Simulated annealing with restrained molecular dynamics using a flexible restraint potential: Theory and evaluation with simulated NMR constraints

DONNA BASSOLINO-KLIMAS,¹ ROBERTO TEJERO,^{2,3} STANLEY R. KRYSTEK,¹
WILLIAM J. METZLER,¹ GAETANO T. MONTELIONE,² AND ROBERT E. BRUCCOLERI¹

¹ Department of Macromolecular Structure, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543

² Center for Advanced Biotechnology and Medicine, and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey, 08854-5638

³ Departamento de Química Física, Facultad de Ciencias Químicas, Universidad de Valencia, Dr. Moliner, 50, 46100-Burjassot (Valencia) Spain

(RECEIVED September 19, 1995; ACCEPTED January 5, 1996)

Abstract

A new functional representation of NMR-derived distance constraints, the flexible restraint potential, has been implemented in the program CONGEN (Bruccoleri RE, Karplus M, 1987, *Biopolymers* 26:137–168) for molecular structure generation. In addition, flat-bottomed restraint potentials for representing dihedral angle and vicinal scalar coupling constraints have been introduced into CONGEN. An effective simulated annealing (SA) protocol that combines both weight annealing and temperature annealing is described. Calculations have been performed using ideal simulated NMR constraints, in order to evaluate the use of restrained molecular dynamics (MD) with these target functions as implemented in CONGEN. In this benchmark study, internuclear distance, dihedral angle, and vicinal coupling constant constraints were calculated from the energy-minimized X-ray crystal structure of the 46-amino acid polypeptide crambin (ICRN). Three-dimensional structures of crambin that satisfy these simulated NMR constraints were generated using restrained MD and SA. Polypeptide structures with extended backbone and side-chain conformations were used as starting conformations. Dynamical annealing calculations using extended starting conformations and assignments of initial velocities taken randomly from a Maxwellian distribution were found to adequately sample the conformational space consistent with the constraints. These calculations also show that loosened internuclear constraints can allow molecules to overcome local minima in the search for a global minimum with respect to both the NMR-derived constraints and conformational energy. This protocol and the modified version of the CONGEN program described here are shown to be reliable and robust, and are applicable generally for protein structure determination by dynamical simulated annealing using NMR data.

Keywords: coupling constants; energy minimization; solution NMR structure determination

It is now well established that three-dimensional structures of small proteins and oligonucleotides can be determined using a combined approach of multidimensional NMR spectroscopy and computational techniques (Wüthrich, 1986; Clore & Gronenborn, 1991, 1994). Recent advances in NMR methodology make it possible to extract a large amount of structural information,

including estimates of interproton distances and vicinal scalar coupling constants, with good precision. Computer programs employing a variety of algorithms have been developed to utilize this information for structure determination. These include distance space methods such as metric-matrix distance geometry (DG) (Havel & Wüthrich, 1984, 1985; Hare & Reid, 1986; Lautz et al., 1989) and real space methods such as variable target function algorithms in dihedral angle space (Braun & Gö, 1985), restrained molecular dynamics (MD) (Kaptein et al., 1985; Brünger et al., 1986), dynamic simulated annealing (SA) (Nilges et al., 1988), and Monte-Carlo (MC) calculations (Bassolino, 1989; Levy et al., 1989; Ulyanov et al., 1993). The DG methods are purely mathematical procedures, whereas the real space methods

Reprint requests to: Gaetano T. Montelione, Center for Advanced Biotechnology and Medicine, and Department of Molecular Biology and Biochemistry, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey, 08854-5638; e-mail: montelione@nmrlab.cabm.rutgers.edu.

Abbreviations: Conf. E., total conformational energy; DG, distance geometry; EL E., electrostatic energy; MC, Monte Carlo; MD, molecular dynamics; SA, simulated annealing; VDW E., van der Waals energy.

include the minimization of a target function that includes both NMR-derived constraints and some form of potential energy function. In general, structures generated with real space methods result in more energetically correct final structures than structures generated exclusively with DG methods. In addition, DG calculations sometimes exhibit incomplete sampling of the solution space of conformations (Metzler et al., 1989), which can be improved by randomized metrization (Havel, 1991a, 1991b). For these reasons, hybrid approaches are often used (see for example Bassolino et al., 1988; Nilges et al., 1988; Havel, 1991b), in which structures generated with DG methods are further refined using one of the real space methods.

Each of these approaches has advantages and disadvantages. The principal advantage of DG methods for small structures is speed. Using modest computer power, literally hundreds of structures can be generated in a few hours. However, for larger structures, DG calculations can be very slow. In either case, the resulting DG structures generally have high conformational energies and usually must be refined further using some form of energy minimization (Bassolino et al., 1988; Nilges et al., 1988; Havel, 1991a, 1991b). MD and SA calculations require much more computer time, especially when generating structures from extended starting conformations. There is also always the question of whether the structures are biased by the choice of starting structure. Another crucial issue in structure generation calculations is convergence. In many cases, less than 50% of the starting structures converge to acceptable final structures that satisfy the NMR-derived constraints and also exhibit low conformational energies.

Although most internuclear distance constraints come from NOESY data, additional information in the form of scalar coupling constant data can be obtained from other types of NMR experiments (Griesinger et al., 1986; Montelione et al., 1989, 1992; Bax et al., 1992; Griesinger & Eggenberger, 1992; Biamonti et al., 1994) and used to determine conformations of polypeptide dihedral angles. Scalar coupling constants are also useful for making stereospecific resonance assignments (Güntert et al., 1989, 1991; Nilges et al., 1990) and confirming secondary structural features (Pardi et al., 1984; Wüthrich, 1986). However, few programs are designed to refine directly against the scalar coupling constant data (Kim & Prestegard, 1990; Torda et al., 1993; Garrett et al., 1994; Constantine et al., 1995a). Most structure generation programs (e.g., EMBED, DG, DG-II, DISMAN, DIANA, and X-PLOR) require converting the scalar coupling data into dihedral angle constraints via Karplus relations (Karplus, 1959; Bystrov, 1976; Pardi et al., 1984). This conversion is complicated by the fact that these Karplus equations are multivalued. Conversion to dihedral angles generally requires choosing one value based either on the most probable conformation from an analysis of Ramachandran-type plots, or by combined analysis of NOE and scalar coupling constant data (Güntert et al., 1989, 1991; Nilges et al., 1990).

CONGEN is a program for molecular structure generation (Brucoleri & Karplus, 1987) that uses several different conformational search algorithms including MD with the CHARMM (Brooks et al., 1983) potential energy function. In this paper, we describe the incorporation of new restraint penalty functions for representing NMR-derived distance, vicinal scalar coupling, and other dihedral angle constraints in CONGEN calculations. Using these flat-bottomed restraint penalty functions, it is possible to incorporate all of the experimental NMR data into re-

strained MD calculations. Other improvements over existing programs (e.g., XPLOR [Brünger, 1992] and GROMOS [Brunne et al., 1993]) for computing NMR structures by SA with MD include the use of a highly flexible and easily tailored penalty function for representing internuclear distance constraints, and the ability to scale terms in the potential function between specific atom pairs by different amounts. These features allow the user to control weights on individual constraints throughout the annealing procedure. The specific form of the distance restraint potential is designed to prevent large constraint violations from dominating the MD trajectory. In addition, there is a target function for restraining dihedral angles directly and another target function that allows the structure to be refined directly against vicinal coupling constant data without conversion into dihedral angle constraints. There is also the flexibility to scale any term in the potential energy function independently. These new features of the CONGEN program are straightforward to use. We have also developed SA-MD protocols with good convergence power.

The effectiveness of the new target functions and SA protocols are demonstrated using the high-resolution X-ray crystal structure of crambin as a benchmark system.⁴ Structures were generated using a simple protocol with various types of constraints. The effects of different starting structures on the convergence frequency and quality of structures have been evaluated. Good sampling of conformational space was obtained in calculations starting from extended polypeptide conformations with atomic velocities (i.e., speeds and directions) assigned randomly from a Maxwellian distribution. A detailed statistical analysis of the generated structures is presented.

Results

Computational methods: Theory

New penalty functions have been added to the program CONGEN (Brucoleri & Karplus, 1987), allowing use of NMR-derived constraints in restrained MD and conformational search calculations. These functions include an NOE-penalty function, a dihedral angle penalty function, and a vicinal coupling constant penalty function. These penalty functions are combined with the CHARMM (Brooks et al., 1983) potential energy function to form a composite optimization function for protein and polypeptide structure generation.

A flat-bottomed NOE-penalty function is used to restrain interatomic distances to ranges defined by upper- and lower-bound interatomic distances estimated from NOE data. This penalty function, shown in Figure 1, is given by:

$$E_{\text{NOE}} = \begin{cases} K_{\text{NOE}}(r_{lb} - r_{ij})^2 & \text{for (I) } r_{ij} < r_{lb} \\ 0.0 & \text{for (II) } r_{lb} \leq r_{ij} \leq r_{ub} \\ K_{\text{NOE}}(r_{ij} - r_{ub})^2 & \text{for (III) } r_{ub} < r_{ij} < r_{sw} \\ -K_{\text{NOE}}(r_{ij} - r_{plat})^2 + 2E_{sw} & \text{for (IV) } r_{sw} < r_{ij} < r_{out} \\ (K_{\text{NOE}} * F_{slope}) + E_{out} & \text{for (V) } r_{ij} < r_{out} \end{cases} \quad (1)$$

⁴ All of the structures computed in this study, together with the corresponding constraint files, will be deposited in the Brookhaven Protein Data Bank for use in benchmarking future versions of CONGEN and related programs.

Fig. 1. A flexible NOE-penalty function. Drawing of the E_{NOE} term versus distance as implemented in CONGEN. The function is harmonic in region I, and flat-bottomed in region II between the lower-bound (r_{lb}) and upper-bound (r_{ub}) distances. The upper-bound restraint begins as a harmonic function in region III, but flattens in regions IV and V.

where r_{ij} is the r^{-6} -weighted average interproton distance, and r_{lb} and r_{ub} are the lower- and upper-bound distance constraints, respectively. The Roman numerals correspond to the labeled regions of Figure 1. The parameter r_{ij} is determined from the n proton-proton interactions contributing to the intensity of the measured NOE using the following equation:

$$r_{ij} = \left[\sum_{k=1}^n \left(\frac{1}{r_{ij,k}} \right)^6 \right]^{-1/6}. \quad (2)$$

This allows for proper evaluation of NOEs involving degenerate protons, which are designated in the constraint input files by H*. A detailed summary of considerations used in generating distance constraints involving degenerate protons from experimental NOE data is presented in the companion paper (Tejero et al., 1996). The parameter r_{sw} is determined from the upper-bound of each distance constraint (r_{ub}), the weighting factor (K_{NOE}), and the maximum force (F_{max}) as follows:

$$r_{sw} = \frac{F_{\text{max}}}{2.0 * K_{\text{NOE}}} + r_{ub}, \quad (3)$$

where F_{max} and K_{NOE} are user-defined parameters assigned individually to each constraint. Thus, r_{sw} is different for each constraint. E_{sw} is the energy at r_{sw} (see Fig. 1), the point between regions III and IV at which the harmonic function switches sign in a transition to the constant force region V. The parameter r_{out} is the outer-bound cutoff, which is determined from r_{plat} [the point where $-K_{\text{NOE}}(r_{ij} - r_{ub})^2$ reaches a plateau value] as follows:

$$r_{plat} = (2 * r_{sw}) - r_{ub}$$

$$r_{out} = r_{plat} - \frac{F_{\text{slope}}}{2.0 * K_{\text{NOE}}}, \quad (4)$$

where F_{slope} is the user-defined slope for the outer region (region V) and E_{out} is the energy at the point r_{out} . The points r_{plat} and r_{out} are chosen (for each constraint) such that the slope between these two points is equal to F_{slope} , in order to assure that the function is continuous and differentiable at r_{out} . This NOE-penalty function is extremely flexible in that the parameters F_{max} , F_{slope} , K_{NOE} , r_{lb} , r_{ub} can be defined independently for each constraint.

This form of the NOE-penalty function has the advantage of allowing the forces from small distance constraint violations to dominate the dynamics, whereas contributions from large constraint violations are attenuated. For large constraint violations, there is only a weak force driving atoms together. In addition, by providing larger forces for the smaller constraint violations, the penalty function allows local structure to form first, followed by increasingly longer-range structure. These effects are controlled by adjusting the values of F_{max} and F_{slope} at each stage of the SA procedure. By using small values of F_{max} and F_{slope} in the initial stages of the protocol, constraints that have extremely large violations initially do not play a significant role in driving the early stages of the dynamics. Then, as the simulation progresses, F_{max} and F_{slope} are increased until all the constraints are satisfied. It may be possible to configure other software packages to use similar flexible penalty functions, but it is not commonly done. Further details on choosing these adjustable parameters during the annealing process are provided in the Methods.

A flat-bottomed penalty function was also developed to provide direct constraints on dihedral angles:

$$E(\Theta) = \begin{cases} K_{\text{dihed}}(\Theta_{lb} - \Theta)^2 & \text{for } \Theta < \Theta_{lb} \\ 0.0 & \text{for } \Theta_{lb} \leq \Theta \leq \Theta_{ub} \\ K_{\text{dihed}}(\Theta - \Theta_{ub})^2 & \text{for } \Theta > \Theta_{ub} \end{cases} \quad (5)$$

where the origin of the dihedral angle difference is set to the midpoint between Θ_{lb} and Θ_{ub} . This choice assumes that the maximum in $E(\Theta)$ is 180° opposite to the center of the minimum of $E(\Theta)$. In this study, the dihedral angle penalty function is used first to keep the ω peptide dihedral angles *trans* and planar at high temperatures and then for the refinement of structures using direct constraints on backbone dihedral angles phi and psi. It can also be used for generating starting structures that already satisfy dihedral angle constraints (Güntert et al., 1989; Nilges et al., 1990) or for refinement against dihedral angle constraints derived from carbon-13 chemical shift data (Celda et al., 1995).

A CONGEN target function has also been developed to include vicinal coupling constant data in a completely general way. These coupling constants are related to the dihedral angles ϕ by Karplus relationships (Karplus, 1959; Bystrov, 1976):

$$^3J(\phi) = C_1 \cos^2 \phi - C_2 \cos \phi + C_3. \quad (6)$$

Although any possible Karplus curve can be used in the CONGEN program, in this paper only $^3J(\text{H}^N\text{-H}^\alpha)$ coupling constants were tested in structure calculations. In this case, ϕ is the dihedral angle defined by $\text{H}^\alpha\text{-C}^\alpha\text{-N-H}^N$ and $C_1 = 6.4 \text{ Hz}$, $C_2 =$

1.4 Hz, and $C_3 = 1.9$ Hz (Pardi et al., 1984). By redefining these coefficients in the input to CONGEN, appropriate Karplus curves can be defined for a wide range of vicinal homo- and heteronuclear coupling constants. This feature allows the user to customize the J-restraint target function as new or improved Karplus relationships are developed. In addition, the capacity to use vicinal coupling constants involving prochiral atoms is available (Constantine et al., 1995a).

In refining against the vicinal scalar-coupling data, the target function:

$$E_{J-c}(J) = \begin{cases} K_{J-c}(J_{lb} - J)^2 & \text{for } J < J_{lb} \\ 0.0 & \text{for } J_{lb} \leq J \leq J_{ub} \\ K_{J-c}(J_{ub} - J)^2 & \text{for } J_{ub} < J \end{cases} \quad (7)$$

allows coupling constant data to be incorporated directly into the structure determination process. The flat-bottomed harmonic form of this penalty function weights equally the entire range of uncertainties in the estimated values of coupling constants, and is therefore distinct from the simple harmonic penalty functions that have been described for other programs (Kim & Prestegard, 1990; Torda et al., 1993; Garrett et al., 1994).

Another feature of CONGEN is the ability to scale any term in the potential energy function independently. In these calculations, the van der Waals and electrostatic terms were scaled down during the initial annealing process. In this "weight annealing" process (see the Methods), weights on the NMR constraint terms are increased, whereas weights on the conformational energy terms are scaled down, allowing the trajectories to evolve primarily under the influence of the NMR constraints. Weight annealing prior to the temperature annealing process was found to provide improved sampling of solution space and to yield structures at the end of the SA process that better satisfied simulated or real NMR constraints (see also Tejero et al., 1996).

The constraints

The high-resolution crystal structure of crambin (ICRN; Hendrickson & Teeter, 1981) refined to 1.5 Å resolution was chosen as a protein structure likely to be near the global energy minimum of the CHARMM potential function. Ideal NOE-type distance constraints were derived from the CONGEN energy-minimized crystal structure of crambin to which hydrogens were added (see the Methods). This "target structure" has van der Waals energy and total conformational energy of -145.0 and -470.9 kcal/mol, respectively. The conformational energy is the sum of all the terms in the potential excluding any constraint terms (see the Methods). This set of 1,657 exact distance constraints, the EXACT set, is summarized in Table 1. Intra-residue distance constraints were excluded from this input constraint file (see Table 1). A second constraint set, RANGES, included only 514 constraints; i.e., 11.2 constraints per residue. This list, meant to simulate distance constraints ≤ 5 Å extracted from high-quality NOESY data, was generated from the EXACT set assuming a $\pm 20\%$ uncertainty in the experimental distance measurement, as described in the Methods. A survey of the distribution of constraints shows that they are distributed uniformly over the whole molecule. The $^3J(\text{H}^N\text{-H}^\alpha)$

constraints were back-calculated from the structure using Equation 6, and upper and lower bounds were computed by adding ± 0.2 Hz to these target values. Backbone dihedral angle constraints were also derived from the structure and the corresponding upper- and lower-bound constraints obtained assuming $\pm 5^\circ$ uncertainty.

Structure generation and analysis

Two sets of 10 crambin structures each were generated from an extended starting structure using either the EXACT or the RANGES constraint sets. These two ensembles of structures are referred to as the extended-EXACT (EE) and extended-RANGES (ER) families, respectively. During the initial weight annealing period, restraint terms were scaled up and energy terms representing van der Waals and electrostatic interactions were scaled down in order to allow the chains to move more freely and to increase the rate of convergence. Weights on these conformational energy terms were then increased back to standard values as the temperature was reduced to 300 K. In the EE set, the maximum value of K_{NOE} was 20 kcal/mol-Å², whereas for the ER set, the maximum value of K_{NOE} was 80 kcal/mol-Å² (see the Methods).

For each class of starting structures, the protocol was repeated until 10 structures converged. A converged structure is defined as one with no violations greater than 0.2 Å. It was found that structures either satisfied the constraints reasonably well (i.e., no violations greater than about 0.2 Å), or had many large violations (sometimes greater than 3.0 Å), and large positive van der Waals energy terms. Structures that did not converge often exhibit local inversion of covalent chirality, whereas converged structures do not. For both sets of constraints, the rate of convergence was on the order of 50–60%.

Structures generated with EXACT constraints

The high quality of the EE crambin structures is indicated by the structural statistics summarized in Table 2, including resid-

Table 1. Summary of input data for crambin structure calculations

Data set	EXACT	RANGES
Total number of distance constraints ^a	1,657	514
Intraresidue constraints [$i = j$]	0	0
Interresidue constraints [$(i - j) = 1$]	552	177
Backbone-backbone constraints	53	70
Backbone-side-chain constraints	97	18
Side-chain-side-chain constraints	402	89
Interresidue constraints [$(i - j) > 1$]	575	148
Backbone-backbone constraints	13	49
Backbone-side-chain constraints	161	34
Side-chain-side-chain constraints	401	65
Interresidue constraints [$(i - j) > 5$]	530	189
Dihedral angle constraints	78	78
Scalar coupling constraints $^3J(\text{H}^N\text{-H}^\alpha)$	39	39

^a Each constraint corresponds to a pair of constraining values, one upper-bound distance (dihedral or coupling) constraint and one lower-bound distance (dihedral or coupling) constraint. In the case of exact constraints, the upper- and lower-bound values are identical.

Table 2. Summary of structural parameters for the target and two sets of calculated crambin structures

	RMS distance constraint violation (Å)	RMSDs from idealized geometry			
		Bond lengths (Å)	Bond angles (deg)	Peptide bonds (deg) (impropers)	Peptide bonds (deg) (omega)
Target (energy-minimized ICRN PDB coordinates)	0.00 (0.00) ^a	0.007 (0.034)	0.68 (3.92)	0.221 (0.87)	7.29 (16.67)
EE —Extended EXACT (final weight K_{NOE} of 20 kcal/mol-Å ²)	0.01 (0.12)	0.008 (0.057)	1.45 (6.50)	2.36 (7.27)	2.39 (3.97)
ER —Extended RANGES (final weight K_{NOE} of 80 kcal/mol-Å ²)	0.11 (0.38)	0.009 (0.05)	1.86 (13.2)	2.19 (8.32)	2.10 (3.06)

^a Values in parentheses are the maximum deviations and/or violations and include bond length and bond angle violations involving hydrogen atoms.

ual RMS deviations (RMSDs) of distance constraints and of bond lengths, bond angles, and peptide bond geometries from ideal values. The deviations from idealized geometries for the calculated structures are similar to those reported for the target X-ray structure (ICRN) and those reported for a recently described 0.8-Å resolution crystal structure of crambin (Teeter et al., 1993). In general, the crambin structures generated from extended starting structures (e.g., set EE) have few significant deviations from idealized geometries. The standard geometry deviations summarized in Table 2 include statistics for bond lengths and bond angles involving hydrogen atoms. The *maximum* bond length and bond angle violations, shown in parentheses in Table 2, are structural parameters that are not usually reported in structure determination papers. Analysis of other refined NMR structures available in the Brookhaven Protein Data Bank (PDB) show maximum bond angle violations of similar magnitudes to those reported here.⁵

A summary of the residual violations and final energies for the EE structures is shown in Table 3. The structures have very low energies and almost no residual violations. In fact, the calculated structures have energies lower than that of the target structure. The few small violations remaining are uniformly distributed along the peptide sequence. Table 4 shows the atomic RMSDs of the generated structures. Statistics for structure superpositions were computed using mathematically averaged coordinates as reference structures. The structures generated have very small RMSDs compared to the average structure (~0.3 Å for all heavy atoms), indicating that the structures in the set are very similar to one another. Average atomic RMSDs of each structure relative to the target structure are also reported in

Table 4. These RMSDs are also small; on the order of ~0.3 Å (for all heavy atoms). Although they exhibit good accuracy and precision, even using an extensive set of exact distance constraints, the calculated structures do not exactly match the target structure. These calculations provide a benchmark for the best we can expect using these SA protocols with ideal data devoid of noise or inconsistencies due to inaccuracies, imprecisions, and dynamical features of real experimental data.

Structures generated with RANGES constraints

Tables 2 and 3 also summarize the structural statistics, residual violations, and energies for the ER structures. Overall, the quality of these ER structures is similar to those of the EE structures. The convergence rate for these ER structures was greater than 60%. As can be seen in Table 3, the ER protocol also results in structures with lower energies than those obtained with any of the other protocols. In particular, protocol ER, using fewer loosened constraints, results in structures with final van der Waals and conformational energies significantly lower than those obtained with protocol EE using EXACT distances. These structures are more “relaxed” energetically, although they exhibit higher RMSDs, relative both to the average structure in this ensemble and to the target structure (Table 4).

Stereo drawings showing all heavy atoms of the 10 EE and ER crambin structures, superimposed on the target structure, are shown in Figure 2. All the structures are visually similar to the target structure. Even when a large number (1,657) of very tight constraints are considered all at once, the dynamical SA protocols described in the Methods generate reasonably good structures. These CONGEN protocols also work well using the RANGES constraint set, representative of what could be obtained from carefully interpreted high-resolution NOESY data.

Structure refinement with vicinal-coupling and dihedral angle constraints

In order to test the consistency of the distance, J, and dihedral angle restraint functions, each of the structures summarized in Table 3 was further refined by simple minimization of the energy and distance-constraint target function supplemented by an additional 39 ³J(H^N-H^α) and 78 dihedral angle constraints. Structural statistics for these refined EE and ER structures are

⁵ For example, in the X-ray crystal structures of crambin (ICRN) (Hendrickson & Teeter, 1981) and bovine pancreatic trypsin inhibitor (4PTI) (Marquart et al., 1983), the RMS (and maximum) bond angle deviations for all heavy atoms are 2.79° (11°) and 2.73° (10°), respectively. For the crystal structure 5PTI of bovine pancreatic trypsin inhibitor with hydrogen atoms (Wlodawer et al., 1984), the RMS (and maximum) bond angle deviations for all heavy and hydrogen atoms are 5.44° (33°). In the solution NMR structures of hirudin (2H1R, 32 structures) (Folkers et al., 1989) and for an antihypertensive antiviral protein (2BDS, 42 structures) (Driscoll et al., 1989), the RMS (and maximum) bond angle deviations for all heavy and hydrogen atoms are 2.39° (33°) and 2.84° (32°), respectively.

Table 3. Summary of violations and final energies for two sets of CRAMBIN structures

	1	2	3	4	5	6	7	8	9	10
EE —Extended EXACT (final weight K_{NOE} of 20 kcal/mol-Å ²)										
>1.0 Å	0	0	0	0	0	0	0	0	0	0
0.5–1.0 Å	0	0	0	0	0	0	0	0	0	0
0.3–0.5 Å	0	0	0	0	0	0	0	0	0	0
0.2–0.3 Å	0	0	0	0	0	0	0	0	0	0
0.1–0.2 Å	1	1	0	1	6	1	1	1	1	1
<0.1 Å	28	30	24	29	43	28	28	30	28	33
Total	29	31	24	30	49	29	29	31	29	34
VDW E. ^a	−167.2	−168.3	−167.4	−168.4	−163.7	−167.6	−168.0	−167.8	−167.8	−171.3
EL E. ^a	−377.9	−379.8	−377.8	−372.5	−382.6	−372.0	−377.4	−377.4	−371.9	−371.0
Conf. E. ^a	−518.1	−519.1	−518.6	−509.1	−515.6	−510.1	−519.1	−517.1	−509.8	−514.1
ER —Extended RANGES (final weight K_{NOE} of 80 kcal/mol-Å ²)										
>1.0 Å	0	0	0	0	0	0	0	0	0	0
0.5–1.0 Å	0	0	0	0	0	0	0	0	0	0
0.3–0.5 Å	0	0	0	0	0	0	0	0	0	0
0.2–0.3 Å	0	0	0	0	0	0	0	0	0	0
0.1–0.2 Å	0	0	0	0	0	0	0	0	0	0
<0.1 Å	35	34	28	33	32	23	29	20	20	25
Total	35	34	28	33	32	23	29	20	20	25
VDW E. ^a	−185.4	−179.9	−176.5	−182.0	−174.0	−181.3	−179.8	−179.0	−181.4	−181.0
EL E. ^a	−519.3	−520.2	−535.9	−506.4	−525.0	−510.4	−519.0	−521.3	−516.8	−505.9
Conf. E. ^a	−695.2	−691.2	−698.0	−676.2	−691.0	−682.9	−683.8	−691.4	−693.0	−683.6

^a All energies are in units of kcal/mol.

presented in Table 5, and the residual constraint violations and final energies are summarized in Table 6. The RMSDs to the average structure and the target structure are presented in Table 4. Comparing Tables 3 and 6 shows that the CONGEN penalty functions for the relatively tight dihedral and J constraints are highly consistent with the penalty function for distance constraints. After optimizing the structures to also satisfy the dihedral and J constraints, the EE and ER structures exhibit similar distributions of residual distance constraint violations and energies only 1–5 kcal/mol higher than structures calculated without dihedral and J constraints. This result demonstrates self-consistency of the individual distance, dihedral, and J restraint

components of the composite penalty function. The absence of large violations of dihedral and J constraints also indicates that there are no local mirror-image conformations in these sets of EE and ER structures. Table 4 shows that the RMSDs to the average and target structures do not change much after refinement, a further indication of the self-consistency of the components of the penalty function.

Discussion

Benchmark studies of structure generation algorithms using simulated data are essential for establishing their inherent accuracy,

Table 4. Atomic RMSDs for superpositions of crambin structures^a

Starting conformation	Distance constraint type	RMSD for backbone atoms (Å)		RMSD for all heavy atoms (Å)	
		Average	Target	Average	Target
EE	EXACT	0.01 (0.01–0.02)	0.11 (0.10–0.18)	0.30 (0.24–0.38)	0.31 (0.15–0.38)
ER	RANGES	0.33 (0.24–0.41)	0.51 (0.43–0.67)	0.53 (0.41–0.61)	0.81 (0.69–0.93)
EE (Refined)	EXACT	0.01 (0.01–0.02)	0.067 (0.06–0.07)	0.1 (0.16–0.28)	0.28 (0.15–0.37)
ER (Refined)	RANGES	0.33 (0.24–0.41)	0.51 (0.40–0.67)	0.53 (0.41–0.61)	0.82 (0.69–0.93)

^a Atomic RMSDs for superpositions of residues 1–46 relative to the average and energy-minimized X-ray crystal (target) structures. The average structure was obtained by averaging the Cartesian coordinates after optimally superimposing the set of structures in each family. The target structure was obtained by energy minimizing with CONGEN the crystal coordinates for crambin (1CRN from the PDB).

Fig. 2. Stereo drawing of all heavy atoms for 10 calculated crambin structures (blue) superimposed on the target structure (red) for the (A) extended-EXACT ensemble and the (B) extended-RANGES ensemble.

precision, and reliability for interpreting macromolecular NMR data (Braun & Gö, 1985; Havel & Wüthrich, 1985; Bassolino et al., 1988; Nilges et al., 1988; Metzler et al., 1989; Havel, 1991b; Suri & Levy, 1993). In this work, we used the high-resolution X-ray crystal structure of crambin as a target in testing new CONGEN restraint functions, improved SA protocols, and different choices of starting conformations. These calculations demonstrate the high reliability and robustness of CON-

GEN for NMR structure analysis, and provide the basis for reliable use of CONGEN in analysis of real NMR data. Indeed, the implementation and SA protocols described here have already been used to determine solution NMR structures of several small proteins including epidermal and type- α transforming growth factors (Tejero et al., 1996), the Z-domain of staphylococcal protein A (Jendeborg et al., 1996), and the major cold shock protein from *Escherichia coli* (Newkirk et al., 1994). The

Table 5. Summary of structural parameters for the two sets of crambin structures refined with distance, dihedral angle, and vicinal coupling constraints, compared to the target structure

	RMS distance constraint violation (Å)		RMSDs from idealized geometry			
			Bond lengths (Å)	Bond angles (deg)	Peptide bonds (deg) (impropers)	Peptide bonds (deg) (omega)
Target (minimized 1CRN PDB coordinates)	0.00	(0.00) ^a	0.007 (0.034)	0.68 (3.92)	0.22 (0.87)	7.29 (16.67)
EE—Extended EXACT (final weight K_{NOE} of 20 kcal/mol-Å ²)	0.02	(0.12)	0.008 (0.058)	1.46 (9.50)	2.10 (7.43)	2.42 (4.20)
ER—Extended RANGES (final weight K_{NOE} of 80 kcal/mol-Å ²)	0.004	(0.08)	0.008 (0.052)	1.86 (13.12)	2.12 (7.92)	2.09 (3.18)

^a Values in parentheses are the maximum deviations and/or violations and include bond lengths and bond angles involving hydrogen atoms.

Table 6. Summary of violations and final energies for two sets of CRAMBIN structures after refinement with distance, scalar coupling, and dihedral angle constraints

	1	2	3	4	5	6	7	8	9	10
EE—Extended EXACT (final weight K_{NOE} of 20 kcal/mol-Å ²)										
>1.0 Å	0	0	0	0	0	0	0	0	0	0
0.5–1.0 Å	0	0	0	0	0	0	0	0	0	0
0.3–0.5 Å	0	0	0	0	0	0	0	0	0	0
0.2–0.3 Å	0	0	0	0	0	0	0	0	0	0
0.1–0.2 Å	1	1	0	1	5	1	1	1	1	1
<0.1 Å	25	26	24	28	42	26	25	28	25	31
Total	26	27	24	29	47	27	26	12	26	33
>20°	0	0	0	0	0	0	0	0	0	0
>10°	0	0	0	0	0	0	0	0	0	0
>5°	0	0	0	0	0	0	0	0	0	0
>0.5 Hz	0	0	0	0	0	0	0	0	0	0
VDW E. ^a	−167.3	−168.3	−167.1	−168.2	−163.6	−167.5	−167.8	−167.5	−167.7	−171.0
EL E. ^a	−377.0	−378.8	−377.0	−371.2	−380.8	−370.3	−375.9	−375.9	−370.3	−369.4
Conf. E. ^a	−517.5	−518.0	−517.5	−507.5	−513.6	−509.9	−516.9	−510.0	−507.5	−511.8
ER Extended RANGES (final weight K_{NOE} of 80 kcal/mol-Å ²)										
>1.0 Å	0	0	0	0	0	0	0	0	0	0
0.5–1.0 Å	0	0	0	0	0	0	0	0	0	0
0.3–0.5 Å	3	0	0	0	0	0	0	0	0	0
0.2–0.3 Å	0	0	0	0	0	0	0	0	0	0
0.1–0.2 Å	0	0	0	0	0	0	0	0	0	0
<0.1 Å	31	34	26	33	30	22	27	18	21	20
Total	34	34	26	33	30	22	27	18	21	20
>20°	0	0	0	0	0	0	0	0	0	0
>10°	0	0	0	0	0	0	0	0	0	0
>5°	1	1	1	2	1	1	2	2	1	2
>0.5 Hz	1	0	0	0	0	0	0	0	0	0
VDW E. ^a	−185.7	−180.5	−176.7	−183.6	−174.7	−181.5	−181.5	−179.1	−182.0	−181.0
EL E. ^a	−517.8	−516.3	−531.0	−498.7	−519.3	−508.1	−509.6	−515.1	−512.7	−501.8
Conf. E. ^a	−692.0	−685.5	−690.8	−666.8	−682.4	−678.0	−673.3	−682.9	−686.3	−677.1

^a All energies are in units of kcal/mol.

results presented in this paper provide the basis for reliable use of CONGEN in future studies of protein structure by NMR.

Using ideal EXACT distance constraints allowed us to identify and correct shortcomings in the SA protocols and restraint functions that would not have been obvious in calculations using less precise real NMR data. For example, although some efforts were made to generate structures from random or pseudo-random starting conformations, these calculations exhibited poor convergence and the resulting structures were in poorer agreement with the target structure than structures generated from extended starting conformations. Accordingly, all of the structures described in these benchmark tests are initiated from extended conformations with randomly assigned initial velocities. In related work using real NMR distance constraints (Tejero et al., 1996), CONGEN structures generated from extended starting conformations (with randomly assigned initial velocities) also exhibit fewer constraint violations and lower energies than structures generated using a hybrid DG/SA algorithm.

A second important result of these benchmark calculations is the observation that, even for mutually self-consistent distance constraints representing a single static structure, there are advantages in loosening the constraints from exact values into ranges of values, while simultaneously penalizing small violations of these loosened constraints using a relatively high value

of K_{NOE} , e.g., 80 kcal/mol-Å². Constraint loosening results in structures that do not violate the exact constraint list significantly and are well-clustered around the target structure, but that exhibit lower values of VDW E and Conf. E than structures computed with exact constraints. Similar results were obtained using real NMR data (Tejero et al., 1996). This observation is consistent with the fact that structures generated in the EE and ER calculations have lower conformational energies than the target structure, indicating some inconsistency between the CHARMM potential energy function used in CONGEN (which does not include solvent effects explicitly) and the crystal structure of crambin. The extended ER structures also exhibited the highest convergence rates (i.e., >60%) of all the tested protocols. These results demonstrate that widened flat-bottomed penalty functions with steep walls can have advantages for restrained energy-optimization calculations when the true structure (or the structure represented by the constraints) does not correspond exactly to the global minimum of the potential energy function. Loosened constraints can also allow molecules to overcome local minima in the search for a global minimum with respect to both the NMR-derived constraints and conformational energy, resulting in improved convergence rates.

The skewed distance-restraint penalty function (Fig. 1) also appears to have significant advantages for SA with restrained

MD. In order to document this, identical sets of constraints were used to generate structures from extended starting conformers using both a typical flat-well harmonic penalty function and the skewed penalty function of Figure 1. Using a flat-well harmonic function, the constraints could not be included all at once because the large distance violations associated with long-range constraints will dominate the dynamics (Bassolino, 1988). The EXACT list of 1,657 constraints was therefore broken into three sets of increasingly longer-range upper-bound values. Using the flat-well harmonic penalty function and progressive incorporation of longer-range constraints, no converged structures were obtained in 100 attempts. On the other hand, the skewed potential exhibited a convergence rate of about 50%. By including long-range constraints with low weight in the early stages of the SA process, the skewed penalty function guides the structure toward the global minimum of the composite constraint and energy potential.

In this work, $^3J(\text{H}^{\text{N}}-\text{H}^{\alpha})$ coupling constraints were also introduced into the structure calculations using a flat-bottomed penalty function that provides no penalty for conformations whose back-calculated coupling constant values lie within lower- and upper-bound J values. This approach is distinct from previously described simple harmonic penalty functions used for J restraints in other programs (Kim & Prestegard, 1990; Torda et al., 1993; Garrett et al., 1994). The flat-bottomed penalty function described here may have significant advantages over a simple harmonic function. In the absence of additional constraints, the flat-bottomed function allows uniform sampling of the whole conformational space consistent with the estimated experimental uncertainty in the coupling constant measurement, whereas simple harmonic penalty functions bias the conformational space of solutions toward the minimum value of the harmonic well and may thus result in overfitting of the scalar-coupling data. Crambin structures that were further refined with $^3J(\text{H}^{\text{N}}-\text{H}^{\alpha})$ and dihedral constraints exhibit distributions of small distance-constraint violations and low values of conformational energies and RMSDs to the target structures, similar to structures generated without these constraints, demonstrating that the $^3J(\text{H}^{\text{N}}-\text{H}^{\alpha})$ and dihedral penalty function implemented in CONGEN is consistent with the distance constraint penalty functions of the program. The implementation and testing of penalty functions for additional vicinal coupling constants [e.g., $^3J(\text{H}^{\alpha}-\text{H}^{\beta})$ and $^3J(^{15}\text{N}-\text{H}^{\beta})$ coupling constants], and an evaluation of the utility of vicinal coupling constants in improving convergence and accuracy of CONGEN structure calculations, will be described at a later date.

Conclusion

The results of this paper show that the new flexible distance constraint potential, as implemented in CONGEN, can be used to successfully generate high-resolution protein structures from sets of conformational constraints similar to those that are obtained from NMR experiments. The new target functions allow all constraints to be included from the beginning of the MD calculations. The EE and ER protocols, beginning with extended starting conformations with randomly assigned initial velocities, provide high-quality structures with a high (i.e., 50–60%) convergence rate. These benchmark calculations further show that, when distance constraints are loosened into ranges defined by upper- and lower-bound values, higher weights on the NOE con-

straint term can be used without compromising overall constraint satisfaction and energetics. This strategy allows structures to be generated that are more relaxed in the potential energy force field. The current implementation of CONGEN also allows use of dihedral angle constraints and J constraints. Successful applications of the general SA protocol described here and the CONGEN computer program to structure determinations of peptides and proteins using real NMR data will be described in detail elsewhere (Newkirk et al., 1994; Jendeborg et al., 1996; Tejero et al., 1996). In addition, the ability to model ensembles of structures with the correct averaging of NOEs and vicinal coupling constants has been added to CONGEN recently. The use of ensemble averaging in the analysis of a β -hairpin-forming peptide will be described elsewhere (Constantine et al., 1995b; Friedrichs et al., 1995).

Methods

Structure calculations were performed using the MD routines implemented in CONGEN (Brooks et al., 1983; Bruccoleri & Karplus, 1987). All calculations were run on a single processor of either a Silicon Graphics 4D/440, 4D/480, or Indigo 2 computer. The displays of the structures employed the Insight graphical display package from Biosym Technologies and structure analysis software developed at Bristol-Myers Squibb or at CABM. The crystal structure of crambin (1CRN, Hendrickson & Teeter, 1981) was obtained from the PDB. Hydrogens were added and the structure was energy-minimized with the CONGEN program for 200 steps of steepest-descent minimization to reduce steric violations introduced by the addition of hydrogens and residual violations due to the different parameter dictionaries. This structure was then used as the target structure from which the constraints were derived and against which all structures were compared.

From the target structure, a list of all interresidue proton-proton distances less than 7.5 Å was compiled. Although NOE data is not usually obtained for protons separated by more than 4–5 Å, the 7.5-Å cutoff was used to critically challenge the robustness of the algorithm. This constraint list assumed all stereospecific resonance assignments could be made, but did not include any intraresidue distance constraints. Hydrogen bond and disulfide constraints were not included. A set of 1,657 exact interproton distances were generated. In the first constraint set, both the upper and lower bounds were set to these target distances. This set of distance constraints is referred to as the EXACT constraint set. A second set of 514 constraints, the RANGES constraint set, was generated from the EXACT constraint list by adding $\pm 20\%$ of the target distance to each upper- and lower-bound constraint (lower-bound limits less than 1.8 Å were set to 1.8 Å) and then limiting the set to distances ≤ 5.0 Å.

Two sets of structures were generated using extended starting structures, i.e., a fully extended starting conformation (all dihedral angles set to 180°) with *trans* peptide bonds. In order to vary the dynamics trajectory for each run, different random number speeds were used to generate random initial atomic velocities (i.e., speeds and directions) from a Maxwellian distribution at 300 K. The calculations were repeated until 10 converged structures (i.e., no distance violations greater than 0.2 Å relative to the EXACT constraint list) were obtained.

Conf. E. was computed as the sum of bond energies, valence bond angle energies, VDW E., hydrogen bond energies, EL E., dihedral angle energies, and improper torsional energies, as defined in the CHARMM parameter libraries (Brooks et al., 1983). Nonbonded interactions were included up to a cutoff of 10.0 Å and the list updated every 20 steps. Electrostatic effects were computed assuming a distance-dependent dielectric equal to the atom-atom distance. Integration of the equations of motion were performed using a Verlet integration algorithm (Verlet, 1967). The time step of the integrator was 1 fs.

Velocity rescaling was employed on an atom-by-atom basis at high temperatures to prevent any one atom from moving too fast in one MD step. This was achieved with the special option TLIMIT in CONGEN. This option applies a velocity-dependent damping factor to any atom whose effective temperature exceeds the limit TLIMIT, normally set to be between two and six times the temperature of the simulation. The TLIMIT option proved to be invaluable in avoiding failures in the numerical integration of the dynamics that can result when atoms have large kinetic energies.

The SA-MD protocol summarized in Figure 3 was employed. This protocol includes an initial heating stage in which the molecule is heated from 300 K to 800 K over an 8-ps period. During this time, the weighting factors on F_{max} and F_{slope} were increased from 1.0 and 0.01 units to 6.0 and 0.1 units, respectively, while the van der Waals and electrostatic terms were scaled from 1.0 to 0.01 units. The system is maintained at 800 K for 23 ps of "weight annealing," during which the weight on the NOE constraint term was also increased to 20 kcal/mol-Å² for the EE set and 80 kcal/mol-Å² for the ER set. In the RANGES protocol,

the "loosened" constraints provide more flexibility for the trajectory within the flat-bottomed penalty function (Fig. 1), and a higher NOE constraint weight resulted in structures with few NOE violations, without compromising the energetics. During these high-temperature dynamics, the weight of constraints on peptide ω dihedral angles was also increased to prevent *cis/trans* flips of the amide bonds.

The system was then slowly cooled at a rate of 50 K per ps to 300 K while the weights on nonbonded interactions were brought back to 1.0 unit. Next, the system was equilibrated at 300 K for 5 ps using normal energy parameters. Finally, a 10-ps MD trajectory was carried out. An average structure was calculated from the last 2–5 ps of this trajectory, and energy-minimized for 4,000 steps using the ABNR minimization algorithm (Brooks et al., 1983) in CONGEN. The resulting atomic coordinates were then stored for analysis. Each of the final structures was further refined by minimization with 39 J-coupling and 78 dihedral angle constraints obtained from the target structure. Weights of 1 kcal/mol-Hz² or 1 kcal/mol-deg², respectively, were used for these constraints. Statistical analyses of structures were performed using both in-house developed code (R. Tejero & G.T. Montelione, unpubl.) and InsightII (Biosym Technologies, Inc.).

Acknowledgments

We thank Stewart Samuels at BMS and George Elkins at CABM for maintaining the computer systems of our research groups, and Dr. Jiri Novotny for encouraging this work and the collaboration. This work was supported by The National Science Foundation (MCB-93557526), The National Institutes of Health (GM-47014), and a National Science Foundation Young Investigator Award (to G.T.M.) R.T. was supported in part by a fellowship from Pharmacia, A.B., Stockholm, Sweden. The CONGEN program and sample input files are available directly from D.B.-K. or R.E.B.

References

- Bassolino DA. 1989. New computational methods for the determinations of solution structures of biopolymers from NMR data. Ph.D. Thesis, Rutgers University.
- Bassolino DA, Hirata F, Kitchen D, Kominos D, Pardi A, Levy R. 1988. Determination of protein structures in solution using NMR data and IMPACT. *Int J Supercomp Appl* 21:41–61.
- Bax A, Max D, Zax D. 1992. Measurement of long-range ¹³C-¹³C J couplings in a 20 kDa protein-peptide complex. *J Am Chem Soc* 114: 6923–6925.
- Biamonti C, Rios CB, Lyons BA, Montelione GT. 1994. Multidimensional NMR experiments and analysis techniques for determining homo- and heteronuclear scalar coupling constants in proteins and nucleic acids. *Adv Biophys Chem* 4:51–120.
- Braun W, Gö N. 1985. Calculation of protein conformations by proton-proton distance constraints: A new efficient algorithm. *J Mol Biol* 186: 611–626.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
- Bruccoleri RE, Karplus M. 1987. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 26:137–168.
- Brunger AT. 1992. *X-PLOR, version 3.1. Manual*. New Haven, Connecticut: Yale University.
- Brunger AT, Clore GM, Gronenborn AM, Karplus M. 1986. Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: Application to crambin. *Proc Natl Acad Sci USA* 83:3801–3805.
- Brunne RM, van Gunsteren WF, Brüschweiler R, Ernst RR. 1993. Molecu-

Fig. 3. Flow chart of general protocol used for restrained optimization by SA with MD in CONGEN.

- lar dynamics simulations of the proline conformational equilibrium and dynamics in Antamanide using the GROMOS force field. *J Am Chem Soc* 115:4764–4768.
- Bystrov VF. 1976. Spin-spin coupling and the conformational states of peptide systems. *Prog NMR Spectroscopy* 10:41–81.
- Celda B, Biamonti C, Arnau MJ, Tejero R, Montelione GT. 1995. Combined use of ^{13}C chemical shift and $^1\text{H}^\alpha\text{-}^{13}\text{C}^\alpha$ heteronuclear NOE data in monitoring a protein NMR structure refinement. *J Biomol NMR* 5: 161–172.
- Clore GM, Gronenborn AM. 1991. Structures of larger proteins in solution: Three and four dimensional heteronuclear NMR spectroscopy. *Science* 252:1390–1399.
- Clore GM, Gronenborn AM. 1994. Structures of larger proteins, protein-ligand, and protein-DNA complexes by multidimensional heteronuclear NMR. *Protein Sci* 3:372–390.
- Constantine KL, Friedrichs MS, Mueller L, Brucoleri RE. 1995a. J-coupling restraint potentials for non-stereospecifically assigned methylene protons and ensemble average calculations. *J Magn Reson Ser B* 108:176–184.
- Constantine KL, Mueller L, Andersen NH, Tong H, Wandler CF, Friedrichs MS, Brucoleri RE. 1995b. Structural and dynamic properties of a β -hairpin-forming linear peptide. I. Modeling using ensemble averaged constraints. *J Am Chem Soc* 117:10841–10854.
- Driscoll PC, Gronenborn AM, Bergess L, Clore GM. 1989. Determination of the 3D solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*. A study using NMR and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* 28:2188–2198.
- Friedrichs MS, Mueller L, Constantine KL, Brucoleri RE. 1995. Structural and dynamic properties of a β hairpin-forming linear peptide. II. ^{13}C NMR relaxation analysis. *J Am Chem Soc* 117:10855–10864.
- Folkers PJM, Clore GM, Driscoll PC, Dodt J, Köhler S, Gronenborn AM. 1989. Solution structure of recombinant hirudin and the Lys-47-Glu mutant. A NMR and hybrid distance geometry-dynamical simulated annealing study. *Biochemistry* 28:2601–2617.
- Garrett DS, Kuszewski J, Hancock TJ, Lodi PJ, Vuister GW, Gronenborn AM, Clore GM. 1994. The impact of direct refinement against three-bond $\text{H}^\alpha\text{-H}^\alpha$ coupling constants on protein structure determination by NMR. *J Magn Reson Ser B* 104:99–103.
- Griesinger C, Eggenberger U. 1992. Determination of proton-proton coupling constants in ^{13}C labeled molecules. *J Magn Res* 97:426–434.
- Griesinger C, Sørensen OW, Ernst RR. 1986. Correlation of connected transitions by 2D NMR spectroscopy. *J Chem Phys* 85:6837–6852.
- Güntert P, Braun W, Billeter M, Wüthrich K. 1989. Automated stereospecific ^1H NMR assignments and their impact on the precision of protein structure determination in solution. *J Am Chem Soc* 111:3997–4004.
- Güntert P, Braun W, Wüthrich K. 1991. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS, and GLOMSA. *J Mol Biol* 217:517–530.
- Hare DR, Reid BR. 1986. Three dimensional structure of a DNA hairpin in solution: Two dimensional NMR studies and distance geometry calculations on d(CGCGTTTTCGCG). *Biochemistry* 25:5341–5350.
- Havel TF. 1991a. The sampling properties of some distance geometry algorithms applied to unconstrained polypeptide chains: A study of 1830 independently computed conformations. *Biopolymers* 29:1565–1585.
- Havel TF. 1991b. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by NMR. *Prog Biophys Mol Biol* 56:143–78.
- Havel TF, Wüthrich K. 1984. Distance geometry program for determining the structures of small proteins and other macromolecules from NMR measurements of intramolecular $^1\text{H}\text{-}^1\text{H}$ proximities in solution. *Bull Math Biol* 46:673–698.
- Havel TF, Wüthrich K. 1985. An evaluation of the combined use of NMR and distance geometry for the determination of protein conformations in solution. *J Mol Biol* 182:281–294.
- Hendrickson WA, Teeter MM. 1981. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulfur. *Nature* 290:107–113.
- Jendeborg L, Tashiro M, Tejero R, Lyons B, Uhlén M, Montelione GT, Nilsson B. 1996. The mechanism of binding staphylococcal protein A to immunoglobulin G does not involve helix unwinding. *Biochemistry* 35:22–31.
- Kapteyn R, Zuiderweg ERP, Scheek RM, Boelens R, van Gunsteren WF. 1985. A protein structure from NMR data – lac Repressor headpiece. *J Mol Biol* 182:179–182.
- Karplus M. 1959. Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys* 30:11–15.
- Kim Y, Prestegard JH. 1990. Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins Struct Funct Genet* 8:377–385.
- Lautz J, Kessler H, Blaney JM, Scheek RM, van Gunsteren WF. 1989. Calculating three dimensional molecular structures from atom-atom distance information: Cyclosporin A. *Intl J Pept Prot Res* 33:281–288.
- Levy RM, Bassolino DA, Kitchen DB, Pardi A. 1989. Solution structures of proteins from NMR data and modeling: Alternative folds for neutrophil peptide 5. *Biochemistry* 28:9361–9372.
- Marquart M, Walter J, Drenth J, Bode W, Huber R. 1983. The geometry of the reactive site of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr B* 39:480–490.
- Metzler WJ, Hare DR, Pardi A. 1989. Limited sampling of conformational space by the distance geometry algorithm: Implications for structures generated from NMR data. *Biochemistry* 28:7045–7052.
- Montelione GT, Emerson SD, Lyons BA. 1992. A general approach for determining scalar coupling constants in polypeptides and proteins. *Biopolymers* 32:327–334.
- Montelione GT, Winkler ME, Rauenbuehler P, Wagner G. 1989. Accurate measurements of long range heteronuclear coupling constants from homonuclear 2D NMR spectra of isotope enriched proteins. *J Magn Reson* 82:198–204.
- Newkirk K, Feng W, Jiang W, Tejero R, Emerson SD, Inouye M, Montelione GT. 1994. Solution NMR structure of the major cold shock protein (CspA) from *E. coli*: Identification of a binding epitope for DNA. *Proc Natl Acad Sci USA* 91:5114–5118.
- Nilges M, Clore GM, Gronenborn AM. 1988. Determination of three-dimensional structures of proteins from inter-proton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett* 229:317–324.
- Nilges M, Clore GM, Gronenborn AM. 1990. ^1H -NMR stereospecific assignments by conformational data-base searches. *Biopolymers* 29:813–822.
- Pardi A, Billeter M, Wüthrich K. 1984. Calibration of the angular dependence of the amide proton-C-alpha proton coupling constants in a globular protein. Use of $^3J_{\text{HN},\alpha}$ for identification of helical secondary structures. *J Mol Biol* 180:741–751.
- Suri AK, Levy RM. 1993. Estimation of interatomic distances in proteins from NOE spectra at longer mixing times using an empirical two-spin equation. *J Mag Res Ser B* 101:320–324.
- Teeter MM, Roe SM, Heo NH. 1993. Atomic resolution (0.83 Å) crystal structure of the hydrophobic protein crambin at 130 K. *J Mol Biol* 230: 292–311.
- Tejero R, Bassolino D, Brucoleri R, Montelione GT. 1996. Simulated annealing with restrained molecular dynamics using CONGEN: Application in energy minimization of the NMR solution structures of epidermal and type- α transforming growth factors. *Protein Sci* 5:578–592.
- Torda AE, Brunne RM, Huber T, Kessler H, van Gunsteren WF. 1993. Structure refinement using time-averaged J-coupling constant restraints. *J Biomol NMR* 3:55–66.
- Ulyanov NB, Schmitz U, James TL. 1993. Metropolis Monte Carlo calculations of DNA structure using internal coordinates and NMR distance restraints: An alternative method of generating a high-resolution solution structure. *J Biomol NMR* 3:547–568.
- Verlet L. 1967. Computer experiments on classical fluids. I. Thermodynamic properties of Lennard-Jones molecules. *Phys Rev* 159:98–103.
- Wlodawer A, Walter J, Huber R, Sjölin L. 1984. Structure of BPTI – Results of joint neutron and X-ray refinement of crystal form II. *J Mol Biol* 180:301–329.
- Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: Wiley & Sons.