotype (the *pksX* gene cluster) with phenotype (antibiotic and likely other activities), are central to chemical biology, and as this example demonstrates, comparative NMR-based approaches such as DANS should be generally useful for their characterization (**Figure 12**).

#### 9.06.3.2.3 *Identification of signaling molecules in* Caenorhabditis elegans *through DANS*

The utility of DANS for the identification of signaling molecules in eukaryotes was recently demonstrated with the identification of a mating pheromone in the nematode *C. elegans*.[23] *C. elegans* is an important model organism for biomedical research, and a systematic characterization of structures and functions of small molecules in *C. elegans* will be critical for advancing our understanding of many biological processes.[69]

Earlier work had shown that three glycosides of the dideoxysugar ascarylose are part of a male-attracting pheromone that is produced by *C. elegans* hermaphrodites.[22,43,70] These compounds, the ascarosides ascr#2, ascr#3, and ascr#4, showed strong synergism as mating signals: mixtures of ascarosides were potently active at concentrations at which individual components effected no response.[22] Although biologically fascinating, the ascarosides' synergistic properties resulted in tremendous logistical challenges for their identification through activity-guided fractionation, as this required to combinatorially recombine chromatographic fractions in order to assess activity. Despite these efforts, biological testing of mixtures of ascr#2, ascr#3, and ascr#4 at

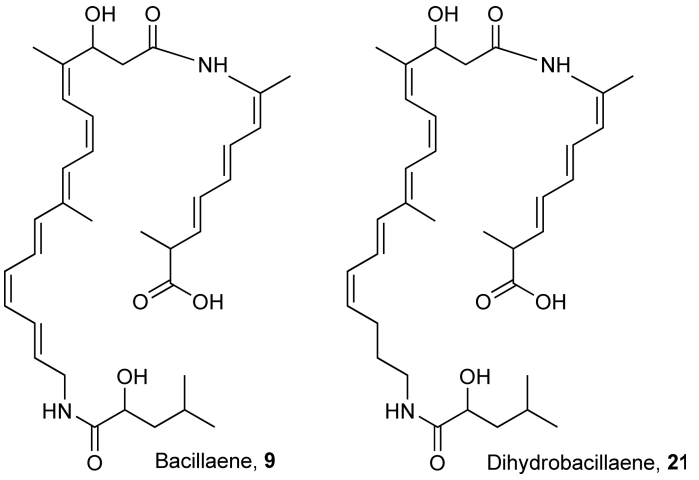


**Figure 12** Compounds identified through DANS *Bacillus subtilis*.

physiological concentrations did not fully reproduce activity of the original pheromone extracts, and it seemed likely that important components of the mating pheromones remained to be identified (**Figure 13**).

For the purpose of identifying missing components of the mating pheromone, the *C. elegans* mutant strain *daf-22* offered a unique opportunity. *daf-22*-derived metabolite extracts had been shown to have little dauer-inducing activity and are not significantly active in the male attraction assay. Therefore, a careful comparison of the *daf-22* metabolome with that of wild-type worms should reveal the missing *daf-22*-dependent pheromone components among compounds present in wild-type worms but absent in *daf-22*. As in the examples described in the preceding sections, this comparison was accomplished through DANS based on largely unfractionated metabolite extracts that represent highly complex mixtures of many hundred metabolites.[23] For differential analysis of the dqfCOSY spectra, the *daf-22*-derived spectrum was superimposed onto the wild-type spectrum, again using an algorithm that suppressed signals present in both mutant and wild-type spectra. As a result, only signals present in the wild-type spectrum but entirely absent from the *daf-22* spectrum remained unaltered in the overlay (**Figure 14**).

DANS-based comparisons of *daf-22* and wild-type metabolite extracts revealed several partial structures representing compounds produced only by wild type but not *daf-22* worms, including several previously unknown compounds. These compounds represented far <0.1% of the entire metabolite mixture and therefore further characterization through HSQC or HMBC was not possible based on spectra of the unfractionated metabolite extracts. However, the differentially produced compounds were easily identified after partial

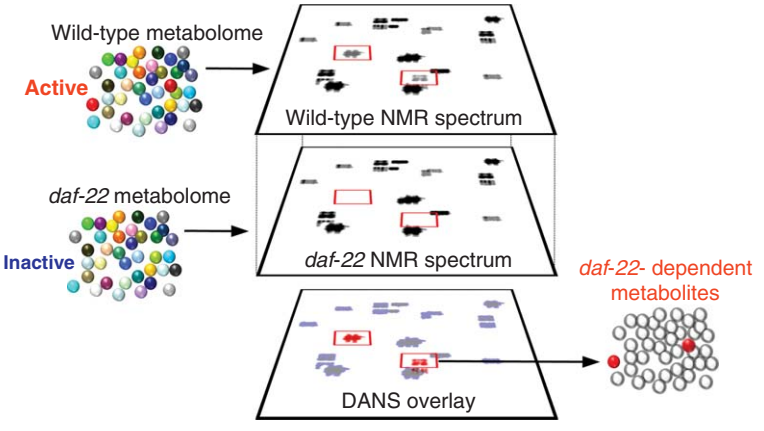


**Figure 13** DANS-based comparison of *Caenorhabditis elegans* wild-type and *daf-22* mutant metabolomes.[23]

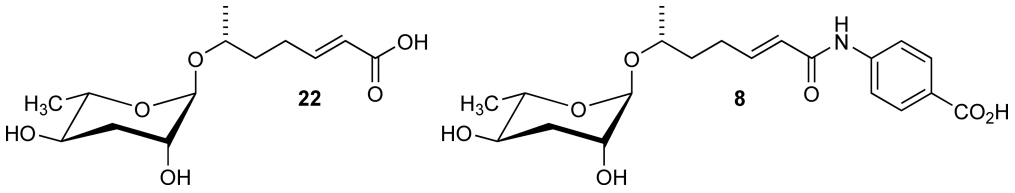


**Figure 14** Components of the *Caenorhabditis elegans* mating signal identified through DANS.[23]

chromatographic purification, using additional 2D NMR spectroscopy and MS, as the ascarosides ascr#7 (**22**) and ascr#8 (**8**). In total, the DANS-based comparison of *C. elegans* wild-type and *daf-22* metabolite extracts led to the identification of four novel ascarosides, three of which were shown to function as mating pheromones or regulators of developmental timing. Ultimately, these investigations allowed to fully reconstitute the male-attracting activity of wild-type pheromone extract to that of a *daf-22* mutant.[23]

One significant problem for any comparison of metabolite mixtures is that metabolism is strongly dependent on environmental conditions, and even small changes in temperature, nutrient conditions, or other factors can induce significant changes in relative concentrations of compounds. To minimize the impact of such variations, the algorithm used for DANS in this study was chosen in such a way that it would highlight only cases where a compound is completely absent (given the detection limit of the NMR spectroscopic equipment) from the *daf-22* spectra. Increase of NMR-spectroscopic sensitivity, or consideration of metabolites whose biosynthesis is less strongly *daf-22*-dependent, could reveal additional compounds relevant for phenotypic differences between wild-type and *daf-22* worms.

This study showed that comparative NMR spectroscopic methods such as DANS can be used to dissect changes in small molecule production in response to genetic manipulation, and that this approach could complement or replace activity-guided fractionation for identifying biologically relevant small molecules. The primary benefit of DANS lies in the ability to quickly obtain structural information for metabolites that may represent good candidates for further evaluation in a specific biological context.

### 9.06.3.3 Complex Mixture Analysis by NMR

Covariance NMR data processing developed by the Bruschweiler laboratory leads to high-resolution symmetric 2D datasets, even with relatively low-resolution acquisition in the indirect dimension.[71,72] Bruschweiler's group has developed an efficient approach called COLMAR to identify individual components in complex biological covariance NMR spectra.[73] COLMAR is freely available through a Web Portal developed and maintained by the Bruschweiler laboratory (http://spin.magnet.fsu.edu/). The input dataset for COLMAR is a covariance processed 2D NMR spectrum. Originally, COLMAR was developed for homonuclear TOCSY spectra, but the Bruschweiler laboratory is adding other options for the analysis of 2D $^{13}$C-HSQC–TOCSY datasets. The heart of COLMAR is an algorithm called DemixC, which deconvolutes covariance TOCSY spectra and extracts 1D spectral traces that represent individual spin systems with minimal likelihood of overlap and thus, individual compounds.[13,74] Although they are a probabilistic measure of nonoverlapping spin systems, the 1D traces from DemixC look like 1D NMR spectra and can be analyzed similar to 1D NMR spectra of pure compounds. The final component of COLMAR is an efficient database matching algorithm called COLMAR Query.[75,76] Chemical shifts from the DemixC traces are screened against the BMRB or other metabolomics spectral database.[77] The output of COLMAR Query is a ranked list of the highest scoring compounds with the best matches to known compounds in the database. COLMAR represents an efficient way to semiautomatically identify known compounds from a complex mixture, because it only requires a single 2D TOCSY spectrum as input.

As mixture analysis by NMR is developed, it is increasingly critical to improve NMR small molecule databases. There are currently three main publically accessible small-molecule databases available with NMR data, the Biological Magnetic Resonance Data Bank (BMRB: http://www.bmrb.wisc.edu/),[78,79] the Madison Metabolomics Consortium Database (MMCD: http://mmcd.nmrfam.wisc.edu/),[80] and the Human Metabolome Database (HMDB: http://www.hmdb.ca/).[81] These databases each support searching

experimental NMR databases for matches to experimental spectra. The MMCD and HMDB both extensively link to other databases, and MMCD has chemical shift prediction protocols that can aid identification. Importantly, they contain experimental NMR spectra that can be downloaded and compared with experimental mixtures. New tools such as MetaboMiner[82] are being developed to analyze experimental NMR data of unknown mixtures with library spectra from all of the databases. The BMRB database has extensive datasets with raw time-domain data that can be freely downloaded and analyzed. The MMCD directly utilizes the BMRB experimental data, and efforts are being made to put the experimental data from the HMDB into the BMRB database. The BMRB accepts referenced and assigned NMR data from users, so the database is steadily growing. NMR databases are less developed than their GC–MS counterparts, and there are several technical issues related to referencing, solution conditions, scalar couplings, and specific types of NMR experiments and detected nuclei that make NMR more complex than mass-spectrometry databases. As NMR small molecule databases develop, complex mixture analysis by NMR will become more and more important and routine.

#### 9.06.3.4 Metabolomics/Metabonomics

Another very powerful approach to complex mixture analysis by NMR has been developed for biomarker discovery. The Nicholson group has developed computational tools and approaches to identify small-molecule metabolites that change in response to some perturbation such as the use of a drug or from disease.[83] Statistical correlation spectroscopy (STOCSY) is a powerful approach that utilizes the natural variation found in all biological samples to find biomarkers.[27] STOCSY is based on a very simple concept: standard 1D NMR spectra are recorded on a large number of samples, and the NMR signals in these spectra are then statistically correlated by comparing their amplitudes between samples. By statistically correlating the amplitudes of chemical shifts from individual spectra, resonances that are from the same compound or biosynthetic pathway can be identified. Furthermore, the Nicholson group has developed approaches that can be used to discriminate between correlations from the same compound versus correlations from the same metabolic pathway.[84]

Much of the development and most applications of STOCSY concerned very complex mixtures such as human urine or blood plasma. For example, STOCSY can be used to compare groups of control versus diseased or drug-treated individuals in order to discover compounds that are unique biomarkers of the condition of interest.[85–87] A relatively simple extension of STOCSY, called statistical heterospectroscopy (SHY), allows for correlating different types of datasets, such as NMR and MS or even microarray data collected on the same samples.[28] SHY could aid in structure identification from complex mixtures by correlating NMR and mass-spectrometry data to assign molecular weights to compounds with known chemical shifts. One great advantage of the STOCSY approach to complex mixture analysis is that key resonances can be efficiently identified as potential biomarkers. In other words, it can help to find the proverbial needle in a haystack from a large mixture of compounds.

## 9.06.4 Methods to Improve Sensitivity

Much of the technical development of NMR over the past half century has focused on improving sensitivity. The fundamental problem is the low starting Boltzmann polarization that arises from the low energies of nuclear spin transitions. Several methods have been developed to improve the sensitivity or S/N in NMR. One major approach is through pulse sequence development to optimize the efficiency and information content of NMR spectra through manipulating the spin physics; some of the more important experiments for small molecules were described above.

NMR frequencies are directly proportional to the magnetic field by the basic equation, $\omega_0 = -\gamma B_0$, which relates the frequency ($\omega_0$) to the applied field ($B_0$) by the gyromagnetic ratio ($\gamma$). This simple equation drives the development and purchase of larger and larger magnets, because the S/N goes up as the resonance frequency goes up. The exact increase in S/N depends on many factors, especially differential rates of relaxation at different field strengths, but it is commonly accepted that the S/N increases approximately as $B_0{}^{1.5}$–$B_0{}^{1.75}$.[88] Unfortunately, the price of big magnets also increases significantly as the field strength increases.

For example, a 950 MHz (22.3 T) superconducting system costs around $8 million whereas a 500 MHz (11.7 T) is closer to $500 000. The biggest magnets also require considerable physical infrastructure and space, making the highest field systems difficult for most users to acquire, maintain, and operate. In the future, NMR facilities might become more like X-ray synchrotron facilities with very large magnets at a few major sites that can provide remote access to users. Some major magnet facilities, such as the National High Magnetic Field Laboratory, also have resistive or hybrid (resistive plus superconducting) magnets that currently can reach field strengths up to 45 T (for hybrid resistive and superconducting magnets) and require large power supplies and other infrastructure. While these low homogeneity magnets are not yet suitable for routine NMR, there is a possibility that methods will be developed to better utilize these for high-resolution studies.[89]

A far more practical solution to improve S/N for most natural products chemists is with the NMR probe. Through the radio frequency (RF) coil, the NMR probe is the interface between the sample and the spectrometer, and it is used to both excite nuclear spins and detect the electrical signals generated by precessing spins. For a fraction of the cost required to purchase a magnet, a fairly routine 500 or 600 MHz system can provide outstanding S/N for small molecules with the right choice of probe. The basic requirements for a probe are that they have an electrical conductor oriented to deliver a magnetic field $B_1$ that is perpendicular to the static field $B_0$, and there are several ways to do this. Standard commercial probes that are sold with virtually every NMR system have coils that are made from copper wire and wound in a geometry to deliver a horizontal $B_1$ magnetic field while accommodating a 5 mm vertically loaded NMR tube. This system has been used successfully for many years, because 5 mm tubes allow for approximately 600 μl of liquid for analysis and provide good S/N for samples with concentrations of about 1 mmol $l^{-1}$ on modern instruments using a standard probe. For a molecule with a molecular weight of 500 Da, an investigator would need about 300 μg to get good results with a standard 5 mm NMR probe using a 600 MHz spectrometer.

For challenging studies, it is helpful to consider two different types of sample limitations, mass and solubility limited. Natural product studies are often mass limited because of challenges associated with the collection and isolation of samples. In contrast, they are often not solubility limited, because a wide range of organic solvents can be employed. In contrast, studies with proteins or other biological macromolecules often are solubility limited, but relatively large quantities of samples can often be produced. The worst scenario for NMR is when a sample is both mass and solubility limited. Although there is some overlap, the type of sample will often dictate the choice in NMR probe technology that can best solve the problem.

Many natural product samples are severely mass limited, and it is difficult or impossible to isolate enough material to achieve the necessary concentration in a 5-mm tube. Several methods can be worked out to improve the situation with mass limited samples. Perhaps the simplest is to use 5-mm NMR tubes with susceptibility matched plugs that reduce the need for excess sample outside of the active volume of the probe. Samples need to be long enough to extend beyond the coil to avoid edge effects that severely degrade the homogeneity of the field. Susceptibility plugs allow most of the sample to be positioned in the center of the probe, but they require careful loading and positioning to avoid air bubbles that will degrade line shapes. This will bring the sample requirements of a 500 Da compound to about 150 μg in a standard probe at 600 MHz. Although susceptibility matched NMR tubes are useful in optimizing the use of the available sample, they can be difficult to shim and, as a result, the line shapes are sometimes compromised, which can lower overall S/N, especially in HMBC spectra. Of all 2D spectra routinely used for small molecule structure elucidation, HMBC spectra generally have the lowest S/N, and, unfortunately, signal strength in HMBC spectra is also strongly dependent on $^1$H line shapes.

### 9.06.4.1 Specialized NMR Probes

For studies at common magnetic field strengths such as 11.7 T (500 MHz) or 14.1 T (600 MHz), there are three main ways to improve S/N beyond a standard 5-mm probe. The simplest is to make the coils smaller, as the mass sensitivity of an NMR measurement increases roughly in inverse proportion to the diameter of the coil. A second very popular approach is to cool the entire coil and preamplifier in order to reduce the noise, thus increasing the S/N. A third approach utilizes material that conducts electricity more efficiently than copper wire.

### 9.06.4.2 Signal-to-Noise Issues

S/N values are routinely used in NMR, especially when shopping for a new spectrometer or probe. One would think that this ratio of two numbers would be an unambiguous and objective way to compare systems, but unfortunately, it is not so straightforward. First, major NMR vendors use different algorithms to estimate noise, and several additional definitions of noise are used in the literature. Second, the thickness of the walls of NMR tubes can influence S/N measurements, especially as the tube diameter decreases. Not all probe and spectrometer manufacturer's use the same standards. It is most common among conventional top-loading tube systems to use 0.1% ethylbenzene in $CDCl_3$, but solenoidal flow systems typically report S/N values using 10 mmol $l^{-1}$ sucrose in $D_2O$. Finally, when working with very small volumes, solvent volatility can play a role in manufacturing consistent sealed standards. For example, when evaluating the performance of a 1-mm probe, we found differences as large as 10% between two factory-sealed samples of 0.1% ethylbenzene in $CDCl_3$. In short, S/N is a useful guide but needs to be interpreted with great care, especially when informing decisions on major purchases.

### 9.06.4.3 Small Coils

The sensitivity of an NMR coil is defined as the $B_1$ field per unit current, and this is inversely proportional to the diameter of the coil,[88] so smaller coils have greater mass sensitivity. Any type of coil can be made smaller, and standard saddle coil or Helmholtz designs that accommodate vertically loaded tubes are commercially available as small as 1 mm in diameter. However, depending on details of the coil geometry, solenoid coils can provide between 2 and 3 times greater sensitivity than a standard saddle coil.[88] Solenoid coils pose two challenges, both of which have been nicely addressed. First, the horizontal orientation of the solenoid in the main $B_0$ field causes severe distortions to the field. Uncorrected, this leads to poor NMR line shape and big losses in sensitivity. Andrew Webb, Jonathan Sweedler, and colleagues solved this problem by surrounding the coil in a fluid that has the same magnetic susceptibility as copper wire.[90] Using this approach, high-quality NMR spectra can be obtained from extremely small volumes of sample by using small coil diameters.[90–92] The second problem involves sample handling; the horizontal orientation of the coil makes standard NMR tubes and sample loading impossible. In principle, samples can be placed in sealed capillary tubes and inserted into the coils, but this is cumbersome and requires that the probe be removed from the magnet for each sample change. Moreover, sealing the tubes without introducing air bubbles is difficult. A much more practical solution is to connect tubing to flow the samples into and out of the coil. This loading scheme can be as simple as a syringe attached to tubing or as complex as the output of a chromatographic separation.

Integrated systems that utilize 1-mm solenoidal microcoil probes and various sample-loading methods are available commercially from Protasis. The utility of commercially available solenoidal microprobes for the analysis of mass-limited natural products has recently been reviewed.[10] Examples for natural products applications include the identification of 13 new steroids from only 50 specimens of the firefly *Lucidota atra* (e.g., **4** in **Figure 2**).[4] These analyses were carried out on only partially purified samples, each containing 20–100 μg of up to three steroids. In direct comparison to using a 5-mm inverse-detection room temperature probe and susceptibility plugs (Shigemi tubes), the use of a solenoidal microprobe provided an up to threefold gain in S/N while maintaining very high spectral quality.

These small-volume systems are good choices for either high-throughput semiautomated analysis or in environments with multiple users, because the probes can be easily switched with other standard probes.[10] S/N comparisons between conventional tube systems and flow solenoids are especially problematic. However, based on comparisons between a 1-mm cryogenic HTS probe[93] and values in the literature from a 1-mm room temperature solenoid,[94] about 30 μg of a 500 Da sample would give comparable results to a 1 mmol $l^{-1}$ sample in a 5-mm warm copper probe.

### 9.06.4.4 Cooling the Electronics

A second approach to increasing S/N through NMR probe design involves reducing the coil and receiver noise. Significant advances have been made during the past decade in cryogenically cooling the coils, electrical

circuits, and preamplifiers in order to reduce the thermal noise associated with the measurement.[95] Major commercial NMR vendors offer cryogenically cooled probes, and these are very effective, even with standard copper wire and coil geometries that allow top-loading samples. All of these probes thermally isolate the sample from the coils, which are cooled to about 20 K. Sample temperatures can be regulated in modest ranges around room temperature, so biological samples are easily analyzed. Although the increase in S/N is dependent upon the dielectric properties of the sample, an increase of about 4× is not unusual. Most cryogenic probes accommodate 5-mm tubes, and with sample tubes that are large, the conductivity of the solvent can have a significant influence on S/N with these probes. For organic solvents and low-salt aqueous buffers, cryogenic probes deliver the best results. However, even moderate salt concentrations can seriously degrade their performance. The salt dependence worsens with increasing field strengths and with larger diameter samples. A 500 Da sample in an organic solvent would require roughly 75 µg for good NMR spectra in a 5-mm cryoprobe using a standard 600 MHz spectrometer. One should note that the required sample concentration would only be about 250 µmol $l^{-1}$, whereas in order to get similar performance with a room-temperature probe one would have to use a much higher concentration of about 1 mmol $l^{-1}$. Large-volume cryogenic probes are excellent choices for samples with limited solubility, explaining their widespread use in biomolecular NMR.

The primary disadvantages of cryogenic probes are that they are very expensive, require more physical infrastructure like chilled water lines than a conventional probe, and are difficult to install and remove. Most facilities with cryogenic probes keep them in dedicated instruments and only remove them for maintenance or repair. This works well for groups with similar samples and needs. Cryogenic probes typically have fixed frequencies and cannot replace the flexibility of broadband probes for unusual nuclei.

### 9.06.4.5 High-Temperature Superconducting Coils

Copper-based material is the most common conductor for NMR probes. However, there are other choices, which can provide better sensitivity through improved current carrying capacity. High-temperature superconducting material, specifically YBCO (yttrium barium copper oxide), has been used since the early 1990s in NMR coils. The first HTS probe was designed and built at Conductus (Sunnyvale, CA) in the 1990s.[96] HTS coils are constructed by depositing YBCO onto planar surfaces and inductively coupling them to a copper RF circuit.[97] These coils have a much higher quality factor ($Q$) than cooled copper coils and as a result have been shown to provide significantly higher S/N in NMR than achieved by cold copper coils. The drawbacks and challenges to HTS coils include poor filling factors due to the flat coil geometry, difficulty and cost in construction, and difficulty in tuning flat wafers to multiple frequencies required for biological NMR. However, the benefits of HTS probe technology are significant: for the same temperature and coil diameter, planar HTS coils can increase the S/N by up to a factor of 2 over copper wire.

Combining HTS materials with cryogenic cooling and small coil size can result in very sensitive NMR probes. The National High Magnetic Field Laboratory, Bruker Biospin, and University of Florida recently collaborated to design and build a 1-mm cryogenic probe with HTS coils.[93] This probe uses top-loading glass tubes with sample volumes between 5 and 10 µl, depending on the wall thickness. This probe has an S/N value of close to 300 for 0.1% ethylbenzene, which is about 20× more mass sensitive than a commercial 5-mm warm copper probe (S/N ~ 1000 for 0.1% ethylbenzene). Thus, a 500 Da sample would require about 10 µg for good results. However, the concentration would increase to about 2 mmol $l^{-1}$, so the smallest probes are not appropriate for concentration-limited samples. This 1-mm HTS probe has been used in several natural products studies, including the analysis of insect defensive secretions with a single or very few insects,[12,13,98–100] identification of a component of the *C. elegans* mating pheromone,[22] identification of glycosylated pheromones suspensoside A (**23**) and suspensoside B from male Caribbean fruit flies,[101] and several marine natural product identifications (**Figure 15**).[7,102,103]

A 1.7 mm cryogenic probe is now available commercially, and this appears to provide excellent results. Dalisay *et al.*[5,6] reported the identification of several new natural products of very low abundance from marine sponges of the genus *Phorbas*, including the tetrachloro polyketide muironolide A (**24**) and another polyketide, hemi-phorboxazol A (**2**). These structures were determined based on samples of only 90 µg of muironolide and 16.5 µg of hemiphorboxazol A.

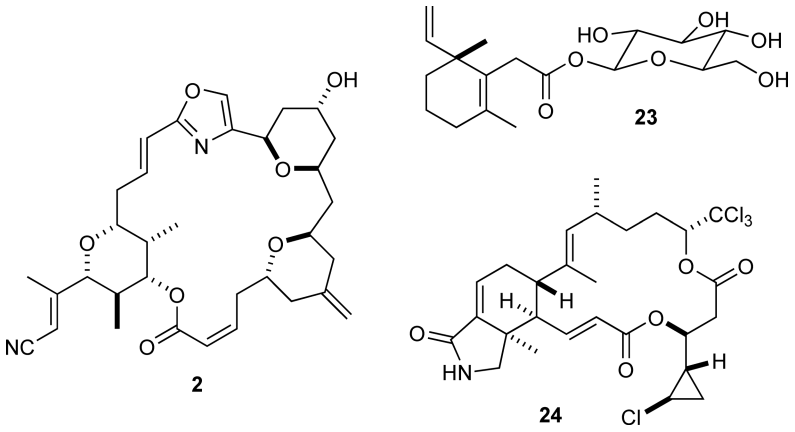


**Figure 15** Examples for natural products identified using small-volume cryogenic probes.

The limitations of small-volume cryogenic probes are similar to standard cryogenic probes. If the coils are made from HTS material, there are additional challenges related to the fact that the coils need to be on planar surfaces that are not fully optimized to the geometry of cylindrical sample tubes and the glass vacuum tubes needed to isolate the cryogenic coil temperatures from the sample at room temperature.

### 9.06.4.6 Probe Summary

There are many choices of probes, and the best one depends on the amount of sample, the solubility of the sample, the number of different types of users of the system, and the budget. The most flexible probes in terms of accommodating a wide range of samples and users are standard 5-mm room temperature probes. Smaller diameter probes have higher mass sensitivity but because their sample volumes are dramatically smaller, they are not optimal for samples with limited solubility. The highest sensitivity probes are small, cryogenically cooled, and utilize high-conducting HTS materials, but these are expensive and suffer from general limitations of cumbersome cryogenic systems and small volumes. A good bet for general biomolecular NMR and some natural product work is a 5-mm cryogenically cooled probe; however, all cryogenic probes are expensive and more difficult to change with other probes such as broadband for nonproton-detected studies.

### 9.06.4.7 Dynamic Nuclear Polarization

Dynamic nuclear polarization (DNP) is a rapidly developing technique that achieves significantly higher S/N than conventional NMR spectroscopy by transferring the very large polarization of electrons to nuclei at low temperatures. Much of the development in DNP focused on solid state samples and frozen liquid samples,[104] and techniques to directly polarize solutions in high magnetic fields are now being developed.[105] Under the right conditions, DNP enhancements can reach several orders of magnitude,[104,106] but there are some limitations of DNP that need to be overcome before it could be widely applied to natural product studies. First, DNP only lasts as long as the $T_1$ of the polarized nucleus, so most applications are short 1D $^{13}C$ experiments, although rapid acquisition 2D methods have been demonstrated on hyperpolarized samples.[107] Second, nuclei with short $T_1$ times are difficult or impossible to detect, and therefore identification of unknown compounds could be challenging without additional analyses using conventional NMR spectroscopy. Third, radicals need to be added to samples to provide a source of electrons, and investigators might be unwilling to add these to very precious natural product samples that took considerable time to isolate and purify. Finally, most solution DNP studies employ samples that were polarized in the frozen state and then thawed,[106] and most studies have used aqueous solutions. Modified protocols and polarizing agents would need to be developed for natural products analysis in organic solvents. Despite these current limitations, the future of DNP for major enhancements of NMR S/N is very promising, and natural products chemists should follow the developments of this field.

## 9.06.5 Outlook

It is an exciting time for natural products chemistry. Analytical tools are now available that significantly reduce the amount of sample needed for structure determination. Studies that required heroic efforts and years to isolate enough material for NMR spectroscopic analysis a few decades ago can now be done with two, three, or more orders of magnitude less material today. This not only makes natural products research much more efficient, more importantly, it opens up possibilities for entirely new lines of scientific inquiry, involving individual variation, population chemical biology, and much more extensive examination of the influence of genetic or environmental factors in natural product expression levels.

Because of the advances in analytical technology, natural products chemists now have tremendous opportunities to take full advantage of 'omics'-type approaches. Genomics and proteomics technologies and databases make it much more feasible to both study and manipulate the biosynthesis of important natural products. This will allow a more complete understanding of basic biological processes but also will enable more efficient drug development that uses natural products as a starting point.

Computational power and databases constantly improve and will allow the design of more and more comprehensive approach to biological problems. Metabolomics has recently emerged as a key component of 'systems biology', and with new analytical, computational, and information technology, metabolomics is evolving into a central hub that connects many of the other 'omics' to better understand biology, and consequently, human health. As analytical and computational tools improve, the distinction between natural products chemistry and metabolomics is becoming less and less clear. Natural products represent both downstream products and upstream regulators of metabolic pathways, and as we learn more about these interactions, we learn more about function and possible applications of natural products. One of the current challenges in metabolomics studies is 'biomarker identification', a new term for natural products chemistry. As traditional natural products studies become more integrated with the 'omics' and as metabolomics becomes more focused on identifying key metabolites, the two fields will become less and less distinct.

### Abbreviations

| | |
|---|---|
| **2D-NMR** | two-dimensional NMR |
| **COLMAR** | complex liquid mixture analysis by NMR |
| **COSY** | correlation spectroscopy |
| **DANS** | differential analysis of 2D NMR spectra |
| **DECODES** | diffusion-encoded spectroscopy |
| **DNP** | dynamic nuclear polarization |
| **DOSY** | diffusion-ordered spectroscopy |
| **dqfCOSY** | double-quantum-filtered correlation spectroscopy |
| **E.COSY** | exclusive correlation spectroscopy |
| **gCOSY** | gradient correlation spectroscopy |
| **HMBC** | heteronuclear multiple-bond correlation |
| **HMQC** | heteronuclear multiple-quantum correlation |
| **HOHAHA** | homonuclear Hartmann–Hahn correlation |
| **HSQC** | heteronuclear single-quantum correlation |
| **INADEQUATE** | incredible natural abundance double quantum transfer experiment |
| **MS** | mass spectrometry |
| **NMR** | nuclear magnetic resonance |
| **NOESY** | nuclear Overhauser effect spectroscopy |
| **ROESY** | rotating-frame Overhauser effect spectroscopy |
| **SHY** | statistical heterospectroscopy |
| **STOCSY** | statistical correlation spectroscopy |
| **TOCSY** | total correlation spectroscopy |

## References


1. P. F. Wiley; K. Gerzon; E. H. Flynn; M. V. Sigal; O. Weaver; U. C. Quarck; R. R. Chauvette; R. Monahan, *J. Am. Chem. Soc.* **1957**, *79*, 6062–6070.
2. P. F. Wiley; M. V. Sigal; O. Weaver; R. Monahan; K. Gerzon, *J. Am. Chem. Soc.* **1957**, *79*, 6070–6074.
3. P. F. Wiley; R. Gale; C. W. Pettinga; K. Gerzon, *J. Am. Chem. Soc.* **1957**, *79*, 6074–6077.
4. M. Gronquist; J. Meinwald; T. Eisner; F. C. Schroeder, *J. Am. Chem. Soc.* **2005**, *127*, 10810–10811.
5. D. S. Dalisay; T. F. Molinski, *Org. Lett.* **2009**, *11*, 1967–1970.
6. D. S. Dalisay; B. I. Morinaka; C. K. Skepper; T. F. Molinski, *J. Am. Chem. Soc.* **2009**, *131*, 7552–7553.
7. D. S. Dalisay; E. W. Rogers; A. S. Edison; T. F. Molinski, *J. Nat. Prod.* **2009**, *72*, 732–738.
8. T. F. Molinski, *Curr. Opin. Drug Discovery Dev.* **2009**, *12*, 197–206.
9. J. Meinwald; D. F. Wiemer; T. Eisner, *J. Am. Chem. Soc.* **1979**, *101*, 3055–3060.
10. F. C. Schroeder; M. Gronquist, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 7122–7131.
11. M. Yoshida; M. Murata; K. Inaba; M. Morisawa, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14831–14836.
12. A. T. Dossey; S. S. Walse; J. R. Rocca; A. S. Edison, *ACS Chem. Biol.* **2006**, *1*, 511–514.
13. F. Zhang; A. T. Dossey; C. Zachariah; A. S. Edison; R. Bruschweiler, *Anal. Chem.* **2007**, *79*, 7748–7752.
14. H. Barjat; G. A. Morris; S. Smart; A. G. Swanson; S. C. R. Williams, *J. Magn. Reson. Ser. B* **1995**, *108*, 170–172.
15. K. Bleicher; M. F. Lin; M. J. Shapiro; J. R. Wareing, *J. Org. Chem.* **1998**, *63*, 8486–8490.
16. M. F. Lin; M. J. Shapiro, *J. Org. Chem.* **1996**, *61*, 7617–7619.
17. M. F. Lin; M. J. Shapiro; J. R. Wareing, *J. Org. Chem.* **1997**, *62*, 8930–8931.
18. M. F. Lin; M. J. Shapiro; J. R. Wareing, *J. Am. Chem. Soc.* **1997**, *119*, 5249–5250.
19. M. G. Lin; M. J. Shapiro, *Anal. Chem.* **1997**, *69*, 4731–4733.
20. A. E. Taggi; J. Meinwald; F. C. Schroeder, *J. Am. Chem. Soc.* **2004**, *126*, 10364–10369.
21. F. C. Schroeder; A. E. Taggi; M. Gronquist; R. U. Malik; J. B. Grant; T. Eisner; J. Meinwald, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14283–14287.
22. J. Srinivasan; F. Kaplan; R. Ajredini; C. Zachariah; H. T. Alborn; P. E. Teal; R. U. Malik; A. S. Edison; P. W. Sternberg; F. C. Schroeder, *Nature* **2008**, *454*, 1115–1118.
23. C. Pungaliya; J. Srinivasan; B. W. Fox; R. U. Malik; A. H. Ludewig; P. W. Sternberg; F. C. Schroeder, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7708–7713.
24. R. A. Butcher; F. C. Schroeder; M. A. Fischbach; P. D. Straight; R. Kolter; C. T. Walsh; J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1506–1509.
25. F. Schroder; V. Sinnwell; H. Baumann; M. Kaib, *Chem. Commun.* **1996**, 2139–2140.
26. F. Schroder; V. Sinnwell; H. Baumann; M. Kaib; W. Francke, *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 77–80.
27. E. Holmes; O. Cloarec; J. K. Nicholson, *J. Proteome Res.* **2006**, *5*, 1313–1320.
28. D. J. Crockford; E. Holmes; J. C. Lindon; R. S. Plumb; S. Zirah; S. J. Bruce; P. Rainville; C. L. Stumpf; J. K. Nicholson, *Anal. Chem.* **2006**, *78*, 363–371.
29. D. B. Kell; M. Brown; H. M. Davey; W. B. Dunn; I. Spasic; S. G. Oliver, *Nat. Rev. Microbiol.* **2005**, *3*, 557–565.
30. J. K. C. Nicholson; L. John; C. John; E. Holmes, *Nat. Rev. Drug Discov.* **2002**, *1*, 153–161.
31. B. J. Blaise, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19808.
32. Y. Wang; J.r. Utzinger; J. Saric; J. V. Li; J. Burckhardt; S. Dirnhofer; J. K. Nicholson; B. H. Singer; R. Brun; E. Holmes, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6127–6132.
33. S. Rochfort, *J. Nat. Prod.* **2005**, *68*, 1813–1820.
34. A. E. Derome, *Nat. Prod. Rep.* **1989**, *6*, 111–141.
35. W. F. Reynolds; R. G. Enriquez, *J. Nat. Prod.* **2002**, *65*, 221–244.
36. E. E. Kwan; S. G. Huang, *Eur. J. Org. Chem.* **2008**, 2671–2688.
37. T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, 1st ed.; Pergamon: Amsterdam: New York, 1999.
38. S. L. Robinette; F. Zhang; L. Bruschweiler-Li; R. Bruschweiler, *Anal. Chem.* **2008**, *80*, 3606–3611.
39. F. Zhang; R. Bruschweiler, *ChemPhysChem* **2004**, *5*, 794–796.
40. S. P. Rucker; A. J. Shaka, *Mol. Phys.* **1989**, *68*, 509–517.
41. A. Bax; D. G. Davis, *J. Magn. Reson.* **1985**, *65*, 355–360.
42. A. Bax, *Methods Enzymol.* **1989**, *176*, 151–168.
43. R. A. Butcher; M. Fujita; F. C. Schroeder; J. Clardy, *Nat. Chem. Biol.* **2007**, *3*, 420–422.
44. C. Griesinger; O. W. Sorensen; R. R. Ernst, *J. Magn. Reson.* **1987**, *75*, 474–492.
45. B. Vogeli; L. Yao; A. Bax, *J. Biomol. NMR* **2008**, *41*, 17–28.
46. F. C. Schroeder; T. Tolasch, *Tetrahedron* **1998**, *54*, 12243–12248.
47. T. D. Claridge; I. Perez-Victoria, *Org. Biomol. Chem.* **2003**, *1*, 3632–3634.
48. D. Neuhaus; M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 2nd ed.; Wiley-VCH: New York, 2000.
49. A. Bax; R. Freeman; T. A. Frenkiel, *J. Am. Chem. Soc.* **1981**, *103*, 2102–2104.
50. W. Francke; F. Schroeder; F. Walter; V. Sinnwell; H. Baumann; M. Kaib, *Liebigs Ann.* **1995**, *0*, 965–977.
51. J. T. Arnold; S. S. Dharmatti; M. E. Packard, *J. Chem. Phys.* **1951**, *19*, 507.
52. M. E. Packard; J. T. Arnold, *Phys. Rev.* **1951**, *83*, 210–211.
53. R. R. Ernst, *Biosci. Rep.* **1992**, *12*, 143–187.
54. N. Bross-Walch; T. Kuhn; D. Moskau; O. Zerbe, *Chem. Biodivers.* **2005**, *2*, 147–177.
55. P. L. Rinaldi, *Analyst* **2004**, *129*, 687–699.
56. F. Schroeder; S. Franke; W. Francke; H. Baumann; M. Kaib; J. M. Pasteels; D. Daloze, *Tetrahedron* **1996**, *52*, 13539–13546.
57. F. C. Schroeder; S. R. Smedley; L. K. Gibbons; J. J. Farmer; A. B. Attygalle; T. Eisner; J. Meinwald, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13387–13391.

58. F. C. Schroder; J. J. Farmer; A. B. Attygalle; S. R. Smedley; T. Eisner; J. Meinwald, *Science* **1998**, *281*, 428–431.
59. J. McCormick; Y. Li; K. McCormick; H. I. Duynstee; A. K. van Engen; G. A. van der Marel; B. Ganem; J. H. van Boom; J. Meinwald, *J. Am. Chem. Soc.* **1999**, *121*, 5661–5665.
60. S. Grzesiek; H. Dobeli; R. Gentz; G. Garotta; A. M. Labhardt; A. Bax, *Biochemistry* **1992**, *31*, 8180–8190.
61. F. C. Schroeder; D. M. Gibson; A. C. Churchill; P. Sojikul; E. J. Wursthorn; S. B. Krasnoff; J. Clardy, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 901–904.
62. F. E. Koehn, *Prog. Drug Res.* **2008**, *65*, *175,* 177–210.
63. K. Biemann, *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1254–1272.
64. L. Smith; J. Novak; J. Rocca; S. McClung; J. D. Hillman; A. S. Edison, *Eur. J. Biochem.* **2000**, *267*, 6810–6816.
65. V. H. Wysocki; K. A. Resing; Q. Zhang; G. Cheng, *Methods* **2005**, *35*, 211–222.
66. K. Adermann; H. John; L. Standker; W. G. Forssmann, *Curr. Opin. Biotechnol.* **2004**, *15*, 599–606.
67. F. C. Schroeder; D. M. Gibson; A. C. L. Churchill; P. Sojikul; E. J. Wursthorn; S. B. Krasnoff; J. Clardy, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 901–904.
68. J. Moldenhauer; X. H. Chen; R. Borriss; J. Piel, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 8195–8197.
69. F. C. Schroeder, *ACS Chem. Biol.* **2006**, *1*, 198–200.
70. P. Y. Jeong; M. Jung; Y. H. Yim; H. Kim; M. Park; E. Hong; W. Lee; Y. H. Kim; K. Kim; Y. K. Paik, *Nature* **2005**, *433*, 541–545.
71. R. Bruschweiler; F. Zhang, *J. Chem. Phys.* **2004**, *120*, 5253–5260.
72. R. Bruschweiler, *J. Chem. Phys.* **2004**, *121*, 409–414.
73. F. Zhang; L. Bruschweiler-Li; S. L. Robinette; R. Bruschweiler, *Anal. Chem.* **2008**, *80*, 7549–7553.
74. F. L. Zhang; R. Bruschweiler, *ChemPhysChem* **2004**, *5*, 794–796.
75. S. L. Robinette; F. Zhang; L. Brüschweiler-Li; R. Brüschweiler, *Anal. Chem.* **2008**, *80*, 3606–3611.
76. D. A. Snyder; F. Zhang; S. L. Robinette; L. Brüschweiler-Li; R. Brüschweiler, *J. Chem. Phys.* **2008**, *128*, 052313.
77. B. R. Seavey; E. A. Farr; W. M. Westler; J. L. Markley, *J. Biomol. NMR* **1991**, *1*, 217–236.
78. E. L. Ulrich; H. Akutsu; J. F. Doreleijers; Y. Harano; Y. E. Ioannidis; J. Lin; M. Livny; S. Mading; D. Maziuk; Z. Miller; E. Nakatani; C. F. Schulte; D. E. Tolmie; R. Kent Wenger; H. Yao; J. L. Markley, *Nucleic Acids Res.* **2008**, *36*, D402–D408.
79. J. L. Markley; M. E. Anderson; Q. Cui; H. R. Eghbalnia; I. A. Lewis; A. D. Hegeman; J. Li; C. F. Schulte; M. R. Sussman; W. M. Westler; E. L. Ulrich; Z. Zolnai, *Pac. Symp. Biocomput.* **2007**, *12*, 157–168.
80. Q. L. Cui; A. Ian; A. D. Hegeman; M. E. Anderson; J. Li; C. F. Schulte; W. M. Westler; H. R. Eghbalnia; M. R. Sussman; J. L. Markley, *Nat. Biotechnol.* **2008**, *26*, 162–164.
81. D. S. Wishart; D. Tzur; C. Knox; R. Eisner; A. Guo; N. Young; D. Cheng; K. Jewell; D. Arndt; S. Sawhney; C. Fung; L. Nikolai; M. Lewis; M. Coutouly; I. Forsythe; P. Tang; S. Shrivastava; K. Jeroncic; P. Stothard; G. Amegbey; D. Block; D. Hau; J. Wagner; J. Miniaci; M. Clements; M. Gebremedhin; N. Guo; Y. Zhang; G. E. Duggan; G. D. Macinnis; A. M. Weljie; R. Dowlatabadi; F. Bamforth; D. Clive; R. Greiner; L. Li; T. Marrie; B. D. Sykes; H. J. Vogel; L. Querengesser, *Nucleic Acids Symp. Ser.* **2007**, *35*, D521–D526.
82. J. Munger; B. D. Bennett; A. Parikh; X.-J. Feng; J. McArdle; H. A. Rabitz; T. Shenk; J. D. Rabinowitz, *Nat. Biotechnol.* **2008**, *26*, 1179–1186.
83. J. K. Nicholson; J. C. Lindon, *Nature* **2008**, *455*, 1054–1056.
84. A. Couto Alves; M. Rantalainen; E. Holmes; J. K. Nicholson; T. M. Ebbels, *Anal. Chem.* **2009**, *81*, 2075–2084.
85. A. D. Maher; O. Cloarec; P. Patki; M. Craggs; E. Holmes; J. C. Lindon; J. K. Nicholson, *Anal. Chem.* **2009**, *81*, 288–295.
86. A. D. Maher; D. Crockford; H. Toft; D. Malmodin; J. H. Faber; M. I. McCarthy; A. Barrett; M. Allen; M. Walker; E. Holmes; J. C. Lindon; J. K. Nicholson, *Anal. Chem.* **2008**, *80*, 7354–7362.
87. E. Holmes; R. L. Loo; O. Cloarec; M. Coen; H. R. Tang; E. Maibaum; S. Bruce; Q. Chan; P. Elliott; J. Stamler; I. D. Wilson; J. C. Lindon; J. K. Nicholson, *Anal. Chem.* **2007**, *79*, 2629–2640.
88. D. I. Hoult; R. E. Richards, *J. Magn. Reson.* **1976**, *24*, 71–85.
89. B. Shapira; K. Shetty; W. W. Brey; Z. H. Gan; L. Frydman, *Chem. Phys. Lett.* **2007**, *442*, 478–482.
90. D. L. Olson; T. L. Peck; A. G. Webb; R. L. Magin; J. V. Sweedler, *Science* **1995**, *270*, 1967–1970.
91. D. Raftery, *Anal. Bioanal. Chem.* **2004**, *378*, 1403–1404.
92. A. G. Webb, *J. Pharm. Biomed. Anal.* **2005**, *38*, 892–903.
93. W. W. Brey; A. S. Edison; R. E. Nast; J. R. Rocca; S. Saha; R. S. Withers, *J. Magn. Reson.* **2006**, *179*, 290–293.
94. D. L. Olson; J. A. Norcross; M. O'Neil-Johnson; P. F. Molitor; D. J. Detlefsen; A. G. Wilson; T. L. Peck, *Anal. Chem.* **2004**, *76*, 2966–2974.
95. H. Kovacs; D. Moskau; M. Spraul, *Prog. Nuclear Magn. Reson. Spectrosc.* **2005**, *46*, 131–155.
96. W. W. Brey; W. Anderson; W. H. Wong; L. F. Fuks; V. Y. Kotsubo; R. S. Withers, Nuclear Magnetic Resonance Probe Coil. U.S. Patent 5,565,778, 1996.
97. W. A. Anderson; W. W. Brey; A. L. Brooke; B. Cole; K. A. Delin; L. F. Fuks; H. D. W. Hill; M. E. Johanson; V. Kotsubo; R. Nast; R. S. Withers; W. H. Wong, *Bull. Magn. Reson.* **1995**, *17*, 98–102.
98. A. T. Dossey; S. S. Walse; A. S. Edison, *J. Chem. Ecol.* **2008**, *34*, 584–590.
99. A. T. Dossey; S. S. Walse; O. V. Conle; A. S. Edison, *J. Nat. Prod.* **2007**, *70*, 1335–1338.
100. B. Wang; A. T. Dossey; S. S. Walse; A. S. Edison; K. M. Merz, Jr., *J. Nat. Prod.* **2009**, *72*, 709–713.
101. S. S. Walse; F. Lu; P. E. Teal, *J. Nat. Prod.* **2008**, *71*, 1726–1731.
102. S. Matthew; P. J. Schupp; H. Luesch, *J. Nat. Prod.* **2008**, *71*, 1113–1116.
103. J. C. Kwan; J. R. Rocca; K. A. Abboud; V. J. Paul; H. Luesch, *Org. Lett.* **2008**, *10*, 789–792.
104. A. B. Barnes; G. D. Paepe; P. C. van der Wel; K. N. Hu; C. G. Joo; V. S. Bajaj; M. L. Mak-Jurkauskas; J. R. Sirigiri; J. Herzfeld; R. J. Temkin; R. G. Griffin, *Appl. Magn. Reson.* **2008**, *34*, 237–263.
105. M. J. Prandolini; V. P. Denysenkov; M. Gafurov; B. Endeward; T. F. Prisner, *J. Am. Chem. Soc.* **2009**, *131*, 6090–6092.
106. J. H. Ardenkjaer-Larsen; B. Fridlund; A. Gram; G. Hansson; L. Hansson; M. H. Lerche; R. Servin; M. Thaning; K. Golman, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 10158–10163.
107. L. Frydman; D. Blazina, *Nat. Phys.* **2007**, *3*, 415–419.

## Biographical Sketches

Arthur S. Edison obtained a B.S. in chemistry from the University of Utah, where he studied monoterpenes isolated from southern Utah sagebrush by NMR. He completed his Ph.D. in biophysics from the University of Wisconsin, Madison, where he developed and applied NMR methods for peptide and protein structural studies under the supervision of John Markley and Frank Weinhold. In 1993, Dr. Edison joined the laboratory of Anthony O. W. Stretton at the University of Wisconsin as a Jane Coffin Childs postdoctoral fellow where he investigated the role of neuropeptides in the nervous system of the parasitic nematode *Ascaris suum.* He joined the faculty at the University of Florida and the National High Magnetic Field Laboratory in 1996 and is currently the Director of Chemistry & Biology at the NHMFL. Dr. Edison's current research is in technology development for high-sensitivity NMR and natural product discovery in nematodes and other invertebrates. Dr. Edison is the recipient of the 1997 American Heart Association Robert J. Boucek Award, a CAREER Award from the National Science Foundation in 1999, and, with his postdoctoral scientist Aaron Dossey, the Beal award for the best publication of the year in the *Journal of Natural Products* in 2007.

Frank C. Schroeder studied chemistry and physics at the University of Hamburg, where he worked under the guidance of Wittko Francke. He received his doctorate in 1998 for studies on structures and functions of insect-derived natural products, which included the serendipitous discovery of a group of structurally complex ant alkaloids, the myrmicarins. During his graduate studies, he developed a deep appreciation for NMR spectroscopy as a tool in natural products chemistry and metabolomics. He continued to develop new analytical methodology for characterizing structures and functions of small molecule metabolites as a postdoc and later research associate with Jerrold Meinwald at Cornell University and Jon Clardy at Harvard Medical School. In August 2007, he joined the faculty of Cornell

University's Boyce Thompson Institute and the Cornell Department of Chemistry and Chemical Biology.

Dr. Schroeder's research aims to develop NMR spectroscopy-based approaches that complement or enhance traditional methodology by enabling detailed characterization of small molecule metabolites in complex biological samples, with regard to both chemical structure and biological function. His current work focuses on a comprehensive structural and functional annotation of the metabolome of the model organism *Caenorhabditis elegans*.