M a t h e m a t i c a
B a l k a n i c a

New Series Vol. ..., 200.., Fasc. .....

On the Definition of Artificial Backbones for the Discretizable Molecular Distance Geometry Problem

Antonio Mucherino, Carlile Lavor, Leo Liberti, Nelson Maculan

Presented by Antonio Mucherino

Finding the conformation of a molecule is one of the major challenges in chemistry and biology. Information obtained by NMR experiments can be used to provide estimates of some of the distances between the atoms forming the molecule. The conformation of the molecule can be found by solving the corresponding distance geometry problem. In this work, we focus our attention on protein conformations. We show how an artificial backbone of atoms can be defined for exploiting data from NMR in order to reformulate the distance geometry problem as combinatorial. We formally prove that this artificial backbone can only contain hydrogen atoms, and we introduce a particular ordering for such hydrogens. Computational experiments on a set of artificially generated instances are presented.

1. 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. Indeed, this is one of the major challenges in chemistry and biology.

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*

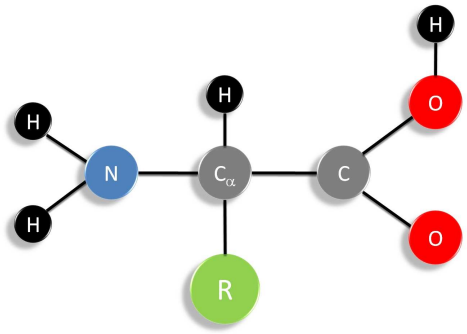

Figure 1: The general structure of an amino acid.

of each amino acid. All these parts define the so-called *backbone* of the protein. 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 during protein synthesis, some of the atoms of their common parts are lost, while 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.

A way for discovering the conformation of a molecule is through experiments of Nuclear Magnetic Resonance (NMR). Indeed, the information obtained by this experimental technique can be used for estimating the distances between pairs of atoms forming a given molecule [5]. The distances that can be estimated have particular properties. First, these distances are usually not larger than 6Å. This limits the number of available distances and makes the problem of finding the molecular conformation harder. Moreover, in a protein backbone formed by carbons, nitrogens, oxygens and hydrogens, most of the known distances refer to a kind of atom only: the hydrogen.

The focus of this paper is on the following problem: can we find the coordinates of the backbone atoms of a protein starting from the distances (found through NMR) between some pairs of its hydrogens? This is a distance geometry problem. Since we consider molecules, and in particular proteins, this problem is known in the literature as the Molecular Distance Geometry Problem (MDGP) [3].

Let

$$X = \{x_1, x_2, \ldots, x_n\}$$

be a protein conformation, where x_i is the generic vector (x_i^1, x_i^2, x_i^3) of coordinates for 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):

$$LDE(X) = \frac{1}{m} \sum_{\{i,j\}} \frac{||x_i - x_j|| - d_{ij}}{d_{ij}},$$

where m is the total number of known distances ($m = |E|$). 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.

Different methods have been proposed over time for solving this global optimization problem. One of the difficulties to be faced is that the LDE function (and even other penalty functions used in this context) has many local minima, where a method for optimization can get stuck at. In order to overcome this problem, in [11,12], for example, the used penalty function is approximated by smoother functions converging to the original function. In this way, the search is guided torward the global optimum. Another method for the MDGP makes use of a penalty function which can be seen as the difference of two convex functions [1], and particular techniques for *d.c.* optimization are used. More recently, in [4], a Population Basin Hopping method is employed, in which basic concepts (such as the ones of *funnel* and *funnel bottom*) are used, as in many other methods for molecular conformations. Note that many of these methods are based on a continuous representation of the problem, and that deterministic methods are often employed. However, there are meta-heuristic algorithms particularly designed for the solving the MDGP, such as, for example, the SPE algorithm [16]. For a survey on methods and algorithms for the MDGP, the reader is refereed to [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 problem. In this way, the search domain is reduced to a discrete set, and an ad-hoc algorithm can be used for solving the combinatorial problem. Computational experiments, presented for example in [7,8,10], showed that the combinatorial approach to the MDGP is more efficient than the continuous one. We refer to this combinatorial reformulation of the MDGP as the Discretizable Molecular Distance Geometry Problem (DMDGP).

In this paper, we will show how to build an artificial backbone of atoms that satisfies the necessary assumptions for having the combinatorial reformulation. We will prove that only hydrogen atoms must be included in the artificial backbone in order to have the assumptions satisfied. As a consequence, the problem of finding the coordinates of all the backbone atoms can be divided in two stages. In the first one, the coordinates of the hydrogens belonging to the protein backbone can be obtained by solving a DMDGP, where an artificial backbone is defined and considered. Then, the coordinates of all the other atoms can be identified by solving a different MDGP, where the coordinates of the hydrogens and some distances, known a priori, are exploited. This paper mainly focuses on the first stage, and preliminary studies can be found in [9].

Note that we will work in the simplified case in which the provided distances can be considered as accurate. The work here presented can be extended in order to manage experimental errors. For example, the strategy presented in [13] for handling inaccurate distances could be used, as well as the strategy presented in [14] in which wrong distances are automatically discarded.

The paper is organized as follows. In Section 2, we will give an outline of 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 a DMDGP. In Section 3, we will show how to generate an artificial backbone of atoms that satisfies the necessary assumptions, and we will prove that only hydrogens can be included in the artificial backbone. In Section 4, computational experiments on instances related to artificial backbones are shown. Finally, in Section 5, we end with some conclusions.

2. The Branch and Prune algorithm

Let us suppose that a set of distances between pairs of atoms of a protein backbone have been obtained through NMR experiments. Let $G = (V, E, d)$ be a weighted undirected graph, where

- there is a vertex $i \in V$ associated to each atom of the protein backbone, in a given ordering;

- there is an edge $(i, j) \in E$ if and only if the distance between i and j is known;
- the weights d associated to the edges provide the numerical values of the known distances.

The MDGP is the problem of finding a function $x : G \to \Re^3$ such that the molecular conformation

$$X = \{x_i : i \in V\}$$

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 zero. Assumption 1 is harder to be satisfied. As already observed, when 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. A binary tree of atomic positions can be defined and explored with the aim of finding solutions to the problem. Since the search domain is a binary tree, the problem to be solved is a combinatorial optimization problem, to which we refer to as DMDGP. For more details, we refer the reader to [7,8,10].

The Branch and Prune (BP) algorithm [10] is an exact algorithm for the DMDGP. In the algorithm, the binary tree of possible solutions is explored,

```
BP (i, n, d)
for k = 1,2 do
   compute the k^th atomic position for the i^th atom: x_i^k;
   check the feasibility of the atomic position x_i^k:
   if (| ||x'_i − x_j|| − d_ij | < ε, ∀j < i) then
      the atomic position x_i^k is feasible;
      if (i = n) then
         a solution is found;
      else
         BP (i + 1,n,d);
      endif
   else
      the current branch is pruned;
   endif
endfor
```

Figure 2: The BP algorithm.

where the search proceeds by placing one atom 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 computationally demanding.

Figure 2 provides a sketch of the BP algorithm. The first atom can be placed in the origin of the coordinate system. The second and the third atom (see [10] for more details) can also be uniquely defined, so that solutions that can be obtained by translating and rotating other solutions are avoided. Then, the BP algorithm is invoked iteratively, starting from the atomic position 4. The input parameters of the algorithm 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

$$| \, ||x_i - x_j|| - d_{ij} \, | < \varepsilon, \quad \forall j < i,$$

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.

3. Building artificial backbones

Let G be the graph associated to an instance of the DMDGP. Let us suppose that all the available distances regarding the backbone of a protein (including the atoms N, C_α, C and all the hydrogens) are considered. The majority of these distances need to be detected by NMR in order to solve the problem, whereas some of them are known a priori. As an example, all the bond lengths are already known. Moreover, if an atom is bound to two atoms, the angle between the two chemical bonds is known, and it can be exploited for computing the distance between the two atoms bound to the same one. For example, all the distances $d_{i-1,i}$ and $d_{i-2,i}$ related to the sequence of atoms $N-C_\alpha-C$ can be computed.

In order to reformulate the problem as a combinatorial problem, we need that the two necessary assumptions are satisfied. The probability of having Assumption 2 unsatisfied is zero, and therefore we do not consider it in the following. Assumption 1 requires that, for each atom x_i, there are at least three edges in E that precede the vertex i in the given ordering and that are incident to i. We will show that some of the vertices that are contained in G cannot satisfy this assumption.

Let G_H be the subgraph of G in which at least three edges (j, i), with $j < i$, are incident to each vertex $i \in V_H$ such that $i > 3$. We will refer to the set of atoms associated to G_H as *artificial backbone*.

Theorem 1 *In the hypothesis that only distances between hydrogens are found through NMR experiments, the artificial backbones satisfying Assumption 1 have only hydrogen atoms associated to the vertices $i > 3$.*

Proof. Let i be a vertex of V_H such that $i > 3$. Let us suppose that the corresponding atom x_i is not a hydrogen. Since it is not a hydrogen, the distances between x_i and other atoms of the artificial backbone cannot be detected by NMR, and therefore they need to be known a priori. In the case in which the pairs of atoms (x_{i-3}, x_{i-2}), (x_{i-2}, x_{i-1}) and (x_{i-1}, x_i) are chemically bound, we are able to get the maximum possible information regarding the distances. Indeed, all the distances $d_{i-3,i-2}$, $d_{i-2,i-1}$ and $d_{i-1,i}$ are bond lengths, and hence they are known. Moreover, x_{i-3} and x_{i-1} are both bound to x_{i-2}, and then their relative distance $d_{i-3,i-1}$ can be computed. Similarly, the distance between x_{i-2} and x_i can also be computed. The last distance that must be known in order

to have Assumption 1 satisfied is the distance $d_{i-3,i}$. However, this distance cannot be known a priori, because it depends on the torsion angle among the atoms x_{i-3}, x_{i-2}, x_{i-1} and x_i. Therefore, Assumption 1 cannot be satisfied.

If we suppose that some of the pairs of atoms (x_{i-3}, x_{i-2}), (x_{i-2}, x_{i-1}) and (x_{i-1}, x_i) are not chemically bound, then also other needed distances may be absent. As a consequence, Assumption 1 can never be satisfied if x_i is not a hydrogen. This proves that only hydrogens can be assigned to vertices of the graph G_H such that $i > 3$, in the hypothesis that only distances between hydrogens are found through NMR experiments. ∎

Note that only atoms x_i with $i > 3$ are considered in the theorem, and therefore atoms that are not hydrogens may be associated to the vertices 1, 2 or 3 of G_H. Moreover, the majority of the considered distances must come from NMR experiments, even if some of the distances between hydrogens can be known a priori (this is quite rare). A hydrogen can be bound to only one atom, but more than one hydrogen can be bound to the same atom (see for example a molecule of water: H_2O). In the latter case, the distance between the two hydrogens can be computed, because they are bound to a common atom. However, this information could also be obtained by NMR, because the two hydrogens are close enough to be detected. Therefore, the distance between these two hydrogens is usually detected with all the others by the experimental technique.

Recall that 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 that defines the 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) [15]. 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 to the hydrogen bound to C_α with the symbol HA.

We also 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

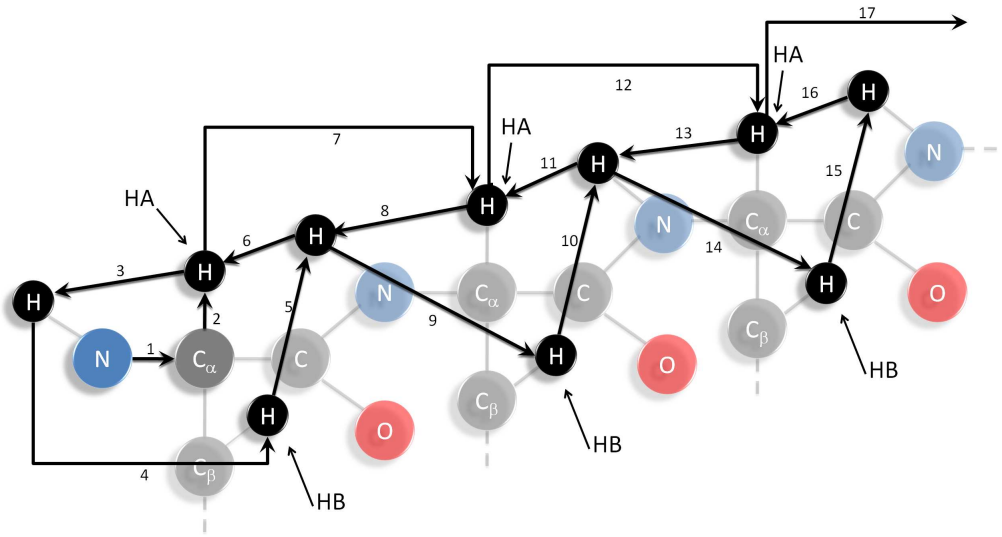

Figure 3: An artificial backbone providing an ordering such that the assumptions for the DMDGP are satisfied. Note that some of the hydrogens are considered twice and that the considered ordering is specified through the labels associated to the edges.

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.

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 $\{x_{i-3}, x_{i-2}, x_{i-1}, x_i\}$. As a consequence, some of the relative distances between the hydrogens 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, hence, 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.

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). As Theorem 1 shows, in order to have the necessary assumptions satisfied, all the atoms with a rank greater than 3 must be hydrogens, whereas the first three atoms can be

of any kind. The distances between these two atoms and their following three atoms on the artificial backbone are known a priori, and therefore they do not need to be detected experimentally by NMR. We decided to add these two atoms because they, together with the first hydrogen H, define a common coordinate system for all the hydrogens and the other backbone atoms.

Note that, if the artificial backbone in Figure 3 is considered, there are no distances between hydrogens that are known a priori, but all of them need to be found by NMR. Indeed, we have three different kinds of hydrogens: H, HA and HB. H is bound to N, and there are no other hydrogens bound to the same N. The same observation can be made for HA, which is bound to C_α. Finally, HB is taken from the side chain of the amino acid, and, depending on the kind of amino acid, it could be bound, together with other hydrogens, to the same carbon atom. However, these other hydrogens are not considered (for each C_β, only one of its hydrogens is chosen), and therefore their relative distances are not needed.

The LDE function is usually used for evaluating the quality of the solutions to the DMDGP. However, when artificial backbones are used, there are divisions by zero when the LDE function is evaluated. Thus, in the experiments showed in the next section, we will consider a modified version of the LDE function, in which the terms that contain the divisions by zero are discarded.

4. Computational experiments

We will show in this section how instances of the DMDGP related to artificial backbones can be efficiently solved by the BP algorithm. 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.

The instances we consider are generated from known conformations of proteins. Protein conformations can be downloaded from a public database, the Protein Data Bank (PDB) [2]. They are stored in text files called `pdb` files, where, among other information, the coordinates of the atoms forming the molecule are specified. In order to generate an instance of the DMDGP, we used the following procedure:

- we downloaded the `pdb` file of a given protein conformation;
- we extracted all the atoms considered in the artificial backbone;
- we sorted the atoms as described in Figure 3;

protein name	n	m	LDE	time
1a11	125	1680	2.79e-15	0.00
1bbl	185	2241	3.94e-15	0.00
1k1v	205	2676	4.46e-15	0.01
1jkz	230	2968	8.05e-15	0.36
1bqx	385	5096	1.38e-14	0.02
1b4c	460	6105	4.40e-15	0.04
2hsy	520	7057	6.79e-14	0.06
1itm	650	9562	6.98e-14	0.03
1ngl	895	11760	4.78e-14	63.94
1a23	945	13839	3.08e-14	0.71
2ron	1210	16378	2.26e-14	1.69
1d8v	1315	18526	4.59e-14	0.19
1q8k	1500	21401	2.70e-13	24.11
1ezo	1850	25299	1.29e-13	94.60

Table 1: The BP algorithm applied to a set of artificial backbones obtained from known protein conformations.

- we computed the distances between all the atoms;
- we kept all the distances smaller than 6Å.

We performed this procedure on a set of protein conformations (for more details, the reader is referred to [9]).

All the instances we generated belong to the class of instances of the DMDGP. We applied the BP algorithm for solving such instances, and the computational experiments are shown in Table 1. In the table, n is the number of atoms included in the instance. It is always a number which is divisible by 5, because each amino acid of the considered artificial backbone contains exactly 5 hydrogens (two of them are considered twice). The cardinality of the set of edges E corresponds to the number m of known distances. 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 the table. Finally, the CPU time (in seconds) is given for each experiment.

The experiments show that the BP algorithm is very efficient in finding solutions of the DMDGP in terms of quality of the solutions and CPU time, as already shown in previous works. In these experiments, each solution consists of the set of coordinates of the hydrogen atoms H, HA and HB of the artificial backbones. Such coordinates, together with some distances known a priori between the hydrogens and the other backbone atoms, could be used for finding

the entire conformation of the protein backbones.

5. Conclusions

We presented an artificial backbone of atoms associated to protein backbones for which the assumptions for the DMDGP are satisfied. We proved that such an artificial backbone can only contain hydrogen atoms, exception made for the first three atoms in the sequence. We showed 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 formulate the problem as a DMDGP.

In order to identify the coordinates of all the backbone atoms (including the sequence $N - C_\alpha - C$), another MDGP could be formulated, where all the coordinates of the hydrogen atoms and some distances known a priori can be exploited. We are currently working on this MDGP. Our hope is to present in future publications a method which is able to reconstruct the whole protein backbone from the information obtained through NMR experiments only.

Acknowledgments

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

References

[1] L.T.H. An, P.D. Tao, *Large-Scale Molecular Optimization from Distance Matrices by a D.C. Optimization Approach*, SIAM Journal on Optimization **14**, 2003, 77–114.

[2] 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**, 2000, 235–242.

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

[4] A. Grosso, M. Locatelli, F. Schoen, *Solving Molecular Distance Geometry Problems by Global Optimization Algorithms*, Computational Optimization and Applications **43**, 2009, 23–27.

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

[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, 2009, 2305–2311.

[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, 2009, 804–805.

[9] C. Lavor, A. Mucherino, L. Liberti, and N. Maculan, *Computing Artificial Backbones of Hydrogen Atoms in order to Discover Protein Backbones*, IEEE Conference Proceedings, International Multiconference on Computer Science and Information Technology (IMCSIT09), Workshop on Computational Optimization (WCO09), Mragowo, Poland, October 2009.

[10] L. Liberti, C. Lavor, and N. Maculan, *A Branch-and-Prune Algorithm for the Molecular Distance Geometry Problem*, International Transactions in Operational Research, **15**, No 1, 2008, 1–17.

[11] J.J. Moré, Z. Wu, *Smoothing Techniques for Macromolecular Global Optimization*. In: G. Di Pillo and F. Gianessi, (Eds.), Nonlinear Optimization and Applications, Plenum Press, New York, 1996, 297–312.

[12] J.J. Moré, Z. Wu, *Global Continuation for Distance Geometry Problems*, SIAM Journal on Optimization **7**, 1997, 814–836.

[13] A. Mucherino, C. Lavor, *The Branch and Prune Algorithm for the Molecular Distance Geometry Problem with Inexact Distances*, Proceedings of World Academy of Science, Engineering and Technology (WASET), International Conference on Computer, Electrical, and Systems Science, and Engineering (CESSE09), International Conference on Computational Biology (ICCB09), Venice, Italy, October 2009.

[14] 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, 2009, 333–340.

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

[16] H. Xu, S. Izrailev, D.K. Agrafiotis, *Conformational Sampling by Self-Organization*, Journal of Chemical Information and Computer Sciences **43**, 2003, 1186–1191.