

METHOD AND APPARATUS FOR CONFORMATIONALLY ANALYZING MOLECULAR FRAGMENTS

COPYRIGHT NOTICE

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority from the following U.S. Provisional Application, the disclosure of which, including all appendices and all attached documents, is incorporated by reference in its entirety for all purposes:

U.S. Provisional Patent Application Serial No. 60/048,944, in the name of Andrew S. Smellie and Steven L. Teig, entitled, "Method, Apparatus, and Article of Manufacture for Conformationally Analyzing a Molecule," filed June 24, 1997.

Further, this application makes reference to commonly owned, co-pending U.S. Provisional Patent Application Serial No. _____, in the name of Jonathan W. Greene and John Mount, entitled, "Method and System for Search of Implicitly Described Virtual Libraries," filed March 27, 1998 (attorney docket number 18590-0002000) which is incorporated herein by reference for all purposes.

BACKGROUND OF THE INVENTION

The present invention relates generally to the determination of molecular structure of chemical compounds by using computer based methods to analyze subspecies of the molecule, then combining the results from these analyses to determine the properties of the molecule.

Molecules have one or more three-dimensional structures. A molecule's structure determines its chemical, physical and bio-active properties. Scientists use a set of convenient parameters, such as bond length, bond angle and torsion angles, to describe the organization of atoms within a molecule that give rise to its molecular structure.

Researchers in the pharmaceutical field, for example, have sought for some time for a way to systematically analyze molecular structures of chemical compounds in order to determine their suitability as medicines. A conformation is the spatial arrangement of the atoms in a molecule at any point in time that results from rotation of parts of the molecule about covalent bonds and the "bending" of bond angles. Researchers in other fields also desire to search for chemical compounds having desirable attributes by analyzing fragments of molecules with a computer based method, rather than subjecting samples of the compounds to chemical analyses in a laboratory.

In a commonly owned, co-pending U.S. Provisional Patent Application Serial No. _____, entitled, "Method and System for Search of Implicitly Described Virtual Libraries," Greene and Mount describe a method for searching chemical fragment libraries for molecules using search keys including conformation and pose of molecules. While this is an important contribution to the field of drug research, there is no method taught for automatically determining a conformation for a chemical substance based upon conformation data about its constituents.

What is needed is a method of determining a conformation for a molecule based upon information about the structure of its constituents.

10

SUMMARY OF THE INVENTION

The present invention provides techniques for improved automated determination of molecular information. More particularly, the present invention provides a method for analyzing the structure of compounds.

15

According to an embodiment of the present invention, a method for determining a conformation for a molecular structure is provided. The method includes a variety of steps such as computationally decomposing the molecular structure into fragments. A step of normalizing each of the fragments in order to form normalized fragments is also included. The step of determining at least one or possibly many conformers for each normalized fragment is included in the method. Finally, the step of combining at least a first conformer and a second conformer in order to derive the molecular structure is performed. Other conformers may also be included in the combination.

20

Some embodiments will also include a step of searching for one or more conformers fragments in a library. If the fragment is found in the library, then its corresponding conformers are read from the library. Otherwise, the method includes an additional step of storing the fragment and conformer information it produces in the determining step into the library for subsequent analyses.

25

In another aspect according to the present invention, a method for determining a conformation for a molecular structure is provided. The method includes a variety of steps such as computationally decomposing the molecular structure into a plurality of fragment molecular structures. A step normalizing the plurality of fragments to form a plurality of normalized fragments is also included. A step of determining at least one of a plurality of conformers for each normalized fragment can be included. Select embodiments also include a step of associating a plurality of internal coordinates with each conformer. One such set of internal coordinates are torsion models, such that each of the torsional models possess a plurality of characteristic angle values corresponding to each chemical bond representation in the conformer. Then a step of determining an energy difference level for each conformer may be performed. This energy difference level represents an incremental amount of potential energy above a specified nominal energy value required to maintain each chemical bond representation at each of the plurality of characteristic angle values in the torsional model. A step of selecting for each torsional model, a conformer having a corresponding minimum energy difference level is performed. The corresponding minimum energy difference level is selected from the energy difference levels computed for each of the torsional models, for each characteristic angle value.

30

35

A step of determining a maximum energy difference value for each characteristic angle value from among the corresponding minimum energy difference levels is included. The step of selecting from the plurality of characteristic angle values, a plurality of likely angle values for each torsional model is included in the method. Also, a step for selecting a plurality of candidate conformers from the plurality of conformers so that each candidate conformer corresponds to a torsional model having at least one characteristic angle value in the plurality of likely angle values can be part of the method. Another step is choosing at least two candidate conformers from the plurality of candidate conformers and combining the candidates to produce a conformation for the molecular structure. In some embodiments, candidate conformers are chosen such that the energy difference level of each of the candidate conformers is less than a specified cutoff energy level.

In another aspect of the present invention, a computer programming product for determining a conformation for a molecular structure is provided. The computer programming product includes a variety of computer code for performing a plurality of functions, such as code for automatically decomposing said molecular structure into a plurality of fragments. Code for normalizing each fragment to form a plurality of normalized fragments may also be included. The computer programming product can also include code for determining for each normalized fragment, at least one conformer. Code for combining at least a first conformer and a second conformer to derive the conformation for the molecular structure is also included. Finally, the computer programming product also includes a computer readable storage medium for storing the codes.

In another aspect of the present invention, a computer programming product for determining a low energy conformation for a molecular structure is provided. The computer programming product includes a variety of computer code for performing a plurality of functions, such as code for decomposing the molecular structure into a plurality of fragments. Code for normalizing the plurality of fragments to form a plurality of normalized fragments is also included. The computer programming product can also include code for determining for each normalized fragment, at least one conformer having at least one chemical bond representation, each chemical bond representation interconnecting at least two atoms in the conformer. Code for associating with each conformer a plurality of internal coordinates. One such set of internal coordinates are torsion models, each torsional model having a plurality of particular characteristic angle values corresponding to each chemical bond representation in the conformer is also included. The product can also include code for determining for each conformer, an energy difference level representing an incremental amount of potential energy above a first nominal energy value required to maintain each chemical bond representation at each of the characteristic angle values in the torsional model. Also, code for selecting, for each torsional model and each characteristic angle value, a conformer having a corresponding minimum energy difference level selected from the energy difference levels computed for each conformer, forming a plurality of corresponding minimum energy difference levels can be included. Code for determining, for each characteristic angle value, a maximum energy difference value from among the corresponding minimum energy difference levels, forming a plurality of maximum energy difference values is part of the computer programming product. Code for selecting a plurality of likely angle values from among the plurality of characteristic angle values, for each torsional model, such that the plurality of likely

angle values in the torsional model corresponds to a second nominal energy value selected from the plurality of maximum energy difference levels is also included. Also, code for selecting a plurality of candidate conformers from the plurality of conformers, such that each candidate conformer corresponds to a torsional model having at least one characteristic angle value in the plurality of likely angle values can be included.

5 Finally, code for combining at least two candidate conformers chosen from the plurality of candidate conformers to produce a conformation for said molecular structure is included. The candidate conformers are chosen such that the energy difference level of each of said at least two candidate conformers is less than a specified cutoff energy level. Further, the at least two candidate conformers are chosen such that they have at least one common torsional model. The computer programming product also includes a computer readable
10 storage medium which holds the various codes.

In a yet further aspect of the present invention, a method for determining likely torsion values for conformers where each conformer has a plurality of particular chemical bond representations includes a variety of steps such as associating a plurality of internal coordinates with each conformer. One such set of internal coordinates are torsion models in a plurality of conformers. A step of determining an
15 energy difference level, for each conformer is also included in the method. A step of selecting a conformer, having a corresponding minimum energy difference level for each torsional model, for each characteristic angle value, is part of the method. Further, the method can include a step of determining a maximum energy difference value for each characteristic angle value from among the corresponding minimum energy difference levels. A step of selecting a plurality of likely angle values from among the plurality of
20 characteristic angle values such that the plurality of likely angle values in the torsional models corresponds to a nominal energy value

Numerous benefits are achieved by way of the present invention over conventional techniques. In some embodiments, the present invention provides more automated methods of analyzing molecular structures using a computer than many of the manual techniques heretofore known. The present
25 invention can also provide a library of fragments which is without theoretical limit as to size or scope. Some embodiments according to the invention can consider entire molecular contexts in determining a molecular structure based upon fragment conformers. Select embodiments according to the invention may be more robust than those using known techniques. These and other benefits are described throughout the present specification and more particularly below.

30 The invention will be better understood upon reference to the following detailed description and its accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1A depicts a simplified block diagram of a representative hardware embodiment
35 according to the invention;

Fig. 1B depicts a functional perspective of the representative hardware embodiment according to the invention;

Figs. 2A-2E depict perspective views of a representative molecular structure having various conformations;

Fig. 3 depicts a graph of potential energy of a representative molecular system through the course of a rotation about a particular molecular bond;

Figs. 4A-4D depict flowcharts of representative processing steps in select embodiments according to the invention;

5 Fig. 5 depicts an example molecule;

Fig. 6 depicts an example fragment molecule of the molecule in Fig. 5;

Fig. 7 depicts a graph representation of the molecule of Fig. 5 with its molecular structure shown as a collection of nodes and edges;

Figs. 8A-8F depict possible fragments of the graph of Fig. 7;

10 Fig. 9 depicts a hypothetical molecular structure being broken into overlapping fragments;

Fig. 10 depicts conformation sets for each of fragments of Fig. 9;

Fig. 11 depicts the intersection of conformation sets of Fig. 10;

Fig. 12 depicts a sample molecule containing a ring structure;

Fig. 13 depicts the demarcation of a ring within the molecule of Fig. 12;

15 Fig. 14 depicts some of the fragments that are formed during decomposition of the molecule of Fig. 12;

Fig. 15 depicts tables used in conformational analysis and intersection of ring nodes;

Fig. 16 depicts an example molecular structure computationally decomposed into a plurality of fragments; and

20 Figs. 17-22 depict simplified representative tables used in conformational analysis and intersection.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

25 1.0 Introduction

The present invention provides techniques for determining conformers of molecules based upon information about component molecular fragments. Systems according to the present invention enable researchers and scientists to identify promising candidate compounds in the search for new and better substances.

30 1.1 Hardware Overview

The method for determining a molecule's conformers is implemented in the C++ programming language and is operational on a computer system such as shown in Fig. 1A. This invention may be implemented in a client-server environment, but a client-server environment is not essential. Fig. 1A shows a conventional client-server computer system that includes a server 20 and numerous clients, one of which is shown as client 25. The use of the term "server" is used in the context of the invention, wherein the server receives queries from (typically remote) clients, does substantially all the processing necessary to formulate responses to the queries, and provides these responses to the clients. However, server 20 may

itself act in the capacity of a client when it accesses remote databases located at another node acting as a database server.

The hardware configurations are in general standard and will be described only briefly. In accordance with known practice, server 20 includes one or more processors 30 that communicate with a number of peripheral devices via a bus subsystem 32. These peripheral devices typically include a storage subsystem 35, comprised of memory subsystem 35a and file storage subsystem 35b. Storage subsystem 35 is disposed to hold computer programs (e.g., code or instructions) and data. Other peripheral devices include a set of user interface input and output devices 37, and an interface to outside networks, which may employ Ethernet, Token Ring, ATM, IEEE 802.3, ITU X.25, Serial Link Internet Protocol (SLIP) or the public switched telephone network. This interface is shown schematically as a "Network Interface" block 40. It is coupled to corresponding interface devices in client computers via a network connection 45.

Client 25 has the same general configuration, although typically with less storage and processing capability. Thus, while the client computer could be a terminal or a low-end personal computer, the server computer is generally a high-end workstation or mainframe, such as a SUN SPARCTM server. Corresponding elements and subsystems in the client computer are shown with corresponding, but primed, reference numerals.

The user interface input devices typically includes a keyboard and may further include a pointing device and a scanner. The pointing device may be an indirect pointing device such as a mouse, trackball, touch pad, or graphics tablet, or a direct pointing device such as a touch screen incorporated into the display. Other types of user interface input devices, such as voice recognition systems, are also possible.

The user interface output devices typically include a printer and a display subsystem, which includes a display controller and a display device coupled to the controller. The display device may be a cathode ray tube (CRT), a flat-panel device such as a liquid crystal display (LCD), or a projection device. The display controller provides control signals to the display device and normally includes a display memory for storing the pixels that appear on the display device. The display subsystem may also provide non-visual display such as audio output.

The memory subsystem typically includes a number of memories including a main random access memory (RAM) for storage of instructions and data during program execution and a read only memory (ROM) in which fixed instructions are stored. In the case of Macintosh-compatible personal computers the ROM would include portions of the operating system; in the case of IBM-compatible personal computers, this would include the BIOS (basic input/output system).

The file storage subsystem provides persistent (non-volatile) storage for program and data files, and typically includes at least one hard disk drive and at least one floppy disk drive (with associated removable media). There may also be other devices such as a CD-ROM drive and optical drives (all with their associate removable media). Additionally, the computer system may include drives of the type with removable media cartridges. The removable media cartridges may, for example be hard disk cartridges, such as those marketed by SyQuest and others, and flexible disk cartridges, such as those marketed by Iomega. One or more of the drives may be located at a remote location, such as in a server on a local area network or at a site of the Internet's World Wide Web.

In this context, the term "bus subsystem" is used generically so as to include any mechanism for letting the various components and subsystems communicate with each other as intended. With the exception of the input devices and the display, the other components need not be at the same physical location. Thus, for example, portions of the file storage system could be connected via various local-
5 area or wide-area network media, including telephone lines. Similarly, the input devices and display need not be at the same location as the processor, although it is anticipated that the present invention will most often be implemented in the context of PCs and workstations.

Bus subsystem 32 is shown schematically as a single bus, but a typical system has a number of buses such as a local bus and one or more expansion buses (e.g., ADB, SCSI, ISA, EISA, MCA, NuBus, or PCI), as well as serial and parallel ports. Network connections are usually established through a device
10 such as a network adapter on one of these expansion buses or a modem on a serial port. The client computer may be a desktop system or a portable system.

Fig. 1B is a functional diagram of the computer system of Fig. 1A. Fig. 1B depicts a server 20, and a representative client 25 of a multiplicity of clients that may interact with the server 20 via
15 the internet 45 or any other communications method. Blocks to the right of the server are indicative of the processing components and functions that occur in the server's program and data storage indicated by block 35a in Fig. 1A. A TCP/IP "stack" 44 works in conjunction with Operating System 42 to communicate with processes over a network or serial connection attaching Server 20 to internet 45. Web server software 46 executes concurrently and cooperatively with other processes in server 20 to make data objects 50 and 51
20 available to requesting clients. A Common Gateway Interface (CGI) script 55 enables information from user clients to be acted upon by web server 46, or other processes within server 20. Responses to client queries may be returned to the clients in the form of a Hypertext Markup Language (HTML) document outputs which are then communicated via internet 45 back to the user.

Client 25 in Fig. 1B possesses software implementing functional processes operatively
25 disposed in its program and data storage as indicated by block 35a' in Fig. 1A. TCP/IP stack 44', works in conjunction with Operating System 42' to communicate with processes over a network or serial connection attaching Client 25 to internet 45. Software implementing the function of a web browser 46' executes concurrently and cooperatively with other processes in client 25 to make requests of server 20 for data objects 50 and 51. The user of the client may interact via the web browser 46' to make such queries of the
30 server 20 via internet 45 and to view responses from the server 20 via internet 45 on the web browser 46'.

2.0 Torsion and Conformation

Fig. 2A depicts a model of a molecule 101, in this example an ethane molecule. Ethane molecule 101 comprises the carbon atoms 110 and 120 and hydrogen atoms 115a, 115b, 115c, 125a, 125b,
35 and 125c. Carbon atom 110 is bound to hydrogen atoms 115a, 115b, and 115c via single bonds 116a, 116b, and 116c, respectively, forming a tetrahedron with carbon atom 110 at its center. Carbon atom 120 is bound to hydrogen atoms 125a, 125b, and 125c via single bonds 126a, 126b, and 126c, respectively, forming a second tetrahedron. Carbon atom 110 is bound to carbon atom 120 via single bond 130. Single bond 130 is subject to torsion, whereby each tetrahedron rotates relative to the other

Figs. 2B through 2E depict two conformations of ethane molecule 101. Fig. 2B depicts a perspective view of molecule 101 in a staggered conformation. In the staggered conformation, bond 130 is subjected to a torsion such that each hydrogen atom 115a, 115b, and 115c is positioned the maximum angular distance from each hydrogen atom 125a, 125b and 125c. Fig. 2C depicts a Newman projection of molecule 101 in a staggered conformation. It will be noted that, in the staggered conformation, each hydrogen atom 115a, 115b, and 115c is positioned 60 degrees from each hydrogen atom 125a, 125b and 125c in the plane perpendicular to the longitudinal axis of bond 130. This angle is referred to as the "dihedral angle" or the "angle of torsion."

Fig. 2D depicts a perspective view of molecule 101 in an eclipsed conformation. In the eclipsed conformation, bond 130 is subjected to a torsion such that each hydrogen atom 115a, 115b, and 115c is positioned the minimum angular distance from each hydrogen atom 125a, 125b and 125c. Fig. 2E depicts a Newman projection of molecule 101 in an eclipsed conformation. It will be noted that, in the eclipsed conformation, each hydrogen atom 115a, 115b, and 115c is positioned coincident to the same position of each hydrogen atom 125a, 125b and 125c in the plane perpendicular to the longitudinal axis of bond 130. That is, the angle of torsion is zero degrees.

Torsions are an example of an internal coordinate representation of the molecule whereby with knowledge of the torsion angle about the bond 130 and coordinates of each tetrahedron centered at carbon atoms 110 and 120, coordinates for all the atoms may be determined. In general, the representation of a conformation by its internal coordinates can be used to generate atomic coordinates for each atom in the molecule (modulo rotation and translation). There are many types of internal coordinates, well-known to those skilled in the art, that can be used in conjunction with the invention, but, in the preferred embodiment, torsion angles are the internal coordinate system used.

In the course of a complete 360-degree rotation, the molecule alternates between three instances of the staggered conformation and three instances of eclipsed conformation. The three instances of staggered conformation occur at 60 degrees, 180 degrees and 300 degrees. The three instances of eclipsed conformation occur at zero (or 360) degrees, 120 degrees, and 240 degrees.

Experimental observations have shown that torsion angles are easier to change than bond angles, which in turn are easier to deform than bond lengths. This observation lead to the "fixed valence approximation" in which bond lengths and bond angles are assumed to be invariant, leaving only torsional angles as determiners of a molecule's structure. Even plastic molecular models hold bond angle and bond length fixed.

Generally, the number of conformers to be determined for a particular molecule may be calculated by the expression $N = s^b$, where s is the number of samples of angular torsion to be examined for a bond, and b is the number of bonds whose torsion is to be sampled. For example, to calculate conformers for ethane molecule 101, given a sampling at 0, 60, 120, 180, 240 and 300 degrees, six samples are made of a single bond. Therefore, the number of conformers is $N = s^b = 6^1 = 6$. Calculating conformers for a simple molecule such as ethane is computationally inexpensive. However, as the complexity of a molecule increases, the cost of calculating its conformers increases exponentially. For example, to calculate the conformers of a molecule having only 20 rotatable bonds and sampling each bond's torsion angle at 60-

degree increments (i.e., six samples over 360 degrees of angle), $N = 3^b = 6^{20} = 3,656,158,440,062,980$ conformers must be calculated. Increasing the number of samples to every 10 degrees (i.e., 36 samples over 360 degrees of angle) on a molecule having 100 rotatable bonds would require the generation of 4.268×10^{155} conformers. The calculation and evaluation of potential preferred conformers as required in present-day drug discovery using conventional systematic searching methods clearly is beyond the capabilities of even the fastest available computers. An alternative that has been tried is the use of random selection to find promising conformers by weighted chance, and then exploring similar conformers systematically. This approach, however, by its very nature risks missing useful conformers that were not randomly selected.

Additional information regarding conformation and torsion may be found in Eliel, Stereochemistry of Carbon Compounds, Ch. 6, pp. 124-179 (1962), the disclosure of which is hereby incorporated by reference.

3.0 Potential Energy of a Conformer

Fig. 3 depicts the potential energy of the ethane molecule of Fig. 2A through the course of a 360-degree rotation. Plotted line 310 graphically depicts the potential energy as a function of the angle of torsion through the course the rotation. Each point on the plotted line corresponds to a possible conformation of the ethane molecule 101. The points 330a, 330b and 330c correspond to the torsion angles of 60, 180, and 300 degrees, respectively, and represent the three staggered conformations. The points 320a, 320b, and 320c correspond to the torsion angles of 120, 240 and 360 degrees, respectively, and correspond to the three eclipsed conformations. It will be seen that the three staggered conformations represent the conformations at the energy minima and are therefore the most preferred conformations.

4.0 Process Flow of the Specific Embodiments

Fig. 4A depicts a flowchart 401 of simplified process steps in a particular representative embodiment according to the invention for determining a conformation for a molecular structure. In a step 410, the molecular structure is computationally decomposed into a plurality of fragment molecular structures. Then, in a step 420, each one of the plurality of fragment molecular structures is normalized in order to form a plurality of normalized fragments. Subsequently, in a step 430, at least one of a plurality of conformers is determined for each normalized fragment produced in step 420. Finally, in a step 440, at least a first conformer and a second conformer determined in step 420, are combined in order to derive a conformation for the entire molecular structure.

Fig. 4B depicts a flowchart 403 of simplified process steps in another embodiment according to the invention. In this embodiment, in a step 410, the molecular structure is computationally decomposed into a plurality of fragment molecular structures. Then, in a step 420, each one of the plurality of fragment molecular structures is normalized in order to form a plurality of normalized fragments. Next, in a step 432, a search is performed for one or more normalized fragments in a library. Then, in a decisional step 434, a determination is made whether the fragment was located in the library. If the fragment is found in the library, then in a step 436, its corresponding conformers are read from the library.

Otherwise, in a step 430, at least one of a plurality of conformers is determined for each normalized fragment produced in step 420. Next, in a step 438, the fragment and conformer information produced in step 430 is stored in the library for subsequent analyses. Finally, in a step 440, at least a first conformer and a second conformer determined in step 430, or found in the library in step 436 are combined in order to
5 derive a conformation for the entire molecular structure.

Fig. 4C depicts a flowchart 405 of simplified process steps in another embodiment according to the invention. In this embodiment, in a step 450, the molecular structure is computationally decomposed into a plurality of fragment molecular structures. Then, in a step 452, each of the plurality of fragments is normalized to form a plurality of normalized fragments. Next, in a step 454, at least one of a
10 plurality of conformers is determined for each normalized fragment determined in step 452. Each conformer in the plurality of conformers has at least one of a plurality of chemical bond representations, and each chemical bond representation interconnects at least two atoms in the conformer. In a step 456, a plurality of internal coordinates (i.e., torsional models) is associated with each conformer, each of the torsional models possessing a plurality of characteristic angle values corresponding to each chemical bond representation in the conformer. Then, in a step 458, for each conformer, an energy difference level is determined. This energy
15 difference level represents an incremental amount of potential energy above a specified nominal energy value required to maintain each chemical bond representation at each of the plurality of characteristic angle values in the torsional model. Next, in a step 460, for each torsional model, for each characteristic angle value, a conformer having a corresponding minimum energy difference level, is selected. The corresponding
20 minimum energy difference level is selected from the energy difference levels computed for each of the conformers in each of the fragments. In a step, 462, a maximum energy difference value is determined for each characteristic angle value from among the corresponding minimum energy difference levels. This forms a plurality of maximum energy difference values corresponding to the plurality of conformers. Then, in a step 463, a plurality of likely angle values is selected from among the plurality of characteristic angle
25 values such that the plurality of likely angle values in the torsional models corresponds to a nominal energy value. This nominal energy value is determined by adding an incremental energy to the maximum energy difference values formed in step 462. Next, in a step 464, a plurality of candidate conformers is selected from the plurality of conformers, each candidate conformer corresponds to a torsional model having at least one characteristic angle value in the plurality of likely angle values determined in preceding step 463. Finally, in a step 466, at least two candidate conformers chosen from the plurality of candidate conformers
30 are combined to produce a low energy conformation for the molecular structure. Candidate conformers are chosen such that the energy difference level of each of the candidate conformers is less than a specified cutoff energy level. Further, the candidate conformers are chosen such that they have at least one common torsional model.

Fig. 4D depicts a flowchart 407 of simplified process steps in another embodiment according to the invention. In this embodiment, in a step 470, a plurality of internal coordinates (i.e. torsional models) is associated with each conformer in a plurality of conformers. Each of the torsional models possesses a plurality of characteristic angle values corresponding to each chemical bond
35 representation in the conformer. Then, in a step 472, for each conformer, an energy difference level is

determined. This energy difference level represents an incremental amount of potential energy above a specified nominal energy value required to maintain each chemical bond representation at each of the plurality of characteristic angle values in the torsional model. Next, in a step 474, for each torsional model, for each characteristic angle value, a conformer having a corresponding minimum energy difference level, is selected. The corresponding minimum energy difference level is selected from the energy difference levels computed for each of the conformers in each of the fragments. In a step 475, a maximum energy difference value is determined for each characteristic angle value from among the corresponding minimum energy difference levels. This forms a plurality of maximum energy difference values corresponding to the plurality of conformers. Then, in a step 476, a plurality of likely angle values is selected from among the plurality of characteristic angle values such that the plurality of likely angle values in the torsional models corresponds to a nominal energy value. This nominal energy value is determined by adding an incremental energy to the maximum energy difference values formed in step 475.

4.1 Decomposition

Select embodiments according to the invention employ a process of computationally decomposing the molecule under analysis into a plurality of fragments. The fragment size is selected to allow for a reduced number of calculated conformers for the fragment, while still maintaining a sufficient length so that the fragment conformers produced accurately represent the possible configurations of that fragment within the entire molecule.

In the general case, it is not obvious how the molecular decomposition should be done. In other works, molecules have been fragmented at bonds which are obvious to those skilled in art. One such decomposition is that of polypeptides into amino acid sequences (described in Vasquez and Scheraga, Biopolymers, Vol 24, pp.1437-1447, which is incorporated herein by reference in its entirety for all purposes). Other decompositions have partitioned the molecule into rings and chains (see, for example, Sadowski and Gasteiger, Chem Rev., 1993, 93, pp 2567-2581 and Leach, Prout and Dolata, J.Comp.-Aided Mol.Des.,4(1990), pp271-282). Where such decompositions are not obvious, as is the case for general molecules, other fragmentation scheme must be considered.

A particularly useful fragment size is a fragment comprising up to 4 rotatable bonds. A fragment of this size includes the five atoms bound to the rotatable bonds, plus every atom present in the molecule that is bonded to any of the five atoms. Such a fragment is said to have a path length of length 7. At times in this specification, the process of moving, counting, measuring, etc., along a path from one atom or bond to the next is referred to as "walking" the fragment or path length.

In a fragment having a path-length of 7, four of the bonds are subject to rotation in a conformational analysis of the fragment. The remaining two bonds, while subject to rotation within the molecule, have little effect on the conformation of the fragment, and need not be rotated in a conformational analysis of the fragment. It is known that, for drug-sized molecules, the overwhelming number of close atomic contacts occur between atoms that are connected by a path length of 7 or closer. Thus, if conformers

are constructed that eliminate steric clashes within a length 7, it is unlikely that a molecule made up of such conformers will have any steric clashes at all.

By limiting a fragment to consideration of up to four rotatable bonds, the number of potential conformers that need to be analyzed for the fragment is greatly reduced. As will be discussed below with respect to conformer generation, to generate a conformer for the fragment each bond is sampled either 8, 18, or 36 times, depending on the types of atom groups attached to the bond. The atom groups result from hybridization of atomic orbitals and is discussed further below. Therefore, exhaustive calculation of a particular fragment will generate only between 8^4 (4096) and 36^4 (1,679,616) potential conformer configurations for the fragment. This range of potential configurations is well within a practical computational capabilities of widely available and inexpensive computer systems.

The use of a fragment in which rotation around four bonds is analyzed therefore allows for a reduced number of calculated conformers for the fragment and yet has a sufficient length so that the fragment conformers produced accurately represent the possible configurations of that fragment within the entire molecule.

Fig. 7 depicts a graphic representation 701 of the molecule of Fig. 5 with its molecular structure described as a plurality of nodes and edges. A node represents a collection of atoms. An edge represents a freely rotatable bond between two atoms. Within a single node, no two atoms are connected by an edge. In other words, there are no freely rotatable bonds between any two atoms represented by a node. Nodes 710, 725, 730, 732, 734, 736, and 738 correspond to atoms 510, 525, 550, 530, 540, 560, and 570, and their related outlying univalent atoms, respectively. Edges 750, 752, 754, 756, 758 and 760 correspond to bonds 554, 534, 542, 562, 571 and 572, respectively. Because outlying bonds to univalent atoms such as bond 680 are not significantly rotatable and will not be analyzed in a conformational analysis, the atoms bonded by these bonds are represented by the node that also represents the atom to which they are bound. For example, node 710 represents nitrogen atom 510 and hydrogen atoms 511 and 512, because bond 562 is rotatable and delimits the node.

A molecule represented as a graph may be broken up into a series of paths, each path having a length equal to or less than a selectable maximum number of atoms, which may be five atoms or fewer in select embodiments. These atoms, combined with all atoms bonded to them, form the path that has a maximum path length of seven or fewer atoms. Figs. 8A through 8F depict six possible fragments of graph 501 having a length 7 or less. Fig. 8A depicts a first possible fragment 801, comprising nodes 730, 734, 736, 738, and 725 connected by edges 750, 754, 758, and 760. Fig. 8B depicts a second possible fragment 802, comprising nodes 732, 734, 736, 738 and 725 connected by edges, 752, 754, 758, and 760. Fig. 8C depicts a third possible fragment 803, comprising nodes 732, 734, 736 and 710 connected by edges 752, 754 and 756. Likewise, Figs. 8D through 8F depict additional possible fragments.

To further optimize the conformation generation process, fragments that are known not to provide additional information are excluded from further consideration. For example, in a scenario where a path of length 7 walks into a ring, walking more than two bonds around the ring means that, within the local context of the fragment, the ring constrains the end points of the path so they can never interact. The bonds that are in the ring are not freely rotatable, but are constrained by the conformations of the ring structure as a

whole. The conformations of the path in isolation are constrained when a path intersects a ring. This interaction between paths and rings is explained more fully below.

4.2 Normalization

5 Once the fragments have been automatically generated from the decomposition of the molecule as described herein, the fragments are normalized to capture some of their context in the environment of the molecule. In general, a fragment in isolation will have a different conformational model than a fragment in context of a molecule because of the influence of the surrounding environment. In a present preferable embodiment, each fragment isolated from the molecule is stubbed, or *normalized*, to
10 simulate the rest of the molecule environment. Consider the fragment 806 of Fig. 8F. This contains nodes 730, 734 and 732 which correspond to atoms 550, 540 and 530 (with their associated univalent atoms) respectively. This fragment is shown in Fig 6, where atoms 650, 640 and 630 correspond to original atoms 550, 540 and 530 respectively. Retaining atoms one level out means that atom 660 in Fig 6 corresponds to original atom 560. To simulate the environment of the whole molecule, and satisfy valency, hydrogen atoms
15 661, 662 and 663 are added to model atoms 561, 570 and 510 respectively. This embodiment normalizes the fragment by stubbing out the molecular environment with hydrogen. Other embodiments may not stub at all, or stub with an arbitrarily large fragment intended to simulate the rest of the molecule, without departing from the scope of the invention.

4.3 Conformer Generation

20 Select embodiments according to the invention include a step of determining a set of potential conformers for each fragment. A database can be employed to obtain sets of conformers that have been generated previously and to store sets of conformers that are being generated for the first time. Embodiments using the database include an additional step reducing each fragment to a text string that
25 describes the atomic structure of the fragment for use as a database search key. The text string is put in canonical form so that the same text string is generated, even if there are variations in the fragment, such as reversed order of atoms, and so forth, that do not affect its conformational analysis. In this case, the generation of the canonical text string comprises of two steps: (1) placing the atoms of the fragment in a canonical order; (2) using this canonical order to generate a unique text string.

30 Canonical orderings are generated by first generating a simple hash key for each atom in the fragment. This key is formed by encoding the atom's simple properties (e.g., atomic number, charge, hybridization, aromaticity) to make the hash key. Each key is examined for unique atoms within the fragment. If any are found, they are placed in the canonical ordering in lexicographic order. If two or more atoms have the same simple hash key, their neighbors are recursively examined to update the hash key.
35 Then the keys are re-examined to look for unique atoms, which are placed in the canonical ordering. This process is repeated until all distinguishable atoms have been identified. If two atoms are truly the same in the fragment, they are arbitrarily given unique names.

 Once a canonical ordering is generated, the canonical text string is generated by encoding each atom's simple properties and storing this encoding in canonical order. Then, for each atom in canonical

order, all bonds from this atom (in canonical order) are encoded and written to the canonical string. Thus, the canonical text string is formed by encoding information about the atoms and bonds in canonical order.

After canonical text string is formed, it is used as a key in a call to a database management system (DBMS) to retrieve a previously determined set of conformers for the fragment, as indicated by Step 436 in Fig. 4B. In the first iteration, there will generally be no previously stored conformers. As a result, the DBMS will indicate that the set was not found. Select embodiments may include a supply of conformational models of frequently used fragments.

For example, although fragment 801 and fragment 802 are different fragments within the molecule, they differ only in that, in fragment 801, node 730 represents atom 630, while in fragment 802, node 732 represents atom 632. However, both atom 630 and atom 632 are carbon atoms tetrahedrally bound to three hydrogen atoms. Therefore, a fragment conformation set calculated for fragment 801 is equally applicable to fragment 802. When it is necessary to determine a conformation set for fragment 802, the previously calculated conformation set form 801 may be used.

If no previously stored conformer set exists for the fragment, a conformer set is generated for the fragment. The conformer set may be generated using any prior art means. In one embodiment, the systematic search method is the preferable means for generating conformers. In systematic search, a fixed set of search positions is established for each rotatable bond. Then each combination of torsions is systematically searched. Coordinates of the atoms are generated for each combination of torsions, and the energy of the conformer is measured. Conformers having an energy value within N kcals of a global minimum energy value are retained. Energy level N is obtained by multiplying the number of rotatable bonds by a factor of 1.0-4.0. Systematic search methods are more fully described in Lipton and Still, *The Multiple Minimum Problem in Molecular Modelling. Tree Searching in Internal Coordinate Space*, Journal of Computational Chemistry, 1989, Vol.9, No.9, pp.343-55, which is incorporated herein by reference in its entirety for all purposes.

The systematic search is performed on a number of different degrees of torsion. By way of example, in a representative embodiment, the number of tests can be dependent upon the nature of the atom groups on either end of the bond subjected to torsion which in turn depends on hybridization of atomic orbitals that occur in bonding. For example, in a representative embodiment, if both atom groups are of order sp^3 , then about 18 samples are taken. If both atom groups are of order sp^2 with a single bond, then some 24 samples may be taken. If both atom groups are of order sp^2 with a double bond, then approximately 6 samples are taken. If one atom group is of order sp^3 and the other is of order sp^2 , then about 36 samples are taken. The number and position of the points sampled may be controlled by parameters that may be specified at execution time.

In the example representative embodiment, the sp^3 - sp^3 combination, may be sampled at some major points, such as maybe three points, of approximately 60, 180, and 300 degrees, because these are low-energy areas for sp^3 - sp^3 configuration. Samples may be taken at major points about 0, 120, and 240 degrees. In certain crowded molecules, an eclipsed configuration in one sp^3 - sp^3 pair may result in a lower overall energy level for the molecule taken as a whole. In addition, samples can also be taken at minor points located anywhere from about -10 degrees to about +10 degrees about each major point. This

additional sampling captures conformers that would not be of lowest energy if the bond were considered alone, but may be of lower energy due to the interactions of other atoms in the fragment. Consequently, for the sp^3-sp^3 combination, samples may be taken in groups of points. For example, three groups of three points each, may be taken at approximately 50, 60, and 70 degrees; 170, 180, and 190 degrees; 290, 300, and 310 degrees; 350, 0, and 10 degrees; 110, 120, and 130 degrees and 230, 240, and 250 degrees, for a total of 18 samples. Other sampling regimes employing differing number or points of samples may be used without departing from the scope of the invention.

Further, in the example representative embodiment, the sp^2-sp^3 combination may be sampled at some major points, such as maybe six points at approximately 0, 60, 120, 180, 240 and 300 degrees. Points may be selected because the point corresponds to a low-energy area for an sp^2-sp^3 configuration. Samples can also be taken at major points of about 30, 90, 150, 210, 270, and 330 degrees. In certain crowded molecules, these values may result in a lower overall energy in the context of the molecule. In addition, samples can be taken at minor points located anywhere from about -10 degrees to about +10 degrees about each major point. This additional sampling captures conformers that would not be of lowest energy if the bond were considered alone, but may be of lower energy due to the interactions of other atoms in the fragment. For example, in the sp^2-sp^3 combination, samples may be taken in 12 groups of three points each. This sampling regime provides for sampling at approximately 350, 0, and 10 degrees; 20, 30, and 40 degrees; 50, 60, and 70 degrees; 80, 90, and 100 degrees; 110, 120, and 130 degrees; 140, 150, and 160 degrees; 170, 180, and 190 degrees; 200, 210, and 220 degrees; 230, 240, and 250 degrees; 260, 270, and 280 degrees; 290, 300, and 310 degrees and 320, 330, and 350 degrees, for a total of 36 samples. Other sampling regimes employing differing number or points of samples may be used without departing from the scope of the invention.

Further, in the example representative embodiment, the sp^2-sp^2 combination joined by a single bond, may be sampled at some major points, such as maybe eight major points at approximately 0, 60, 120, 180, 240, and 300 degrees. These points are selected because they are low-energy areas for an sp^2-sp^2 single bond configuration. Samples can also be taken at major points around 90 and 270 degrees. In certain crowded molecules, these values may result in a lower overall energy in the context of a molecule. In addition, samples can be taken at oh say 16 minor points located anywhere from about -10 degrees to about +10 degrees about each major point. For example, samples may be taken in 8 groups of three points each. Accordingly, samples could be taken at approximately 350, 0 and 10 degrees; 170, 180, and 190 degrees; 50, 60, and 70 degrees; 110, 120, and 130 degrees; 230, 240, and 250 degrees; 290, 300, and 310 degrees, 80, 90, and 100 degrees; and 260, 270, and 280 degrees, for a total of 24 samples. Other sampling regimes employing differing number or points of samples may be used without departing from the scope of the invention.

Further, in the example representative embodiment, the sp^2-sp^2 combination joined by a double bond, may be sampled at some major points, such as 2 major points at approximately 0 and 180 degrees. These points are selected because they are low-energy areas for sp^2-sp^2 double bond configuration. In addition, samples can be taken at 4 minor points located anywhere from about -10 degrees to about +10 degrees about each major point. For example, samples may be taken in 2 groups of 3 points each.

Accordingly, samples could be taken at approximately 350, 0, and 10 degrees; and at 170, 180, and 190 degrees, for a total of 6 samples. Other sampling regimes employing differing number or points of samples may be used without departing from the scope of the invention.

Possible fragment conformers having bonds rotated in the amount indicated are calculated.

- 5 If the calculated fragment conformer is of low energy, the number of the major points used to obtain the conformer is saved in a table. The major point is saved, even if the low-energy conformer was calculated at a minor point about the major point. If the calculated fragment conformer is of high energy, the conformer is considered unstable and not likely to be a part of a preferred conformer for the molecule, and is discarded. The conformer is considered to be low-energy if it is within N kcals of the global minima; that is, within N
10 kcals of the lowest energy conformation calculated so far for the fragment. N is obtained by multiplying the number of rotatable bonds by a scaling factor, which may range anywhere from 0 to 2.0 kcals/bond. In a current preferable embodiment, N is generally 0.5 kcals/bond. The scaling factor can be higher (e.g., 1.5 kcals/bond) when it is desirable to cover the conformational space more completely.

- Once the fragment conformer set has been calculated, it is stored in the database. This
15 makes the conformer set available for future computations that require determination of a conformer set for a fragment derived from a similar molecule, and may also be used in future executions for a totally different molecule, provided that the totally different molecule contains a fragment matching the canonical form of the stored fragment. For example, to determine a conformer set for fragment 803, the previously stored conformer set for fragment 801 may be used, bypassing the repeated calculation of what would be an
20 identical conformer set.

- Fig. 9 depicts a graph 900 of a hypothetical molecule being broken into overlapping fragments 901, 902, and 903. Molecule 900 is of atom length 9, with seven nodes shown and omitting representation for additional hydrogen atoms bound to atoms represented by node 930 and node 942 (not shown). Graph 900 comprises seven nodes 930, 932, 934, 936, 938, 940, and 942, and six edges 910, 912,
25 914, 916, 918, and 920. Fragment 901 comprises five nodes 930, 932, 934, 936, and 938, and four edges 910, 912, 914, and 916. Fragment 902 comprises five nodes 932, 934, 936, 938, and 940, and four edges 912, 914, 916, and 918. Fragment 903 comprises five nodes 934, 936, 938, 940 and 942, and four edges 914, 916, 918, and 920.

- Fig. 10 depicts conformation sets for each of fragments 901, 902, and 903. Tables 1001, 1002, and 1003 each depict a hypothetical conformation set for fragments 901, 902, and 903, respectively. Each table comprises a plurality of columns, each of which corresponds to an edge in the graph fragment, which in turn corresponds to a rotatable bond in the atom being conformationally analyzed. Thus, table 1001, representing a conformation set for fragment 901, comprises columns 1010-1, 1012-1, 1014-1, and 1016-1, corresponding to edges 910, 912, 914, and 916, respectively. Likewise, table 1002, representing a
35 conformation set for fragment 902, comprises columns 1012-2, 1014-2, 1016-2, and 1018-2, corresponding to edges 912, 914, 916, and 918, respectively. Finally, table 1003, representing a conformation set for fragment 903, comprises columns 1014-3, 1016-3, 1018-3, and 1020-3, corresponding to edges 914, 916, 918, and 920, respectively.

Each row in the table represents a conformation for the fragment. For example, table 1001 contains nine rows, indicating nine possible low-energy conformations for fragment 901. The first row in table 1001 contains the sequence 3-0-8-11, which corresponds to a particular possible low energy conformation for fragment 901. This particular sequence indicates a low-energy conformation where the bond corresponding to edge 910 has a torsion measured at point 3, the bond corresponding to edge 912 has a torsion measured at point 0, the bond corresponding to edge 914 has a torsion measured at point 8, and the bond corresponding to edge 916 has a torsion measured at point 11. The particular degree value for the torsion will depend on the nature of the bond (i.e., sp^2-sp^2 , sp^2-sp^3 or sp^3-sp^3).

The value of the tabular form will become apparent in the discussion of fragment intersection below.

4.3 Fragment Intersection

Conformation sets corresponding to two overlapping fragments may be intersected to produce a conformation set corresponding to a larger fragment comprising both fragments. This process may be iterated or combined until the conformation sets for all fragments in the molecule have been intersected. The resulting conformation set represents a conformation set for the entire molecule.

Fig. 11 depicts the intersection of conformation sets of Fig. 10. Table 1101 represents the intersection of table 1001 and table 1002. Column 1012-1 in table 1001 represents a torsion value for the bond corresponding to edge 912 in a conformer for fragment 901. Similarly, column 1012-2 in 1002 represents a torsion value for the bond corresponding to edge 912 in a conformer for fragment 902. A similar relationship is found between column 1014-1 in table 1001 and 1014-2 in table 1002, and between column 1016-1 in table 1001 and 1016-2 in table 1002. That is, each column pair represents a column of torsion values for a particular edge common to both fragments. Columns 1014-1 and 1014-2 correspond to edge 914, and columns 1016-1 and 1016-2 correspond to edge 916.

If a conformer row in table 1001 is found having values for columns 1012-1, 1014-1, and 1016-1 equal to the values for columns 1012-2, 1014-2, and 1016-2 for a conformer row in table 1002, then the corresponding conformers for the two fragments are identical for the bonds shared by the two fragments. Therefore, a new row may be constructed, comprising the shared values and one value from each selected row. The result will be a row whose order is one greater than the order of the rows from which it is derived. For example, if four-column table 1001 has three columns in common with four-column table 1002, the intersection of the table will have five columns: the three common columns and one additional column from each table.

The operation is clearly shown with an example. By inspection of table 1001, it will be seen that row 1040 has values of 3, 4, and 5 for columns 1012-1, 1014-1 and 1016-1, respectively. Likewise, by inspection of table 1002, it will be seen that row 1050 has values 3, 4, and 5 for columns 1012-2, 1014-2, and 1016-2, respectively. Accordingly, an intersection of the two rows may be constructed, having a values of 11, 3, 4, 5, and 10. The value 11 counts from unshared column 1010-1 in row 1040. The values 3, 4, 5 come from common columns 1012-1, and 1012-2, 1014-1, and 1014-2, and 1016-1 and 1016-2 for rows 1040 and 1050. The value 10 comes from unshared column 1018-3 in row 1050.

The resulting sequence 11, 3, 4, 5, 10 represents a low energy conformer for a fragment of length 2-8 comprising bonds 910, 912, 914, 916, and 918.

A complete intersection of tables 1001 and 1002 is depicted as table 1101. Row 1120 represents the operation just described. It will be noted that there are three other rows resulting from the intersection of tables 1001 and 1002. Row 1041 has values for 1012-1, 1014-1 and 1016-1 in common with columns 1012-2, 1014-2, and 1016-2 for rows 1051 and 1053, thus resulting in rows 1121 and 1122 in table 1101. Similarly, rows 1042 and 1054 intersect to form row 1123.

Table 1101, which now describes low energy conformers for an 8 length fragment comprising bonds 910 through 918 may be further intersected with table 1003, which describes low-energy conformers for 7 length fragment 903 comprising bonds 914 through 920. The two tables have three bonds in common. Columns 1014-3, 1016-3, and 1018-3 in table 1003 and columns 1114-12, 1116-12, and 1118-12 in table 1101 each correspond to edges 914, 916, and 918, respectively. Therefore, the two tables may be intersected to form a third table 1102. Table 1102 describes a fragment of length 9 comprising bonds 910, 912, 914, 916, 918, and 920. Row 1122 may be intersected with row 1060 (common values 4, 9, and 2) to form row 1140; row 1123 may be intersected with row 1061 (common values 10, 11, and 8) to form row 1141.

As noted above, table 1102 describes a fragment of length 9 comprising bonds 910, 912, 914, 916, 918, and 920. Referring to Fig. 9, it will be seen that this fragment actually is equivalent to the entire 9 molecule. Thus, with the intersection to form table 1102, conformational analysis of the entire table is complete. That is, table 1102 comprises a list of the preferred conformations for the molecule being analyzed.

It will be noted that the intersection of the tables may be efficiently performed as a relational join of tables in a relational database. In addition, it will be noted that the values in each column are restricted to a range of 0 through 11. This permits each value to be encoded in a 4-bit hexadecimal digit (a half-byte, or "nibble"). The advantage of such encoding is that up to eight values in common may be loaded into a single 32-bit (4-byte) word, as is common to many commonly available inexpensive computers. Consequently, this approach allows for very rapid comparison of up to eight column-rows in a single computer instruction. Although the comparison may be by a COMPARE instruction or the like, in many architectures, a bit-wise comparison may be performed for greater speed. For example, a 32-bit word containing up to eight column-rows from one table may be combined with a 32-bit word containing the corresponding column-rows from another table by means of an exclusive OR operation. A zero result indicates that all column-row values correspond, and a non-zero result indicates that at least one column-row does not correspond. Since exact identification of a non-corresponding set of column-rows is not necessary to the operation of the invention, this approach allows for extremely rapid evaluation of multiple columns to determine whether a particular selection of rows is found to intersect.

4.4 Ring Structures

In an alternative embodiment, a method by which ring structures contained within molecular structures can be analyzed will now be described.

Figure 12 depicts a sample molecule 1201 containing a ring structure. The ring structure comprises nodes (representing atoms) 1210, 1212, 1214, 1216 and 1218 and edges (representing bonds) 1220, 1222, 1224, 1226 and 1228. In addition, nodes 1230, 1232, 1234, 1238, and 1240 and edges 1250, 1252, 1254, 1256, 1258, and 1260 are said to be outside the ring structure.

For purposes of decomposition, the ring may be considered a node capable of multiple conformations. In this way, the conformational variation of ring systems can be explored. Figure 13 depicts the demarcation of a ring as a ring node 1310 within molecule 1201. Figure 14 depicts some of the fragments that are formed during decomposition of molecule 1201. Fragments 1410 and 1420 are ordinary linear fragments not containing a ring structure. It will be noted that fragment 1410 comprises one node 1214 within the ring structure, and that fragment 1420 comprises two nodes 1212 and 1214 and one edge 1222 within the ring structure, and that neither of these conditions requires special treatment.

Where paths walk into a ring, special handling is required if any rotatable bond in the path is also a bond in the ring. In this case, a separate algorithm is used to prune the conformations of the path by the conformations of the ring structure. Only conformations of the path that have torsions compatible with torsions in the ring are retained. As described previously, paths are only permitted to walk a maximum of two bonds into a ring, so at most, two torsions from the ring can be used to prune the conformations of the paths.

Figure 15 depicts tables used in conformational analysis and intersection of ring nodes. Table 1501 represents a ring node and the conformations that it may assume. Table 1501 contains one row for each potential low-energy conformation that the ring structure, taken alone, could assume. Columns 1510 represent the points corresponding to angles of torsion, just as the columns in fragment conformation sets such as 1001, 1002, and 1003 do. In addition, table 1501 comprises an additional column 1520 that identifies a conformation number associated with the low-energy conformation of the ring.

Table 1502 is a fragment conformation table similar to fragment conformation tables 1001, 1002 and 1003. Table 1502 may be intersected with table 1501 to produce table 1503. Row 1530 in table 1501 contains point values 5, 6, and 4. Likewise, row 1532 in table 1502 contains the same point values 5, 6, and 4. In this example, it is assumed that the point values 5, 6, and 4 in both rows each describe the same edges corresponding to the same molecular bonds. Therefore, the two rows may be intersected to form row 1534 in table 1503.

In such an intersection, the resulting table includes not only the point values from both rows that are pertinent to the new fragment, but also the conformation number of the ring conformation in which the matching values were found. Since the ring is capable of assuming multiple conformations, the inclusion of the conformation number prevents row 1503 from being improperly intersected with a row from another table describing another fragment, where that other table may have the same point values from the ring structure, but from a different conformation.

4.5 Torsional Analysis for Resolving Strain in Conformers

Fig. 16 depicts an example of a molecular structure, 1600, which is computationally decomposed into a plurality of fragments, 1601, 1602, 1603, 1604, 1605 and 1606. The notation F(A1, A2,

...An) is used to represent a fragment derived from a linear atom sequence A1 through An. Each normalized fragment derived from the molecular structure is conformationally analyzed in isolation.

Conformational models of each fragment comprise the torsions of low-energy conformations of the fragments. In a particular embodiment, these conformational models may be represented as a table, where each row represents a conformer of the fragment and each column represents an internal coordinate in the fragment. Here, we use torsions as the internal coordinates. In these tables, central torsions (i.e. the torsions defined by the atoms in the defining path of the fragment) are listed. A global minimum energy value is computed for each conformer for each fragment and stored with the conformer in the table. In a current preferred embodiment, conformers are sorted by relative energy above the global minimum for the fragment. An energy cut-off may be employed in the generation of these fragment conformational models in order to control the number of conformers retained. Conformational energies are measured using a standard molecular mechanics force field which estimates the energy of the conformer as a function of its internal coordinates. In a current embodiment, the DREIDING force field is used in its entirety. For further information regarding the DREIDING technique, reference may be had to Mayo, Olafson and Goddard, J.Phys.Chem (1990) 94, pp8897-8909 which is incorporated herein by reference for all purposes.

A conformational model of the entire molecular structure is constructed by intersecting the torsion tables of each of the fragments. Table 1 - table 6 in Figs. 17 - 19 list the conformational models of the example fragments 1601, 1602, 1603, 1604, 1605 and 1606 shown in Fig. 16.

Certain torsions may be eliminated from consideration because these torsions provide no information about the conformation. For example, torsions that lie within a ring structure have a structure that is completely determined by the conformation of the ring. For example, consider torsions 13-12-11-16 and 14-13-12-11 in Fig. 16. These correspond to fragments that lie exclusively within the phenyl ring of the example, which comprises atoms 11-12-13-14-15-16. Given the complete list of conformations of the fragment F(2,1,11,12,13) in table 1 of Fig. 17, conformers that do not have torsions consistent with the phenyl ring can be eliminated.

In a particular embodiment, a mechanism of a Fragment-Torsion-Value table ("FTV table") and a subsequent Torsion-Value (T-V table) is employed in order to resolve information about torsion angle strains among the conformers. Fig. 21 depicts a simplified representative FTV table, table 9, having rows, such as a row 2101, representing "variable" torsions and columns, such as a column 2102, indexed by characteristic torsion angle values for each torsion. An FTV table may be constructed incrementally by considering each conformational model of the constituent fragments in sequence. As used herein, a "variable torsion" is any exocyclic torsion crossed by a fragment. In certain embodiments, the FTV table is three dimensional, having a typical entry represented by FTV(f,b,v), wherein f is a particular fragment, b is a particular variable torsion and v is the value of a particular torsion. A separate FTV table can be provided for each fragment, in which each entry is the incremental energy above a particular nominal value, such as a global minimum energy for the fragment, of the first conformer that has the torsion b at the value v. The entry represents the way that this fragment can place a particular torsion at a particular value at the lowest incremental energy.

The quantity $TV(v,b)$ is the maximum of all FTV table entries, $FTV(f,v,b)$, for all fragments f . It is the least strained way of achieving the torsion b at the value v over all fragments f . The simplified FTV table, table 9 of Fig. 21, has entries for an example torsion 12-11-1-2. For every fragment in which this torsion appears, the incremental energy above the global minimum of the first conformer that has this torsion at a particular value v is set in the FTV table. In row 2101 of table 9, the first conformer of the fragment 1601 in table 1, fragment $F(2,1,11,12,13)$, that sets the torsion value to 30 degrees for torsion 12-11-1-2 is 0.9 kcals above the global minimum for the fragment. This may be represented as, when $f = F(2,1,11,12,13)$, $b = \text{torsion 12-11-1-2}$, and $v = 30 \text{ degrees}$, $FTV(f,b,v) = 0.9$. When all such entries are determined for every bond and every value for every fragment, the TV entries for every torsion value may be determined by taking the maximum of each column in table 9.

The TV table entries for any bond, after considering all fragments that cross the bond, express the lowest energy which will be expended in order for any fragment to set the torsion to a particular characteristic angle value. For example, row 2103 of table 9 predicts that torsion 12-11-1-2 prefers values of 30 degrees and 150 degrees and has a particularly large incremental energy when the torsion angle is 0 degrees. At least one fragment, $F(6,5,1,11,16)$, in the molecule possessed the requisite information about a high-energy H-H interaction between the phenyl rings when this torsion is planar. It is this fragment that dominates the entries in the TV table when torsion 12-11-1-2 is planar.

4.6 Minimizing Energy of a Conformer's Bonds

The quantity $TV(v,b)$, is the maximum incremental energy required for the first conformer to have torsion b at value v , over all fragments that contain torsion b . For all values v for a particular torsion b , there is one TV entry for every such value v . The $TV(v,b)$ quantities may be derived from the $FTV(f,v,b)$ entries. An example of TV entries for torsion 12-11-1-2 are contained in row 2103 of table 9.

Since each $TV(v,b)$ entry represents the lowest energy state in which some fragment had a torsion b at value v , it is not worth considering values v that require too large an incremental energy above the global minimum for the row. Row 2103 of table 9 indicates that for torsion 12-11-1-2 to have a value of 0 degrees, some fragment's incremental energy must rise to at least 7.1 kcals above its own global minimum. However, to achieve a value of 30 degrees or 150 degrees, the largest incremental energy that any fragment must have is 2.4 kcals and 3.0 kcals respectively.

Sorting the FTV table by incremental energy can facilitate the selecting of TV values for torsions likely to appear in the final conformational model. All torsion values within a nominal energy threshold can be retained, while those exceeding the threshold are discarded. A current preferable threshold value is 0.4 kcals per rotatable bond in the fragments. In select embodiments, the threshold may be controlled by a user. In the example of table 9, torsions having characteristic angle values of 30 degrees and 150 degrees will be retained. However, the characteristic angle value of 0 degrees is correctly eliminated, because at least one fragment possessed sufficient contextual information to determine that a value of 0 degrees was a poor choice for torsion 12-11-1-2.

The TV table processing limits the characteristic angle values that any torsion can assume. In our example, any conformer of a fragment that does not have a value surviving the cutoff analysis is no

longer considered. This is illustrated in table 10, where table 1 and table 5 are intersected at an energy cutoff (for each individual table) of 2.0 kcals. From conformer 1701 of table 1, it can be seen that lowest energy conformation has torsion 12-11-1-2 set to a value of 0 degrees. However, this value was eliminated during the TV table processing, so conformer 1701 is eliminated from further consideration. The new global
5 minimum energy conformer 1702 of table 1 is the second conformer in the table. Conformer 1702 and conformer 1703 are included in subsequent join processing. The result of the intersection of these conformers is depicted in table 11.

5.0 Conclusion

10 In conclusion the present invention provides for a method for dynamically determining conformers for a synthesizable molecular structure based upon information about its constituents. In some embodiments, the present invention provides more automated methods of analyzing molecular structures using a computer than many of the manual techniques heretofore known. The present invention can also provide a library of fragments which is without theoretical limit as to size or scope. Some embodiments
15 according to the invention can consider entire molecular contexts in determining a molecular structure based upon fragment conformers. Select embodiments according to the invention may be more robust than those heretofore known in the art.

20 Other embodiments of the present invention and its individual components will become readily apparent to those skilled in the art from the foregoing detailed description. As will be realized, the invention is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the spirit and the scope of the present invention. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive. It is therefore not intended that the invention be limited except as indicated by the appended claims.

WHAT IS CLAIMED IS:

1. A computer based method for determining a conformation for a molecular
 5 structure comprising the steps:

automatically decomposing said molecular structure into a plurality of fragments, including
 a first fragment and a second fragment, such that at least one of the fragments is not an amino acid;

normalizing each of said plurality of fragments to form a plurality of normalized fragments
 including a first normalized fragment corresponding to said first fragment and a second normalized fragment
 10 corresponding to said second fragment;

determining for each normalized fragment in said plurality of normalized fragments, at
 least one of a plurality of conformers including a first conformer corresponding to said first normalized
 fragment and a second conformer corresponding to said second normalized fragment, said first conformer
 having at least one of a first plurality of internal coordinates, said second conformer having at least one of a
 15 second plurality of internal coordinates; and

combining said at least one internal coordinate from said first plurality of internal
 coordinates of said first conformer and said at least one internal coordinate from said second plurality of
 internal coordinates of said second conformer to derive said conformation for said molecular structure.

20 2. The method of claim 1 wherein the determining step further comprises:
 searching in a library of normalized fragments for at least one matching fragment, said
 matching fragment being identical to at least one normalized fragment in said plurality of normalized
 fragments; thereupon,

if said matching fragment is found, using conformer information associated with said
 25 matching fragment as said at least one of said plurality of conformers, otherwise, computing said at least one
 of said plurality of conformers of the normalized fragment and storing said at least one of said plurality of
 fragments in said library.

30 3. The method of claim 1 wherein the combining step further comprises:
 relationally joining said first conformer and said second conformer.

4. The method of claim 3 wherein the automatically decomposing step further
 comprises:

35 automatically determining said first fragment and said second fragment to overlap
 maximally, wherein said at least one of said first plurality of internal coordinates is not contained in said
 second plurality of internal coordinates, or said at least one of said second plurality of internal coordinates is
 not contained in said first plurality of internal coordinates.



5. The method of claim 1 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

5 6. The method of claim 1 wherein fragments are represented by a plurality of nodes and edges.

7. The method of claim 6 wherein each of said plurality of nodes represents a collection of atoms in a molecule.

10 8. The method of claim 6 wherein each of said plurality of edges represents a molecular bond.

15 9. The method of claim 8 wherein said automatically decomposing step further comprises the step of:

enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

20 10. The method of claim 9 wherein M is 2 and P is 7.

11. A computer programming product for determining a conformation for a molecular structure comprising:

25 code for automatically decomposing said molecular structure into a plurality of fragments, including a first fragment and a second fragment, such that at least one of the fragments is not an amino acid; code for normalizing each of said plurality of fragments to form a plurality of normalized fragments including a first normalized fragment corresponding to said first fragment and a second normalized fragment corresponding to said second fragment;

30 code for determining for each normalized fragment in said plurality of normalized fragments, at least one of a plurality of conformers including a first conformer corresponding to said first normalized fragment and a second conformer corresponding to said second normalized fragment, said first conformer having at least one of a first plurality of internal coordinates, said second conformer having at least one of a second plurality of internal coordinates;

35 code for combining said at least one internal coordinate from said first plurality of internal coordinates of said first conformer and said least one internal coordinate from said second plurality of internal coordinates of said second conformer to derive said conformation for said molecular structure; and a computer readable storage medium for storing the codes.

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED



12. The computer programming product of claim 11 wherein the code for determining further comprises:

code for searching in a library of normalized fragments for at least one matching fragment, said matching fragment being identical to at least one normalized fragment in said plurality of normalized fragments; thereupon,

if a matching fragment is found, using conformer information associated with said matching fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for said at least one of said plurality of conformers and storing said conformer information for at least one of said plurality of fragments in said library.

13. The computer programming product of claim 11 wherein the code for combining further comprises:

code for relationally joining said first conformer and said second conformer.

14. The computer programming product of claim 13 wherein the code for automatically decomposing further comprises:

code for automatically determining said first fragment and said second fragment to overlap maximally, wherein said at least one of said first plurality of internal coordinates is not contained in said second plurality of internal coordinates, or said at least one of said second plurality of internal coordinates is not contained in said first plurality of internal coordinates.

15. The computer programming product of claim 11 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

16. The computer programming product of claim 11 wherein fragments are represented by a plurality of nodes and edges.

17. The computer programming product of claim 16 wherein each of said plurality of nodes represents a collection of atoms in a molecule.

18. The computer programming product of claim 16 wherein each of said plurality of edges represents a molecular bond.

19. The computer programming product of claim 18 wherein said code for automatically decomposing further comprises:

code for enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

20. The computer programming product of claim 19 wherein M is 2 and P is 7.

- 5 21. An apparatus for determining a conformation for a molecular structure comprising:
a processor operatively disposed to perform steps comprising:
automatically decomposing said molecular structure into a plurality of fragments, including
a first fragment and a second fragment, such that at least one of the fragments is not an amino acid;
normalizing each of said plurality of fragments to form a plurality of normalized fragments
10 including a first normalized fragment corresponding to said first fragment and a second normalized fragment
corresponding to said second fragment;
determining for each normalized fragment in said plurality of normalized fragments, at
least one of a plurality of conformers including a first conformer corresponding to said first normalized
fragment and a second conformer corresponding to said second normalized fragment, said first conformer
15 having at least one of a first plurality of internal coordinates, said second conformer having at least one of a
second plurality of internal coordinates; and
combining said at least one internal coordinate from said first plurality of internal
coordinates of said first conformer and said at least one internal coordinate from said second plurality of
internal coordinates of said second conformer to derive said conformation for said molecular structure.
- 20 22. The apparatus of claim 21 wherein the determining step further comprises:
searching in a library of normalized fragments for at least one matching fragment, said
matching fragment being identical to at least one normalized fragment in said plurality of normalized
fragments; thereupon,
25 if a matching fragment is found, using conformer information associated with said matching
fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for
said at least one of said plurality of conformers and storing said conformer information for at least one of
said plurality of fragments in said library.
- 30 23. The apparatus of claim 21 wherein the combining step further comprises:
relationally joining said first conformer and said second conformer.
24. The apparatus of claim 23 wherein the automatically decomposing step further
comprises:
35 automatically determining said first fragment and said second fragment to overlap
maximally, wherein said at least one of said first plurality of internal coordinates is not contained in said
second plurality of internal coordinates, or said at least one of said second plurality of internal coordinates is
not contained in said first plurality of internal coordinates.



25. The apparatus of claim 21 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

5 26. The apparatus of claim 21 wherein fragments are represented by a plurality of nodes and edges.

27. The apparatus of claim 26 wherein each of said plurality of nodes represents a collection of atoms in a molecule.

10 28. The apparatus of claim 26 wherein each of said plurality of edges represents a molecular bond.

15 29. The apparatus of claim 28 wherein said automatically decomposing step further comprises the step of:

enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

20 30. The apparatus of claim 29 wherein M is 2 and P is 7.

31. An apparatus for determining a conformation for a molecular structure comprising:
means for automatically decomposing said molecular structure into a plurality of fragments,
including a first fragment and a second fragment, such that at least one of the fragments is not an amino acid;

25 means for normalizing each of said plurality of fragments to form a plurality of normalized fragments including a first normalized fragment corresponding to said first fragment and a second normalized fragment corresponding to said second fragment;

means for determining for each normalized fragment in said plurality of normalized fragments, at least one of a plurality of conformers including a first conformer corresponding to said first normalized fragment and a second conformer corresponding to said second normalized fragment, said first conformer having at least one of a first plurality of internal coordinates, said second conformer having at least one of a second plurality of internal coordinates; and

30 means for combining said at least one internal coordinate from said first plurality of internal coordinates of said first conformer and said at least one internal coordinate from said second plurality of internal coordinates of said second conformer to derive said conformation for said molecular structure.

32. The apparatus of claim 31 wherein the means for determining further comprises:



means for searching in a library of normalized fragments for at least one matching fragment, said matching fragment being identical to at least one normalized fragment in said plurality of normalized fragments; thereupon,

if a matching fragment is found, using conformer information associated with said matching fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for said at least one of said plurality of conformers and storing said conformer information for at least one of said plurality of fragments in said library.

33. The apparatus of claim 31 wherein the means for combining further comprises:
means for relationally joining said first conformer and said second conformer.

34. The apparatus of claim 33 wherein the means for automatically decomposing further comprises:

means for automatically determining said first fragment and said second fragment to overlap maximally, wherein said at least one of said first plurality of internal coordinates is not contained in said second plurality of internal coordinates, or said at least one of said second plurality of internal coordinates is not contained in said first plurality of internal coordinates.

35. The apparatus of claim 31 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

36. The apparatus of claim 31 wherein fragments are represented by a plurality of nodes and edges.

37. The apparatus of claim 36 wherein each of said plurality of nodes represents a collection of atoms in a molecule.

38. The apparatus of claim 36 wherein each of said plurality of edges represents a molecular bond.

39. The apparatus of claim 38 wherein said means for automatically decomposing further comprises:

means for enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

40. The apparatus of claim 39 wherein M is 2 and P is 7.



41. A molecule having a corresponding molecular structure derived using a computer based method, said computer based method substantially similar to the method of claim 1.

42. The molecule of claim 41 wherein the determining step further comprises:
5 searching in a library of normalized fragments for at least one matching fragment, said matching fragment being identical to at least one normalized fragment in said plurality of normalized fragments; thereupon,

10 if a matching fragment is found, using conformer information associated with said matching fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for said at least one of said plurality of conformers and storing said conformer information for at least one of said plurality of fragments in said library.

43. The molecule of claim 41 wherein the combining step further comprises:
15 relationally joining said first conformer and said second conformer.

44. The molecule of claim 43 wherein the automatically decomposing step further comprises:

20 automatically determining said first fragment and said second fragment to overlap maximally, wherein said at least one of said first plurality of internal coordinates is not contained in said second plurality of internal coordinates, or said at least one of said second plurality of internal coordinates is not contained in said first plurality of internal coordinates.

45. The molecule of claim 41 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized
25 fragment.

46. The molecule of claim 41 wherein fragments are represented by a plurality of nodes and edges.

30 47. The molecule of claim 46 wherein each of said plurality of nodes represents a collection of atoms in a molecule.

48. The molecule of claim 46 wherein each of said plurality of edges represents a molecular bond.

35 49. The molecule of claim 48 wherein said automatically decomposing step further comprises the step of:



automatically enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

5 50. The molecule of claim 49 wherein M is 2 and P is 7.

51. A computer based method for determining a conformation for a molecular structure, wherein said conformation is one of a plurality of low energy conformations of said molecular structure, said method comprising the steps:

10 decomposing said molecular structure into a plurality of fragments, including a first fragment and a second fragment;

 normalizing each of said plurality of fragments to form a plurality of normalized fragments including a first normalized fragment corresponding to said first fragment and a second normalized fragment corresponding to said second fragment;

15 determining for each normalized fragment in said plurality of normalized fragments, at least one of a plurality of conformers including a first conformer corresponding to said first normalized fragment and a second conformer corresponding to said second normalized fragment, wherein each conformer in said plurality of conformers has at least one of a plurality of chemical bond representations, each chemical bond representation interconnecting at least two atoms in said conformer;

20 associating with each conformer a plurality of internal coordinates, each of said internal coordinates having a plurality of characteristic values corresponding to each chemical bond representation in said conformer;

 determining, for each conformer, an energy difference level, said energy difference level representing an incremental amount of potential energy above a first nominal energy value, said incremental amount of potential energy required to maintain each chemical bond representation at each of said plurality of characteristic values;

25 for each internal coordinate, for each characteristic value in said plurality of characteristic values, selecting a conformer having a corresponding minimum energy difference level, said corresponding minimum energy difference level selected from said energy difference level computed for each conformer, to form a plurality of corresponding minimum energy difference levels;

30 for each characteristic value, determining a maximum energy difference value from among said plurality of corresponding minimum energy difference levels, to form a plurality of maximum energy difference values;

35 for each internal coordinate, selecting a plurality of likely values from said plurality of characteristic values, wherein said plurality of likely values in said internal coordinate correspond to a second nominal energy value selected from said plurality of maximum energy difference levels;

 selecting a plurality of candidate conformers from said plurality of conformers, each candidate conformer in said plurality of candidate conformers having at least one internal coordinate with a corresponding characteristic value in said plurality of likely values; and



combining at least two candidate conformers chosen from said plurality of candidate conformers to produce a conformation for said molecular structure, said candidate conformers chosen such that the energy difference level of each of said at least two candidate conformers is less than a specified cutoff energy level, said at least two candidate conformers further chosen from said plurality of candidate conformers such that said at least two conformers have at least one common internal coordinate.

52. The method of claim 51 wherein the determining for each normalized fragment step further comprises:

searching in a library of normalized fragments for at least one matching fragment, said matching fragment being identical to at least one normalized fragment in said plurality of normalized fragments; thereupon,

if a matching fragment is found, using conformer information associated with said matching fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for said at least one of said plurality of conformers and storing said conformer information for at least one of said plurality of fragments in said library.

53. The method of claim 51 wherein the combining at least two candidate conformers step further comprises:

relationally joining a first candidate conformer and a second candidate conformer chosen from said plurality of candidate conformers.

54. The method of claim 53 wherein the automatically decomposing step further comprises:

automatically determining said first fragment and said second fragment to overlap maximally, wherein said at least one of said first plurality of internal coordinates is not contained in said second plurality of internal coordinates, or said at least one of said second plurality of internal coordinates is not contained in said first plurality of internal coordinates.

55. The method of claim 51 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

56. The method of claim 51 wherein fragments are represented by a plurality of nodes and edges.

57. The method of claim 56 wherein each of said plurality of nodes represents a collection of atoms in a molecule.



58. The method of claim 56 wherein each of said plurality of edges represents a molecular bond.

59. The method of claim 58 wherein said decomposing step further comprises the step of:

automatically enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

60. The method of claim 59 wherein M is 2 and P is 7.

61. The method of claim 51 wherein said first nominal energy value is a global minimum energy value for said fragment, said method further comprising the step of:
automatically determining said global minimum energy value for said fragment.

62. The method of claim 61 wherein said automatically determining said global minimum energy value for said fragment further comprises:

determining a quantity of rotatable bonds in said fragment; and
multiplying said quantity by a per bond energy factor.

63. The method of claim 62 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

64. The method of claim 51 wherein said second nominal energy value is an energy threshold value for said fragment, said method further comprising the step of:
automatically determining said energy threshold value for said fragment.

65. The method of claim 64 wherein said automatically determining said energy threshold value for said fragment further comprises:

determining a quantity of rotatable bonds in said fragment; and
multiplying said quantity by a per bond cut-off energy.

66. The method of claim 65 wherein said per bond cut-off energy is 0.4 kcals per rotatable bond.

67. A computer programming product for determining a conformation for a molecular structure, wherein said conformation is one of a plurality of low energy conformations of said molecular structure, said computer programming product comprising:



code for decomposing said molecular structure into a plurality of fragments, including a first fragment and a second fragment;

code for normalizing each of said plurality of fragments to form a plurality of normalized fragments including a first normalized fragment corresponding to said first fragment and a second normalized fragment corresponding to said second fragment;

code for determining for each normalized fragment in said plurality of normalized fragments, at least one of a plurality of conformers including a first conformer corresponding to said first normalized fragment and a second conformer corresponding to said second normalized fragment, wherein each conformer in said plurality of conformers has at least one of a plurality of chemical bond representations, each chemical bond representation interconnecting at least two atoms in said conformer;

code for associating with each conformer a plurality of internal coordinates, each of said internal coordinates having a plurality of characteristic values corresponding to each chemical bond representation in said conformer;

code for determining, for each conformer, an energy difference level, said energy difference level representing an incremental amount of potential energy above a first nominal energy value, said incremental amount of potential energy required to maintain each chemical bond representation at each of said plurality of characteristic values;

code for selecting, for each internal coordinate, for each characteristic value in said plurality of characteristic values, a conformer having a corresponding minimum energy difference level, said corresponding minimum energy difference level selected from said energy difference level computed for each conformer, to form a plurality of corresponding minimum energy difference levels;

code for determining, for each characteristic value, a maximum energy difference value from among said plurality of corresponding minimum energy difference levels, to form a plurality of maximum energy difference values;

code for selecting for each internal coordinate, a plurality of likely values from said plurality of characteristic values, wherein said plurality of likely values in said internal coordinate correspond to a second nominal energy value selected from said plurality of maximum energy difference levels;

code for selecting a plurality of candidate conformers from said plurality of conformers, each candidate conformer in said plurality of candidate conformers having at least one internal coordinate with a corresponding characteristic value in said plurality of likely values;

code for combining at least two candidate conformers chosen from said plurality of candidate conformers to produce a conformation for said molecular structure, said candidate conformers chosen such that the energy difference level of each of said at least two candidate conformers is less than a specified cutoff energy level, said at least two candidate conformers further chosen from said plurality of candidate conformers such that said at least two conformers have at least one common internal coordinate; and,

a computer readable storage medium for storing the codes.



68. The computer programming product of claim 67 wherein the code for determining for each normalized fragment further comprises:

code for searching in a library of normalized fragments for at least one matching fragment, said matching fragment being identical to at least one normalized fragment in said plurality of normalized fragments; thereupon,

if a matching fragment is found, using conformer information associated with said matching fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for said at least one of said plurality of conformers and storing said conformer information for at least one of said plurality of fragments in said library.

69. The computer programming product of claim 67 wherein the code for combining at least two candidate conformers further comprises:

code for relationally joining a first candidate conformer and a second candidate conformer chosen from said plurality of candidate conformers.

70. The computer programming product of claim 67 wherein the code for relationally joining further comprises:

code for determining a maximally overlapping molecular structure from said first candidate conformer and said second candidate conformer.

71. The computer programming product of claim 67 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

72. The computer programming product of claim 67 wherein fragments are represented by a plurality of nodes and edges.

73. The computer programming product of claim 72 wherein each of said plurality of nodes represents a collection of atoms in a molecule.

74. The computer programming product of claim 72 wherein each of said plurality of edges represents a molecular bond.

75. The computer programming product of claim 74 wherein said code for automatically decomposing further comprises:

code for automatically enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.



76. The computer programming product of claim 75 wherein M is 2 and P is 7.

77. The computer programming product of claim 67 wherein said first nominal energy value is a global minimum energy value for said fragment, said computer programming product further comprising:

code for automatically determining said global minimum energy value for said fragment.

78. The computer programming product of claim 77 wherein said automatically determining said global minimum energy value for said fragment further comprises:

code for determining a quantity of rotatable bonds in said fragment; and

code for multiplying said quantity by a per bond energy factor.

79. The computer programming product of claim 78 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

80. The computer programming product of claim 67 wherein said second nominal energy value is an energy threshold value for said fragment, said computer programming product further comprising:

code for automatically determining said energy threshold value for said fragment.

81. The computer programming product of claim 80 wherein said automatically determining said energy threshold value for said fragment further comprises:

code for determining a quantity of rotatable bonds in said fragment; and

code for multiplying said quantity by a per bond cut-off energy.

82. The computer programming product of claim 81 wherein said per bond cut-off energy is 0.4 kcals per rotatable bond.

83. An apparatus for determining a conformation for a molecular structure, wherein said conformation is one of a plurality of low energy conformations of said molecular structure, said apparatus comprising:

a processor operatively disposed to perform the steps of:

decomposing said molecular structure into a plurality of fragments, including a first fragment and a second fragment;

normalizing each of said plurality of fragments to form a plurality of normalized fragments including a first normalized fragment corresponding to said first fragment and a second normalized fragment corresponding to said second fragment;

determining for each normalized fragment in said plurality of normalized fragments, at least one of a plurality of conformers including a first conformer corresponding to said first normalized



fragment and a second conformer corresponding to said second normalized fragment, wherein each conformer in said plurality of conformers has at least one of a plurality of chemical bond representations, each chemical bond representation interconnecting at least two atoms in said conformer;

associating with each conformer a plurality of internal coordinates, each of said internal
5 coordinates having a plurality of characteristic values corresponding to each chemical bond representation in said conformer;

determining, for each conformer, an energy difference level, said energy difference level
10 representing an incremental amount of potential energy above a first nominal energy value, said incremental amount of potential energy required to maintain each chemical bond representation at each of said plurality of characteristic values;

for each internal coordinate, for each characteristic value in said plurality of characteristic
values, selecting a conformer having a corresponding minimum energy difference level, said corresponding
minimum energy difference level selected from said energy difference level computed for each conformer, to
form a plurality of corresponding minimum energy difference levels;

15 for each characteristic value, determining a maximum energy difference value from among
said plurality of corresponding minimum energy difference levels, to form a plurality of maximum energy
difference values;

for each internal coordinate, selecting a plurality of likely values from said plurality of
characteristic values, wherein said plurality of likely values in said internal coordinate correspond to a second
20 nominal energy value selected from said plurality of maximum energy difference levels;

selecting a plurality of candidate conformers from said plurality of conformers, each
candidate conformer in said plurality of candidate conformers having at least one internal coordinate with a
corresponding characteristic value in said plurality of likely values; and

25 combining at least two candidate conformers chosen from said plurality of candidate
conformers to produce a conformation for said molecular structure, said candidate conformers chosen such
that the energy difference level of each of said at least two candidate conformers is less than a specified
cutoff energy level, said at least two candidate conformers further chosen from said plurality of candidate
conformers such that said at least two conformers have at least one common internal coordinate.

30 84. The apparatus of claim 83 wherein the determining for each normalized fragment
step further comprises:

searching in a library of normalized fragments for at least one matching fragment, said
matching fragment being identical to at least one normalized fragment in said plurality of normalized
fragments; thereupon,

35 if a matching fragment is found, using conformer information associated with said matching
fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for
said at least one of said plurality of conformers and storing said conformer information for at least one of
said plurality of fragments in said library.



85. The apparatus of claim 83 wherein the combining at least two candidate conformers step further comprises:

relationally joining a first candidate conformer and a second candidate conformer chosen from said plurality of candidate conformers.

5

86. The apparatus of claim 85 wherein the relationally joining step further comprises: determining a maximally overlapping molecular structure from said first candidate conformer and said second candidate conformer.

10

87. The apparatus of claim 83 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

15

88. The apparatus of claim 83 wherein fragments are represented by a plurality of nodes and edges.

89. The apparatus of claim 88 wherein each of said plurality of nodes represents a collection of atoms in a molecule.

20

90. The apparatus of claim 88 wherein each of said plurality of edges represents a molecular bond.

91. The apparatus of claim 90 wherein said decomposing step further comprises the step of:

25

enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

92. The apparatus of claim 91 wherein M is 2 and P is 7.

30

93. The apparatus of claim 83 wherein said first nominal energy value is a global minimum energy value for said fragment, said processor of said apparatus further disposed to perform the steps of:

automatically determining said global minimum energy value for said fragment.

35

94. The apparatus of claim 93 wherein said automatically determining said global minimum energy value for said fragment further comprises:

determining a quantity of rotatable bonds in said fragment; and
multiplying said quantity by a per bond energy factor.



95. The apparatus of claim 94 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

5 96. The apparatus of claim 83 wherein said second nominal energy value is an energy threshold value for said fragment, said processor of said apparatus further disposed to perform the steps of: automatically determining said energy threshold value for said fragment.

10 97. The apparatus of claim 96 wherein said automatically determining said energy threshold value for said fragment further comprises:
determining a quantity of rotatable bonds in said fragment; and
multiplying said quantity by a per bond cut-off energy.

15 98. The apparatus of claim 97 wherein said per bond cut-off energy is 0.4 kcals per rotatable bond.

99. An apparatus for determining a conformation for a molecular structure, wherein said conformation is one of a plurality of low energy conformations of said molecular structure, said apparatus comprising:

20 means for decomposing said molecular structure into a plurality of fragments, including a first fragment and a second fragment;

means for normalizing each of said plurality of fragments to form a plurality of normalized fragments including a first normalized fragment corresponding to said first fragment and a second normalized fragment corresponding to said second fragment;

25 means for determining for each normalized fragment in said plurality of normalized fragments, at least one of a plurality of conformers including a first conformer corresponding to said first normalized fragment and a second conformer corresponding to said second normalized fragment, wherein each conformer in said plurality of conformers has at least one of a plurality of chemical bond representations, each chemical bond representation interconnecting at least two atoms in said conformer;

30 means for associating with each conformer a plurality of internal coordinates, each of said internal coordinates having a plurality of characteristic values corresponding to each chemical bond representation in said conformer;

35 means for determining, for each conformer, an energy difference level, said energy difference level representing an incremental amount of potential energy above a first nominal energy value, said incremental amount of potential energy required to maintain each chemical bond representation at each of said plurality of characteristic values;

means for selecting for each internal coordinate, for each characteristic value in said plurality of characteristic values, a conformer having a corresponding minimum energy difference level, said



corresponding minimum energy difference level selected from said energy difference level computed for each conformer, to form a plurality of corresponding minimum energy difference levels;

means for determining for each characteristic value, a maximum energy difference value from among said plurality of corresponding minimum energy difference levels, to form a plurality of

5 maximum energy difference values;

means for selecting for each internal coordinate, a plurality of likely values from said plurality of characteristic values, wherein said plurality of likely values in said internal coordinate correspond to a second nominal energy value selected from said plurality of maximum energy difference levels;

10 means for selecting a plurality of candidate conformers from said plurality of conformers, each candidate conformer in said plurality of candidate conformers having at least one internal coordinate with a corresponding characteristic value in said plurality of likely values; and

means for combining at least two candidate conformers chosen from said plurality of candidate conformers to produce a conformation for said molecular structure, said candidate conformers chosen such that the energy difference level of each of said at least two candidate conformers is less than a
15 specified cutoff energy level, said at least two candidate conformers further chosen from said plurality of candidate conformers such that said at least two conformers have at least one common internal coordinate.

100. The apparatus of claim 99 wherein the means for determining for each normalized fragment further comprises:

20 means for searching in a library of normalized fragments for at least one matching fragment, said matching fragment being identical to at least one normalized fragment in said plurality of normalized fragments; thereupon,

if a matching fragment is found, using conformer information associated with said matching fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for
25 said at least one of said plurality of conformers and storing said conformer information for at least one of said plurality of fragments in said library.

101. The apparatus of claim 99 wherein the means for combining at least two candidate conformers further comprises:

30 means for relationally joining a first candidate conformer and a second candidate conformer chosen from said plurality of candidate conformers.

102. The apparatus of claim 99 wherein the means for relationally joining further comprises:

35 determining a maximally overlapping molecular structure from said first candidate conformer and said second candidate conformer.



103. The apparatus of claim 99 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

5 104. The apparatus of claim 99 wherein fragments are represented by a plurality of nodes and edges.

105. The apparatus of claim 104 wherein each of said plurality of nodes represents a collection of atoms in a molecule.

10 106. The apparatus of claim 104 wherein each of said plurality of edges represents a molecular bond.

15 107. The apparatus of claim 106 wherein said means for decomposing further comprises:
means for enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

20 108. The apparatus of claim 107 wherein M is 2 and P is 7.

109. The apparatus of claim 99 wherein said first nominal energy value is a global minimum energy value for said fragment, said apparatus further comprising:
means for automatically determining said global minimum energy value for said fragment.

25 110. The apparatus of claim 109 wherein said means for automatically determining said global minimum energy value for said fragment further comprises:
means for determining a quantity of rotatable bonds in said fragment; and
means for multiplying said quantity by a per bond energy factor.

30 111. The apparatus of claim 110 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

35 112. The apparatus of claim 99 wherein said second nominal energy value is an energy threshold value for said fragment, said apparatus further comprising:
means for automatically determining said energy threshold value for said fragment.

113. The apparatus of claim 112 wherein said means for automatically determining said energy threshold value for said fragment further comprises:



means for determining a quantity of rotatable bonds in said fragment; and
means for multiplying said quantity by a per bond cut-off energy.

114. The apparatus of claim 113 wherein said per bond cut-off energy is 0.4 kcals per
5 rotatable bond.

115. A molecule having a corresponding molecular structure derived using a computer
based method, said computer based method substantially similar to the method of claim 51.

116. The molecule of claim 115 wherein the determining for each normalized fragment
10 step further comprises:

searching in a library of normalized fragments for at least one matching fragment, said
matching fragment being identical to at least one normalized fragment in said plurality of normalized
fragments; thereupon,

15 if a matching fragment is found, using conformer information associated with said matching
fragment as said at least one of said plurality of conformers, otherwise, computing conformer information for
said at least one of said plurality of conformers and storing said conformer information for at least one of
said plurality of fragments in said library.

117. The molecule of claim 115 wherein the combining at least two candidate
20 conformers step further comprises:

relationally joining a first candidate conformer and a second candidate conformer chosen
from said plurality of candidate conformers.

118. The molecule of claim 117 wherein the relationally joining step further comprises:
25 determining a maximally overlapping molecular structure from said first candidate
conformer and said second candidate conformer.

119. The molecule of claim 115 wherein each conformer in said plurality of conformers
30 has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized
fragment.

120. The molecule of claim 115 wherein fragments are represented by a plurality of
nodes and edges.

121. The molecule of claim 120 wherein each of said plurality of nodes represents a
35 collection of atoms in a molecule.



122. The molecule of claim 120 wherein each of said plurality of edges represents a molecular bond.

123. The molecule of claim 122 wherein said decomposing step further comprises the step of:
enumerating fragments based upon a path including nodes and edges, said path having a characteristic length said characteristic length being the number of atoms in said path, said characteristic length having a minimum value of M, said characteristic length having a maximum value P.

124. The molecule of claim 115 wherein M is 2 and P is 7.

125. The molecule of claim 115 wherein said first nominal energy value is a global minimum energy value for said fragment, said molecule further comprising:
automatically determining said global minimum energy value for said fragment.

126. The molecule of claim 125 wherein said automatically determining said global minimum energy value for said fragment further comprises:
determining a quantity of rotatable bonds in said fragment; and
multiplying said quantity by a per bond energy factor.

127. The molecule of claim 126 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

128. The molecule of claim 115 wherein said second nominal energy value is an energy threshold value for said fragment, said molecule further comprising:
automatically determining said energy threshold value for said fragment.

129. The molecule of claim 128 wherein said automatically determining said energy threshold value for said fragment further comprises:
determining a quantity of rotatable bonds in said fragment; and
multiplying said quantity by a per bond cut-off energy.

130. The molecule of claim 129 wherein said per bond cut-off energy is 0.4 kcals per rotatable bond.

131. A method for determining likely torsion values for each conformer in a plurality of conformers, each conformer having a plurality of particular chemical bond representations, said method comprising the steps:



associating with each conformer a plurality of internal coordinates, each of said internal coordinates having a plurality of characteristic values corresponding to each chemical bond representation in said conformer;

determining, for each conformer, an energy difference level, said energy difference level
5 representing an incremental amount of potential energy above a first nominal energy value, said incremental amount of potential energy required to maintain each chemical bond representation at each of said plurality of characteristic values;

for each internal coordinate, for each characteristic value in said plurality of characteristic values, selecting a conformer having a corresponding minimum energy difference level, said corresponding
10 minimum energy difference level selected from said energy difference level computed for each conformer, to form a plurality of corresponding minimum energy difference levels;

for each characteristic value, determining a maximum energy difference value from among said plurality of corresponding minimum energy difference levels, to form a plurality of maximum energy difference values;

15 for each internal coordinate, selecting a plurality of likely values from said plurality of characteristic values, wherein said plurality of likely values in said internal coordinate correspond to a second nominal energy value selected from said plurality of maximum energy difference levels.

20 132. The method of claim 131 wherein the internal coordinates further comprise torsions.

133. The method of claim 131 wherein said first nominal energy value is a global minimum energy value for said fragment, said method further comprising the step of:

25 automatically determining said global minimum energy value for said fragment.

134. The method of claim 133 wherein said automatically determining said global minimum energy value for said fragment further comprises:

30 determining a quantity of rotatable bonds in said fragment; and
multiplying said quantity by a per bond energy factor.

135. The method of claim 134 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

35 136. The method of claim 131 wherein said second nominal energy value is an energy threshold value for said fragment, said method further comprising the step of:

automatically determining said energy threshold value for said fragment.



137. The method of claim 136 wherein said automatically determining said energy threshold value for said fragment further comprises:

determining a quantity of rotatable bonds in said fragment; and
multiplying said quantity by a per bond cut-off energy.

138. The method of claim 137 wherein said per bond cut-off energy is 0.4 kcals per rotatable bond.

139. A computer programming product for determining likely torsion values for each conformer in a plurality of conformers, each conformer having a plurality of particular chemical bond representations, said computer programming product comprising:

code for associating with each conformer a plurality of internal coordinates, each of said internal coordinates having a plurality of characteristic values corresponding to each chemical bond representation in said conformer;

code for determining, for each conformer, an energy difference level, said energy difference level representing an incremental amount of potential energy above a first nominal energy value, said incremental amount of potential energy required to maintain each chemical bond representation at each of said plurality of characteristic values;

code for selecting for each internal coordinate, for each characteristic value in said plurality of characteristic values, a conformer having a corresponding minimum energy difference level, said corresponding minimum energy difference level selected from said energy difference level computed for each conformer, to form a plurality of corresponding minimum energy difference levels;

code for determining for each characteristic value, a maximum energy difference value from among said plurality of corresponding minimum energy difference levels, to form a plurality of maximum energy difference values;

code for selecting for each internal coordinate, a plurality of likely values from said plurality of characteristic values, wherein said plurality of likely values in said internal coordinate correspond to a second nominal energy value selected from said plurality of maximum energy difference levels; and
a computer readable storage medium for storing the codes.

140. The computer programming product of claim 139 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

141. The computer programming product of claim 139 wherein said first nominal energy value is a global minimum energy value for said fragment, said computer programming product further comprising:

code for automatically determining said global minimum energy value for said fragment.



142. The computer programming product of claim 141 wherein said code for automatically determining said global minimum energy value for said fragment further comprises:

code for determining a quantity of rotatable bonds in said fragment; and
code for multiplying said quantity by a per bond energy factor.

143. The computer programming product of claim 142 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

144. The computer programming product of claim 139 wherein said second nominal energy value is an energy threshold value for said fragment, said computer programming product further comprising:

code for automatically determining said energy threshold value for said fragment.

145. The computer programming product of claim 144 wherein said code for automatically determining said energy threshold value for said fragment further comprises:

code for determining a quantity of rotatable bonds in said fragment; and
code for multiplying said quantity by a per bond cut-off energy.

146. The computer programming product of claim 145 wherein said per bond cut-off energy is 0.4 kcals per rotatable bond.

147. An apparatus for determining likely torsion values for each conformer in a plurality of conformers, each conformer having a plurality of particular chemical bond representations comprising:

a processor operatively disposed to perform the steps of:

associating with each conformer a plurality of internal coordinates, each of said internal coordinates having a plurality of characteristic values corresponding to each chemical bond representation in said conformer;

determining, for each conformer, an energy difference level, said energy difference level representing an incremental amount of potential energy above a first nominal energy value, said incremental amount of potential energy required to maintain each chemical bond representation at each of said plurality of characteristic values;

for each internal coordinate, for each characteristic value in said plurality of characteristic values, selecting a conformer having a corresponding minimum energy difference level, said corresponding minimum energy difference level selected from said energy difference level computed for each conformer, to form a plurality of corresponding minimum energy difference levels;

for each characteristic value, determining a maximum energy difference value from among said plurality of corresponding minimum energy difference levels, to form a plurality of maximum energy difference values; and



for each internal coordinate, selecting a plurality of likely values from said plurality of characteristic values, wherein said plurality of likely values in said internal coordinate correspond to a second nominal energy value selected from said plurality of maximum energy difference levels.

5 148. The apparatus of claim 147 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

10 149. The apparatus of claim 147 wherein said first nominal energy value is a global minimum energy value for said fragment, said processor of said apparatus further disposed to perform the steps of:

 automatically determining said global minimum energy value for said fragment.

15 150. The apparatus of claim 149 wherein said automatically determining said global minimum energy value for said fragment further comprises:

 determining a quantity of rotatable bonds in said fragment; and
 multiplying said quantity by a per bond energy factor.

20 151. The apparatus of claim 150 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

25 152. The apparatus of claim 147 wherein said second nominal energy value is an energy threshold value for said fragment, said processor of said apparatus further disposed to perform the steps of:
 automatically determining said energy threshold value for said fragment.

 153. The apparatus of claim 152 wherein said automatically determining said energy threshold value for said fragment further comprises:

30 determining a quantity of rotatable bonds in said fragment; and
 multiplying said quantity by a per bond cut-off energy.

 154. The apparatus of claim 153 wherein said per bond cut-off energy is 0.4 kcals per rotatable bond.

35 155. An apparatus for determining likely torsion values for each conformer in a plurality of conformers, each conformer having a plurality of particular chemical bond representations comprising:

 means for associating with each conformer a plurality of internal coordinates, each of said internal coordinates having a plurality of characteristic values corresponding to each chemical bond representation in said conformer;



means for determining, for each conformer, an energy difference level, said energy difference level representing an incremental amount of potential energy above a first nominal energy value, said incremental amount of potential energy required to maintain each chemical bond representation at each of said plurality of characteristic values;

5 means for selecting for each internal coordinate, for each characteristic value in said plurality of characteristic values, a conformer having a corresponding minimum energy difference level, said corresponding minimum energy difference level selected from said energy difference level computed for each conformer, to form a plurality of corresponding minimum energy difference levels;

means for determining for each characteristic value, a maximum energy difference value
10 from among said plurality of corresponding minimum energy difference levels, to form a plurality of maximum energy difference values; and

means for selecting for each internal coordinate, a plurality of likely values from said plurality of characteristic values, wherein said plurality of likely values in said internal coordinate correspond to a second nominal energy value selected from said plurality of maximum energy difference levels.

15 156. The apparatus of claim 155 wherein each conformer in said plurality of conformers has at least one of a plurality of bonds, each bond interconnecting at least two atoms in said normalized fragment.

20 157. The apparatus of claim 155 wherein said first nominal energy value is a global minimum energy value for said fragment, said apparatus further comprising:
means for automatically determining said global minimum energy value for said fragment.

25 158. The apparatus of claim 157 wherein said means for automatically determining said global minimum energy value for said fragment further comprises:
means for determining a quantity of rotatable bonds in said fragment; and
means for multiplying said quantity by a per bond energy factor.

30 159. The apparatus of claim 158 wherein said per bond energy factor ranges from 1.0 kcals to 4.0 kcals per rotatable bond.

35 160. The apparatus of claim 155 wherein said second nominal energy value is an energy threshold value for said fragment, said apparatus further comprising:
means for automatically determining said energy threshold value for said fragment.

161. The apparatus of claim 160 wherein said means for automatically determining said energy threshold value for said fragment further comprises:
means for determining a quantity of rotatable bonds in said fragment; and
means for multiplying said quantity by a per bond cut-off energy.



162. The apparatus of claim 161 wherein said per bond cut-off energy is 0.4 kcals per rotatable bond.

