1 and 0.5, respectively). Also noteworthy is that in this structure O6-H1 is hydrogen bonded to a solvent molecule instead of to O4. The other hydroxyls are also hydrogen bonded to water molecules, as expected. In the binding region, 5 hydrogen-bonded water molecules are observed (2 to O2, 1 to O3, 2 to O6-H1), consistent with the analysis above.

Conclusions

Conformational searching of FK506 in isolation revealed the existence of a large number of low-lying energy minima. Both cis and trans amide isomers are found among these minima and appear to favor different conformations of nearby torsions. Since two of these torsions, about C1-C2 and C8-C9, are in the binding region of FK506, these preferences may be relevant to the preferential binding of the trans isomer to FKBP.

Molecular dynamics simulation in water over 200 ps revealed some interesting conformational behavior. Notable observations are the flexibility of the binding region of FK506 and the sampling of two conformations of the macrocyclic ring. The solution structure, as obtained from the simulation, correlates well with the X-ray crystal structure, even to the extent of positional fluctuations of the individual atoms. Hydration of FK506, as analyzed through hydrogen bonding, reflects significant variations in the exposure of potential hydrogen-bonding sites. There are on average 5 hydrogen bonds between water molecules and FK506 in the critical binding region.

Acknowledgment. Gratitude is expressed to the National Institutes of Health for support of this work. We also thank Dr. Julian Tirado-Rives for computational assistance and Professor Stuart L. Schreiber for valuable discussions.

Registry No. FK506, 104987-11-3.

Supplementary Material Available: Complete listing of AMBER parameters for FK506 and stereopictures and tables of coordinates of the structures found in the conformational search (25 pages). Ordering information is given on any current masthead page.

Elucidation of Solution Structures by Conformer Population Analysis of NOE Data

Clark Landis* and Viyola S. Allured†

Contribution from the Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53705, and the Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309. *Received June 14, 1991*

Abstract: The elucidation of molecular structures from NOE data may be complicated by the presence of multiple solution conformations. Conformer population analysis (CPA) is a new procedure for obtaining solution structures from NOE data that accommodates multiple conformations and avoids some of the biases inherent in the construction of single structures from NOE distance matrices. The CPA method utilizes trial conformers generated by molecular modeling techniques and solves for the set of conformer populations that yield the most significant best-fit to the NOE data. Simulations of transient NOE experiments and steady state NOE experiments illustrate the method and the factors that influence the qualities of the results.

I. Introduction

NMR methods for the elucidation of molecular structures in solution[1] are popular and powerful because the methods do not require special molecular properties (such as single-crystal growth required by crystallography or significant vapor pressures as required for electron-diffraction work) and because the methods probe the structure in the liquid phase, the most common reaction medium. These methods have been applied extensively to the determination of peptides and small proteins.

Structural information arises from NMR experiments of two types: the NOE-based techniques (NOESY,[2] ROESY,[3] and related methods[4]), which express internuclear distance information, and scalar coupling measurements,[5] which are sensitive to torsion angles. Whereas the scalar coupling measurements yield information about the local structure, the NOE-based methods uniquely reveal gross (tertiary) structural features in addition to local features. Thus, modern NMR methods for protein sequencing and the determination of secondary and tertiary structure strongly emphasize ^{1}H NOE measurements.[1]

The interpretation of NOE information commonly follows one of two courses:[6] (1) a "best-fit" approach or (2) a molecular construction approach. The former viewpoint employs a predefined collection of different structural possibilities, or conformers, and seeks to distinguish which structures are consistent, either collectively or individually, with the NMR data. For this approach the NOE intensities of individual, predefined conformers are computed and then compared with observed intensities. The most elaborate applications of this approach attempt to maximize the fit of observed and calculated NOE intensities by varying the populations of the conformers. A disadvantage of this method

* University of Wisconsin.
† University of Colorado.

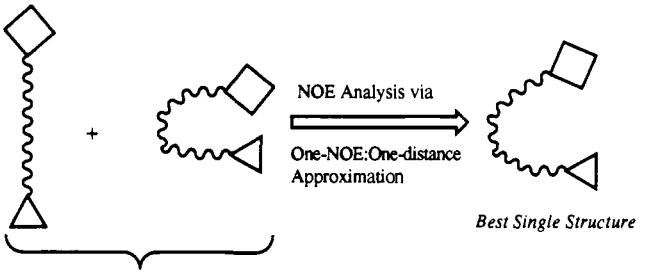
(1) (a) Wutrich, K. *Acc. Chem. Res.* **1989**, *22*, 36. (b) Wutrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(2) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546-4553.
(3) (a) Bothner-By, A. A.; Stephens, R. L.; Lee, J.; Warren, C. D.; Jeauloz, R. W. *J. Am. Chem. Soc.* **1984**, *106*, 811-813. (b) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1986**, *68*, 568-574. (c) Kessler, H.; Griesinger, C.; Kerssebaum, R.; Wagner, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1987**, *109*, 607-609. (d) Griesinger, C.; Ernst, R. R. *J. Magn. Reson.* **1987**, *75*, 261-271.
(4) (a) Dobson, C. M.; Olejniczak, E. T.; Poulsen, F. M.; Ratcliff, R. G. *J. Magn. Reson.* **1982**, *48*, 97. (b) Bother-By, A. A.; Noggle, J. H. *J. Am. Chem. Soc.* **1979**, *101*, 5152. (c) Kessler, H.; Oschkinat, H.; Griesinger, C.; Bermel, W. *J. Magn. Reson.* **1986**, *70*, 106-133. (d) Williamson, M. P.; Neuhaus, D. *J. Magn. Reson.* **1987**, *72*, 369. (e) Noggle, J. H.; Schirmer, R. E. *The Nuclear Overhauser Effect*; Academic Press: New York, 1971.
(5) (a) Bystrov, V. F. *Prog. Nucl. Magn. Reson. Spectrosc.* **1976**, *10*, 41-81. (b) Pardi, A.; Billeter, M.; Wurtrich, K. *J. Mol. Biol.* **1984**, *180*, 741. (c) Haasnoot, C. A. G.; DeLeeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783.
(6) Williamson, M. P.; Neuhaus, D. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; VCH: New York, 1989.

0002-7863/91/1513-9493$02.50/0

is the requirement of predefined conformations; i.e. the method does not direct the construction of molecular structures based on NOE information.

The molecular construction viewpoint[7] is founded on the assumption that each NOE cross-relaxation is associated with a single interproton distance—i.e. a one cross-relaxation:one distance hypothesis is made. This leads to the expression of the NOE information as a single, "average" structure. From full relaxation matrix analysis (or less accurately, the independent spin pair approximation) of the NOE data an interproton distance matrix is constructed which can be used as a matrix of constraints for molecular dynamics, molecular mechanics, or distance geometry calculations. Resulting from these calculations is an average structure which best satisfies the combination of NOE constraints and other geometrical constraints (internal coordinate constraints, van der Waals radii, etc).

An obvious difficulty associated with the construction of a molecular structure from an NOE-derived matrix of distance constraints is the possibility that more than one conformer may exist in solution.[8] The presence of multiple conformations invalidates the one cross-relaxation:one distance assumption. If these conformers are very different the average structure may have little utility because it may be so strained that its probability of existing is very low. Furthermore, because of the r^{-6} weighting of the NOE measurement the average structure is highly biased to be most similar to the conformer with the shortest interproton distances. Consider the simple situation illustrated below with equal population of two conformers (linear and bent) of a short-chain molecule. Because of the strong (r^{-6}) distance dependence of the NOE interaction, any structural analysis which attempts to fit the experimental data with a single structure will be highly biased toward the structure that gives crosspeaks between the chain termini—i.e. toward the bent structure.



Another common situation that invalidates the one cross-relaxation:one distance approximation is the accidental overlap of otherwise inequivalent resonances. Because the observed cross-relaxation represents a summation of two different cross-relaxations, the assumption that a single distance applies is accurate only when just one of the overlapping resonances is involved in the cross-relaxation.

The limitations of the molecular construction approach summarized in the preceding two paragraphs suggest that, in principle, the alternative "best-fit" approach is a more realistic and less biased method of extracting structural information from NOE data. However, the practical implementation of this approach raises a number of questions. How are the predefined conformations generated? Do the predefined conformations span all reasonably accessible conformational space? Does the group of predefined conformations contain redundancies (or how do we determine if two conformations are similar)? How does one determine the conformer populations which yield the best-fit to the data? What

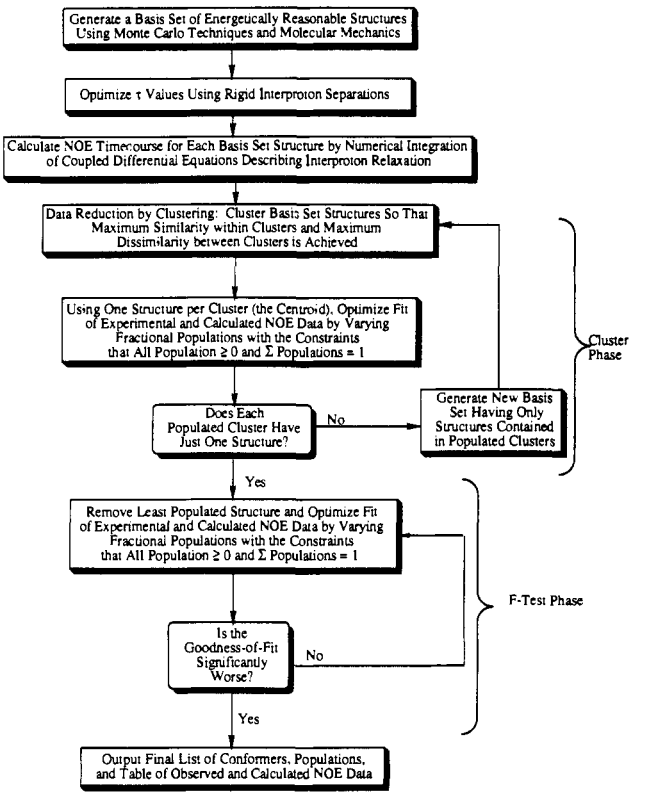


Figure 1. Flowchart for the analysis of NOE data by the conformer population analysis (CPA) technique.

is the significance of the best-fit conformer populations? Is the solution the most parsimonious description of the data? What are the influences of conformer interconversion kinetics on the solutions?

As an initial attempt at resolving some of these questions we have developed a new "best-fit" method, conformer population analysis (CPA), for the generation of solution structures from NOE data. The subsequent sections describe the premises and computational techniques of the method followed by two illustrations of its application to small-molecule structures.

II. CPA Computational Method

The CPA procedure comprises four stages (Figure 1): (1) creation of a collection of conformers, (2) calculation of direct and cross-relaxation rates for each conformer, (3) fitting and refining conformer populations, and (4) significance testing of the results. The object of the procedure is to obtain the *best fit* to the observed NMR data using the *smallest number of significant* conformers. In this paper we will emphasize the last two stages of the method because the generation of conformers and the backcalculation of NOE timecourses have received extensive discussion in the literature.

A. Conformer Generation. A fundamental premise of our procedure is that empirical force field methods are capable of producing a collection of structures that span all the accessible conformational space of the molecule. Methods[9] for generating the structures may include systematic searches of torsional angles, Monte Carlo methods operating on either torsional angles or atomic Cartesian coordinates, distance geometry methods with internal coordinate and non-bond constraints only, structures taken from quenched molecular dynamics, or structures derived from sequence homology modeling. The technique used in searching is unimportant so long as the resulting collection of structures is capable of representing the available low-energy conformers. For small molecules, systematic searches have the advantage of searching the conformational space most completely; however, these methods may not be practical for large molecules.

In order to span the accessible conformational space, large numbers (>100) of structures may be required depending on the size of the molecule and the degree of structural refinement desired. Conformers

(7) (a) Williamson, M. P.; Havel, T. F.; Wutrich, K. *J. Mol. Biol.* **1985**, *182*, 295. (b) Boelens, R.; Koning, T. M. G.; van der Marel, G. A.; van Boom, J. H.; Kaptein, R. *J. Magn. Reson.* **1989**, *82*, 290. (c) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1990**, *87*, 475–487. (d) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1988**, *79*, 493.

(8) For examples of multiconformational interpretations of solution NMR data see: (a) Inman, W.; Crews, P. *J. Am. Chem. Soc.* **1989**, *111*, 2822–2829. (b) Kessler, H.; Griesinger, C.; Lautz, J.; Muller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393–3396.

(9) (a) Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. D. *J. Am. Chem. Soc.* **1990**, *112*, 1419. (b) Summers, N. L.; Karplus, M. *J. Mol. Biol.* **1989**, *210*, 785.

that are very high in energy may be excluded outright on the basis of energetics. For medium-sized molecules (<150 atoms) we typically exclude from consideration all conformers that have total energies more than 15 kcal/mol (a user-selectable amount) greater than the lowest energy conformer. In principle, one could weight the conformers according to their energies, but this would require that the energy differences be accurate to better than ca. 1 kcal/mol—an unrealistic expectation. In general, we choose not to overly bias the fitting of NMR data with empirical force field energies and implement these energies as a threshold, only, for eliminating unreasonable structures.

B. Computation of NOE Timecourses. A multiconformational approach to analyzing NMR data must make some assumptions concerning the rates of conformer interconversion. For the purposes of this discussion all conformer interconversion rates are assumed to be slow compared with T_1. Hence, the observed NOE timecourse is considered to be the population-weighted sum of the individual NOE timecourses for each conformer. Of course, in many cases this assumption may be unrealistic.

$$I_j(t) = \sum_a \rho_a I_j(t) \tag{1}$$

Conformer exchange kinetics that are fast on the T_1 time scale but slower than molecular rotation require weighted averaging of exchange rates rather than intensities. Corrections for conformational averaging of rates rather than intensities are most important for structures in which distant groups undergo cross-relaxation through conformational transitions of mediating groups, such as phenyl ring flips.[10] In this paper we limit the discussion to slow conformer interconversion rates because of the numerical integration methods employed for the computations. In future versions of the program, the implementation of matrix methods will permit an accurate description of NOE intensities through the full range of conformer exchange kinetics.

Computation of the NOE timecourse for the transient NOE and the steady state NOE experiments are used for the examples of this paper. In addition, our current versions of the program are capable of simulating nonselective T_1 and truncated NOE experiments. For each structure an NOE timecourse is computed by numerical integration of the coupled

$$\frac{\partial I_j}{\partial t} = -\sum_{n \neq j}(\omega_{0jn} + 2\omega_{1jn} + \omega_{2jn})(I_j - I_{0j}) - \sum_{n \neq j}(\omega_{2jn} - \omega_{0jn})(I_n - I_{0n}) \tag{2}$$

Solomon equations[11] (eq 2) where the transition probabilities, ω_{xjn}, depend on the interproton separations and the correlation times (τ). Alternatively, matrix methods could be used for the backcalculation of the NOE intensities. Because methyl groups may have internal motions whose rates are competitive with molecular rotations, three correlation times (τ values) are defined: methyl–methyl, methyl–non-methyl, and non-methyl–non-methyl. The magnetization of each proton is computed as a function of time, and the individual magnetizations of equivalent or overlapping resonances are summed to give the group resonance intensities.

C. Conformer Population Fitting and Refinement. With the assumption of slow conformer interconversion, the best fit of computed and observed NOE's could be obtained by solving for the conformer populations in the set of linear equations describing the magnetization intensities for each observed resonance as a function of time. However the method of conformer generation described previously produces (1) a large number of trial conformers (variables) that may exceed the number of structurally sensitive intensities (observables), and (2) many of the trial structures may be so similar that singularities occur in solving the linear equations. Therefore, we have adopted a clustering procedure for data reduction and the elimination of structural redundancy.

Clustering[12] is a procedure that groups conformers into clusters by maximizing the similarities of data within a cluster and by minimizing the similarities between different clusters. Clustering requires the adoption of a measure of the similarity of two structures, i.e. the distance between two structures in conformational space. Various schemes for measuring distances (D) have been used; in this study we have adopted the Euclidean distance measurement. Because we wish to cluster structures together on the basis of their similarities *as seen by the NOE experiment*, we measure the Euclidean distance using the difference between the interproton separations raised to the −6 power (r_{ij}^{-6}) for all proton pairs exhibiting cross-relaxations in the NOE spectrum. Our programs contain options for r_{ij} and r_{ij}^{-3} based clustering as well.

$$D_{ab} = \left(\sum_{ij}^{\text{NOE pairs}} (r_{ija} - r_{ijb})^2\right)^{1/2} \tag{3}$$

Conformer structures were clustered by a two-step process: an initial configuration was generated by the hierarchical method followed by application of the K-means clustering algorithm. This combination of clustering algorithms has been shown to be robust for recovering the true clusters in the presence of random noise.[13] In general, the optimal number of clusters to be used cannot be defined. For our applications, an obvious goal of clustering is to reduce the data so that the number of clusters is less than the total number of NOE observations (the product of the number of cross-peaks and the number of mixing times). In practice we use a number of clusters (clustering level) that is slightly less than the number of NOE observations and explore the influence of changing the clustering level empirically. The most desirable situation is to use the maximum clustering level consistent with computational stability. The advantage of clustering is that data reduction is accomplished with homogeneous (minimum bias) sampling of the accessible conformational space (vide infra).

Using a reduced basis set corresponding to the conformers which are most representative of each cluster (i.e. closest to each cluster centroid), the calculated NOE timecourse is fit to the observed data by least-squares optimization[14] of the conformer populations under the constraints that the fractional population of each conformer is positive and the sum of the conformer populations is unity. From the fitting a collection of cluster centroids and populations results; many of these populations are likely to be very small (due to the strong distance dependence of the NOE interaction).

Refinement of the conformer populations is accomplished by expanding the basis set of conformers to include all members of clusters with fractional populations greater than a threshold value (usually 0.0001). The resultant basis set is clustered and the conformer populations are optimized. Each cycle of clustering–fitting–expansion is a refinement cycle with an attendant increase in the goodness-of-fit between computed and observed data. Refinement is continued until the number of structures in the active basis set is less than the number of requested clusters. If convergence is not reached, the clustering level is increased in increments of one until convergence occurs.

D. Significance Testing. Central to any multiconformational approach for analyzing NOE data is the question of whether the data justifies the use of multiple conformations. The Hamilton version of the F-test[15] is one method for testing the significance of improvements in fitting that occur as the number of variable parameters is increased. In our implementation, the number of conformers used in the fit is systematically lowered by removing the least populated conformer in the basis set, re-optimizing the populations, and comparing the ratio of R factors according to Hamilton's method. Typically we test for significance at the (i.e. with an emphasis on removing insignificant conformers) 20% level (α = 0.2, or a 20% probability of incorrectly rejecting the null hypothesis). Resulting from this procedure is a set of structures and populations that have been reduced to the minimum, justifiable number of conformers.

III. Application of the CPA Method

CPA has the advantage of permitting multiconformational analyses of solution NOE data. Our emphasis in developing this technique has been the application of systematic methods for considering large numbers of structural alternatives (so as to lessen bias) in attempting to generate the most parsimonious, yet realistic, description of molecular structures in solution. Naturally, questions arise concerning how well the methods work. Is the sampling method free of bias (i.e. is the accessible conformational space sampled homogeneously)? Is the method convergent? Does the method reveal an authentic description of the solution conformers? How sensitive is the method to random noise? To what extent is the final solution dependent on the starting point? Because these questions cannot be answered a priori, we have performed a series of simple simulations in order to test these concerns. Although some of the assumptions (e.g. slow conformer interconversion) made in these simulations are not necessarily realistic, these simulations play an important role in the testing of the methods because the answers are known exactly.

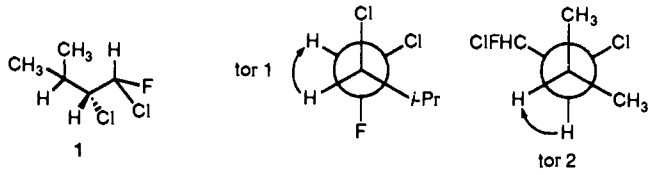
(10) Koning, T. M. G.; Boelens, R.; Kaptein, R. *J. Magn. Reson.* **1990**, *90*, 111–123.
(11) Solomon, I. *Phys. Rev.* **1953**, *92*, 953.
(12) Everitt, B. *Cluster Analysis*; Heinemann Educational Books: New York, 1980.
(13) Milligan, G. W. *Psychometrika* **1980**, *45*, 325–340.
(14) Lawson, C. L.; Hanson, R. J. *Solving Least Squares Problems*; Prentice-Hall: New York, 1974.
(15) Hamilton, W. C. *Acta Crystallogr.* **1965**, *18*, 502.

Table I. Fitting Results for Simulated NOE Data Using Conformational Population Analysis with Clustering and F-Test Significance Testing

			results following F-test phase[a]				
cluster level	noise level	correlation times[b]	fractional populations	tor 1	tor 2	strain E[c]	*R* factor[d]
12	0.0	τ_{1-1} = 1.0 ns	0.24	55	67	2.8	0.0001
basis set of		τ_{1-3} = 0.98 ns*	0.25	51	51	2.9	
50 random		τ_{3-3} = 1.01 ns*	0.26	171	−70	2.93	
structures			0.16	176	−98	4.33	
			0.08	−178	−77	3.16	
9	0.0	τ_{1-1} = 1.0 ns	0.19	55	67	2.8	0.0002
basis set of		τ_{1-3} = 0.98 ns*	0.02	58	81	3.9	
50 random		τ_{3-3} = 1.01 ns*	0.26	51	51	2.9	
structures			0.38	171	−70	2.9	
			0.09	176	−98	4.3	
			0.06	−65	−156	5.5	
9	0.03	τ_{1-1} = 1.0 ns	0.14	58	81	3.9	0.01
basis set of		τ_{1-3} = 0.98 ns*	0.38	51	51	2.9	
50 random		τ_{3-3} = 1.01 ns*	0.48	173	−84	2.6	
structures							
9	0.10	τ_{1-1} = 1.0 ns	0.19	58	81	3.9	0.034
basis set of		τ_{1-3} = 0.98 ns*	0.34	51	51	2.9	
50 random		τ_{3-3} = 1.01 ns*	0.19	176	−98	4.3	
structures			0.29	173	−83	2.7	
9	0.03	τ_{1-1} = 1.0 ns	0.16	58	81	3.9	0.01
basis set of 50		τ_{1-3} = 0.98 ns*	0.35	51	51	2.9	
random plus 2		τ_{3-3} = 1.01 ns*	0.49	174	−79	2.6	
true structures							
distance	0.03	τ_{1-1} = 1.0 ns	1.0	130	−31	29.2	0.02
constrained		τ_{1-3} = 1.05 ns*					
structure[e]		τ_{3-3} = 0.97 ns*					
2 true		τ_{1-1} = 1.0 ns	0.5	53	59	2.6	
structures		τ_{1-3} = 1.0 ns	0.5	174	−79	2.6	
		τ_{3-3} = 1.0 ns					

[a] Structures were tested at the 50% confidence level. [b] An asterisk indicates that the correlation times were optimized using fixed interproton separations of the isopropyl group. [c] The CHARM-calculated potential energy. [d] *R* factor = $(\sum_i(|I_i|_{obs} - |I_i|_{calc})^2/\sum_i(|I_i|_{obs})^2)^{1/2}$ where $|I_i|_{obs}$ and $|I_i|_{calc}$ are the observed and calculated magnetization intensities, respectively. [e] Structure obtained using distance constraints in CHARMM energy minimization. Distances were obtained by initial cross-relaxation rates according to the independent spin pair approximation.

A. Rotamer Simulations. In order to mimic a realistic analysis, we chose for simulation a halogenated alkane (**1**) for which em-



pirical force field computations suggested multiple, low-energy torsional conformations (rotamers). Systematic, adiabatic searches of the two torsion angles (tor 1 and tor 2) revealed three rotamers of equal stability (within 0.1 kcal/mol). Using two of these rotamers at equal populations, transient NOE timecourse "data" were generated by numerical integration of the Solomon equations with all τ = 1.0 ns. At these correlation times and a 500-MHz Larmor frequency, the molecule exhibits negative NOE's and significant spin diffusion. Next a basis set of 50 trial structures was generated by random torsion angle sampling followed by partial energy minimization using the Boltzman jump routines incorporated in QUANTA;[16] the basis set did not include the two "true" structures used to simulate the "data". This set of data and trial structures poses a rigorous test of the analysis method because most of the NOE interactions are not particularly sensitive to the torsion angles; i.e. the interproton distance matrices for the two conformers are not dramatically different.

Using this basis set, CPA analysis was performed with systematic examination of the influences of (1) random noise, (2) using exact τ values vs fitted τ values, (3) the clustering level, and (4) the presence of the "true" structures. Selected results are listed in Table I along with the results for the structure obtained using distance-constrained minimization.

The gross features of the results presented in Table I demonstrate that, under all conditions, multiple conformations yielded *significantly* better fits to the NOE "data" than a single conformation even after elimination of conformers by the F-test procedure. For example, a single conformer was built using distance-constrained energy minimization with distances based on initial (0.001 s) rates. Not only are the multiple conformation fits significantly better than the fit obtained using a structure constructed from distance constraints, but the structure produced by constrained minimization actually corresponds to an energy maximum on the torsional potential energy surface. Thus, for this simulation the distance constrained structure actually has little probability of existing in solution.

An important concern in fitting solution NOE data is mitigating the effects of bias. In part this is accomplished by using large, randomly generated collections of trial structures that are energetically reasonable. An additional requirement is that the data reduction process samples the conformational space in a homogeneous (unbiased) way. The sampling of structures at the initial data reduction step and subsequent sampling steps is illustrated in Figure 2. Note that the initial cluster distribution is not weighted by the density of structures in a given region of conformer space but, rather, is determined by the overall range of occupied conformational space. Thus, the clustering method samples homogeneously.

In the absence of noise, better fits are obtained with larger clustering levels. This is understood readily, since small clustering levels correspond to severe data reduction, hence, limited degrees of fitting freedom. For the data used in these simulations a total of 80 data points were used (five inversion sites × four cross-relaxation sites × four time points per inversion) and the least-squares fitting routine was found to be stable at a clustering level of 50 (i.e. each trial structure is its own cluster). However, for applications of this method to larger, more flexible molecules it is less likely that the number of data points would exceed the number of trial conformers. Thus, clustering levels should be set to as large a value as the least-squares problem will allow.

Another concern is the convergency of the method; do different starting points yield the same solution? In this simulation, different

(16) QUANTA Program, Copyright Polygen Corporation, 1990.

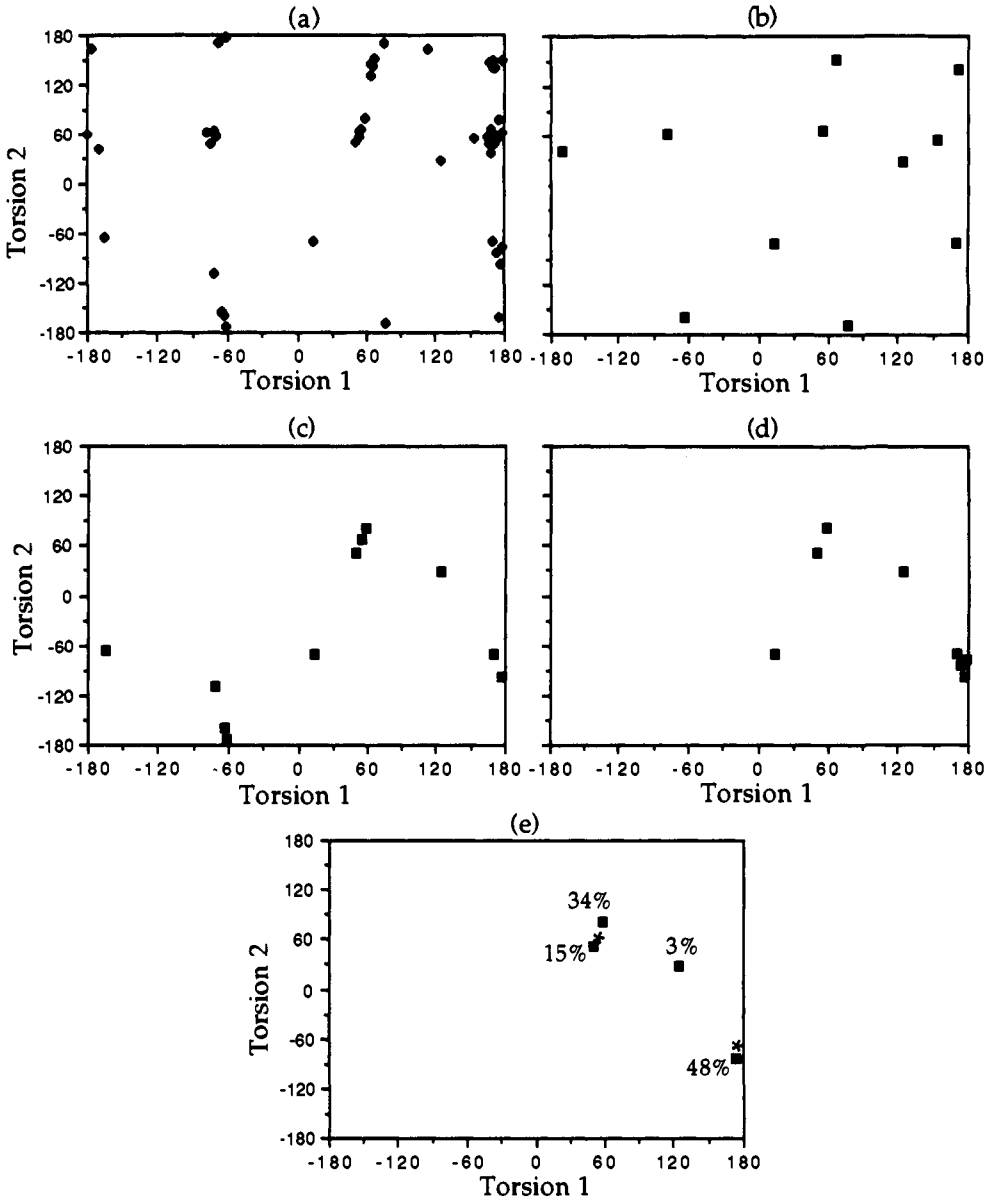


Figure 2. Progress of refinement during CPA analysis of the rotamers of **1**. (a) The original distribution of structures in the basis set. (b) Structures best representing cluster centroids after data reduction of the original basis set into 11 clusters based on r^{-6} distances. (c) Structures resulting from the first refine–expand–cluster cycle. (d) Structures resulting from the last refine–expand–cluster cycle. (e) Best solutions to the multiconformational structure problem after the elimination of insignificant conformers using the F-test procedure. Percentages represent the percent contribution of each structure to the total solution structure and the asterisks mark the two conformers used to generate the data.

starting points are created by using different clustering levels. Although fits obtained with different initial clustering levels have different R factors and slightly different populations of conformers, all starting points yield solutions that clearly are distributed about the two "true" conformers. Thus, we conclude that the method is adequately convergent.

The principal effect of adding random noise is to decrease the number of significant conformers. Due to the presence of noise, the small increase in the goodness-of-fit achieved by adding more conformers is not sufficiently large to warrant the increase in the number of fitting variables. The gross distribution of structures, however, is not highly sensitive to random noise.

As with most optimizations based on comparisons between models and experimentally derived quantities, systematic error is particularly troublesome. A primary source of systematic noise in the NOE experiment can come from errors in the rotational correlation time.[17] For all simulated fits, rotational correlation times were determined by least-squares fitting using a single trial conformer. These values exhibited little deviation (±3%) from the actual values regardless of the trial conformer used in the fitting. Optimizations using a conformer ensemble rather than a single trial conformer gave similar results. However, manually setting larger offsets from the true τ values results in conformer distributions that are skewed significantly from the true distributions. For example, when a τ value of 0.5 ns is used, the method finds just one significantly populated conformer. Thus, well-determined rotational correlation times are required if detailed structural analyses of NOE data are to be carried out, particularly when the rotational correlation times are intermediary between the fast and slow tumbling regimes.

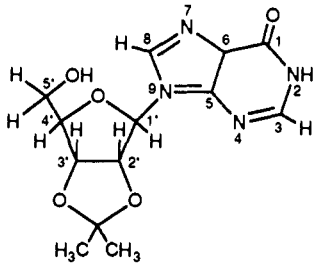
A final question concerning this set of simulations is if the true structures are members of the trial basis set, will the CPA method find them? A series of tests were performed in which the trial basis set contained the 50 randomly generated conformers plus the two "true" structures used to generate the data. In the absence of noise and using the exact τ values, the true solution (i.e. equal population of the two true conformers, only) was reproduced when the clustering level was 12 or higher. With lower clustering levels only one of the true conformers was found. The interpretation

(17) (a) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035. (b) Keepers, J. W.; James, T. L. *J. Magn. Reson.* **1984**, *57*, 404.

of these results is straightforward; when severe data reduction is used, fewer clusters are being used to represent the conformational space. Thus, there is a greater probability that a true structure may be included in a cluster whose centroid is sufficiently far from the true conformation that the cluster is not populated in the initial fitting stages. This emphasizes the need to inspect the influence of clustering level on the quality of the final fit. However, even when one of the true structures was discarded due to low clustering levels, the final solutions clearly were distributed around the true structures.

In the presence of random noise, the method is less sensitive to the difference between the true structures and nearby conformers so that a spread of structures around the true conformers is observed. Finally, in the presence of systematic noise (errors in τ) the true conformers are no longer the best solution and, hence, are not populated.

B. The Conformation of Inosine. Some of the aspects that differentiate the CPA method from other "best-fit" methods are illustrated well by reanalysis of NOE data for the nucleoside derivative 2′,3′-isopropylideneinosine (**2**). In a series of elegant



2

studies presaging the current interest in quantitative conformational analysis by NOE, Noggle et al.[18] used steady state NOE to examine the conformations of 2′,3′-isopropylideneinosine. It is important to note that steady-state measurements do not yield structural information as readily as transient methods. Thus, the NOE data for 2′,3′-isopropylideneinosine may not be sufficient to reveal the actual solution conformations. Rather, the point of performing the analysis is to illustrate the use of the CPA method on real data and to contrast the method with the Noggle analysis.

Assuming that the observed spectrum of inosine represents contributions from both the syn and anti conformers and that the molecule has a correlation time in the fast tumbling limit, Noggle et al. computed the width, population, and position of Gaussian distributions representing each of the two conformers by a least-squares procedure. Their computations encompassed two rate regimes: (1) fast conformer interconversion compared with T_1 and (2) slow conformer interconversion compared with T_1.

On the basis of the results of their simulations, Noggle et al. concluded that 2′,3′-isopropylideneinosine exists in solution as two conformers, approximately 80% syn and 20% anti. Interestingly, their results were not sensitive to either the conformer exchange rates or inclusion of external relaxation processes. However, their results did not exhibit substantial sensitivity to the structures and populations of the conformers used in the measurements. These features suggest that analysis by the CPA method, which does not (at this time) include conformer exchange but does compute conformer populations and structures, is appropriate.

A basis set of structures for the CPA method was created by adiabatic (fully relaxed) torsion driving about the C1′–N9 bond in 10° increments using the CHARM[19] computational package. As shown in Figure 3, two distinct minima were found about the torsion angles of 116° and −24°, approximately corresponding to anti and syn conformers, respectively. This generated a total of 36 trial conformers.

Next the steady-state NOE intensities for all protons were computed by numerical integration of the Solomon equations for

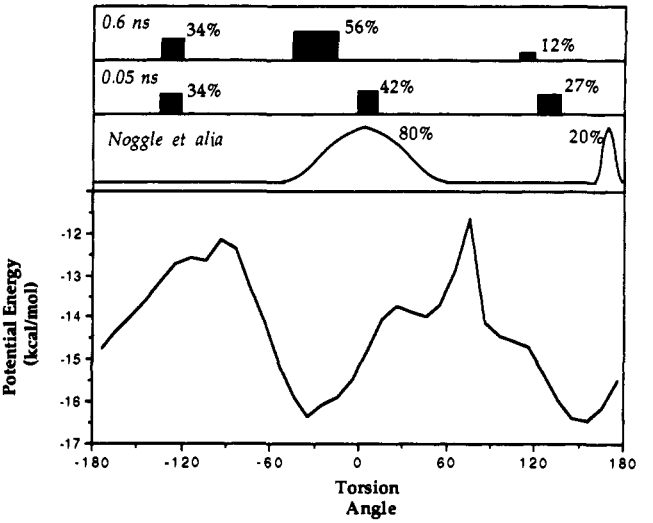


Figure 3. A comparison of CHARMM-computed energies (lowest box), conformer population distributions determined by Noggle et al. (second box), and CPA-computed conformer distributions at 0.05 ns (third box) and 0.6 ns (fourth box) correlation times determined for **2** as a function of torsion angles based on the original steady-state NOE data of Noggle et al.

a 100-MHz field strength at both 10- and 20-s irradiation times of the sites reported by Noggle et al. Because the steady-state NOE depends on the rotational correlation times, we explored the influence of correlation times on the computed NOE intensities over the range of 0.05 ns (fast tumbling) to 0.8 ns in 0.1-ns increments. The best fit to experimental data was obtained at a correlation time of 0.6 ns. At all correlation times no significant difference was seen for 10- and 20-s irradiation periods.

Clustering and refinement was performed using the CPA routines operating with an initial basis set of 36 structures, the set of 12 observed data points used by Noggle et al., and a clustering level of 10. The clustering phase was followed by significance testing using Hamilton's version of the F-test. Sampling of the effects of clustering mode and level on the quality of the fit and on the conformer populations demonstrated little dependence of the results on the starting point.

As shown in Figure 3, the CPA method puts forth three significantly populated conformers in order to obtain the best fit between experimental and computed NOE data. The root-mean-square deviation calculated from observed NOE intensities is 0.029. In contrast, the analysis of Noggle et al. *assumed* two conformer distributions. The populations and distributions of the conformers determined by the CPA technique exhibit some dependence on the rotational correlation time used. The longer correlation time results in a greater population of the syn conformer.

Two difficulties in the analysis of **2** are that the true conformational distributions are not known and the steady-state experiment is relatively poor in distance information compared with kinetic NOE methods. Thus, the results of these analyses serve to contrast two different analysis techniques. Whereas the Noggle treatment assumes, in this case in agreement with force field computations, that two energetic minima exist on the torsional energy surface, the number of conformers that can be utilized by the CPA method is limited by the number of data points only. As compared with the force field results, the conformer distributions suggested by Noggle are slightly offset from the force field minima and the population of the anti conformer is low. The CPA method obtains the best fit to the experimental steady-state NOE data using populations and torsions for the syn conformer which agree with force field computations. For the anti conformer, however, the geometries of the two anti conformers straddle either side of the force field minimum for the anti conformer.

If one were to consider the force field computations to be most representative of the true solution energetics, a hazardous presumption, one would conclude that the penalty for the confor-

(18) Schirmer, R. E.; Davis, J. P.; Noggle, J. H.; Hart, P. A. *J. Am. Chem. Soc.* **1972**, *94*, 2561.

(19) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187–217.

mational flexibility of the CPA method is that the best fit may be obtained with conformers that do not represent energetic minima. By constraining their search to just two Gaussian distributions, Noggle et al. obtain better matches of torsion angles between force field and NOE results but poorer populations.

IV. Summary

Conformer population analysis is a method for the elucidation of solution structures from NOE data that is an alternative to building structures from interproton distance constraint matrices. Compared with distance-constrained methods, CPA has the advantage of permitting multiple conformations in the fitting solution and, hence, is capable of removing biases inherent in the assumption of a single structure. Another advantage is that the method directly works with NOE intensities, an important distinction because NOE experiments produce intensity data, not distances.

The most significant disadvantage of the CPA technique is that it does not direct the construction of a molecular structure; a group of predefined conformations is required. Although molecular modeling techniques are capable of generating large sets of energetically reasonable conformations, it is obvious that the structures produced will reflect the inaccuracies of the force fields utilized (this is true of constrained methods as well, although the constraints commonly are less rigid). For large molecules (e.g. protons) the CPA method may be impractical because the generation of adequate trial structures using internal coordinate randomization techniques alone is unreasonable. However, other techniques, such as homology modeling, could be used to generate trial structures for proteins.

Because large sets of trial structures are required to sample conformational space thoroughly, data reduction techniques are needed. The clustering algorithms used in CPA accomplish data reduction by sifting through large data sets and identifying a smaller subset of representative structures. Refinement occurs by identifying clusters that are promising and focusing the fitting procedures on the members of those clusters. Convergence in this process is not assured, but in practice the NOE is sufficiently sensitive to structure that the method readily compresses its search to a small number of conformers. A final phase of significance testing produces the most parsimonious solution justified by the data itself.

Perhaps the most useful application of a multiconformational technique is in concert with molecular construction methods based on the one cross-relaxation:one distance approximation. A reasonable procedure may be to use the structure built by a molecular construction technique as a starting point for conformation searching by modeling techniques.

In the current implementation, the CPA technique is deficient in not considering the influence of conformer exchange and rapid internal motions, such as methyl group rotation. However, these limitations are imposed principally by the numerical integration schemes employed; a subsequent version will use more versatile matrix methods and will include the influence of conformer exchange and internal rotation kinetics. Future work will focus on the inclusion of symmetry and torsional information into the fitting procedure and on the experimental examination of solution structures of molecules having multiple conformations.

Acknowledgment. We thank Dr. Art Pardi for his help and encouragement. Partial support of this work was supplied through a generous gift from the Monsanto Company.

Registry No. 1, 136587-47-8; **2**, 2140-11-6.

Gas-Phase Formation of Four Isomeric $C_4H_4^{\cdot+}$ Ions. Ionic Isomer Quantitation with Neutralization–Reionization Mass Spectrometry

Mei-Yi Zhang, Barry K. Carpenter, and Fred W. McLafferty*

Contribution from the Department of Chemistry, Baker Laboratory, Cornell University, Ithaca, New York 14853-1301. *Received January 28, 1991*

Abstract: Four $C_4H_4^{\cdot+}$ isomers yield characteristic mass spectra by neutralization, dissociation of these neutrals, and product reionization to both positive ($^+NR^+$ spectra) and negative ($^+NR^-$ spectra) ions. Such measurements show that a wide variety of unsaturated/cyclic hydrocarbon ions, and of ionized 2,4,6-cycloheptatrien-1-one, yield $C_4H_4^{\cdot+}$ of the composition ~68% methylenecyclopropene ($\mathbf{c}^{\cdot+}$) and ~32% vinylacetylene ($\mathbf{a}^{\cdot+}$). This appears to result from the dissociation of a common intermediate, the excited benzene ion; for the $C_6H_8^{\cdot+}$ precursors, this is consistent with an early proposal of Franklin. This most stable $\mathbf{c}^{\cdot+}$ isomer can also be formed in even higher purity from ionized 6,7-benzo-3-methylenetricyclo[3.2.2.0^{2,4}]nona-6,8-diene, while the isomer $\mathbf{a}^{\cdot+}$ can also be formed by loss of H_2O from ionized 3-butyn-1-ol. Butatriene ions ($\mathbf{b}^{\cdot+}$) are formed by Br_2 loss from ionized 1,4-dibromo-2-butyne, but in addition the $C_4H_4^{\cdot+}$ products contain ~44% cyclobutadiene ions ($\mathbf{d}^{\cdot+}$). The $\mathbf{d}^{\cdot+}$ ions also result from dissociation of ionized 7,8-benzotricyclo[4.2.2.0^{2,5}]deca-3,7,9-triene and 3,4-diiodocyclobutene, and $\mathbf{d}^{\cdot+}$ is the main $C_4H_4^{\cdot+}$ product from 1,2- and 1,4-benzoquinone; with both, $\mathbf{d}^{\cdot+}$ is accompanied by ~14% $\mathbf{a}^{\cdot+}$, suggesting a common cyclopentadienone$^{\cdot+}$ intermediate. For each of 13 precursors the values for the $^+NR^+$ and $^+NR^-$ measurements differ from their average by 1.8 ± 2.6% (absolute), consistent with relatively high accuracy for NR isomeric analysis of gaseous ions.

A wide variety of studies over the past two decades[1-8] have sought the isomeric identity of the $C_4H_4^{\cdot+}$ radical cations found as a common product of the unimolecular dissociation of ionized unsaturated/cyclic hydrocarbons such as benzene.[5a,9] Although

(1) (a) Franklin, J. L.; Carroll, S. R. *J. Am. Chem. Soc.* **1969**, *91*, 6564-6569. (b) Rosenstock, H. M.; McCulloh, K. E.; Lossing, F. P. *Int. J. Mass Spectrom. Ion Phys.* **1977**, *25*, 327-341.

(2) Baer, T.; Willet, G. D.; Smith, D.; Phillips, J. S. *J. Chem. Phys.* **1979**, *70*, 4076-4085.

(3) Lifshitz, C.; Gibson, D.; Levsen, K.; Dotan, I. *Int. J. Mass Spectrom. Ion Phys.* **1981**, *40*, 157-165.

(4) Ausloos, P. *J. Am. Chem. Soc.* **1981**, *103*, 3931-3932.

(5) (a) Borchers, F.; Levsen, K. *Org. Mass Spectrom.* **1975**, *10*, 584-594. (b) Levsen, K.; Hilt, E. *Liebig's Ann. Chem.* **1976**, 257-268. (c) Wagner-Redeker, W.; Illies, A. J.; Kemper, P. R.; Bowers, M. T. *J. Am. Chem. Soc.* **1983**, *105*, 5719-5724. (d) Staley, S. W.; Norden, T. D. *Ibid.* **1989**, *111*, 445-449. (e) Ray, J. C., Jr.; Danis, P. O.; McLafferty, F. W.; Carpenter, B. K. *Ibid.* **1987**, *109*, 4408-4409. (f) Turecek, F.; Hajicek, J.; Maquin, F.; Stahl, D. *New J. Chem.* **1987**, *11*, 793-796.

0002-7863/91/1513-9499$02.50/0