How is an NMR structure best defined? An analysis of molecular dynamics distance-based approaches

David A. Pearlman

Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, MA 02139-4211, U.S.A.

Received 20 May 1993
Accepted 29 July 1993

Keywords: Refinement methods; Time-averaged restraints; FK506; Molecular dynamics

SUMMARY

Model studies on the macrocyclic immunosuppressive agent FK506 challenge traditional approaches to defining a structure from data collected during a 2D NMR experiment. A variety of joint molecular dynamics/NMR-distance refinement methodologies are characterized. From the results it is clear that the traditional presentation of an NMR structure as a single representative minimized conformation or as a fairly tight envelope of conformers best meeting the imposed restraints can be misleading; a greater emphasis is required on dynamics and on the fact that an NMR structure represents a time average.

INTRODUCTION

At the most rigorous level of detail, almost all chemical phenomena are extraordinarily complex. As a result, our understanding of such phenomena is generally couched in simplifications and models. This can be quite useful if the model is a good one. But it is critical that one be aware of the assumptions and failings of the model.

Sometimes a model becomes so familiar and intuitive that it blinds us to a better scientific understanding of the underlying phenomena. In many ways, the crystallographic (Blundell and Johnson, 1976) and multidimensional Nuclear Magnetic Resonance (NMR) (Wüthrich, 1986) techniques which are now used to generate the majority of structure-based molecular information have become overly reliant on such convenient-yet-misleading models. In both techniques, the result is typically presented as a single conformation, which is taken to represent the 'structure' of the molecule being studied. But, in fact, most molecules (and nearly all macromolecules) are very flexible (Petsko and Ringe, 1984; Elber and Karplus, 1987; Kessler et al., 1988; Smith, 1991), and the presentation of a single snapshot as the structure belies the true behavior. A detailed understanding of conformational variability is critical to generating a model which can be reliably used to interpret other experiments and as a nucleus for subsequent detailed studies. Structural chemists are certainly aware of this conformational variability, and to some extent it is reflected in the...
B-factors determined during crystallographic refinement (Blundell and Johnson, 1976) and in the envelope of structures typically determined during NMR refinement (Clore and Gronenborn, 1989; Wagner et al., 1992). Yet in most discussions, and in most studies incorporating these structures (e.g., molecular modeling) the concept of a single ‘structure’ has been too appealing and comfortable to dismiss.

Here we probe the failings of this traditional description, and then determine how the structure from a multidimensional NMR experiment can better be refined and described. NMR structure determinations commonly consist of three phases (Wüthrich, 1990; James and Basus, 1991). In the first phase the data, a series of interatomic spin relaxation measurements, is collected. In the second phase, the relaxation rates are converted into interatomic distance estimates (‘NOE distances’). In the third phase the distances are used to determine a ‘structure’. It is this third phase that will be the focus of this paper.

BACKGROUND

Two methods are used to generate a 3D structure from a set of interatomic distances. The first is Distance Geometry (DG) (Crippen and Havel, 1988; Havel, 1991), which is very efficient at producing random structures consistent with a distance matrix. The second is molecular mechanics (minimization, Molecular Dynamics (MD), Monte Carlo, Simulated Annealing) (Burkert and Allinger, 1982; Kaptein et al., 1988; Brunger and Karplus, 1991). Molecular mechanics techniques are good at producing low-energy structures which balance the information contained in the distance set with optimization of the potential energy surface representing the molecule, but are inefficient at producing an initial guess from the distance information. Optimally, DG and molecular mechanics are used together (Liu et al., 1992). In this case, the MD phase is pivotal in refining the final ‘structure’. A variety of MD simulations are run, subject to restraints based on the NMR-derived distances, and the consensus of these runs is reported as the structure. MD is an integration of Newton’s equation:

\[
F = -\frac{dV}{dx} = ma = m \frac{dx}{dt}^2
\]

(1)

giving the molecular conformation over time \(x(t)\). The integration is carried out for a specified non-zero temperature, allowing the system to sample conformational space. Here \(F\) is the force, \(V\) is the potential energy, \(x\) is a Cartesian coordinate, \(m\) is an atomic mass, and \(a\) is the acceleration. The complexity of \(F\) means we have to integrate this equation step-wise (in steps of 0.5–2 fs). Computer resource considerations therefore limit us to total simulations of on the order of picoseconds or nanoseconds.

In conventional MD-based NMR refinement (Nilges et al., 1988), the set of NMR-derived distances \(r_{\text{NOE}}\) is imposed on the model structure using a harmonic potential term of the form:

\[
V_{\text{NOE}} = K_i (r_{\text{model}}(t) - r_i)^2 \quad r_{\text{model}}(t) < r_i
\]

(2a)

\[
V_{\text{NOE}} = 0 \quad r_i \leq r_{\text{model}}(t) \leq r_u
\]

(2b)

\[
V_{\text{NOE}} = K_u (r_{\text{model}}(t) - r_u)^2 \quad r_u < r_{\text{model}}(t)
\]

(2c)