ChemComm

Cite this: *Chem. Commun.*, 2011, **47**, 12476–12478

www.rsc.org/chemcomm

COMMUNICATION

^{13}C-isotope labelling for the facilitated NMR analysis of a complex dynamic chemical system†

Marta Dal Molin, Giulio Gasparini, Paolo Scrimin, Federico Rastrelli* and Leonard J. Prins*

Received 25th August 2011, Accepted 28th September 2011
DOI: 10.1039/c1cc15295e

^{13}C-isotope labelling is presented as a novel tool for the study of complex chemical systems. ^{13}C-isotope labelling permits the quantification of all 26 members of a dynamic library from a single ^{13}C NMR spectrum without the need for advanced instrumentation or sophisticated experimental protocols.

Inspired by the astonishing achievements of Nature, chemists are now involved in the construction of functional synthetic systems able to match up to the complexity of natural systems.[1] Not only this entails the self-assembly of supramolecular systems,[2] but it also involves the study of properties emerging from artificial networks.[3] Evidently, the achievable complexity of supramolecular constructs and networks is strongly dependent on the availability of analytical tools able to inform on the structure of the supramolecule or the concentration of each network component.

NMR spectroscopy is an attractive analytical tool for studying complex chemical systems. It is a non-invasive technique thus permitting the direct analysis of the complete mixture under equilibration conditions. In addition, the signals are very sensitive to small structural changes.[4] However, in the case of proton NMR spectroscopy this is counterbalanced by the fact that the chemically-shifted Larmor frequencies of ^{1}H vary within a limited range of values (typically 10 ppm), which most often results in serious signal overlap at average magnetic fields (*i.e.* 400 MHz). Previously, we have shown that the use of ^{1}H–^{13}C HSQC spectroscopy can alleviate this problem, but high concentrations and advanced experimental protocols were required to study a chemical system of moderate complexity (8 components).[5] The limited use of NMR spectroscopy as an analytical tool within the context of systems chemistry is quite remarkable, considering that NMR is widely accepted as *the* tool for studying complex biomolecular structures.[6] A common route to sensitivity enhancement in protein NMR is the insertion of isotopic labels (notably ^{15}N and ^{13}C) at key positions in the molecular structure.[7] Isotope labelling in small organic molecules is far less popular, because their much lower complexity simply does not require the insertion of isotope labels and sensitivity issues arising from low compound concentrations are typically absent. However, as chemistry is heading towards the study of increasingly complex chemical systems, it becomes evident that sophisticated tools, such as targeted isotope labelling, for studying complex structures will turn out to be essential for NMR screening. In particular systems relying on an imine-type exchange reaction, which is one of the most frequently applied dynamic covalent bonds,[8,9] are highly suitable for ^{13}C isotope labelling of the iminic carbon for the following reasons. Firstly, the iminic position is intrinsically advantageous because the chemical shift of the iminic carbon is very sensitive to small structural variations. Secondly, the widespread chemically-shifted Larmor frequencies of ^{13}C (up to 200 ppm and more) combined with (proton) broadband decoupling will effectively minimise the occurrence of signal overlap, even at moderate magnetic fields. Thirdly, being an endogenous label, the ^{13}C-isotope does not affect the thermodynamic stability or the structure of the (supra)molecule under investigation, something that cannot be excluded when heteroatoms (such as ^{19}F or ^{31}P) are inserted as exogenous NMR labels. Finally, the iminic carbon atom can be labelled by exploiting a reasonably simple chemistry.

Here, we show that ^{13}C-labelling permits the straightforward identification and quantification of each member of a chemical system containing up to 26 components. This study was aimed at determining how neighbouring spectator groups affect the thermodynamic stability of a series of hydrazones and imines ($\mathbf{P_{1-4}X}$) in which an intramolecular interaction exists between a phosphonate and an ammonium group (Fig. 1).[10] In prior studies we have shown that this information is of relevance for the development of a model that mechanistically mimics a serine protease.[11,12] The key issue is the self-selection of functional groups by the phosphonate target, which acts as a transition state analogue for the hydrolysis of a carboxylic ester. In enzymes remote spectator groups may appear inert on first sight. Yet, by subtly altering the conformation of the enzyme they may actually play an important role in determining the molecular recognition event occurring in the active site.[13] We were challenged by the idea of quantifying the contribution of different spectator groups to the thermodynamic stability of a single compound from a direct analysis of the dynamic system and with a minimal amount of experiments. So far, the application of dynamic combinatorial chemistry for the

Department of Chemical Sciences, University of Padova, Via Marzolo 1, 35131 Padova, Italy.
E-mail: federico.rastrelli@unipd.it, leonard.prins@unipd.it
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c1cc15295e

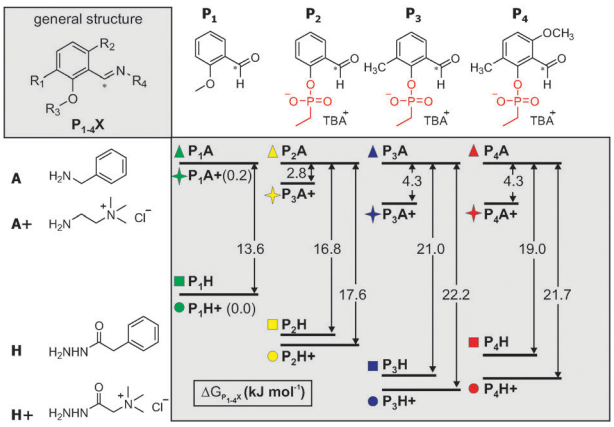

Fig. 1 Relative thermodynamic stabilities of the ^{13}C-isotopically labelled imines $\mathbf{P_{1–4}(A/A+)}$ and hydrazones $\mathbf{P_{1–4}(H/H+)}$ as determined from ^{13}C NMR competition experiments (the asterisk indicates a ^{13}C nuclide). The symbols correspond to the indicated signals in Fig. 2 and 3. The relative thermodynamic stabilities of compounds $\mathbf{P_{1–4}X}$ are arbitrarily normalized on the energy level of the imines $\mathbf{P_{1–4}A}$, which is the thermodynamically least stable species in each series. The differences in free energy (ΔG; kJ mol^{-1}) are calculated from experiments performed in CD_3OD at 298 K.

determination of quantitative structure–activity relationships (QSAR) has received very little attention.[14–16]

The library we have chosen is particularly challenging for two reasons. First, it includes two sub-libraries (imines and hydrazones), which are expected to have very different intrinsic stabilities.[17] Second, the hydrazones are present as a mixture of geometric isomers, which cannot be resolved by chromatographic means. The direct analysis of such a complex system under equilibration conditions would be impossible with any technique other than NMR. Aromatic aldehydes $\mathbf{P_1–P_4}$ were prepared with various degrees of substitution of the aromatic ring. Within this set, compound $\mathbf{P_1}$, lacking the phosphonate group, serves as an internal reference to assess the intrinsic stabilities of the formed imines and hydrazones. All compounds $\mathbf{P_1–P_4}$ contain a ^{13}C labelled aldehyde group, which was easily introduced in the *ortho*-position to a MOM-protected phenol *via* quenching of the BuLi-activated aromatic ring with a slight excess of ^{13}C-DMF (see ESI†). Hydrazides **H** and **H**+ and amines **A** and **A**+ were chosen as complementary components for hydrazone and imine formation, respectively, because they allowed an assessment of the importance of the dynamic covalent bond in terms of rigidity and spacer length. ^{13}C-labelling of the iminic carbon provides a single characteristic resonance for each compound in the ^{13}C NMR spectrum which can now be recorded with sensitivity close to that of a ^{1}H NMR spectrum. The characteristic resonances for the iminic carbon fall within the 140–170 ppm range compared to the 8–9.5 ppm range for the iminic proton. This much larger interval permits the study of mixtures of far higher complexity. This is evidenced by the observation that all 26 components (including isomers) of the library obtained by mixing all scaffolds $\mathbf{P_1–P_4}$, both hydrazides **H** and **H**+ *and* both amines **A** and **A**+ gave well-resolved signals in the ^{13}C NMR spectrum (Fig. 2a). The overlapping signals in the ^{1}H NMR spectrum indicate that the complexity of this system is well above the resolution limit for ^{1}H NMR spectroscopy.

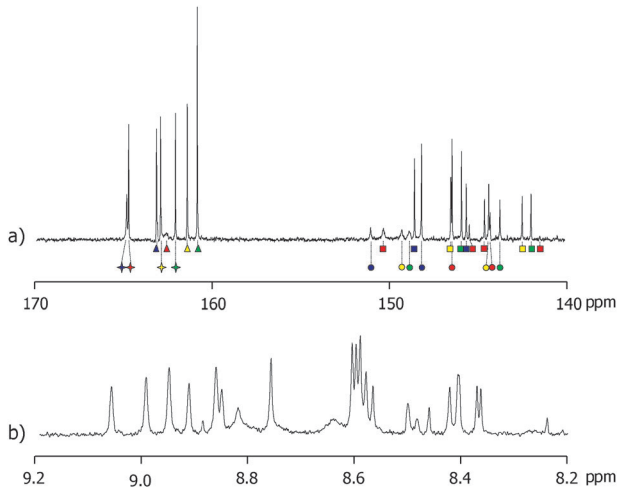

Fig. 2 (a) ^{13}C NMR spectrum (150 MHz, ^{1}H decoupled DEPT-90) and (b) ^{1}H NMR spectrum (600 MHz, ^{13}C decoupled) of the mixture of scaffolds $\mathbf{P_1–P_4}$, both hydrazides **H** and **H**+ and both amines **A** and **A**+ at thermodynamic equilibrium (CD_3OD, 298 K). The symbols refer to the structures in Fig. 1.

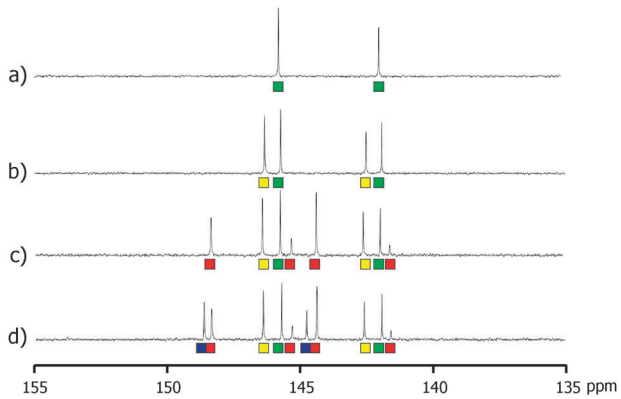

Fig. 3 Signal assignment based on the sequential addition of scaffolds $\mathbf{P_1–P_4}$ (2.5 mM each) to a solution of hydrazide **H** (25 mM) in CD_3OD at 298 K (600 MHz). Scaffolds were added in the following order: (a) $\mathbf{P_1}$ (▪); (b) $\mathbf{P_2}$ (▪); (c) $\mathbf{P_4}$ (▪); (d) $\mathbf{P_3}$ (▪).

In the ^{13}C spectrum 18 species out of 26 were quantified in just 20 minutes, whereas a thorough quantification of all signals was possible in 12 hours. Species present at concentrations down to 0.1 mM could be detected. Signals were assigned to each library member in a straightforward manner by gradually increasing the complexity of the mixture. Thus, scaffolds $\mathbf{P_1–P_4}$ were added sequentially to a solution containing an excess of either one of the complementary components **A**, **A**+, **H**, or **H**+. This is illustrated in Fig. 3a–d for the addition of scaffolds $\mathbf{P_1–P_4}$ (2.5 mM each) to hydrazide **H** (25 mM). Competition experiments were then started by mixing these sub-libraries, followed by ^{13}C NMR spectroscopy. In order to gain the highest possible sensitivity, we exploited polarization transfer by the use of the DEPT pulse sequence[18] (see ESI† for details on the advantages of the DEPT sequence in our case).

The full system analysis permitted a quantification of all 26 species present, but the obtained data were difficult to interpret in terms of relative thermodynamic stability. These problems

do not originate from the analytical technique itself, but rather from the fact that near stoichiometric amounts of hydrazides were required to allow the simultaneous coexistence of the thermodynamically much less stable imines.[17] The absence of a buffering amount of free hydrazide introduces a cross-talk between the various equilibria in the sense that the preferred selection of **H**+ over **H** by one scaffold will also affect the hydrazone ratio for the other scaffolds. Such a cross-talk is detrimental for a proper quantification of the relative thermodynamic stability of each species. For that reason, the analysis was repeated on the separate hydrazone and imine libraries, containing 18 and 8 components, respectively. In addition, individual competition experiments between $\mathbf{P_{1-4}A}$ and $\mathbf{P_{1-4}H}$ were used to correlate the energy scales of the imines and hydrazones (see ESI†). From these studies it emerges in a straightforward manner how substituents on the scaffolds $\mathbf{P_1}$–$\mathbf{P_4}$ affect the thermodynamic stability of the corresponding hydrazones and imines (Fig. 1). The interpretation of these results requires some considerations. First, in contrast to our previous studies using $\mathbf{P_1}$ and $\mathbf{P_2}$,[5,11,19] the competition experiments involving different scaffolds now occur simultaneously in the same NMR tube in a parallel set of equilibria. On one hand, this guarantees identical screening conditions but, on the other hand, it may complicate the analysis in the case of crosstalk between the platforms (see above). This appears not to be the case, since the measured differences in thermodynamic stabilities between hydrazones $\mathbf{P_1H}$–$\mathbf{P_1H}$+ (0.0 kJ mol^{-1}) and $\mathbf{P_2H}$–$\mathbf{P_2H}$+ (0.8 kJ mol^{-1}) are nearly identical to those obtained previously from separate experiments (0.2 kJ mol^{-1} and 0.7 kJ mol^{-1}).[5,11] Second, it should be noted that these experiments do not reveal information on the thermodynamic stabilities *between* the different scaffolds.‡ In fact, all energy levels have been arbitrarily normalized on imines $\mathbf{P_{1-4}A}$, which are taken as a reference within each scaffold series. Third, in this preliminary analysis we have not taken hydrazone isomerism into account and have treated each component $\mathbf{P_{1-4}X}$ as a single species.§

This analysis provides a quantitative insight into the correlation between structure and thermodynamic stability. A first glance immediately reveals some key features. First, the anticipated stabilizing interactions between phosphonate and ammonium groups in both $\mathbf{P_2A}$+ (2.6 kJ mol^{-1}) and $\mathbf{P_2H}$+ (0.8 kJ mol^{-1}), which is further enhanced upon the insertion of a methyl group in the *ortho*-position to the phosphonate group (scaffold $\mathbf{P_3}$). Second, the anticipated much higher intrinsic stability of hydrazones compared to imines. Nonetheless, after a closer look it emerges that the overall picture is actually rather complicated. For example, substituents affect the intrinsic stability of imines and hydrazones ($\Delta G_{\mathbf{P_2H}-\mathbf{P_2A}}$ = 16.8 kJ mol^{-1} against 13.6 kJ mol^{-1} for $\Delta G_{\mathbf{P_1H}-\mathbf{P_1A}}$) and the presence of the methoxy substituent in $\mathbf{P_4}$ has a destabilizing effect in particular on the hydrazones $\mathbf{P_4H}$ and $\mathbf{P_4H}$+.§ It is evident that such an interplay between stabilizing and destabilizing energetic contributions always plays a role in determining the composition of a dynamic system. However, such information can only emerge if the thermodynamic stability of all system components can be determined.

In summary, these results show that ^{13}C isotopic labelling of a dynamic library is a very attractive tool to study complex dynamic systems by NMR spectroscopy. Compared to a previous NMR study on a similar non-labelled system, ^{13}C-labelling permits an increase in complexity from 8 to 26 signals without the need for bidimensional ^{1}H–^{13}C HSQC spectroscopy, specialized NMR cryo-probes or complex data-processing protocols.[5] Most importantly, much lower compound concentrations were required (0.1 mM compared to 10 mM) and a total of 18 out of 26 signals could be quantified in just 20 minutes, implying that the kinetics of these systems can be studied with enhanced time resolution, too. It is anticipated that the combination of advanced NMR protocols with ^{13}C-labelling will enable the analysis of mixtures of even higher complexity than reported here. As we have shown, the ability to quantify the concentrations of all components of the system provides a unique possibility to determine quantitative structure–activity relationships in a straightforward manner.

Financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC Starting Grant agreement no. 239898 and COST (CM0703) is acknowledged.

Notes and references

‡ In the presence of an excess of amines and hydrazides there is no competition between scaffolds for imine- or hydrazone bond formation.

§ A full account including a treatment of hydrazone isomerism and all energetic contributions involved will be reported in due course. Preliminary results indicate that steric hindrance in scaffold $\mathbf{P_4}$ induces the formation of two non-planar isomers in $\mathbf{P_4H}$ in addition to the regular two isomers originating from *E*/*Z* isomerism of the amide bond, which are common to all hydrazones.

1 G. M. Whitesides and R. F. Ismagilov, *Science*, 1999, **284**, 89–92.
2 J. M. Lehn, *Angew. Chem., Int. Ed. Engl.*, 1990, **29**, 1304–1319.
3 R. F. Ludlow and S. Otto, *Chem. Soc. Rev.*, 2008, **37**, 101–108.
4 W. C. Pomerantz, E. B. Hadley, C. G. Fry and S. H. Gellman, *ChemBioChem*, 2009, **10**, 2177–2181.
5 G. Gasparini, B. Vitorge, P. Scrimin, D. Jeannerat and L. J. Prins, *Chem. Commun.*, 2008, 3034–3036.
6 J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance and N. J. Skelton, *Protein NMR spectroscopy: principles and practice*, 2nd edn, Academic Press, 2007.
7 L. P. McIntosh and F. W. Dahlquist, *Q. Rev. Biophys.*, 1990, **23**, 1–38.
8 S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders and J. F. Stoddart, *Angew. Chem., Int. Ed.*, 2002, **41**, 898–952.
9 C. D. Meyer, C. S. Joiner and J. F. Stoddart, *Chem. Soc. Rev.*, 2007, **36**, 1705–1723.
10 L. J. Prins and P. Scrimin, *Angew. Chem., Int. Ed.*, 2009, **48**, 2288–2306.
11 G. Gasparini, L. J. Prins and P. Scrimin, *Angew. Chem., Int. Ed.*, 2008, **47**, 2475–2479.
12 G. Gasparini, M. Dal Molin and L. J. Prins, *Eur. J. Org. Chem.*, 2010, 2429–2440.
13 A. Fersht, *Structure and Mechanism in Protein Structure*, W.H. Freeman and Company, 1999.
14 M. G. Woll and S. H. Gellman, *J. Am. Chem. Soc.*, 2004, **126**, 11172–11174.
15 M. K. Chung, K. Severin, S. J. Lee, M. L. Waters and M. R. Gagne, *Chem. Sci.*, 2011, **2**, 744–747.
16 R. F. Ludlow, J. Liu, H. X. Li, S. L. Roberts, J. K. M. Sanders and S. Otto, *Angew. Chem., Int. Ed.*, 2007, **46**, 5762–5764.
17 F. A. Carey and R. J. Sundberg, *Advanced Organic Chemistry*, vol. A, Springer, 2008.
18 D. M. Doddrell, D. T. Pegg and M. R. Bendall, *J. Magn. Reson.*, 1982, **48**, 323–327.
19 G. Gasparini, F. Bettin, P. Scrimin and L. J. Prins, *Angew. Chem., Int. Ed.*, 2009, **48**, 4546–4550.

