Journal of Biomolecular NMR, 8 (1996) 49–66
ESCOM

J-Bio NMR 368

FINGAR: A new genetic algorithm-based method for fitting NMR data

David A. Pearlman

Vertex Pharmaceuticals Inc., 130 Waverly Street, Cambridge, MA 02139-4242, U.S.A.

Received 27 February 1996
Accepted 24 May 1996

Keywords: Refinement; Genetic algorithm; Adenosine; FK506

Summary

A new NMR refinement method, FINGAR (FIt NMR using a Genetic AlgoRithm), has been developed, which allows one to determine a weighted set of structures that best fits measured NMR-derived data. This method shows appreciable advantages over commonly used refinement methods. FINGAR generates an ensemble of conformations whose average reproduces the experimental NMR-derived restraints. In addition, a statistical importance weight is assigned to each of the conformations in the ensemble. As a result, one is not limited to simply presenting an envelope of sampled conformers. Instead, one can subsequently focus on a select few conformers of high weight. This is critical, because many structural analyses depend on using discrete conformations, not simply averages or ensembles. The genetic algorithm used by FINGAR allows one to simultaneously and reliably fit against many restraints, and to generate solutions which include as many conformations with non-zero weights as are necessary to generate the best fit. An added benefit of FINGAR is that because the time-consuming step in this method needs only to be performed once, in the beginning of the first run, numerous FINGAR simulations can be performed rapidly.

Introduction

Multi-dimensional NMR as a technique for determining three-dimensional atomic structures is now well-established. Although there are variations, the approach employed most frequently for NMR structural studies consists of three parts (Wüthrich, 1990): (i) collect the NMR data; (ii) convert the NMR/NOE data into a set of interatomic distances and combine these with any measured vicinal ^{3}J-coupling constants; (iii) use these distances and coupling constants to generate a 3D set of atomic coordinates (refinement). In many ways, part three of this procedure – converting NMR-derived parameters into useful and reliable three-dimensional structures – remains most problematic. This is, in large part, because these parameters reflect not a single conformation, but an ensemble of conformers. Both generating and representing this ensemble in a useful form has proven to be difficult (Pearlman, 1994a; Van Gunsteren et al., 1994).

The most commonly applied method for NMR refinement ('standard refinement') has been to use a sampling procedure – typically either distance geometry (DG) (Crippen and Havel, 1988) or molecular dynamics (MD) (Burkert and Allinger, 1982; Brunger and Karplus, 1991) or a combination of these – to generate molecular conformations that are consistent with the set of experimentally derived distances and J-coupling constants. In both approaches, restraints are fulfilled in a static sense. That is, any particular single MD or DG simulation is an attempt to satisfy as well as possible the restraints with a single structure. MD-based NMR refinement is performed (Clore et al., 1986; Nilges et al., 1988) by adding a set of restraint terms to the standard potential energy function that dictates the MD trajectory:

$$E_{total} = E_{potential} + \sum_{restraints} K_i(r_i(t) - rNMR_i)^2 \qquad (1)$$

where $r_i(t)$ is the instantaneous value of restraint r at any time 't' during the MD simulation, $rNMR_i$ is the NMR-derived target restraint value, and K_i is a force constant. Frequently, a flat region is defined to account for errors in the NMR-derived restraints. In this case, the restraint term is set to 0 for any value $r_i(t)$ within a given tolerance of $rNMR_i$, and is quadratically restrained to $rNMR_i$, as in Eq. 1, outside of this range.

During an MD simulation, kinetic energy allows the

system to pass over local energy minima and (hopefully) find a reasonably low value of the E_{total} energy function. Most commonly, the temperature of the system is first raised, to facilitate passing over very large energy barriers, and then lowered to a very small value, to leach potential and kinetic energy out of the system and focus on a local energy minimum ('simulated annealing'). In many cases, the energy of the system is subsequently minimized. By this approach, a single conformation that (hopefully) satisfies the added restraints well is determined. This process is usually repeated many times, with different sets of initial randomly assigned velocities or different starting conformations, to produce an envelope of possible solutions to the problem. The simulation is rerun multiple times for two reasons: (i) even with high-temperature simulated annealing, these searches frequently have difficulty finding the global minimum; repeating the simulation helps to guard against getting trapped in a high-energy local minimum; and (ii) there may be several conformations that fulfill the restraints equally well (or give equivalently low values for the energy function), due to underdetermination of the problem. In such a case, we want to generate a representative envelope of equivalent solutions.

With DG, the process is similar: the simulation is run numerous times, with different random number seeds, in an attempt to determine structures which fulfill the NOE-derived restraints (given as ranges of acceptable distances). DG is extremely efficient at generating conformations that adhere to the specified restraints, but it does not take into account the potential energy, except in a very crude fashion. Consequently, the structures generated may satisfy the restraints well, but be of relatively high energy. For this reason, energy-based refinement (MD or minimization) is frequently performed on the structures generated in a DG run. Alternatively, the ensemble of DG structures can be analyzed directly to provide an idea of the variance in conformation that is defined principally by the restraints, covalent geometry, and simple steric repulsions. However, correlation between the distribution of conformers in the DG ensemble and the true conformations in solution is often weak (Yang and Havel, 1993).

Regardless of whether the DG structures are subsequently energy refined or not, the end result from any standard MD or DG refinement run is a *single* conformation which fulfills the restraint and energy (if any) criteria. The ensemble of structures produced from multiple runs – frequently presented superimposed as an envelope of the possible structure – does not change this fact. The envelope merely illustrates a series of conformations which, *individually*, satisfy the NMR data.

The problem with this approach is that the distances and coupling constants one derives from an NMR experiment represent an averaging over the ensemble of conformers that are sampled while the experiment is being run. For example, from a NOE experiment one derives not the interproton distance r, but rather $\langle r^{-6}\rangle^{-1/6}$, which represents the average over conformations sampled while data are being collected (Wagner and Wüthrich 1979; Tropp, 1980). As a result, it is incorrect to force any single conformation to satisfy the restraints. To do so risks generating an envelope of conformers that is too tight, and individual conformations that are high in energy and distorted. These problems have afflicted many of the NMR-derived structures that have been reported (Torda et al., 1990; Pearlman and Kollman, 1991; Van Gunsteren et al., 1994).

Significant improvement in the quality of the refined results can be achieved by incorporating ensemble averaging into the refinement procedure: instead of requiring that a single conformer satisfies the experimental restraint information, we require that the distances/torsions *averaged* over a set of conformations satisfy the restraints. For MD-based refinement, this leads to the method of 'time-averaged refinement' (Torda et al., 1989,1993), in which one runs an MD simulation subject to restraints of the form:

$$E_{distance} = \sum_{\text{measured NOEs}} K_r \left(\langle r_{MD}^{-6}\rangle^{-1/6} - r_{NMR}\right)^2 \qquad (2)$$

$$E_J = \sum_{\text{measured J}} K_J \left(\langle J_{MD}\rangle - J_{NMR}\right)^2 \qquad (3)$$

where $\langle r_{MD}^{-6}\rangle^{-1/6}$ is the reciprocal-weighted average of the distance over the MD trajectory, r_{NMR} is the distance derived from the experimentally measured NOE data, $\langle J_{MD}\rangle$ is the weighted average of the J-coupling constant over the MD trajectory, and J_{NMR} is the experimental J-coupling value. The results are presented as both a time-averaged structure, and as an envelope of sampled conformations.

Time-averaged refinement has been compared in detail to standard refinement (Torda et al., 1990; Pearlman and Kollman, 1991; Schmitz et al., 1993; Pearlman, 1994a,b; Gonzalez et al., 1995), and the results argue quite convincingly that the time-averaged protocol is superior to the traditional method. In particular, compared to standard refinement, time-averaged refinement does a better job at reproducing the 'true' atomic fluctuations and averaged conformation, and the accuracy of the results is less sensitive to refinement variables like restraint weights or estimated restraint errors (Pearlman and Kollman, 1991; Pearlman, 1994a,b).

But despite these advantages to standard refinement, time-averaged refinement is not ideal. In particular, any restraint applied during the MD trajectory can affect the motions of the molecule being refined, generally in an unrealistic fashion (Scheek et al., 1995). For time-averaged refinement, this is a particular problem for the slower motions in the molecule (e.g. those with a characteristic time of greater than 10–20 ps). In addition, because the averaged values used in time-averaged refinement change

very slowly, and because time-averaged refinement is typically implemented using nonconservative forces, the atomic fluctuations predicted during such refinement tend to be too large, increasingly so with larger restraint force constants (Pearlman, 1994a; Nanzer et al., 1995). Time-averaged refinement also has problems reproducing experimentally derived restraints that represent the average of two or more conformations separated by moderate or larger energy barriers (Pearlman and Kollman, 1991; Pearlman, 1994b; Fennen et al., 1995). Finally, while the results of time-averaged refinement are considerably better than those from standard refinement, the averaged structure and envelope of sampled conformers are not necessarily ideal for use in subsequent molecular modeling applications. Better would be a set of important, dominant, conformers. Although the simultaneous simulation of several molecular replicates has been proposed as a means to improve sampling about large energy barriers during time-averaged refinement (Fennen et al., 1995), this modification does not solve the other problems with the method, and introduces its own set of limitations.

So, although time-averaged MD demonstrates many advantages relative to standard refinement, the need remains for a better method of NMR refinement. Ideal characteristics of such a method would include the ability to: (i) generate a solution from an ensemble of structures free of restraint bias; (ii) specify a set of distinct important conformers and weight their relative importance; (iii) incorporate appropriately weighted averaging in calculating the internal values that are fit to the NMR-derived restraints; (iv) derive a solution relatively quickly; and (v) incorporate a robust algorithm for deriving the refined solution.

In response to this need, we have developed the new method FINGAR (short for FIt NMR data using a Genetic AlgoRithm). In this method, unrestrained and weakly restrained MD simulations are run to sample conformational space. Snapshots are stored periodically during the MD trajectories. Subsequently, these snapshots are clustered to reduce the number of structures down to a smaller group of distinct 'basis structures'. Finally, a genetic algorithm (GA) is applied (Goldberg, 1989), where the relative weights of the basis structures are allowed to change in order to determine a weighted ensemble that balances reproducing the experimentally derived restraints with a low average potential energy. This method has been tested against data for several model systems, and it appears to be able to both generate refined models in excellent agreement with experiment, and avoid the drawbacks associated with time-averaged refinement.

FINGAR is not the first program to attempt to fit NMR data to a discrete packet of structures weighted by their relative contributions (Kim and Prestegard, 1990; Bruschweiler et al., 1991; Landis and Allured, 1991; Nikiforovich et al., 1993; Yang and Havel, 1993; Landis et al., 1995; Ulyanov et al., 1995). But many earlier attempts to do so have relied on local optimization procedures (Press et al., 1989) to determine an optimized fit for the multivariate problem. Such procedures suffer from the local-minimum problem (the refined answer is a relative minimum close to the starting guess, rather than the absolute minimum), and this problem becomes more severe as the number of fitted variables increases. In an attempt to circumvent this problem, some studies have performed exhaustive searches on small numbers of fitted variables (Bruschweiler et al., 1991; Landis et al., 1995). That is, all combinations of a specified, small number of structures from the basis set are individually best fit to the NMR data, and the subset or coefficients yielding the best agreement with the NMR data are taken as the solution. The problem with this approach is the resulting combinatorial explosion (e.g. combinations of 100 basis structures taken 4 at a time yields nearly 4 million sets; combinations of 100 basis structures taken 5 at a time yields 75 million sets), which typically limits one to fits of less than structures. Alternately, one can attempt to severely reduce the number of structures that must be fit using clustering methods (Kim and Prestegard, 1990; Landis and Allured, 1991; Landis et al., 1995), thereby reducing the unknowns/equation ratio to a point where direct methods like linear least squares (Lawson and Hanson, 1974) can be used to determine a solution directly. But this approach runs the real risk of omitting important structures before the fitting occurs. Another approach has been to increase the effective number of observables by performing refinement against raw 'pixelated' NMR data, rather than derived quantities such as integrated intensities or distances (Yang and Havel, 1993). This also results in an overdetermined problem that can be solved directly by linear least-squares methods, although one may encounter difficulties because many of the data used will reflect appreciable random and/or systematic errors. Of these earlier studies, only one (Ulyanov et al., 1995) utilized a global optimization procedure. In this study, quadratic programming was used in an attempt to minimize a relaxation-rate-based residual. Of the methods published, this one seems the most promising for fitting integrated NOE intensities, NOE-derived distances or J-coupling data to realistic numbers of potential structures. In general, local optimization methods are suboptimal or impractical if the best fit requires a moderate number of non-zero-weighted basis structures. As will be shown, the power of the GA allows one to reliably fit dozens or even hundreds of structural weights simultaneously, provided one has enough experimental data to define a fit.

Finally, we note that the FINGAR method, wherein weights are assigned to members of an ensemble of static structures, differs significantly from another approach recently described for determining the relative populations of conformers (Bonvin and Brunger, 1995; Kemmink and

Scheek, 1995). In this alternate approach, MD sampling (or annealing) is run in parallel on several pre-chosen (and presumably noninterconverting) conformations of the same molecule. Weights are assigned to the various conformers, and the averaged values of restrained variables over the weighted ensemble of conformations is used in the penalty functions. Separate MD simulations are run with a variety of different weight distributions in order to determine the optimal set, and cross-validation (Brunger et al., 1993) can be used to avoid overfitting the problem. However, this method can be quite time-consuming, since each change in the number of members of the ensemble or their weighting requires another MD simulation. Because of this, the number of conformations that can practically be fit is relatively small. In addition, it was recently shown that NOE distances may not contain enough information to reliably assess the relative weights of the conformations when refinement is performed in this manner (Bonvin and Brunger, 1996). This is due to the fact that the conformations can change as the relative weights change. In contrast, because the basis-set structures do not change, it is expected that FINGAR will not suffer this drawback, provided the basis set is reasonably defined and there are sufficient restraint data to define the fit.

Methods

To run FINGAR, one must first obtain a set of experimental distance/J-coupling restraints. Next, a set of structures which represent the 'basis set' for the calculation must be derived. The basis set is a set of structures which should include representatives of all the important conformations that contributed to the experimental ensemble from which the restraints were derived. FINGAR will ultimately attach weights to all the structures in the basis set, and weights of zero can be assigned, so it is not a problem if structures with negligible contributions to the experimental ensemble are also included in the basis set.

The capabilities of FINGAR have been investigated using two types of test simulations. In the first type, refinement is performed against a set of restraints which are generated from a randomly assigned set of weights and a fabricated set of basis sets. Because the restraints are derived entirely from the weights and basis sets, the system is self-consistent, and the restraints can be perfectly fit if we correctly reproduce the set of weights used to generate them. We call this the 'exact test'. In the second type of test simulation, a set of MD-generated basis-set structures are fit to a set of restraints derived from an unrestrained MD trajectory on the same molecule. In this type of simulation, we can only approximate the solution, since our restraint set was not derived directly from the basis set. This test more directly approximates how FINGAR will be used for refinement of actual NMR data. We call this the 'model test'.

For the 'exact test', we first generated a series of basis sets. A basis set is defined by the values of the restrained variables for that set. Here, values were randomly assigned from the interval [2.5,5.5] Å to represent restrained distances. Between 50 and 1000 restraint distances were specified for each basis set, depending on the run. A total of between 50 and 1000 basis sets were defined, also depending on the run. A weight was then assigned to each of the basis sets, choosing random weights on the interval [0,1] for approximately 90–95% of the structures, and [0,10] for the remaining 5–10%. The weights were then normalized to total 1.0. In this way, we approximated a real situation, where the ensemble will be dominated by a subset of the structures in the basis set. Finally, potential energies were assigned to each basis set, based on the Boltzmann factor that would result in the assigned weight. In principle, if FINGAR is successful, it will be able to exactly reproduce the weight vector used to generate the data.

For the 'model tests', two systems were examined: the 126-atom macrocycle FK506 (Schreiber, 1991) with 66 restraints representing the 66 NOE intra-FK506 distances that were determined experimentally for an FK506–FKBP complex (Lepre et al., 1992); and the nucleoside (ribo)-adenosine (Saenger, 1984), with 5 J-coupling restraints representing the five torsion angles whose J-coupling values can, in principle, be measured. These two systems were chosen because, for each, both standard and time-averaged refinement procedures have previously been applied and evaluated in detail (Pearlman, 1994a,b), and because they test two different facets of the refinement. The first one tests the ability of the method to refine against a set of distance restraints. The second one tests the ability of the method to refine against torsional (J-coupling) restraints, in a case where the J-coupling values represent the average of two distinct rotational conformers.

Each 'model test' requires that a basis set of conformations be generated for the system of interest. The method chosen for generating the basis set was to run MD simulations on the molecule to be refined, periodically storing a snapshot of the molecule. While this could be done in any number of ways, and in the general case one would probably wish to run several MD simulations starting from different starting configurations, for the purposes of the studies here it was sufficient to run starting from a single known conformation. Structures generated during MD at 300 K are preferable to those from minimization or another static optimization procedure, because they include the small-but-sometimes-significant internal distortions characteristic of a molecule at room temperature. All MD calculations were run using AMBER/SANDER, v. 4.1 (Pearlman et al., 1995). All MD simulations were run at 300 K, in vacuo, using a 1-fs timestep. Force-field parameters were as described previ-

ously (Pearlman, 1994a,b). In each case, two simulations were run. The first one was standard MD with no restraints. The second one was identical to the first, except that time-averaged restraints were imposed with a small force constant of 1 kcal/mol Å^2 and an exponential damping factor $\tau = 20$ ps (Torda et al., 1989). In general, this second run will be desirable, because entirely free MD may not fully sample the conformational space implied by the experimental restraints. A run using low-weight time-averaged restraints will help ensure that certain regions of conformational space in the vicinity of those required to satisfy the restraints will be sampled. Note that because we always include structures from an unrestrained run, and because inclusion of restraints always forces the potential energy to increase, at least slightly, relative to the same conformation without restraints, conformations from the restrained trajectory are only used when they are necessary to better fit the NMR data and also do not appear in the unrestrained run. In principle, one could assure good sampling by running several trajectories starting from a variety of structures generated using another form of conformational sampling, such as DG or torsional driving. In this case, one might be able to omit the time-averaged simulation altogether. A prerequisite for success with only unrestrained trajectories is that no conformers important for the NMR fit correspond to calculated potential energies which are so high that these conformers would not be sampled during any standard temperature MD run. A critical advantage to FINGAR is that it is not necessary that we sample all important conformers in their appropriate ratios during the sampling procedures. All that is necessary is that we sample each of them at least once. The appropriate weights will be assigned later during the fitting phase. For the FK506 case, both the unrestrained and restrained simulations were run for 1.0 ns, archiving the coordinates every 0.6 ps. For the adenosine case, the unrestrained trajectory was run for 1.5 ns, archiving every 0.5 ps, while the restrained trajectory was run for 0.5 ns, archiving every 0.5 ps.

The 3333–4000 snapshot structures archived for each of the systems were then clustered to provide a manageable set of basis structures. A divisive hierarchical clustering procedure was used (Massart and Kaufman, 1983), in which each existing cluster (initially the entire population is considered one cluster) is split into two until the maximum distance between any two members of any cluster (the 'diameter') is less than a specified threshold. The similarity index used in the clustering was the 'distance matrix error', D_{ab}, which is given as (Havel, 1990):

$$D_{ab} = \left(\frac{2}{N(N-1)} \sum_{i<j} \left(d_{ij}^{a} - d_{ij}^{b} \right)^2 \right)^{1/2} \quad (4)$$

where N is the number of heavy atoms in the structure, the sum is over all pairs of heavy atoms, d_{ij} is the distance between atoms i and j, and a and b are the two conformations being compared. Results from the 'exact test' (see below) indicated that, for a set of 50 restraints – typical of the number determined for a small molecule – FINGAR could reasonably determine proper weights for a basis set of at least 100 structures. A basis set of 100 structures also appeared sufficient to provide a reasonably thorough sampling of the conformational spaces of the molecules examined herein. Thus, for clustering, the maximum cluster diameter was set in order to generate approximately 100 clusters. For each cluster, the member with the lowest potential energy was chosen as the representative of that cluster. For each cluster representative, the values of all the restrained internal coordinates and the potential energy (determined using the same force field used to perform the MD) were calculated and stored to a file, INTVAL.

For the 'model-test' FINGAR runs, input consisted of the INTVAL file, a file consisting of the restraint definitions, and a file which contains various run control options. (For 'exact-test' runs, only the file of run control options is required.) FINGAR uses a genetic algorithm to optimize the following fitness function:

$$E_{tot} = K_{pot} \langle E_{pot} \rangle + K_{bond} E_{bond} + K_J E_J + K_{\sigma E} E_{\sigma E} \quad (5)$$

Each term in this expression represents the weighted ensemble average over the basis-set structures. The GA optimization changes the relative weights of the structures in the basis set, w(i). The w(i) values are normalized to total 1.0 before the fitness function is evaluated:

$$\langle E_{pot} \rangle = \sum_{NBASIS} w(i)\, E_{pot}(i) \quad (6)$$

$$E_{bond} = \sum_{NREST} (\langle r(n) \rangle - r_{NMR}(n))^2 \quad (7)$$

$$E_J = \sum_{NREST} (\langle J(n) \rangle - J_{NMR}(n))^2 \quad (8)$$

$$E_{\sigma E} = (\sigma(E_{pot}) - \sigma(E_{pot,0}))^2 \quad (9)$$

where $E_{pot}(i)$ is the potential energy of basis set i, as calculated by AMBER/SANDER, NBASIS is the number of basis sets, NREST is the number of NMR-defined restraints, and $\langle r(n) \rangle$ is the weighted average value of restraint n over all basis sets i:

$$\langle r(n) \rangle = \left(\sum_{NBASIS} w(i)\, r(i,n)^{-N} \right)^{-1/N} \quad (10)$$

where r(i,n) is the value of distance restraint n for basis set i, and N is typically 6. The reciprocal weighting reflects the relationship between the measured NOE-derived distance $r_{NMR}(n)$ and the experimental ensemble of dis-

tances contributing to $r_{NMR}(n)$ (Wagner and Wüthrich, 1979; Tropp, 1980). <J(n)> is the weighted average value of J-coupling restraint n over all basis sets i:

$$<J(n)> = \sum_{NBASIS} w(i)\, J(i,n) \qquad (11)$$

where J(i,n) is the value of the nth restrained J for basis set i. J(i,n) can be related to the underlying torsion angle τ through the Karplus relationship (Karplus, 1959):

$$J = A\cos^2(\tau) + B\cos(\tau) + C \qquad (12)$$

For adenosine, A = 10.2, B = –0.8 and C = 0.0 (Davies, 1978). Finally, the $E_{\sigma E}$ term is included to ensure that the standard deviation in the value of E_{pot}, $\sigma(E_{pot})$, is reasonable. Typically, a value centered on the range of 2–4 RT is used (Atkins, 1990). FINGAR optionally allows one to specify that $E_{\sigma E}$ is 0 for a range $\sigma(E_{pot,0}) \pm \delta\sigma(E_{pot,0})$ and non-zero outside that range.

Unless noted otherwise, for the simulations reported here, $K_{pot} = 1.0$, $K_{bond} = 200.0$, $K_J = 20.0$, $K_{\sigma E} = 1.0$ and $\sigma(E_{pot,0})$ is set to 3.5 ± 1.5 RT, where R is the gas constant and T is 300 K. For the 'exact-test' simulations, $\sigma(E_{pot,0})$ is set to the actual value, as determined from the true weight vector and assigned potential energies.

For the GA (Goldberg, 1989), a 'population' of IPOP solutions is randomly generated and then allowed to change in search of an optimal value of the fitness function. Each member of the population is a 'weight vector', of length NBASIS, consisting of a set of weights {w(i), i = 1, NBASIS}. Each of the weights w(i) for a particular member can be termed a 'codon'. The weight vector for each member of the population can be different, and the member of the population with the weight vector that results in the lowest value of the fitness function E_{tot} (Eq. 5) is the best solution to the problem. Optimization occurs as changes ('cross-over', 'mutation' and 'replication') are systematically applied to the population. Each cycle of these procedures is termed a 'generation'.

Because the values of $E_{pot}(i)$, r(i) and J(i) are all precalculated prior to starting FINGAR, the fitness evaluations are very fast. This means it is practical to run the GA with a relatively large population and number of generations. The simulations reported here were run with a population (IPOP) of 1000 members for 500 generations. Each generation consists of the following: (i) IPOP cross-overs are performed on the parent population P_0. A cross-over consists of exchanging the contiguous series of weight-vector elements $m_1 \rightarrow m_2$ in population member i with elements $m_1 \rightarrow m_2$ in population member j. Members i and j, and starting and ending codons m_1 and m_2 are chosen randomly from the parent population for each cross-over. This creates a daughter population P_1; (ii) mutate a specified fraction (here 0.005) of the codons of population P_1. Mutated codons are chosen randomly from the entire population. A mutation consists of changing the weight of the chosen codon. The change is chosen according to the function $(-1)^N(1 - e^{-\delta C/RT_{mut}})$, where δC is a real randomly chosen on [0,1], N is an integer randomly chosen on [1,2], R is the gas constant, and T_{mut} is 300 K. If a change would result in w(i) < 0, w(i) is set to 0.0. This creates daughter populations P_2; (iii) the weights for each member of the population are normalized so that their sum is equal to 1.0; (iv) the fitness of each member of the population P_2 is evaluated; (v) the next generation population is formed. This consists of the single most fit member from the last generation plus the most fit members of the current population. Members of P_2 are replicated for the next generation on the basis of their fitness. The number of replicates of each member is given by:

$$NREP(i) = (IPOP - 1)e^{-E_{tot}(i)/RT_{rep}} / \sum_{IPOP} e^{-E_{tot}(i)/RT_{rep}} \qquad (13)$$

subject to the condition that no more than 80% of the next generation population can be replicates of any single member of P_2 and with T_{rep} set to 300 K. If after replicating according to Eq. 12 the number of members of the next generation is not equal to IPOP (either due to round-off or due to the maximum-replication condition), the population is brought to IPOP members by adding or removing randomly chosen members of P_2.

All weights are forced to be integral multiples of 0.001. This granularity is imposed at all stages, including renormalization.

After the GA is completed, the weight vector that gives the lowest value E_{tot} for the fitness function can be used to perform several types of structural analysis. In addition to direct examination of the basis-set structures with the highest weights, one can generate a weighted average conformation. Since the basis-set structures used here were determined from an in vacuo MD simulation, we must first superimpose the basis sets on a common structure (the basis set with the highest weight) to place them in a common reference frame. We can then perform simple weighted coordinate averaging to produce an averaged conformation. Rms fluctuations of the atomic positions about this averaged conformation can subsequently be determined, again weighted according to the best result from the GA.

Results

The 'exact test'

Before proceeding to test FINGAR against MD-derived model data (restraints and basis sets), we chose to characterize the program's ability to fit the synthesized 'exact-test' data. The advantage of these test data is that because the restraints are generated directly from a weighted sum of the basis sets, we know the exact solution (the weight vector) and can directly compare any results to it. By

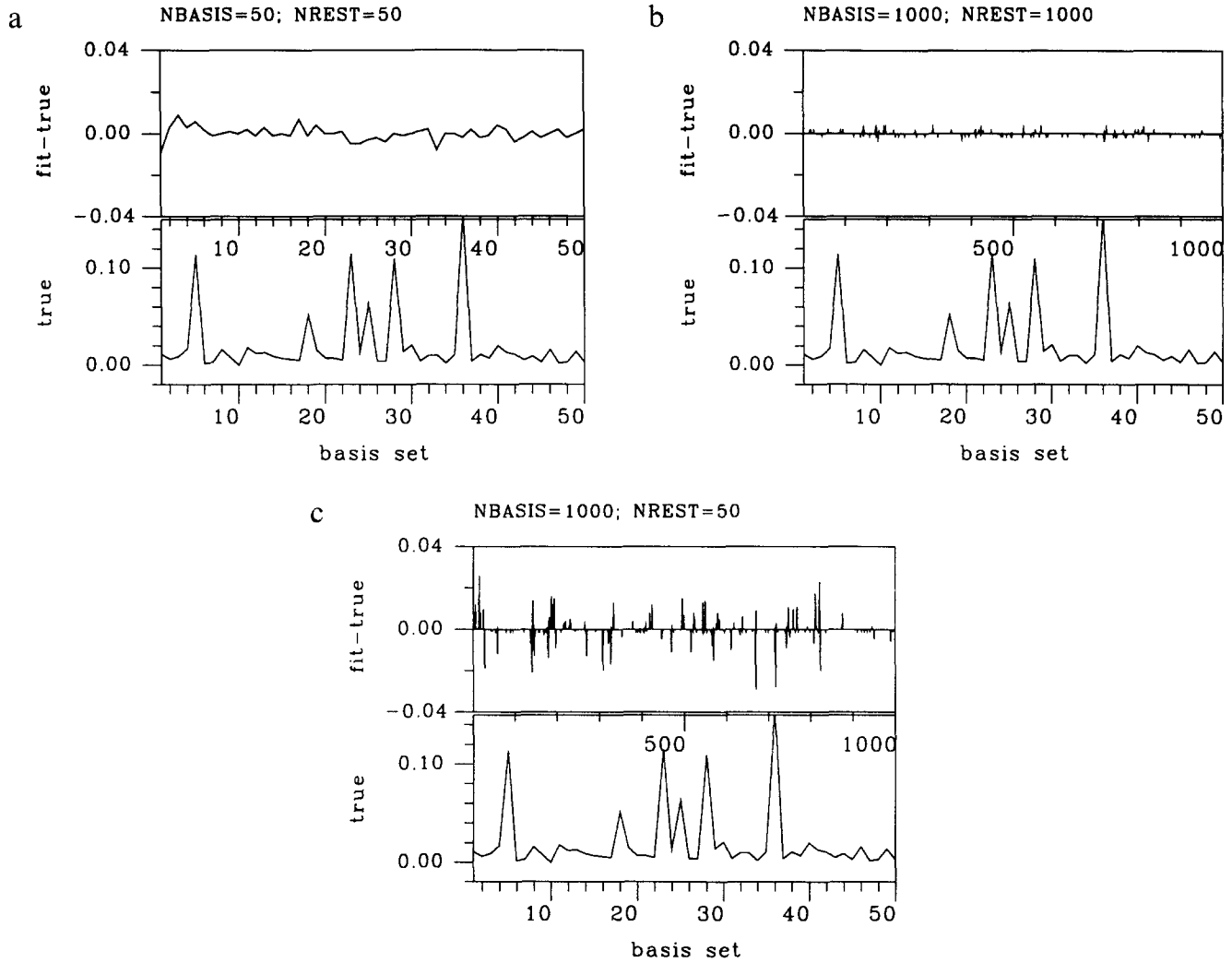

Fig. 1. (a) Lower spectrum: the true weight vector for the 50-basis-set/50-restraint 'exact-fit' simulation. Upper spectrum: the difference in weight between the true vector and the fitted vector as a function of basis set for this simulation (fit–true). Results are for the most-fit member of a population of 1000 members after 500 generations. The restraint set consisted of distances randomly assigned on the interval [2.5,5.5] Å. (b) The true and (fit–true) curves for the 1000-basis-set/1000-restraint 'exact-fit' simulation. (c) The true and (fit–true) curves for the 1000-basis-set/50-restraint 'exact-fit' simulation.

contrast, for the 'model-test' data we do not have, a priori, the 'correct' weight vector for comparison, and the results will reflect the ability (or lack thereof) of the chosen basis set of conformations to appropriately represent the ensemble from which the restraints were generated.

FINGAR 'exact-test' runs were performed in pairs. For each pair, the simulations differed in the number of synthesized distance restraints used: the first one used 50 distance restraints, while the second one used a number of distance restraints equal to the number of basis sets. Each pair of runs used a different number of basis sets: 50, 100, 200, 500 or 1000. The simulations were performed in pairs to test two things: firstly, how does the fit degrade as the number of fit variables (basis sets) becomes much larger than the number of experimental data (restraints), and secondly, is the GA able to generate a good fit even for large numbers of fit variables when the number of experimental data are equal to the basis set (thereby defining a unique solution). The distance restraints used in each fit were generated as the $\langle r^{-6} \rangle^{-1/6}$-weighted average of the distances in the basis sets.

Results for these simulations are presented in Fig. 1 and Table 1. Figure 1a compares the target (true) weight vector with the difference between the true and fit vectors for the run using 50 basis sets and 50 restraints. In Fig. 1b, the fit for the run using 1000 basis sets and 1000 restraints is shown. Figure 1c presents the fit for the severely undetermined case of 1000 basis sets and only 50 restraints. It is apparent from Figs. 1a and 1b and Table 1 that in all cases, even where 1000 weights must be fit, FINGAR does an excellent job reproducing the true weight vector when the number of restraints equals the number of basis sets. The maximum and rms weight differences for these fits (Table 1) are uniformly low. These results reflect the power of GA as a search technique for this problem.

TABLE 1
RESULTS OF 'EXACT-TEST' FINGAR SIMULATIONS

Basis sets	Restraints	FINGAR (fit–true)	
		Rms	Maximum
50	50	0.0032	0.0090
100	50	0.0041	0.0160
100	100	0.0017	0.0060
200	50	0.0034	0.0170
200	200	0.0011	0.0060
500	50	0.0050	0.0410
500	500	0.0006	0.0020
1000	50	0.0032	0.0290
1000	1000	0.0004	0.0020

All simulations were run with a population of 1000 members for 500 generations. The results listed are for the most-fit member of each final population.

By comparison, the ability of FINGAR to generate a fit in accord with the true vector degrades significantly with the number of basis sets when the experimental restraints do not fully determine the answer. The fit to 50 restraints is excellent for a set of 50 basis sets, as expected. The fit using 100 basis sets and 50 restraints (Table 1) is also good, though not as good as with the overdetermined problem. The maximum weight difference between the fit and actual weight vectors is 0.0160, as compared to 0.0090 for the 50-basis-set/50-restraint-set. The maximum weight difference continues to increase as the number of basis sets increases and the number of restraints remains at 50. These results merely reflect the fact that when the problem is undetermined, numerous solutions can be obtained that meet the provided restraint criteria well but differ from the 'true' set. If we run any of these undetermined simulations with a different random number seed, we obtain a different answer, but typically one which is a similarly good (or poor) approximation to the true solution. Note that the inability to determine the true fit for the underdetermined problem does not reflect a weakness in the GA method. The underdetermined runs result in rms distance-restraint violations that are the same as their fully determined counterparts (between 0.00 and 0.01 Å). However, without sufficient distance restraints, the chances of determining the solution that minimizes both distance-restraint violations *and* agreement with the true vector are small – increasingly so as the problem becomes more undetermined.

In any case, these results provide evidence that FINGAR is capable of generating a reliable fit of around 100 basis sets to a set of 50 independent NMR-derived restraints. As more basis sets are fit without increasing the amount of NMR-derived restraint information, undertermination of the solution becomes increasingly significant. It should be noted that undetermination is not necessarily a problem when fitting NMR data; it *does* mean that in such a case the derived solution may not be unique or may be determined in significant part by the force field being used.

The results indicate that the 500 generations being performed should be more than enough to produce a good optimized solution. Figure 2 presents the rms difference between the refined and true basis weight vectors as a function of generation number. As is typical of GAs, the fitness improves quickly in the early generations, but undergoes a much more gradual improvement in later stages.

FK506 'model-test' refinement

Having satisfied ourselves that FINGAR works as expected, we proceeded to apply the program to two model test systems that have been examined in detail before. The first of these is the 126-atom macrocycle FK506. Previously, time-averaged refinement was shown to generate a more reliable and realistic representation of the true molecular conformation and flexibility of this molecule than is possible using standard refinement (Pearlman, 1994a). However, even the time-averaged results suffered some disadvantages, particularly in that the motions of many parts of the molecule are overestimated with this technique. In addition, the time-averaged results (as well as those from standard refinement) should be viewed in terms of a molecular average and envelope of sampled structures, rather than in terms of the individual snapshots that are more easily usable in subsequent molecular studies.

The previous study was carried out by performing unrestrained MD on FK506 to generate a set of $\langle r^{-3} \rangle^{-1/3}$-weighted distances between all proton pairs for which resonances were observed in an experimental NMR study of FK506 complexed to binding protein FKBP. These distances constituted the 'experimental' restraints used to then refine FK506. A comparison of the results from the

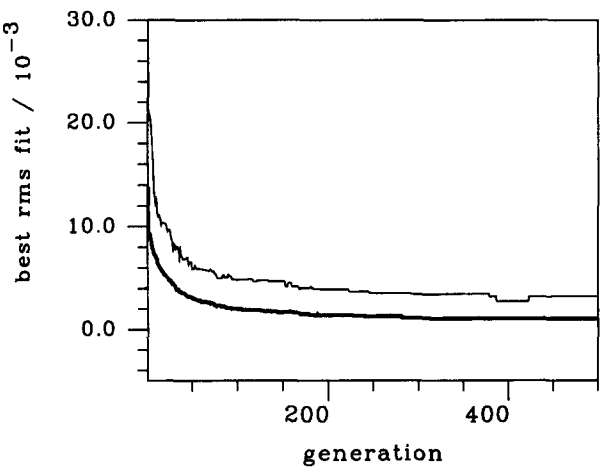

Fig. 2. Rms difference between the true and most-fit weight vectors as a function of generation. Thin line: the 50-basis-set/50-restraint simulation in Fig. 1; thick line: the 200-basis-set/200-restraint simulation. These curves are representative of all the simulations listed in Table 1.

refinement to the unrestrained trajectory allow an evaluation of the refinement method unbiased by the force field (since both the 'experimental' restraint set and the refinement use the same force field). And, because the restraints are generated from a model calculation, we know with total accuracy what the correct refined mean structure and associated motion about that mean should be. In the earlier study, $\langle r^{-3} \rangle^{-1/3}$ averaging, rather than $\langle r^{-6} \rangle^{-1/6}$, was used to be consistent with a suggestion that this is the appropriate weighted averaging to perform for MD when the simulation is short relative to the correlation time for angular fluctuations (Kessler et al., 1988).

The same set of MD-derived restraints was used in the FINGAR refinement of FK506. For consistency $\langle r^{-3} \rangle^{-1/3}$ averaging was used for the FINGAR refinement. No flat-well regions were associated with the restraints. The results obtained are directly comparable to the unrestrained trajectory that was originally used to generate the restraints. Using a maximum cluster diameter of 0.93, 3333 MD trajectory snapshots (see above) were reduced to 104 clusters. The 104 structures with the lowest potential energy in their respective cluster were taken as the basis set for these FINGAR runs. A population of 1000 members was run for 500 generations. At the end of the simulation, average and rms NMR distance restraint violations of 0.05 and 0.07 Å, respectively, had been obtained, with an averaged potential energy of 38.20 kcal/mol. These measures are considerably more favorable than the corresponding measures for standard MD refinement (0.19/0.36 Å and 61.20 kcal/mol, respectively) with the

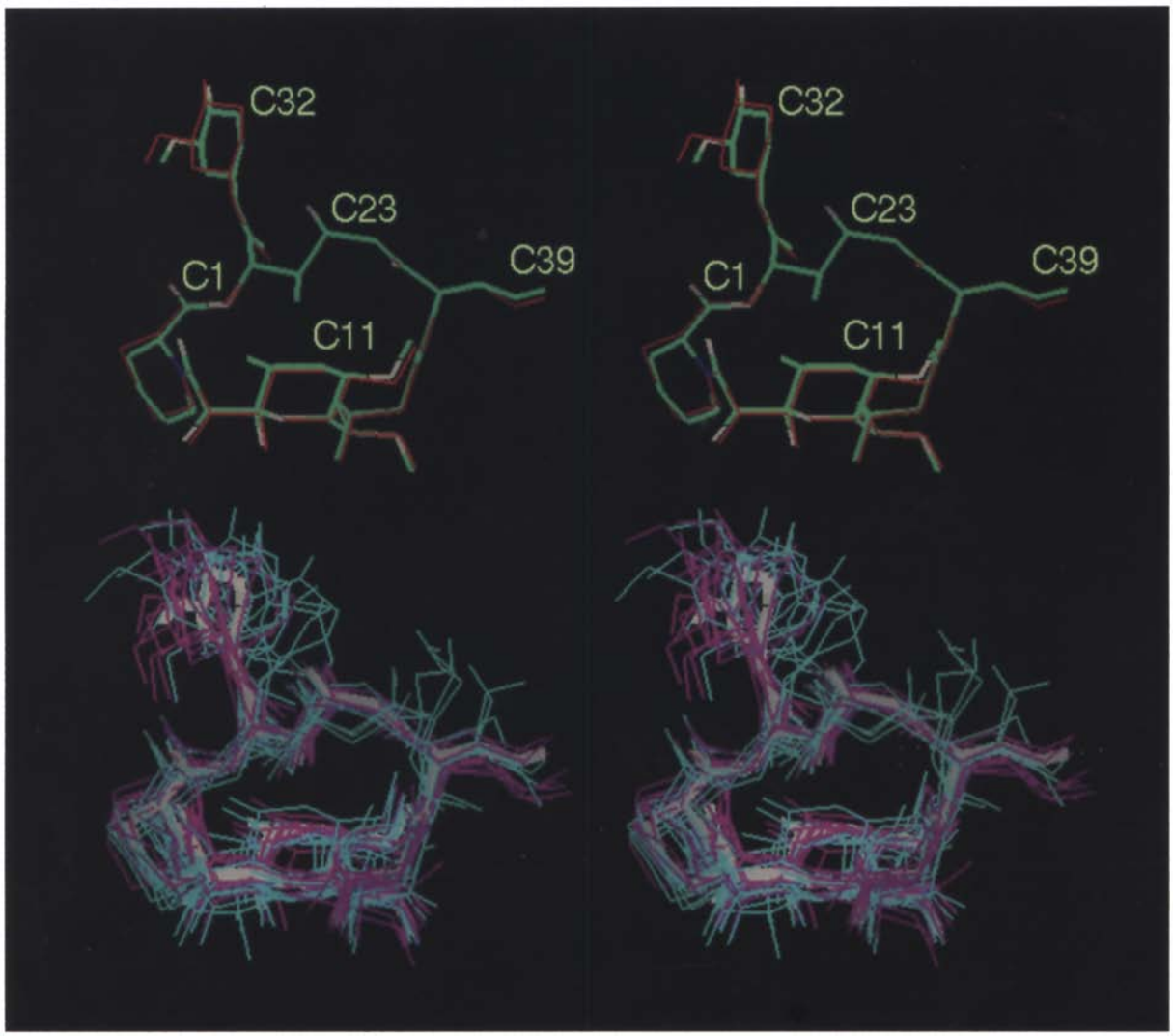

Fig. 3. (Upper pictures) A comparison of the FINGAR fit and true-averaged FK506 structures. The true-averaged FK506 structure is shown in heavy lines. Carbon atoms are green, oxygen atoms are white, nitrogen is blue. Hydrogens are not shown. The FINGAR-generated structure is shown in thin red lines. The FINGAR structure is calculated from the weighted average of the basis-set structures for the most-fit member of the population after a FINGAR run. (Lower pictures) The ensemble of basis-set structures with non-zero weights that contributed to the best solution. The eight structures with weights > 5% are shown in magenta, while the remaining nine structures are shown in cyan. The true-averaged FK506 structure is shown in heavy white lines.

same set of restraints. The restraint violations are the same as those obtained from time-averaged refinement, and with a substantially lower average potential energy (Table 2). A plot of E_{tot} versus the generation of the best solution in the generation (not shown) is essentially flat past 200–250 generations, indicating the best solution has easily converged by generation 500. The final solution consists of 17 basis structures with non-zero weights. Of these, only eight contribute more than 5% to the solution. The ensemble of basis-set structures with non-zero weights is shown in Fig. 3, with those structures contributing more than 5% to the solution in magenta, and the remainder in cyan. It can be seen that this ensemble widely samples the FK506 conformational envelope, especially in the allyl (around C39) and cyclohexyl 'Northwest Corner' (around C32) regions.

Both the weighted mean structure and the weighted rms atomic fluctuations about the mean position can be calculated for the most-fit member of the population after running FINGAR. These can be compared directly with the corresponding properties for the unrestrained MD trajectory from which the restraints were originally derived (the 'experimental data'). These comparisons are presented in Figs. 3 and 4, respectively. As can be seen, the mean structure is nearly indistinguishable from the experimental one, and the rms fluctuations curve is also in excellent accord with the experimental data. In addition, in Fig. 4 we have plotted the rms fluctuations that arise when only the eight structures contributing > 5% to the average are included (dashed line). While this curve is also in acceptable agreement with experiment, the low-weight structures do perceptibly improve agreement. There are two places in the curves, atoms 99 and 118, where there is notable disagreement between the FINGAR-refined and experimental curves. These correspond to the carbon atoms C40 and C43, both part of pendant methyl

TABLE 2
REFINEMENT OF FK506

Method	Restraint violation		$<E_{pot}>$
	Rms	Average	
FINGAR[a]	0.07	0.05	38.20
Standard MD[b]	0.36	0.19	61.20
Time-averaged MD[b]	0.07	0.05	54.99

Violations are in Å. Energies are in kcal/mol. All simulations used the same set of 66 restraints distances derived from an unrestrained MD run.

[a] The FINGAR simulation was run with a population of 1000 members for 500 generations. The results listed are for the most-fit member of the final population. $<r^{-3}>^{-1/3}$ distance averaging was used to be consistent with the averaging used to generate the target restraints.

[b] The standard and time-averaged MD simulations were run for 1 ns with 5-kcal/mol restraint weights. For the time-averaged run, $<r^{-3}>^{-1/3}$ averaging was employed with an exponential damping-factor constant of $\tau = 10$ ps.

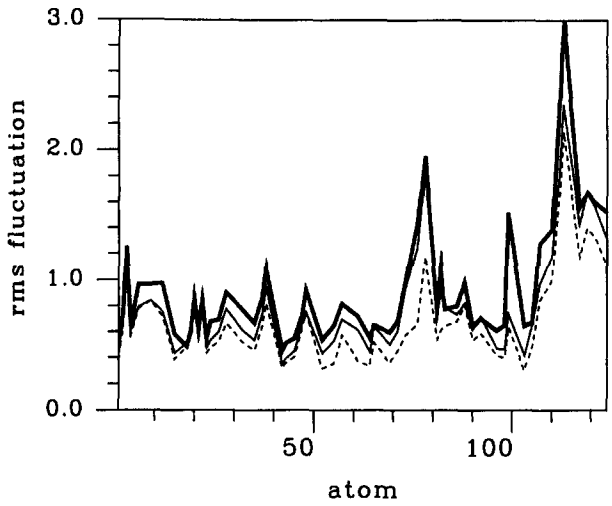

Fig. 4. Rms fluctuations (Å) about mean atomic position for heavy atoms of FK506. The heavy solid-line curve represents the 'true' fluctuations, taken from the unrestrained MD trajectory from which the NMR restraints were derived. The thin solid line represents the fluctuations calculated from the weighted average of the 17 structures in the basis set with non-zero weights after a FINGAR run. The dashed line represents the fluctuations calculated from the weighted average of the eight structures in the basis set with weights > 0.05 after a FINGAR run. The FINGAR run results are for the most fit member of the population.

groups of FK506. That the rms fluctuation fit for FINGAR is much improved relative to either standard or time-averaged refinement is demonstrated in Fig. 5. There, the rms fluctuation curves for standard (short dashed line), time-averaged (long dashed line), and FINGAR (thin solid line) refinements are compared to the 'experimental' curve (thick solid line). In contrast to the FINGAR curve, the standard refinement results severely underestimate atomic mobility, while the time-averaged results overestimate it. The atomic mobility predicted by FINGAR is very close to the experimental curve. Clearly, FINGAR does the best job at reproducing the true rms fluctuations.

Adenosine 'model-test' refinement

A more challenging problem for both standard and time-averaged refinement has been refining against J-coupling data for adenosine (Pearlman, 1994b). As with FK506, 'experimental' data (here, J-coupling constants) were generated from an unrestrained MD simulation. Results from refinement using these J-coupling restraints can then be compared with the unrestrained trajectory. Five J-coupling restraints were employed, corresponding to the five torsions about which J-coupling information could, in principle, be collected experimentally for this molecule: $J_{H1'H2'1}$, $J_{H2'1H3'}$, $J_{H3'H4'}$, $J_{H4'H5'1}$, and $J_{H4'H5'2}$ (Van de Ven and Hilbers, 1988; Kim et al., 1992). These five J-coupling constants are related to two conformational variables: the phase angle of pseudorotation, P (which describes the conformation of the furanose sugar moi-

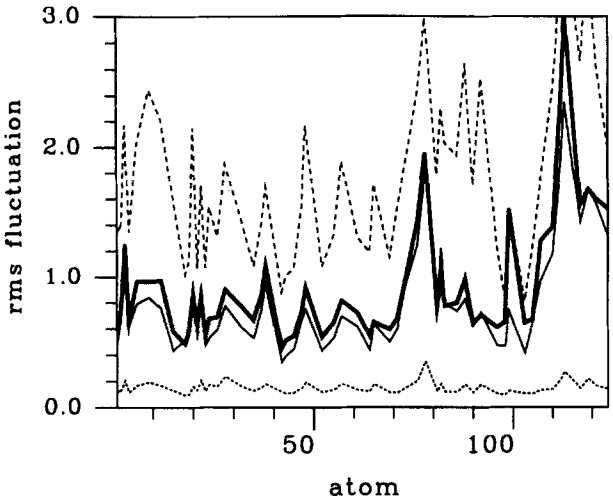

Fig. 5. Rms fluctuations (Å) about mean atomic position for heavy atoms of FK506. Here, the rms fluctuations calculated using standard MD refinement (short dashed line), time-averaged MD refinement (long dashed line), and FINGAR refinement (thin solid line) are compared with the true fluctuations from the unrestrained MD trajectory (heavy solid line). The MD-derived rms fluctuations are from simulations using modest restraint weights ($K_{restraint}$ = 5.0 kcal/mol) and the same restraint target values as those used for the FINGAR run.

ety), and the torsion angle γ (C3'-C4'-C5'-O5') (Pearlman, 1994b).

The reason this refinement presents a more difficult challenge for both MD-based refinement methods, is that during the unrestrained MD trajectory leading to the J-restraints, the γ-torsion, from which two of the J-coupling restraints are determined, undergoes several large conformational transitions between *t* (~180°) and g^+ (~60°) (Fig. 6). The related averaged J-coupling restraints reflect this. Note, however, that the period for the transition is rather long. The periodicity is lost when the averaging is performed, and does not appear explicitly or implicitly in the restraints themselves. Therefore, when these restraints are applied during MD-based time-averaged refinement, the conformation undergoes relatively rapid oscillations in order to attempt to reproduce the experimentally derived values. For two disparate conformations such as *t* and g^+, this can lead to artificial distortions in the molecular structure and severe errors in the calculated rms atomic motions. In fact, this is what has been observed for this system. While the time-averaged refinement values are still significantly better than those from standard refinement, they are far from ideal (Table 3).

Such a situation is well suited to the FINGAR approach. This is because the time domain plays no part in FINGAR refinement. The potential energy barrier between two low-energy conformers – and the related rate of interconversion between them – is irrelevant. FINGAR merely fits the relative weights of the two low-energy conformations. As long as the basis set adequately samples the low-energy conformations available to the molecule, it makes no difference what the barriers to, or rates of, interconversion among those conformers are.

The J-coupling restraint approach was previously characterized in detail for MD-based refinement of the adenosine system (Pearlman, 1994b). The same set of five averaged J-coupling restraints used in the previous study were applied in our FINGAR refinement of adenosine; 4000 MD trajectory snapshots were reduced to 106 clusters using a maximum cluster diameter of 0.385. The lowest potential energy structure in each cluster was placed in the basis set. FINGAR refinement was run on a population of 1000 members for 500 generations. This refinement resulted in 33 basis structures with non-zero weights. However, of these structures only seven contributed by more than 5%, and together these seven structures account for 65% of the total ensemble. There are 18 structures from the ensemble with weights of ≥1%, and these contribute 95% of the total. The weighted averaged FINGER-determined adenosine structure is nearly identical to the true average, as can be seen in Fig. 7. Also shown in Fig. 7 is the ensemble of basis set structures with non-zero weights contributing to the solution. Those basis sets with weights > 5% are shown in magenta, and the remaining basis sets are shown in cyan. Appreciable sampling of the sugar conformation and γ-torsion can be observed from this ensemble. A number of conformations are also seen where the adenine base adopts an alternate conformation about the glycosyl (C1'-N9) bond. All of these are among the structures contributing ≤1% to the ensemble.

The results from the FINGAR refinement of adenosine are presented in tabular form in Table 3. The five J-coupling restraint values determine the values of the conformation of the furanose sugar, represented numerically by

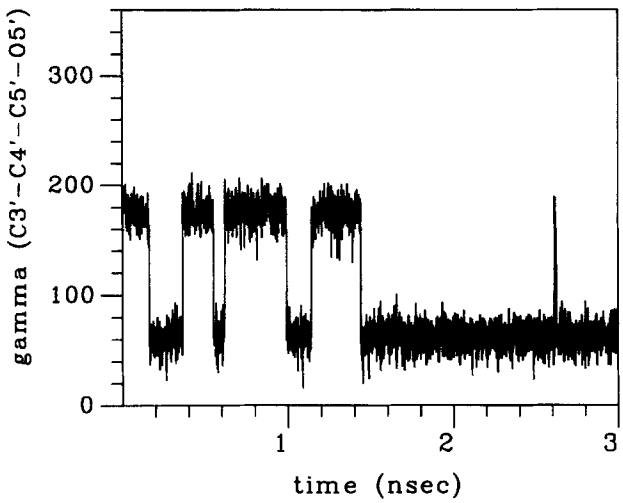

Fig. 6. Trajectory of the γ(C3'-C4'-C5'-O5') torsion angle (in degrees) in adenosine during the unrestrained MD trajectory that was used to generate the J-coupling restraints. The γ-value is related to two of the J-coupling restraints through the following Karplus relationships: (i) $J_{H4'H5'1} \approx 10.2 \cos^2(\gamma - 120°) - 0.8 \cos(\gamma - 120°)$ and (ii) $J_{H4'H5'2} \approx 10.2 \cos^2(\gamma) - 0.8 \cos(\gamma)$.

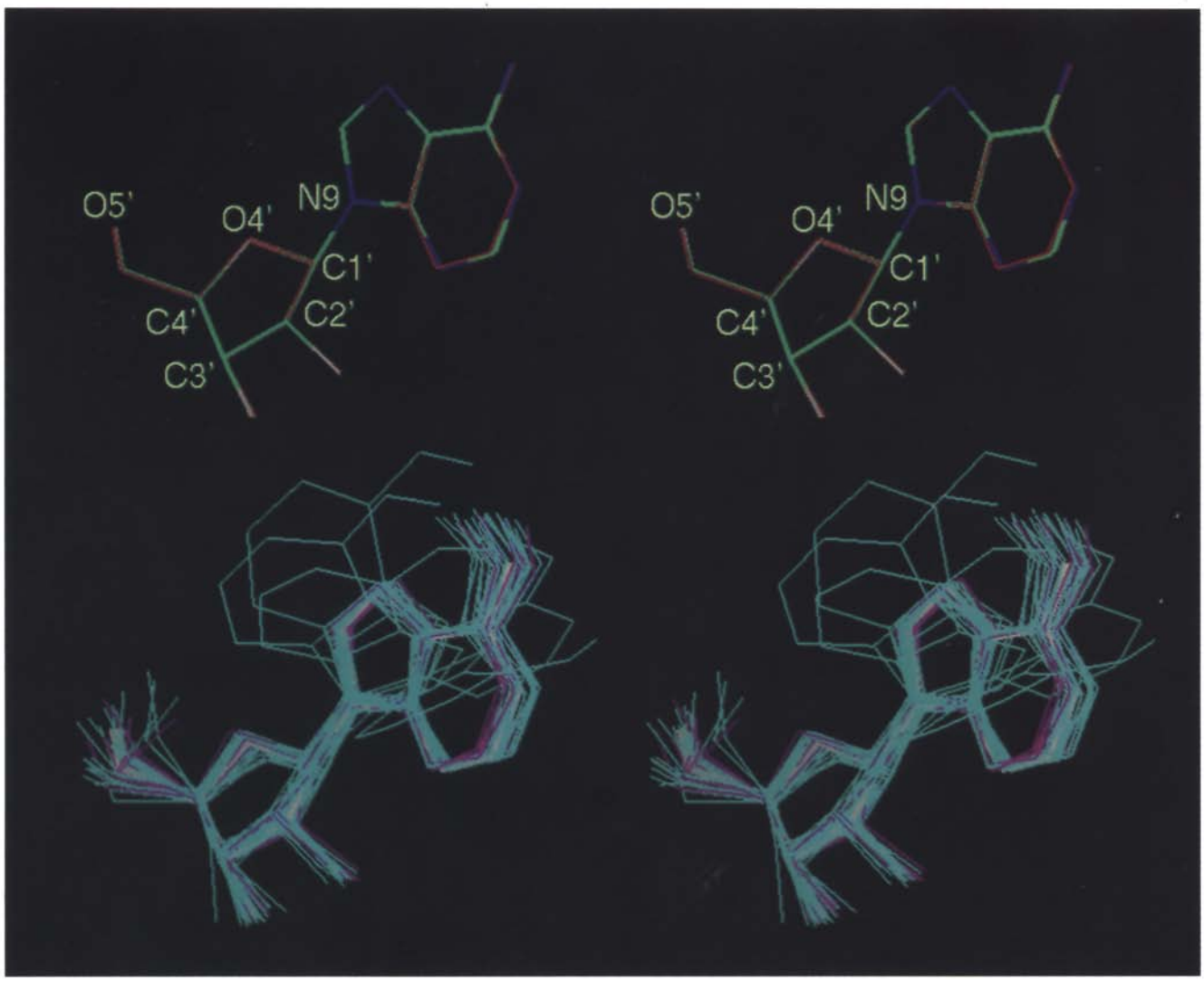

Fig. 7. (Upper pictures) A comparison of the FINGAR fit and true-averaged adenosine structures. The true-averaged adenosine structure is shown in heavy lines. Carbon atoms are green, oxygen atoms are white, nitrogen is blue. Hydrogens are not shown. The FINGAR-generated structure is shown in thin red lines. The FINGAR structure is calculated from the weighted average of the basis-set structures for the most-fit member of the population after a FINGAR run. (Lower pictures) The ensemble of basis-set structures with non-zero weights that contributed to the best solution. The seven structures with weights > 5% are shown in magenta, while the remaining 26 structures are shown in cyan. The true-averaged adenosine structure is shown in heavy white lines.

the phase angle of pseudorotation, P, and of torsion angle γ. The averaged values of P and γ, and the associated rms fluctuations, are given in Table 3 for the unrestrained MD trajectory that lead to the J-coupling restraints (the 'experimental' values), for the FINGAR refinement, for standard refinement, and for time-averaged refinement. As is seen, FINGAR refinement leads to a much better agreement in rms angle fluctuations than either of the MD-based refinement methods. In addition, only FINGAR is capable of generating an averaged value of γ in reasonable agreement with the experimental one. Finally, the averaged value of the potential energy using FINGAR is at least 15 kcal/mol lower than that obtained with either MD-based refinement method. The weighted solution produced by FINGAR shows no J-coupling restraint violations. In contrast, both of the MD-based methods yielded residual restraint violations.

It is worth noting that the FINGAR-derived average structure reproduces the appropriate value of the χ-torsion connecting the sugar to the base, despite the fact that no restraints were applied to this linkage. In fact, this is to be expected, given the fact that the calculated potential energies used in the FINGAR refinement are identical to those used in the MD simulation from which the experimental restraints were derived. In a situation where no restraint data are used, one expects the force field itself to dictate conformation preferences, unless the restraints are artificially coloring the MD trajectory. Referring to Table 3, it is seen that the averaged value of χ is properly reproduced with all refinement methods. However, the afore-

TABLE 3
REFINEMENT OF ADENOSINE

Method	Restraint violation		$\langle E_{pot} \rangle$	$\langle P \rangle$[a]	$\langle \gamma \rangle$[b]	$\langle \chi \rangle$[c]
	rms	average				
FINGAR[d]	0.00	0.00	−4.87	141.6 (28.8)	99.7 (56.6)	−11.9 (20.2)
Standard MD[e]	0.33	0.26	10.53	191.1 (11.1)	60.9 (8.1)	−9.4 (15.5)
Time-averaged MD[e]	0.48	0.31	40.22	139.2 (35.0)	76.6 (96.3)	−11.3 (50.0)
(Unrestrained)[f]			2.13	138.5 (28.4)	93.5 (55.3)	−11.9 (17.8)

Energies are in kcal/mol. Angles are in °. All simulations used the same set of 5 vicinal ^{3}J-coupling values derived from an unrestrained MD run. Violations are J-coupling violations and are unitless. Values in parentheses are rms fluctuations.

[a] The phase angle of pseudorotation, which is related to the conformation of the furanose ring (Altona and Sundaralingam, 1972), which is in turn related to the torsion angles described by $J_{H1'H2'1}$, $J_{H2'1H3'}$ and $J_{H3'H4'}$.

[b] Torsion angle γ(C3'-C4'-C5'-O5'), which is related to J-couplings $J_{H4'H5'1}$ and $J_{H4'H5'2}$.

[c] Torsion angle χ(O4'-C1'-N9-C8).

[d] The FINGAR simulation was run with a population of 1000 members for 500 generations. The results listed are for the most-fit member of the final population.

[e] The standard and time-averaged MD simulations were run for 3 ns with 4-kcal/mol restraint weights.

[f] The unrestrained simulation that served as the 'experimental' data from which the J-coupling restraints used in the refinements were derived.

mentioned tendency of time-averaged refinement to exaggerate atomic motion, clearly results in anomalously large motion about this bond. The other refinement approaches predict motion about this bond that agrees with the unrestrained trajectory.

The superiority of FINGAR to either MD-based refinement method is dramatically illustrated in Fig. 8. There, the rms atomic fluctuations about the mean position are shown for the unrestrained/experimental simulation (thick solid line), for standard MD refinement (short dashed line), for time-averaged MD refinement (long dashed line), and for FINGAR refinement (thin solid line). FINGAR refinement leads to a solution which is in excellent agreement with experiment, and which is superior to either of the MD-based techniques. Once again, as with FK506, we see that standard refinement underestimates the rms fluctuations, while time-averaged refinement overestimates them. It is worth noting that while the standard refinement fluctuations curve is in somewhat better agreement with experiment than the time-averaged fluctuations curve, the averaged conformation from standard refinement is in relatively poor agreement with experiment (Table 1). The FINGAR results are in good agreement by both sets of criteria.

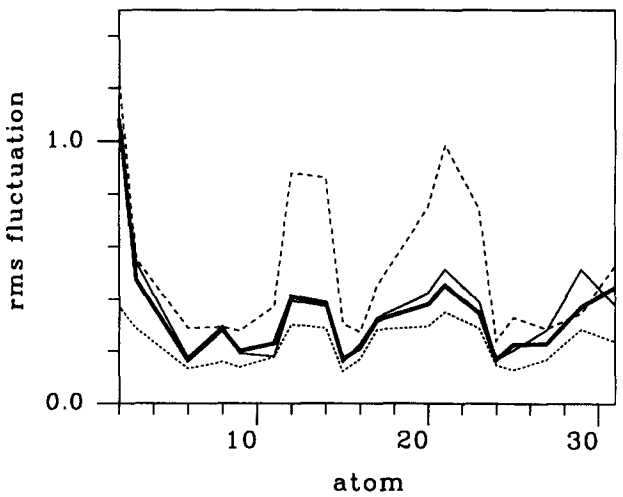

Fig. 8. Rms fluctuations (Å) about mean atomic position for heavy atoms of adenosine. The rms fluctuations calculated using standard MD refinement (short dashed line), time-averaged MD refinement (long dashed line), and FINGAR refinement (thin solid line) are compared with the true fluctuations from the unrestrained MD trajectory (heavy solid line). The MD-derived rms fluctuations are from simulations using modest restraint weights ($K_{restraint}$ = 4.0 kcal/mol) and the same restraint target values as those used for the FINGAR run.

How do errors affect the FINGAR refinement?

The above-presented studies demonstrate the power of FINGAR for refinement against error-free data, and against basis sets that contain the necessary structures. Important related questions concern how well the method works when these conditions are not met. That is, how are the FINGAR-refined results affected both by the errors associated with data collected in actual NMR experiments and by the absence of important conformations in the basis set?

To test the effects of errors in NMR-derived restraint distances, we ran a series of FINGAR FK506 refinement simulations where a flat region was associated with the target distance. That is, the exact restraint in Eq. 7 was replaced by the function (Eq. 14):

$$E_{bond} = \sum_{NBASIS} \begin{cases} K_b(\langle r(n)\rangle - r_{NMR}(n)_l)^2 \\ \qquad \text{for } \langle r(n)\rangle < r_{NMR}(n)_l \\ 0 \quad \text{for } r_{NMR}(n)_l \le \langle r(n)\rangle \le r_{NMR}(n)_u \\ K_b(\langle r(n)\rangle - r_{NMR}(n)_u)^2 \\ \qquad \text{for } r_{NMR}(n)_u < \langle r(n)\rangle \end{cases} \quad (14)$$

with

$$r_{NMR}(n)_l = r_{NMR}(n) - \Gamma(r_{NMR}(n)) \quad (15a)$$

and

$$r_{NMR}(n)_u = r_{NMR}(n) + \Gamma(r_{NMR}(n)) \quad (15b)$$

where $\Gamma(r_{NMR}(n))$ is the 'error' associated with restraint n. The flat region is typically used in refinement calculations to account for uncertainties in the experimental target distance. A series of FINGAR simulations have been run. In each, $\Gamma(r_{NMR}(n))$ is fixed at a different value between 0 and 3σ, where σ is the rms fluctuation in the associated target distance $r_{NMR}(n)$ in the unrestrained MD trajectory from which the target distance was derived. In total, 31 FINGAR simulations, corresponding to 0.1σ intervals between 0 and 3σ, have been performed. Other FINGAR simulation parameters were as described above.

For each FINGAR simulation, the averaged structure and rms atomic fluctuations about the averaged structure have been calculated. The averaged structures, the rms atomic fluctuations, and $<E_{pot}>$ vary systematically with Γ, as shown in Fig. 9. Encouragingly, all three quantities vary in the expected manner as Γ increases and the information content of the restraints decreases. For example, $<E_{pot}>$ decreases as Γ increases, until $\Gamma = 1.4\sigma$, after which point it levels off to about 36 kcal/mol (dashed curve). One expects that the averaged energy will diminish as the restraints are loosened, since restraints typically bias the system to different (higher calculated energy) minima than would be chosen by the force field itself. That is, restraints are only necessary because the force field or model itself is imperfect. By the time $\Gamma = 3\sigma$ is reached, the restraints basically impart no information to the refinement. This can be seen by performing a FINGAR refinement where K_{bond} is set to zero, so that the refined ensemble depends only on the potential energies of the basis sets. The results from this refinement (not shown) are identical to those for the $\Gamma = 3\sigma$ refinement. Of course, the value of Γ at which the restraints no long impart information to the refinement is case-dependent, and will vary with how well the potential energies assigned to the individual basis-set members agree with the experimental ensemble implicit in the restraints themselves. The fact that this occurs at a relatively low multiple of σ for this system reflects the fact that the 'experimental' restraints were derived from a MD simulation that used the same potential energy force field as that used to calculate the individual energies of the basis sets.

A parameter more sensitive to the quality of the refined results is the rms of atomic fluctuations about the mean position. As noted above, the FINGAR approach produces atomic fluctuations in much better agreement with the experimental data than either standard refinement or time-averaged refinement. In Fig. 9, RMSXX, the rms difference between the rms atomic fluctuations for FINGAR-refined results and those for the experimental data:

$$RMSXX = \sum_{\text{heavy atoms}} ((rms_{refined} - rms_{exp})^2 / N)^{1/2} \quad (16)$$

is plotted versus Γ (thick solid curve). As can be seen, RMSXX steadily increases with Γ until a plateau at around $\Gamma = 2.3\sigma$ is reached. The implications of this plot are twofold: firstly, that the restraints are necessary to properly reproduce the experimental data – that is, the good fit between experiment and refinement is not simply due to the self-consistent nature of this study. Secondly, that the refinement process is not introducing artifacts as the restraints become tighter. As Γ deceases, RMSXX also decreases. This is in contrast to both the standard and time-averaged refinement, where very tight restraints can result in artifacts in the refined parameters. Two additional notes on the RMSXX plot are necessary. Firstly, the modest jaggedness of the curve between $\Gamma = 0$ and $\Gamma = 2.2\sigma$ reflects the fact that while the GA in FINGAR is quite an efficient optimizer, it is still possible to obtain a solution which is highly optimized, near the global minimum, and perfectly acceptable, but which is not itself the global minimum. It is likely that we are obtaining such near-global-minimum solutions in our FINGAR runs, which explains the jaggedness. Secondly, large fluctuations appear in the RMSXX curve beyond $\Gamma = 2.2\sigma$. These do not reflect inadequacies in the refinement, but rather the fact that for such large values of Γ, there are

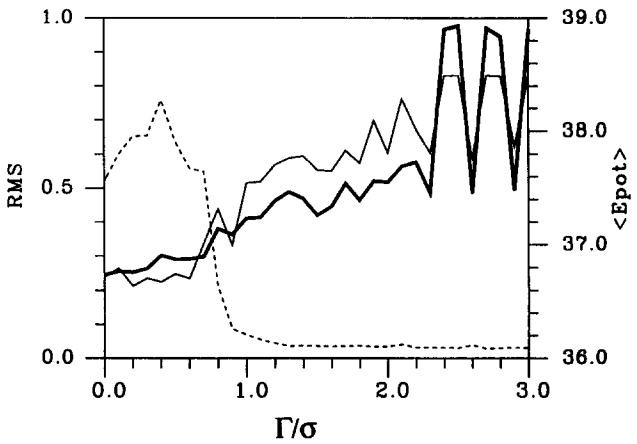

Fig. 9. Weighted average properties of the best GA solution for the FK506 FINGAR refinement as a function of the flat region error estimate associated with each of the restraints. The flat region is of width 2Γ, centered on the true restraint value ($r_{NMR}(n) \pm \Gamma$). The value of Γ is given in units of σ, the rms fluctuation in $r_{NMR}(n)$ in the unrestrained MD run from which the target distances were derived. Thin solid curve: rms difference in atomic position for the heavy atoms of the weighted average structure from refinement versus the true average structure. The rms is calculated over all heavy atoms after best-fit superposition; thick solid curve: RMSXX, the rms difference between the mean-squared atomic motions of all heavy atoms about the refined averaged structure and the true rms motion of their counterparts; thin dashed line: the weighted average value of E_{pot}.

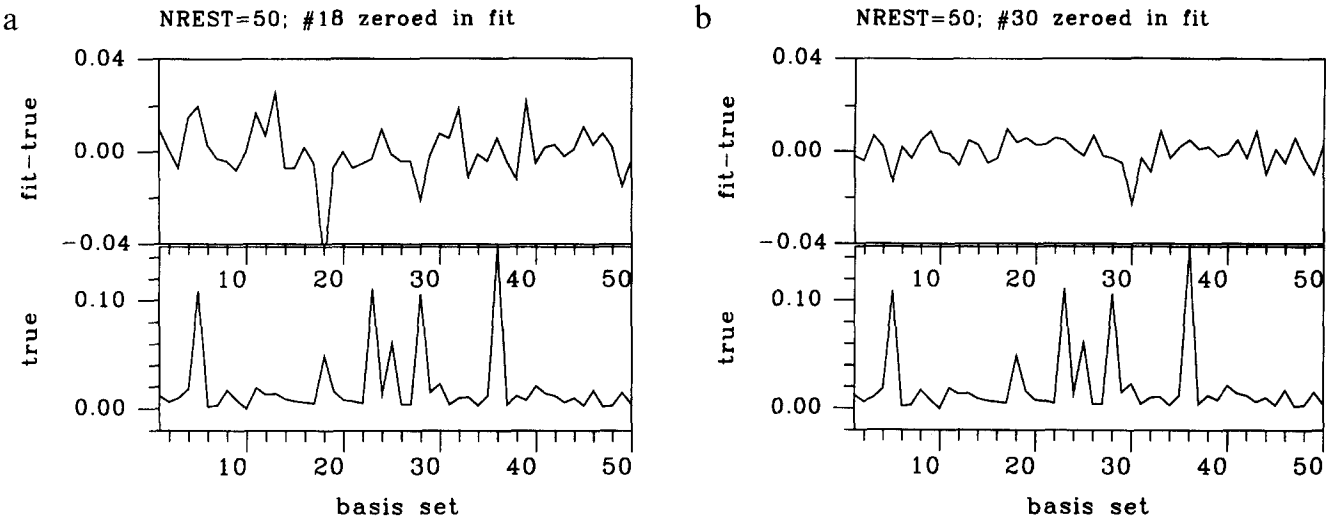

Fig. 10. (a) Lower spectrum: the true weight vector for the 50-basis-set/50-restraint 'exact-fit' simulation where the weight of the basis set making the sixth-largest contribution to the true solution (number 18; w(i) = 0.049) was forced to remain at zero for members of the population during refinement. Upper spectrum: the difference in weight between the true vector and the fitted vector as a function of basis set for this simulation (fit–true). (b) The true and (fit–true) curves for a 50-basis-set/50-restraint 'exact-fit' simulation where the weight of the basis set making the seventh-largest contribution to the true solution (number 30; w(i) = 0.023) was forced to remain at zero for members of the population during refinement. The simulations represented by (a) and (b) are identical to that in Fig. 1a, except that during refinement one fitted basis-set weight was forced to zero. Results are for the most-fit member of a population of 1000 members after 500 generations. The restraint set consisted of distances randomly assigned on the interval [2.5,5.5] Å.

two solutions which give nearly identical values of the fitness function (the same values of $<E_{pot}>$ and no restraint violations), but which result in quite different atomic fluctuations. The solution that is obtained will depend in large part on the random number series used for the simulation. More generally, because of differences in the refined solution that can arise using different random number seeds, in practice it would generally be a good idea to run a FINGAR simulation more than once to assure that the GA optimizer has properly converged.

Also plotted in Fig. 9 is the rms difference in atomic positions between the averaged conformation from each refinement and the experimental average (thin solid curve). As seen for RMSXX, the difference between the refined and experimental values increases (agreement becomes poorer) as the information content of the restraints is reduced. The same comments about the modest jaggedness and fluctuations at large values of Γ made for RMSXX apply for this curve as well. Note that a direct comparison can be made between this curve and Fig. 6 of a previous study of FK506 using standard and time-averaged refinement. In that study, it was shown that due to refinement biases in both of these refinement methods, agreement between the averaged structures *improves* (rms difference decreases) as Γ increases. That is, as the information content of the restraint goes up, so do biases to the averaged structure. By contrast, with FINGAR, the averaged structure is closer to the true average as the information content increases. The FINGAR behavior is, of course, more desirable.

In addition to errors in target restraint distances, it is possible within the FINGAR method to encounter errors due to inadequacies in the basis set. It would be quite difficult to characterize these in the general case, but it is instructive to consider how refinement results are affected when an important basis-set structure is missing. This is best determined by performing a couple of additional 'exact' test simulations. The 50-basis-set/50-restraints simulation presented in Fig. 1a was rerun, but with the weight of a single preselected basis set fixed at 0.0 in all members of the population during refinement. Since in the exact-fit runs the restraints are calculated directly from the basis sets, we can determine how the entire fit is affected when a basis set that contributed to the target restraints (which is not replicated by other members of the basis set) is missing. Figure 10 presents the results of two simulations: one where the weight of the basis set that contributed the 6th-largest amount to the restraints (number 18; w(i) = 0.049) is forced to zero in the refined population; and one where the weight of the basis set that contributed the 7th-largest amount to the restraints (number 30; w(i) = 0.023) is forced to zero. The weight zeroed out in the former simulation is the smallest of the six large weights that dominate the weight profile for this run. This is representative of a simulation where a critical basis set is missing. The weight zeroed out in the latter simulation is still of moderate importance, but is significantly smaller. As can be seen by comparison to Fig. 1a, removal of only one important basis set has appreciable affects on the refined weights of the entire population. When the missing basis set is of lesser importance, the overall affect on the refined weight vector is considerably

decreased, but the effect is still spread throughout the population. Clearly, the take-home message here is that one should be as careful as possible to ensure that the basis set includes representatives of all conformers that could potentially have contributed to the measured NMR data.

Discussion

A new genetic-algorithm-based method, FINGAR, for refining molecular structures from NMR-derived distance and angular restraints is presented and applied to several model systems, including two systems previously examined in detail using standard MD and time-averaged MD refinement techniques. The results clearly demonstrate that FINGAR is a superior method to these for NMR refinement. Compared to MD-based refinement methods, FINGAR is capable of yielding a weighted ensemble of structures which has smaller restraint violations and a lower averaged energy, and with associated rms atomic fluctuations in much better accord with experiment. In addition, the properties of structures refined using FINGAR edge closer to experiment as the errors associated with the restraints decrease, as is desired in a refinement method. With MD-based methods, biases due to including the restraints during MD can cause the error to increase as the associated errors decrease.

Aside from the obvious advantages in the refined ensemble, FINGAR sports additional advantages over other techniques in use. Firstly, FINGAR yields a set of weights that define the ensemble fitting the NMR data. The weights provide a straightforward map of the most important conformations, and those basis-set structures with the largest weights can be used in further modeling studies. In the most-commonly used NMR refinement methods, one can only generate an ensemble of conformations or an average conformation. While such representations are instructive, they are not nearly as readily usable for most applications.

Another advantage of FINGAR is that the most time-consuming part of the simulation – generation of the basis set – is only carried out once. After the basis set is generated, one needs only to evaluate the values of the restraints and potential energy of each structure (roughly 100 energy function evaluations for the simulations here) to run FINGAR with any restraint set. FINGAR itself, because the fitness function is calculated from these predetermined quantities, is relatively fast. This is quite advantageous as one will frequently modify the restraint set (remove erroneous restraints, change restraint bounds) and rerun the refinement, particularly in the early stages of structure refinement. By contrast, traditional DG and MD-based refinement methods can be quite time-intensive to rerun with a new restraint set. The time saving is also enjoyed when one needs to rerun to try different refinement conditions, either in terms of parameters controlling the GA, or if one wishes to change the relative weights of the terms in the fitness function. Although we have chosen to perform our refinements in the distance domain, the FINGAR method would be equally applicable to refinements directly against NOE intensities. In such a case, the time savings would be even more dramatic, because the calculation of these intensities is quite CPU-intensive (Borgias and James, 1988; Yip and Case, 1989; Nilges et al., 1991), and, again, these intensities would only need to be calculated once for each member of the basis set before running the actual GA fitting. Similarly, one can envision straightforward extension of the method to allow a weighted fit of basis structures to crystallographic data.

While other methods have been described that attempt to derive a weighted fit from an ensemble of basis sets, most have been based on local optimization methods or ad hoc reduction in the number of basis sets fit. As a result, they are limited in their ability to determine weights for the number of basis sets that may be required to generate an optimal fit (or when one cannot a priori define a relatively small basis set). A recent study (Ulyanov et al., 1995) described an alternate global optimization method which is applicable to many basis sets. This method appears to be considerably more robust than traditional local optimization methods, although not enough data were presented to assess the relative strength of the optimization procedure compared to a GA.

It should be stressed that here FINGAR has been tested for small-molecule refinement in cases where one is able to generate a decent basis set using straightforward MD sampling, followed by hierarchical clustering. Generating an appropriate basis set is the trickiest part of the method, since there is no guarantee that any particular simulation will adequately sample conformational space. The best general prescription may be to first run a number of distance-geometry simulations to generate an ensemble of structures that samples conformational space while fulfilling the restraints. Subsequently, each of these can be used as a starting point for a modest amount of molecular dynamics at 300 K, during which snapshot structures are stored for subsequent clustering and building of the basis set. Unrestrained MD simulations might be complemented by very low restraint weight time-averaged MD simulations, as was done here, to further improve sampling of conformational space. The need for restrained refinement will depend on the quality of the force field and our ability to sample conformational space using DG, MD, or whatever sampling technique is chosen. For small molecules with a limited number of rotatable torsions, one could conceivably generate starting points for the MD simulations from a rigid-body search along all these torsions. Sampling to produce a basis set can also be performed using other techniques. A recent study described an interesting minimization-based method for

generating a potential basis set of structures consistent with distance restraints (Bruschweiler et al., 1991). Unfortunately, the structures generated will then fail to reflect the natural molecular distortions appropriate for the temperature at which the NMR experiment was run. MD does not have this problem. In any case, sampling is no more an issue for FINGAR than for any other NMR refinement method.

It is also worth noting that while the clustering method used here worked quite well for the small molecules being considered, any method in which one is able to tune the number of resulting clusters would be acceptable. For the work here, it was sufficient to use the distance-matrix error, applied to all heavy atoms, as our clustering criterion. But calculating the distance-matrix error is the most CPU-intensive portion of the clustering procedure, and the time required for this calculation increases with the square of the number of atoms involved in the clustering. For larger molecules, we would need to either restrict the comparison to a smaller subset of the atoms (e.g. protein backbone atoms), or else use an alternate clustering criterion such as the rms difference. In the end, the key to the success of FINGAR is not the particular set of procedures used to sample conformational space and to generate clusters, but simply that such methods are successfully employed to generate an acceptable basis set.

We chose to use a clustering diameter that resulted in approximately 100 basis sets, because our work with the 'exact-test' simulations indicated FINGAR was capable of generating a solution in excellent accord with the true solution when the number of basis sets (fitted weight variables) was in this range – even when the number of restraints was smaller than the number of basis structures. However, even for larger basis sets, FINGAR is able to generate a good fit, provided sufficient experimental restraint data are available. Therefore, there should be no problem fitting to a larger basis set in cases where it appears desirable or necessary. Use of a larger basis set is most likely to become an issue when FINGAR is applied to bigger systems, such as proteins. In creating a larger basis set, one should try to ensure that not too much redundancy is being included in the set. Although clustering should generate a basis set of conformers that span conformational space, it is not unusual to have pairs of conformers in the basis set that are rather similar to one another. This occurs because clustering methods frequently must attempt to create boundaries in regions where no obvious boundaries exist (e.g. for an evenly distributed set of data). In such a case, two points close to one another can be assigned to different clusters if the cluster boundary runs between them. Redundancy in the basis set does not present a problem in running FINGAR, in finding a solution that best fulfills the NMR-derived criteria, or in calculating the averaged conformation or rms atomic motions. But it may somewhat blur the interpretation of the weight data.

Note that the number of restraints obtained in an experiment is not necessarily an accurate measure of how much independent data one has. Restraints, particularly those which are short-range, can sometimes impart little independent information to the refinement process. This occurs either when a restraint defines a distance which is never violated because of steric or other large energetic preferences, or when the distance defined by the restraint is already selected by other restraints in the list. For the 'exact-test' runs here, the restraints were independent, owing to the random method by which distances in the basis sets were defined. But in real experiments, one must be more careful in deciding how many basis sets are reasonably justifiable relative to the amount of information in the restraint set.

More generally, the results here provide strong evidence that FINGAR will be broadly applicable to small-molecule structure refinement. But it is less conclusive how well the method will work for large systems such as proteins. Again, this comes down to a question of whether an appropriate basis set can be generated. The method will certainly work *in principle* for proteins every bit as well as for small molecules: the GA fitting process has been demonstrated herein to be robust even for large numbers of variables. However, it will likely be challenging to generate a tractable set of basis structures, given the large number of independent variables (and hence conformations) potentially available. The number of experimental restraints for a high-resolution protein NMR structure is typically at least an order of magnitude larger than for a small molecule. Therefore, we can reasonably expect to increase the number of refined variables (basis sets) by a similar factor over the number used for a small molecule. Based on the work here, that might mean a basis set of a thousand or more structures. Whether a basis set of this size is sufficient to acceptably characterize the conformational space normally accessible to a protein remains to be tested. One can always further increase the basis set, but then approaches a situation where the refined results will be increasingly dependent on the force field itself. Note that this problem is identical to that encountered with *any* type of NMR refinement. The number of independent degrees of freedom for a protein is huge, and the number of independent NMR-derived restraints rarely overdetermines the problem. The difference is that in FINGAR one makes an explicit decision as to how many refined variables (basis sets) will be used, while in other methods one is generally limited to refining all the degrees of freedom in the system. That is, FINGAR calls explicit attention to the potential undetermination of the problem, while with other NMR refinement methods inadequacies in the the observables/variables ratio can be buried in the protocol.

Although there are minor issues to be dealt with regarding generation of the basis set, one thing is clear: where applicable, the GA-based FINGAR method presents a marked improvement over previous refinement techniques. FINGAR incorporates the ensemble averaging advantage of time-averaged MD refinement over standard (DG or MD) refinement. FINGAR also corrects several problems with time-averaged (and standard) refinement: (i) the ensemble is formed from structures with no (or very low) restraints applied, so geometric distortions are avoided; (ii) the rms atomic displacements predicted are in very good accord with experiment; (iii) large barrier heights and slow interconversion rates do not present a problem as long as the basis set is properly formed; and (iv) interpretation and utilization of the results is straightforward, since the result is a weighted ensemble of discrete structures. Furthermore, FINGAR can simultaneously fit the entire set of basis-set weights, offering a considerable advantage over previous attempts at assigning weights to discrete structures. In summary, there appear to be many reasons to suggest that the FINGAR method should be employed in future NMR-based refinements.

Acknowledgements

We thank Dr. Ajay, Dr. Christopher Lepre, and Dr. Jeff Peng for helpful discussions and for critical reading of this manuscript.

References

Altona, C. and Sundaralingam, M. (1972) *J. Am. Chem. Soc.*, **94**, 8205–8212.

Atkins, P.W. (1990) *Physical Chemistry*, Freeman, New York, NY.

Bonvin, A.M.J.J. and Brunger, A.T. (1995) *J. Mol. Biol.*, **250**, 80–93.

Bonvin, A.M.J.J. and Brunger, A.T. (1996) *J. Biomol. NMR*, **7**, 72–76.

Borgias, B.A. and James, T.L. (1988) *J. Magn. Reson.*, **79**, 493–512.

Brunger, A.T. and Karplus, M. (1991) *Acc. Chem. Reson.*, **24**, 54–61.

Brunger, A.T., Clore, G.M., Gronenborn, A.M., Saffrich, R. and Nilges, M. (1993) *Science*, **261**, 328–331.

Bruschweiler, R., Blackledge, M. and Ernst, R.R. (1991) *J. Biomol. NMR*, **1**, 3–11.

Burkert, U. and Allinger, N.L. (1982) *Molecular Mechanics, American Chemical Society Monographs*, Vol. 177, American Chemical Society, Washington, D.C.

Clore, G.M., Nilges, M., Sukumaran, D.K., Brunger, A.T., Karplus, M. and Gronenborn, A.M. (1986) *EMBO J.*, **5**, 2729–2735.

Crippen, G.M. and Havel, T.F. (1988) *Distance Geometry and Molecular Conformation*, Wiley, New York, NY.

Davies, D.B. (1978) *Prog. NMR Spectrosc.*, **12**, 135–225.

Fennen, J., Torda, A.E. and Van Gunsteren, W.F. (1995) *J. Biomol. NMR*, **6**, 163–170.

Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley, Reading, MA.

Gonzalez, C., Stec, W., Reynolds, M.A. and James, T.L. (1995) *Biochemistry*, **34**, 4969–4982.

Havel, T.F. (1990) *Biopolymers*, **29**, 1565–1585.

Karplus, M. (1959) *J. Chem. Phys.*, **30**, 11–15.

Kemmink, J. and Scheek, R.M. (1995) *J. Biomol. NMR*, **5**, 33–40.

Kessler, H., Griesinger, C., Lautz, J., Muller, A., Van Gunsteren, W.F. and Berendsen, H.J.C. (1988) *J. Am. Chem. Soc.*, **110**, 3393–3396.

Kim, S.-G., Lin, L.-J. and Reid, B.R. (1992) *Biochemistry*, **31**, 3564–3574.

Kim, Y. and Prestegard, J.H. (1990) *Protein Struct. Funct. Genet.*, **8**, 377–385.

Landis, C. and Allured, V.S. (1991) *J. Am. Chem. Soc.*, **113**, 9493–9499.

Landis, C.R., Luck, L.L. and Wright, J.M. (1995) *J. Magn. Reson.*, **109**, 44–59.

Lawson, C.L. and Hanson, R.J. (1974) *Solving Least-Squares Problems*, Prentice-Hall, New York, NY.

Lepre, C.A., Thomson, J.A. and Moore, J.M. (1992) *FEBS Lett.*, **302**, 89–96.

Massart, D.L. and Kaufman, L. (1983) *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, Wiley, New York, NY.

Nanzer, A.P., Van Gunsteren, W.F. and Torda, A.E. (1995) *J. Biomol. NMR*, **6**, 313–320.

Nikiforovich, G.V., Prakash, O., Gehrig, C.A. and Hruby, V.J. (1993) *J. Am. Chem. Soc.*, **115**, 3399–3406.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.

Nilges, M., Habazettl, J., Brunger, A.T. and Holak, T.A. (1991) *J. Mol. Biol.*, **219**, 499–510.

Pearlman, D.A. and Kollman, P.A. (1991) *J. Mol. Biol.*, **220**, 457–479.

Pearlman, D.A. (1994a) *J. Biomol. NMR*, **4**, 1–16.

Pearlman, D.A. (1994b) *J. Biomol. NMR*, **4**, 279–299.

Pearlman, D.A., Case, D.A., Caldwell, J.C., Ross, W.S., Cheatham III, T.E., DeBolt, S., Ferguson, D.M., Seibel, G.L. and Kollman, P.A. (1995) *Comput. Phys. Commun.*, **91**, 1–41.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1989) *Numerical Recipes*, Cambridge University Press, New York, NY.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York, NY.

Scheek, R.M., Van Nuland, N.A.J., De Groot, B.L. and Amadei, A. (1995) *J. Biomol. NMR*, **5**, 106–111.

Schmitz, U., Ulyanov, N.B., Kumar, A. and James, T.L. (1993) *J. Mol. Biol.*, **234**, 373–389.

Schreiber, S.L. (1991) *Science*, **251**, 283–287.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1989) *Chem. Phys. Lett.*, **157**, 289–294.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Torda, A.E., Brunne, R.M., Huber, T., Kessler, H. and Van Gunsteren, W.F. (1993) *J. Biomol. NMR*, **3**, 55–66.

Tropp, J. (1980) *J. Chem. Phys.*, **72**, 6035–6043.

Ulyanov, N.B., Schmitz, U., Kumar, A. and James, T.L. (1995) *Biophys. J.*, **68**, 13–24.

Van de Ven, F.J.M. and Hilbers, C.W. (1988) *Eur. J. Biochem.*, **178**, 1–38.

Van Gunsteren, W.F., Brunne, R.M., Gros, P., Van Schaik, R.C., Schiffer, C.A. and Torda, A.E. (1994) *Methods Enzymol.*, **239**, 619–654.

Wagner, G. and Wüthrich, K. (1979) *J. Magn. Reson.*, **33**, 675–680.

Wüthrich, K. (1990) *J. Biol. Chem.*, **265**, 22059–22062.

Yang, J.-X. and Havel, T.F. (1993) *J. Biomol. NMR*, **3**, 355–360.

Yip, P. and Case, D.A. (1989) *J. Magn. Reson.*, **83**, 643–648.