(19) **Europäisches Patentamt**
European Patent Office
Office européen des brevets

(11) **EP 1 488 364 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
05.11.2008 Bulletin 2008/45

(51) Int Cl.:
G06F 19/00 (2006.01)

(21) Application number: **02718498.5**

(22) Date of filing: **26.03.2002**

(86) International application number:
PCT/IN2002/000064

(87) International publication number:
WO 2003/081510 (02.10.2003 Gazette 2003/40)

(54) **METHOD AND SYSTEM TO BUILD OPTIMAL MODELS OF 3-DIMENSIONAL MOLECULAR STRUCTURES**

VERFAHREN UND SYSTEM ZUM ERSTELLEN VON OPTIMALEN MODELLEN FÜR DREIDIMENSIONALE, MOLEKULARE STRUKTUREN

PROCEDE ET SYSTEME DE CONSTRUCTION DE MODELES OPTIMAUX DE STRUCTURES MOLECULAIRES TRIDIMENSIONNELLES

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(43) Date of publication of application:
22.12.2004 Bulletin 2004/52

(73) Proprietor: **Council of Scientific and Industrial Research**
New Delhi 110 001 (IN)

(72) Inventors:
• **GAUTHAM, N.**
Department of Crystallography and
Chennai 600 025
Tamil Nadu (IN)
• **VENGADESAN, K.**
Department of Crystallography and
Chennai 600 025
Tamil Nadu (IN)

(74) Representative: **Lamb, Martin John Carstairs**
Marks & Clerk
90 Long Acre
London WC2E 9RA (GB)

(56) References cited:
WO-A-99/26901

• **GAUTHAM N ET AL: "Global Search for Optimal Biomolecular Structures using Mutually Orthogonal Latin Squares: A Novel Algorithm." CURRENT SCIENCE (BANGALORE), vol. 63, no. 9-10, 1992, pages 560-564, XP008015566 ISSN: 0011-3891 cited in the application**
• **JOHNSON D M ET AL: "ORTHOMORPHISMS OF GROUPS AND ORTHOGONAL LATIN SQUARES. I" CANADIAN JOURNAL OF MATHEMATICS, UNIVERSITY OF TORONTO, TORONTO,, CA, vol. 13, 1961, pages 356-372,ABSTRACTS, XP002922312 ISSN: 0008-414X**

EP 1 488 364 B1

Description

Field of the invention

5   **[0001]**   The present invention relates to a method to build 3-dimensional models of molecular structures corresponding to the lowest value of the potential energy function from knowledge of their chemical structures. The present invention specifies a method that may be used to build a model of the three-dimensional structure of any molecule including, but not limited to, peptides, proteins, nucleotides, nucleic acids, antibiotics, saccharides, etc. The system uses Mutually Orthogonal Latin Squares to search through conformational space to obtain the best conformation. The system may

10   also be used to obtain the optimum of any function, and not only those related to biomolecular structure. The method is incorporated into a computer and this system forms the present invention.

Background of the invention

15   **[0002]**   United States Patent No. 5,884,230 to Srinivasan et al. titled 'Method and system for protein modeling' discloses a method in a computer system for modeling a three-dimensional structure of a model protein in which the modeling is based upon a three-dimensional structure of a template protein and an amino acid sequence alignment of the model protein and the template protein. For each amino acid in the model protein, when the template protein has an amino acid aligned with the amino acid of the model protein, the position of the backbone atom of the amino acid aligned with

20   the amino acid of the model protein is established based on the position of a topologically equivalent backbone atom in the aligned amino acid of the template protein. However, this method works only, when the structure of a model protein is already known. Therefore it is not applicable to smaller molecules such as oligopeptides, or to molecules that are chemically different, such as nucleic acids, saccharides, antibiotics, and other organic molecules. Even in the case of protein structure, the method is not applicable to *ab initio* protein structure determination, in other words, the method

25   cannot be applied if there is no model structure from which to start.
       [0003]   United States Patent No. 6,188,956 to Mayo, et al. titled 'Apparatus and method for automated protein design' relates relates to apparatus and methods for quantitative protein design and optimization. The method receives a known protein backbone structure and builds the conformations of the side chains using the Dead-End Elimination method. As above, this method also would work only when the structure of a model protein is already known. Therefore it is not

30   applicable to smaller molecules such as oligopeptides, or to molecules that are chemically different, such as nucleic acids, saccharides, antibiotics, and other organic molecules. Even in the case of protein structure, the method not applicable to *ab initio* protein structure determination, in other words, if there is no model structure from which to start, the method cannot be applied since it requires an already known protein backbone conformation.
       [0004]   '*Combinatorial Mathematics*', H. J. Ryster, a textbook of combinatorial mathematics gives a definition of Mutually

35   Orthogonal Latin Squares together with an equation to construct them. However, there is no explanation of how the equation is to be incorporated into a computer program. Also there is no mention of any application of Mutually Orthogonal Latin Squares (MOLS) to find the best value of any function, including the potential energy function. Also there is no mention of any application to determination of the three-dimensional structure of a molecule structure. Also it only describes how to distribute a set of symbols so as to form a set of MOLS, but it does not describe any method to replace

40   the symbols with the values of the parameters that define any function. Also it does not describe how the calculations are to be analyzed after the construction of the MOLS.
       [0005]   '*Introduction to Combinatorial Mathematics*', by C. L. Liu is another textbook reference on combinatorial mathematics which gives a definition of Mutually Orthogonal Latin Squares together with an equation to construct them. Again however, there is no explanation of how the equation is to be incorporated into a computer program. Also there is no

45   mention of any application of Mutually Orthogonal Latin Squares (MOLS) to find the best value of any function, including the potential energy function. Also there is no mention of any application to determination of the three-dimensional structure of a molecule. Also it only describes how to distribute a set of symbols so as to form a set of MOLS, but it does not describe any method to replace the symbols with the values of the parameters that define any function. Also it does not describe how the calculations are to be analyzed after the construction of the MOLS.

50   **[0006]**   '*Experimental Design and Its Statistical Basis*', D. J. Finney is a textbook on experimental design which gives a method to use Mutually Orthogonal Latin Squares in agricultural and pharmaceutical experiments. However, there is no teaching herein of how MOLS are to be constructed, nor does it specify any computer program to do so, nor does it mention any application to determination of molecular structure or to finding the best value of a function. Also the analyses of the calculations made in each sub square as stated in this book are more complicated than the one used in the present

55   invention.
       [0007]   'Genetic Algorithms and Protein Folding' by S. Schulze-Kremer, Methods in Molecular Biology (D Webster, ed.) 2000 describes describes the application of genetic algorithms in protein structure prediction. However, the method described fails for *ab initio* structure prediction. That is, if there is no previous information about the secondary structure,

the method does not work. Also the method does not use MOLS in any form in the calculations, and hence is different from the present invention. 'Global Search for Optimal Biomolecular Structures using Mutually Orthogonal Latin Squares' by N. Gautham and Z. A. Rafi, Current Science, 1992 is a very preliminary report of the possible use of MOLS in structure prediction. The method was however, crude and not properly formulated. It did not incorporate any equation to predict the MOLS, and it did not specify how the symbols of the MOLS are to be related to the values of the variables. It did not specify in detail the method used for analyzing the values calculated in the sub squares. It was not applied to pentapeptides.

[0008] 'Revised algorithms for build-up procedure for predicting protein conformations by energy minimization' by K. D. Gibson and H. A. Scheraga, Journal of Computational Chemistry, 1987 discloses 'Build-Up Procedure' to predict protein conformations. This method carries out a truncated search, relaying on the dominance of short-range interactions. Thus, it finds local minima of short fragments by an exhaustive energy-minimization procedure. The number of conformations of fragments that must be energy minimized and stored at each step increases exponentially. The method becomes unmanageable for polypeptide chains about 20 amino acid residues. Also there is no mention at all of the use of MOLS in any form in the procedure. 'On the multiple-minima problem in the conformational analysis of polypeptides' I. Backbone degrees of freedom for a perturbed a-helix' by L. Piela and H. A. Scheraga, Biopolymers, 1987 discloses the use of 'Self Consistent Electrostatic Field', based on the idea that peptide group dipoles in native conformation must have approximately optimal orientations in the electric field generated by the whole molecule and its surrounding solvent. The method does not use all the other energy terms. It is therefore specific for only the electrostatic field. It cannot be applied to small molecules such as peptides, etc. Also there is no mention at all of the use of MOLS in any form in the procedure.

[0009] 'Monte Carlo-Minimization approach to the multiple-minima problem in protein folding' by Z. Li and H. A. Scheraga, Proceedings of the National Academy of Sciences, 1987 discloses a stochastic approach for global optimization of polypeptides and proteins that combines the strength of the Metropolis Monte Carlo method in global combinatorial optimization with that of conventional energy minimization to find local minima. This is not a deterministic search procedure and therefore will not ensure the best structure every time. Also it is very computationally intensive and therefore very expensive in terms of computer time. Also there is no mention at all of the use of MOLS in any form in the procedure and hence this method is different from the present invention. 'New developments of the electrostatically driven Monte Carlo method - Test on the membrane bound portion of melittin' by D. R. Ripoll, A. Liwo and H. A. Scheraga, Biopolymers, 1998 discloses an iterative procedure for searching the conformational hypersurface of polypeptides. This is not a deterministic search procedure and therefore will not ensure the best structure every time. Also it is very computationally intensive and therefore very expensive in terms of computer time. Also there is no mention at all of the use of MOLS in any form in the procedure.

[0010] 'The multiple-minima problem in the conformational analysis of molecules-Deformation of the potential energy hypersurface by the diffusion equation method', L. Piela, J. Kostrowicki and H. A. Scheraga, Journal of Physical Chemistry, 1989 discloses the use of the basic idea of the 'Diffusion Equation Method' to deform the multivariable function that represents the potential energy in such a manner as to make the shallow wells disappear gradually, while other potential wells grow at their expense. The method is not easily applied in practice to complicated structures. This is not a deterministic search procedure and therefore will not ensure the best structure every time. Also it is very computationally intensive and therefore very expensive in terms of computer time. Also there is no mention at all of the use of MOLS in any form in the procedure.

[0011] 'Molecular dynamics on deformed potential energy hypersurfaces' by J. Pillardy and L. Piela, Journal of Physical Chemistry, 1995. The 'Distance Scaling Method' is a procedure to deform the potential energy hypersurface. The procedure is an extension of the 'diffusion equation method' and suffers from the same drawbacks. This is not a deterministic search procedure and therefore will not ensure the best structure every time. Also' it is very computationally intensive and therefore very expensive in terms of computer time. Also there is no mention at all of the use of MOLS in any form in the procedure.

[0012] 'Conformational analysis of the 20-residue membrane-bound portion of melittin by conformational space annealing' by J. Lee and H. A. Scheraga, Biopolymers, 1998. combines essential aspects of the build-up procedure and a genetic algorithm. It searches the whole conformational space in its early stages and then narrows the search to smaller regions with low energy. However, Searching the whole of conformational space requires large amounts of computer time. Therefore it is very computationally intensive and therefore very expensive in terms of computer time. Also there is no mention at all of the use of MOLS in any form in the procedure.

[0013] 'Protein structure can be predicted by global optimization of a potential energy function' by A. Liwo, J. Lee, D. R Ripoll, J. Pillardy and H. A. Scheraga, Proceedings of the National Academy of Sciences, 1999 uses a Hierarchical Approach based on two recent developments, a united-residue force field (UNRES) and the CSA method. An extensive conformational search is carried out with the CSA using the UNRES force field. This suffers from the same drawbacks as the 'Conformational space Annealing method'. In other words, it is very computationally intensive and therefore very expensive in terms of computer time. Also there is no mention at all of the use of MOLS in any form in the procedure.

[0014] 'Taboo Search: An approach to the Multiple Minima Problem', D. Cvijiovic and J. Klinowski, Science, 1995 describes a method, Taboo Search (TS), based on the Glover's taboo search for discrete functions, of solving the multiple minima problem for continuous functions. The method does not indicate any application to molecular structure. Also it is very computationally intensive and therefore very expensive in terms of computer time. Also there is no mention at all of the use of MOLS in any form in the procedure.

[0015] 'TRUST: A deterministic algorithm for Global Optimization' by J. Barthen, V. Protopopescu and D. Reister, Science, 1997 suggests an approach, Terminal Repeller Unconstrained Sub-energy Tunneling algorithm (TRUST) to solving continuous global optimization problems. The method however does not indicate any application to molecular structure and is also very computationally intensive. Therefore the method is expensive in terms of computer time. Also there is no mention of the use of MOLS in the procedure.

Objects of the invention

[0016] The main object of the present invention is to provide a method to build optimal 3-dimensional models of molecular structures from knowledge of their chemical structures, which obviates the above-mentioned drawbacks, as detailed below.

[0017] Another object of the present invention is to provide a method and system to build minimum-energy structures of biological macromolecules such as proteins and nucleic acids.

[0018] Another object of the present invention is to apply the method of the invention to find the best value of a function.

[0019] Another object of the present invention is to apply the method of the invention to find the best value of a potential energy function of a molecule.

[0020] Another object of the present invention is to make the method of structure determination simple and straight-forward and inexpensive, in terms of computer time, to use.

[0021] Another object of the present invention is to apply the method of the invention to a wide variety of problems in which a physical or chemical system is defined as a function of a set of parameters and the invention is used to find the best configuration of the system by finding the best set of values of the parameters which give the best value of the function.

[0022] Another object of the present invention is to provide a method which can be easily be combined with other known methods such as gradient search, simulated annealing, etc.

[0023] Another object of the present invention is to make it easily programmable to run on parallel computers, thus saving valuable time.

Summary of the invention

[0024] Accordingly the present invention provides a method for determining the lowest value of the conformational potential energy function of a molecule useful for building models of 3-dimensional molecular structures using a new algorithm defining the best conformation as the one corresponding to the lowest value for the conformational potential energy function, wherein the said method comprises the steps of:

> a) providing representation of a molecule in terms of the coordinates of the atoms comprising the molecule;
> b) selecting a set of conformations chosen from all possible conformations using a combinatorial arrangement n Mutually Orthogonal Latin Squares (MOLS) of order N, wherein said MOLS are constructed based upon conformational parameters in molecule, there being N^2 sub squares in each set of MOLS, where N is equal to n or m, whichever is larger, with n being number of parameters that define conformation of the molecule and m being the number of values sampled along each conformational parameter and the step size being R/m, with R being the range of the parameter, and each sub square corresponding to one possible conformation of the said molecule;
> c) filling up the sub squares of the MOLS with conformations of the said molecule, with one conformation in each sub square, the conformation in any given sub square being represented by the values of the conformational parameters in that sub square;
> d) choosing the values of the conformational parameters in the sub squares by first specifying the number of parameters, n, and the range of each of the n parameters, and the step size at which the values of each parameter is sampled, there being a total of m steps along each parameter, so that $\theta_{r,s}$, r = 1 to n and s = 1 to m are specified as the input set of values for the conformational parameters;
> e) labeling the sub squares of the said MOLS by indices (u,t), u and t being integer numbers, u = 1 to N and t = 1 to N, and the value of the rth torsion angle in a given sub square (u,t) being chosen from the input set of values $\theta_{r,s}$ as $\theta^r_{u,t} = \theta_{r,s}$, in which the index u is given by u = {(t-1) (r-1) + (s-1) } modulo (N);
> f) repeating the procedure of filling up the sub squares with the values of the parameters chosen for all values of r = 1 to n, s = 1 to m, and t = 1 to N, with u being calculated each time by the formula given in step (e), by which at the end of the procedure, each sub square of the N^2 sub squares of the said MOLS will be filled with a set of n

values, each value corresponding to one of the n conformational parameters, thus the set of n values in each sub square defining one conformation of the said molecule, there being N^2 conformations;

g) calculating the potential energy function by known methods for each of N^2 conformations of the said molecule as represented in the N^2 sub squares of the said set of MOLS, these functions being represented as

$$V_{u,t} = V(\theta_{u,t}^r)$$

h) analyzing the set of N^2 values of the potential energy function in order to determine the best value of each conformational parameter;

i) analysing by determining the average value of the potential energy function for each value of each parameter, by taking the average of the values of the potential energy function at all the N conformations where that value of the conformational parameter is used in defining the conformations, there being thus n x m average values, the average values being given by

$$\langle V \rangle_s^r = \frac{\sum\limits_t V(\theta_{u,t}^r) \exp(-V(\theta_{u,t}^r)/kT)}{\sum\limits_t \exp(-V(\theta_{u,t}^r)/kT)}$$

j) scanning the set of n x m average values of the potential energy function to obtain, for each conformational parameter, the value of that parameter that corresponds to the lowest average value of the potential energy function, which is followed by using set of n values of the n conformational parameters, one value to each parameter, to define best conformation of said molecule having lowest value of potential energy function.

k) confirming the results or locating another optimal structure, if required, by choosing a different way of arranging the values of the conformational parameters and repeating steps e to k.

[0025]   In one embodiment of the invention, the method comprises obtaining optima of several simple non-linear functions each with a single known optimum in parameter range.

[0026]   In another embodiment of the invention, the method is carried out by finding the best value of different types of functions of many parameters.

[0027]   In a further embodiment of the invention, the method is carried out by obtaining the best structure of small molecules such as peptides, nucleotides, etc.

[0028]   In a further embodiment of the invention, the method is carried out by obtaining the best structure of biological macromolecules such as proteins and nucleic acids, etc.

[0029]   In another embodiment of the invention, the method runs in parallel on several computers simultaneously to further reduce the computation time.

Brief description of the accompanying drawings

[0030]   Figure 1 is an example of a set of Mutually Orthogonal Latin Squares (MOLS) in order to better explain the method. The diagram shows an example of three MOLS of order 7, i.e. n = 3, m = 7, N = 7. Symbols in the first Latin Square: a1, a2, a3, a4, a5, a6, a7. Each of these is repeated 7 times to give a total of 49 symbols, which have been arranged in an LS. Symbols in second LS: b1, b2, b3, b4, b5, b6, b7. The second LS is orthogonal to the first. Note that every pairing of a symbol from the first Square with one from the second occurs exactly once. Symbols in third LS: c1, c2, c3, c4, c5, c6, c7. This is orthogonal to both the other Squares. For clarity in this figure 3 different sets ofN symbols have been used. One can use the same set of N symbols and construct N-1 MOLS of order N. One of sub-squares of the set of MOLS has been highlighted, its symbols are a7 of the first LS, b1 of the second LS and c5 of the third LS. In the present application, each such sub-square represents a possible conformation of the molecule. The MOLS method requires the potential function to be evaluated at all these N^2 points in the conformation space.

[0031]   Figure 2a gives the structure of the molecule pentapeptide [Met[5]] enkephalin (Tyr-Gly-Gly-Phe-Met) as obtained by an application of the present MOLS method as described in example 1. The value of the potential energy function for this structure is -13.61 kcal/mole.

[0032]   Figure 2b shows for comparison the structure of the molecule as reported by Z. Li and H. A. Scheraga in the Proceedings of the National Academy of Sciences (USA), Volume 84, pages 6611-6615. It is clear from the comparison that the two structures are the same and that the present MOLS method successfully identifies the correct structure of the molecule. The present MOLS method identified this structure and 99 other structures in a total computation time of

65 seconds on a Pentium III processor running at 650 MHz. This time is far less than the time taken by the procedure of Li and Scheraga.

[0033]   Figure 3a shows another structure of the molecule pentapeptide [Met[5]] enkephalin (Tyr-Gly-Gly-Phe-Met) as obtained by application of the present MOLS method as described in example 1. Value of potential energy function for this structure is -8.53 kcal/mole.

[0034]   Figure 3b shows for comparison the experimentally determined structure of the molecule as reported by Griffin, et al., in Proceedings of the National Academy of Sciences (CSA), volume 83, pages 3272-3276. It is clear from the comparison that the two structures are the same and that the present MOLS method successfully identifies the correct structure of the molecule. The present MOLS method identified this structure and 99 other structures, including the one shown in drawing 2 above, in a total computation time of 65 seconds on a Pentium III processor running at 650 MHz.

[0035]   Figure 4 shows Ramachandran plot for one hundred different structures, including the structures shown in Figure 2 and in Figure 3 above, obtained by repeating the method with a different set of input values each time. The drawing clearly shows that all structures are conformationally best structures as tested by the Ramachandran map, since their conformations all correspond to the allowed regions of the Ramachandran map.

[0036]   Figure 5a and 5b show two minimum energy structures of the pentapeptide $(Aib)_5$ as obtained by the present MOLS method (in green). The values of the potential energy function for these structures are 5.19 kcal/mole and 9.41 kcal/mole respectively. For comparison these two structures are superimposed on the structures (in red) of the same molecule as obtained experimentally by Benedetti et al. and reported in Journal of the Chemical Society, volume 104, pages 2437-2444, and by Shamala et al. and reported in Journal of Chemical Society, Chemical Communications, year 1978 page 996. The comparison shows the present MOLS method yields correct structures. These two structures along with 998 others were identified in total time of less than 1 hour using Pentium III processor at 650 MHz.

[0037]   Figure 6 shows the Ramachandran plot for one thousand different structures, including the structures shown in Figure 5 above, obtained by the present MOLS method, by repeating the method with a different set of input values each time. The drawing clearly shows that all structures are conformationally best structures as tested by the Ramachandran map, since their conformations all correspond to the allowed regions of the Ramachandran map. These 1000 structures were obtained in a total time of less than 60 minutes on Pentium III processor running at 650 MHz.

[0038]   Figure 7 gives a detailed flowchart to explain the procedure.

Detailed description of the invention

[0039]   The method of the invention is best explained in terms of a series of steps. The construction of the set of MOLS comprises a Latin Square (LS) of order N defined as a set of N symbols, arranged in a N x N square, such that each symbol occurs exactly once in every row and once in every column. Two LS are orthogonal if, when they are superimposed, each symbol of the first square occurs once, and only once, with each symbol of the second square. A set of MOLS is a set of LS, every pair of which is orthogonal (Figure.1). It has been shown in the prior art that if N is a prime power, one can construct N-1 MOLS of order N, as follows. The N symbols (belonging to a Galois field) are written as b_i, i = 1 to N. If is the $(i, j)^{th}$ element of the LS labeled by e, then

$$a_{ij}^e = b_e x\, b_i + b_j \qquad\qquad (1)$$

for i = 1 to N; j = 1 to N; e = 1, N-1. In the present algorithm a correspondence is made between the symbols b and the conformational parameters of the molecule. For example, these may be the torsion angles, θ_i, i = 1 to n, which characterize the conformation of an oligopeptide. The best conformation of the molecule is defined to be the one, which has the minimum value of the potential energy function, which is a function of the conformational parameters. In other words if the set of conformational parameters that will yield a minimum value to the potential energy function is found, then that set of conformational parameters will correspond to the best conformation of the molecule. Each conformational parameter is specified as being capable of taking up m different values in a particular range. If r is the index for these conformational parameters and s is the index for the values taken up by each parameter, then $\theta_{r,s}$, r = 1 to n; s = 1 to m are the input set of values, and the search for the best conformation has to be carried out among the n^m combinations of these values. Thus, for example, $\theta_{1,1}$ could refer to the first value of the first parameter, $\theta_{1,2}$ to the second value of the first parameter, $\theta_{2,1}$ to the first value of the second parameter, $\theta_{2,2}$ to the second value of the second parameter, and so on up to $\theta_{n,m}$, which would refer to the last value of the last parameter. Obviously there would be a set of m x n input values. To construct the set of MOLS, each conformational parameter is set to correspond to one LS, and apply equation (1) to the indices of the conformational parameters. The indices r and s are considered to form n Galois sets, each with m elements, i.e. the index pair (r,s) will form the sth element in the rth set. The sub-squares of the MOLS are labeled by the pair of indices (u,t), with both u and t taking up integer values from 1 to N, the value of the rth torsion angle in the

sub-square (u,t) may be chosen from the input set of values $\theta_{r,s}$ as

$$\theta^r_{u,t} = \theta_{r,s} \qquad\qquad (2)$$

where the index u is given by

$$u = \{(t-1)(r-1) + (s-1)\}\,\text{modulo}(\,N)$$

The value of N is chosen such that (i) N-1 $\geq$ maximum (m,n) and (ii) N is a prime power.
This last condition requires that in practice the step size in the sampling grid is chosen to be R/N, where R is the specified range for the conformational parameters. The application of this procedure for all values of r = 1 to n; s = 1 to N and t = 1 to N will result in a set of n MOLS of order N. These MOLS, when superposed, will give N^2 sub-squares, each sub-square containing a set of values for the n conformational parameters (Figure 1). Each set of values of the parameters, labeled by the index pair (u,t), represents one possible conformation of the polypeptide chain and, therefore, a point in the n dimensional conformational space. The method of the present invention shows that sampling the potential surface of the molecule at the N^2 points specified by the sub-squares, out of the possible n^m, is sufficient to build a map of the entire space, and that this map can then be used to perform a rapid search for the optimum. The hypothesis is supported by the fact that the set of MOLS implicitly contains every possible two-dimensional projection of the n-dimensional space. The next step is the sampling of the energy hyper-surface in conformational space, which is carried out by calculating the potential energy

$$V_{u,t} = V(\theta^r_{u,t})$$

at each of the N^2 points u = 1 to N; t = 1 to N. After this step is completed and the N^2 values of the potential energy function have been calculated, the next task is to recover the energy map of the conformational space. To accomplish this n one-dimensional representations of the variation of the potential energy function V along each of the conformational parameters is constructed. The effect of setting a particular conformational parameter to a particular value is estimated by taking the average of V over the N points where that value of that parameter occurs in the MOLS, i.e.

$$\langle V \rangle^r_s = \frac{\sum_t V(\theta^r_{u,t})\exp(-V(\theta^r_{u,t})/kT)}{\sum_t \exp(-V(\theta^r_{u,t})/kT)}$$

where k is Boltzmann's constant and T is the temperature. Once again, equation (2) is used to calculate the value of u for each value of r, s and t. Since the number of steps m along each parameter are set as equal to N, there will be N such average values for each of the r = 1 to n. conformational parameters. The N average values for a given parameter will form a representation of the behavior of the potential V as a function of that parameter. The final step is an inspection of each one-dimensional representation, which will therefore reveal the optimum value for the respective torsion. Thus, if

$$V^r_{s=\omega_r} = \underset{\text{over s}}{\text{optimum}}\ (\langle V \rangle^r_s) \qquad\qquad (3)$$

then θ_{r,ω_r} is the optimum value for the parameter r. The set of optima $\theta_{r,\omega}$, r = 1 to n will then define the optimum conformation of the molecule. This completes the procedure.

[0040]   The indices r and s used in the construction of the MOLS may be assigned to the m values of the n parameters in $(m!)^n$ different ways. According to the invention no matter which way the assignment is made, the procedure will result in an optimal structure. Therefore, by choosing different assignments of the angles, the calculations may be repeated to check for the consistency of the results obtained.

Novelty vis-à-vis prior art:

[0041]

5   1. Mutually Orthogonal Latin Squares (MOLS) have not been previously used to find the best molecular or biomolecular structure.
2. MOLS have not been previously used to find the best value of a function.
3. MOLS have not been previously used to find the global best value of a function.
4. Equation 2 has not been previously stated or used.
10   5. The method of taking averages as given in equation (3) has not been previously used.

[0042]   The following examples are given by way of illustration of the present invention and should not be construed to limit the scope of the present invention.

15   **Example 1**

[0043]   As an example the method of the invention is applied to pentapeptide [Met[5]] enkephalin (Tyr-Gly-Gly-Phe-Met). Ignoring the side chain torsion angles, which were kept fixed at the values found in the literature, there are 8 backbone torsion angles. These are considered the conformational parameters, with range between 0 and 360°. To
20   sample each of them at intervals of 10° would require 36 points along each dimension. The total search space is thus 36^8 or about 2.8 x 10^{12} points and an exhaustive search for the optimal structure would entail the calculation of the energy at each of these points to choose the best one. In the present method the energy is calculated at only 36^2 points. However since the order of MOLS is always a prime number and since 36 is not a prime number, in reality 37^2 points are searched. As the first step the set of values for each torsion angle that is to be considered is specified. Thus the
25   values for the first torsion angle may be 0°, 10°, 20°, 30°, .360°. Similarly the second torsion angle has to be considered at the values 0°, 10°, .360°. Likewise the sampling points of the other torsion angles are also specified. Thus, the values $\theta_{r,s}$ for r = 1,8; s = 1, 37 are specified. Equation (2) is now used to distribute 37 copies of these 37 values of each of the 8 torsion angles among the 37^2 sub squares so as form a set of 8 Mutually Orthogonal Latin Squares of order 8. The 8 values in each sub square will then form one possible conformation of the molecule. Note that some of these conformations
30   may be 'impossible' conformations, in the sense that the atoms may interpenetrate. However, the present invention allows one to derive the optimal conformation from this set of 37^2 conformations. The next step in the procedure is to calculate the semi-empirical potential energy at each of these 37^2 points. Once these energies have been calculated, the next step comprises ascertaining the average energy for each value of each torsion angle. For example, for the first value of the first torsion angle this is 0°. There are 37 conformations of the molecule in which the value of the first torsion
35   angle has been set to 0° and the values of the potential energy function of all these conformation have been determined in the procedure above. The Boltzmann weighted average of these 37 energy values is taken, and thus a single number obtained which specifies the effect of setting this angle to this value, i.e. the effect of setting the first torsion to the value 0°. Similarly the average energy when the first torsion is set to 10°, 20°, etc is obtained. It is important to note that it is the property of MOLS that each setting is repeated 37 times (or N times, where N is the order of the MOLS). Since there
40   are 37 settings for the torsion, there will be 37 average energies. The next step comprises inspecting these 37 averages and identifying one setting of this torsion that has yielded the lowest average energy. This will be the optimum value for this torsion angle. The procedure is repeated for the next torsion angle and the optimum value for this also determined in an exactly similar manner as for the first one. Note that again, due to the property of the MOSL, each value of the second torsion also will be repeated 37 times. The whole procedure is again repeated for the other torsion angles. At
45   the end of this step the invention has individually determined the best setting for each of the 8 torsion angles. When the molecule is now constructed with these 8 torsion angle values, identified by the invention, this will be the optimal conformation of [Met[5]] enkephalin. The entire procedure is repeated by choosing $\theta_{r,s}$ in a different way. Now instead of specifying $\theta_{1,1} = 0.0$, $\theta_{1,2} = 10.0$ etc, the specified values for $\theta_{1,1} = 10.0$, $\theta_{1,2} = 0.0$ etc. The assignments are made using a random number generator. The steps are repeated to yield another optimal conformation for the molecule. This may
50   be repeated again and again to yield as many optimal structures as required. In this example, one hundred optimal structures were generated. Each structure was the result of the application of the above procedure to a different input set of torsion angles $\theta_{r,s}$, which was obtained by a different assignment of the indices to the torsion angle values. The assignments were chosen using a random number generator. The energies corresponding to all these structures are low, the highest being 72.3 kcal/mole and the lowest -14.0 kcal/mole, with an average of -7.34 kcal/mole for all 100
55   structures. The best structure is shown in Figure 2 and Figure 3 and the conformational Ramachandran Plot of all 100 structures shown in Figure 4.

Example 2

[0044] As another example the invention, to shown its application for arbitrary functions and not just for potential energy functions based on conformational parameters, the method is applied to an artificial function with a known minimum as follows:

$$F = 2^{(a-5)} \times 3^{|b-2|} \times 4^{|c-9|} \, 5^{|d-6|} \times 6^{|e-8|} \times 6^{(f-8)} \times 5^{(g-6)} \times 4^{(h-9)} \times 3^{(i-2)} \times 2^{|j-5|} \times (k-7.1)^4$$

[0045] This is a function of the 11 variables a, b, c, d, e, f, g, h, i, j, k all of them sampled in the range 1.0 to 13.0, in steps of 1.0. The minimum of the function in this space is at a = 1.0, b = 2.0, c = 9.0, d = 6.0, e = 8.0, f= 1.0, g = 1.0, h = 1.0, i = 1.0, j = 5.0, k = 7.0. A set of 11 MOLS of order 13 is constructed using the values 1.0, 2.0, ... 13.0 as range of values for each of the 11 variables. The function is evaluated at each of the points specified by the 13 x 13 sub squares. Averages are taken and the minimum is identified as 1.0, 1.0, 8.0, 5.0, 8.0, 2.0, 1.0, 1.0, 1.0, 5.0, 7.0. This result is very close to the theoretical minimum given above.

Example 3

[0046] In this example, the molecule chosed comprises the pentapeptide (Aib)$_5$. An in example 1, the number of free torsion angles in this molecule is 8. Again these are chosen as the conformational parameters, with the range set to 0 to 360°. At steps of 10° there are 37 points along each dimension and order of the MOLS is chosen as 37. The 37^2 conformational points are specified using MOLS, the potential energy functions are calculated, averages are taken and optimal conformation is identified as described in example 1. This is repeated 1000 times in a total computer time of about 1hour for all 1000 conformations. All of them are good conformations as seen by the low energy values. Some of them correspond to well known structures such as alpha helices or turns. Figure 5 shows the best two structures and Figure 6 shows the Ramachandran conformational plot of all 1000 structures.
Summary of observations: The invention works very well for different types of functions. Particularly in molecular conformation analyses, the invention allows one to map the entire conformational space in a reasonable amount of time, to identify physically reasonable structures.

The main advantages of present invention are

[0047]

1. The main advantage of the present invention is that it provides a method to build optimal 3-dimensional models of molecular structures from knowledge of their chemical structures, which obviates the drawbacks of the previous methods.
2. Another advantage of the present invention is that it provides a method and system to build minimum-energy structures of biological macromolecules such as proteins and nucleic acids.
3. Another advantage of the present invention is that it may be applied to find the best value of a function.
4. Another advantage of the present invention is that it may be applied to find the best value of a potential energy function of a molecule.
5. Another advantage of the present invention is that it is simple and straightforward and inexpensive, in terms of computer time, to use.
6. Another advantage of the present invention is that it may be applied to a wide variety of problems in which a physical or chemical system is defined as a function of a set of parameters and the invention is used to find the best configuration of the system by finding the best set of values of the parameters which give the best value of the function.
7. Another advantage of the present invention is that it may be combined with other known methods such as gradient search, simulated annealing, etc.
8. Another advantage of the present invention is that it may be easily programmed to run on parallel computers, thus saving valuable time.

Claims

1. A computer-implemented method for determining the lowest value of the conformational potential energy function of a molecule useful for building models of 3-dimensional molecular structures using a new algorithm defining the

best conformation as the one corresponding to the lowest value for the conformational potential energy function, wherein the said method comprises the steps of:

(a) providing representation of a molecule in terms of the coordinates of the atoms comprising the molecule, by known methods;

(b) selecting a set of conformations chosen from all possible conformations using a combinatorial arrangement of n Mutually Orthogonal Latin Squares (MOLS) of the order N, wherein said MOLS are constructed based upon the conformational parameters in the molecule, there being N^2 sub squares in each set of MOLS, where N is equal to n or m, whichever is larger, with n being the number of parameters that define the conformation of the molecule and m being the number of values sampled along each conformational parameter and the step size being R/m, with R being the range of the parameter, and each sub square corresponding to one possible conformation of the said molecule;

(c) filling up the sub squares of the MOLS with conformations of the said molecule, with one conformation in each sub square, the conformation in any given sub square being represented by the values of the conformational parameters in that sub square;

(d) choosing the values of the conformational parameters in the sub squares by first specifying the number of parameters, n, and the range of each of the n parameters, and the step size at which the values of each parameter are sampled, there being a total of m steps along each parameter, so that we specify that $\theta_{r,s}$, r = 1 to n and s = 1 to m are the input set of values for the conformational parameters;

(e) labeling the sub squares of the said MOLS by indices (u, t), u and t being integer numbers, u = 1 to N and t = 1 to N, and the value of the rth torsion angle in a given sub square (u, t) being chosen from the input set of values $\theta_{r,s}$, as $\theta^r_{u,t}, = \theta_{r,s}$, in which the index u is given by u = {(t-1) (r-1) + (s-1)} modulo (N);

(f) repeating the procedure of filling up the sub squares with the values of the parameters chosen for all values of r = 1 to n, s = 1 to m, and t = 1 to N, with u being calculated each time by the formula given in step (e), by which at the end of the procedure, each sub square of the N^2 sub squares of the said MOLS will be filled with a set of n values, each value corresponding to one of the n conformational parameters, thus the set of n values in each sub square defining one conformation of the said molecule, there being N^2 conformations;

(g) calculating the potential energy function by known methods for each of N^2 conformations of the said molecule as represented in the N^2 sub squares of the said set of MOLS, these functions being represented as

$$V_{u,t} = V(\theta^r_{u,t})$$

(h) analyzing the set of N^2 values of the potential energy function in order to determine the best value of each conformational parameter;

(i) analyzing by determining the average value of the potential energy function for each value of each parameter, by taking the average of the values of the potential energy function at all the N conformations where that value of the conformational parameter is used in defining the conformations, there being thus n x m average values, the average values being given by

$$\langle V \rangle^r_s = \frac{\sum_t V(\theta^r_{u,t}) \exp(-V(\theta^r_{u,t})/kT)}{\sum_t \exp(-V(\theta^r_{u,t})/kT)}$$

(j) scanning the set of n x m average values of the potential energy function to obtain, for each conformational parameter, value of that parameter that corresponds to the lowest average value of the potential energy function, which is followed by using the set of n values of the n conformational parameters, one value to each parameter, to define the best conformation of the said molecule having lowest value of the potential energy function; and

(k) confirming the results or locating another optimal structure, if required, by choosing a different way of arranging the values of the conformational parameters and repeating steps (e) to (k).

2. A computer-implemented method as claimed in claim 1, wherein the method comprises obtaining the optima of several simple non-linear functions each with a single known optimum in the parameter range.

3. A computer-implemented method as claimed in claim 1, wherein the method is carried out by finding the best value of different types of functions of many parameters.

4. A computer-implemented method as claimed in claim 1, wherein the molecule comprises a small molecule selected from peptides and nucleotides.

5. A computer-implemented method as claimed in claim 1 wherein the molecule comprises a biological macromolecule selected from proteins and nucleic acids.

6. A computer-implemented method as claimed in claim 1 wherein the method runs in parallel on several computers simultaneously to further reduce the computation time.

Patentansprüche

1. Rechnerimplementiertes Verfahren zum Bestimmen des niedrigsten Wertes der Konformationspotentialenergie-funktion eines Moleküls, verwendbar zum Erstellen von Modellen der 3-dimensionalen Molekularstruktur, unter Verwendung eines neuen Algorithmus, der die beste Konformation als diejenige definiert, die dem niedrigsten Wert für die Konformationspotentialenergiefunktion entspricht, wobei das Verfahren die folgenden Schritte umfasst:

(a) Bereitstellen einer Darstellung eines Moleküls in Form der Koordinaten der Atome, die das Molekül ausma-chen, durch bekannte Verfahren,
(b) Wählen eines Satzes von Konformationen, ausgewählt aus allen möglichen Konformationen unter Verwen-dung einer kombinatorischen Anordnung von n zueinander rechtwinkligen lateinischen Quadraten (Mutually Orthogonal Latin Squares - MOLS) der Ordnung N, wobei die MOLS auf der Grundlage der Konformationspa-rameter in dem Molekül aufgebaut sind, wobei es in jedem Satz von MOLS N^2 Unterquadrate gibt, wo N gleich n oder m ist, je nachdem, welches größer ist, wobei n die Zahl der Parameter ist, welche die Konformation des Moleküls definieren, und m die Zahl von längs jedes Konformationsparameters abgetasteten Werten ist und die Schrittgröße R/m ist, wobei R der Bereich des Parameters ist, und jedes Unterquadrat einer möglichen Konformation des Moleküls entspricht,
(c) Auffüllen der Unterquadrate der MOLS mit Konformationen des Moleküls, mit einer Konformation in jedem Unterquadrat, wobei die Konformation in einem beliebigen gegebenen Unterquadrat durch die Werte der Kon-formationsparameter in diesem Unterquadrat dargestellt wird,
(d) Auswählen der Werte der Konformationsparameter in den Unterquadraten **dadurch**, dass zuerst die Zahl von Parametern, n, und der Bereich jedes der n Parameter und die Schrittgröße, mit der die Werte jedes Parameters abgetastet werden, spezifiziert werden, wobei es eine Gesamtzahl von m Schritten längs jedes Parameters gibt, so dass wir spezifizieren, dass $\theta_{r,s}$, r = 1 bis n und s = 1 bis m der eingegebene Satz von Werten für die Konformationsparameter sind,
(e) Bezeichnen der Unterquadrate der MOLS durch Indices (u, t), wobei u und t ganze Zahlen sind, u = 1 bis N und t = 1 bis N, und der Wert des r-ten Drehwinkels in einem gegebenen Unterquadrat (u, t) aus dem eingegebenen Satz von Werten $\theta_{r,s}$, als $\theta^r_{u,t}$, = $\theta_{r,s}$, ausgewählt wird, worin der Index u durch u = {(t-1)(r-1) + (s-1)} Modulo (N) gegeben ist,
(f) Wiederholen des Vorgangs des Auffüllens der Unterquadrate mit den Werten der ausgewählten Parameter für alle Werte von r = 1 bis n, s = 1 bis m und t = 1 bis N, wobei u jedes Mal durch die in Schritt (e) angegebene Formel berechnet wird, wodurch am Ende des Vorgangs jedes Unterquadrat der N^2 Unterquadrate der MOLS mit einem Satz von n Werten gefüllt sein wird, wobei jeder Wert einem der n Konformationsparameter entspricht, wobei folglich der Satz von n Werten in jedem Unterquadrat eine Konformation des Moleküls definiert, wobei es N^2 Konformationen gibt,
(g) Berechnen der Potentialenergiefunktion durch bekannte Verfahren für jede von N^2 Konformationen des Moleküls, wie sie in den N^2 Unterquadraten des Satzes von MOLS dargestellt sind, wobei diese Funktionen dargestellt werden als

$$V_{u,t} = V(\theta^r_{u,t}),$$

(h) Analysieren des Satzes von N^2 Werten des Potentialenergiefunktion, um den besten Wert jedes Konforma-tionsparameters zu bestimmen,

(i) Analysieren durch Bestimmen des Mittelwerts der Potentialenergiefunktion für jeden Wert jedes Parameters, **dadurch**, dass das Mittel der Werte der Potentialenergiefunktion an allen N Konformationen genommen wird, wobei dieser Wert des Konformationsparameters beim Definieren der Konformationen verwendet wird, wobei es folglich n x m Mittelwerte gibt, wobei die Mittelwerte gegeben sind durch

$$(V)_s^r = \frac{\sum_t V(\theta_{u,t}^r) \exp(-V(\theta_{u,t}^r)/kT)}{\sum_t \exp(-V(\theta_{u,t}^r)/kT)},$$

(j) Abtasten des Satzes von n x m Mittelwerten der Potentialenergiefunktion, um, für jeden Konformationsparameter, einen Wert dieses Parameters zu erhalten, der dem niedrigsten Mittelwert der Potentialenergiefunktion entspricht, was gefolgt wird durch Verwenden des Satzes von n Werten der n Konformationsparameter, eines Werts für jeden Parameter, um die beste Konformation des Moleküls mit dem niedrigsten Wert der Potential-energiefunktion zu definieren, und

(k) Bestätigen der Ergebnisse oder Lokalisieren einer anderen optimalen Struktur, falls erforderlich, durch Aus-wählen einer anderen Weise der Anordnung der Werte der Konformationsparameter und Wiederholen der Schritte (e) bis (k).

2. Rechnerimplementiertes Verfahren nach Anspruch 1, wobei das Verfahren umfasst, die Optima mehrerer einfacher nichtlinearer Funktionen jeweils mit einem einzigen bekannten Optimum in dem Parameterbereich zu erhalten.

3. Rechnerimplementiertes Verfahren nach Anspruch 1, wobei das Verfahren durch Finden des besten Werts von unterschiedlichen Typen von Funktionen von vielen Parametern ausgeführt wird.

4. Rechnerimplementiertes Verfahren nach Anspruch 1, wobei das Molekül ein kleines, aus Peptiden und Nukleotiden ausgewähltes, Molekül umfasst.

5. Rechnerimplementiertes Verfahren nach Anspruch 1, wobei das Molekül ein aus Proteinen und Nukleinsäuren ausgewähltes biologisches Makromolekül umfasst.

6. Rechnerimplementiertes Verfahren nach Anspruch 1, wobei das Verfahren parallel gleichzeitig auf mehreren Rech-nern läuft, um die Berechnungszeit weiter zu verringern.

Revendications

1. Procédé mis en oeuvre par ordinateur pour déterminer la valeur la plus faible de la fonction d'énergie potentielle conformationnelle d'une molécule utile pour construire des modèles de structures moléculaires tridimensionnelles en utilisant un nouvel algorithme définissant la meilleure conformation comme étant celle correspondant à la valeur la plus faible pour la fonction d'énergie potentielle conformationnelle, dans lequel ledit procédé comprend les étapes de :

(a) fourniture d'une représentation d'une molécule en termes des coordonnées des atomes constituant la mo-lécule, au moyen de procédés connus ;

(b) sélection d'un jeu de conformations choisies parmi toutes les conformations possibles en utilisant un arran-gement combinatoire de n Carrés Latins Mutuellement Orthogonaux (MOLS) de l'ordre N, dans lequel lesdits MOLS sont construits sur la base des paramètres conformationnels de la molécule, étant donné qu'il y a N^2 sous-carrés dans chaque jeu de MOLS, où N est égal à n ou m en fonction de celui qui est le plus grand, n étant le nombre de paramètres qui définissent la conformation de la molécule et m étant le nombre de valeurs échantillonnées sur l'amplitude de chaque paramètre conformationnel et la dimension de marche étant de R/m, R étant la plage du paramètre, et chaque sous-carré correspondant à une conformation possible de ladite molécule ;

(c) remplissage des sous-carrés des MOLS avec des conformations de ladite molécule, une seule conformation étant prévue dans chaque sous-carré, la conformation dans un quelconque sous-carré donné étant représentée par les valeurs des paramètres conformationnels dans ce sous-carré ;

(d) choix des valeurs des paramètres conformationnels dans les sous-carrés en tout d'abord spécifiant le nombre n de paramètres et la plage de chacun des n paramètres, et la dimension de marche selon laquelle les valeurs de chaque paramètre sont échantillonnées, étant entendu qu'il y a un total de m marches pour l'étendue de chaque paramètre, de telle sorte que nous spécifions que $\theta_{r,s}$, r = 1 à n et s = 1 à m sont le jeu d'entrées des valeurs pour les paramètres conformationnels;

(e) étiquetage des sous-carrés desdits MOLS avec des indices (u, t), u et t étant des nombres entiers, u = 1 à N et t = 1 à N, et les valeurs du r-ième angle de torsion dans un sous-carré donné (u, t) étant choisies à partir du jeu d'entrées des valeurs $\theta_{r,s}$, en tant que $\theta^r_{u;t} = \theta_{r,s}$, dans lequel l'indice u est donné par u = {(t-1)(r-1)+(s-1)}modulo(N);

(f) répétition de la procédure de remplissage des sous-carrés avec les valeurs des paramètres choisies pour toutes les valeurs de r = 1 à n, s = 1 à m, et t = 1 à N, u étant calculé chaque fois par la formule donnée dans l'étape (e), et ainsi, à la fin de la procédure, chaque sous-carré des N^2 sous-carrés desdits MOLS sera rempli avec un jeu de n valeurs, chaque valeur correspondant à un des n paramètres conformationnels, et ainsi le jeu de n valeurs dans chaque sous-carré définissant une conformation de ladite molécule, étant donné qu'il y a N^2 conformations;

(g) calcul de la fonction d'énergie potentielle par des procédés connus pour chacune des N^2 conformations de ladite molécule telles que représentées dans les N^2 sous-carrés dudit jeu de MOLS, ces fonctions étant représentées en tant que:

$$V_{u,t} = V(\theta^r_{u,t})$$

(h) analyse du jeu de N^2 valeurs de la fonction d'énergie potentielle afin de déterminer la meilleure valeur de chaque paramètre conformationnel;

(i) analyse en déterminant la valeur moyenne de la fonction d'énergie potentielle pour chaque valeur de chaque paramètre, en prenant la moyenne des valeurs de la fonction d'énergie potentielle pour toutes les N conformations pour lesquelles la valeur des paramètres conformationnels est utilisée en définissant les conformations, étant entendu qu'il y a ainsi n x m valeurs moyennes, les valeurs moyennes étant données par:

$$\langle V \rangle^r_s = \frac{\sum_t V(\theta^r_{u,t}) \exp(-V(\theta^r_{u,t})/kT)}{\sum_t \exp(-V(\theta^r_{u,t})/kT)}$$

(j) balayage du jeu de n x m valeurs moyennes de la fonction d'énergie potentielle afin d'obtenir, pour chaque paramètre conformationnel, la valeur du paramètre qui correspond à la valeur moyenne la plus faible de la fonction d'énergie potentielle, ceci étant suivi par l'utilisation du jeu de n valeurs des n paramètres conformationnels, à raison d'une valeur pour chaque paramètre, afin de définir la meilleure conformation de ladite molécule présentant la valeur la plus faible de la fonction d'énergie potentielle; et

(k) confirmation des résultats ou localisation d'une autre structure optimale, si requis, en choisissant une façon différente d'arranger les valeurs des paramètres conformationnels et en répétant les étapes (e) à (k).

2. Procédé mis en oeuvre par ordinateur selon la revendication 1, dans lequel le procédé comprend l'obtention de l'optimum de plusieurs fonctions non linéaires simples dont chacune a un optimum connu unique dans la plage de paramètres.

3. Procédé mis en oeuvre par ordinateur selon la revendication 1, dans lequel le procédé est mis en oeuvre en trouvant la meilleure valeur pour différents types de fonctions de plusieurs paramètres.

4. Procédé mis en oeuvre par ordinateur selon la revendication 1, dans lequel la molécule est constituée par une petite molécule choisie parmi des peptides et des nucléotides.

5. Procédé mis en oeuvre par ordinateur selon la revendication 1, dans lequel la molécule est constituée par une macromolécule biologique choisie parmi des protéines et des acides nucléiques.

6. Procédé mis en oeuvre par ordinateur selon la revendication 1, dans lequel le procédé se déroule en parallèle sur plusieurs ordinateurs simultanément de façon à réduire davantage le temps de calcul.

Figure 1

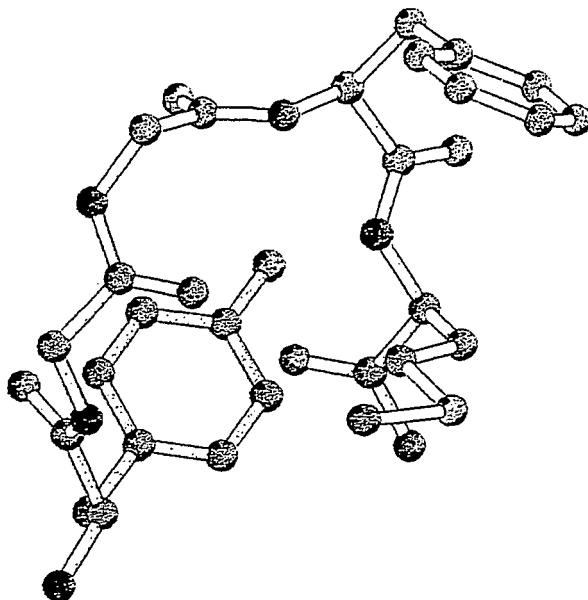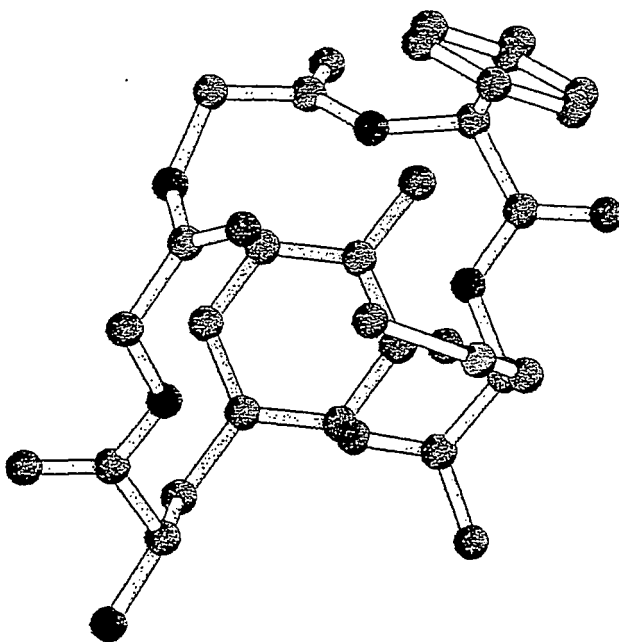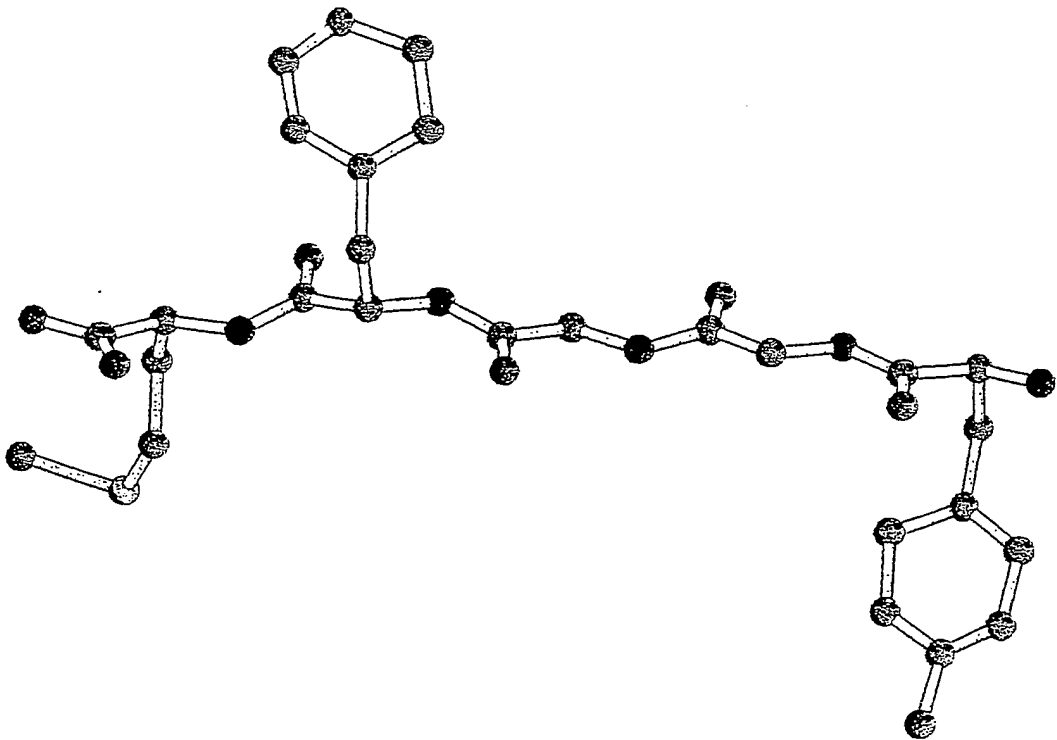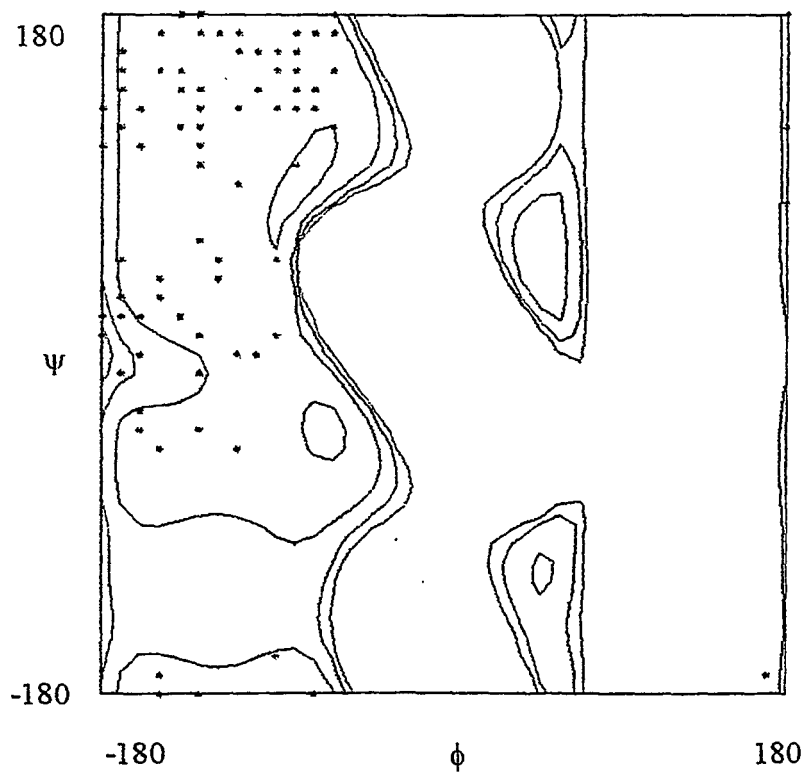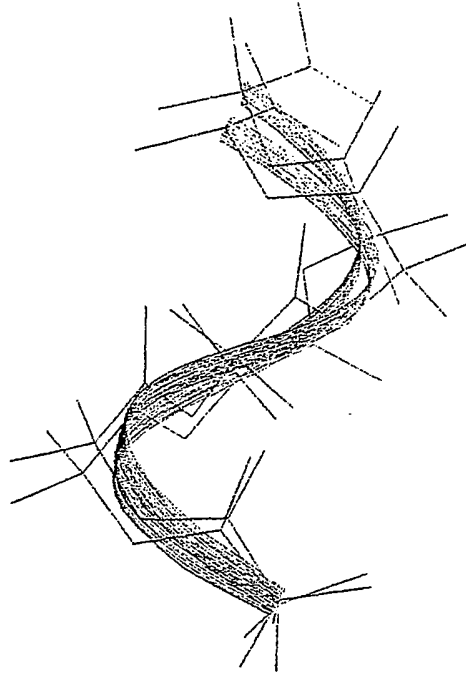|						
a1 b1 c1	a2 b2 c2	a3 b3 c3	a4 b4 c4	a5 b5 c5	a6 b6 c6	a7 b7 c7
a2 b7 c6	a3 b1 c7	a4 b2 c1	a5 b3 c2	a6 b4 c3	a7 b5 c4	a1 b6 c5
a3 b6 c4	a4 b7 c5	a5 b1 c6	a6 b2 c7	a7 b3 c1	a1 b4 c2	a2 b5 c3
a4 b5 c2	a5 b6 c3	a6 b7 c4	a7 b1 c5	a1 b2 c6	a2 b3 c7	a3 b4 c1
a5 b4 c7	a6 b5 c1	a7 b6 c2	a1 b7 c3	a2 b1 c4	a3 b2 c5	a4 b3 c6
a6 b3 c5	a7 b4 c6	a1 b5 c7	a2 b6 c1	a3 b7 c2	a4 b1 c3	a5 b2 c4
a7 b2 c3	a1 b3 c4	a2 b4 c5	a3 b5 c6	a4 b6 c7	a5 b7 c1	a6 b1 c2

Figure 2a

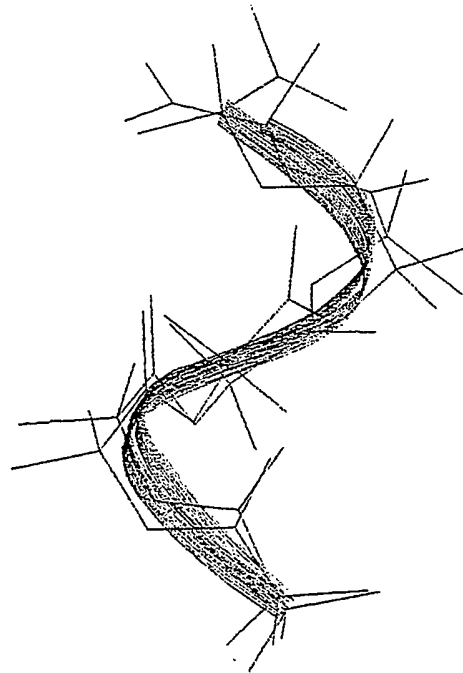Figure 2b

Figure 3a
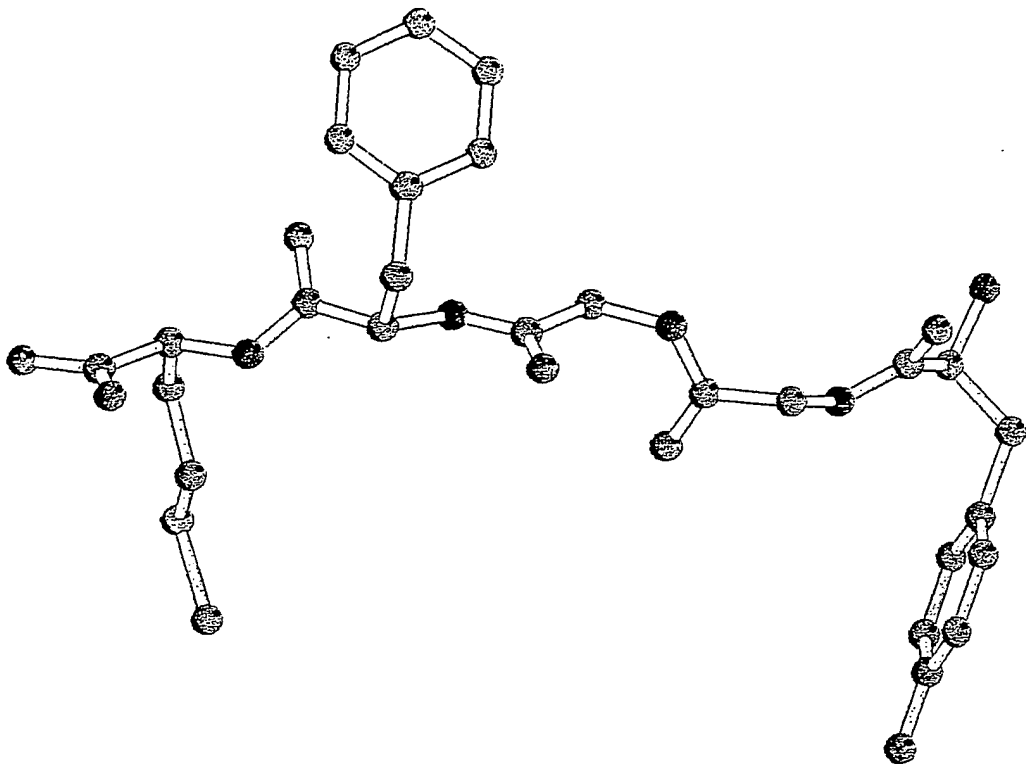
Figure 3b

Figure 4

Figure 5a

Figure 5b

Figure 6

Figure 7
Flow Chart

Start

Step a: Provide a representation of the molecule in terms of the atomic coordinates of the atoms in the molecule (by known methods)

Step b: Select a set of conformations, using MOLS, by first specifying the conformational parameters to be varied and by specifying the step size and range of sampling along each parameter

Step c: Fill up the sub squares of the MOLS with conformations of the molecules.

Step d: Choose the range and step size and the parameters to generate the set of input values $\theta_{r,s}$, r = 1 to n; s = 1to m

C

Step e: Label the sub squares with indices (u, t) and fill up sub squares with, $\theta^r_{u,t} = \theta_{r,s}$, where u = {(t-1)(r-1) + (s-1)} modulo (N)

A

Figure 7 (contd)

A

Step f: Repeat the procedure of filling up the sub squares
for all values of r = 1 to n, s = 1 to m and t = 1 to N, the
sub squares will correspond each to one conformation of
the molecule

Step g: Calculate the value of the potential energy by
known methods of each of the conformation of the
molecule defined in the sub squares.

Step h: Analyze the values of the potential energy function
to determine the best conformation

Step i: The analysis in the step h is carried out by
determining the average values of the energy of each value
of the each parameters

Step j: Scan the n x m average values determined in the step i
above to obtain the best value of the parameter. The set of best
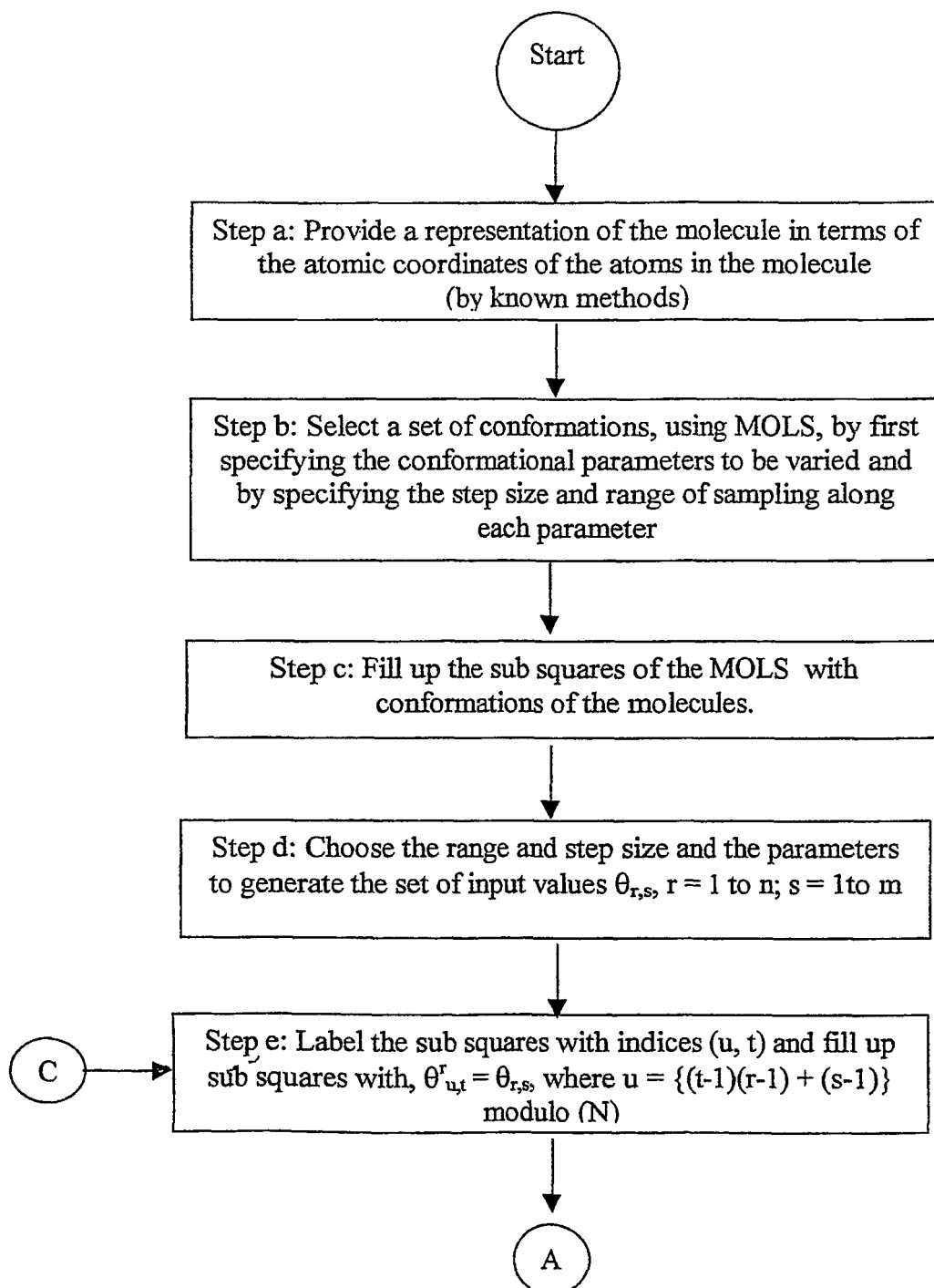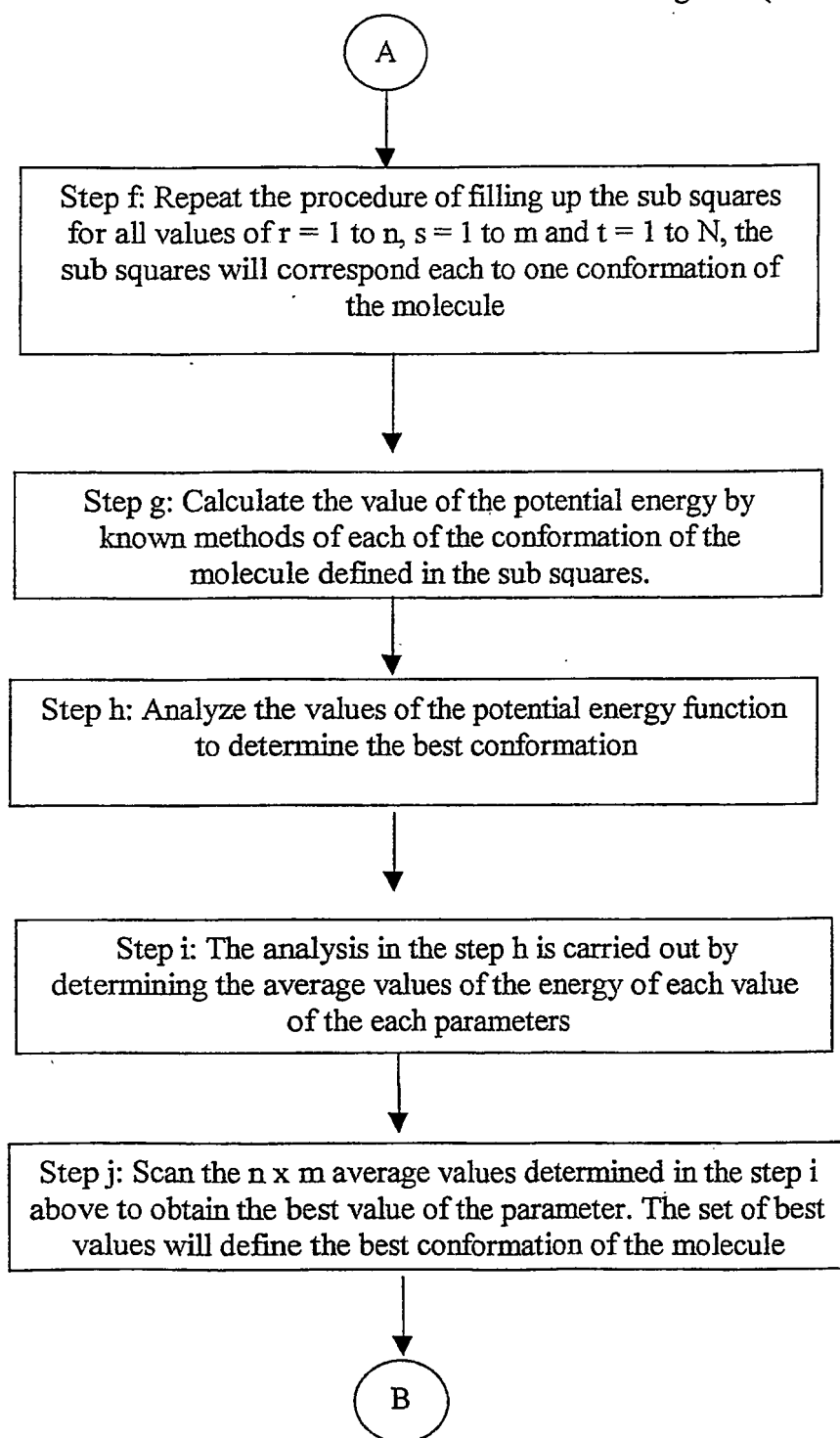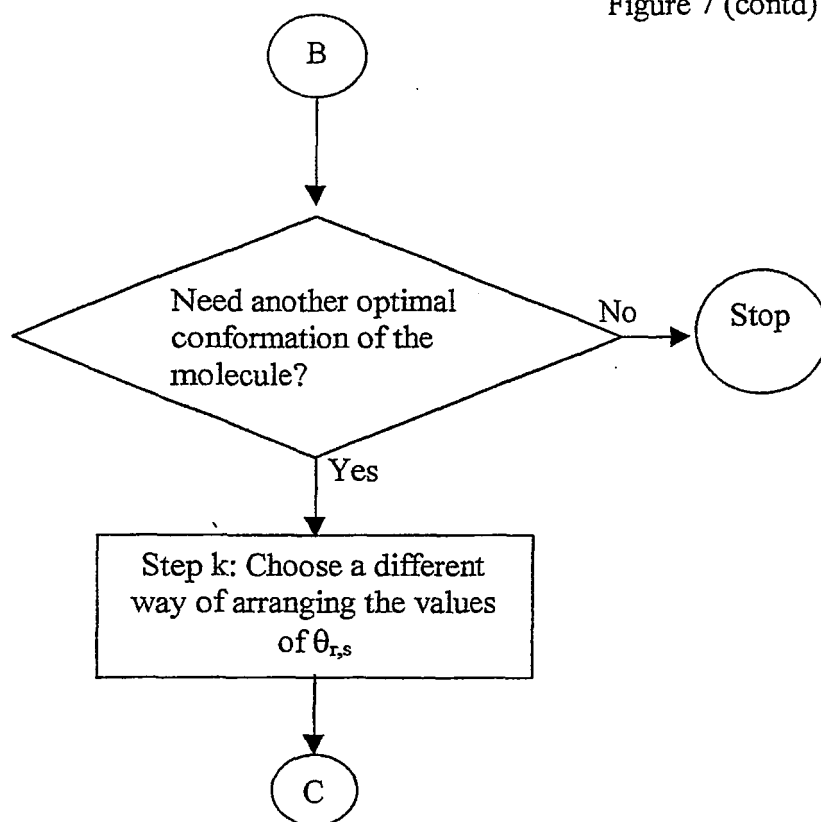values will define the best conformation of the molecule

B

Figure 7 (contd)

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 5884230 A, Srinivasan **[0002]**
- US 6188956 B, Mayo **[0003]**

Non-patent literature cited in the description

- Genetic Algorithms and Protein Folding. **S. SCHULZE-KREMER.** Methods in Molecular Biology. 2000 **[0007]**
- **N. GAUTHAM ; Z. A. RAFI.** Global Search for Optimal Biomolecular Structures using Mutually Orthogonal Latin Squares. *Current Science,* 1992 **[0007]**
- **K. D. GIBSON ; H. A. SCHERAGA.** Revised algorithms for build-up procedure for predicting protein conformations by energy minimization. *Journal of Computational Chemistry,* 1987 **[0008]**
- **L. PIELA ; H. A. SCHERAGA.** On the multiple-minima problem in the conformational analysis of polypeptides' I. Backbone degrees of freedom for a perturbed a-helix. *Biopolymers,* 1987 **[0008]**
- **Z. LI ; H. A. SCHERAGA.** Monte Carlo-Minimization approach to the multiple-minima problem in protein folding. *Proceedings of the National Academy of Sciences,* 1987 **[0009]**
- **D. R. RIPOLL ; A. LIWO ; H. A. SCHERAGA.** New developments of the electrostatically driven Monte Carlo method - Test on the membrane bound portion of melittin. *Biopolymers,* 1998 **[0009]**
- **L. PIELA ; J. KOSTROWICKI ; H. A. SCHERAGA.** The multiple-minima problem in the conformational analysis of molecules-Deformation of the potential energy hypersurface by the diffusion equation method. *Journal of Physical Chemistry,* 1989 **[0010]**
- **J. PILLARDY ; L. PIELA.** Molecular dynamics on deformed potential energy hypersurfaces. *Journal of Physical Chemistry,* 1995 **[0011]**
- **J. LEE ; H. A. SCHERAGA.** Conformational analysis of the 20-residue membrane-bound portion of melittin by conformational space annealing. *Biopolymers,* 1998 **[0012]**
- **A. LIWO ; J. LEE ; D. R RIPOLL ; J. PILLARDY ; H. A. SCHERAGA.** Protein structure can be predicted by global optimization of a potential energy function. *Proceedings of the National Academy of Sciences,* 1999 **[0013]**
- **D. CVIJIOVIC ; J. KLINOWSKI.** Taboo Search: An approach to the Multiple Minima Problem. *Science,* 1995 **[0014]**
- **J. BARTHEN ; V. PROTOPOPESCU ; D. REISTER.** TRUST: A deterministic algorithm for Global Optimization. *Science,* 1997 **[0015]**
- **Z. LI ; H. A. SCHERAGA.** *Proceedings of the National Academy of Sciences (USA,* vol. 84, 6611-6615 **[0032]**
- **GRIFFIN et al.** *Proceedings of the National Academy of Sciences (CSA,* vol. 83, 3272-3276 **[0034]**
- **BENEDETTI et al.** *Journal of the Chemical Society,* vol. 104, 2437-2444 **[0036]**
- **SHAMALA et al.** *Journal of Chemical Society, Chemical Communications,* 1978, 996 **[0036]**