Complexation-Induced Changes in ^{1}H NMR Chemical Shift for Supramolecular Structure Determination

Christopher A. Hunter* and Martin J. Packer[a]

Abstract: Complexation-induced changes in ^{1}H NMR chemical shift have been used to determine high-resolution three-dimensional structures of supramolecular complexes. The approach has been validated for a system in which the X-ray crystal structure of the complex can be compared with the optimised NMR structure, and the agreement is remarkably good (rmsd = 0.37 Å). Determination of the solution structure of an H-bonded zipper complex is used to demonstrate the power of the approach in systems for which alternative tools for structure determination are not useful.

Keywords: computer chemistry · genetic algorithm · NMR spectroscopy · ring current · structure elucidation · supramolecular chemistry

Introduction

Structure determination remains a major problem in supramolecular chemistry. If complexes are sufficiently stable, then it is often possible to obtain high-resolution structural information by means of X-ray crystallography in the solid state. However, it is extremely difficult to crystallise more weakly bound complexes, and for these systems the solid-state structure may not reflect the situation in solution.[1] The solution-phase method that has proved most useful for obtaining structural information on supramolecular complexes is ^{1}H NMR. Intermolecular NOEs provide information about protons that are close in space, and complexation-induced changes in chemical shift (CIS) provide information about functional-group interactions, for example, interactions with aromatic rings or hydrogen-bonding sites. With a few notable exceptions,[2–6] NOE and CIS data are generally used in a qualitative manner to produce low-resolution pictures of the structures of intermolecular complexes. Some attempts have been made to quantitate NOE data to provide more accurate structural information, but this approach is limited by the number of NOEs that can be observed in low molecular weight often symmetrical supramolecular systems. In this paper, we discuss a quantitative approach to the use of CIS data in supramolecular structure determination. In the field of protein NMR structure determination, quantitative methods have been developed to use folding-induced changes in ^{1}H NMR chemical shift in the structure optimisation process,[7–12] and we have found that these methods show considerable promise for obtaining high-resolution three-dimensional structures of supramolecular complexes in solution.

[a] Prof. C. A. Hunter, M. J. Packer
Krebs Institute for Biomolecular Science, Department of Chemistry
University of Sheffield, Sheffield S3 7HF (UK)
Fax: (+44) 114-2738673
E-mail: c.hunter@sheffield.ac.uk

Computational Methods

We have adapted the approach of Williamson and Asakura who have developed a method for predicting the chemical shifts of protons in proteins.[8, 9, 11, 12] The main points are reiterated here.

The chemical shift of a proton is given by Equation (1). For CIS values, clearly $\sigma_{diamagnetic}$ is irrelevant, since the covalent connectivity of the system is unaltered on complexation. It is the other three terms which concern us,

$$\sigma = \sigma_{diamagnetic} + \sigma_{anistropy} + \sigma_{electric\ field} + \sigma_{ring\ current} \quad (1)$$

and these can be calculated by the use of models from the literature. In the current work, we deal with amides and aromatic rings, but the method could be generalised to other functional groups. For amides, the anisotropy effect of the C=O and C–N bonds is given by Equation (2). The values of $\Delta\chi_1$ and $\Delta\chi_2$ used were -25.7×10^{-30} cm^3 and -13.5×10^{-30} cm^3 for the C=O bond and -20.6×10^{-30} cm^3 and -13.2×10^{-30} cm^3 for the C–N bond.[13, 14] The other parameters are defined in Figure 1a. The electric field effect of the amide group was calculated according to Equation (3).

$$\sigma_{anistropy} = (1/3r^3)\,[\Delta\chi_1(1-3\cos^2\theta) + \Delta\chi_2(1-3\sin^2\theta\sin^2\gamma)] \quad (2)$$

$$\sigma_{electric\ field} = \varepsilon_1\{\Sigma(q_i/r_i^2)\cos\theta_i\} + \varepsilon_2[\{\Sigma(q_i/r_i^2)\cos\theta_i\}^2 + \{\Sigma(q_i/r_i^2)\sin\theta_i\}^2] \quad (3)$$

The values of ε_1 and ε_2 used here were -2.0×10^{-12} esu and -1.0×10^{-18} esu.[14] The charges, q, were -1.980 for the amide oxygen, $+1.365$ for the carbon, -1.133 for the nitrogen and $+0.988$ for the hydrogen.[15] The other parameters are defined in Figure 1b.

The ring current shifts were calculated by means of the Haigh-Mallion model [Eq. (4)].[16] The value of B used here was 5.46×10^{-6} Å, and f is the ring current intensity factor: 1.0 for six-membered rings and 0.55 for five-

  0947-6539/99/0506-1891 $ 17.50+.50/0

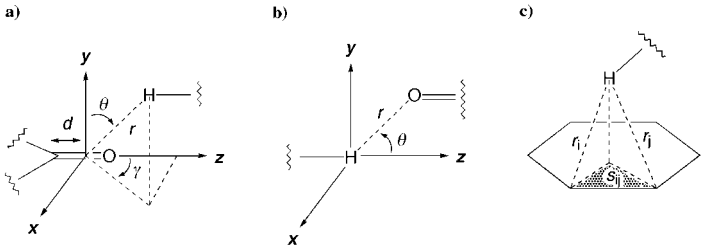

Figure 1. Definition of parameters used in the calculation of CIS values. a) Parameters for calculation of bond anisotropy effects, illustrated for an the amide carbonyl group. We used values of d of 1.10 Å for the amide C=O bond and 1.13 Å for the amide C–N bond.[14] b) Parameters for calculation of electric field effects, illustrated for an amide carbonyl group. c) Parameters for calculation of ring current effects.

membered rings.[7] The other parameters are defined in Figure 1c. Thus the CIS is given by Equation (5).

$$\sigma_{\text{ring current}} = fB\{\Sigma s_{ij}(r_i^{-3} + r_j^{-3})\} \quad (4)$$

$$\Delta\delta = \sigma_{\text{anistropy}} + \sigma_{\text{electric field}} + \sigma_{\text{ring current}} \quad (5)$$

We wrote a simple Fortran program to calculate these CIS values, $\Delta\delta$. This program operates in two different modes: it can be used to calculate CIS values for a given input structure (or set of structures) of the complex (fixed mode); or it can be used to perform a conformational search to determine the three-dimensional structure of a complex for which the calculated CIS values best match a set of experimentally determined CIS values (search mode). These operations are described in detail below.

Fixed mode: We used two input files: an atom-label file that contained information about atoms which belong to amide groups and aromatic rings, magnetically equivalent molecules and protons, and the experimentally observed CIS values; and a Macromodel structure file or multi-structure file that contained the structure(s) to be evaluated.[17] For each structure in the Macromodel input file, Equations (2)–(5) were used to calculate $\Delta\delta$ for each proton in the complex. The values for magnetically equivalent protons were then averaged. The difference between the experimental CIS value and the calculated value was determined for each NMR signal, and the root mean square (RMS) difference between the calculated and experimental CIS values, $R_{\Delta\delta}$, provided a quantitative measure of how closely the structure in the Macromodel input file matched the structure of the complex in solution. The output files contained $\Delta\delta$ for each proton, the averaged value for magnetically equivalent protons, the difference between that value and the experimental CIS, and the overall RMS difference. When the Macromodel input file was a multi-structure file, the structures were ranked based on $R_{\Delta\delta}$, so that it was straightforward to locate the structure for which the calculated CIS value best matched the experimental value. For example, the multi-structure output file from a molecular mechanics conformational search can be tested for how well the structures agree with the NMR CIS data.

Search mode: A genetic algorithm was used to minimise the RMS difference between the calculated and experimental CIS values, $R_{\Delta\delta}$, as a function of the relative position and orientation of the molecules and the internal torsion angles. The genetic algorithm (GA) was implemented with the use of the SUGAL program package.[18] SUGAL provides a generic set of routines that simulate most of the common GA methods.[19, 20] This allowed us to experiment extensively with the set up of the GA in order to optimise the speed and accuracy of the search procedure. In order to use a genetic algorithm, it is necessary to define a fitness function and coding for the problem. The fitness function is defined such that a high fitness value corresponds to a good solution. The coding is used to construct binary chromosomes that represent possible solutions to the problem, each with a distinct fitness. The GA then uses crossover and mutation operations on the chromosomes in order to evolve solutions of high fitness.[20]

The GA attempts to maximise the fitness, and a good solution will have a small $R_{\Delta\delta}$ value, so the fitness function was defined as $R_{\text{expt}}/R_{\Delta\delta}$, in which R_{expt} is the RMS of the experimentally observed CIS values. The chromosomes coded for a set of intermolecular variables (the three global rotations and three global translations for each molecule to be moved) and a set of intramolecular variables (the internal torsion angles to be varied). The variables were all coded as integers in the binary chromosome in order to provide a compact representation. Conversion of the code to a fitness value then required the integer variables to be read from the chromosome and converted to angles or distances. Three types of input file were used: an atom-label file as in the fixed mode; a Macromodel structure file for each molecule in the complex (here only two molecules are considered); a file which specifies the torsion angles that are to be varied, the molecules that are to be moved and any constraints (see later). For any given chromosome, the intramolecular variables were used to change the internal torsion angles by the specified amount, and the intermolecular variables were used to specify the relative positions and orientations of the individual molecules in the complex. The $\Delta\delta$ values were then calculated as in the fixed mode to give $R_{\Delta\delta}$.

The size of the integers used in the coding determines the resolution of the search. If each integer is coded as a four bit string, fifteen possible integers are possible (from -7 to $+7$) and hence fifteen values of each angle or distance are sampled. We generally used eight bit integers, giving a total of 255 possible values. Typically the intermolecular distance was searched in the range ± 10 Å, which implies that it was sampled at increments of 0.08 Å, and a $\pm 180^\circ$ range for the rotations gave increments of 1.4°. The resolution can be increased by increasing the length of the chromosomes (which slows the search down) or restricting the size of the search space. A rank-based generational GA was used with a population of typically 100 chromosomes. In each generation, all but the fittest chromosome could be replaced, subject to the restriction that new chromosomes could enter the population only if they were better than the worst current member. This ensured that the mean fitness did not decrease during the optimisation. The reproduction operators used were single-point crossover and uniform mutation.[18] The mutation operator was allowed to mutate integers over the full range of accessible values with an average of one mutation per chromosome. Rank-based selection was used to reduce the possibility of premature convergence: chromosomes were selected for reproduction based on their relative rank within the population rather than their absolute fitness.[19] This was achieved by the use of rank-linear normalisation with a bias of 5, so that the fittest chromosome was given a ranked fitness five times larger than that of the least fit chromosome, with the remainder spaced evenly between. The total fitness was then normalised. Annealing was also used to maintain diversity in the population: annealing allows inferior chromosomes to enter the population if the factor $\exp\{-(f_{\text{old}} - f_{\text{new}})/T_{\text{anneal}}\}$ is greater than a randomly generated number between 0 and 1; f_{old} and f_{new} are the fitnesses of the chromosomes and T_{anneal} is the annealing temperature.[18] A high annealing temperature allows chromosomes of very low fitness to enter the population and hence maintains population diversity during the run. Diversity is essential to ensure that all regions of the solution space are explored. The annealing temperature was set at 500 in all runs.

The size of the search space was set to ensure that all possible complex conformations could be sampled, for example, the intermolecular distance was typically searched in the range ± 10 Å even though many conformations within this range will not be good candidate solutions. If the molecules are not in van der Waals contact, it is unlikely that a change of torsion angles will improve $R_{\Delta\delta}$, and so the searches converged very slowly. There are two solutions to this problem: restrict the search space by introducing interatomic distance constraints or once a reasonable structure has been found, restart the GA and perform a local search in the region of this structure. Both strategies have been implemented as described below.

Three types of distance constraint were introduced to restrict the search space: user defined interatomic distances, NOE constraints and van der Waals (VDW) clashes. Interatomic distance constraints can be used for example for cyclic systems, in which one bond is allowed to break in order to let the conformation of the ring vary freely. The constraint discriminates against structures that do not bring the atoms involved within bonding distance to enable ring closure. An optimum value for the interatomic distance, d_{con}, was specified along with a range, $\pm\varepsilon$, in which no penalty was incurred. If this constraint was violated, the penalty term in Equation (6) was added to $R_{\Delta\delta}$, in which d_{ij} is the distance between the two constrained atoms in the structure.

$$\text{Penalty} = \{(d_{ij} - d_{\text{con}})/\varepsilon\}^2 \quad (6)$$

0947-6539/99/0506-1892 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 6

NOE constraints dramatically improve the speed with which conformational searches converge, because they impose a significant restriction on the size of the conformational space to be searched. In this case, a maximum allowed separation, d_{NOE}, was defined for the two protons involved. The interatomic distance was calculated for all magnetically equivalent pairs of protons, and the shortest separation in this set of distances, $d_{min,}$ was found. If $d_{min} < d_{NOE}$, then the constraint was satisfied and no penalty was applied. If this constraint was violated, the penalty term in Equation (7) was added to $R_{\Delta\delta}$.

$$\text{Penalty} = (d_{min} - d_{NOE})^2 \quad (7)$$

For the VDW clash constraints, two distances were specified: the minimum allowed separations for non-hydrogen atoms, d_{inter} for intermolecular clashes and d_{intra} for intramolecular clashes. If the distance between any two non-hydrogen atoms d_{ij} was less than the relevant constraint, the penalty term in Equation (8) was added to $R_{\Delta\delta}$. Typically, values of 2 Å and 1 Å were used for d_{inter} and d_{intra}, respectively, so that only severe VDW clashes were eliminated, and the search space was not biased by the choice of these values. To avoid problems with very short interatomic separations, the maximum value for the VDW penalty was fixed at 100.

$$\text{Penalty} = 10/d_{ij} \quad (8)$$

A windowing procedure was also implemented to improve the convergence of the search and the resolution of the final structure. For example, the GA was started from a random conformation for which the intermolecular distance varied between ±10 Å and the orientation angles between ±180°. When the fitness reached 2 (i.e., $R_{\Delta\delta}/R_{expt} = 0.5$), the search space was halved in all dimensions, and the GA was restarted from the optimum structure from the previous run. This was repeated when the fitness reached 4, 8, 16 and so on. The effect was to reduce the search space at every restart and improve the resolution of the structure. No stopping condition was placed on the calculation except the number of generations, which was typically set to 500. In our experience, a fitness of 10 or more corresponds to a good solution.

It is of interest to consider the size of search space that the GA is exploring and whether we could use any other optimisation technique to tackle this problem. The problem is made discrete by using an integer representation for the angles and distances. If we consider the case in which there are 10 intramolecular torsion angles, this gives us a total of 16 variables. If we use an eight bit integer representation, this gives 255 possible values for each variable and hence $(255)^{16}$ possible conformations to explore. This is of the order of 10^{38} conformations, a number far too large for an exhaustive search. Genetic algorithms, however, have been shown to perform extremely well on problems in which the search space increases exponentially with the number of variables.[20] We have also found that $R_{\Delta\delta}$ shows large changes for relatively small conformational motions, so the fitness landscape is quite flat with a spike corresponding to the optimum structure. Genetic algorithms are ideally suited to the optimisation of such functions, where standard optimisation techniques such as simplex or analytic derivative methods would struggle. Hence, both the size of the search space and the nature of the function to be optimised favoured the use of a genetic algorithm. By optimising the precise set up of the genetic algorithm, we have been able to implement an efficient search method which would be difficult by other means.

Results and Discussion

Evaluation of the method: We tested the approach on a simple supramolecular complex, for which we have both a high-resolution three-dimensional structure from X-ray crystallography and accurate ^{1}H NMR CIS values in solution (Figure 2, Table 1).[21] This complex contains only the functional groups discussed above, and the molecules are relatively rigid, so it makes an ideal test case. Using the X-ray crystal structure geometry and the fixed mode described above, we calculated an $R_{\Delta\delta}$ value of 0.044 ppm (Table 1). The glycine anhydride CH_2 signal gave the largest error in $\Delta\delta$ of 0.3 ppm. This discrepancy reflects approximations inherent in the method: the experimental CIS values are the time-average of all conformations thermally accessible to the complex at room temperature, and we have calculated CIS values for a single rigid structure; the calculations use approximate models and a set of empirical parameters from the literature.

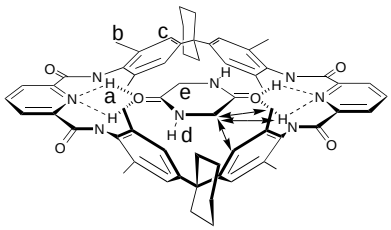

Figure 2. The complex formed between a macrocyclic tetraamide receptor and glycine anhydride in chloroform. Intermolecular NOEs identified in ROESY experiments and the proton labelling scheme are shown.

Table 1. Complexation-induced changes in chemical shift [ppm] in chloroform for the macrocycle complex in Figure 2 (see Figure 2 for proton labelling scheme).

Proton	Experiment	X-Ray crystal structure	NMR structure
a	+1.50	+1.42	+1.48
b	0.00	−0.07	−0.05
c	+0.04	+0.03	+0.01
d	−1.19	−1.11	−1.18
e	−1.19	−1.47	−1.18
$R_{\Delta\delta}$	–	0.044	0.013

However, the calculated CIS values are extremely sensitive to very small changes in conformation. We therefore used the search mode to find the structure for which the calculated CIS values provide the best match to the experimental values for this complex. The two individual molecules were constructed with Macromodel and the energy was minimised with the MM3 force-field.[17] These two structure files were then used as the input starting structures for the conformational search. The molecules were allowed to translate (±10 Å) and rotate (±180°) relative to one another, and the only torsions that were allowed to vary (±180°) were those which permitted rotation of the four aromatic side walls of the host along their axis as shown in Figure 3. The three NOE constraints shown in Figure 2 were introduced with d_{NOE} set to 5 Å. With a population of 50, a windowed search converged to a fitness

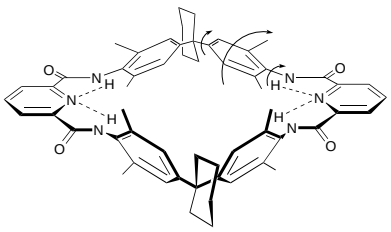

Figure 3. Torsion angles that are allowed to vary in the GA conformational search on the macrocycle complex shown in Figure 2. All four symmetry-related xylyl groups are allowed to rotate on their axes by allowing coupled changes in two torsion angles as indicated.

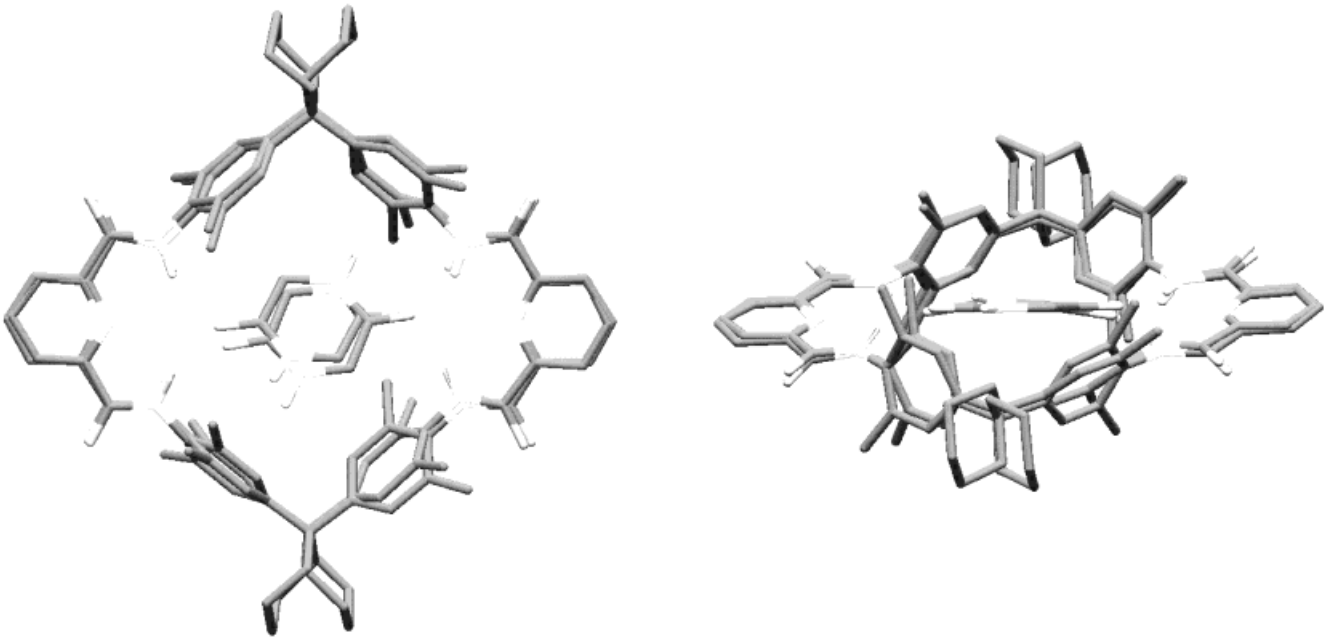

Figure 4. Two different views of the optimised NMR structure of the macrocycle complex superimposed on the X-ray crystal structure. The RMS difference in the positions of the non-hydrogen atoms (excluding the cyclohexyl groups) is 0.37 Å.

of greater than 32 after 2020 generations giving the structure shown in Figure 4 ($R_{\Delta\delta}$ = 0.013 ppm). The $\Delta\delta$ values for individual signals are listed in Table 1: all agree with the experimental values to within 0.1 ppm. An overlay of the optimised NMR structure and the X-ray structure shows very good agreement (Figure 4). If the cyclohexyl rings, which are on the outside of the complex and not defined by the NMR data, are excluded, the RMS difference between the non-hydrogen atom positions is only 0.37 Å.

Although the CIS values calculated for the X-ray structure do not exactly match the experimental values, the experimental CIS values can be used to dock the two molecules together to obtain a high-resolution structure that is in very good agreement with the X-ray structure. The reason is that a small deviation in geometry produces a large change in the calculated CIS value. Thus it appears that NMR CIS data can be used to determine the three-dimensional structure of complexes of this type to a reasonably high degree of accuracy.

Application to structure determination: We now address the problem of structure determination. We originally started this work in an effort to solve a real structure determination problem in our laboratory. We have been working on the supramolecular zipper system shown in Figure 5 for some years.[22–24] Evidence for the structure was obtained from limiting CIS values from ^{1}H NMR titrations (Table 2), intermolecular NOEs (Figure 5a) in deuterochloroform and the X-ray crystal structure of a model compound that represents one half of the zipper complex (Figure 5b).

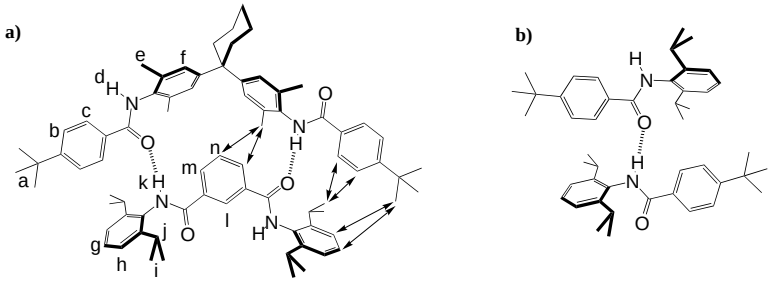

Figure 5. a) The structure of the zipper complex deduced from the CIS values in Table 2 and the intermolecular NOEs indicated. The proton labelling scheme is also shown. b) Illustration of the intermolecular interactions present in the X-ray crystal structure of a related model compound.

Table 2. Complexation-induced changes in chemical shift [ppm] in chloroform for the zipper complex in Figure 5 (see Figure 5 for proton labelling scheme).

		Molecular mechanics structures			
Proton	Experiment	MM3	AMBER	OPLS	NMR structure
a	−0.07	−0.05	0.00	0.00	−0.04
b	−0.51	−0.52	−0.23	−0.20	−0.46
c	−0.26	−0.31	−0.09	+0.05	−0.24
d	+1.13	+1.07	−0.06	+0.17	+1.13
e	−0.03	−0.06	−0.06	−0.09	−0.01
f	+0.22	+0.04	−0.08	−0.08	+0.19
g	+0.06	+0.16	−0.05	−0.07	+0.09
h	−0.01	+0.09	−0.14	−0.11	0.00
i	−0.15	−0.09	−0.09	−0.07	−0.16
j	−0.04	+0.08	+0.13	+0.15	+0.05
k	+1.39	+0.87	+1.12	+1.29	+1.39
l	+0.04	−0.11	−0.24	−0.16	−0.04
m	−0.42	−0.95	−1.17	−0.95	−0.42
n	−1.63	−1.93	−1.27	−1.10	−1.65
$R_{\Delta\delta}$		0.036	0.066	0.057	0.010

0947-6539/99/0506-1894 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 6

We also used Macromodel to carry out Monte Carlo conformational searches on this complex with three different force fields, MM3, AMBER and OPLS, using the continuum dielectric model for chloroform solvation (these force fields are possibly not the best that can be applied to this system, but they are the standards available in the Macromodel package).[17] The optimum MM3 structure is similar to that shown in Figure 5 with four edge-to-face aromatic interactions and two hydrogen-bonds (Figure 6a), but OPLS and AMBER produce a different structure, which is shown in Figure 6b. There are still two hydrogen bonds, but the aromatic rings are twisted to give three stacking interactions with the fourth pair of aromatic rings separated by a large distance. OPLS and AMBER do find the MM3 structure, but it is 18 kJ mol^{-1} above the OPLS optimum structure and 14 kJ mol^{-1} above the AMBER optimum structure. The difference between the two structures in Figure 6 is not enormous, but for our purposes it was significant. We have been using the zippers to measure the magnitudes of the two terminal aromatic interactions using chemical double-mutant cycles.[25–27] The thermodynamic analysis strongly suggested that both interactions contribute 1.0–1.5 kJ mol^{-1} to the free energy of complexation, but the OPLS/AMBER structure cast some doubt on what we were actually measuring. How then can we distinguish between the two possible conformations?

In principle, NOE data could resolve the problem. We have not carried out a quantitative NOE experiment to derive distances, but Table 3 shows that this might be difficult to achieve. Short inter-proton distances are listed for the two types of structure. All of the short contacts in the MM3 structure (Figure 7a, entries a–n in Table 3) are also short contacts in the OPLS/AMBER structure, and so the NOEs that are actually observed are consistent with both structures (see bold-face entries in Table 3). There are some very short contacts in the OPLS/AMBER structure that do not occur in the MM3 structure (Figure 7b, entries w–z in Table 3). These NOEs are not observed experimentally, which is encouraging

Table 3. Inter-proton distances [Å] in the zipper complexes (see Figure 7 for NOE labelling scheme). The distances in the OPLS structure are very similar to those listed for the AMBER structure.

NOE	MM3 structure	AMBER structure
a	5.1	5.0
b	4.6	4.7
c	**2.7**	**3.1**
d	4.2	4.2
e	**2.5**	**2.0**
f	2.9	2.8
g	4.0	3.7
h	**2.6**	**2.5**
i	5.1	5.4
j	**3.2**	**3.5**
k	3.5	3.7
l	3.1	3.1
m	**3.0**	**2.6**
n	**2.8**	**2.5**
w	3.4	3.0
x	4.5	3.9
y	5.1	2.1
z	6.1	2.6

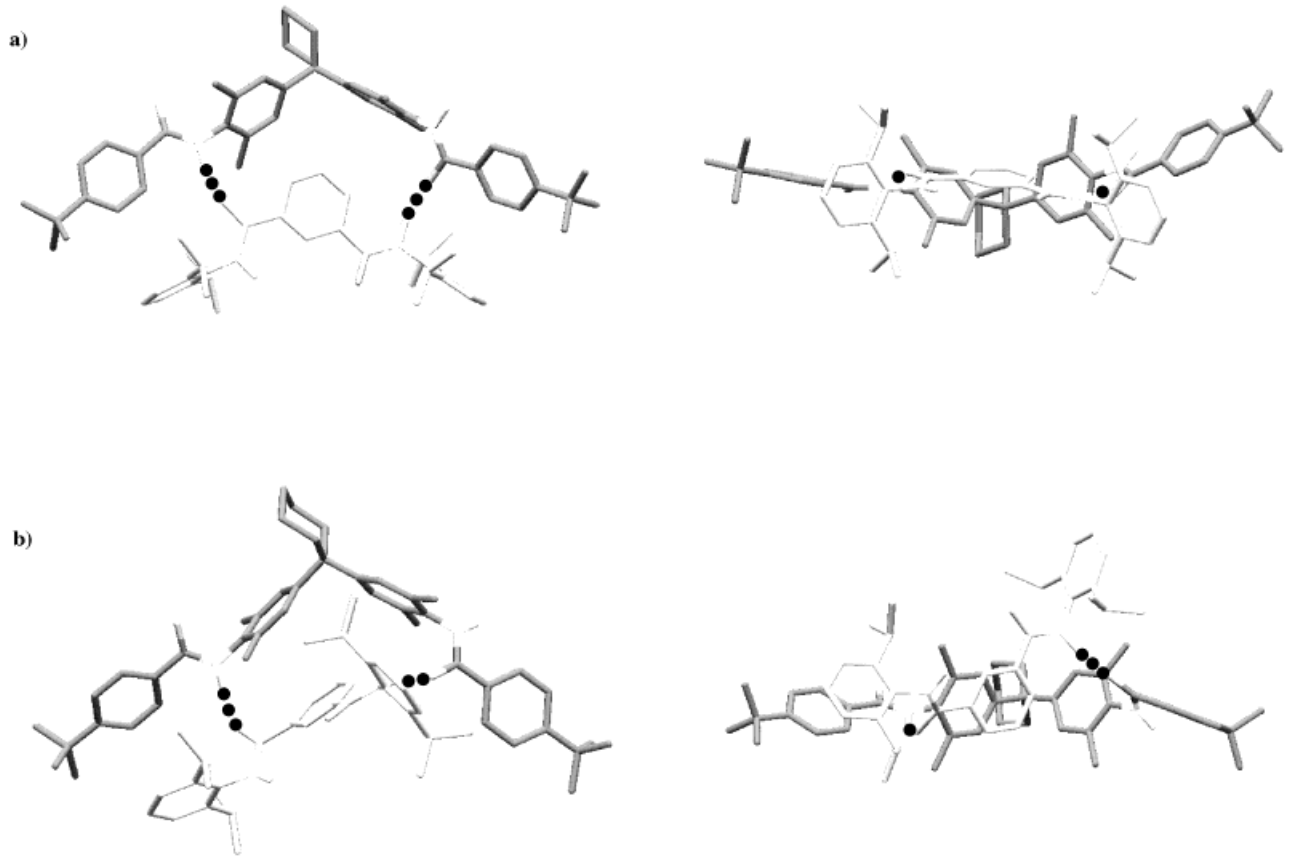

Figure 6. Two different views of the molecular mechanics structures of the zipper complex. a) The optimum MM3 structure. b) The optimum AMBER/OPLS structure.

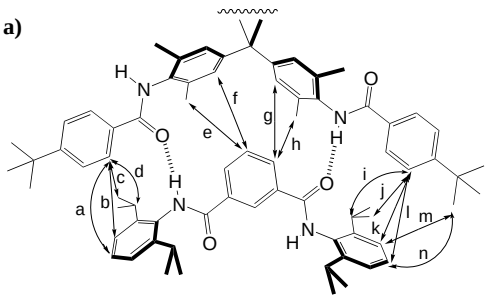

Figure 7. Short intermolecular distances found in the molecular-mechanics structures in Figure 6 (values for these distances are listed in Table 3). a) All inter-proton distances below 5.1 Å in the MM3 structure. b) Short inter-proton distances found exclusively in the AMBER/OPLS structures.

but does not provide very strong evidence for the MM3 structure.

We therefore turned to the CIS values, since there is a clear difference in the way the aromatic rings interact, and this should manifest itself in a difference in the ring-current-induced changes in chemical shift. Use of the fixed mode and the optimised molecular mechanics structures gave the $\Delta\delta$ values shown in Table 2. Clearly, the MM3 structure matches the experimental CIS values much more closely than the OPLS and AMBER structures.

We also used the search mode to find the conformation for which the calculated CIS values best match the experimental values. The two individual molecules were built with Macromodel and the energy was minimised with the MM3 forcefield, and these two structure files were used as the input. The molecules were allowed to translate (± 10 Å) and rotate ($\pm 180^\circ$) relative to one another, and Figure 8 shows the

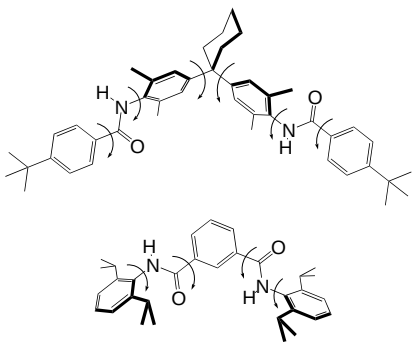

Figure 8. Torsion angles that were allowed to vary in the GA conformational search on the zipper complex.

torsion angles that were allowed to vary ($\pm 180^\circ$). The six NOE constraints shown in Figure 5 were introduced with d_{NOE} set to 5 Å. With a population of 200, a windowed search converged to a fitness of 18 after 3000 generations giving the structure shown in Figure 9 ($R_{\Delta\delta} = 0.010$ ppm). The $\Delta\delta$ values for individual signals are listed in Table 2: all agree with the experimental values to within 0.1 ppm. The optimised NMR structure is essentially identical to the MM3 structure (Figure 9): the only difference is in the orientation of one of the central aromatic rings of the bisaniline component of the complex, but all four edge-to-face aromatic interactions are clearly present. This confirms that the conformation illustrated in Figure 5 best represents the solution structure of this complex.

Conclusions

We have developed software for calculating complexation-induced changes in ^{1}H NMR chemical shift for supramolecular complexes, and shown how these data can be used for supramolecular structure determination in solution. The approach has been validated for a system for which we have an X-ray crystal structure to compare with the optimised NMR structure, and the agreement is remarkably good. The systems discussed here are relatively rigid and have a limited number of functional groups. It remains to be seen how well the method will work for more complex and more flexible systems.[10–12] However, the results presented show that for some systems at least, the method provides a powerful tool for determining high-resolution three-dimensional structures of supramolecular complexes in solution.

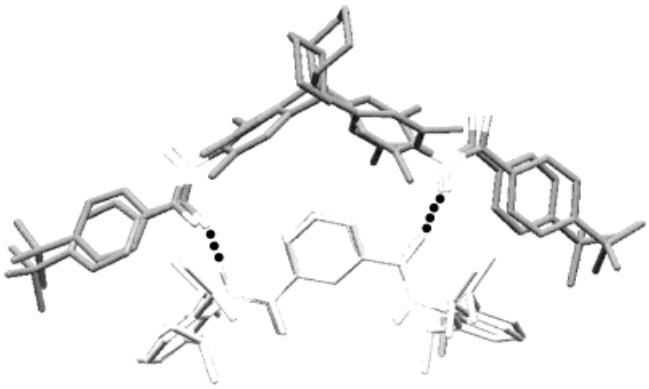

Figure 9. The optimised NMR structure of the zipper complex superimposed on the MM3 structure.

Acknowledgment

We thank the BBSRC (MJP) and the Lister Institute (CAH) for financial support. We also acknowledge the contribution of Dr. M. P. Williamson and thank him for many helpful discussions.

[1] X. Chi, A. J. Guerin, R. A. Haycock, C. A. Hunter, L. D Sarson, *J. Chem. Soc. Chem. Commum.* **1995**, 2567–2569.
[2] R. J. Abraham, G. R. Bedford, D. McNeillie, B. Wright, *Org. Mag. Res.* **1980**, *14*, 418–425.
[3] R. J. Abraham, K. M. Smith, *J. Am. Chem. Soc.* **1983**, *105*, 5734–5741.
[4] R. J. Abraham, C. J. Medforth, *Mag. Res. Chem.* **1988**, *26*, 803–812.
[5] P. Leighton, J. A. Cowan, R. J. Abraham, J. K. M. Sanders, *J. Org. Chem.* **1988**, *53*, 733–740.
[6] H.-J. Schneider, F. Hachret, V. Rudiger, H. Ikeda, *Chem. Rev.* **1998**, 1755–1785.

0947-6539/99/0506-1896 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 6

[7] K. Osapay, D. A. Case, *J. Am. Chem. Soc.* **1991**, *113*, 9436–9444.
[8] M. P. Williamson, T. Asakura, E. Nakamura, M. Demura, *J. Biomol. NMR* **1992**, *2*, 83–98.
[9] M. P. Williamson, T. Asakura, *J. Mag. Res. B* **1993**, *101*, 63–71.
[10] L. Szilagyi, *Prog. Nucl. Magn. Reson. Spectrosc.* **1995**, *27*, 325.
[11] M. P. Williamson, J. Kikuchi, T. Asakura, *J. Mol. Biol.* **1995**, *247*, 541–546.
[12] M. Iwadate, T. Asakura, M. P. Williamson, *Eur. J. Biochem.* **1998**, *257*, 479–487.
[13] R. F. Zurcher, *Prog. Nucl. Magn. Reson. Spectrosc.* **1967**, *2*, 205–257.
[14] T. Asakura, I. Ando, A, Nishioka, *Makromol. Chem.* **1977**, *178*, 1111–1132.
[15] J. G. Vinter, *J. Comput. Mol. Des.* **1994**, *8*, 653–668.
[16] C. W. Haigh, R. B. Mallion, *Prog. Nucl. Magn. Reson. Spectrosc.* **1980**, *13*, 303–344.
[17] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comput. Chem.* **1990**, *11*, 440–467.
[18] A. N. Hunter, *SUGAL Genetic Algorithms Package—Version 2.1*, University of Sunderland, UK, **1995**.
[19] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, **1991**.
[20] D. E. Goldberg, *Genetic Algorithms in Search*, *Optimisation and Machine Learning*, Addison-Wesley, **1989**.
[21] H. Adams, F. J. Carver, N. J. Osborn, C. A. Hunter, *Chem. Commun.* **1996**, 2529–2530.
[22] A. P. Bisson, F. J. Carver, C. A. Hunter, J. P. Waltho, *J. Am. Chem. Soc.* **1994**, *116*, 10292–10293.
[23] A. P. Bisson, C. A. Hunter, *Chem. Commun.* **1996**, 1723–1724.
[24] A. P. Bisson, C. A. Hunter, J. C. Morales, K. Young, *Chem. Eur. J.* **1998**, *4*, 845–851.
[25] H. Adams, F. J. Carver, C. A. Hunter, J. C. Morales, E. M. Seward, *Angew. Chem.* **1996**, *108*, 1628–1631; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1542–1544.
[26] H. Adams, K. D. M. Harris, G. A. Hembury, C. A. Hunter, D. Livingstone, J. F. McCabe, *Chem. Commun.* **1996**, 2531–2532.
[27] F. J. Carver, C. A. Hunter, E. M. Seward, *Chem. Commun.* **1998**, 775–776.

Received: September 11, 1998 [F1342]