MILLSIAN 2.0: A molecular modeling software for structures, charge distributions, and energetics of biomolecules

W. Xie, R. L. Mills,[a] W. Good, A. Makwana, B. Holverstott, and N. Hogle
Millsian, Inc., 493 Old Trenton Road, Cranbury, New Jersey 08512, USA

(Received 12 January 2011; accepted 24 February 2011; published online 24 March 2011)

Abstract: In this paper, we describe the implementation and test cases for a new molecular modeling software package, called MILLSIAN 2.0, designed for modeling the three-dimensional structures, charge distribution, and energetics of biomolecules of pharmaceutical interests. MILLSIAN's predictions of molecular properties are based on the solved parameters (i.e., bonds, angles, dihedral angles, and charge distributions) from classical physics principles of Mills' work. Charge distribution on the spheroidal molecular orbital is directly transferable in the functional groups. Similar to the implementation of force fields, an atom typing scheme is employed to facilitate the recognition of the functional groups previously solved from classical physics by Mills (The Grand Unified Theory of Classical Physics, available at http://blacklightpower.com/theory/book.shtml). The implementation of MILLSIAN 2.0 is extensively tested against the available experimental data. Remarkable agreement between MILLSIAN predictions and experiments has been observed. © *2011 Physics Essays Publication*. [DOI: 10.4006/1.3567145]

Résumé: Dans cet article nous décrivons l'exécution et les essais d'un nouveau progiciel de modélisation moléculaire intitulé MILLSIAN 2.0 élaboré pour la modélisation des structures 3D, la distribution de charges et les énergétiques biomoléculaires d'intérêt pharmaceutique. Les prédictions de Millsian sur les propriétés moléculaires sont basées sur les paramètres résolus (par ex. les liaisons, les angles, les angles dihedrales et les distributions des charges) des principes de physique classique du travail de Mills. La distribution des charges sur l'orbital moléculaire sphéroïdal est directement transférable dans les groupes fonctionnels. Semblablement à l'exécution des champs de force, un schéma de modèle atomique est employé pour faciliter la reconnaissance des groupes fonctionnels précédemment résolus à partir de la physique classique de Mills. L'exécution de MILLSIAN 2.0 est testée largement contre les données expérimentales disponibles. Une concordance remarquable des prévisions de Millsian et les expérimentations a été observée.

Key words: Molecular Modeling; Molecular Structure; Classical Physics; Dipole Moment; Bond Moment; Electron Density.

I. INTRODUCTION

Computer simulation has become a powerful tool for the predictions of molecular structure, visualization, dynamics, and chemical reactions.[1] Of fundamental importance to the reliability of predicted molecular properties is its energy function. Based on classical physics principles, Mills solved for the first time the closed form solution of bound electrons in atoms and molecules.[2–13] The predicted molecular and atomic properties including structure parameters, energies, and spectroscopic properties are in remarkable agreement with experimentally observed data.

Molecular simulation for biological molecules has become an important area of interest in pharmaceutical industry and academic research. Popular software packages that are used to model biomolecules include CHARMM,[14] AMBER,[15] and OPLS.[16] Currently, the biomolecular modeling packages employ the simple empirical functions to describe the interactions within the molecular systems.[14–16] For example, the atoms are treated as dimensionless points carrying some certain point charges in space. Therefore, the electrostatic interaction between atoms is described by the Coulombic energy of formal point charges. To prohibit atoms collapsing into each other, the empirical short-range repulsion terms are added into the energy functions, usually in the Lennard-Jones form to increase the computational speed. Bonds and angles are treated as harmonic springs with parametrized force constants and equilibrium values fitted to reproduce the experimental vibrational spectrum and molecular structure. Furthermore, it was found that in the force field the rotation about bonds cannot be adequately described by the simple Coulomb interactions between formal charges and short-range repulsion terms; thus empirical cosine functions are usually incorporated into the energy functions.

Although force fields based on simple empirical energy functions do not represent the physical molecular systems, it has been a practical tool for the simulation of large biomolecular systems, which normally requires millions of repeated energy and force evaluations.[17] In fact, the electrons are not explicitly treated in the force fields and all the empirical energy components are fitted to the experimental data.

[a]Author to whom correspondence should be addressed. rmills@blacklightpower.com

Therefore, the limitation of force fields is due to its inherent simplification of energy functions. For example, the reliability of force fields depends largely on the optimization of the parameters in the energy function. In principle, force fields can only be applied to the molecules with parameters fitted to reproduce the experimental data. Practically, this is a significant limitation for the modeling of biomolecular systems because the force field parametrization is very time consuming. Furthermore, the force field development depends largely on the availability of high quality experimental data. Due to the large variety of functional groups in pharmaceutical target molecules, the applicability of force fields is limited and its reliability becomes questionable for molecules, which consist of functional groups of distant relevance to the parametrized functional groups in the force fields.

Aside from its intrinsic limitations due to the over simplification of energy functions in order to enhance the speed of calculations, the molecular modeling using force fields highly depends on the quality of input three-dimensional (3D) molecular structures. Although structure generation of simple molecules has normally been incorporated in most of the molecular modeling packages using the force fields,[14–16] the structures of most complicated molecules, such as cyclic molecules with fused and bridged rings, are challenging to calculate in these packages. A popular program for structure generation of arbitrary molecules, CORINA, is based on a database search and the assumption that the structure factors depend mainly on the hybridization states and covalent radius of atoms.[18,19]

A powerful molecular modeling software thus should satisfy the following requirements: (1) be fast in computing the molecular properties, (2) be applicable to as many different classes of molecules as possible, (3) require a minimum amount of effort in parametrization, and (4) quickly generate reliable 3D structure. To the best of our knowledge, none of the current molecular modeling packages satisfies all of the four requirements. MILLSIAN software is designed to meet the above requirements by utilizing an accurate and efficient classical physics model together with a careful design of the software architecture. The rest of this paper is organized as follows: in Sec. II, we discuss the theoretical background of MILLSIAN 2.0; Sec. III outlines the MILLSIAN software architecture and initial structure generation; Sec. IV focuses on the optimization methods and molecular dynamics implemented in MILLSIAN; results from MILLSIAN 2.0 and discussion are presented in Sec. V; finally, Sec. VI provides concluding remarks about the implementation and performance of MILLSIAN 2.0.

II. THEORETICAL BACKGROUND

The grand unified theory of classical physics (GUT-CP) has been developed by Mills in a series of peer-reviewed publications.[2–13] Readers are directed to the references for details of the method. In this paper, we only outline the fundamental principles used in the MILLSIAN software. In GUT-CP, the bound electrons are treated as two-dimensional (2D) surfaces moving in the potential enforced by the nuclei.[2] The electron distribution, which satisfies the nonradiation requirement, is a solution to the Maxwell's equation. Using GUT-CP, Mills solved the structure, spectroscopy, energetics, and charge distribution of a series of functional groups that can be utilized to construct complicated molecules.[2–13] A list of the organic functional groups solved by classical physics has been presented in a previous publication.[20]

In GUT-CP, a molecular orbital (MO) is a 2D spheroidal surface that satisfies the nonradiation condition and Maxwell's equations. The spheroidal MO is characterized by three fundamental parameters: the major and minor axes and the ellipse of the MO. The charge distribution on the MO is solved within the GUT-CP framework and validated against the bond moments and dipole moments of the molecules. The charge distribution of the MOs of solved functional groups is directly transferable to the molecules consisting of that functional group. Electrostatic interactions in the molecule can thus be calculated from the charge distribution.

MILLSIAN models a molecule by first identifying the functional groups and assigning the bond parameters and charge distribution to each bond in the molecule. The total bond energy is treated as a summation of the energy of each individual bond in the molecule. This approach has been proven to be extremely successful by comparing the predicted heat of formation for a series of molecules with the experimentally observed ones. The errors of the results obtained from the linear combination of bonds were typically less than 0.1%,[20] whereas those of prior approaches were much greater as summarized in Fig. 1. This approach adopted in MILLSIAN significantly reduced the complexity of the energy function and makes the calculation extremely fast.

III. DESIGN OF MILLSIAN SOFTWARE

The fundamental unit used in MILLSIAN is the atom. Each chemical element rendered in MILLSIAN corresponds to a color at the same scale as in most of molecular visualization software. However, the charge density of atomic orbitals has been solved in GUT-CP. A bond in MILLSIAN is formed by an atom pair connected via a spheroidal molecular orbital and is rendered according to the charge distribution on the spheroid. Similarly, a bond angle is formed between two bonds that share a common atom center. To facilitate the definition of bond and angle types, MILLSIAN assigns each atom a uniquely defined atom type. Once the atom types are defined, bond and angle types can be easily identified by the atom types. Bond order is needed to specify a bond corresponding to the selected molecule to be solved and rendered. Based on these rules, the functional groups can be specified by atoms with appropriately assigned atom types and the bond connections between atoms. Bond and dihedral angles are generated automatically from bond connectivity. The following is an example of the functional group definition in MILLSIAN 2.0:

- FUNC CH3
- ATOM C CH3
- ATOM H1 H
- ATOM H2 H
- ATOM H3 H
- ATOM DUM DUM

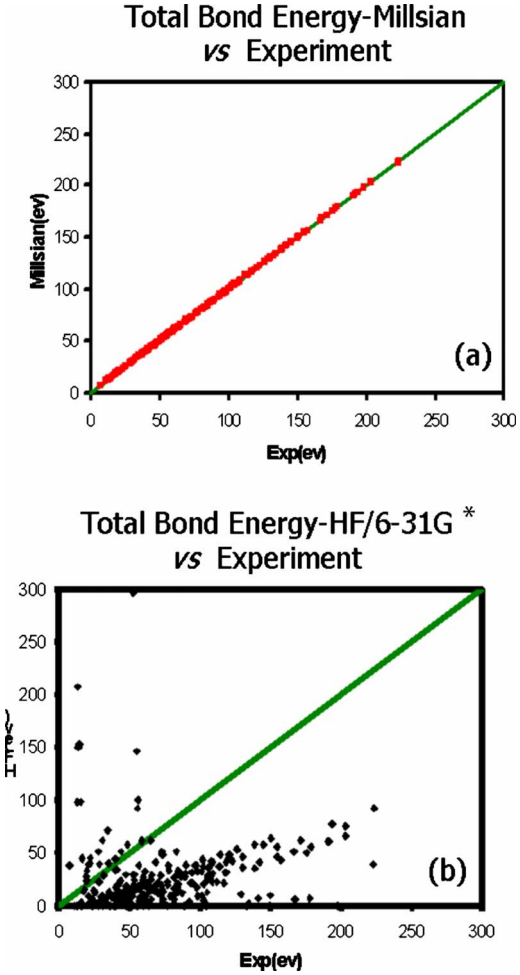

FIG. 1. (Color online) The total bond energies of exact classical solutions of 415 molecules generated by MILLSIAN 1.0 and those from a modern quantum mechanics-based program, Spartan's precomputed database using 6-31G* basis set at the Hartree–Fock level of theory, were compared to experimental values. (a) The MILLSIAN results were consistently within an average relative deviation of about 0.1% of the experimental values. (b) In contrast, the 6-31G* results deviated over a wide range of relative error, typically being >30%–150% with a large percentage of catastrophic failures, depending on functional group type and basis set.

- SINGLE C H1 C H2 C H3 C DUM

In the above example, the "FUNC" keyword defines a functional group called "CH3." The functional group name can be any string that uniquely identifies this functional group. "ATOM" keyword defines the atoms in the functional group. The first string following "ATOM" keyword is the unique name of the atom in this functional group, which is followed by its atom type. In the above case, the carbon atom is of atom type "CH3," and hydrogen atoms are of type "H." The last atom in CH3 functional group is a "dummy" to designate an empty valence to be ultimately filled by a bonding partner when CH3 is joined to another functional group. The introduction of dummy atoms is purely for the convenience of specifying an open bond axis used in making molecules from functional groups. By definition, any bond that has a dummy atom is an open bond axis and can be joined with other functional groups. Finally, the "SINGLE" keyword specifies a single bond. Similarly, "DOUBLE," "TRIPLE," and "AROMAT" are used to specify double,

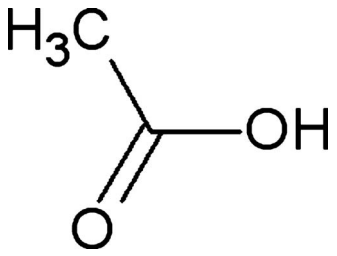

FIG. 2. Bond connectivity of acetic acid. The atom types are determined iteratively using the atom types of all neighbors connected to the center atom.

triple, and aromatic bonds. The first bond is between carbon atom C and hydrogen atom H1, and the bond type is thus CH3 H with bond order of one as specified by the "SINGLE" keyword. In the bond parameter table, we have the following entry for this bond: H CH3 1 1.649 20 1.109 74 4.163 95 0.370 00, where "H CH3" defines the bond made by atoms of type "H" and "CH3;" "1" denotes the bond order is 1, what follows are the major axis of the spheroid (units of the Bohr radius a_0), bond length (Å), bond energy (eV), and the bond dipole moment (units of D). Minor axis is calculated from the bond length and major axis. A positive bond moment means the dipole moment of the bond is in the direction of H to CH3, while a negative bond moment specifies a dipole moment in the reverse direction. Once the parameters for all the bonds in a molecule are known, the total bond energies of the molecule can be readily calculated by summing up all the bonds in the molecule.

Previously solved bond angles are defined or designated in a similar way. For example, the angle between H–C–H in the methyl group is defined as following: H CH3 H 109.5, where "H CH3 H" defines an angle made by atoms of type "H," "CH3," and "H," followed by the angle of 109.5°.

Finally, the dihedral angles formed by three connected MOs are calculated as discussed in the "Dihedral angle" and "General dihedral angle" sections in Chap. 15 of Ref. 2.

A. Determination of the atom types from connectivity

Suppose we parse the molecule in Fig. 2, acetic acid, for which only the connectivity between atoms is known. In the atom type definition table, there are the following entries:

- LEVEL 1
- TYPE CMH3
- MAIN C
- SINGLE H H H*
- TYPE CH3
- MAIN C
- SINGLE H H H C
- TYPE OK
- MAIN O
- DOUBLE C
- TYPE CEST
- MAIN C
- DOUBLE O
- SINGLE O*
- TYPE OH

- MAIN O
- SINGLE H*
- LEVEL 2
- TYPE OACD
- MAIN OH
- SINGLE H CEST
- TYPE CCA
- MAIN CEST
- DOUBLE OK
- SINGLE OH*
- LEVEL 3
- TYPE OCA
- MAIN OEST
- DOUBLE CCA

wherein each level assigns a higher degree of specificity to the groups of bonding atoms in terms of the functional group classification. For example, consider the group of atoms O, C, and O having one double and one single bond that are designated as such in "LEVEL 1" and then to carbonyl to O single bond in "LEVEL 2," followed by carboxylic acid in the next higher level, "LEVEL 3." The atom type symbol follows the "TYPE" keyword; "MAIN" specifies an atom center that the type is being defined; "SINGLE," "DOUBLE," "TRIPLE," or "AROMAT" specify the atoms connected to the center via single, double, triple, or aromatic bonds, respectively; the " *" symbol in the above representations designates a "wildcard" atomic center, any atom in a bond of a particular type to be determined from the identity of the bonding partner. In the cases that involve incumbent nonphysical resonance structures, MILLSIAN inputs a series of integer bond orders of different types and outputs the appropriate noninteger bond order previously solved analytically[2] and designates the atom type accordingly. In the beginning, all atom types are initialized by their chemical element names. For example, hydrogen, carbon, and oxygen atoms are of the types "H," "C," and "O," respectively. By definition, the term "dummy atom" designated type "DUM" corresponds to a single bond empty valence of an atom center that is filled when the bonding partner to an atom center is selected. In a reiterative manner, the updated atom types and connectivity are used to determine the atom types and connectivity in the next step by formulating an altered sequence of atom types and connectivity that more closely satisfies the structure input by the user. The reiteration is performed until all the inherent bonding rules are satisfied with a match to the structure. Specifically, the valence calculated from the connectivity, the bond orders, and all atom types, except for wildcards defined in the atom type definitions, should match exactly to the input structure.

In the implementation of the operations of each LEVEL to arrive at further specificity of the functional groups and their connectivity, the algorithm eliminates arbitrary general cases in favor of specific cases. If an atom is either arbitrary (i.e., a wildcard) or a specific atom type, then the latter atom type will be used. For example, in LEVEL 1 of an iterative algorithm of multiple LEVELS, wherein the definition of atom types are assigned, first, the elements such as C and the H's of CH3 functional group are designated as atom types. Then, the C (C1) in the methyl group is updated to "CH3" according to the definition of type "CH3" comprising an atom type "C" connected to three atoms of type H by single bonds. This is outputted over the more arbitrary case. For example, in a more general designation of a methyl functional group "CMH3" bound to an arbitrary atom type (i.e., wildcard designated by * in the representation list), CMH3 is excluded in favor of the result of CH3 bound to a specific atom type. In the example *supra*, according to priority rules, carbon C1 will be of type CH3. The second C (C2) becomes type "CEST;" the O (O1) with double bond to C has type "OK" and the O (O2) connected to C via single bond is of type "OH." After searching in the first level of atom type definitions, the atom types for C1, C2, O1, and O2 are updated. Then, the algorithm implements a search of the atom types defined at LEVEL 2. Once finished, the atom types are updated for the following two atoms: C2 is of type "CCA" and O3 is of type "OACD." Advancing to searching in LEVEL 3, O1 is updated to type "OCA." Once there are no more updates to the atom types, the search is deemed converged. The atom types for all atoms are listed in the following manner:

- H1 H
- H2 H
- H3 H
- H4 H
- C1 CH3
- C2 CCA
- O1 OCA
- O2 OACD

The above atom types are consistent with the following functional group definitions:

- FUNC CH3
- ATOM C CH3
- ATOM H1 H
- ATOM H2 H
- ATOM H3 H
- ATOM DUM DUM
- SINGLE C H1 C H2 C H3 C DUM
  and
- FUNC ACID
- ATOM C CCA
- ATOM O1 OCA
- ATOM O2 OACD
- ATOM H H
- ATOM DUM1 DUM
- SINGLE O2 C O2 H DUM1 C
- DOUBLE C O1

Note that the use of wildcards is very convenient since certain atom types can be identified by their connectivity to these empty valences in addition to possible other established connectivity. This is of special importance when an exact match cannot be found for some atoms, since the higher level atom type definitions arise from their prioritization over iterations of the algorithm corresponds in more refined atom types in the output structure. Thus, the iteration will assign an optimum atom type to the atom that does not

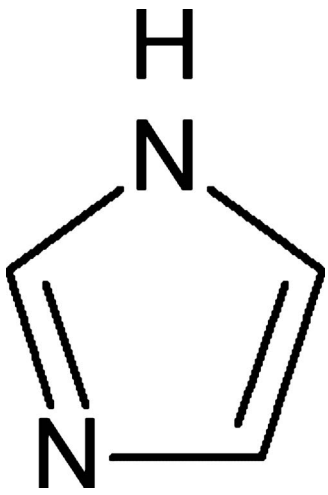

FIG. 3. Bond connectivity of imidazole. The atom types of the atoms on the ring are determined by matching the ring to ring templates in the MILLSIAN library.

have an exact match. Introduction of new functional groups can be conveniently achieved by adding a definition of the corresponding new atom types.

Using this iterative method, the change to a more complicated functional group comprised of simpler functional groups arising from substitution at distant atoms to one or more simpler groups can be reflected in the atom types, i.e., the type of O1 was updated to OCA because O2 is connected to H. Finally, all atom types are defined, and the associated parameters can be found in the bond table, where the following entries are given:

- H CH3 1 1.649 20 1.109 74 4.163 95 0.370 00
- CCA OCA 2 1.299 07 1.206 28 7.806 60 2.230 00
- CCA OACD 1 1.734 90 1.394 02 4.419 25 0.760 00
- OACD H 1 1.264 30 0.971 65 4.410 35 −1.510 00
- CCA CH3 1 2.047 40 1.514 37 4.431 10 0.000 00

From these parameters, we can calculate the total bond energy, bond length, molecular orbital parameters, charge distribution, and dipole moment.

B. Define parameters for cyclic templates

In principle, the atom types for atoms on a ring can be defined in the same manner as that for linear molecules. Given that the atom types depend mainly on the atomic elements and their connectivity on the ring, the types of all atoms are directly defined in a single ring definition. To illustrate this strategy, consider the five-membered ring of imidazole. The bond connectivity of imidazole is displayed in Fig. 3. In the atom type table for cyclic molecules, the following entries are given:

- RING IMID
- ATOM C1 C CPYR
- ATOM N1 N NIMD
- ATOM C3 C CPYR
- ATOM C4 C CPYR
- ATOM N2 N NPYR
- SINGLE N1 C3 C4 N2 N2 C1
- DOUBLE C1 N1 C3 C4

where the keyword "RING" specifies the ring definition. This is followed by the name of the ring "IMID." The ATOM keyword is followed by three strings that specify the atom name used to define the connectivity within the ring, then the chemical element and atom type, respectively. For aromatic rings, the "ISAROM" flag is set to be "TRUE." By default, this flag is not set and the ring is not aromatic. Once a cyclic molecule is read into the MILLSIAN, the atom types of all atoms including those on the ring are assigned using the iterative method for linear molecules. In this case, type H is assigned to all hydrogen. The connectivity and the chemical elements in imidazole match the above definition for ring IMID, so we have the following atom types on the ring:

- C1 CPYR
- N1 NIMD
- C3 CPYR
- C4 CPYR
- N2 NPYR

which is consistent with the definition of the functional group:

- FUNC IMID
- ATOM C1 CPYR
- ATOM H1 H
- ATOM C2 CPYR
- ATOM H2 H
- ATOM N NPYR
- ATOM HN H
- ATOM C3 CPYR
- ATOM H3 H
- ATOM N2 NIMD
- SINGLE C1 H1 C2 H2 N HN C3 H3 C2 N C3 N C1 N2
- DOUBLE C1 C2
- DOUBLE C3 N2

Once the types of the atoms on the ring are determined, the iterative method determines the atom types due to the substitution on the ring. Substituted rings are defined in a similar way with the keyword "SRIN." For example, the 1,3,5-substituted benzene ring having three chlorine atoms is defined in the following manner:

- SRIN BENZ
- ISAROM TRUE
- ATOM CL1 CL CACL
- ATOM C2 CAH1 CAH1
- ATOM CL3 CL CACL
- ATOM C4 CAH1 CAH1
- ATOM CL5 CL CACL
- ATOM C6 CAH1 CAH1
- AROMAT CL1 C2 CL3 C4 CL5 C6 C2 CL3 C4 CL5 C6 CL1

In the next LEVEL of ring designations, the element names are replaced by atom types, which may involve substitution of the standard ring. In case of substitution, the type of the nonhydrogen element name replaces the element name

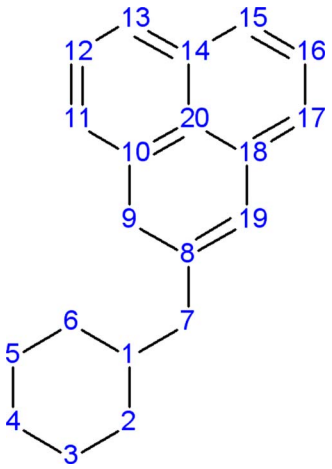

FIG. 4. (Color online) A cyclic molecule with fused rings that can be partitioned into two ring clusters. Rings 1-2-3-4-5-6, 8-9-10-20-18-19, 10-11-12-13-14-20, and 14-15-16-17-18-20 are the smallest set of the smallest rings (SSSR) for this molecule.

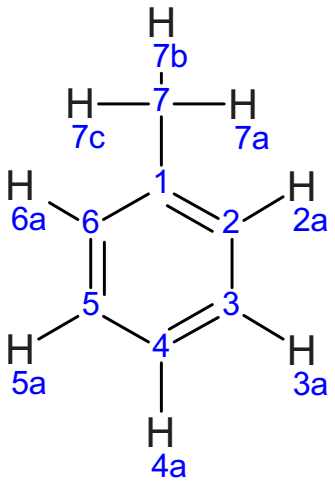

FIG. 5. (Color online) The terminal atoms that are not in any ring can be removed from the molecule to facilitate the search for SSSR. C7 and all hydrogen atoms are terminal atoms in methylbenzene.

of the atom on the ring for which the corresponding nonhydrogen atom substitutes. The final column lists the atom types for all the atoms on the ring. In the case of 1,3,5-chlorobenzene, the atom types are first assigned CAH1 from the definition of benzene ring. Then the atoms connected to atom of type "CL" is updated to CACL. Different levels of ring substitution can be used similarly to define atom types for linear functional groups; however, this was found to be unnecessary for the current version of MILLSIAN.

C. Smallest set of smallest rings (SSSR)

For the proper assignment of atom types, we need to find the rings in the molecule in order to determine the correct atom types for atoms in the rings. As it will become apparent *infra*, the SSSR is also necessary for the generation of 3D coordinates for cyclic molecules. For fused and bridged rings, the atom types are determined for the basic rings, and they will generate all the rings in the molecule. Then, the basic rings are modified for substitution where applicable. The molecule in Fig. 4 has a number of possible rings. However, only 4 six-membered rings are needed to generate all other rings in this molecule. For example, ring 10-11-12-13-14-15-16-17-18-20 can be generated by combining rings 10-11-12-13-14-20 and 14-15-16-17-18-20. These 4 six-membered rings are defined as SSSR for this molecule.

The algorithm to find the SSSR is illustrated in the following with pseudocode:

```
Loop over all atoms
    At atom i, search along the bonds until it goes back to i again
    If it does not go back to i, i is not in a ring, otherwise, the smallest
    ring that contains i is the first sequence that goes back to i.
    If there is a smallest ring, check if it is already in SSSR.
End loop over all atoms
```

Using the above algorithm, the following rings are identified: 1-2-3-4-5-6, 8-9-10-20-18-19, 10-11-12-13-14-20, 14-15-16-17-18-20.

A terminal atom is an atom that participates in only one bond. To reduce the computational cost searching for SSSR, terminal atoms are removed iteratively as illustrated in the following example for methylbenzene (Fig. 5). In the case demonstrated in Fig. 4, all the hydrogen atoms are terminal atoms; thus, they can be removed in the first step of the iteration. Once the hydrogen atoms in the methyl group are removed, C7 becomes a terminal atom and can be removed. This process is continued until there are no more terminal atoms left in the molecule. In the above case, only atoms 1–6 are left after following this procedure.

D. Determination of bond orders and aromaticity

MILLSIAN supports the Protein Data Bank (PDB) file format for the input of the molecules. PDB file does not store the information about the bond orders that are necessary for the determination of atom types in MILLSIAN. The assignment of bond orders for molecules input from the PDB format is based on the recognition of functional groups or templates as well as the analysis of bond lengths. If a protein functional group, amino acid, or other fragment generally known as a "residue" in the PDB file matches one in the MILLSIAN functional group or template library, then the connectivity and bond orders are assigned according to those stored in the MILLSIAN library. Currently, the MILLSIAN library comprises the following residues as basis elements of complex structures to conveniently assign connectivity and bond order:

- *Amino acid residues:* ALA, ARG, ASN, ASP, CYS, GLN, GLU, GLY, HIS, ILE, LEU, LYS, MET, PHE, PRO, SER, THR, TRP, TYR, and VAL.
- *Nucleotides:* A, U, G, C, DA, DT, DG, and DC.
- *Heterogens:* HEM, NAG, ACE, ASX, GLX, EOH, GOL, HYP, MOH, OXY, PCA, and PER.

The residue identifier in the PDB file is used to match the residues stored in the MILLSIAN library. If a residue does not match any residue in the library, then the connectivity and bond orders are assigned based on the distance of atom

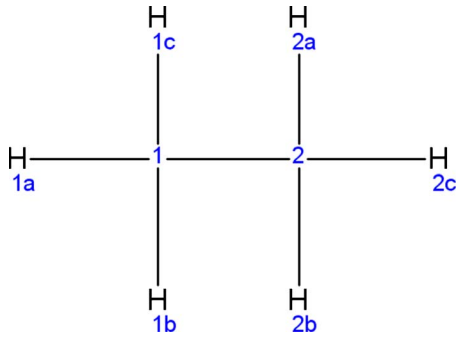

FIG. 6. (Color online) To generate the coordinates for ethane, C1 is moved to the origin, and H1c is then moved to the *x*-axis. The coordinates for all other atoms are generated using the bond length and the angles. Dihedral angles are used to minimize the repulsion between groups in the molecule.

pairs. The advantage of using distance-based bond order assignment is that it can be applied to residues that are not in the library. The disadvantage is that the distance-based bond order assignment may generate unsatisfactory bond orders. For example, if the 3D coordinates have a relatively large uncertainty, the connectivity and bond orders may not be generated correctly. To reduce this problem, MILLSIAN implements a bond order "clean up" method based on the normal valence of the chemical elements. The algorithm is iterative and is illustrated in the following pseudocode:

```
Assign the bond orders based on distance
Loop over all atoms
    Check the used valence of atom i
    While the used valence is more than the allowed valence, reduce the
    highest bond order by one. If there are more than one bond of the
    same highest order, then reduce the bond order of the bond which has
    the largest bond length.
End loop over all atoms
Redo the above loop until no bond orders are updated.
```

The above algorithm will solve the problems for some of the bonds. However, some unsatisfactory bond orders may still exist. MILLSIAN 2.0 implements a graphical user interface to allow the users to interactively update the bond orders of the desired bonds. Once the bond orders are determined, the next step is to assign the aromaticity to the rings present in the molecule. Hukel's rule is used to assign the aromaticity.

E. Initial structure generation

Generating initial molecular structure is an essential step in molecular modeling. Some molecular input formats, i.e., SMILES, do not contain the information about the 3D coordinates of the molecule. Therefore, an important task is to generate the 3D coordinates along with the connectivity information about the molecule. In the current implementation, the coordinates for acyclic and cyclic molecules are generated using different algorithms due to the extra constraints put on cyclic molecules.

1. Acyclic molecules

Initial structure generation for linear molecules is illustrated using ethane exemplarily displayed in Fig. 6. Select any atom, such as atom C1 and designate its center as the origin coordinate position (0.0, 0.0, 0.0). Then check the first atom connected to C1, which is H1a. Find the bond length parameters for the bond H1a-C1 (bond type H CH3) and put H1a on the *x*-axis. Next, check the second atom connected to C1, which is H1b. Find the parameters for bond H1b-C1 (bond type H CH3) and angle H1a-C1-H1b (angle type H CH3 H), and then put H1b on the *xy*-plane. Finally, the positions for all of the other atoms connected to C1 are solved using the bond length and two angles between the atoms already solved. If C1 is a chiral center, then there are two possible positions that satisfy the bond and angle constrains corresponding to the two different enantiomers involving the solution of atom H1c. When solving the position for H2a, a dihedral angle of 180° for H1a–C1–C2–H2a is used to minimize the repulsion resulting in a gauche conformation. Heavy atoms have higher priority over hydrogen atoms when the dihedral angle assignment is determined. The algorithm loops over the solved atoms coordinates until all the atom positions are solved. The pseudocode is as follows:

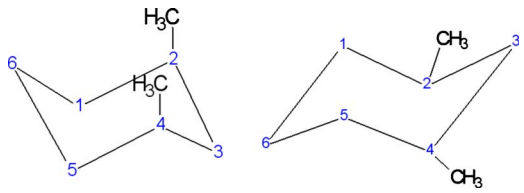

FIG. 7. (Color online) Two chair conformations of cyclohexane become equivalent when there are no substitutions.

```
Assign (0.0, 0.0, 0.0) to the first atom. Flag it as solved.
    Calculate the coordinates for the first atom connected to it using the
    bond length parameter and align it on the x-axis. Flag it as solved.
    Calculate the coordinates for the second atom connected to it using
    bond length and angle parameters and put it on the xy-plane. Flag it
    as solved.
Repeat until positions of all atoms are solved
    Calculate the coordinates for the second atom connected to it using
    bond length and angle parameters and put it on the xy-plane. Flag it
    as solved.
    Calculate the coordinates for all other atoms connected to it using two
    bond angles and bond length.
End
```

The above algorithm will generate reasonable coordinates for molecules without rings. However, for the molecules with rings, the dihedral angle information must be included in order to generate reasonable cyclic structures as discussed in the next section.

2. Cyclic molecules

Rings in the molecule enforce constraints on the structure. For example, the chair conformation of a six-membered ring has dihedral angles at about (60,−60,60,−60,60,−60). In fact, the similar dihedral angle series of (−60,60,−60,60, −60,60) also specifies the chair conformation (Fig. 7). In the case where the six-membered ring is not substituted, these two chair conformations are equivalent. However, when substitutions are made at positions on the ring, the chair confor-

mations are no longer equivalent. In MILLSIAN, the chair conformations corresponding to different dihedral angle series are denoted as "+" and "−" conformations to distinguish between the two.

Distinguishing between the two chair conformations is especially important to generate the fused or bridged cyclic molecules as the neighboring rings impose constraints on the conformation. When connectivity and bond order information for a molecule is parsed into the MILLSIAN, the SSSR is identified using the algorithm discussed *supra*. Then, the connectivity between the rings is generated and used to partition the rings into ring clusters, a group of one or more fused or bridged rings separated by others by a linear chain. For example, the molecule in Fig. 4 contains two ring clusters. Atoms 1–6 belong to one ring cluster and atoms 8–20 belong to the other ring cluster. A back tracing algorithm, trial and error fit to one of the conformations, is employed to determine the conformation of the rings in the ring cluster. Coordinates for atoms not in any ring clusters are generated using the algorithms for acyclic molecules discussed *supra*.

IV. OPTIMIZATION METHODS IN MILLSIAN 2.0

With an initial geometry of the molecule, the energy penalty function is constructed according to the following equation:

$$E_p = \sum_{\text{bond}} k_b(r - r_{\text{eq}})^2 + \sum_{\text{angle}} k_a(\theta - \theta_{\text{eq}})^2 + \frac{1}{2}\sum_{ij} \frac{\int \rho_i \rho_j}{4\pi\varepsilon r_{ij}}, \tag{1}$$

where the first summation denotes the sum of the squared errors due to the deviation from presolved bond lengths, weighted by a positive number k_b, and the second summation denotes the sum of the squared errors of the deviation from the presolved angles, weighted by a positive number k_a. The last term accounts for the nonbonded interactions due to the charge distribution of the molecule. Equation (1) is a function of the Cartesian coordinates of all atoms in the system. The optimum structure can be found by minimizing the energy penalty function:

$$\frac{\partial E_p}{\partial X_i} = 0. \tag{2}$$

A. Steepest descent algorithm

The first method implemented in MILLSIAN to solve the coupled nonlinear equation is the steepest descent algorithm which solves the following equation:

$$f(x) = 0 \tag{3}$$

by using the iterative procedure.

Starting with an initial guess x_0, calculate the derivative of $f(x_0)$ denoted by $f'(x_0)$. Update the variables according to the following equation:

$$x_1 = x_0 - f'(x_0)\Delta s$$

$$x_2 = x_1 - f'(x_1)\Delta s$$

$$\vdots$$

$$x_n = x_{n-1} - f'(x_{n-1})\Delta s, \tag{4}$$

where Δs is the step size for minimization. Once $|x_n - x_{n-1}| \leq \varepsilon$, which is a predefined tolerance, convergence is achieved. Otherwise, the above steps are repeated.

The steepest descent algorithm can be extended to the minimization of the penalty energy by using internal coordinates such as the dihedral angles. In order to use dihedral angles as the variables in the minimization, the only change needed is to calculate the derivative of energy with respect to the dihedral angles φ using the following relationship:

$$\frac{\partial f}{\partial \varphi} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \varphi}. \tag{5}$$

B. Adopted basis Newton–Raphson

The adopted basis Newton–Raphson method (ABNR) was originally implemented in CHARMM by States.[21] The algorithm was described in a publication by Brooks *et al.* with its extension to minimize a path using the nudged elastic bond method[22] wherein the details for this method are given. In the ABNR method, Newton–Raphson algorithm is applied to a subspace of the coordinate vector spanned by the displacement coordinates of the last few minimization steps. The second derivative matrix is constructed numerically from the change in the gradient vectors and is inverted by an eigenvector analysis allowing the routine to recognize and avoid saddle points in the energy surface. At each step, the residual gradient vector is calculated and used to add a steepest descent step onto the Newton–Raphson step, and incorporating a new direction into the basis set. ABNR usually converges faster than the steepest descent method.

C. Simulated annealing algorithm

The steepest descent algorithm is likely to give local minima that satisfy the condition having the gradient of Eq. (3) close to zero. The simulated annealing algorithm avoids the local minima by introducing a random perturbation to the function. Simulated annealing has been implemented in MILLSIAN 2.0 using the internal coordinates. The algorithm is outlined in the following:

1. Starting with an initial set of dihedral angles, calculate the energy penalty E_p according to Eq. (1).
2. Generate a set of random displacements of dihedral angles and calculate E_p for this new configuration.
3. If the E_p is lower than the one from the previous structure, accept this random move.
4. If the E_p is higher than the one from the previous structure, accept this random move with the following probability:

$$P = \exp(-\Delta E/k_b T),$$

where ΔE is the difference between the current and previous structure E_p, k_b is the Boltzmann constant, and T is the speci-

fied temperature. If the temperature is high, the energy of Eq. (1) is more likely to overcome the barrier to get out of the local minimum.

5. Decrease the temperature.
6. Repeat steps 2–5 until predefined convergence criterion has been satisfied.

Simulated annealing is an effective way to avoid being trapped in the local minima, but it cannot be applied to the torsion angles in the rings for cyclic molecules because rotation around a bond in the ring will not change the conformation of the ring. This leads to the need for the coupled molecular dynamics and steepest descent method outlined *infra*.

D. VERLET algorithm

The VERLET algorithm is used to carry out molecular dynamics simulations of the molecular system. Given E_p for the interaction between functional group and molecules, the pair wise additive forces can be evaluated by taking the derivative of the energy penalty function:

$$F_i = -\frac{\partial E_p}{X_i}. \tag{6}$$

Once the force on each atomic site is calculated, the phase space of the system can be sampled according to the VERLET algorithm derived from Newton's equation of motion:

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a(t)\Delta t^2 \tag{7}$$

and the velocities can be updated by the following equation:

$$v(t) = \frac{1}{2\Delta t}[r(t+\Delta t) - r(t-\Delta t)], \tag{8}$$

where t denotes the time during which molecular dynamics occur and Δt is the time step. The time step used in MILLSIAN 2.0 is $\Delta t = 1$ fs.

E. Constrained molecular dynamics: SHAKE algorithm

In order to obtain meaningful statistics of a physical system, the molecular dynamics needs to be able to sample the phase space at a certain time scale (for example, nanosecond time scale for thermodynamical properties). To achieve a large time step for the simulation, the high frequency modes can be fixed during molecular dynamics, (i.e., the functional groups solved using classical physics can be treated as rigid bodies). The SHAKE algorithm,[23] which is widely used in constrained molecular dynamics, can be used to fix a certain number of degrees of freedom of the system.

The constraints on the n bond lengths take the following form:

$$\sigma_k^{(t)} = (r_{ij}^{(t)})^2 - d_k^2 = 0, \tag{9}$$

where $r_{ij}^{(t)}$ is the distance between atoms i and j, while d_k is the constrained bond length. These constraint equations are added into the energy penalty function in the equation of motion such that the constrained force is given by

$$F_i^c = F_i + \frac{\partial}{\partial X_i}\sum_k^n \lambda_k \sigma_k^{(t)}, \tag{10}$$

where the last term is the gradient due to the energy constraints. Adding the constraints should not change the forces acting on each atom because $\sigma_k^{(t)}$ will be zero when the constraints are satisfied. With the constraints, the VERLET algorithm reads

$$\begin{aligned} r(t+\Delta t) &= 2r(t) - r(t-\Delta t) + a(t)\Delta t^2 \\ &\quad + m_i^{-1}\Delta t^2 \frac{\partial}{\partial X_i}\sum_k^n \lambda_k \sigma_k^{(t)} \\ &= \hat{r}(t+\Delta t) + m_i^{-1}\Delta t^2 \frac{\partial}{\partial X_i}\sum_k^n \lambda_k \sigma_k^{(t)}, \end{aligned} \tag{11}$$

where the first three terms are the updates of positions due to the unconstrained dynamics and the last term is due to the constraints on the bonds. To satisfy the constraints at time step $t+\Delta t$, the Lagrange multipliers must be chosen such that

$$\sigma_k^{(t+\Delta t)} = (r_{ij}^{(t+\Delta t)})^2 - d_k^2 = 0. \tag{12}$$

This requires solving a system of nonlinear equations

$$\begin{aligned} \sigma_k^{(t+\Delta t)} = \Bigg\| & \hat{r}_i^{(t+\Delta t)} - \hat{r}_j^{(t+\Delta t)} + m_i^{-1}\Delta t^2 \frac{\partial}{\partial X_i}\sum_k^n \lambda_k \sigma_k^{(t)} \\ & - m_j^{-1}\Delta t^2 \frac{\partial}{\partial X_j}\sum_k^n \lambda_k \sigma_k^{(t)} \Bigg\|^2 - d_k^2 = 0 \end{aligned} \tag{13}$$

simultaneously for n unknown Lagrange multipliers.

This system of n nonlinear equations can be solved by Newton's method, where the solution vector λ is updated using

$$\lambda_n = \lambda_{n-1} - \mathbf{J}^{-1}\sigma, \tag{14}$$

where the matrix $\mathbf{J}$ is the Jacobian of the equations σ_k,

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \sigma_1}{\partial \lambda_1} & \frac{\partial \sigma_1}{\partial \lambda_2} & \cdots & \frac{\partial \sigma_1}{\partial \lambda_n} \\ \frac{\partial \sigma_2}{\partial \lambda_1} & \frac{\partial \sigma_2}{\partial \lambda_2} & \cdots & \frac{\partial \sigma_n}{\partial \lambda_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sigma_n}{\partial \lambda_1} & \frac{\partial \sigma_n}{\partial \lambda_2} & \cdots & \frac{\partial \sigma_n}{\partial \lambda_n} \end{pmatrix}. \tag{15}$$

In practical implementation, the positions and the Lagrange multipliers are updated according to the following equation:

$$\begin{aligned} \lambda_k^n &= \frac{\sigma_k}{\partial \sigma_k / \partial \lambda_k^{n-1}}, \\ r_i^n &= r_i^{n-1} + \lambda_k^n \frac{\partial \sigma_k}{\partial r_k}, \end{aligned} \tag{16}$$

until all constraints are satisfied to a predefined tolerance.

F. Coupled molecular dynamics and steepest descent method

The above methods are efficient for the geometry optimization of most of the molecules. However, for a molecule that has very complicated ring systems where the generated initial coordinates are far away from minimum, the optimization is likely to fail to converge. In order to optimize the conformation of the rings, a minimization method with Cartesian coordinates needs to be used. Molecular dynamics is used to sample the conformational space of the molecule with a complicated ring structure, resulting in reasonable rings that are close to the minimum. Once a reasonable conformation is generated, the structure can be further optimized using steepest descent or ABNR method. In the coupled molecular dynamics and minimization method, a conformation with smallest value of the energy penalty such as that from molecular dynamics given in Eq. (1) is used as the initial input for steepest descent or ANBR minimization. This method has been successfully used to generate the geometry for cyclic molecules with fused and bridged rings.

V. RESULTS AND DISCUSSIONS

In a previous paper,[20] we presented the total bond energies for a series of molecules containing C, H, O, N, P, S, Si, B, F, Cl, Br, I, and metals calculated from MILLSIAN based on the linear summation of bond energies. Excellent agreement has been obtained between the calculated and experimental data. The average relative error of only 0.1% of the experimental heat of formation is obtained, which significantly outperforms existing methods. This observation shows that the functional groups are indeed transferable. Interestingly, the transferability of functional groups is a direct result of the correctness of the exact physical solutions of the electronic structure given by Mills' classical physics approach. The detailed discussions about the MILLSIAN calculated total bond energy compared to experimental data and other approaches namely Hartree–Fock methods with 3-21G and 6-31G* basis sets can be found in an earlier paper.[20]

As a direct test for the electron distribution, dipole moments of a series of molecules were calculated using the vector sum of bond moments solved previously using classical physics. Table 16.17 in Chap. 16 of GUT-CP2 lists the bond moments solved along with the bond moments extrapolated from the experimental dipole moment of molecules. Overall, the predicted bond moments are in good agreement with the experimental value. The calculation of the dipole moment by vector sum of bond moments is permitted by the localization of the electrons in the molecules to the spheroidal MOs of each corresponding bond instead of being distributed over the entire space as in quantum mechanics. The bond moments listed in Table 16.17 were used to calculate the dipole moments of a series of organic molecules that span over 13 common organic functional groups. Table I shows dipole moments calculated using MILLSIAN compared to the experimental measurements. The overall accuracy of predicted dipole moments are remarkable, given that the experimental uncertainty involved in measuring dipole moment is larger for those measured in the condensed phase. The

TABLE I. Dipole moment of the molecules calculated from the vector sum of bond moments obtained from the closed form classical physics solutions.

Molecules	Expt.[a]	Calc.	Error
1-bromobutane	2.08	1.96	−0.12
Bromoethane	2.04	1.97	−0.07
1-bromoheptane	2.16	1.95	−0.21
1-bromopentane	2.16	1.95	−0.21
1-bromopropane	2.18	1.95	−0.23
2-bromopropane	2.21	2.01	−0.20
1-chlorobutane	2.05	1.97	−0.08
Chloroethane	2.05	1.97	−0.08
1-chloropentane	2.16	1.96	−0.20
1-chloropropane	1.95	1.96	0.01
2-chloropropane	2.17	1.98	−0.19
Fluoroethane	1.94	1.93	−0.01
1-fluoropropane	2.05	1.94	−0.11
2-fluoropropane	1.96	1.94	−0.02
Iodoethane	1.98	1.84	−0.14
1-iodopropane	2.04	1.81	−0.23
Methanol	1.69	1.72	0.03
Ethanol	1.69	1.73	0.04
1-propanol	1.65	1.71	0.06
2-propanol	1.69	1.71	0.02
1-butanol	1.6	1.73	0.13
Isobutanol	1.64	1.72	0.08
Methylamine	1.29	1.41	0.12
Dimethylamine	1.03	1.21	0.18
Ethylamine	1.22	1.40	0.18
Propylamine	1.17	1.33	0.16
Butyl ethyl ether	1.24	1.33	0.09
Butyl vinyl ether	1.25	1.39	0.14
Dibutyl ether	1.17	1.33	0.16
Diethyl ether	1.098	1.33	0.23
Dimethylether	1.3	1.32	0.02
Dipentyl ether	1.2	1.28	0.08
Dipropyl ether	1.21	1.28	0.07
Ethyl methyl ether (*trans*)	1.17	1.37	0.20
Ethyl vinyl ether	1.26	1.38	0.12
Isopropyl methyl ether	1.247	1.33	0.09
Methyl propyl ether (*trans-trans*)	1.107	1.35	0.24
Acetaldehyde	2.75	2.84	0.09
Butanal	2.72	2.84	0.12
2,2-dimethylpropanal	2.66	2.85	0.19
2-methylpropenal	2.68	2.85	0.17
Propanal	2.72	2.83	0.11
2-propynal	2.78	2.81	0.03
Acetone	2.88	2.82	−0.06
2-butanone	2.779	2.83	0.05
2,6-dimethyl-4-heptanone	2.66	2.85	0.19
2,4-dimethyl-3-pentanone	2.74	2.81	0.07
2-heptanone	2.59	2.83	0.24
3-heptanone	2.78	2.84	0.06
2-hexanone	2.66	2.83	0.17
2-pentanone	2.7	2.82	0.12
3-pentanone	2.82	2.83	0.01
Acetic acid	1.7	1.48	−0.22
Butanoic acid	1.65	1.48	−0.17
Formic acid	1.425	1.49	0.06
Hexanoic acid	1.13	1.48	0.35
Methacrylic acid	1.65	1.52	−0.13
Pentanoic acid	1.61	1.50	−0.11
Pentanoic acid (*cis*)	1.46	1.50	0.04

TABLE I. (*Continued.*)

Molecules	Expt.[a]	Calc.	Error
Acetamide	3.68	3.69	0.01
N,*N*-dimethylacetamide	3.7	3.16	−0.54
N,*N*-dimethylformamide	3.82	3.13	−0.69
Formamide	3.73	3.88	0.15
N-methylacetamide	4.3	3.92	−0.38
N-methylformamide	3.83	4.03	0.20
N-methylpropanamide	3.61	3.95	0.34
Acetonitrile	3.93	4.09	0.17
Butanenitrile (*gauche*)	3.91	4.08	0.17
2,2-dimethylpropanenitrile	3.95	4.10	0.15
2-methylpropanenitrile	4.29	4.08	−0.21
Pentanenitrile	4.12	4.09	−0.03
Propanenitrile	4.05	4.09	0.04
1-butanethiol	1.53	1.54	0.01
Ethanethiol (*trans*)	1.58	1.54	−0.04
Methanethiol	1.52	1.54	0.02
2-methyl-2-propanethiol	1.66	1.53	−0.13
1-propanethiol (*trans*)	1.6	1.52	−0.08
2-propanethiol (*gauche*)	1.53	1.53	0.00
Dibutyl sufide	1.61	1.59	−0.02
Diethyl sufide (*trans-gauche*)	1.591	1.59	0.00
Dimethyl sufide	1.554	1.61	0.05
Ethyl methyl sufide (*gauche*)	1.593	1.61	0.02
Ethyl methyl sufide (*trans*)	1.56	1.61	0.05
Dimethyl sulfoxide	3.96	3.77	−0.19
Nitromethane	3.46	3.23	−0.23
1-nitropropane	3.66	4.25	0.59
2-nitropropane	3.73	3.35	−0.38

[a]Experimental dipole moment is obtained in Ref. 40 and 41.

average error for the dipole moment is only 0.01D with a standard deviation of 0.19D and the unsigned error is only 0.14D.

The potential energy surface for the rotation about a bond plays an important role in understanding the conformations of a molecule. Based on the charge distribution calculated from classical physics, the energy surface for rotation about a bond can be calculated. For the rigid rotation, the single point energy is calculated for each step during the rotation. In practice, the molecule is allowed to relax during the rotation that implies the minimization that needs to be carried out with the constrained dihedral angle. The ABNR algorithm discussed above is used for the constrained minimization. The constraints for the dihedral angles are effectively enforced by removing the gradients, which modify the constrained dihedral angle. Figures 8–10 show the calculated energy surface for ethane, butane, and *N*-methylacetamide. The latter is usually used as the smallest peptide model in the validation of protein force fields. In ethane, three global minima occur at the H–C–C–H dihedral angle of 60°, 180°, and 300° corresponding to the staggered conformation. Three global maxima occur at the dihedral angle of 0°, 120°, and 240° corresponding to the eclipsed conformation of ethane. The calculated rotation barrier is 3.22 kcal/mol in good accord with 2.93 kcal/mol estimated from the infrared torsional spectrum.[24] The rotation of butane around the C–C bond shows a global minimum at 180°, which corresponds to the anticonformation. Two local minima occur at about 60° and 300° corresponding to the gauche conformation. Unsurprisingly, the global maximum occurs at 0° and 360° (syn conformation), where the maximum repulsion between two methyl groups occurs. Two local maxima occur at the eclipsed conformation. The conformation energy of gauche conformation is 0.38 kcal/mol compared to the experimental estimates of 0.5–0.9 kcal/mol from different groups.[25–31] The conformation energies of eclipsed and syn conformation are 3.56 kcal/mol and 6.77 kcal/mol, respectively. The experimental

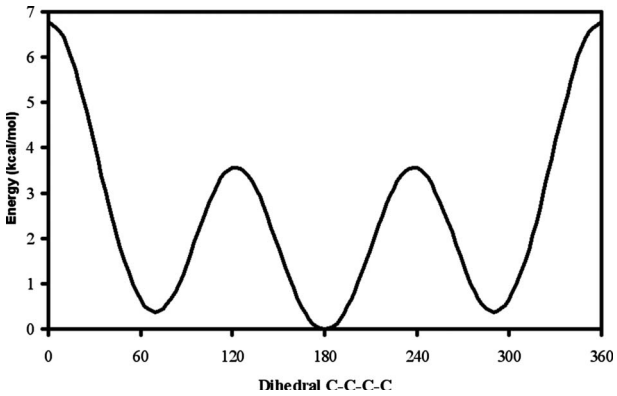

FIG. 9. The conformation energy profile of butane by rotating about the C–C bond in the middle. The molecule is allowed to relax during the rotation.

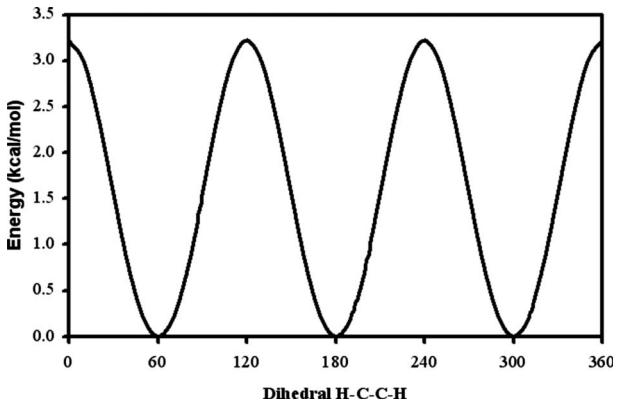

FIG. 8. The conformation energy profile of ethane by rotating about the C–C bond. The molecule is allowed to relax during the rotation.

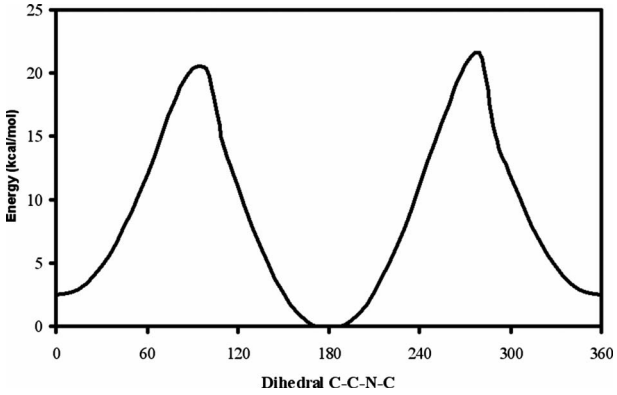

FIG. 10. The conformation energy profile of *N*-methylacetamide by rotating about the peptide bond. The dihedral angle of C–C–N–C is used to scan the conformation. The molecule is allowed to relax during the rotation.

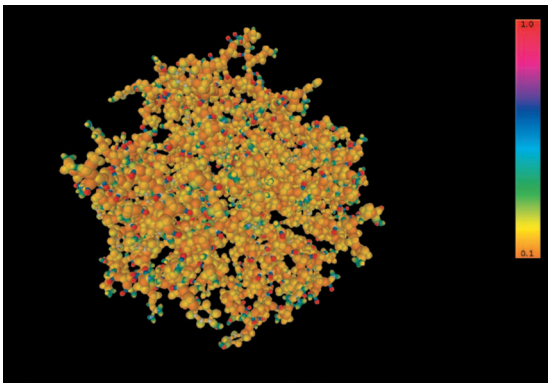

FIG. 11. (Color) Electron density of human insulin hexamer rendered in MILLSIAN 2.0 with electron density scale.

estimates for these two conformations are 3.49–3.70 and 4.37–6.1 kcal/mol.[32–36] The scattered distribution of the experimental observations suggests the considerable uncertainty involved with the experimental data. Figure 10 shows the rotation energy surface for *N*-methylacetamide with a global minimum at 180° corresponding to the *trans*-peptide bond. The *trans*-peptide bond is about 2.48 kcal/mol more favorable compared to the *cis*-peptide bond at the dihedral angles of 0° and 360°, in good accord with the experimental observations of 2.3 and 2.48 kcal/mol.[37–39] The calculated rotation barrier height is 21.47 kcal/mol as compared to 19.8 ± 1.8 kcal/mol from the NMR experiment.[14,39]

Based on the classical physics, MILLSIAN software has an intrinsic advantage over other molecular modeling packages in the requirement of computational power and speed. Electron distributions can be calculated and rendered in a real-time scale for molecules consisting of thousands of atoms such as protein, DNA, and RNA. Figure 11 shows an example of electron distribution for a protein, human insulin hexamer, generated from MILLSIAN 2.0. Remarkably, it was completely rendered in less than 30 s using a slow personal computer.

VI. CONCLUSION

A molecular modeling software called MILLSIAN was developed for the structure, energetics, and charge distributions of molecules of pharmaceutical interest. Based on classical physics, MILLSIAN outperforms other existing molecular modeling software on the prediction of total bond energies for a wide range of molecules. An average error of only 0.1% has been observed for about 700 molecules containing C, H, O, N, P, S, Si, B, F, Cl, Br, I, and metals. The functional groups solved using the classical physics principles span a wide range of molecules that are of special interest in drug discovery.

In MILLSIAN 2.0, the dipole moment was calculated by the vector sum of bond moments, which results directly from the classical physics bond moment solutions due to the localization of electron distributions to the spheroidal MOs of each bond. The tests for a series of organic molecules demonstrated the remarkable accuracy of classical physics in calculating the charge distribution. The unsigned error of dipole moment is about 0.14D, which is in part due to the experimental uncertainty in measuring the dipole moments.

Since the functional groups are shown to be very transferable, the modeling of large complex biomolecules can be achieved by the modeling of the linear combination of the fundamental functional groups solved from classical physics. For example, the charge distribution and heat of formation can be easily calculated for protein and nucleic acids by considering the functional groups. With the electron distribution available, other molecular properties such as the potential energy surfaces can be calculated.

[1]M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Oxford University Press, London, 1987).
[2]R. L. Mills, The Grand Unified Theory of Classical Physics, available at http://blacklightpower.com/theory/book.shtml.
[3]R. L. Mills, Phys. Essays **16**, 433 (2003).
[4]R. L. Mills, Phys. Essays **20**, 403 (2007).
[5]R. L. Mills, Phys. Essays **18**, 321 (2005).
[6]R. L. Mills, Phys. Essays **17**, 342 (2004).
[7]R. L. Mills, Phys. Essays **19**, 225 (2006).
[8]R. L. Mills, Phys. Essays **21**, 103 (2008).
[9]R. L. Mills, Annales de la Fondation Louis de Broglie **30**, 129 (2005).
[10]R. L. Mills, Int. J. Hydrogen Energy **27**, 565 (2002).
[11]R. L. Mills, Int. J. Hydrogen Energy **26**, 1059 (2001).
[12]R. L. Mills, Int. J. Hydrogen Energy **25**, 1171 (2000).
[13]R. L. Mills, *29th Conference on High Energy Physics and Cosmology Since 1964*, Lago Mar Resort, Fort Lauderdale, FL, 14–17 December 2000 (Kluwer Academic, Dordecht/Plenum, New York), pp. 243–258.
[14]A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos,S. Michnick, T. Ngo, D. T.Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[15]W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[16]W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[17]A. D. MacKerell, Jr., J. Comput. Chem. **25**, 1584 (2004).
[18]J. Gasteiger, C. Rudolph, and J. Sadowski, Tetrahedron Comput. Method **3**, 537 (1990).
[19]J. Sadowski, C. Rudolph, and J. Gasteiger, Anal. Chim. Acta **265**, 233 (1992).
[20]R. L. Mills, B. Holverstott, W. Good, N. Hogle, and A. Makwana, Phys. Essays **23**, 153 (2010).
[21]B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[22]J.-W. Chu, B. L. Trout, and B. R. Brooks, J. Chem. Phys. **119**, 12708 (2003).
[23]J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[24]S. Weiss and G. E. Leroi, J. Chem. Phys. **48**, 962 (1968).
[25]S. Tsuzuki, T. Uchimaru, and K. Tanabe, J. Mol. Struct.: THEOCHEM **366**, 89 (1996).
[26]A. L. Verma, W. F. Murphy, and H. J. Bernstein, J. Chem. Phys. **60**, 1540 (1974).
[27]W. S. Bradford, S. Fitzwater, and L. S. Bartell, J. Mol. Struct. **38**, 185 (1977).
[28]J. R. Durig and D. A. C. Compton, J. Phys. Chem. **83**, 265 (1979).
[29]D. A. C. Compton, S. Montrero, and W. F. Murphy, J. Phys. Chem. **84**, 3587 (1980).
[30]R. K. Heenan and L. S. Bartell, J. Chem. Phys. **78**, 1270 (1983).
[31]M. Räsänen and V. E. Bondybey, J. Chem. Phys. **82**, 4718 (1985).
[32]K. S. Pitzer, Chem. Rev. **27**, 39 (1940).
[33]K. S. Pitzer, J. Chem. Phys. **12**, 310 (1944).
[34]K. Ito, J. Am. Chem. Soc. **75**, 2430 (1953).
[35]K. S. Pitzer, J. Chem. Phys. **8**, 711 (1940).
[36]S. S. Chen, R. C. Wilhoit, and B. J. Zwolinski, J. Phys. Chem. Ref. Data **4**, 859 (1975).

[37]A. Radzicka, L. Pedersen, and R. Wolfenden, Biochem. **27**, 4538 (1988).
[38]S. Ataka, H. Takeuchi, and M. Tasumi, J. Mol. Struct. **113**, 147 (1984).
[39]T. Drakenberg and S. Forsen, Chem. Commun. (Cambridge) **15**, 1404 (1971).
[40]D. R. Lide, *CRC Handbook of Chemistry and Physics* (CRC, Boca Raton, FL/Taylor & Francis, New York, 2005).
[41]A. L. McClellan, *Tables of Experimental Dipole Moments* (Rahara Enterprises, El Cerrito, CA, 1963).