Chapter 1

Drug Discovery and Development: The Role of NMR

Jeremy R. Everett

Medway Metabonomics Research Group, University of Greenwich, Chatham Maritime, Kent ME4 4TB, UK

1.1 INTRODUCTION TO DRUG DISCOVERY AND DEVELOPMENT AND THE ROLE OF NMR

Drug discovery and development [also known as research and development (R&D)] is at a crossroads. Over the past 10 years, many major pharmaceutical companies have reduced the size of their R&D organizations as they struggled to achieve the increased outputs expected of them in this postgenomic era. Recent analyses have suggested that the common strategy of companies merging and acquiring other companies has not led to the improvements in output that were expected.[1] Indeed, until recently, the number of new drugs being launched per year was falling in spite of increasing expenditure on R&D.[2] The key reasons for the low productivity across the industry have been the very significant project attrition at all stages of the R&D process due to lack of efficacy and lack of safety in new drug candidates. This attrition has been particularly acute in the critical Phase II clinical trials, where drugs are tested in patients for the first time. We are now in a new era of drug discovery and development, with far greater involvement from academic laboratories and better cooperation between companies and academia, as well as between companies themselves, especially in precompetitive areas such as the development of new safety testing methodologies.

Before moving on to consider the role of NMR in this challenging environment, it is worthwhile to consider the different stages of the drug R&D process. These stages can be represented as follows:

1. Disease selection
2. Target selection and validation (the target is the name for the biological macromolecule in the human body that the drug is designed to interact with to produce clinical benefit: this is usually a protein)
3. Hit discovery against that target
4. Hit-to-lead optimization

NMR in Pharmaceutical Sciences. Edited by Jeremy R. Everett, John C. Lindon, Ian D. Wilson, and Robin K. Harris

ISBN: 978-1-118-66025-6
Also published in eMagRes (online edition)
DOI: 10.1002/9780470034590.emrstm1389

5. Lead-to-drug optimization
6. Phase I drug safety studies in human volunteers
7. Phase II drug efficacy and safety studies in small cohorts of patients
8. Phase III efficacy and safety studies in larger patient populations
9. Drug launch
10. Postlaunch support

The most critical stage (and the most difficult) is stage 2: target selection and validation. For most small molecule drug discovery projects, the choice of which protein target to prosecute (via its inhibition, agonism, antagonism, partial agonism, etc.) is absolutely vital. Many projects fail right from this stage (although the project team will not know this for some years!) because the protein target chosen will not produce sufficient clinical benefit in the patient, or will only produce clinical benefit together with unacceptable side effects. The reason for these problems is that many drug discovery targets are unprecedented; that is, at the time of launch of the project, there are no Phase II data showing that modulation of that target gives clinical benefit in the absence of side effects. Drug discovery project teams therefore have to make key go/no-go decisions on target selection using less-reliable genomics or animal data.

If drug discovery and development are in difficulties, NMR, by contrast, is not. In the 70 years or so since its discovery, NMR spectroscopy has progressed quickly to become the most powerful method of molecular structure elucidation in both the solution state and the amorphous or heterogeneous solid state, and an indispensable tool in chemistry and biochemistry. In addition, the more recent developments of whole body MRI and magnetic resonance spectroscopy (MRS) have enabled the influence of NMR to spread from chemistry and biochemistry to biology and medicine, so that NMR is now pivotal to drug R&D, and the array of NMR applications stretches across each of the 10 stages of the R&D process listed earlier, up to and beyond drug launch.

The reason that NMR spectroscopy has become so dominant in solution-state molecular structure elucidation is that the information it provides is rich, coherent, and highly stable and reproducible over time. Information from proton, carbon-13, and fluorine-19 NMR chemical shifts, for instance, can be interpreted in terms of the chemical environment of the atoms giving rise to those NMR signals. In addition, the phenomenon of spin–spin coupling enables the establishment of through-bond connectivities between different atoms in a molecule in a way that is unobtainable with any other technique. Finally, the relaxation properties of the nuclear spins under study give information on the motions, dynamics, and through-space connectivities in molecules that is critical for a thorough understanding of their properties and behavior in solution.

The drawbacks to the use of NMR principally arise from its lack of sensitivity relative to some other techniques such as mass spectrometry (MS). It is actually appropriate to tackle this principal issue with NMR spectroscopy straight away, as tremendous progress has been made in this area. First of all, superconducting magnets of increasing field strength have been developed over the past 20 years. In the 1980s, NMR spectrometers operating at 9.4 T (resonance frequency of 400 MHz for protons) were common. Magnets operating at 18.8 T (800 MHz for protons) emerged in the 1990s, and the first 1 GHz system was installed in France in 2009. As the signal-to-noise ratio of a spectrum increases approximately with the 3/2 power of the field strength used, these increases in field strength have resulted in significant gains in sensitivity (a doubling in sensitivity from 600 to 900 MHz for protons) but, unfortunately, at the expense of a considerable rise in spectrometer costs. Therefore, other developments have been pursued. One of these is the development of cryocooled probes, where the receiver coil and preamplifier are cooled down to temperatures as low as about 20 K using liquid helium. This has resulted in sensitivity gains of about three- to four-fold, relative to running the same sample on a conventional probe at room temperature. Although technically demanding, switching to cryoprobes is much more cost effective in terms of sensitivity increase than moving to higher magnetic field strengths. A further development has been in the area of microcoil probes. Reducing the size of the receiver coils in NMR probes is particularly important for samples that have limited mass or volume, as, for constant coil length-to-volume ratio, sensitivity is inversely proportional to coil diameter. A standard 5 mm NMR probe may require a sample volume of about 600 μl, whereas a microcoil probe may allow measurements on volumes as low as about 1 μl. This approach has also been exploited for the coils in flow-detection experiments such as liquid chromatography–nuclear magnetic resonance (LC–NMR), capillary electrophoresis–nuclear magnetic resonance (CE–NMR), or even lab-on-a-chip.[3] Significant improvements in spectrometer performance have also ensued from the development of pulse field gradient technologies that allow for the selection

or nonselection of resonances with much greater fidelity than previous phase-cycling approaches. In the solid-state nuclear magnetic resonance (SSNMR) arena, the development of rapid-spinning technologies combined with new homonuclear decoupling methods has enabled the acquisition of high-quality ^{1}H SSNMR spectra with reduced $^{1}H-^{1}H$ dipolar coupling and increased chemical shift resolution. An array of solid-state homonuclear and heteronuclear 2D NMR experiments, based on correlations via J-coupling and dipolar coupling, are now available, meaning that the study of molecular structure in the solid state can employ many of the same 2D NMR approaches used in the solution state. Finally, and importantly, new data collection methodologies have emerged, including nonlinear data sampling and ultrafast pulsing, that can reduce experiment times by orders of magnitudes.[4] All of these advances have meant that low sensitivity is no longer a major issue for modern NMR spectroscopy. See Chapters 2 and 3, as these both have further important information.

In the remainder of this article, we will cover the principal uses of NMR in drug R&D with a greater focus on new and emerging areas.

1.2 NMR SPECTROSCOPY IN DRUG DISCOVERY AND DEVELOPMENT

1.2.1 Molecular Structure Elucidation by NMR Spectroscopy

1.2.1.1 Structure Confirmation by NMR Spectroscopy

The key area of application for NMR spectroscopy in drug R&D is, and has always been, its role in small molecule structure elucidation. There are several aspects to this area, including the structure confirmation of small molecules made by medicinal chemistry teams in drug discovery and by synthetic chemistry teams in drug development. This work is relatively routine, assuming that the molecules are of low molecular weight and reasonably pure, especially now that modern NMR processing software allows the automated zero-filling, apodization, FT, baseline correction, integration, and multiplet analysis of simple proton NMR spectra. Indeed, much of this work is done on automated or open-access systems for nonexpert users of the spectrometer. Some NMR software such as MNova[5] will also allow the user to predict the NMR spectrum of the molecule under study and give an indication of the match between the predicted and experimental spectra, thus giving increased confidence in the structure confirmation, assuming that the prediction calculations are precise enough (see Chapters 6, 11, and 14).

1.2.1.2 Structure Elucidation by NMR Spectroscopy: Natural Products

The de novo structure elucidation of natural products, when limited information is available on the expected structure, is much more challenging than structure confirmation of expected molecules. Typically, the structure confirmation of a molecule made by an established synthetic chemistry methodology can be performed with a simple 1D proton NMR spectrum (provides an identity check and purity information) and a mass spectrum of some kind (identity check), while ultraviolet (UV) spectroscopy (provides chromophore of the molecule), infrared (IR) spectroscopy (provides functional groups in the molecule), and elemental analysis (provides elemental formula and confirmation of good purity) information is optional but sometimes acquired. By contrast, the structure elucidation of a genuine unknown will require UV, IR, MS, and an array of NMR data. Typically, for a small molecule (of <500 Da molecular weight), the NMR data acquired would include the information shown in Table 1.1.

As an example of the use of NMR spectroscopy in natural product structure elucidation, Figure 1.1 shows the scanning electron micrograph of a streptomyces organism isolated from a soil sample under the River Mole in Surrey, UK. A whole series of novel anthelmintic (antiparasitic worm) milbemycin natural products were isolated from this organism,[6] including VM44857 shown in Figure 1.1(b). The expansion of the 2D ^{1}H COSY-45 NMR spectrum in Figure 1.1(c) shows some remarkable long-range proton-to-proton connectivities that were used to help piece fragments of the molecular structure together. Note in particular the four-bond and five-bond connectivities to CH_3-26 from H2, H3, and H5.

If a reasonable amount of the natural product is available, carbon-13 NMR spectroscopy could be used to directly interrogate the chemical environments of the carbon atoms in the molecule, but these experiments are significantly less sensitive and more time consuming, owing to the low natural abundance and low sensitivity of carbon-13 NMR compared with proton NMR.

Table 1.1. The different NMR experiments generally required for the elucidation of the molecular structure of an unknown small molecule and the information provided by those experiments

NMR experiment	Information provided
1D ^{1}H NMR	• Number of protons • Chemical type of protons • Proton-to-proton coupling connectivities in simple cases • Information on molecular conformations and stereochemistry through values of coupling constants
2D ^{1}H COSY (correlated spectroscopy)	Proton-to-proton coupling connectivities typically over two to four bonds, sometimes more
2D ^{13}C, ^{1}H HSQC	Carbon-13 to proton coupling connectivities over one bond
2D ^{13}C, ^{1}H HMBC	Carbon-13 to proton coupling connectivities over two to three bonds, allowing connections to be made over the so-called spectroscopically silent centers such as oxygen atoms and carbonyl groups

HSQC, heteronuclear single quantum coherence spectroscopy; HMBC, heteronuclear multiple bond correlation spectroscopy.

A major issue with drug discovery using natural product approaches, apart from the work in elucidating the structure of the molecules, has been the issue of dereplication, i.e., how do you know when you are working on a genuinely novel natural product and not one that has already been discovered previously? Some databases are available but this area is problematic. Approaches that combine classical natural product discovery with metabolite profiling methods are being pursued. Essentially, these approaches use multivariate statistical methods borrowed from metabonomics to identify the metabolites responsible for differences in biological activity between different, partially purified natural product extracts, avoiding the need for the isolation of pure, individual natural products initially (see Chapter 13).[7,8]

1.2.1.3 Structure Elucidation by NMR Spectroscopy: Impurities and Degradation Products

The structure elucidation challenges here are very similar to those for natural products and consequently very similar approaches are taken. The issue is that drugs can degrade in surprising ways and extremely unexpected structures can arise as degradation products. A good example of this is the study of the degradation of the penicillins oxacillin and nafcillin in aqueous solution.[9]

A golden yellow precipitate was observed during the study of the degradation of both oxacillin and nafcillin (Figure 1.2). A combination of UV, IR, MS, and, critically, 2D NMR methodologies was used to establish that the degradation products were unusual thietan-2-one structures formed by dehydration and rearrangement in aqueous solution. The critical information to piece the structure together was provided by long-range hydrogen-to-carbon-13 connectivities that established the connections between carbons 2 to 8. An added complication to the structure elucidation was that these thietan-2-ones exist as a mixture of two isomers, *cis*- and *trans*-, in slow exchange about the C6C7 double bond. These structures were considered so unusual when proposed from the NMR data that X-ray crystallography of the nafcillin degradation product and two independent chemical syntheses of the oxacillin degradation product were used to confirm their identities. This work took several months: the identification of the thietan-2-ones was achieved in 48 h by a combination of spectroscopic methods but principally by NMR![9]

Before leaving this section, we should note that one of the great powers of NMR spectroscopy is its ability to discriminate between different isomers and tautomers of a compound. This could be seen in the abovementioned example where two isomers of the thietan-2-one degradation product about the C6C7 double bond were in dynamic exchange with one another. However, many other forms of isomerism are readily determined by NMR spectroscopy, including ring and chain positional isomers, diastereomers formed by two or more chiral centers in a molecule, sugar anomers, and *cis* and *trans* amide isomers.

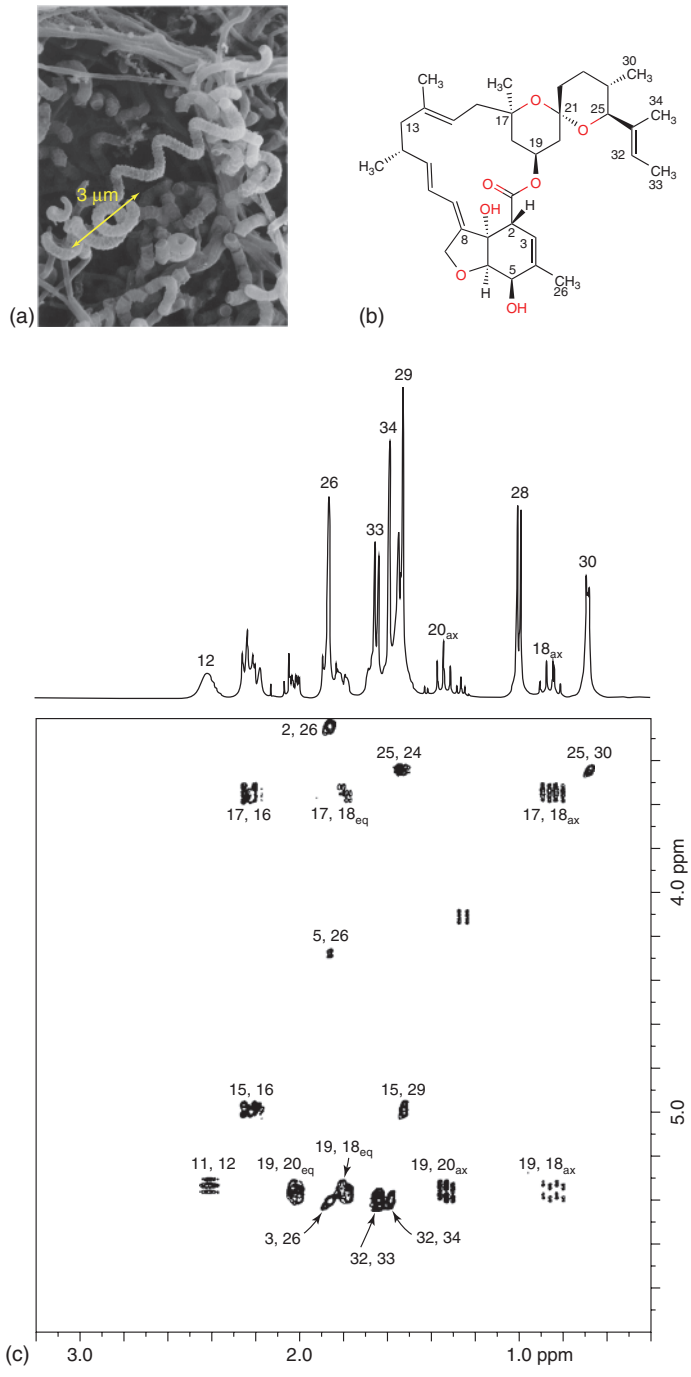

Figure 1.1. (a) A scanning electron micrograph of the *Streptomyces hygroscopicus aureolacrimosus* organism (code E225), isolated from an underwater soil sample in the River Mole, Surrey, that produced a number of milbemycins including VM44857. (b) The molecular structure of VM44857. (c) A portion of the 400 MHz 2D ^{1}H COSY-45 NMR spectrum of VM44857, displayed as a contour plot, showing some key long-range connectivities that helped elucidate the molecule structure. (Reproduced with permission from Ref. 6. © Nature Publishing Group, 1990)

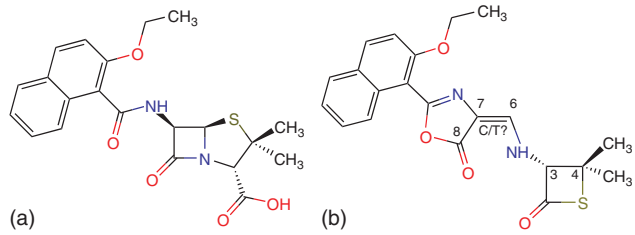

Figure 1.2. The molecular structure of nafcillin free acid (a) and nafcillin thietan-2-one degradation product (b). Note that the double bond between C6 and C7 in nafcillin thietan-2-one is in equilibrium between cis and trans forms

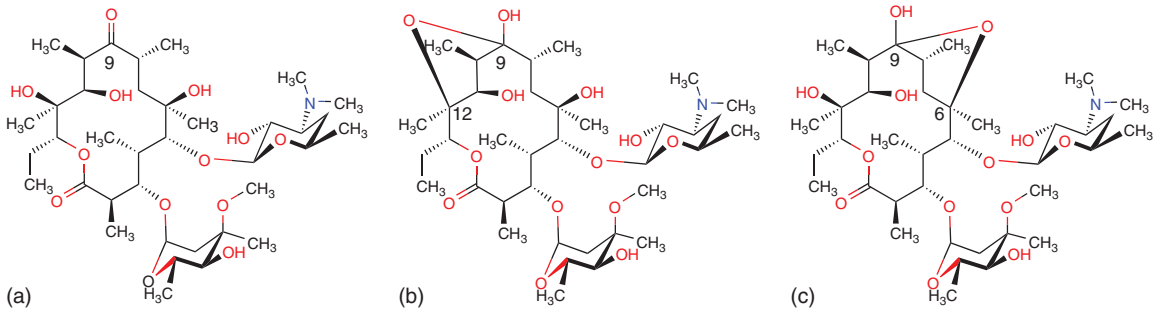

Figure 1.3. Three tautomers of erythromycin A: the C9 ketone tautomer (a), the 9,12-cyclic hemiacetal (b), and the 6,9-cyclic hemiacetal (c). Note that some bond lengths are distorted for clarity

In addition, NMR spectroscopy is uniquely powerful in establishing the tautomeric form in which drugs exist in solution, and this can be critical in the defense or assertion of drug patent claims (see Chapter 27). As an example of tautomerism in drug molecules, the antibiotic erythromycin A exists as a mixture of the dominant C9 ketone form (Figure 1.3) and two hemi-acetals; one 6,9-cyclic hemiacetal (9-deoxo-6-deoxy-9-hydroxy-6,9-epoxyerythromycin A) and one 9,12-cyclic hemiacetal (9-deoxo-12-deoxy-9-hydroxy-9,12-epoxyerythromycin A) formed by the interaction of the hydroxyl groups on C6 and C12, respectively, with the C9 ketone and subsequent cyclic hemiacetal formation.[10]

Finally, it is even possible to distinguish enantiomers using NMR spectroscopy. Ordinarily, NMR experiments are conducted in an achiral environment, and the signals from the two enantiomeric forms of a molecule, e.g., 2*R*, 3*R*-butanediol and 2*S*, 3*S* butanediol, will be identical (Figure 1.4).

However, in a chiral environment, the NMR signals of the two enantiomers will be different. This can be achieved by the addition of a chiral shift reagent to the solution under study. Note that there is a third isomer of

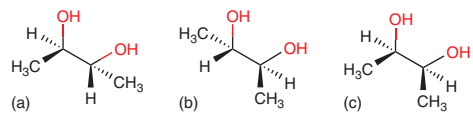

Figure 1.4. The molecular structures of 2*R*, 3*R*-butanediol (a), 2*S*, 3*S*-butanediol (b), and meso-butanediol (c).

2,3-butanediol: the meso isomer. This is a diastereomer of the 2*R*, 3*R* and 2*S*, 3*S* enantiomers and is distinguishable from both of them in an achiral environment. Note that for the meso isomer, the stereochemistry is 2*R*, 3*S* or 2*S*, 3*R*: these two molecules are identical! See Chapter 17 for additional information on this important area.

1.2.2 Additional Important Roles for NMR Spectroscopy in Drug Discovery and Development

There are many other significant roles for NMR spectroscopy in addition to molecular structure elucidation

that are of key importance for drug discovery and development.

1.2.2.1 Conformational Analysis of Drug Molecules

In the previous section, there is a discussion of the phenomenal ability of NMR spectroscopy to distinguish between the various different isomeric forms that molecules can have. These isomers typically require bond breaking and making in order to convert one isomer into another. However, there are other forms of a molecule that can be interconverted by the simple act of rotation around single bonds and these are known as *conformational isomers* or *conformers*. One well-known example of this is the conformational equilibrium between the axial and equatorial conformations of 1-methylcyclohexane (Figure 1.5).

Rotation of the cyclohexane ring single bonds will convert the equatorial methyl conformation into the axial methyl conformation, which is less favored and of higher free energy owing to steric interactions with the axial hydrogen atoms at C3 and C5. At room temperature, the two conformations are in rapid exchange with one another, and NMR spectra of solutions of 1-methylcyclohexane at this temperature display a weighted average of the spectra. However, if a solution of 1-methylcyclohexane is cooled to below −120 °C, then the rate of exchange of the two conformations is slowed down to such an extent that separate NMR signals for the two conformations can start to be observed and their relative abundance directly quantified.[11]

Understanding the solution conformations of drug molecules is an important aspect of drug design. If a drug molecule has a great degree of conformational freedom, then there will be an entropic penalty to binding to their protein target.

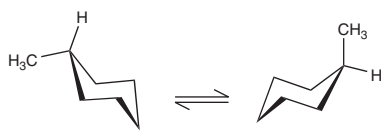

Figure 1.5. The two stable conformations of methylcyclohexane: equatorial methyl conformation (low free energy, left) and axial methyl conformation (higher free energy, right)

On the other hand, if the drug molecule prefers to adopt a solution conformation that is different from that in the protein binding site, then there may be an enthalpic penalty to the binding. Both situations are nonoptimal, and medicinal chemists seek to design molecules that are preorganized into the preferred binding conformation, as this can lead to significant increases in potency. There is now a resurgence of interest in new NMR-based methods for the determination of the experimental solution-state conformations of leads or hits in drug discovery projects as this (i) avoids the need to conduct protein structure determination in the crystalline state (n.b., most structure-based drug design uses X-ray crystallographic approaches as they are faster and more versatile than the corresponding NMR-based approaches), (ii) avoids the need to rely on small-molecule X-ray crystal structures that may represent conformations completely different from those in solution, and (iii) avoids the need to rely on computational chemistry methods that have significant limitations in terms of the force fields available and the ability to treat solvation realistically.[12] The new methods are based on combining traditional NMR information such as NOEs and scalar coupling constants with residual dipolar coupling information arising from measurements in aligned media (see Chapter 16).

1.2.2.2 Quality Control of Small Molecule Drugs

On moving from a drug discovery to a drug development environment, the focus changes to include a greater concern with the quality of the drug molecule. In particular, it is important to know its purity and the levels and identities of the impurities from its synthesis or fermentation, and the products that arise from post-synthesis degradation of the drug (and its impurities!). Many excellent analytical methods are used in this endeavor, including high-performance liquid chromatography (HPLC), ultra performance liquid chromatography (UPLC), liquid chromatography–mass spectrometry (LC–MS), and UPLC–MS. However, chromatographic methods on their own are only of any use if the identities of all of the peaks in the chromatograms are known. Furthermore, even LC–MS and UPLC–MS are of limited use in novel impurity or degradation product identification, as these technologies are generally unable to discriminate between isomers. In these

circumstances, it has been common practice to isolate the impurities and degradation products of interest and to determine their structures using NMR spectroscopy (see Section 1.2.1.3). Increasingly these days, such studies are using hyphenated technologies such as LC–NMR–MS, enabling integrated MS and NMR data to be obtained on each peak of interest in a complex mixture (see Chapter 15).

The need to quantify both drug purity and the levels of impurities and degradation products is served very well by NMR spectroscopy, which is an inherently quantitative technique. Indeed, a whole branch of NMR dedicated to this area is now termed *qNMR* (quantitative nuclear magnetic resonance). See Chapter 5 for details of various approaches to quantification, all of which rely on the fact that (under suitable experimental conditions) the signal area in NMR spectra is directly proportional both to the quantity of the molecule giving rise to that signal in the sample and to the number of nuclei being observed in that signal, e.g., a CH_2 group in a given molecule will give an NMR signal twice as big as a CH group, provided that the experiment is set up correctly. Borrowing methods from metabonomics, qNMR is being applied to increasingly complex mixtures.[13]

1.2.2.3 Quality Control of Biopharmaceuticals

Exactly the same attributes that make NMR spectroscopy such a good technique for the quality control of small molecules also make it an excellent tool for the quality control of biopharmaceuticals, typically protein-based drugs. This area has been reviewed recently,[4] so this article will not go into great detail. The key applications include various methods to verify the molecular structure of the protein, especially features such as the glycosylation pattern for glycoproteins. In addition, the open-window, unbiased observational power of NMR spectroscopy lends itself extremely well to determine product quality and the absence of any adulteration with small molecules or other biomacromolecules.

1.2.2.4 SSNMR Analyses of Leads and Drugs

This is an important area of application of NMR and use of the technology spans from late discovery right through the development phase of drug R&D. The principal use of SSNMR is to characterize the solid-state form of the drug. Many developments have occurred in the past decade in terms of new pulse sequences, faster sample rotation, and higher field spectrometers that have facilitated the success of these applications.[14] A typical use of SSNMR would be to investigate the occurrence of different polymorphs and solvates of a crystalline drug by ^{13}C SSNMR spectroscopy. In principle, each different polymorph or solvate will have a characteristic SSNMR spectrum, with different ^{13}C chemical shifts, owing to the unique packing arrangements of the molecules in the different crystal forms. The systematic exploration of these different forms by high-throughput crystallization methods is an important part of the early drug development process for two reasons. Firstly, it is critical that the solid-state form of the drug with optimal stability, solubility, formulation properties, and purity is taken forward for full development. Secondly, if a third-party competitor discovers a previously overlooked polymorph or solvate with significantly improved properties over the solid-state form of an existing agent, it is possible that they may derive intellectual property on that beneficial form of the molecule, even if they do not hold patent rights on the molecule itself, as first discovered.

It is important to mention here that SSNMR provides important information on the solid-state form of the drug not just in the drug substance but also in the drug product, when it is formulated with all of the excipients, and, of course, SSNMR can also give information on the solid-state form of all of the excipients. In addition, it can provide critical information on the stability of the drug solid-state form and can address the question of whether there is any transition between solid-state forms occurring on storage. Thus, SSNMR provides an important bridge between exploratory polymorph studies on the drug substance in the very early stages of development, right through to the demonstration that the solid-state form of the drug selected for final development is stable, both as the drug substance and as the drug product. SSNMR can, of course, also be used to monitor for degradation directly in the solid state without the need to use solution-state analytical methods.

The power of SSNMR in this area goes well beyond the examples cited earlier, however, and the technology is used in a host of different and important applications (see Chapter 21).

1.3 NEW APPLICATIONS OF NMR SPECTROSCOPY IN PHARMACEUTICAL R&D

1.3.1 Hit Discovery and Hit Follow-up Using NMR Spectroscopy

NMR Spectroscopy has always struggled as a tool for protein structure elucidation, although it has many attractive features such as the ability to provide information on dynamics as well as structure, because X-ray crystallography has been faster, provided higher resolution structures, avoided the requirement for isotope labeling of the proteins and required smaller samples. For example, the current protein data bank (http://www.wwpdb.org) has over 100 000 depositions, of which only about 10% derive from NMR studies, with most of the remainder from X-ray crystallography. However, NMR spectroscopy has become an important tool for hit finding against protein targets, which is itself a very important step close to the beginning of the drug discovery process.

Hit finding in the pharmaceutical industry has been based for a long time upon high-throughput screening (HTS) of large collections of compounds in compound libraries: the so-called screening files. Considerable investment went into HTS in the 1990s and 2000s, and the experience of large companies such as Pfizer was that HTS would deliver hits in >90% of screens and that it would initiate new hit-to-lead drug discovery projects in 50–70% of cases.[15] However, in spite of this success, HTS has acquired a bad press in some quarters, as it is unfairly blamed for the increasing molecular weight and lipophilicity of leads and drugs emerging from pharmaceutical R&D. This is inappropriate as the responsibility for this increase lies with the chemists designing those compounds and with the custodians of the screening files: the old maxim 'garbage in, garbage out' applies very well in HTS. Another issue with HTS is the expense of the screening equipment and material management operations required to support screening on a scale of hundreds of thousands or even millions of compounds in the screening files for each target screened. Here again, progress has been made in two areas to reduce the scale of these screening operations. Firstly, the development of the concept of molecular redundancy[15] has enabled the elimination of compounds that are too similar from the screening file, as these provide no new information and waste resources. Secondly, highly efficient subset screening methods such as plate-based diversity screening (PBDS) enable the coverage of the chemical space of the entire screening file while only screening a fraction of the screening plates: this innovation was developed simultaneously by groups in Pfizer and Novartis.[16] Nevertheless, there is a significant need for additional effective and efficient hit discovery methods, especially as academics become more involved in drug discovery and do not necessarily have access to the enormous resources and infrastructure that the pharmaceutical companies once did. In this context, huge efforts have gone into the development of a range of new hit-finding technologies that use NMR, X-ray crystallography, or other methods to detect binding of molecules to target proteins without recourse to the radioactive or fluorescent labeling of reagents required in conventional HTS.

The hit-finding methods using NMR can be divided into two categories: those detecting changes in the NMR signals from the protein when bound by a ligand and, conversely, those methods that detect changes in the NMR signals of a ligand when it experiences binding with a target protein molecule.

The ligand-detected hit-finding methods use ingenious NMR experiments such as saturation transfer difference (STD) and variants of this method like water–ligand observed via gradient spectroscopy (waterLOGSY) to determine which parts of a ligand are involved in binding with the protein binding site.[17] There are several attractive features to these methods including (i) the ability to screen ligands for binding interactions to a target protein in mixtures, improving the screening efficiency, (ii) the ability to screen at relatively low protein concentration, thus limiting protein reagent supply issues, (iii) the ability to screen the protein without the need for isotope labeling, which can be time consuming and expensive, (iv) the ability to derive at least some structural information on the mode of binding, and (v) the ability to derive binding constants directly from these NMR experiments.

The alternative approach is to titrate mixtures of ligands into a solution of an isotopically labeled protein and observe changes in the chemical shifts of protein resonances in contact with ligand molecules in the binding site. Typically, these experiments will require general or residue-selective ^{15}N, ^{2}H, or ^{13}C labeling of the protein in order to enable the observation of discrete resonances from, e.g., each amide N–H group in a uniformly ^{15}N-labeled protein. The observation of a binding event when screening mixtures will obviously require some deconvolution to determine which member of the ligand mixture was responsible for binding.

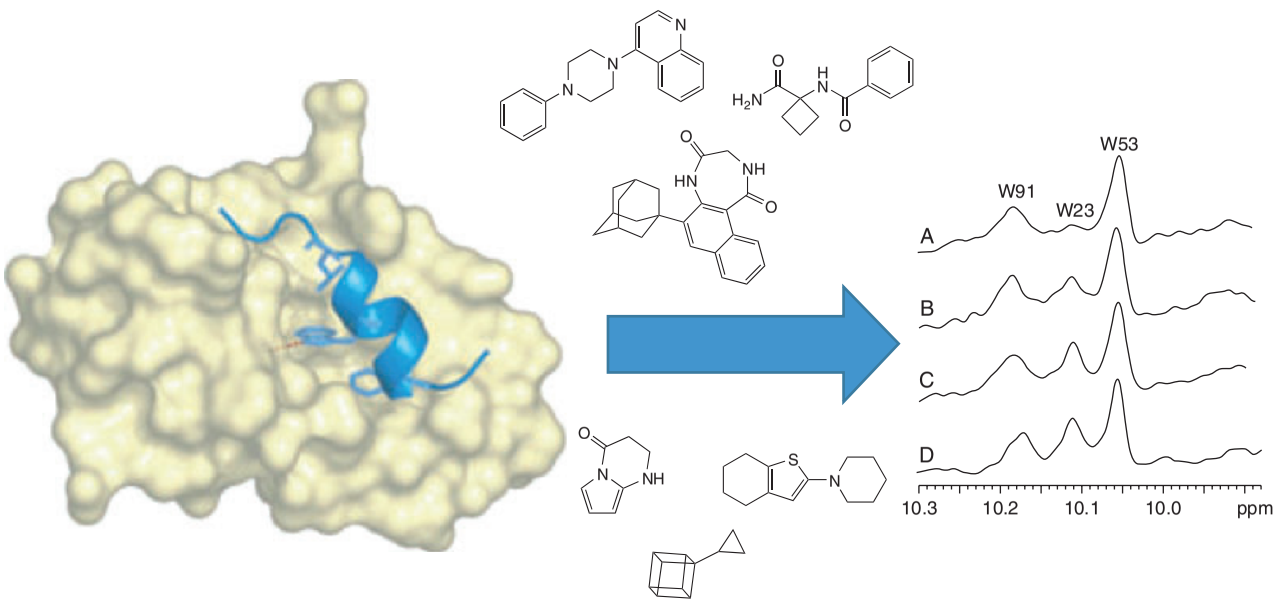

Figure 1.6. The use of antagonist-induced dissociation assay (AIDA) in order to find ligands that disrupt the binding of p53 (blue helix) to protein mdm2 (yellow surface). The proton NMR spectra labeled A to D show the signals for p53 in the presence of increasing ligand concentrations from A (no antagonist) and B to D (increasing concentrations of an antagonist present in the ligand mixture). As the protein–protein complex dissociates, the intensity of the signal due to tryptophan 23 (W23) in p53 increases significantly, as it is freed from its bound, buried position in the complex. (Reproduced with permission from Ref. 20. © John Wiley & Sons, Inc. 2010)

Chemical space is very large. Putting this another way, the number of drug-like, small molecules that can be designed is almost infinite and well beyond any current synthetic methodology. An issue with conventional ligand screening is that an enormous number of different molecules are required to cover a relatively small amount of chemical space. This has given rise to increased interest in fragment screening and then fragment-based drug design (FBDD), in which the molecules screened generally conform to the Rule of 3 rather than the well-known Rule of 5.[18] These fragment molecules are generally below molecular weight 300 Da, with a calculated partition coefficient clogP <3 and less than 3 hydrogen bond donors. The advantages of using fragments are that (i) fragment chemical space is much much smaller than drug space because of the smaller molecular size and it can thus be covered more completely with a reasonably sized library, (ii) fragments tend to be more promiscuous in their binding than larger molecules because they are unencumbered with lots of side chains and appendages that get in the way of a good binding interaction, and (iii) there is a perception that the ligand efficiency (binding energy per heavy atom) of fragments is higher than that of conventional, larger ligands, although in our experience that is not always the case. Utilizing fast pulsing methods, it has been reported that more than 200 2D heteronuclear correlation NMR spectra can be recorded in 24 h for fragment screening, using a 500 MHz NMR spectrometer with a cryoprobe, 40 μM protein concentration and a relatively small, globular protein of 15 kDa.[18]

These general methods can also be applied to the study of protein–carbohydrate interactions[19] and to the study of complexes with membrane proteins[20] (Figure 1.6).

This is an exciting and important area, and the first drug emerging from a fragment-based drug discovery approach, Vemurafenib (Figure 1.7), was approved in 2011 (see Chapter 12).

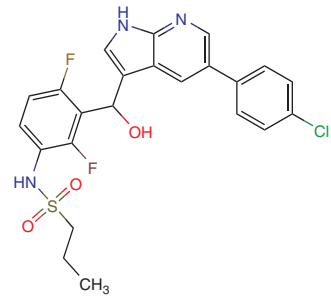

Figure 1.7. The molecular structure of Vemurafenib (V600E mutated BRAF inhibition), a B-Raf inhibitor for late-stage melanoma patients with B-Raf V600E mutation

1.3.2 Metabolic Profiling, Metabonomics, and Pharmacometabonomics

Metabolic profiling is another new area that is increasing rapidly in importance for drug discovery and development. NMR spectroscopy has been an important tool for the study of metabolite profiles in biological fluids since the early 1980s.[21] The applications of the technology are manifold and include the direct study of drug metabolism for relatively high-dosage drugs,[22] the study of drug safety, and the study of disease states,[23] all of prime importance for drug discovery and development.

The increasing importance of metabolic profiling of biofluids was recognized in the late 1990s by the introduction of the term metabonomics to describe this activity[24] and to provide a conceptual framework for combined genomics, proteomics, and metabolic profiling studies that were underway at that time in Pfizer Global R&D. Metabonomics was defined as 'the study of the metabolic response of organisms to disease, environmental change, or genetic modification'. Some confusion has crept into the field with the later definition of the term metabolomics as a 'comprehensive analysis in which all the metabolites of a biological system are identified and quantified', which is quite a tall order, if not impossible.[25]

The increase in importance of metabonomics studies has arisen for several reasons, some of which are technological. Firstly, the development of efficient multivariate statistical methods that enable (i) the rapid analysis of large numbers of complex NMR spectra of biofluids (mainly) from different groups of subjects, (ii) the unbiased determination of differences between those groups based on differences in their spectral properties, and, most importantly, (iii) the unbiased derivation of the spectroscopic variables that drive those differences (see Chapter 7). Secondly, the availability of ever more reliable spectrometers that can perform demanding NMR experiments with excellent water suppression in full automation, often in conjunction with cooled sample changers that can keep samples stable for long periods before analysis, and automated sample preparation, enabling high throughput of samples

for large studies. Thirdly, the development of powerful databases of metabolic data such as the Human Metabolite Database (HMDB, http://www.hmdb.ca), the Biological Magnetic Resonance Data Bank (BMRB, http://www.bmrb.wisc.edu/metabolomics/), and the Birmingham Metabolite Library (BML, http://www.bml-nmr.org) that facilitate the identification of metabolites that occur in the NMR spectra of biological fluids using information from 1D and 2D proton and carbon-13 NMR spectra. In addition, dedicated repositories for the deposition of metabonomics data are emerging such as MetaboLights (http://www.ebi.ac.uk/metabolights/), which will facilitate access to raw data throughout the metabolic profiling community and help set standards for data deposition.

There are also scientific reasons for the increased importance of NMR-based metabolic profiling. Firstly, NMR spectroscopy is an unbiased detector that can provide fully quantitative data on all of the components in a complex mixture that are above the detection threshold (given certain caveats). Secondly, metabolite profiling gives information directly on the actual physiological status of the organism under study, in contrast to, say, genomics, which can only provide information on what might happen to an organism in the future: this is a really critical distinction. Thirdly, metabolic profiling of biological fluids such as urine provides

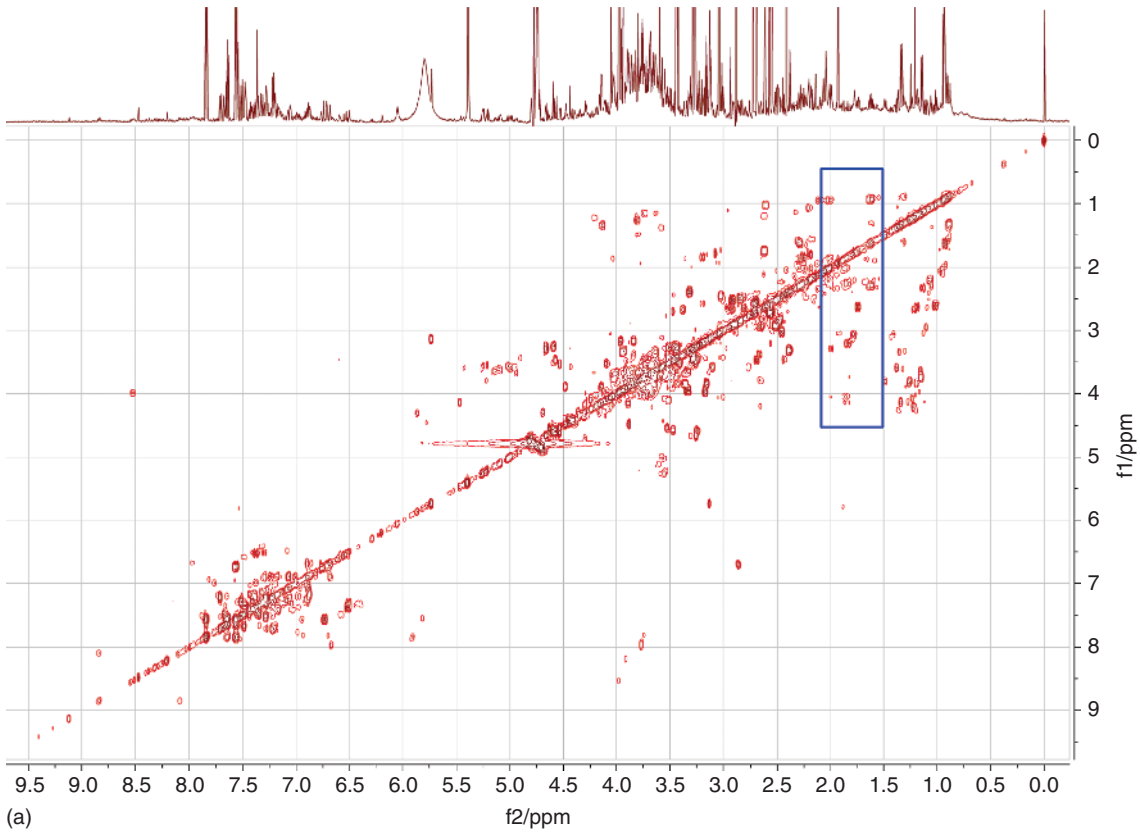

Figure 1.8. (a) The full 600 MHz 2D ^{1}H COSY NMR spectrum of the urine of a C57BL/6 mouse; (b) an expansion of the area of the COSY spectrum shown in the blue box in (a); and (c) the same expansion of the corresponding 600 MHz 2D ^{1}H J-resolved (JRES) spectrum of the same mouse urine, underneath the corresponding 1D ^{1}H NMR spectrum. The six peaks of the multiplet due to *N*-butyrylglycine across f1 are marked with blue arrows. This multiplet is a triplet of quartets but only six of the twelve lines are visible as the two coupling constants are very similar. All spectra are shown as contour plots and are referenced to TSP at 0 [Everett, 2014, unpublished, from work in collaboration with Dorsa Varshavi (University of Greenwich) and with Liz Shephard and Flora Scott, (University College London)]

information not only on the status of the organism and the expression of its genome but also on the activity of the various microbiomes in that organism. Again, this is a critical distinction and a difference between human genomics and human metabonomics studies.

The huge amount of high-quality data on metabolites that can be obtained from NMR spectroscopy studies of biological fluids at only modest NMR field strengths is illustrated in Figure 1.8, which shows representative 2D ^{1}H COSY and 2D ^{1}H J-resolved NMR spectra from a C57BL/6 mouse. In Figure 1.8(a), the full 600 MHz 2D ^{1}H COSY NMR spectrum is plotted underneath the corresponding 1D ^{1}H NMR spectrum.

A staggering array of proton-to-proton connectivity information is present in the 2D ^{1}H COSY NMR spectrum (Figure 1.8a), which becomes clearer on expansion (Figure 1.8b). Note that the expansion still contains a huge amount of data and information, although it represents less than about 3% of the entire spectrum! For instance, only four of the six lines of the triplet of quartets multiplet signal for CH_2-3 of *N*-butyrylglycine (Figure 1.9) are visible in the 1D ^{1}H NMR spectrum at about 1.62 ppm (Figure 1.8b, top). All six peaks are however visible across the second dimension f1 of the 2D ^{1}H J-resolved (JRES) spectrum (Figure 1.8c), and the expanded COSY spectrum clearly shows well-resolved cross-peaks to both the CH_3-4 protons at about 0.925 ppm and the CH_2-2 protons at about 2.28 ppm (blue boxes in Figure 1.8b). More remarkably, the JRES spectrum shows two multiplets with at least 12 peaks each at about 1.845 and about 2.005 ppm, belonging to the CH_2-3 methylene protons of L-2-hydroxyglutaric acid. These peaks are almost invisible in the 1D ^{1}H NMR spectrum owing to the complexity of signals in this region, but are clearly defined across the second dimension, f1, of the JRES spectrum, and their mutual coupling connectivities to the CH-2 methyne proton at about 4.03 can clearly be seen in the COSY spectrum (green boxes).

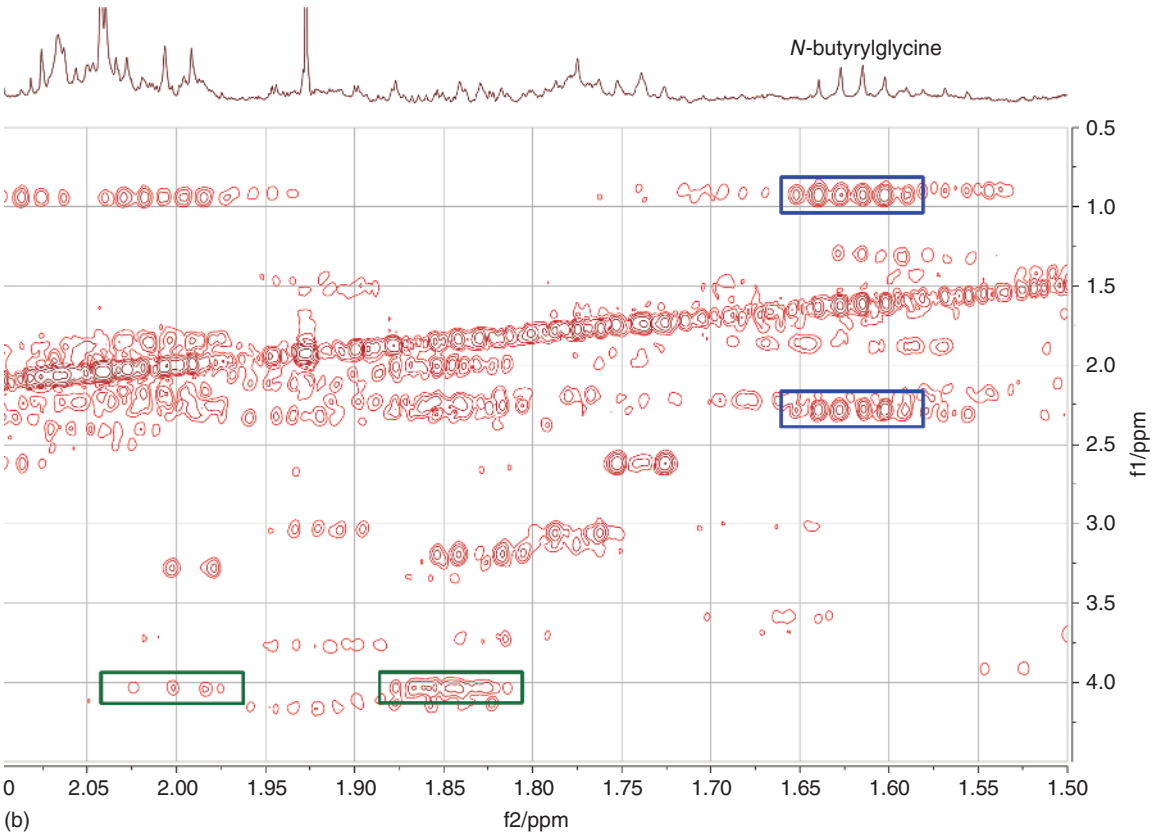

Figure 1.8. (*Continued*)

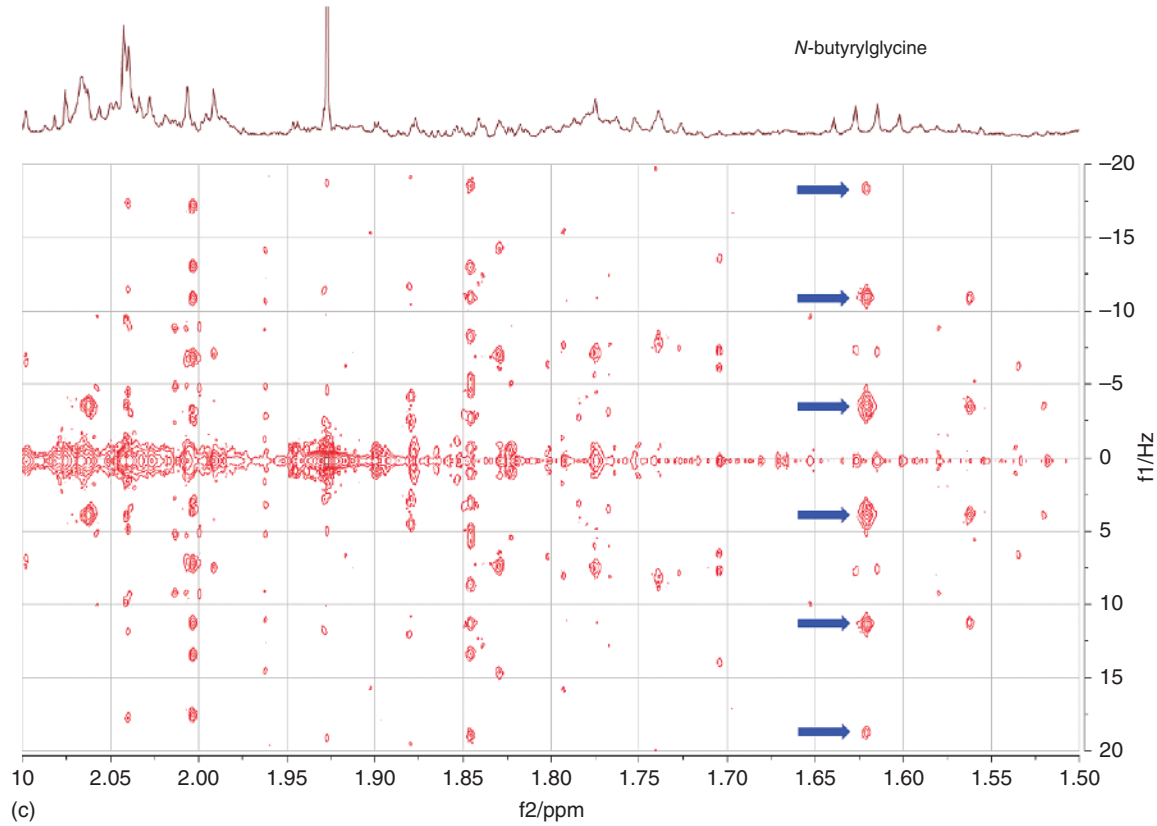

Figure 1.8. (*Continued*)

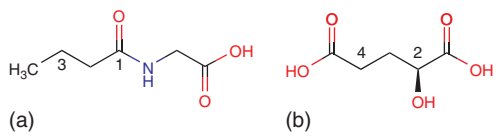

Figure 1.9. The molecular structures of *N*-butyrylglycine (a) and L-2-hydroxyglutaric acid (b)

A major new area of importance in medicine is that of personalized or precision medicine. The concept here is to try to stratify patients so that they receive treatments that will be both efficacious and safe for them. Unfortunately, oftentimes, drugs are either ineffective in certain patient groups or unsafe and this causes significant morbidity and mortality in the Western world. Efforts have been ongoing for decades to use genomic data to stratify patients and to predict the effects that drugs will have on the patients, or that the patients will have on the drug: this is known as *pharmacogenomics*.

Two landmark studies published in 2006 and 2009 demonstrated that *predose* metabolic profiles determined by NMR spectroscopy could be used to predict the outcome of drug treatment in animals in terms of drug metabolism and drug safety,[26] and also that *predose* metabolic profiles could also be used to predict drug metabolism in humans.[27] Moreover, in the latter study, the key biomarker indicating whether the particular drug acetaminophen would be sulfated or glucuronidated in humans was not of human origin, but came from the microbiome! This new area of application of metabolic profiling or metabonomics is called *pharmacometabonomics*[26] by analogy to pharmacogenomics. Many examples of this phenomenon have now been published.

See Chapters 10, 19, 20, 24, and 25, respectively, all of which have extensive additional material.

1.4 *IN VIVO* MRS AND MRI

MRI is now a very well-established technology for imaging in both clinical and preclinical settings, and it can also provide functional information, especially in the brain, through diffusion and flow measurements.[28,29] In one recent development of this area, MRI diffusion experiments are being used to define the human connectome (Figure 1.10), which is the map of neural connections in the brain. This will surely be of use in the development of future drugs for the treatment of central nervous system diseases especially those connected with neurodegeneration.

The term *MRS* is generally used to describe *in vivo* NMR spectroscopy experiments on animals and humans. MRS and MRI are two among a number of technologies being used to provide structural and functional information on various parts of human and animal bodies, particularly the brain, to aid drug discovery and development. MRS has the advantage of being able to measure the levels of a dozen or more high-abundance metabolites in localized tissue regions, such as a particular portion of the brain, in a completely noninvasive and safe manner (no ionizing radiation involved), but suffers from low sensitivity and long acquisition times compared with other technologies. The most sensitive brain imaging technology is positron emission tomography (PET), which detects twin gamma rays emitted from annihilation events associated with positron-emitting isotopes such as carbon-11, fluorine-18, oxygen-15, and nitrogen-13. The PET imaging center needs to be colocated with, or very close to, the particle accelerator producing these isotopes and specialized chemistry facilities are needed to produce isotopically labeled forms of the drug of interest for PET studies. PET studies can give detailed information on the pharmacokinetics (PK) and distribution of labeled drug molecules with picomolar sensitivity. Unfortunately, while very sensitive (up to 10^8 times more sensitive than MRS), PET gives no information on the chemical nature of the molecules giving rise to the signals and the images may represent mixtures of signals from the drug plus any metabolites and degradation products. Single-photon emission computed tomography (SPECT) is an alternative imaging technology that utilizes the emission of single gamma rays from radioactive nuclei such as xenon-133, technetium-99,

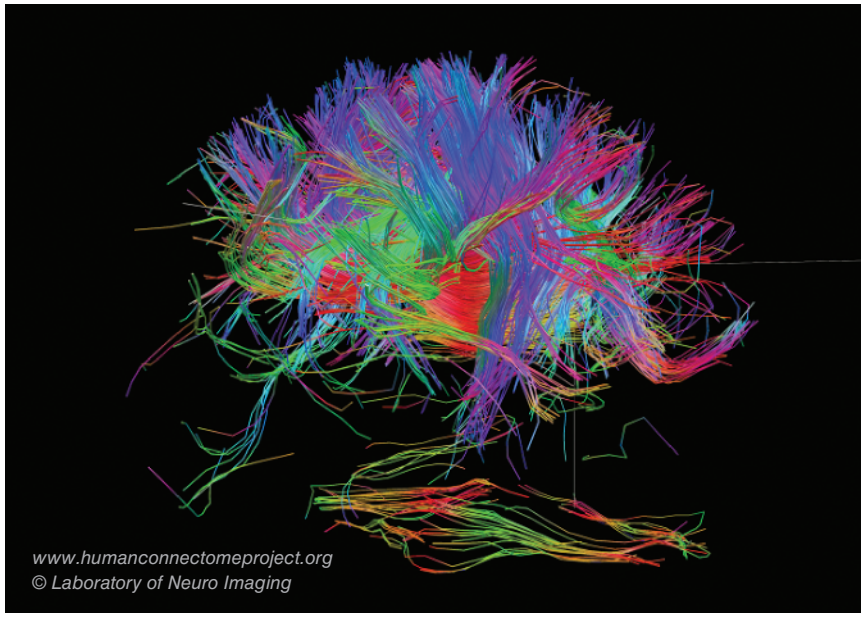

Figure 1.10. White matter fiber architecture of the human brain from diffusion spectral imaging. The brain fibers are color-coded by direction: red = left to right, green = anterior to posterior, and blue = through the brain stem. (Image courtesy of the Laboratory of Neuro Imaging and Martinos Center for Biomedical Imaging, Consortium of the Human Connectome Project: www.humanconnectome.org)

or iodine-123 to give images that can be used to determine the PK and distribution of labeled drugs and ligands. While there is no requirement to be close to a synchrotron facility, the images from SPECT are of lower sensitivity and lower resolution than PET.

Although relatively insensitive, MRS is a field that has been developing since the 1980s and is now well established in preclinical research settings. MRS can provide information complementary to the anatomical information obtained from MRI in animal models of disease, for instance, on changes associated with disease progression or the effects of drug treatment. For example, in mouse models of Alzheimer's disease (AD), changes in the brain levels of *N*-acetylaspartate (NAA), and of *myo*-inositol, can be observed with good sensitivity and precision that enable the effects of treatment to be observed, and this also holds out the promise of translation to patients in the future.[30]

Challenges remain for the full utilization of MRS in clinical settings, however,[31] and some experts believe that this is more likely to occur when performed simultaneously with both PET and MRI. While it is possible to perform joint MRI and MRS studies, currently there are no commercial instruments available enabling the performance of PET and MRI/MRS in a single system.[32] Further technological developments in MRS that increase sensitivity, perhaps including DNP methods, will doubtlessly drive greater utility of MRS in both preclinical and clinical settings. It has to be taken as a given that any technology that can provide rapid, noninvasive, cost-effective, and safe chemical imaging of the human body will find increased usage in the future in drug discovery and development.

A similar situation applies for functional magnetic resonance imaging (fMRI), where in spite of the technology providing significant impact on the understanding of brain functioning and brain system phenotypes in many neurological and psychiatric diseases, there has been minimal benefit so far to individual patients.[28] Again, it is clear that this situation will change in the future with advances in technologies and their application to better diagnose disease states, recommend treatments, and predict outcomes. See Chapters 18 and 26 for further insights and expert review.

1.5 FUTURE DEVELOPMENTS

The technological progress of NMR spectroscopy and MRI over the past 30 years has been astonishing and whole areas of application exist now that were not even conceptualized 20 years ago. The future is notoriously difficult to predict but several areas seem ripe for significant change. Firstly, the identification of the rich array of metabolites present in metabonomics studies is still problematic and requires a significant amount of user knowledge and expertise. It is expected that this area will become increasingly amenable to computer-based expert systems over the next decade. Secondly, the emerging area of pharmacometabonomics is likely to become of much greater future importance for the prediction of drug effects and the delivery of personalized medicine, especially because of its ability to predict drug effects based on environmental (including microbiome) information, as well as genomic-derived information. Thirdly, a new area of predictive metabonomics, of which pharmacometabonomics is just a subset, will emerge, enabling prediction of patient outcomes over time, and as a result of varied interventions, not just drug effects: this has already begun. Finally, improvements in computer power will enable better prediction of NMR spectra based on quantum mechanical methods, giving better confidence in NMR-based structure elucidation by comparisons of actual spectra with precisely calculated spectra of the elucidated molecular structure and viable alternatives. The future is always exciting!

REFERENCES

1. J. L. LaMattina, *Nat. Rev. Drug Discovery*, 2011, **10**, 559.
2. S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, and A. L. Schacht, *Nat. Rev. Drug Discovery*, 2010, **9**, 203.
3. R. M. Fratila and A. H. Velders, *Annu. Rev. Anal. Chem.*, 2011, **4**, 227.
4. D. S. Wishart, *TrAC, Trends Anal. Chem.*, 2013, **48**, 96.
5. Mestrelab Research, The home page for MNova NMR software, Date accessed 5th November 2014, http://mestrelab.com/software/mnova/nmr/
6. G. Baker, R. Dorgan, J. R. Everett, J. Hood, and M. Poulton, *J. Antibiot.*, 1990, **43**, 1069.
7. D. A. Dias, S. Urban, and U. Roessner, *Metabolites*, 2012, **2**, 303.
8. M. Halabalaki, K. Vougogiannopoulou, E. Mikros, and A. L. Skaltsounis, *Curr. Opin. Biotechnol.*, 2014, **25**, 1.

9. K. Ashline, R. Attrill, E. Chess, J. Clayton, E. Cutmore, J. R. Everett, J. Nayler, D. Pereira, W. Smith, J. Tyler, M. Vieira, and M. Sabat, *J. Chem. Soc., Perkin Trans. 2*, 1990, 1559.
10. J. R. Everett, E. Hunt, and J. Tyler, *J. Chem. Soc., Perkin Trans. 2*, 1991, 1481.
11. H. Booth and J. R. Everett, *J. Chem. Soc., Chem. Commun.*, 1976, 278.
12. H. Finch, *Drug Discovery Today*, 2014, **19**, 320.
13. C. Simmler, J. G. Napolitano, J. B. McAlpine, S.-N. Chen, and G. F. Pauli, *Curr. Opin. Biotechnol.*, 2014, **25**, 51.
14. K. Paradowska and I. Wawer, *J. Pharm. Biomed. Anal.*, 2014, **93**, 27.
15. G. A. Bakken, A. S. Bell, M. Boehm, J. R. Everett, R. Gonzales, D. Hepworth, J. L. Klug-McLeod, J. Lanfear, J. Loesel, J. Mathias, and T. P. Wood, *J. Chem. Inf. Model.*, 2012, **52**, 2937.
16. A. S. Bell, J. Bradley, J. R. Everett, M. Knight, J. Loesel, J. Mathias, D. McLoughlin, J. Mills, R. E. Sharp, C. Williams, and T. P. Wood, *Mol. Diversity*, 2013, **17**, 319.
17. O. Cala, F. Guilliere, and I. Krimm, *Anal. Bioanal. Chem.*, 2014, **406**, 943.
18. M. J. Harner, A. O. Frank, and S. W. Fesik, *J. Biomol. NMR*, 2013, **56**, 65.
19. L. Unione, S. Galante, D. Diaz, F. Javier Canada, and J. Jimenez-Barbero, *MedChemComm*, 2014, **5**, 1280.
20. N. Yanamala, A. Dutta, B. Beck, B.van Fleet, K. Hay, A. Yazbak, R. Ishima, A. Doemling, and J. Klein-Seetharaman, *Chem. Biol. Drug Des.*, 2010, **75**, 237.
21. J. Lindon, J. Nicholson, and J. Everett, *Annu. Rep. NMR Spectrosc.*, 1999, **38**, 1.
22. J. R. Everett, K. Jennings, G. Woodnutt, and M. Buckingham, *J. Chem. Soc., Chem. Commun.*, 1984, 894.
23. A.-H. M. Emwas, R. M. Salek, J. L. Griffin, and J. Merzaban, *Metabolomics*, 2013, **9**, 1048.
24. J. Lindon, J. Nicholson, E. Holmes, and J. R. Everett, *Concepts Magn. Reson.*, 2000, **12**, 289.
25. O. Fiehn, *Plant Mol. Biol.*, 2002, **48**, 155.
26. T. Clayton, J. Lindon, O. Cloarec, H. Antti, C. Charuel, G. Hanton, J. Provost, J. Le Net, D. Baker, R. Walley, J. R. Everett, and J. Nicholson, *Nature*, 2006, **440**, 1073.
27. T. A. Clayton, D. Baker, J. C. Lindon, J. R. Everett, and J. K. Nicholson, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 14728.
28. E. T. Bullmore, *Neuroimage*, 2012, **62**, 1267.
29. D. Borsook, R. J. Hargreaves, and L. Becerra, *Expert Opin. Drug Discovery*, 2011, **6**, 597.
30. N. Beckmann, R. Kneuer, H.-U. Gremlich, H. Karmouty-Quintana, F.-X. Ble, and M. Mueller, *NMR Biomed.*, 2007, **20**, 154.
31. B. D. Ross, *Expert Opin. Drug Discovery*, 2013, **8**, 849.
32. W. Wolf, *Pharm. Res.*, 2011, **28**, 490.