Journal of Mathematical Chemistry 15 (1994) 353–366 

# Applications of weighting and chirality strategies for distance geometry algorithms to an enterotoxin peptide analog

W. Glunt, T.L. Hayden and C. Wells

*Department of Mathematics, University of Kentucky, Lexington, KY 40506, USA*

J.G. Shelling and D.J. Ward

*Department of Biochemistry, University of Kentucky, KY 40536, USA*

Received 9 September 1992; revised 18 August 1993

Significant improvements are made to a recent algorithm that finds molecular conformations using distance geometry on nuclear magnetic resonance data. Weighting factors for the nearest approximation of the distance matrix to a data matrix are allowed to vary between iterations of the algorithm. These changes are proportional to the error of the distance between atoms in the configuration and the nuclear magnetic resonance data bounds. The weight changes increase the rate of convergence by an order of magnitude. Penalty functions are proposed to ensure the correct chirality. Numerical results for these modifications and subsequent energy calculations using CHARMm are given for an analog of the heat stable (ST) enterotoxin peptide STh produced by *E.coli* in humans.

## 1. Introduction and background

Distance geometry is one method used to determine the conformation of molecules from nuclear magnetic resonance (NMR) data and other chemical information. A review of chemical distance geometry by G. Crippen may be found in ref. [1]. One may view distance geometry as a means of sampling the space of conformations consistent with the distance constraints, or as a means of generating chemically-reasonable structures (in terms of bond lengths and valence angles) for refinement by energy calculations. Distance geometry has made significant contributions to both the general determination of molecular conformation [2] and pharmaceutical design [3,4].

Our distance geometry algorithms use the following notation and ideas. The chemical and NMR constraints yield a Data Box defined by upper, $u_{ij}$, and lower, $l_{ij}$, bounds so that the distance, $d_{ij}$ between atom $i$ and atom $j$ in a molecule with $n$ atoms should satisfy

© J.C. Baltzer AG, Science Publishers

$$0 < l_{ij} \leqslant d_{ij} \leqslant u_{ij}, \quad 1 \leqslant i,j \leqslant n.$$

We use the bound smoothing algorithms developed by Havel, Crippen, and Easthope [5,6], incorporating basic geometry such as the triangle inequality to lower the upper bounds and raise the lower bounds.

Given the smoothed Data Box, we randomly choose a set of numbers $\delta_{ij}$, called dissimilarities, which obey the bounds

$$l_{ij} \leqslant \delta_{ij} \leqslant u_{ij}, \quad i,j = 1,2,\ldots n.$$

We then seek $n$ atoms (points) in three dimensions, denoted by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, to produce a molecular conformation, and place the coordinates in a $n$ by 3 matrix, X, whose distances $d_{ij}(X)$ approximate $\delta_{ij}$.

An initial starting configuration is found by using an alternating projection algorithm, MAP, (see [7,8]). The next step is to use the initial configuration and a gradient method to minimize a function so that the distances $d_{ij}(X)$ generated by the conformation $X$ are near the dissimilarities $\delta_{ij}$ picked in the Data Box. The loss function, related to the Stress function in multidimensional scaling, to be minimized is

$$\sigma(X) = \tfrac{1}{2} \sum_i \sum_j w_{ij}(d_{ij}(X) - \delta_{ij})^2. \tag{1}$$

If the function $\sigma(X)$ is differentiable at $X$, then (see [9])

$$\nabla\sigma(X) = 2(VX - B(X)X).$$

The elements of $V$ are given by

$$v_{ij} = \begin{cases} -w_{ij}, & i \neq j; \\ \sum_{k \neq i} w_{ik}, & i = j. \end{cases}$$

$B(X) = [b_{ij}(X)]$ is a function of $X$ and is defined by

$$b_{ij}(X) = \begin{cases} -w_{ij}\delta_{ij}/d_{ij}(X), & i \neq j \text{ and } d_{ij}(X) > 0; \\ 0, & i \neq j \text{ and } d_{ij}(X) = 0; \\ -\sum_{k \neq i} b_{ik}(X), & i = j. \end{cases}$$

In order to minimize (1) we use a spectral gradient algorithm, [10]. Its main advantage over other nonlinear gradient algorithms is that no line search is necessary. We denote the local minimum of $\sigma(X)$ (which is a function of the dissimilarities and the weights) by

$$X = P_S([\delta_{ij}], [w_{ij}]).$$

For a given conformation produced by the above algorithm, some of the distances generated by the conformation will not lie in the Data Box. In order to apply alternating projections, we project the distance matrix onto the Data Box with the following algorithm:

Let $X^k$ be a three-dimensional conformation and $[d_{ij}(X^k)]$ be the distance matrix generated by $X^k$. Then

$$\delta_{ij}^k = \begin{cases} d_{ij}(X^k) & \text{if } l_{ij} \leqslant d_{ij}(X^k) \leqslant u_{ij}\,; \\ l_{ij} & \text{if } d_{ij}(X^k) < l_{ij}\,; \\ u_{ij} & \text{if } d_{ij}(X^k) > u_{ij}\,. \end{cases}$$

We denote this projection onto the Data Box by $P_B(X^k)$.

In the next section an alternating minimization algorithm that was used in [10] will be modified.

## ALTERNATING MINIMIZATION ALGORITHM

- Let $[\delta_{ij}^0]$ be an initial set of dissimilarities in the Data Box.
- While $\sigma^k > \epsilon$
- $X^k = P_S([\delta_{ij}^{k-1}], [w_{ij}^{k-1}])$
- $[\delta_{ij}^k] = P_B(X^k)$
- $\sigma^k = \frac{1}{2}\sum_i \sum_j w_{ij}(d_{ij}(X^k) - \delta_{ij}^k)^2$
- End while

The above algorithm is similar to alternating least-squares methods used in multidimensional scaling [11,12] and to alternating projections on convex sets as proposed by Cheney and Goldstein [13]. The goal is to find three-dimensional configurations whose distance matrices are in or near the Data Box.

These algorithms and the modifications discussed in this paper will be applied to a peptide. It is known that peptides and proteins exhibit dynamical motion at room/body temperature, and while proteins fluctuate about a well defined conformation (characterized in terms of its secondary structure), peptides have an ensemble of *different* conformations of similar energy. X-ray structures of proteins are usually similar to those determined by NMR, except for surface groups interacting with the environment. However, for peptides, without an obvious "core", there may be major differences between the crystal and NMR structures.

An investigation using NMR in $D_2O/H_2O$ of the enterotoxin peptide analog STh(6-19) [14] with sequence $Cys^6$-$Cys^7$-$Glu^8$-$Leu^9$-$Cys^{11}$-$Asn^{12}$-$Pro^{13}$-$Ala^{14}$-$Cys^{15}$-$Thr^{16}$-$Gly^{17}$-$Cys^{18}$-$Tyr^{19}$, identified two NOEs involving $Asn^{12}C\alpha$-H. One of these was to Gly Cα-H and the other to Tyr Cδ-H. Neither NOE was stereospecifically assigned, and both appear relatively weak. Although the NMR study [14] found about a 50 : 50 mixture of *cis* and *trans* isomers at the $Asn^{12}$-$Pro^{13}$ peptide bond, the effect was local and only the *trans* form was considered in the simulations discussed here.

Prior to applying the non-linear optimization, we obtain mirror image structures of mixed residue chiralities equally satisfying the constraints. Selection criteria are needed to determine which conformer should be chosen during optimization in the above algorithms. Selection could be based on bound violations and, in the case of proteins, analysis of the fold in terms of secondary structure. The latter is unlikely to be an option for peptides, because even where elements of secondary structure are reported [15] they tend to be β turns (as occur in the crystal structure of the ST analog mpr$^5$-STp(5-17)) [16]. A type I β-turn could transpose to a type I′ turn in a conformer of opposite handedness. In other words, a mixture of positive and negative $\phi$ angles in a conformer with hydrogen bond-stabilized turns may be reasonable. A study of the 21 residue peptide endothelin [17] was able to distinguish between predicted handed conformers by identification of right-handed helicity in solution using circular dichroism.

In this paper, since we have not developed explicit criteria for the selection of the backbone handedness, we fully optimize both mirror image structures. By developing new weight change strategies in this paper, we significantly increase the rate of convergence of a new distance geometry algorithm [10]. Additionally we impose constraints in the algorithm to enforce the correct chirality and also consider volume constraints to enforce the planarity of peptide bonds and aromatic rings. At the end of the section on the weight and chirality strategy we indicate the mathematical insight that our methods can give for solving the difficult conformation problem using penalty terms on the chirality and the balance needed with the changing weights. In the final section, we have used the distance geometry algorithms described above to generate structures of STh(6-19) consistent with the NOE constraints. These structures are compared, before and after energy minimization using CHARMm [18], with the crystal structure of the ST analog mpr$^5$-STp(5-17) [16].

## 2. Weight changes and chirality

### 2.1. UPDATING WEIGHTS

One of the advantages of the spectral gradient algorithm is that the matrix $V$ in the gradient of $\sigma$ is easy to compute from the weights. If one uses the majorization algorithm in [9], one needs to compute a generalized inverse of $V$.

In our approach a weight matrix $W_B$ is first formed from the bounds. We found that

$$w_{ij} = \frac{1}{1 + 10(u_{ij} - l_{ij})}$$

worked well to put more weight on the tightest bounds. Neither the constant 10

nor the linear term in $(u_{ij} - l_{ij})$ is crucial. However, using an exponent larger than two on this difference was not satisfactory.

With this constant weight matrix it was necessary to run 1000 iterations of the alternating projection algorithm to find a configuration with bound violations less than 0.2 angstroms (Å). However if one adjusts the weights every 10 iterations, so that additional weight is added proportional to the violations as follows, then comparable accuracy can be achieved in 140 iterations.

Every 10 iterations form a violation, or error, matrix $E = [e_{ij}]$, (symmetric with zero diagonal) where

$$e_{ij} = \alpha(\max(0, d_{ij} - u_{ij}, l_{ij} - d_{ij}) .$$

Then a new weight matrix is formed as

$$W_{\text{NEW}} = W_{\text{B}} + E .$$

We choose $\alpha$ as (iterations)/10. A strategy that failed is to update the weight matrix by adding the violations to the previous update rather than to the original $W_{\text{B}}$.

## 2.2. CHIRALITY AND PLANARITY CONSTRAINTS

The function to be minimized by the previous algorithm is now modified by the addition of penalty terms to enforce the correct chirality of the $\alpha$ carbon atoms and the planarity of the peptide bonds and the aromatic rings.

Suppose we are given a chiral center (an $\alpha$ carbon). Let $P_i, P_j, P_k, P_l$ be the four atoms bonded to the $\alpha$ carbon to form the tetrahedron. Let the coordinates of the points be denoted for example by $P_i = (x_{i1}, x_{i2}, x_{i3})$. The oriented volume of the tetrahedron is determined by one sixth of the following determinant:

$$v_{ijkl} = \det \begin{bmatrix} 1 & x_{i1} & x_{i2} & x_{i3} \\ 1 & x_{j1} & x_{j2} & x_{j3} \\ 1 & x_{k1} & x_{k2} & x_{k3} \\ 1 & x_{l1} & x_{l2} & x_{l3} \end{bmatrix} .$$

Havel [19] employs a simple quadratic penalty term to enforce the chirality which yields a zero value when the correct orientation is achieved. We propose a different chirality penalty term which will not have zero value (the value will be small), but will have a small gradient with the correct orientation and a large gradient when the orientation is not correct. For each $\alpha$ carbon we propose the following sigmodial penalty term:

$$f = \alpha\left(\pi/2 - \frac{\arctan(v_{ijkl})}{\pi/2 + \epsilon + \arctan(v_{ijkl})}\right) .$$

This penalty term will enforce a positive orientation (positive determinant) on the

volume of the points in the order $P_i, P_j, P_k, P_l$. In case of the opposite chirality, use the same expression but simply switch two rows of the determinant by reordering the points, say $P_j, P_i, P_k, P_l$. Since an interchange of two rows of the determinant changes the sign of the determinant, this will impose the opposite chirality when inserted in the same penalty function.

The parameter $\epsilon$ controls the growth range of the penalty function. For example as $\epsilon \to 0$, the penalty is approximately linear for negative values of the determinant. For small positive $\epsilon$, it is almost linear from zero to a large negative value, and then approaches a large horizontal asymptote. We find $\epsilon = 0.01$ to be a satisfactory value. We use the scaling factor $\alpha$ to avoid overflow and we choose $\alpha = 1/n^2$.

In order to enforce planarity constraints, we force the above determinant to be zero (or small). For this purpose we propose the following penalty function:

$$g = \alpha\left(1 - \frac{1}{1 + (\beta v_{ijkl})^2}\right).$$

The value of $\beta$ which has the effect of choosing the width of the "well" around zero which drives the function to zero is 1.5 in this paper.

For $s \in \{1, 2, 3\}$ and $r \in \{1, \ldots, n\}$, the entries of the gradient of $f$ and $g$ are given by

$$\frac{\partial f}{\partial x_{rs}} = \frac{-\alpha(\pi/2 + \epsilon)}{(\pi/2 + \epsilon + \arctan(v_{ijkl}))^2} \frac{1}{1 + (v_{ijkl})^2} \frac{\partial v_{ijkl}}{\partial x_{rs}},$$

and

$$\frac{\partial g}{\partial x_{rs}} = \frac{2\beta^2 \alpha v_{ijkl}}{(1 + (\beta v_{ijkl})^2)^2} \frac{\partial v_{ijkl}}{\partial x_{rs}}.$$

The term $\partial v_{ijkl}/\partial x_{rs} = 0$ if $r \notin \{i, j, k, l\}$. Otherwise

$$\frac{\partial v_{ijkl}}{\partial x_{rs}} = c^r_{ijkl}(-1)^s v^{rs},$$

where

$$c^r_{ijkl} = \begin{cases} 1 & \text{if } r = i; \\ -1 & \text{if } r = j; \\ 1 & \text{if } r = k; \\ -1 & \text{if } r = l. \end{cases}$$

and $v^{rs}$ is the minor determinant obtained from $v_{ijkl}$ by removing the row and column containing the term $x_{rs}$.

To enforce each chirality and planarity constraint a term corresponding to $f$ or $g$ is added to the original loss function $\sigma$. Then all the previous algorithms apply with $\sigma$ replaced by $\sigma$ plus penalty terms. The final conformation obtained by the

alternating projection algorithm should now have the correct chirality for the $\alpha$ carbon atoms, very small deviations from the desired planar configurations, and distances which obey the bounds, with small error (assuming there exist solutions in the Data Box).

Penalty methods in nonlinear optimization require care in the choice of the penalty function and the coefficients. Our results show a robust choice of parameters for the weight updates and penalty parameters yield good results for the enterotoxin peptide. An examination of the volume expression in the penalty terms shows that a moderate change in each distance in the tetrahedron can induce a large change in the volume. Hence if a structure is undergoing large changes in the initial conformations in the minimization algorithm, then very large changes in the penalty terms may cause overflow problems. There are two reasons this did not occur in this application. First, instead of the usual quadratic penalty used in other distance geometry programs, we employ a function that grows more slowly and the penalty parameter was small. Second, this peptide was small and compact, due to the disulfide bridges, and no large distance changes ever occur. However, when we used the same protocol on the enterotoxin peptide, but with no disulfide bridges and no NOE constraints, we observed very large changes in the volumes and poor performance of the algorithm with the current parameters. Hence, for those peptides with similar compactness only small changes in the weight and chirality parameters should be required to yield good results.

However, for a molecule like 50-L-Alanine with no disulfide bridges and no NOE constraints, large distance changes occur during the minimization. In this case we find that to avoid overflow it is necessary to have both the weight change parameters and the chirality penalty term parameters very small initially and slowly increase these values as the minimization progresses.

## 3. Protocols and numerical results

All the distance geometry numerical experiments were run on IBM 3090-600J using the VS Fortran compiler. The numerical methods are applied to find conformations of the analog STh(6-19) of the peptide enterotoxin STh produced by *E.coli* in humans [20].

The Data Box is generated from the standard chemical bounds (generated in this case by DG-II [19], giving 160 atoms) and the NMR data from Gariépy et al. [14]. Two NOEs were identified involving $Asn^{12}C\alpha$-H; one to $Gly^{17}C\alpha$-H and the other to $Tyr^{19}C\delta$-H. We imposed upper bound constraints in the distance geometry calculations between $Asn^{12}C\alpha$-H and Tyr $C\delta_2$-H of 4.5 Å, and between $Asn^{12}C\alpha$-H and $Gly^{17}C\alpha$ of 5.5 Å. The lower bounds were van der Waals radii. Because the directionality constraint of a pseudoatom representation [21] is missing in our calculations, the $Gly^{17}C\alpha$-H atoms may end up a distance greater than the intended maximum of 4.5 Å from $Asn^{12}C\alpha$-H.

We generated two sets of data, one in which no constraints were placed on the disulfide bonds ($S_1$), and a second set ($S_2$) in which distances were constrained so that the torsion angle was restricted to $\pm 90° \pm 30°$. In the initial starting conformation the mirror image is obtained by changing the sign of the first coordinates of each point in the configuration. The resulting conformations with the same backbone handedness as the crystal structure of the ST analog mpr$^5$-STp (5-17) [16] will be denoted by $(X)$ and that of the opposite handedness by $(-X)$. This procedure is repeated four times with different random dissimilarities to produce five structures in each of the classes $S_1(X), S_1(-X), S_2(X), S_2(-X)$. We do not observe a change in the backbone handedness in the subsequent optimization.

The stopping value in the spectral gradient algorithm was set to $10^{-5}$. All conformations are found with 150 iterations of the alternating minimization algorithm. The root mean square difference (RMSD) in angstroms (Å) for the backbone atoms (calculated using QUANTA 3.3, Molecular Simulations Inc.) between pairs of conformations in each of the four sets had the following ranges. RMSD between the five conformations from $S_1(X)$ ranged between 1.26 and 1.76 Å. Similarly the RMSD range for $S_1(-X)$ was 0.962–1.68 Å; for $S_2(X)$ 1.24–1.66 Å; and for $S_2(-X)$ 1.26–1.57 Å.

The data in table 1 indicates the goodness of the fit of our structures in the Data Box. There were 533 positions in the Data Box with $u_{ij} = l_{ij}$, which we call tight bounds, out of a total of 25 440 bound restrictions. We record for each final configuration: the total number of bound violations (TBV) (the tight bounds are always slightly violated); the maximum bound violation (MBV) in angstroms; the average bound violation (AV) in angstroms times $10^{-4}$; the standard deviation of the bound violations (SD) in angstroms times $10^{-3}$. We also record for the tight bounds: the maximal tight violation (MTV), the average tight violation (ATV) times $10^{-2}$ and the standard tight deviation (STD) times $10^{-2}$.

All of the final conformations have very small bound violations. These violations can be further reduced by more iterations, but we find for this example that 150 iterations yields excellent results. For the conformers with opposite handedness $(S_1(-X), S_2(-X))$ to the crystal structure, at least one (usually both) NOE constraints is always violated. Those with the same handedness as the crystal structure $(S_1(X), S_2(X))$ always fall within the bounds, with one exception (and for only one bound); conformer 3.

The conformations obtained by our distance geometry algorithms were used as starting values for energy minimization. These 160 atom structures were imported into CHARMm22/QUANTA3.3 running on a Silicon Graphics 4D 70/GT Iris workstation. In an allatom representation with charged functional groups this gives 177 atoms and a total charge of +1.0. All interactions were included in the total energy and a distance dependent dielectric (RDIE 30, [22]) was used. Each structure was energy minimized using the methods of steepest descents (50 steps) and then adopted-basis Newton Raphson was used to convergence with a gradient tolerance 1E-5.

Table 1
Bound violations for 20 conformers generated by distance geometry.

| Conf. | TBV | MBV | AV | SD | MTV | ATV | STD |
|---|---|---|---|---|---|---|---|
| *Bound violations for* $S_1(X)$ | | | | | | | |
| 1 | 668 | 0.090 | 4.76 | 4.08 | 0.089 | 1.60 | 1.57 |
| 2 | 676 | 0.115 | 5.15 | 4.57 | 0.111 | 1.72 | 1.88 |
| 3 | 672 | 0.112 | 5.32 | 4.76 | 0.112 | 1.77 | 1.97 |
| 4 | 664 | 0.102 | 4.92 | 4.31 | 0.091 | 1.73 | 1.79 |
| 5 | 660 | 0.122 | 5.05 | 4.64 | 0.114 | 1.69 | 1.93 |
| *Bound violations for* $S_1(-X)$ | | | | | | | |
| 1- | 674 | 0.113 | 4.58 | 4.13 | 0.113 | 1.49 | 1.72 |
| 2- | 663 | 0.097 | 4.18 | 3.68 | 0.089 | 1.34 | 1.47 |
| 3- | 672 | 0.138 | 5.07 | 4.56 | 0.138 | 1.72 | 1.98 |
| 4- | 657 | 0.138 | 5.96 | 5.24 | 0.126 | 2.14 | 2.2 |
| 5- | 676 | 0.117 | 5.48 | 4.59 | 0.117 | 1.86 | 1.85 |
| *Bound violations for* $S_2(X)$ | | | | | | | |
| 1 | 667 | 0.103 | 4.74 | 4.19 | 0.095 | 1.56 | 1.59 |
| 2 | 674 | 0.114 | 5.32 | 4.65 | 0.114 | 1.78 | 1.91 |
| 3 | 676 | 0.126 | 5.33 | 4.76 | 0.126 | 1.77 | 1.96 |
| 4 | 661 | 0.114 | 4.82 | 4.25 | 0.089 | 1.67 | 1.71 |
| 5 | 678 | 0.121 | 5.41 | 4.95 | 0.119 | 1.81 | 2.16 |
| *Bound violations for* $S_2(-X)$ | | | | | | | |
| 1- | 670 | 0.114 | 4.68 | 4.13 | 0.113 | 1.61 | 1.76 |
| 2- | 672 | 0.103 | 4.70 | 4.14 | 0.097 | 1.52 | 1.62 |
| 3- | 681 | 0.095 | 4.46 | 3.74 | 0.072 | 1.49 | 1.44 |
| 4- | 665 | 0.131 | 6.41 | 5.42 | 0.131 | 2.27 | 2.21 |
| 5- | 682 | 0.120 | 5.39 | 4.76 | 0.119 | 1.77 | 1.94 |

A bug in the CHARMm22/QUANTA 3.3 package in the NOE constraints code requires that we select a specific Gly Cα proton and we used Cα-$H_1$. The Tyr $C\delta_2$-H was chosen as before. To be consistent with the distance geometry calculations, we made the optimum distance of the 2 NOE constraints in the energy calculations to be 4.5 Å. The minimum interproton distance was set as 1.8 Å and the maximum was 5 Å. A *scale* value, which is an empirical value biasing the NOE distance constraints over other geometrical considerations, was set to 15.

After energy minimization, for conformers with the same handedness as the crystal, RMSD to the crystal never decreases (see table 2). This might be expected since the NOE distance constraints are not satisfied in the mpr[5] STp(5-17) crystal structure (protons added in QUANTA3.3).

Table 2
Data summary of the 20 structures of STh (6-19) generated with 2 NOE constraints [14] by distance geometry and energy minimization.

| Conf. | Energy (kcal/mol) after EM | RMSD (Å) before/after EM | Disulfide helicity | Number H-bonds | NOE distances (Å) (before) & after EM Asn–Gly/Asn–Tyr |
|---|---|---|---|---|---|
| $S_1(X)$ | | | | | |
| 1 | −20.76 | 2.34/2.57 | + + − | 3 | (3.80)4.52/(2.93)3.73 |
| 2 | −7.14 | 2.15/2.52 | + + − | 0 | (4.76)4.51/(2.74)4.51 |
| 3 | −10.95 | 1.91/2.54 | − + + | 1 | (5.56)4.57/(3.42)4.00 |
| 4 | −18.05 | 2.36/2.53 | − − + | 3 | (3.95)4.55/(3.80)4.27 |
| 5 | −9.63 | 2.26/2.80 | + + + | 2 | (3.82)2.23/(2.49)4.52 |
| $S_1(-X)$ | | | | | |
| 1- | −14.04 | 5.39/4.77 | − − + | 1 | (5.55)4.56/(1.08)4.60 |
| 2- | −17.20 | 5.35/4.71 | + + − | 1 | (5.55)4.58/(4.54)4.60 |
| 3- | −19.47 | 5.42/4.65 | + + + | 3 | (5.53)4.58/(4.28)4.29 |
| 4- | −22.22 | 5.45/4.55 | + − − | 2 | (5.43)4.36/(4.55)3.89 |
| 5- | −14.65 | 5.57/4.34 | + − + | 2 | (5.56)4.64/(4.54)4.08 |
| $S_2(X)$ | | | | | |
| 1 | −18.97 | 2.29/2.43 | + + − | 3 | (4.06)4.06/(2.91)4.27 |
| 2 | −10.32 | 2.14/2.14 | − + + | 1 | (4.72)2.77/(2.75)3.48 |
| 3 | −11.10 | 2.00/2.57 | − + − | 3 | (5.58)4.58/(3.32)3.80 |
| 4 | −11.82 | 2.27/2.33 | − − − | 2 | (4.32)3.08/(3.62)4.53 |
| 5 | −14.14 | 2.21/2.87 | + + − | 2 | (3.85)2.57/(3.05)4.58 |
| $S_2(-X)$ | | | | | |
| 1- | −18.57 | 5.53/4.85 | − + + | 3 | (5.56)4.58/(4.53)4.52 |
| 2- | −24.95 | 5.41/4.88 | − − − | 4 | (5.59)4.56/(4.52)4.62 |
| 3- | −19.79 | 5.47/4.62 | + + + | 2 | (5.54)4.59/(4.55)4.11 |
| 4- | −0.91 | 5.40/4.74 | + − + | 3 | (5.41)4.51/(4.55)4.55 |
| 5- | −13.56 | 5.51/4.69 | − + + | 4 | (5.56)4.30/(4.56)4.51 |

Table 2 shows there were changes in NOE constraint distances, before and after energy minimization. In general, the Asn–Tyr interproton distance increases. This is probably a combination of specifying an optimum distance of 4.5 Å *and* the flexibility of the tyrosine at the C-terminus.

The lowest energy conformer predicted in classes $S_1$ and $S_2$ has opposite handedness to the crystal. The torsion angles for this conformer are given in table 3.

The energy of this conformer is −24.95 kcal/mol, its RMSD to the ST analog crystal structure is 4.88 Å and it has 4 hydrogen bonds (by default QUANTA3.3 criteria). All the hydrogen bonds are backbone-backbone and involve NH..CO pairings in residues 6-14, 15-13, 16-6 and 17-12. The prediction for the $Cys^{15}$ amide is in agreement with the NMR temperature coefficient studies [14], but the amides of residues 16 and 17 were said to be solvent exposed.

Table 3
Torsion angles for the predicted lowest energy conformer of STh(6-19) satisfying the 2 NOE distance constraints.

| Residue # | Amino acid | $\phi$ | $\psi$ | $\chi_1$ | $\chi_2$ | $\chi_3$ |
|---|---|---|---|---|---|---|
| 6 | Cys | n/a | n/a | | | |
| 7 | Cys | 61.2 | −175.9 | | | |
| 8 | Glu | −77.3 | −175.8 | −56.2 | −59.6 | −79.0 |
| 9 | Leu | 63.8 | 77.0 | −52.6 | 174.1 | |
| 10 | Cys | 86.8 | −29.1 | | | |
| 11 | Cys | 56.5 | 78.3 | | | |
| 12 | Asn | −135.3 | 163.4 | −72.3 | −58.2 | |
| 13 | Pro | −43.3 | 129.7 | | | |
| 14 | Ala | 79.2 | −84.9 | 66.0 | | |
| 15 | Cys | −156.3 | 109.9 | | | |
| 16 | Thr | 45.3 | 68.1 | −175.5 | −61.6 | |
| 17 | Gly | −89.1 | 36.8 | | | |
| 18 | Cys | −59.4 | 119.1 | | | |
| 19 | Tyr | −88.1 | n/a | −62.0 | −71.0 | |

Although 6 of the residues have positive $\phi$ angles, the backbone angles of Leu[9], Cys[11] and Thr[16] occur in the stable left-handed helix region of the $\phi/\psi$ map. Ala[14] has angles (79.2°, −84.9°) typical of a $\gamma$ turn; the NMR data [14] supports a $\beta$ turn over the region Ala[14]–Cys[18]. This leaves only Cys[7] (61.2°, −175.9°) and Cys[10] (86.8°, −29.1°) in unfavorable regions. We observe that residue chirality can be corrected during optimization without affecting the backbone, and these cysteines could have undergone chirality correction but be left with D-residue type angles.

We performed an equivalent set of calculations (20 structures) *without* NOE distance constraints and found a lower energy structure (energy −28.81 kcal/mol). Its torsion angles are listed in table 4. This conformer has the same handedness as the crystal structure (RMSD of 2.63 Å) and is predicted to have 5 hydrogen bonds with the 4 backbone-backbone NH..CO residue pairings being 6-12, 8-14, 9-16 and 14-17. The Glu[8] and Ala[14] predictions agree with specific NMR findings [14], while none disagrees. This conformer has close agreement to the torsion angle values derived from experimental coupling constants [14]; Asn[12], Ala[14], Cys[18] and Tyr[19] all have $\phi$ angles near −150°, in line with experiment.

This structure has 2 (distorted) $\beta$ turns. The first is a type VIb turn (optimum angles −120°, 120° and −60°, 0°) around residues Leu[9]–Cys[10]. The second is a distorted III' turn (optimum angles 60°, 30° and 60°, 30°) around Cys[15]–Thr[16]. The latter agrees with NMR data proposing a $\beta$ turn between Ala[14] and Cys[18].

The lowest energy conformer in class $S_1$ has the same disulfide helicity as the ST analog crystal structure (+ − −), otherwise there appears to be a random relationship between disulfide helicity and conformer energy.

Table 4
Torsion angles for the predicted lowest energy conformer of STh(6-19) without NOE constraints.

| Residue # | Amino acid | $\phi$ | $\psi$ | $\chi_1$ | $\chi_2$ | $\chi_3$ |
|---|---|---|---|---|---|---|
| 6 | Cys | n/a | n/a | | | |
| 7 | Cys | 67.2 | 113.6 | | | |
| 8 | Glu | 70.0 | 23.6 | −159.2 | 63.7 | 84.6 |
| 9 | Leu | −147.6 | 147.6 | −66.5 | 169.9 | |
| 10 | Cys | −82.2 | 3.3 | | | |
| 11 | Cys | −62.8 | −47.7 | | | |
| 12 | Asn | −155.6 | 82.3 | −68.1 | −76.9 | |
| 13 | Pro | −49.7 | 117.1 | | | |
| 14 | Ala | −153.6 | 165.4 | 58.5 | | |
| 15 | Cys | 51.0 | 74.1 | | | |
| 16 | Thr | 53.9 | 38.9 | −169.7 | −63.9 | |
| 17 | Gly | 67.2 | −107.8 | | | |
| 18 | Cys | −148.2 | 44.3 | | | |
| 19 | Tyr | −160.4 | n/a | 54.8 | −91.4 | |

## 4. Conclusions

By making weight changes to penalize bound violations within the algorithm we find conformations with small bound violations in approximately one tenth the number of iterations required if constant weights are used. Some variations of changing weights other than every ten iterations showed no significant improvement. The proposed chirality and volume penalty terms produced the correct values (small volume terms for planarity) in each conformation. The addition of the chirality penalty terms roughly doubled the time required to find the final conformation.

For the peptide example in this paper we always achieve a final conformation which has small bound violations and the correct chirality for every initial dissimilarity with an average time of eight minutes and thirty seconds for each conformation. The STh(6-19) peptide is relatively small compared to a protein like Bovine Pancreatic Trypsin Inhibitor (BPTI) for example, and only 2 NOE constraints were incorporated in the distance geometry and energy minimization calculations. However, this study allows us to investigate aspects of protocols for *peptide* simulation and provides feedback for the further development of the algorithms outlined in this paper.

One of these aspects is the selection/rejection of a conformer, during or after optimization. We found we were able to optimize both mirror image conformers each time, with little difference in their bounds violations except that conformers of opposite handedness to the crystal almost always slightly exceeded the NOE con-

straint bounds set (though the final distance was within 5 Å). A selection/rejection option based on secondary structure (as can be used for proteins) requires better understanding of peptide conformational behavior in solution.

We have also investigated the effect in distance geometry calculations of including a constraint on the disulfide $\chi_3$ torsion angle. The inclusion reduces backbone conformational changes upon energy minimization. While the cystine Cα–Cα distance is similar whether the torsion angle is near +90° or −90°, the helicity may generally be a determinant of agonist and antagonist peptide activity at its receptor. This study has application to other peptides, for example oxytocin (1 disulfide) and endothelin (2 disulfides).

An explanation for the occurance of positive $\phi$ angles (especially for Cys residues) comes from our observation that the cystine chirality can be corrected (from D to L) in the optimization stage without changing the orientation of the backbone. We find other residues in our simulations to have positive $\phi$ angles and these are usually involved in β or γ turns. We also find several examples of $\phi/\psi$ angles in the left-handed helix part of the map. Although they are not part of an extended helix, the disulfide bridges in the peptide may provide additional stabilization.

Since the generated conformers satisfy the NOE distance constraints, while the crystal structure does not, we would expect (and find) lower RMSD between conformers within $S_1(X)$ and $S_2(X)$ classes, than from each to the crystal.

We set an "optimum" distance for the NOE constraints of 4.5 Å in the energy calculations (reflecting the observation from NMR data that they were weak), but allowed the proton-proton pairing to assume any value in the range 1.8 to 5 Å. This protocol does not reproduce the time-averaged method for modeling NMR data in dynamics simulations [23], but does allow variation in inter-proton distance during energy minimization.

The atomic representation, the form of the energy potential, and the minimization protocol will all influence the conformations obtained and their relative energies. We used fully charged molecules and a distance-dependent dielectric (RDIE 30, [22]) and find relatively few hydrogen bonds (range 0–4, see table 2); in most cases these are backbone to backbone.

In conclusion, we have used distance geometry and energy minimization methods (including two or zero NOE distance constraints) to predict structures of STh(6-19) having many of the conformational properties determined by NMR. The most energetically favorable conformers *with* constraints are opposite handed to the crystal structure of the ST analog.

## Acknowledgements

The authors thank Tim Havel for making his DG-II algorithm available so that bound smoothing could be applied. The authors gratefully acknowledge the support of the National Science Foundation by grant CHE-9005960.

## References

[1] G.M. Crippen, J. Math. Chem. 6 (1991) 307.
[2] P. Kollman, P. Grotenhuis and M. Lopez, Pure Appl. Chem. 61 (1989) 593.
[3] W. Ripka, W. Sipio and J. Blaney, Lect. Heterocycl. Chem. 9 (1987) 95.
[4] R. Gordon, E. Breuer, F. Padilla, R. Smejkal and P. Chiang, Mol. Pharmacol. 36 (1989) 766.
[5] G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation* (Wiley, New York, 1988).
[6] P. Easthope and T. Havel, Bull. Math. Biol. 51 (1989) 173.
[7] W. Glunt, T.L. Hayden, S. Hong and J. Wells, SIAM J. Matrix Anal. Appl. 11 (1990) 589.
[8] W. Glunt, T.L. Hayden and W.-M. Liu, Bull. Math. Biol. 53 (1991) 769.
[9] J. de Leeuw, J. Classification 5 (1988) 163.
[10] W. Glunt, T.L. Hayden and M. Raydan, J. Comp. Chem. 14 (1993) 114.
[11] J. de Leeuw, in: *Recent Developments in Statistics*, eds. J.R. Barra et al. (North-Holland, Amsterdam, 1977) pp. 133–145.
[12] J. de Leeuw and W. Heiser, in: *Multivariate Analysis - V*, ed. P.R. Krishnaiah (North-Holland, Amsterdam, 1980) pp. 510–522.
[13] W. Cheney and A. Goldstein, Proc. Am. Math. Soc. 10 (1959) 448.
[14] J. Gariepy, A. Lane, F. Frayman, D. Wilbur, W. Robien, G. Schoolnik and O. Jardetzky, Biochemistry 25 (1986) 7854.
[15] W. Johnson, T. Pagano, C. Basson, J. Madri, P. Gooley and I. Armitage, Biochemistry 32 (1993) 268.
[16] H. Ozaki, T. Sato, H. Kubota, Y. Hato, Y. Katsube and Y. Shimonishi, J. Biol. Chem. 266 (1991) 5934.
[17] V. Saudek, J. Hoflack and J. Pelton, Int. J. Peptide Protein Res. 37 (1991) 174.
[18] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus, J. Comp. Chem. 4 (1983) 187.
[19] T.F. Havel, Prog. Biophys. Molec. Biol. 56 (1991) 43.
[20] F. Knoop and M. Owens, J. Pharmacol. Toxicol. Methods 28 (1992) 67.
[21] K. Wüthrich, M. Billeter and W. Braun, J. Mol. Biol. 169 (1983) 949.
[22] C. Brooks, personal communication (1993).
[23] J. de Vlieg and W. van Gunsteren, Methods Enzymol. 202 (1991) 268.