Computing Artificial Backbones of Hydrogen Atoms in order to Discover Protein Backbones

C. Lavor* A. Mucherino† L. Liberti† and N. Maculan‡

*Dept. of Applied Mathematics (IMECC-UNICAMP), State University of Campinas, Campinas-SP, Brazil.
clavor@ime.unicamp.br
†LIX, École Polytechnique, Palaiseau, France.
{mucherino,liberti}@lix.polytechnique.fr
‡COPPE, Systems Engineering, Federal University of Rio de Janeiro, Rio de Janeiro-RJ, Brazil.
maculan@cos.ufrj.br

***Abstract*—NMR experiments are able to provide some of the distances between pairs of hydrogen atoms in molecular conformations. The problem of finding the coordinates of such atoms is known as the molecular distance geometry problem. This problem can be reformulated as a combinatorial optimization problem and efficiently solved by an exact algorithm. To this purpose, we show how an artificial backbone of hydrogens can be generated that satisfies some assumptions needed for having the combinatorial reformulation. Computational experiments show that the combinatorial approach to this problem is very promising.**

I. Introduction

Proteins are important molecules because they perform different functions, often of vital importance, in the cells of the living beings. Their function is determined by the dynamics of the proteins, which depend on their three-dimensional conformation. While finding the chemical composition of a protein molecule is relatively simple, finding its three-dimensional conformation is not an easy task. Nuclear Magnetic Resonance (NMR) is an experimental technique which is able to provide some of the distances between pairs of atoms forming the molecule [5]. These experimentally obtained data can then be used for computing the coordinates (into a given Cartesian system) of all the atoms of the molecule. The problem of finding the conformation of the molecule (i.e. all the coordinates of its atoms), starting from the known distances between pairs of atoms, is referred to as the Molecular Distance Geometry Problem (MDGP) [3]. The focus of this paper is, in particular, on protein molecules.

Over the years, many methods have been proposed for solving the MDGP. Most of them are based on a continuous formulation of the problem. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a protein conformation, where x_i is the i^{th} atom of the protein, in a given ordering. Let E be the set of pairs of atoms whose distance is known. Then, the MDGP can be seen as the problem of finding X such that

$$||x_i - x_j|| = d_{ij} \quad \forall (i,j) \in E,$$

where $|| \cdot ||$ represents the computed distance between two atoms of X, and d_{ij} is the known value of their relative distance. This constraint satisfaction problem is usually reformulated as a global optimization problem. The aim is to minimize an objective function which is able to provide a measure of how much the distances $||x_i - x_j||$, related to a certain conformation X, differ from the known distances d_{ij}, for each $(i,j) \in E$. Different objective functions have been proposed, and one of the most used is the Largest Distance Error (LDE):

$$LDE(\{x_1, x_2, \ldots, x_n\}) = \frac{1}{|m|} \sum_{\{i,j\}} \frac{||x_i - x_j|| - d_{ij}}{d_{ij}}, \quad (1)$$

where m is the total number of known distances. Supposing that a position is given to the n atoms of the conformation X, if the value of the LDE function is 0, then the set of given distances is feasible and the conformation X satisfies all of them. For a survey on methods and algorithms for the MDGP, see [6].

Recently, a new approach to the MDGP has been proposed. In the event that some particular assumptions are satisfied, the global optimization problem associated to the MDGP is reformulated as a combinatorial optimization problem. In this way, the search domain is reduced to a discrete set. Moreover, the computation of the number of solutions contained in the discrete search domain is possible, and it is related to the number of atoms forming the molecule. Computational experiments presented in ([7], [8], [9]) showed that the combinatorial approach to the MDGP is much more efficient than the continuous one. We refer to this combinatorial reformulation of the MDGP as the Discretizable Molecular Distance Geometry Problem (DMDGP).

Proteins are chains of smaller molecules called *amino acids*, which are chemically bound to each other through a subgroup of atoms that each amino acid has in common. We will refer to this subgroup of atoms as the *common part* of each amino acid. All these parts define the so-called *backbone* of the protein.

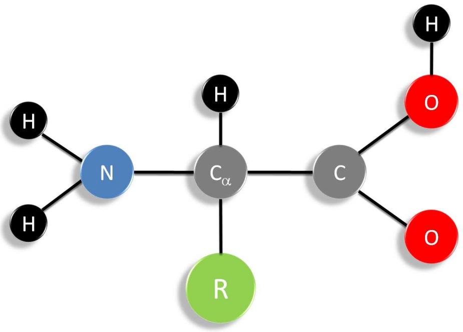

Fig. 1. The general structure of an amino acid.

The general structure of an amino acid is shown in Figure 1. All the atoms of the common part are shown, whereas the circle marked by **R** represents all the others. When two amino acids bind to each other during protein synthesis, some of the atoms of their common parts are lost, and the carbon atom C of the first amino acid binds to the nitrogen N of the second one. Therefore, the protein backbone is finally formed by the sequence of atoms $\mathrm{N} - \mathrm{C}_\alpha - \mathrm{C}$, where oxygen and hydrogen atoms are also attached.

In previous works ([7], [8], [9], [10]), the DMDGP has been tested on instances related to the sequence of atoms $\mathrm{N} - \mathrm{C}_\alpha - \mathrm{C}$ of the protein backbones. As it is also supposed in many related works (see for example ([2], [13])), no distinctions among the different kinds of atoms (N, C, H, O, $\ldots$) were done. The computational experiments showed that instances related to the sequence $\mathrm{N} - \mathrm{C}_\alpha - \mathrm{C}$ can be almost always solved by the combinatorial approach, because the necessary assumptions are satisfied.

In order to perform more realistic experiments, we do not consider in this work the sequence $\mathrm{N} - \mathrm{C}_\alpha - \mathrm{C}$, but rather all the hydrogen atoms of the protein backbones. Indeed, it is very important to make a distinction among the atoms that are contained in protein backbones, because the majority of the distances detected by NMR are distances between pairs of hydrogens. Unfortunately, in general, if only the hydrogens of the protein backbones are considered, then the corresponding instance does not satisfy the assumptions of the DMDGP in the natural ordering given to the atoms.

The focus of this paper is a procedure for building artificial backbones formed by the hydrogen atoms associated to the protein backbones. We will show some orderings that can be given to the hydrogens and how they can make the needed assumptions satisfied. We will work in the simplified case in which the distances can be considered as accurate. The work here presented can be extended in order to manage experimental errors, by integrating, for example, the strategy presented in [10].

The paper is organized as follows. In Section II, we will outline an algorithm for solving the DMDGP, and emphasis will be given to the assumptions that must be satisfied in order to formulate the problem as DMDGP. In Section III, we will show how to generate an artificial backbone of hydrogens that satisfies the necessary assumptions. In Section IV, computational experiments on instances related to artificial backbones are shown. In Section V, we end with some conclusions.

II. The Branch and Prune algorithm

Let us suppose that some of distances between pairs of atoms of a molecule are known. Let $G = (V, E, d)$ be a weighted undirected graph, where

- there is a vertex $v \in V$ associated to each atom of the molecule, in a given ordering;
- there is an edge $(u, v) \in E$ if and only if the distance between u and v is known;
- the weights d associated to the edges provide the numerical values of the known distances.

The Molecular Distance Geometry Problem (MDGP) is the problem of finding a function $x : G \longrightarrow \Re^3$ such that the molecular conformation

$$X = \{x(v) : v \in V\}$$

satisfies all the distances d.

The MDGP can be formulated as a combinatorial problem if the following two assumptions are satisfied:

Assumption 1: all the distances $d_{i-3,i}$, $d_{i-2,i}$ and $d_{i-1,i}$ must be known,

Assumption 2: for each triplet of vertices $\{i-2, i-1, i\}$, the strict triangular inequality

$$d_{i-2,i} < d_{i-2,i-1} + d_{i-1,i}$$

must hold,

for a given ordering of the atoms of the molecule. Assumption 2 is satisfied in most of the cases. Indeed, if, for a certain triplet of consecutive vertices, $d_{i-2,i}$ were perfectly equal to $d_{i-2,i-1} + d_{i-1,i}$, then the corresponding three atoms would be perfectly aligned. The probability for this to happen is almost zero. Assumption 1 is harder to be satisfied. If data from NMR are considered, then only the distances smaller than 6Å are available, and therefore, if some of the distances $d_{i-3,i}$, $d_{i-2,i}$ and $d_{i-1,i}$ are large, then it cannot be detected and Assumption 1 may not be satisfied.

If both assumptions are satisfied, then it is possible to prove that the cosine of the torsion angle among four consecutive atoms $\{x_{i-3}, x_{i-2}, x_{i-1}, x_i\}$ of a protein backbone can be computed. If the atoms x_{i-3}, x_{i-2}, x_{i-1} are already placed into a fixed location, then, by exploiting all the known distances and the value of the torsion angle, the exact position of the atom x_i can be obtained. Unfortunately, the value of the torsion angle is not available, but only its cosine, which brings to two possible values for the angle. Because of this uncertainty, each atom x_i can be placed in two different positions. This allows to reformulate the MDGP as a combinatorial problem, to which we refer to as Discretizable Molecular Distance Geometry Problem (DMDGP). For more details, we refer the reader to ([7], [8], [9]).

Algorithm 1 BP algorithm

```
BP($i$, $n$, $d$)
compute the first atomic position for the $i^{th}$ atom: $x_i'$;
check the feasibility of the atomic position $x_i'$:
if ($|\,||x_i' - x_j|| - d_{ij}\,| < \varepsilon, \forall j < i$) then
  the atomic position $x_i$ is feasible;
  if ($i = n$) then
    a solution is found;
  else
    BP($i + 1$,$n$,$d$);
  end if
else
  the atomic position $x_i'$ is pruned;
end if
compute the second atomic position for the $i^{th}$ atom: $x_i''$;
check the feasibility of the atomic position $x_i''$:
if ($|\,||x_i'' - x_j|| - d_{ij}\,| < \varepsilon, \forall j < i$) then
  the atomic position $x_i'$ is feasible;
  if ($i = n$) then
    a solution is found;
  else
    BP($i + 1$,$n$,$d$);
  end if
else
  the atomic position $x_i''$ is pruned;
end if
```

In the combinatorial reformulation, a binary tree of possible solutions for the DMDGP can be defined. The BRANCH AND PRUNE (BP) algorithm [9] is based on this tree structure. The binary tree of possible solutions is explored starting from its top where the first atom of the conformation is placed, and the search proceeds by placing the following atoms one per time. As soon as a branch of the tree is found to be infeasible, then it is pruned and the search is backtracked. Because of the pruning phase, the size of the tree is reduced quickly and therefore an exhaustive search on the remaining branches is not too computational demanding.

Algorithm 1 provides a sketch of the BP algorithm. The algorithm is invoked iteratively, starting from the atomic position 4. The input parameters are i, the current atom whose position is searched; n, the total number of atoms; d, the set of known distances. One of the solutions to the problem is found when BP(n,n,d) finds one feasible position at least for the last atom of the conformation. The condition $|\,||x_i - x_j|| - d_{ij}\,| < \varepsilon$, for all $j < i$ and where $\varepsilon > 0$ is a given tolerance, represents a pruning test, which we employ for discovering infeasible atomic positions.

We showed in previous works that the BP algorithm is able to efficiently solve instances of the DMDGP. It is important to note that, even though it is able to find solutions of a global optimization problem, the BP algorithm does not exploit any objective function. Once solutions are found by BP, their quality can then be evaluated through, for example, the LDE function (1).

III. CONSTRUCTING ARTIFICIAL BACKBONES

Let us suppose that the sequence of atoms N$-$C$_\alpha$$-$C (defining the protein backbones) and all the hydrogens H which bound to such atoms are considered. Let $G = (V, E, d)$ be the associated weighted undirected graph. Since all the atoms detected by NMR are hydrogens, we can estimate the kind of atoms associated to vertices u and v such that $(u, v) \in E$. There can be indeed two possibilities:

- both the vertices refer to hydrogens, and, in this case, the distance d_{uv} must be computed by NMR;
- at least one of the vertices refers to an hydrogen, and the distance d_{uv} could be known *a priori*, because it may be the length of a chemical bond.

Let us consider the subgraph G_H, such that $G \supset G_H = (V_H, E_H, d_H)$ and such that G_H contains all the vertices in V to which at least two edges are associated. For what observed above, the graph G_H can contain hydrogen atoms only. Therefore, given an ordering on the vertices in V_H for which the assumptions 1 and 2 are satisfied, the MDGP can be formulated as a DMDGP, and solved by the BP algorithm. We will refer to the set of hydrogens associated to the vertices of the graph G_H as *artificial backbone* of hydrogens. We will show how a particular ordering on the vertices of G_H can make assumptions 1 and 2, needed for formulating the problem as combinatorial and applying BP, satisfied.

The main problem we need to solve is the following. The artificial backbone of hydrogens must satisfy both assumptions 1 and 2. As previously observed, there are a few possibilities to have Assumption 2 unsatisfied, and therefore we will not consider it in the following. Inversely, in order to have Assumption 1 satisfied, all the distances $d_{i-3,i}$, $d_{i-2,i}$ and $d_{i-1,i}$, for each i, must be known. Then, in the hypothesis that these distances come from an NMR experiment, they all need to be smaller than 6Å, because NMR can provide only distances smaller than this threshold. Thus, our main problem is to identify an artificial backbone of hydrogens such that the distances $d_{i-3,i}$, $d_{i-2,i}$ and $d_{i-1,i}$ are smaller than 6Å.

A protein is a chain of amino acids. The set of all common parts of the amino acids consists in a sequence of bound atoms which is usually referred to as *protein backbone*. Figure 1 shows the common part of each amino acid (the structure of the *proline* is slightly different, but all the following considerations can be applied anyway). As one can see from Figure 1, there are 4 hydrogens in the common part of each amino acid. However, during the protein synthesis, consecutive amino acids bind to each other through a peptide bond. During this process, one of the hydrogens bound to the nitrogen N and the group OH bound to C separate from the other atoms and form a water molecule (H_2O) [12]. Therefore, the common part of each amino acid in a protein contains two hydrogens only.

We will refer to the hydrogen bound to N with the symbol H, and we will refer to the hydrogen bound to C$_\alpha$ with the symbol HA. The most natural way for building an artificial backbone of hydrogens is to consider the sequence $\mathrm{H} - \mathrm{HA}$ one amino

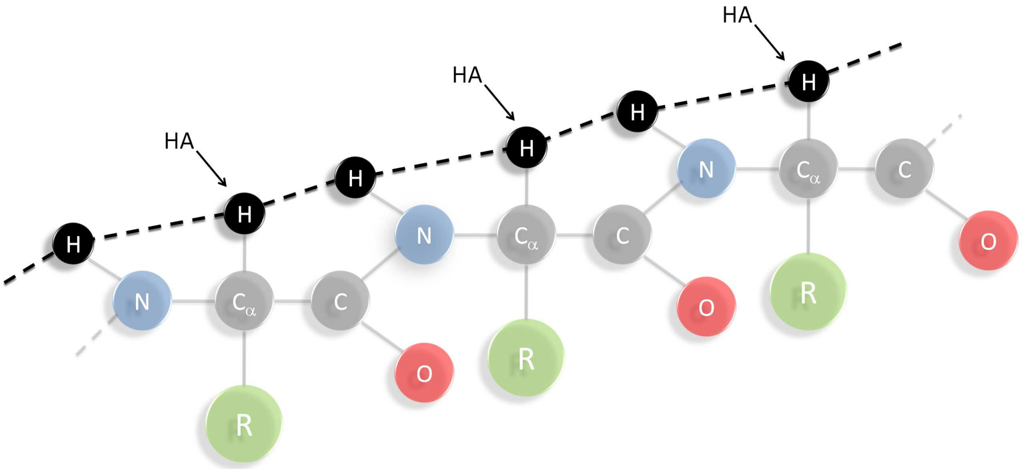

Fig. 2. An artificial backbone created by considering the sequence of hydrogens in the common parts of the amino acids in their natural ordering.

acid at a time, in the ordering defined by the protein backbone (see Figure 2). Unfortunately, this artificial backbone does not satisfy Assumption 1 in most of the cases. Indeed, simple geometric considerations show that the distance between the hydrogen H of the k-th amino acid and the hydrogen HA of the $(k+1)$-st amino acid cannot be smaller than 6Å, except for very particular cases. The same observation holds if the hydrogen HA of the k-th amino acid and the hydrogen H of the $(k+2)$-nd amino acid are considered. Therefore, in general, not all the needed distances are available if this artificial backbone is considered, and then Assumption 1 cannot be satisfied.

In order to overcome this problem, we will consider a third hydrogen for each amino acid. This hydrogen is borrowed from the group **R** of the amino acids, which is also called *side chain* of the amino acid (see Figure 1). We will refer to this hydrogen by the symbol HB. The group **R** is bound to the common part of the amino acid through a carbon atom called C_β. The only exception is given by *glycine*, whose side chain consists in only one hydrogen atom. In the particular case of *glycine*, we consider as third hydrogen the only one that forms its side chain. In general, one hydrogen HB at least is bound to the carbon C_β, and we consider one of them in our artificial backbone.

The artificial backbone we consider is the one in Figure 3. A label is associated to each arrow for specifying the ordering given to the hydrogens. As the figure shows, the artificial backbone considers more than once some of the hydrogens, in order to reduce the relative distances between the hydrogens contained into the quadruplets $\{x_{i-3}, x_{i-2}, x_{i-1}, x_i\}$. Algorithm 2 shows the set of instructions for generating the artificial backbone in Figure 3 starting from a known protein conformation X. We suppose that all the coordinates of the atoms of the protein are stored in a PDB file, which is a standard text file, used for storing the list of coordinates of the atoms forming a protein.

Algorithm 2 reads the information about the hydrogens of the backbone of a protein from a PDB file. Note that more than one hydrogen HA and HB can be found for the same amino acid. The algorithm reads the first HA or HB, and then it substitutes the hydrogens if another HA or HB is found. In our implementation, this substitution is allowed only when the new HA or HB is closer to the previous atoms of the artificial backbone. Algorithm 2 creates an instance of the DMDGP where only the hydrogen atoms of the protein backbones are considered.

The artificial backbones generated by Algorithm 2 have particular properties. Since some of the hydrogens are considered twice, some of the relative distances between them are perfectly zero. If one of the distances between the atoms in the generic triplet $\{x_{i-2}, x_{i-1}, x_i\}$ is zero, then two atoms coincide and, as a consequence, the atoms of the triplet lie on the same straight line (this goes against Assumption 2). For this reason, the artificial backbone is built in a way that only distances d_{ij}, with $j > i+2$, can be zero.

Since there are distances equal to zero, the LDE function (1) cannot be used for evaluating the performances of the BP algorithm on the generated distances, because there would be divisions by zero. Therefore, in the experiments showed in the next section, we will consider a modified version of the LDE function, in which the terms that would contain the divisions by zero are discarded.

Finally, note that the nitrogen atom N and the carbon atom C_α of the first amino acid are also included in the artificial backbone (see Figure 3). The distances between these two atoms and their following three atoms on the artificial backbone are known a priori, and hence they do not need to be detected experimentally. We decided to add these two atoms for the following reason. Once the coordinates of the atoms of the artificial backbone are identified by an algorithm such as BP, then the coordinates of the atoms of the real protein backbone can be computed. The atoms N, C_α and the first H define a common coordinate system for all the hydrogens and the other backbone atoms.

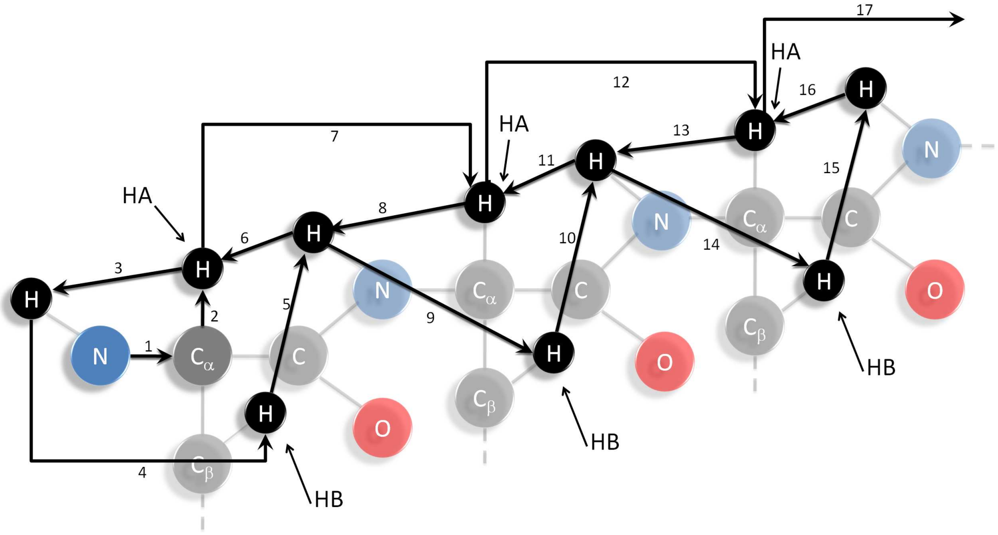

Fig. 3. An artificial backbone providing an ordering such that the assumptions for the DMDGP are satisfied.

IV. Computational Experiments

In this section, instances of the DMDGP generated by applying Algorithm 2 on a set of known proteins are solved by the BP algorithm. We consider a subset of monomeric proteins downloaded from the Protein Data Bank (PDB) ([1], [11]), where all the selected proteins have been experimentally obtained by NMR. All the codes were written in C programming language and all the experiments were carried out on an Intel Core 2 CPU 6400 @ 2.13 GHz with 4GB RAM, running Linux. The codes have been compiled by the GNU C compiler v.4.1.2 with the `-O3` flag.

Table I shows the obtained results. *Protein name* refers to the label given to the considered protein in the PDB. #Sol is the number of found solutions. The LDE function (modified in order to avoid divisions by zero) is used for evaluating the quality of the solutions and the best one is showed in Table I. Finally, the CPU time (in seconds) is given for each experiment.

The number of solutions is at least 8 in each experiment. This is due to the fact that no distances d_{ij}, with $j > i+3$, are given in correspondence with the first two atoms. Then, the first computed atomic positions can never be pruned, leading to multiple solutions. The LDE function indicates that the found solutions are very accurate. This proves that the hydrogens of the protein backbones can be efficiently identified by the BP algorithm if they are organized on a suitable artificial backbone satisfying the necessary assumptions.

It is important to note that the artificial backbones generated by Algorithm 2 can also be identified by exploiting the data obtained by NMR experiments, and the corresponding instances can then be solved by applying BP. In other words, supposing that NMR provided distances between the hydrogens of a protein backbone, and supposing that these distances are accurate, then BP can be used for finding the coordinates of such hydrogens, if they are considered in the ordering given in Figure 3. This is what our computational experiences show. However, we are not able yet to perform experiments in which real data from NMR are considered, because we supposed so far that all the given distances are accurate. The work that is here presented can be extended in order to consider the realistic case in which the given distances are not accurate (see for example [10]). These preliminary experiments show that our way to approach to the problem is promising.

protein name	#Sol	LDE	time
`1a11`	8	2.79e-15	0.00
`1bbl`	8	3.94e-15	0.00
`1k1v`	8	4.46e-15	0.01
`1jkz`	8	8.05e-15	0.36
`1bqx`	8	1.38e-14	0.02
`1b4c`	16	4.40e-15	0.04
`2hsy`	8	6.79e-14	0.06
`1itm`	8	6.98e-14	0.03
`1ngl`	32	4.78e-14	63.94
`1a23`	8	3.08e-14	0.71
`2ron`	16	2.26e-14	1.69
`1d8v`	8	4.59e-14	0.19
`1q8k`	64	2.70e-13	24.11
`1ezo`	8	1.29e-13	94.60

TABLE I
BP (Algorithm 1) applied to the artificial backbones obtained by Algorithm 2.

V. Conclusions

We presented a strategy for building artificial backbones associated to the hydrogens of real backbones of protein molecules. The aim is to find an ordering for the hydrogen atoms, so that the distances usually detected by NMR can be exploited for creating an instance of the DMDGP. This is not trivial, because two particular assumptions must be satisfied in order to generate an instance of the DMDGP.

Algorithm 2 creating artificial backbones

```
procedure(input: PDB file)

# reading information on the hydrogens H, HA and HB
let $n = 0$;
open PDB file;
for (each amino acid in the PDB file) do
   let $n = n + 1$;
   for (each hydrogen atom) do
      let $\ell$ = atom label;
      let $(x, y, z)$ = atom coordinates;
      if ($\ell$ = H) then
         H[$n$] = $(x, y, z)$;
      end if
      if ($\ell$ = HA) then
         HA[$n$] = $(x, y, z)$;
      end if
      if ($\ell$ = HB) then
         HB[$n$] = $(x, y, z)$;
      end if
   end for
end for

# creating the artificial backbone
let X[1] = ($(0, 0, 0)$,'N');
let X[2] = ($(-1.458, 0, 0)$,'CA');
let X[3] = (HA[1],'HA');
let X[4] = (H[1],'H');
let X[5] = (HB[1],'HB');
let $m = 5$;
for ($i = 2, n$) do
   let X[$m + 1$] = (H[$i$],'H');
   let X[$m + 2$] = (HA[$i$],'HA');
   let X[$m + 3$] = (HA[$i - 1$],'HA*');
   let X[$m + 4$] = (H[$i$],'H*');
   let X[$m + 5$] = (HB[$i$],'HB');
   let $m = m + 5$;
end for

# creating an instance for the DMDGP
compute all the distances $d_{ij}$ between all the pairs in
   $$\mathrm{X}[i], i = 1, 2, \ldots, m;$$
keep all the distances $d_{ij}$ such that
   $$j \leq i + 3 \quad \text{or} \quad d(i, j) < 6;$$
the obtained distances form an instance for the DMDGP.
```

We investigated different artificial backbones of hydrogens, and we found a particular ordering that makes the necessary assumptions satisfied in most of the cases. The BP algorithm is used to solve the corresponding instances by providing solutions having a high accuracy. Each solution consists in a set of coordinates for the hydrogens of the protein backbones.

The results discussed in this paper are very promising, because they show how experimentally obtained data can be used for identifying the conformations of the proteins. Indeed, once the coordinates of the hydrogens have been computed by BP, the remaining of the backbone atoms, i.e. the sequence of atoms $\mathrm{N} - \mathrm{C}_\alpha - \mathrm{C}$, can be built by exploiting some geometric observations. We are currently working on a method for automatically reconstructing the whole protein backbone which exploits the information on the positions of its hydrogens.

Acknowledgments

The authors would like to thank the Brazilian research agencies FAPESP and CNPq, the French research agency CNRS and École Polytechnique, for financial support.

References

[1] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *The Protein Data Bank*, Nucleic Acids Research **28**, 235–242, 2000.

[2] P. Biswas, K.-C. Toh, and Y. Ye, *A Distributed SDP Approach for Large-Scale Noisy Anchor-Free Graph Realization with Applications to Molecular Conformation*, SIAM Journal on Scientific Computing **30**, 1251–1277, 2008.

[3] G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation*, John Wiley & Sons, New York, 1988.

[4] Q. Dong, Z. Wu, *A Linear-Time Algorithm for Solving the Molecular Distance Geometry Problem with Exact Inter-Atomic Distances*, Journal of Global Optimization **22**, 365–375, 2002.

[5] T.F. Havel, *Distance Geometry*, D.M. Grant and R.K. Harris (Eds.), Encyclopedia of Nuclear Magnetic Resonance, Wiley, New York, 1701-1710, 1995.

[6] C. Lavor, L. Liberti, and N. Maculan, *Molecular distance geometry problem*, In: Encyclopedia of Optimization, C. Floudas and P. Pardalos (Eds.), 2^{nd} edition, Springer, New York, 2305–2311, 2009.

[7] C. Lavor, L. Liberti, and N. Maculan, *Discretizable Molecular Distance Geometry Problem*, Tech. Rep. q-bio.BM/0608012, arXiv, 2006.

[8] C. Lavor, L. Liberti, A. Mucherino, and N. Maculan, *On a Discretizable Subclass of Instances of the Molecular Distance Geometry Problem*, ACM Conference Proceedings, 24^{th} Annual ACM Symposium on Applied Computing (SAC09), Hawaii USA, 804–805, 2009.

[9] L. Liberti, C. Lavor, and N. Maculan, *A Branch-and-Prune Algorithm for the Molecular Distance Geometry Problem*, International Transactions in Operational Research **15** (1), 1–17, 2008.

[10] A. Mucherino, L. Liberti, C. Lavor, and N. Maculan, *Comparisons between an Exact and a MetaHeuristic Algorithm for the Molecular Distance Geometry Problem*, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO09), Montréal, Canada, July 2009.

[11] Protein Data Bank: `http://www.rcsb.org/pdb/`

[12] T. Schlick, *Molecular Modelling and Simulation: an Interdisciplinary Guide*, Springer, New York, 2002.

[13] D. Wu and Z. Wu, *An Updated Geometric Build-Up Algorithm for Solving the Molecular Distance Geometry Problem with Sparse Distance Data*, Journal of Global Optimization **37**, 661–673, 2007.